A conformation-specific approach to native top-down mass spectrometry

Hannah M. Britt¹, Aisha Ben-Younis¹, Nathanael Page¹,², Konstantinos Thalassinos¹,³*

1 Institute of Structural and Molecular Biology, Division of Biosciences, University College London, London, WC1E 6BT, United Kingdom
2 LGC Group, Teddington, TW11 0LY, United Kingdom
3 Institute of Structural and Molecular Biology, Birkbeck College, University of London, London, WC1E 7HX, United Kingdom
* Contact email: k.thalassinos@ucl.ac.uk

ABSTRACT

Native top-down mass spectrometry is a powerful approach for analyzing proteoforms, but one which largely disregards protein conformation. We report a new ion mobility-enabled method for performing native top-down MS in a conformation-specific manner. Our approach identified conformation-linked differences in backbone dissociation which simultaneously inform upon proteoform variations and provide structural insights. We also illustrate that our method can be applied to protein-ligand complexes, either to identify components or to probe ligand-induced structural changes.

MAIN

True understanding of biological function at the molecular level requires thorough examination of protein variation resulting from differences in genetics, protein structure or post-translational modifications (PTMs).¹ Top-down MS (TDMS) is an invaluable tool for the in-depth characterization of these protein variants, termed proteoforms. In TDMS, ion activation is used to dissociate the covalent bonds making up the protein backbone.² Under the right conditions, TDMS can provide almost complete sequence coverage, and therefore comprehensively characterize PTMs, single nucleotide polymorphisms (SNPs), and splice or truncation variants.

Historically, TDMS has been performed under denaturing conditions, making the protein backbone more accessible for fragmentation and thus maximizing sequence coverage. Recently, however, native top-down mass spectrometry (nTDMS) has come to the fore, in which proteins are introduced into the mass spectrometer using native sample preparation and soft ionization conditions to retain their solution-state characteristics. The major benefit of nTDMS is that protein-protein and protein-ligand interactions remain intact, and therefore proteoforms are characterized within this context.

While protein fragmentation patterns observed in nTDMS have provided exquisite proteoform detail to date, recently there have been attempts to derive structural information from them.³ ⁴ ⁵ One major limitation of this approach, however, is that the instruments used are unable to separate co-existing conformers, such that any fragmentation patterns observed will be a composite of the various conformers. While this is not an issue for primary structural characterization, it becomes problematic when using the fragmentation patterns to report on protein conformation. This inability to link top-down information to individual conformers hampers biological interpretation of the data, especially when structural differences are key to protein function or disease pathogenesis. Proteins with multiple conformations which have different activity are treated as a single species, rather than being studied individually for their unique characteristics. Similarly, in the case of protein misfolding, distinctive features of incorrectly folded proteins would be indistinguishable from the correctly folded analog.

Here we present a new native ion mobility top-down mass spectrometry (nIM-TDMS) approach that overcomes these limitations, and which can provide fragmentation patterns which correlate to specific protein conformational states. Our approach
was implemented on a cyclic ion mobility (cIM)-enabled quadrupole time-of-flight (ToF) instrument with post-mobility electron capture dissociation (ECD) capabilities, shown in Figure 1a. The geometry of this instrument allows nIM-TDMS of proteins and their complexes through three workflows, detailed in Figure 1b.

We applied nIM-TDMS to study the protein calmodulin, which undergoes conformational changes upon binding to calcium and small peptides. The protein was electrosprayed from native-like conditions, and the quadrupole used to isolate the 10+ charge state (Figure 2a), which exhibits a mixed calcium-binding occupancy (Supplementary Figure 1) reflecting the protein's state within its native environment. Ion mobility separation of this charge state revealed two conformational families, as demonstrated by the arrival time distributions (ATD) in Figure 2b. Using the tandem cIM capabilities, each conformational family was then isolated from the rest of the population, and subjected to ECD. The two families exhibited different dissociation patterns, with the extended one showing considerably more fragmentation (c, z, and y ions) at 67% inter-residue cleavage compared to 27% for the compact one (Figure 2c). The fragmentation of the extended population was spread over a greater proportion of the sequence, while the compact structure's dissociation was predominantly at the protein termini. In comparison, performing a similar experiment without conformer selection (Supplementary Figure 2) resulted in a fragmentation pattern corresponding to 63% inter-residue cleavage, with overlapping fragments from both families. This highlights the necessity of using a conformation-specific approach to correctly attribute fragmentation findings to protein structure.

By performing top-down fragmentation on isolated conformational families, we were able to directly link the fragmentation patterns to structural elements of the protein. We highlight the bonds that fragmented in the PDB structures representative of compact and extended forms of calmodulin (Figure 1c). The central region of the protein, specifically residues 60-114, had a greater number of fragmentation sites in the extended conformer than the compact form, particularly z and y ions. This finding suggests that the additional fragmentation can be directly linked to the increased solvent accessibility of the central helix of calmodulin in the extended structure. Interestingly, within this protein region is a known calcium-binding motif between residues 93-100. Only the extended conformer showed c and z fragmentation in this region, suggesting that this site is used for calcium binding within the compact structure but not in the extended form. This hypothesis fits with both the PDB structure and the intact mass measurement (Supplementary Figure 3), which suggests less than 4 calcium occupancy for this state, compared to complete occupancy for the compact conformer. Three other known calcium-binding sites showed similar fragmentation between the two structures, consistent with comparable occupation and location outside the solvent-accessible central region. These sites did show slightly reduced fragmentation compared to surrounding regions, relative to their proximity to the protein termini where high levels of bond dissociation are expected, supporting the idea that within these regions the backbone would be protected by calcium binding. We found no strong correlation between either amount of fragmentation or ion type with the specific higher-order protein structures. Instead, we attribute any differences to be predominantly driven by solvent accessibility and non-covalent interactions.

Binding of ligands often plays a critical role in modulating a protein’s function. We, therefore, developed two additional workflows, shown in Figure 1b, which enable in-depth interrogation of such species. The first workflow, conformer-specific top-down of a protein-ligand complex, is comparable to the single-protein approach, but with ECD performed directly on the intact complex. Results of this experiment are demonstrated for 1:1 (protein:ligand) complexes of calmodulin with two ligands, melittin and calmodulin-dependent protein kinase II (CaMKII) fragment 290-309 (Figure 2g). As the protein:ligand complexes occupy a lower charge state distribution than the free protein, the species selected were 8+ rather than 10+, which had a noticeable effect on the fragmentation efficiency, as is expected with electron-based dissociation techniques. It is still possible, however, to identify regions of the protein backbone which show reduced fragmentation in the presence of ligands, for example the N-terminal residues 13-46, residues 106-114, and some C-terminal residues from 130-146. In the canonical calmodulin-ligand structure (PDB 1C1M) these regions show close proximity or H-bonding to the peptide binding pocket, suggesting solvent-accessibility reductions and non-covalent interactions drive the reduced fragmentation. A small difference in the dissociation pattern for the two ligand bound species is observed between residue 135-138, which could be attributed to the reduced length of CaMKII peptide (19 aa) compared to melittin (26 aa) altering the solvent accessibility in this region.

In the last nIM-TDMS workflow, it is possible to characterize endogenous ligands and protein-variants within protein-ligand complexes, in an approach similar to that of the native-omics workflow first described by Gault and colleagues. Following m/z and mobility isolation of a calmodulin-melittin complex (Figure 2d) the complex is collisionally dissociated upon injection into the mobility cell, causing the complex to break down into its component parts. Unlike the native-omics approach, in which multi-stage MS experiments are performed based upon a m/z isolation, our workflow instead performs species selection based upon conformation, Figure 2e. This mobility isolation of specific conformational species works particularly well for ligands such as peptides, organic molecules, and lipids, which exhibit a significantly different arrival time compared to
proteins. After mobility selection, the ligand was subjected to further fragmentation (ECD or CID) for characterization, enabling successful identification of melittin from the calmodulin-ligand complex (Figure 2f). We foresee the approach being similarly applicable for identifying unknown endogenous ligands.

Using nIM-TDMS we have demonstrated the usefulness of a conformation-specific approach for nTDMS analysis of proteins and their complexes. Instead of sampling across the entire conformational landscape, we can selectively isolate specific protein conformers, which enables us to link the intrinsic structural and proteoform information obtained from nTDMS experiments to specific protein conformations. Our conformation-specific approach can also provide structural insights and component characterization for protein complexes all within the same experiment, making it invaluable in enabling a true biological understanding of systems where proteins exhibit multiple structures, and transient or low abundant structural intermediates. We believe that nIM-TDMS has great potential in the structural characterization of traditionally challenging families of proteins and complexes, such as intrinsically disordered proteins, glycoproteins, membrane proteins, or proteins that are prone to misfolding and aggregation. nIM-TDMS can also provide amino acid level, conformation specific data for integrative structural biology workflows, supplying information which is currently underused due to difficulties in obtaining it by classic structural or biophysical methods, but which is required to provide a truly accurate picture of the systems studied.10

REFERENCES


METHODS

Sample Preparation
Calmodulin from bovine testes was purchased as a lyophilised powder (Sigma, UK) and stored in aliquots at -20°C as a 40 µM stock solution in 10 mM ammonium acetate. For analysis, stock solution was diluted in 10 mM ammonium acetate to a final concentration of 5 µM. Melittin from honey bee venom (Sigma, UK) and Calmodulin-dependent Protein Kinase II fragment 290-309 (Sigma, UK) were stored as stock solutions at 1 mM in 10 mM ammonium acetate at -20°C. For all ligand binding experiments the peptides were added to calmodulin at a final concentration of 7.5 µM, giving a 1:1.5 protein:ligand ratio.

Mass Spectrometry
Mass spectrometry experiments were performed on a SELECT SERIES Cyclic IMS QToF (Waters Corp., UK) fitted with a post-mobility ECD modification (e-MSion, US). The instrument was operated in sensitivity mode, and calibrated using sodium iodide (NaI) to within 1 ppm over the m/z range 0-8000. Samples were infused into the instrument using nano electrospray (nESI) capillaries prepared in house using a Flaming-Brown P97 micropipette puller, and gold-coated with a Quorum Q150RS sputter coater. Parameters used for analyses were: Capillary Voltage 1.4 kV; Sampling Cone 40 V; Trap Collision Energy 8 V; Transfer Collision Energy 6 V.

Ion Mobility
Ion mobility experiments were also performed on the ECD-modified SELECT SERIES Cyclic IMS QToF (Waters Corp., UK) using the MS parameters detailed above. For these experiments the instrument was operated in mobility mode with parameters: Automatic ADC; 1 Push Per Bin; Racetrack Bias 70 V; Twave Height 28 V; Twave Velocity 375 ms⁻¹. For classic native top-down experiments in which no conformer selection was required, the following cyclic control sequence was used: Inject – Separate for 5 ms – Eject Ions and Acquire. For conformer selection experiments the cyclic sequence used was as follows: Inject – Separate for 5 ms – Eject Ions – Eject to Pre-Store – Eject Ions – Reinject from Pre-Store – Separate for 5 ms – Eject Ions and Acquire. Timings for the Eject Ions and Eject to Pre-Store steps were determined for specific conformational families based upon the observed arrival time distribution, those being Eject 3.5 ms - Eject to Pre-Store 2 ms - Eject Ions 22.5 ms for the compact conformer of calmodulin, and Eject Ions 8 ms - Eject to Pre-Store 2 ms – Eject Ions 18 ms for the extended analogue. Further information regarding the geometry of the SELECT SERIES Cyclic IMS QToF (Waters Corp., UK) and the preparation of cyclic control sequences for protein analysis are described in detail elsewhere.

Electron Capture Dissociation
Electron capture dissociation was performed at the post-mobility ECD cell (e-MSion, US), with the voltages across the block (L1; L2; LM3; L4; FB; LM5; L6; L7) tuned using reported e and z ions for calmodulin in order to give the optimum dissociation whilst retaining ion transmission. The values used were L1 1.0; L2 -30.0; LM3 10.0; L4 12.0; FB 3.4; LM5 8.5; L6 -30.0; L7 1.0. The Transfer Collision Energy was increased to 16 V, providing supplemental activation to reduce non-specific fragment binding. For calmodulin-peptide complexes the FB was dropped to 2.9 but all other parameters remained the same.

Data Processing
Native mass spectra and IM data were processed in MassLynx v4.2 (Waters Corp.), and deconvoluted using UniDec v4.3.0. ECD data were processed using UNIFI (Waters Corp., UK). Raw spectra over the m/z range 200-8000 were deconvoluted for 30 iterations using the BayesSpray module operated in intact protein mode with a ToF resolution of 25,000. The resulting [M+H]+ ion masses were then matched within 20 ppm error to theoretically generated BY and CZ* fragment ions for bovine calmodulin (UniProt P62157) using ProSight Lite. Reported PTMs for bovine calmodulin, N-terminal methionine loss, N-terminal acetylation, and K115 tri-methylation, were also applied to the sequence during matching.

REFERENCES


**ACKNOWLEDGEMENTS**

H.M.B. is supported by the Wellcome Trust (209250/Z/17/Z). Mass spectrometry instrumentation in K.T.’s lab has been supported by Wellcome Trust grants 221521/Z/20/Z and 104913/Z/14/Z. The authors would like to thank colleagues at Waters Corporation and e-MSion, who developed the ECD-modified cyclic ion mobility mass spectrometer, for their valuable insight and support. Particular thanks to Darren Hewitt, Jason Wildgoose, Dale Cooper-Shepherd, and Jim Langridge for their advice in setup and optimization of instrument parameters.

**COMPETING INTERESTS**

The authors declare no competing interests.
SUPPLEMENTARY INFORMATION

A supporting document is provided with the manuscript, which contains all Supplementary Figures detailed in the text, and Supplementary Tables detailing the matched fragment ions with error tolerances for each dataset discussed in the main text.
Figure 1. (a) Schematic of the cIM quadrupole time-of-flight instrument used to implement the nIM-TDMS approach. The instrument is based on the commercial cIM platform, but has an additional post-mobility ECD modification. (b) Depending on whether a protein or protein-ligand complex is analysed, different workflows can be followed. For a protein with more than one conformation, tandem ion mobility can be used to isolate each conformer for ECD fragmentation, thus obtaining conformer-specific top-down data (blue line). For protein-ligand complexes, two workflows exist. ECD fragmentation of the complex, which can reveal structural changes due to ligand binding (orange line), or dissociation of the complex followed by isolation of each component using ion mobility, followed by ECD or CID fragmentation (red line). This workflow allows the proteoform characterisation as well as identification of the bound ligand.
Figure 2. (a-c) Analysis of Calmodulin using nIM-TDMS. From the native mass spectrum of Calmodulin (a), the +10 charge state was selected for subsequent analysis. At least two distinct conformers can be observed in the arrival time distribution (ATD) for this charge state (b), each of which were isolated for ECD fragmentation, here shown as the blue and yellow ATDs between 60 and 70 msec. ECD fragmentation data is displayed against the calmodulin sequence in circular plot (c), with coloured rectangles indicating different secondary structural elements (alpha helices – orange, beta strands – red, EF-hand – blue, Ca binding – cyan). The three main ion fragment series (c, z, y ions) are shown as concentric circle slices with different colours. The outermost circle shows the c fragments for the compact and extended conformers, followed by the z ions for the compact and extended conformers, and the innermost circle shows the y ions for the compact and extended conformers. Each fragment is represented as a coloured dot, with the colour indicating the normalised fragment ion intensity. In the centre of the circle, the PDB structures representing the compact (1PRW) and extended conformers (3CLN) are shown, with fragmentation sites mapped in purple. (d-g) Analysis of Calmodulin-ligand complexes using nIM-TDMS. From the Calmodulin: melittin (1:1) complex native mass spectrum (d), the +8 charge state was quadrupole-isolated and the complex dissociated by increasing the injection energy upon introduction to the mobility region. The ligand (orange) was then mobility selected (e) and subjected to ECD fragmentation, allowing identification of melittin through fragment annotation (f). Results from the alternative protein-ligand workflow are presented for apo, melittin, and calmodulin-dependent protein kinase II (CaMKII) fragment 290-309 bound fragments in a circular plot (g). The PDB structure for the canonical calmodulin-ligand complex (1C1M) is shown in the centre of the circle along with the fragment sites (purple) obtained from both the apo and ligand-bound experiments. It should be noted that the apo form in panel g contains fewer fragments than the corresponding data in panel c, as a lower charge state (+8) was used here compared to panel c (+10).
Native MS

Dissociation

Mobility separation

Mobility separation

Tandem ion mobility

ECD

Conformer-specific Top-Down

ECD and/or CID

Proteoform characterisation and ligand identification

Protein-ligand complex