Molecular basis for structural heterogeneity of an intrinsically disordered protein bound to a partner by combined ESI-IM-MS and modeling

Reference:
D' Urzo Annalisa, Konijnenberg Albert, Rossetti Giulia, Habchi Johnny, Li Jinyu, Carloni Paolo, Sobott Frank, Longhi Sonia, Grandori Rita.- Molecular basis for structural heterogeneity of an intrinsically disordered protein bound to a partner by combined ESI-IM-MS and modeling
Journal of the American Society for Mass Spectrometry / American Society for Mass Spectrometry - ISSN 1044-0305 - 26:3(2015), p. 472-481
Full text (Publishers DOI): http://dx.doi.org/doi:10.1007/s13361-014-1048-z
To cite this reference: http://hdl.handle.net/10067/1243630151162165141

Molecular basis for structural heterogeneity of an intrinsically disordered protein bound to a partner by combined ESI-IM-MS and modeling

Annalisa D'Urzo[1], Albert Konijnenberg[2], Giulia Rossetti[3,4], Johnny Habchi[5,6], Jinyu Li[3,7], Paolo Carloni[3], Frank Sobott[2,8], Sonia Longhi[5,6]* and Rita Grandori[1]*

[1]*Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126 Milan, Italy*

[2]*Biomolecular & Analytical Mass Spectrometry group, Department of Chemistry, University of Antwerp, 2020 Antwerpen, Belgium*

[3]*Computational Biophysics, German Research School for Simulation Sciences, and Computational Biomedicine, Institute for Advanced Simulation IAS-5 and Institute of Neuroscience and Medicine INM-9, Forschungszentrum Jülich, 52425 Jülich, Germany*

[4]*Jülich Supercomputing Centre, Forschungszentrum Jülich, 52425 Jülich, Germany*

[5]*Aix-Marseille Université, Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257, 13288, Marseille, France*

[6]*CNRS, AFMB UMR 7257, 13288, Marseille, France*

[7]*Institute of Biochemistry and Molecular Biology, RWTH Aachen University, 52057 Aachen, Germany*

[8]*Center for Proteomics (CFP-CeProMa), University of Antwerp, 2020 Antwerpen, Belgium*

* Address reprint requests to:

Rita Grandori. Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126 Milan, Italy. E-mail: rita.grandori@unimib.it

Sonia Longhi. Aix-Marseille Université, Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257, 13288, Marseille, France.
CNRS, AFMB UMR 7257, 13288, Marseille, France
E-mail: sonia.longhi@afmb.univ-mrs.fr

Submitted to J. Am. Soc. Mass Spec. on July 17th, 2014

Abstract

Intrinsically disordered proteins (IDPs) form biologically active complexes that can retain a high degree of conformational disorder, escaping structural characterization by conventional approaches. An example is offered by the complex between the intrinsically disordered N_{TAIL} domain and the phosphoprotein X domain (P_{XD}) from measles virus (MeV). Here we detect distinct conformers of the complex by electrospray ionization-mass spectrometry (ESI-MS) and ion mobility (IM) techniques yielding estimates for the solvent-accessible surface area (SASA) in solution and the average collision cross section (CCS) in the gas phase. Computational modeling of the complex in solution, based on experimental constraints, provides atomic-resolution structural models featuring different levels of compactness. The resulting models indicate high structural heterogeneity. The intermolecular interactions are predominantly hydrophobic, not only in the ordered core of the complex, but also in the dynamic, disordered regions. Electrostatic interactions become involved in the more compact states. This system represents an illustrative example of a hydrophobic complex that could be directly detected in the gas phase by native mass spectrometry. This work represents the first attempt to modeling the entire N_{TAIL} domain bound to P_{XD} at atomic resolution.

Introduction

The last decade has witnessed an extension to the protein structure-function paradigm, with the progressive understanding of the functional importance of intrinsically disordered proteins

(IDPs) or regions (IDRs). These proteins or protein regions lack ordered secondary and tertiary structure under physiological conditions and exist in solution as dynamic and heterogeneous conformational ensembles [1-3]. Approximately 40% of the human proteins are predicted to contain at least one disordered segment of at least 30 amino acids, with as many as 25% of them being likely disordered from start to end [4]. Predictions on representative genomes from the three kingdoms of life (i.e. bacteria, archaea and eukaryotes) confirmed the ubiquitous character of structural disorder, in spite of significant differences in its relative amount in the three domains [5]. The extent of protein structural disorder tends to increase with biological complexity. This trend could be related to the typical involvement of IDPs in signalling and regulation [6-8].

The structural plasticity of IDPs allows for recognition and binding of multiple partners, resulting in pleiotropic roles of these proteins. Many cases have been described in the literature, in which IDPs acquire ordered conformations upon binding to partners or ligands. Folding coupled to binding can pertain either to specific segments or to the whole protein mettere qui [14]. Complete folding can lead to well-structured complexes that can be analysed by conventional techniques, such as X-ray crystallography [9-11]. However, increasing evidence shows that many IDPs retain a high degree of structural disorder even in the bound state. These "fuzzy" complexes [12] are stabilized by short, ordered recognition elements, referred to as molecular recognition elements (MoREs), and a large number of highly unstable contacts, leading to a cloud of interconverting conformations around a structured core [9, 12-16]. This staccato-type of interactions is much more difficult to characterize than stable interactions of folded complexes. Yet, it is thought to be relevant for biological function. Fuzzy regions within complexes can harbour regulatory post-translational modification sites, or can mediate interactions with additional partners. They can even directly or indirectly interfere with recognition elements, promoting or inhibiting binding. In addition, fuzziness provides a way to reduce entropic penalty, thereby affording enhanced affinity [15, 17].

Hence, description of the conformational ensembles characterizing fuzzy IDP complexes is essential to the understanding of the molecular processes by which IDPs establish their functional networks. However, the highly heterogeneous nature of IDPs and the fuzziness that is often observed in their bound state makes their structural characterization very challenging. Such investigation demands the combined application of various biophysical methods capable of capturing conformational heterogeneity and identifying metastable states.

In this regard, native mass spectrometry based on nano-ESI sources has emerged as a powerful approach, allowing detection of coexisting conformers with distinct global compactness [18][Kaltashov 2008]. The average charge state of each component yields an estimate of SASA for the structure in solution, at the moment of transfer to the gas phase [19, 20]. Hyphenation with IM measurements adds a further dimension to species separation and offers estimates of the CCS for each detected structure in the gas phase [21-24]. These techniques conjugate the exceptional analytical power of mass spectrometry with structural description and, therefore, are particularly well suited to the challenges posed by conformational studies on IDPs [Chen 2014; Saikusa 2013; Beveridge 2013]. The development of atomic-resolution models requires, however, complementation of experimental data by computational simulations. Theoretical approaches have been improved to effectively model disordered conformational ensembles based on structural information derived from native mass spectrometry and other biophysical analyses [25-27].

We herein combine ESI-MS, ESI-IM-MS and computational methods to describe the N_{TAIL}-P_{XD} complex from measles virus (MeV). While N_{TAIL} is an intrinsically disordered domain of 125 residues [28-31], P_{XD} is folded into an ordered three-helix bundle [32]. Interaction between N_{TAIL} and P_{XD} is crucial for MeV replication, as it allows recruitment of the viral RNA-dependent RNA polymerase onto the nucleocapsid template (Longhi, Curr Topics Microbiol Immunol 2009; Habchi and Longhi, Mol Biosysts 2012). Upon binding to P_{XD}, N_{TAIL} undergoes induced folding localized in an 18-residue α-helical region. This α-helical molecular recognition element (α-MoRE) [30, 31, 33, 34] associates to P_{XD} forming a four-helix bundle [32, 35]. The N_{TAIL}-P_{XD} interaction mostly relies on hydrophobic contacts [30, 35, 36] and is characterized by a K_D in the sub-micromolar range, as consistently indicated by surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC) experiments [33, 37, 38]. Previous studies on the N_{TAIL}-P_{XD} complex by small-angle X-ray scattering (SAXS) and NMR showed that the majority of N_{TAIL} remains disordered in the bound state, thus supporting the fuzzy and heterogeneous nature of this complex [30, 33, 39]. Although hydrophobic complexes could be expected to dissociate in the gas phase, detection of such complexes by MS has been reported in some instances [40-42]. The results of this study show that the MeV N_{TAIL}-P_{XD} complex is amenable to MS analysis and that distinct conformational states can be detected. Experimental data are used to validate structural modeling. The results highlight the conformational freedom of the complex and the hydrophobic nature of its interface.

Materials and methods

Expression and purification of MeV N_{TAIL} and P_{XD} proteins

The expression constructs in the pDEST14 vector, allowing expression of N-terminally (N_{TAIL}) or C-terminally (P_{XD}) hexahistidine tagged forms of the MeV proteins under the control of the T7 promoter, have already been described [33]. Expression and purification of MeV N_{TAIL} and P_{XD} were carried out as described [33], except that the final gel filtration step was carried out using 10 mM ammonium acetate pH 6.5 as elution buffer.

Mass spectrometry

Nano-ESI-MS analyses were performed on a hybrid quadrupole time of flight (Q-TOF) mass spectrometer (QSTAR Elite; ABsciex, Framingham, MA) equipped with a nano-electrospray ionization sample source. Metal-coated borosilicate capillaries (Proxeon, Odense, Denmark), with medium-length emitter tips of 1-µm internal diameter, were used to infuse the samples. The instrument was calibrated by the standard Renin-inhibitor solution (ABSciex, Framingham, MA) on the intact molecular ion $[M+2H]^{2+}$ (879.97 Da) and its fragment $[M+H]^{1+}$ (110.07 Da). Data were acquired with ion spray voltage 1,200 V and declustering potential 80 V, and were averaged over 2-min acquisitions. The interface was kept at room temperature (interface heater off). Pure preparations of the N_{TAIL} and P_{XD} protein fragments in 10 mM ammonium acetate pH 6.5 were stored at -20 °C and diluted to the indicated final concentrations in the same buffer before analyses. Samples were incubated at room temperature for 10 minutes before measurements.

Mass spectra from Figure 1A and Figure 2A were transformed from x = m/z to x = z as abscissa axis and data point were fitted by Gaussian functions. The analysis was performed by the software Origin7 (Originlab, Northampton,MA,USA) [43, 44].

Nano-ESI-IM-MS analyses were performed on a Synapt G2 HDMS instrument (Waters, Milford, MA, USA) by direct infusion, under non-denaturing conditions. The samples were injected at the indicated protein concentrations in 10 mM ammonium acetate pH 6.5 at room temperature. Instrument parameters were carefully optimized to minimize disruption of non-covalent interactions. The following instrumental settings were applied: spray voltage 1,400 V, sampling cone voltage 50 V, trap collision energy 4 V, transfer collision energy 0 V, trap DC bias 45 V, backing pressure 2.98 mbar, trap pressure 5×10^{-2} mbar, IMS pressure 3.09 mbar. Instrument calibration for CCS measurements was performed using concanavalin A, albumin, β-lactoglobulin and cytochrome c as standards while referring to the absolute CCS values determined on a modified Synapt G1 as described by Bush et al. [45]

Structural modeling. Homology modeling was performed to build an initial structure of the N_{TAIL}-P_{XD} complex for molecular simulations. The folded core was modeled on the crystallographic structure of the chimeric protein (PDB 1T6O) [35]. The missing regions (residues 401-485 and 505-525 of N_{TAIL}) were split into fragments and used to screen the structure database searching for suitable templates, following the approach by Bowie et al. [46, 47]. Ten templates (Figure S1), showing from a minimum of 28% sequence identity over 41 residues to a maximum of 56% over 17 residues to the target, were identified, reaching a cumulative 96.5% sequence coverage of the N_{TAIL} disordered arms. These homology-derived restraints were used to generate 2000 models of the N_{TAIL}-P_{XD} complex by the MODELLER 9v9 package [48]. Most (94.5%) of the resulting structures have favourable backbone conformations [49] for more than 80% of the residues, according to the Procheck program [50]. The lowest-DOPE (discrete optimized protein energy) score [51] model was selected as the starting structure for the following molecular simulations.

Replica-exchange Monte Carlo (REMC) simulations were carried out by the PROFASI package (PROtein Folding and Aggregation SImulator) [52-54], with 8 replicas performed at distinct temperatures between 298K and 348K. The energy distributions at these temperatures (Figure S2) show optimal overlap between consecutive temperatures of the explored range (Figure S2). Such an observation is relevant in order to guarantee a sufficient number of replica exchanges between neighbouring replicas, thus allowing an efficient sampling of the conformational space of the protein [55]. Dihedral and distance restraints [56] were applied to the regions represented in the crystallographic structure [35]. A total of 5.0×10^6 cycles of simulations were performed, according to a previously described protocol [25]. Models obtained at 298K were extracted, constituting an ensemble of 2831 conformations.

The chemical shifts (CSs) of the N_{TAIL} moiety for each conformation of the ensemble were calculated using the SHIFTX package [57] (Figure S3). The values of solvent accessible surface area (SASA) were calculated by the tools implemented in GROMACS 4.5.5 [58]. The structures showing the best correlation with either experimental CSs or experimental SASA were identified. The circular dichroism (CD) spectra for each conformation were calculated by the DichroCalc webserver [59], using the parameters derived from both *ab initio* [60] and semi-empirical calculations [61]. All the figures for structures visualization were drawn using PyMOL (Molecular Graphics System, Version 1.3, Schrödinger LLC).

Results and discussion

N_{TAIL}. The MeV $\mathrm{N_{TAIL}}$ protein was analyzed by nano-ESI-MS under non-denaturing conditions (Figure 1A). The spectrum shows a very broad, bimodal charge-state distribution (CSD), indicating structural heterogeneity consistent with the intrinsically disordered nature of this protein. Nonetheless, a compact conformation can be identified in the high m/z region of the spectrum, centered on the 9+ charge state. Such an extent of ionization approaches the behavior of normally folded proteins of the same size. The rest of the spectrum is characterized by a broad peak envelope that spans the region between the 10+ and the 24+ ions, with main charge state 16+. This component corresponds to disordered conformations, characterized by low compactness. Deconvolution by Gaussian fitting is shown in Figure 1B. The results indicate that, in addition to the compact and disordered conformations, a third component of intermediate compactness can be identified. Although the relative amounts of the distinct components depend somewhat on the type of instrument and the parameter settings (as generally true), the intermediate species was detected with good reproducibility under the conditions described here.

The compact and the intermediate conformations are depleted under acidic conditions (Figure 1C) and disappear upon further addition of 50% acetonitrile (Figure 1D), while the disordered component slightly shifts to higher charge states. This response to denaturing conditions suggests that the N_{TAIL} protein populates collapsed and partially folded states that can be destabilized by acids and organic solvents. The isotopically averaged molecular weight by mass deconvolution (data not shown) yields a value of 14,630.94 ± 0.17 Da, in close agreement with the value calculated on the basis of the amino acid sequence, including the initial methionine (14631.7 Da).

P_{XD}. Figure 2 shows the nano-ESI-MS spectra of P_{XD} preparations. Although P_{XD} folds into a helical domain, it displays a bimodal CSD under non-denaturing conditions, with main charge states 6+ and 8+ (Figure 2A). Gaussian fitting of the two components is reported in Figure 2B. This result suggests that the three-helix bundle of the isolated P_{XD} fragment tends to open to a less compact structure, at least under the here employed conditions. The 8+ component does not represent the fully denatured state, since it converts progressively into a more highly charged component (9+) by the addition of formic acid and acetonitrile (Figures 3C and 3D). The experimental mass (6,554.28 ± 0.04 Da) corresponds to the value calculated from the amino acid sequence of the protein without the initial methionine (6,555.66 Da). The conformational heterogeneity adopted by MeV P_{XD} is in agreement with previous biophysical and structural studies on the homologous phosphoprotein X domains from members of the closely related *Rubulavirus* genus, indicating that these domains span a structural continuum, ranging from compact to largely disordered states in solution [62, 63].

N_{TAIL}-P_{XD}. The interaction between N_{TAIL} and P_{XD} was investigated by mixing equimolar amounts of the two proteins and infusing the sample after 10-min incubation at room temperature without agitation. The results obtained with concentrations of either 10 μM or 20 μM of both proteins are shown in Figure 3. The spectra are dominated by the peaks of the free proteins, but signals specific to the N_{TAIL}-P_{XD} complex become evident at the higher protein concentration. The measured mass of the complex (21187.4 ± 0.61 Da) corresponds closely to the sum of the theoretical values expected for N_{TAIL} with the initial methionine and P_{XD} without the initial methionine (21,187.46 Da), indicating a 1:1 stoichiometry.

The complex disappears upon applying denaturing solvent conditions or high declustering potentials (data not shown). These results show that the MeV N_{TAIL}-P_{XD} complex, which is stabilized in solution by hydrophobic interactions, is preserved, at least partially, during the electrospray process and gas-phase ion separation, allowing detection by ESI-MS in a concentration-dependent manner. The expected fraction of bound protein at the concentrations shown in Figure 3B, based on solution affinity data [33, 37, 38], is 93%. Although relative amounts of different species cannot be quantified by ESI-MS, and although the buffer conditions are not identical to previous solution studies, these results suggest that significant dissociation of the complex occurs under electrospray conditions.

The CSD of the complex-specific peaks is quite broad and, itself, bimodal, suggesting coexistence of at least two distinct conformational states of the bound species. This conformational heterogeneity likely reflects the persistence of a conspicuous degree of disorder in the complex [30, 31, 33]. The high-charge component of the complex (18+) would then correspond to an "open" conformation, in which the disordered arms of N_{TAIL}, upstream and downstream to the α-MoRE, fluctuate maintaining high solvent accessibility. The low-charge component (11+), instead, likely represents a compact or "closed" conformation of the complex, in which the N_{TAIL} arms collapse onto the surface of the folded partner, as suggested by the low extent of ionization, which approaches the expected value for folded globular proteins of the same mass (9.46+) [19]. The average SASA value of the "open" and "closed" states can be calculated from the empiric relation with the average charge state [19], resulting in estimated values of 16100 Å^2 and 9500 Å^2, respectively.

It is worth noting that this dual nature of the N_{TAIL}-P_{XD} complex did not emerge from previous experimental studies Habchi and Longhi, Mol Biosysts 2012. Thus, the question arises as to whether this result reflects a real conformational heterogeneity in solution, captured only by the exquisite sensitivity of native MS, or an altered conformational ensemble induced by the electrospray conditions. To this regard, it should be noted that CSDs generally reflect structural compactness and conformational heterogeneity in solution, before or during the transfer to the gas phase. Thus, hypothetical conformational rearrangements leading to the low-charge component in the CSDs should take place at the level of ESI droplets. The fact that CSDs of IDPs detected by nano-ESI-MS and their response to solvent conditions are protein specific, as for instance indicated by opposite pH dependence and distinct effects of organic solvents, [19, 64-66] Lambrughi Sic1 suggests that they reflect real conformational properties rather than artifacts induced by the experimental conditions. However, it could not be ruled out [14] spostare in intro that poorer kinetic trapping of disordered conformations compared to folded structures expose IDPs to more significant conformational changes inside ESI droplets, affecting in turn

CSDs. Even more relevant structural rearrangements could be expected to take place in the gas phase, after electrospray and downstream to protonation reactions. Such structural rearrangements would not affect the CSDs but could be captured by ion-mobility measurements. (Rebecca Beveridge, Kamila Parcholarz, Jason Kalapothakis, Cait MacPhee, Perdita Barran, 2014. "The Use of Mass Spectrometry to Determine the Disordered Content of Proteins". The Biophysical Society Meeting on Disordered Motifs and Domains in Cell Control. Dublin, Ireland).

In order to get further insight on the compact state of the complex, we further investigated the conformational space of the complex by ion mobility on a Waters Synapt G2 HDMS instrument under non-denaturing conditions comparable to the MS measurements reported above. The IM-MS spectrum obtained with a mixture of 20 μM N_{TAIL} and 20 μM P_{XD} is reported in Figure 4. The CSDs are similar to those observed on the QSTAR Elite instrument, although with higher relative amounts of the compact form. The additional separation of analytes by IM allows detection of distinct structural species even at the same charge state. In this case, two slightly different conformations are found to populate the compact state of the N_{TAIL}-P_{XD} complex (charge states 8+ to 10+). Average CCS of 1326 and 1422 Å^2 can be derived for the two main peaks detectable for the 8+ charge state. This result suggests that the compact state of the complex is characterized by further structural heterogeneity. Thus, while its overall ionization suggests a collapsed structure, its arrival-time distribution reveals distinct peaks rather than the single peak typically observed for normally folded proteins. The presence of multiple conformers, although with different CCS values, is observed also for the 9+ and 10+ ions. The predominant conformer displays progressively larger CCS values (1422, 1466, 1607 Å^2) as the charge state increases from 8+ to 10+. This trend is consistent with the lowest charge state (8+ in this case) corresponding to the most compact species in the original ensemble and to the species with lowest interference of Coulomb repulsions upon transfer to the gas phase. Interestingly, at higher charge states (figure 4C), the structurally heterogenous nature of the complex is maintained, whereas for a fully disordered protein such as alpha-synuclein this converges into a single arrival time distribution for the higher charge states [ref]. Altogether, these results suggest that both the open and closed states of the complex should be described as conformational ensembles. The high variability of CCS values revealed by the IM profiles could represent a distinct feature of IDPs compared to normally folded, globular structures. Further studies will be needed to clarify to which extent this variability reflects solution and/or gas-phase properties.

Modeling the N_{TAIL}-P_{XD} *complex.* The structural models of the N_{TAIL}-P_{XD} complex were generated by combining homology modeling and an all-atom Monte Carlo based simulation tool, PROFASI [52-54], which allows exploring the conformational ensemble of the complex in implicit solvent (see Methods for details). The quality of the conformational sampling was established by the agreement with available experimental data, namely NMR CSs of N_{TAIL} bound to P_{XD} [30] and CD of the N_{TAIL}-P_{XD} complex [37]. Quite good agreement can be observed between calculated CSs and those measured by solution NMR measurements [30] (Figure S3).

The calculated CD spectra of the conformational ensemble are similarly close to the experimental ones [37], reproducing the minimum at 208 nm, a shoulder at 222 nm and a positive maximum at 192 nm for the complex, although a well-defined local minimum at 222 nm is missing (Figure S4). Finally, the SASA mean value calculated from the models is 17920 ± 690 Å^2, indicating that the conformational ensemble well describes the open state of the complex detected by ESI-MS.

Three models, namely Model 1, Model 2 and Model 3, were selected from the conformational ensemble, based on best agreement with the CSs from NMR or the SASA from ESI-MS (Figure 5, see Methods for selection criteria). Although these models are not representative of the entire N_{TAIL}-P_{XD} conformational ensemble, they are useful to inspect distinct interaction networks that could drive protein compaction and can also provide hints for targeting the complex in a conformation-specific manner by small molecules. Thus, contact analysis was performed, in order to investigate the nature of the interactions involving either the α-MoRE or the disordered regions that were missing from the crystallographic structure. Schematic diagrams of the interface between N_{TAIL} and P_{XD} are shown in Figure S5. The interface between the α-MoRE and P_{XD} shows minor rearrangements relative to the X-ray structure, which could reflect structure dynamics in solution and/or differences between the chimeric protein and the intermolecular complex. Nevertheless, the interface within the four-helix bundle remains almost exclusively hydrophobic in Model 1 and Model 2, while it shows some additional electrostatic interactions in Model 3. Analysis of the contacts established by the fuzzy regions (see Figure S5) unveiled their prevalently hydrophobic nature, suggesting that the interactions between P_{XD} and N_{TAIL} are dominated by hydrophobic interactions, even outside the folded, globular core. A few hydrogen bonds and salt-bridges seem to give additional contribution to the binding of the two proteins. Thus, it seems that not only the helical core of the complex, but also the surrounding disordered cloud is dominated by hydrophobic interactions, although hydrogen bonding and electrostatic interactions seem to play a role in stabilizing the more compact states. The prevalently hydrophobic nature of the N_{TAIL}-P_{XD} complex is in agreement with previous reports highlighting that, although IDPs are characterized by a low content in hydrophobic residues, not sufficient to drive formation of a hydrophobic core, hydrophobic residues often play a dominant role in mediating physiologically relevant protein-protein interactions [67].

Conclusions

This study combines information derived from biophysical investigation and molecular simulations, in order to develop high-resolution structural models for the fuzzy complex between N_{TAIL} and P_{XD}. The results show that the staccato-type interactions characterizing this molecular ensemble are mainly of hydrophobic nature. While previous atomic models focused only on the α-MoRE region of N_{TAIL}, this study combines experimental data and computational modeling to generate models of the entire N_{TAIL}-P_{XD} complex at atomic resolution. By including the disordered arms in the molecular modeling, it was possible to extend interaction analysis to the dynamic regions of the complex, which are likely involved in biological function. The information that can be derived from this analysis sheds light onto the nature of the interactions that drive protein compaction and sets the basis for targeting specific conformations of the complex by small-molecules inhibitors.

The good agreement between SASA values derived from CSD analysis and solution models for the open state of the N_{TAIL}/P_{XD} complex suggests that the high-charge component of MS spectra reflects structural properties of the heterogeneous conformational ensemble and gives, in turn, further support to the structural details derived from the models. The MS results reported here also point out, for the first time, a compact state of the complex. IM/MS data indicate that this component is characterized by high structural heterogeneity in the gas phase. The broad arrival time distributions observed for the low charge-state component could be a distinct feature of collapsed states of IDPs, as opposed to natively folded, globular structures. Since compact conformations were not detected by previous experimental investigation, further studies will be required to tell whether they represent a minor population of the solution ensemble or a collapsed species induced by the electrospray process. Normally, charge states and CCS are closely linked to each other, but, if the protein can re-structure after the charge states are generated (i.e. after ESI), there can be a discrepancy between the two. Systematic comparison of solution and gas-phase conformational ensembles of IDPs will be needed to clarify this point.

Acknowledgments

The authors thank Carlo Santambrogio for expert assistance and helpful discussions. This work was carried out with the financial support of the CNRS and of the Agence Nationale de la Recherche, specific programs "Physico-Chimie du Vivant" (ANR-08-PCVI-0020-01) and "ASTRID" (ANR-11-ASTR-003-01) to S.L. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. F.S. is a Francqui Research Professor at UA. The Synapt G2 mass spectrometer is funded by a grant from the Hercules Foundation – Flanders.

References

1. Habchi, J., Tompa, P., Longhi, S., Uversky, V.N.: Introducing Protein Intrinsic Disorder. Chem. Rev. 114, 6561-88 (2014)

2. Tompa, P.: Intrinsically disordered proteins: a 10-year recap. Trends in Biochemical Sciences 37, 509-516 (2012)

3. Uversky, V.N.: A decade and a half of protein intrinsic disorder: Biology still waits for physics. Protein Science 22, 693-724 (2013)

4. Oates, M.E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M.J., Xue, B., Dosztanyi, Z., Uversky, V.N., Obradovic, Z., Kurgan, L., Dunker, A.K., Gough, J.: (DP2)-P-2: database of disordered protein predictions. Nucleic Acids Research 41, D508-D516 (2013)

5. Xue, B., Dunker, A.K., Uversky, V.N.: Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. Journal of Biomolecular Structure & Dynamics 30, 137-149 (2012)

6. Schad, E., Tompa, P., Hegyi, H.: The relationship between proteome size, structural disorder and organism complexity. Genome Biology 12, 1-13 (2011)

7. Dunker, A.K., Obradovic, Z.: The protein trinity - linking function and disorder. Nature Biotechnology 19, 805-806 (2001)

8. Uversky, V.N., Santambrogio, C., Brocca, S., Grandori, R.: Length-dependent compaction of intrinsically disordered proteins. Febs Letters 586, 70-73 (2012)

9. Kovacs, D., Szabo, B., Pancsa, R., Tompa, P.: Intrinsically disordered proteins undergo and assist folding transitions in the proteome. Archives of Biochemistry and Biophysics 531, 80-89 (2013)

10. Fong, J.H., Shoemaker, B.A., Garbuzynskiy, S.O., Lobanov, M.Y., Galzitskaya, O.V., Panchenko, A.R.: Intrinsic Disorder in Protein Interactions: Insights From a Comprehensive Structural Analysis. Plos Computational Biology 5, 1-11 (2009)

11. Lobanov, M.Y., Shoemaker, B.A., Garbuzynskiy, S.O., Fong, J.H., Panchenko, A.R., Galzitskaya, O.V.: ComSin: database of protein structures in bound (complex) and unbound (single) states in relation to their intrinsic disorder. Nucleic Acids Research 38, D283-D287 (2010)

12. Tompa, P., Fuxreiter, M.: Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends in Biochemical Sciences 33, 2-8 (2008)

13. Hazy, E., Tompa, P.: Limitations of Induced Folding in Molecular Recognition by Intrinsically Disordered Proteins. Chemphyschem 10, 1415-1419 (2009)

14. Uversky, V.N.: Unusual biophysics of intrinsically disordered proteins. Biochimica Et Biophysica Acta-Proteins and Proteomics 1834, 932-951 (2013)

15. Fuxreiter, M.: Fuzziness: linking regulation to protein dynamics. Molecular Biosystems 8, 168-177 (2012)

16. Meszaros, B., Dosztanyi, Z., Simon, I.: Disordered Binding Regions and Linear Motifs-Bridging the Gap between Two Models of Molecular Recognition. Plos One 7, 1-13 (2012)

17. Fuxreiter, M., Tompa, Peter: Fuzziness: Structural Disorder in Protein Complexes. Advances in Experimental Medicine and Biology. Springer, New York (2012)

18. 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 41, 1906-1914 (2002)

19. Testa, L., Brocca, S., Grandori, R.: Charge-Surface Correlation in Electrospray Ionization of Folded and Unfolded Proteins. Analytical Chemistry 83, 6459-6463 (2011)

20. Kaltashov, I.A., Mohimen, A.: Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. Analytical Chemistry 77, 5370-5379 (2005)

21. Konijnenberg, A., Butterer, A., Sobott, F.: Native ion mobility-mass spectrometry and related methods in structural biology. Biochimica Et Biophysica Acta-Proteins and Proteomics 1834, 1239-1256 (2013)

22. Beveridge, R., Chappuis, Q., Macphee, C., Barran, P.: Mass spectrometry methods for intrinsically disordered proteins. Analyst 138, 32-42 (2013)

23. 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 6, 281-294 (2014)

24. Woods, L.A., Radford, S.E., Ashcroft, A.E.: Advances in ion mobility spectrometry-mass spectrometry reveal key insights into amyloid assembly. Biochimica Et Biophysica Acta-Proteins and Proteomics 1834, 1257-1268 (2013)

25. Cong, X., Casiraghi, N., Rossetti, G., Mohanty, S., Giachin, G., Legname, G., Carloni, P.: Role of Prion Disease-Linked Mutations in the Intrinsically Disordered N-Terminal Domain of the Prion Protein. Journal of Chemical Theory and Computation 9, 5158-5167 (2013)

26. Dibenedetto, D., Rossetti, G., Caliandro, R., Carloni, P.: A Molecular Dynamics Simulation-Based Interpretation of Nuclear Magnetic Resonance Multidimensional Heteronuclear Spectra of alpha-Synuclein.Dopamine Adducts. Biochemistry 52, 6672-6683 (2013)

27. Wang, Y., Chu, X., Longhi, S., Roche, P., Han, W., Wang, E., Wang, J.: Multiscaled exploration of coupled folding and binding of an intrinsically disordered molecular recognition element in measles virus nucleoprotein. Proceedings of the National Academy of Sciences of the United States of America 110, E3743-E3752 (2013)

28. Longhi, S., Receveur-Brechot, V., Karlin, D., Johansson, K., Darbon, H., Bhella, D., Yeo, R., Finet, S., Canard, B.: The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. Journal of Biological Chemistry 278, 18638-18648 (2003)

29. Bourhis, J.M., Johansson, K., Receveur-Brechot, V., Oldfield, C.J., Dunker, K.A., Canard, B., Longhi, S.: The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner. Virus Research 99, 157-167 (2004)

30. Gely, S., Lowry, D.F., Bernard, C., Jensen, M.R., Blackledge, M., Costanzo, S., Bourhis, J.-M., Darbon, H., Daughdrill, G., Longhi, S.: Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein. Journal of Molecular Recognition 23, 435-447 (2010)

31. Jensen, M.R., Communie, G., Ribeiro, E.A., Jr., Martinez, N., Desfosses, A., Salmon, L., Mollica, L., Gabel, F., Jamin, M., Longhi, S., Ruigrok, R.W.H., Blackledge, M.: Intrinsic disorder in measles virus nucleocapsids. Proceedings of the National Academy of Sciences of the United States of America 108, 9839-9844 (2011)

32. Johansson, K., Bourhis, J.M., Campanacci, V., Cambillau, C., Canard, B., Longhi, S.: Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. Journal of Biological Chemistry 278, 44567-44573 (2003)

33. Bourhis, J.M., Receveur-Brechot, V., Oglesbee, M., Zhang, X.S., Buccellato, M., Darbon, H., Canard, B., Finet, S., Longhi, S.: The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. Protein Science 14, 1975-1992 (2005)

34. Belle, V., Rouger, S., Costanzo, S., Liquiere, E., Strancar, J., Guigliarelli, B., Fournel, A., Longhi, S.: Mapping alpha-helical induced folding within the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by site-directed spin-labeling EPR spectroscopy. Proteins-Structure Function and Bioinformatics 73, 973-988 (2008)

35. Kingston, R.L., Hamel, D.J., Gay, L.S., Dahlquist, F.W., Matthews, B.W.: Structural basis for the attachment of a paramyxoviral polymerase to its template. Proceedings of the National Academy of Sciences of the United States of America 101, 8301-8306 (2004)

36. Bernard, C., Gely, S., Bourhis, J.-M., Morelli, X., Longhi, S., Darbon, H.: Interaction between the C-terminal domains of N and P proteins of measles virus investigated by NMR. Febs Letters 583, 1084-1089 (2009)

37. Shu, Y., Habchi, J., Costanzo, S., Padilla, A., Brunel, J., Gerlier, D., Oglesbee, M., Longhi, S.: Plasticity in Structural and Functional Interactions between the Phosphoprotein and Nucleoprotein of Measles Virus. Journal of Biological Chemistry 287, 11951-11967 (2012)

38. Blocquel, D., Habchi, J., Costanzo, S., Doizy, A., Oglesbee, M., Longhi, S.: Interaction between the C-terminal domains of measles virus nucleoprotein and phosphoprotein: A tight complex implying one binding site. Protein Science 21, 1577-1585 (2012)

39. Kavalenka, A., Urbancic, I., Belle, V., Rouger, S., Costanzo, S., Kure, S., Fournel, A., Longhi, S., Guigliarelli, B., Strancar, J.: Conformational Analysis of the Partially Disordered Measles Virus N-TAIL-XD Complex by SDSL EPR Spectroscopy. Biophysical Journal 98, 1055-1064 (2010)

40. Liu, L., Bagal, D., Kitova, E.N., Schnier, P.D., Klassen, J.S.: Hydrophobic Protein-Ligand Interactions Preserved in the Gas Phase. Journal of the American Chemical Society 131, 15980-15981 (2009)

41. Barylyuk, K., Balabin, R.M., Gruenstein, D., Kikkeri, R., Frankevich, V., Seeberger, P.H., Zenobi, R.: What Happens to Hydrophobic Interactions during Transfer from the Solution to the Gas Phase? The Case of Electrospray-Based Soft Ionization Methods. Journal of the American Society for Mass Spectrometry 22, 1167-1177 (2011)

42. Marcoux, J., Wang, S.C., Politis, A., Reading, E., Ma, J., Biggin, P.C., Zhou, M., Tao, H., Zhang, Q., Chang, G., Morgner, N., Robinson, C.V.: Mass spectrometry reveals synergistic effects of nucleotides, lipids, and drugs binding to a multidrug resistance efflux pump. Proceedings of the National Academy of Sciences of the United States of America 110, 9704-9709 (2013)

43. Dobo, A., Kaltashov, I.A.: Detection of multiple protein conformational ensembles in solution via deconvolution of charge-state distributions in ESI MS. Analytical Chemistry 73, 4763-4773 (2001)

44. Borysik, A.J.H., Radford, S.E., Ashcroft, A.E.: Co-populated conformational ensembles of beta(2)-microglobulin uncovered quantitatively by electrospray ionization mass spectrometry. Journal of Biological Chemistry 279, 27069-27077 (2004)

45. 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. Analytical Chemistry 82, 9557-9565 (2010)

46. Bowie, J.U., Eisenberg, D.: An Evolutionary Approach to Folding Small Alpha-Helical Proteins That Uses Sequence Information and an Empirical Guiding Fitness Function. Proceedings of the National Academy of Sciences of the United States of America 91, 4436-4440 (1994)

47. Simons, K.T., Kooperberg, C., Huang, E., Baker, D.: Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. Journal of Molecular Biology 268, 209-225 (1997)

48. Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., Sali, A.: Comparative protein structure modeling of genes and genomes. Annual Review of Biophysics and Biomolecular Structure 29, 291-325 (2000)

49. Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V.: Stereochemistry of polypeptide chain configurations. Journal of molecular biology 7, 95-9 (1963)

50. Laskowski, R.A., Macarthur, M.W., Moss, D.S., Thornton, J.M.: Procheck - a program to check the stereochemical of protein structures. Journal of Applied Crystallography 26, 283-291 (1993)

51. Shen, M.-Y., Sali, A.: Statistical potential for assessment and prediction of protein structures. Protein Science 15, 2507-2524 (2006)

52. Irback, A., Mitternacht, S., Mohanty, S.: An effective all-atom potential for proteins. PMC biophysics 2, 2-2 (2009)

53. Li, D.-W., Mohanty, S., Irback, A., Huo, S.: Formation and Growth of Oligomers: A Monte Carlo Study of an Amyloid Tau Fragment. Plos Computational Biology 4, (2008)

54. Jonsson, S.A., Mohanty, S., Irback, A.: Distinct phases of free alpha-synuclein-A Monte Carlo study. Proteins-Structure Function and Bioinformatics 80, 2169-2177 (2012)

55. Fukunishi, H., Watanabe, O., Takada, S.: On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. Journal of Chemical Physics 116, 9058-9067 (2002)

56. Irback, A., Mohanty, S.: PROFASI: A Monte Carlo simulation package for protein folding and aggregation. Journal of Computational Chemistry 27, 1548-1555 (2006)

57. Neal, S., Nip, A.M., Zhang, H.Y.,Wishart, D.S.: Rapid and accurate calculation of protein H-1, C-13 and N-15 chemical shifts. Journal of Biomolecular NMR 26, 215-240 (2003)

58. Van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., Berendsen, H.J.C.: Gromacs: fast, flexible, and free. Journal of Computational Chemistry 26, 1701-1718 (2005)

59. Bulheller, B.M., Hirst, J.D.: DichroCalc-circular and linear dichroism online. Bioinformatics 25, 539-540 (2009)

60. Besley, N.A., Hirst, J.D.: Theoretical studies toward quantitative protein circular dichroism calculations. Journal of the American Chemical Society 121, 9636-9644 (1999)

61. Sreerama, N.,Woody, R.W.: Structural composition of beta(I)- and beta(II)-proteins. Protein Science 12, 384-388 (2003)

62. Kingston, R.L., Gay, L.S., Baase, W.S., Matthews, B.W.: Structure of the nucleocapsid-binding domain from the mumps virus polymerase; an example of protein folding induced by crystallization. Journal of Molecular Biology 379, 719-731 (2008)

63. Yegambaram, K., Bulloch, E.M.M., Kingston, R.L.: Protein domain definition should allow for conditional disorder. Protein Science 22, 1502-1518 (2013)

64. Testa, L., Brocca, S., Santambrogio, C., D'Urzo, A., Habchi, J., Longhi, S., Uversky, V.N., Grandori, R.: Extracting structural information from charge-state distributions of intrinsically disordered proteins by non-denaturing electrospray-ionization mass spectrometry. Intrinsically Disordered Proteins 1, 18-24 (2013)

65. Natalello, A., Benetti, F., Doglia, S.M., Legname, G., Grandori, R.: Compact conformations of alpha-synuclein induced by alcohols and copper. Proteins-Structure Function and Bioinformatics 79, 611-621 (2011)

66. Santambrogio, C., Frana, A.M., Natalello, A., Papaleo, E., Regonesi, M.E., Doglia, S.M., Tortora, P., Invernizzi, G., Grandori, R.: The role of the central flexible region on the aggregation and conformational properties of human ataxin-3. Febs Journal 279, 451-463 (2012)

67. Meszaros, B., Tompa, P., Simon, I., Dosztanyi, Z.: Molecular principles of the interactions of disordered proteins. Journal of Molecular Biology 372, 549-561 (2007)

68. Wallace, A.C., Laskowski, R.A., Thornton, J.M.: Ligplot - a program to generate schematic diagrams of protein ligand interactions. Protein Engineering 8, 127-134 (1995)

Figure legends

Figure 1. Nano-ESI-MS spectra of MeV N_{TAIL} (ABSciex QSTAR Elite instrument).

(A) 10 µM MeV N_{TAIL} in 10 mM ammonium acetate, pH 6.5. (B) Gaussian fitting of the spectrum reported in panel A. (C) 10 µM MeV N_{TAIL} in 10 mM ammonium acetate, 1% formic acid (pH ~3). (D) 10 µM MeV N_{TAIL} in 10 mM ammonium acetate, 50% acetonitrile, 1% formic acid (pH ~3).

Figure 2. Nano-ESI-MS spectra of MeV P_{XD}. (ABSciex QSTAR Elite instrument).

(A) 10 µM MeV P_{XD} in 10 mM ammonium acetate, pH 6.5. (B) Gaussian fitting of the spectrum reported in panel A. (C) 10 µM MeV P_{XD} in 10 mM ammonium acetate, 1% formic acid (pH~ 3). (D) 10 µM MeV P_{XD} in 10 mM ammonium acetate, 50% acetonitrile, 1% formic acid (pH ~3).

Figure 3. Nano-ESI-MS spectra of MeV N_{TAIL} and P_{XD} mixtures. (ABSciex QSTAR Elite instrument) Equimolar mixtures in 10 mM ammonium acetate, pH 6.5**. (**A) 10 µM of each protein. (B) 20 µM of each protein. Squares, P_{XD}; circles, N_{TAIL}; stars, complex.

Figure 4. Nano-ESI-IM-MS spectrum of a MeV N_{TAIL} and P_{XD} mixture (Waters Synapt G2 HDMS instrument) (A) Excerpt of the higher m/z region showing complex formation of P_{XD}-N_{TAIL} for an equimolar ratio of 20 µM protein in 10 mM ammonium acetate (pH 6.5). (B) The P_{XD}-N_{TAIL} complex displays a significant degree of structural heterogeneity, based on the arrival

times of the low charge (8-10+) states. The 8-10+ charge states of the P_{XD}-N_{TAIL} complex show at least two different conformations, with corresponding CCSs values that vary over a wide range. (C) Experimentally determined collision cross sections of the P_{XD}-N_{TAIL} complex reveal that it maintains its structurally heterogeneous character over a broad range of charge states. Coloured lines indicate the theoretical collision cross sections for the structural models obtained from MD simulations as discussed in fig 5, with model 1, 2 and 3 represented as red, blue and green lines, respectively.

Figure 5. Structural models of the MeV N_{TAIL}- P_{XD} complex.

Cartoon representation of selected models obtained by PROFASI (Model 1, Model 2, and Model 3) for the disordered component. Blue, N_{TAIL}; orange, P_{XD}; grey, superimposed X-ray structure (PDB ID: 1T6O) [35]. The comparison between the calculated and experimental CSs of the N, C, C_α, and C_β atoms of N_{TAIL} bound to P_{XD} in each selected models is shown in Figure S3. (B) Schematic diagrams of the N_{TAIL}- P_{XD} interface for the X-ray structure (PDB ID: 1T6O) [35] and for each of the three models shown in panel A generated by the LIGPLOT package [68] which produces schematic 2D representations of protein-ligand complexes. Hydrophobic interactions are indicated by spokes (color code) and hydrogen bonds by dashed green lines. Bond lengths (Å) are shown. The N_{TAIL} residues belonging to the α-MoRE (488-499) are labeled by stars.

Figures

Figure 1

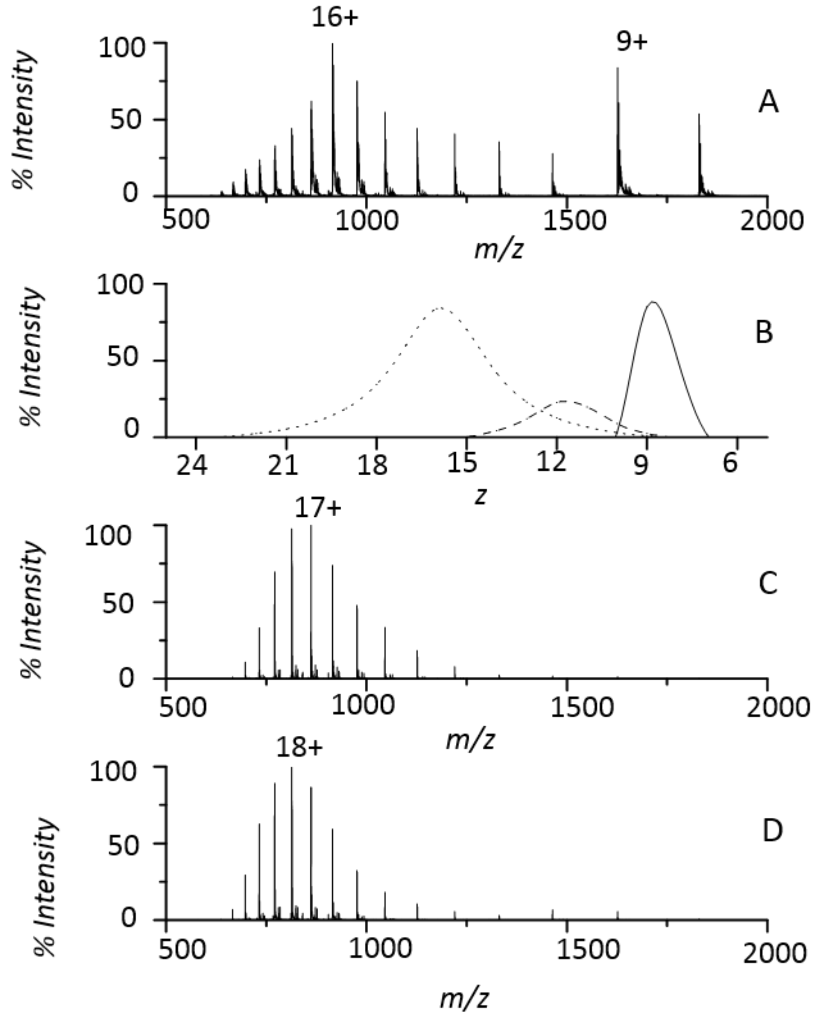

16+
9+
100
50
0
% Intensity
A
500
1000
1500
2000
m/z
100
50
0
% Intensity
B
24
21
18
15
12
9
6
z
17+
100
50
0
% Intensity
C
500
1000
1500
2000
m/z
18+
100
50
0
% Intensity
D
500
1000
1500
2000
m/z

Figure 2

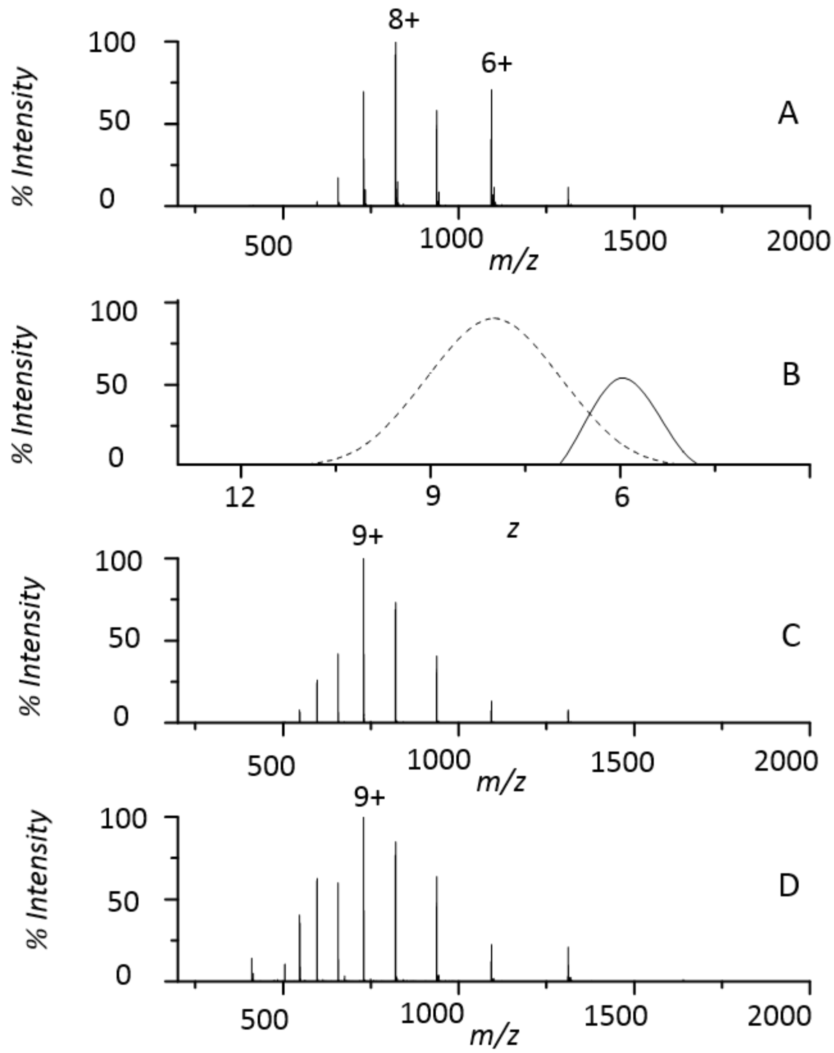

8+
6+
100
50
0
% Intensity
A
500
1000
1500
2000
m/z
100
50
0
% Intensity
B
12
9
6
z
9+
100
50
0
% Intensity
C
500
1000
1500
2000
m/z
9+
100
50
0
% Intensity
D
500
1000
1500
2000
m/z

Figure 3

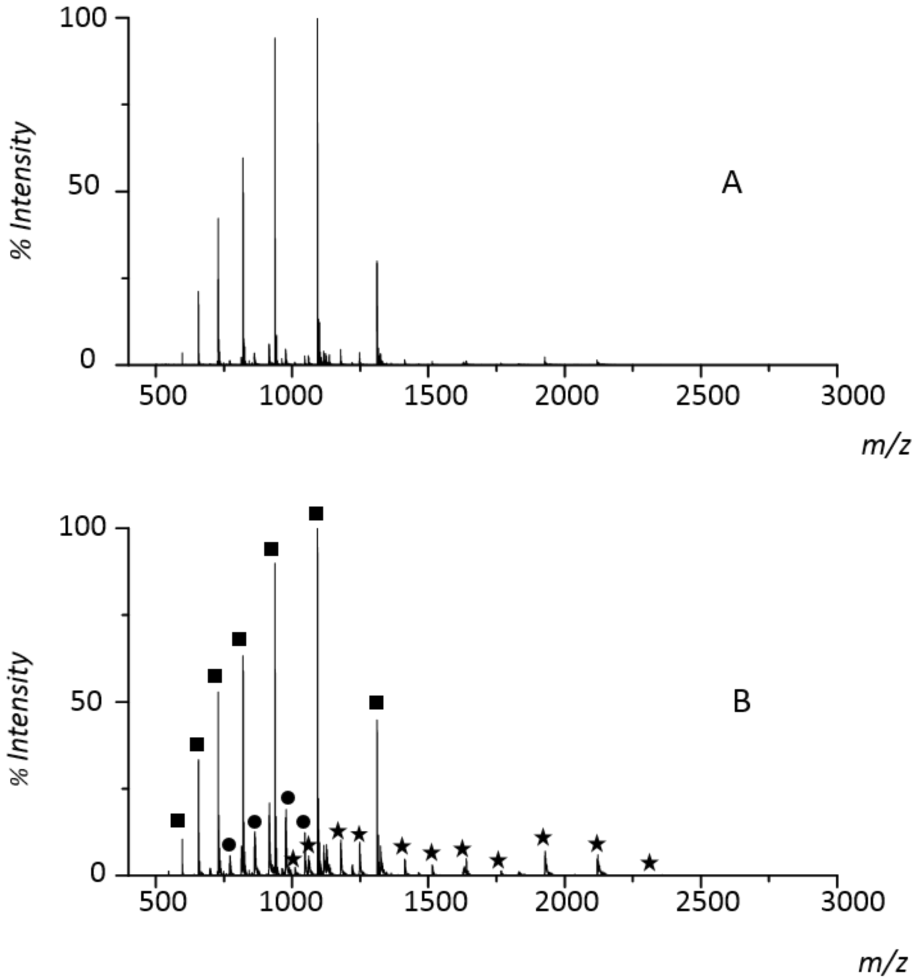

100
50
0
% Intensity
A
500
1000
1500
2000
2500
3000
m/z
100
50
0
% Intensity
B
500
1000
1500
2000
2500
3000
m/z

Figure 4

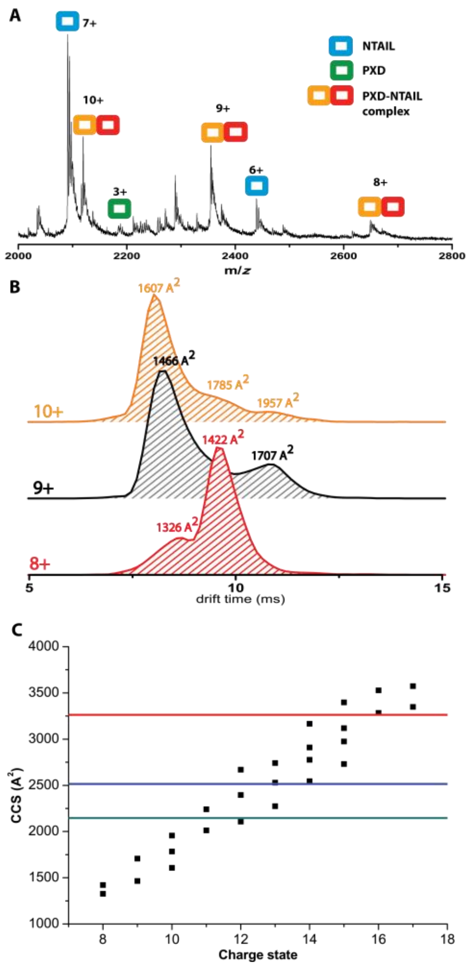
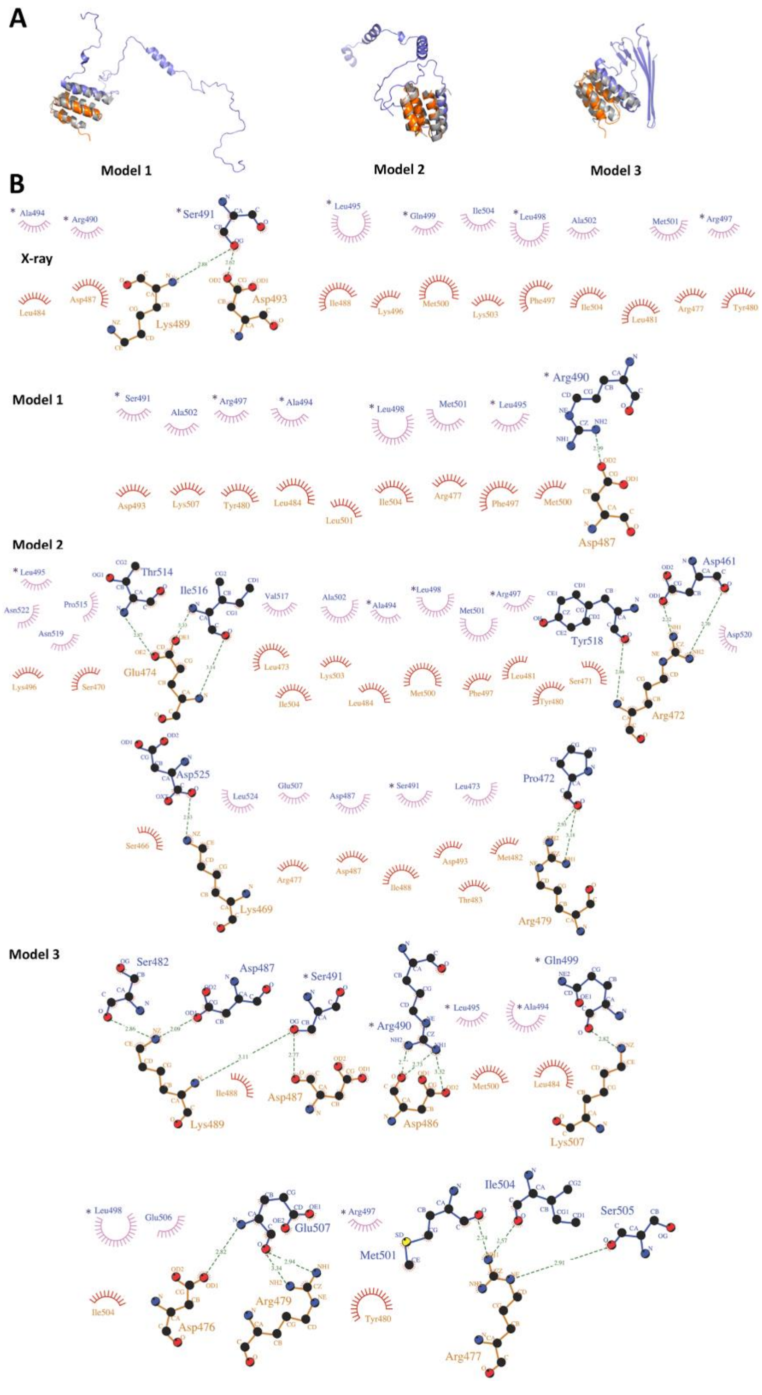
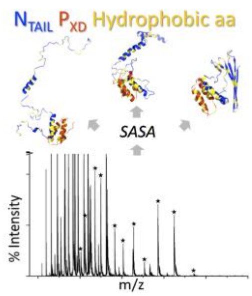

A
7+
NTAIL
PXD
PXD-NTAIL complex
10+
9+
3+
6+
8+
2000
2200
2400
2600
2800
m/z
B
1607 A2
1466 A2
1785 A2
1957 A2
10+
1422 A2
1707 A2
9+
1326 A2
8+
5
10
15
drift time (ms)
C
4000
3500
3000
2500
2000
1500
1000
CCS (A2)
8
10
12
14
16
18
Charge state

Figure 5



Graphical Abstract



N_{TAIL} P_{XD} Hydrophobic aa
SASA
% Intensity
m/z