

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/282619198>

# Ensemble Methods Enable a New Definition for the Solution to Gas-Phase Transfer of Intrinsically Disordered Proteins

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · OCTOBER 2015

Impact Factor: 12.11 · DOI: 10.1021/jacs.5b06027

---

READS

42

## 4 AUTHORS, INCLUDING:



[Antoni Borysik](#)

King's College London

13 PUBLICATIONS 226 CITATIONS

[SEE PROFILE](#)



[Mainak Guharoy](#)

Vlaams Instituut voor Biotechnologie

23 PUBLICATIONS 420 CITATIONS

[SEE PROFILE](#)



[Peter Tompa](#)

Vlaams Instituut voor Biotechnologie

132 PUBLICATIONS 8,543 CITATIONS

[SEE PROFILE](#)

# Ensemble Methods Enable a New Definition for the Solution to Gas-Phase Transfer of Intrinsically Disordered Proteins

Antoni J. Borysik,<sup>\*,†,II</sup> Denes Kovacs,<sup>‡,II</sup> Mainak Guharoy,<sup>‡</sup> and Peter Tompa<sup>\*,‡,§</sup>

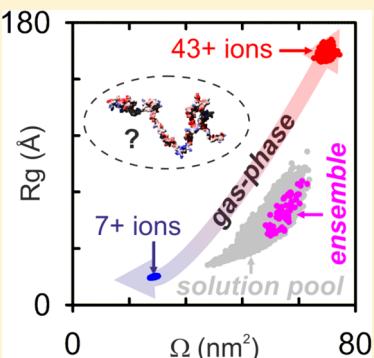
<sup>†</sup> King's College London, Department of Chemistry, Britannia House, 7 Trinity Street, London SE1 1DB, U.K.

<sup>‡</sup>VIB Structural Biology Research Centre (SBRC), Vrije Universiteit Brussel, Pleinlaan 2, Brussels B-1050, Belgium

<sup>§</sup>Institute of Enzymology, Research Centre for Natural Sciences of the Hungarian Academy of Sciences, 1117 Budapest, Hungary

## Supporting Information

**ABSTRACT:** Intrinsically disordered proteins (IDPs) are important for health and disease, yet their lack of net structure precludes an understanding of their function using classical methods. Gas-phase techniques provide a promising alternative to access information on the structure and dynamics of IDPs, but the fidelity to which these methods reflect the solution conformations of these proteins has been difficult to ascertain. Here we use state of the art ensemble techniques to investigate the solution to gas-phase transfer of a range of different IDPs. We show that IDPs undergo a vast conformational space expansion in the absence of solvent to sample a conformational space 3–5 fold broader than in solution. Moreover, we show that this process is coupled to the electrospray ionization process, which brings about the generation of additional subpopulations for these proteins not observed in solution due to competing effects on protein charge and shape. Ensemble methods have permitted a new definition of the solution to gas-phase transfer of IDPs and provide a roadmap for future investigations into flexible systems by mass spectrometry.



## INTRODUCTION

Intrinsically disordered proteins (IDPs) do not fold to unique structures but instead sample a vast number of different conformations under native conditions.<sup>1</sup> Recent predictions estimate that ~30% of all eukaryotic proteins contain disordered regions and that disorder is overrepresented in many proteins linked to human disease.<sup>2</sup> The prevailing challenge in understanding IDPs is to relate how the function of these proteins arises from the disordered state. The most useful structural depictions of IDPs must therefore define these proteins in terms of ensembles. This has required the development of new methods that can describe structural disorder by allowing for the presence of many coexisting conformations from the average solution property. Modern advances in ensemble methods have contributed most significantly to our understanding of IDPs.<sup>3–8</sup> These methods can be applied with a number of different techniques to provide different depictions of IDP structure at the local or global level. In recent years the literature has been dominated by ensemble approaches in both hybrid<sup>9,10</sup> or isolated<sup>11,12</sup> workflows.

Electrospray ionization mass spectrometry (ESI-MS) provides an attractive alternative to ensemble methods to describe the structure and dynamics of IDPs. ESI-MS is not hampered by structural averaging and therefore overcomes the challenge of describing structural flexibility by directly differentiating between closely related structures at equilibrium. This property of ESI-MS allows it to distinguish between different protein folds simultaneously. The characteristically broadened charge state distributions observed for disordered proteins are

interpreted via unbiased contributions from different structural subpopulations in solution.<sup>13–15</sup> These subpopulations can be extracted and quantified directly from ESI-MS charge state distributions by linear deconvolution to provide a unique snapshot of the conformational space that is sampled by a protein that has multiple structures.

Ion-mobility MS (IMS) also provides complementary low resolution structural information on each structure within an ensemble in the form of a collision cross section ( $\Omega$ ). In contrast to globular proteins, which form compact gas-phase structures, IDPs generate  $\Omega$  consistent with extended geometries more in line with the elongated states expected for these proteins in solution.<sup>16–18</sup> Through IMS it has been shown that disordered proteins display a characteristic degree of structural heterogeneity in the gas-phase that is not observed for folded proteins. This suggests that IDPs retain a memory of their structural flexibility even in the absence of solvent. Pioneering experiments by Jarrold and co-workers suggested that flexible structures survive in the gas-phase allowing their solution conformations to be captured and characterized directly by IMS.<sup>19</sup> Overall, gas-phase methods appear uniquely suited to meet the challenge of describing the conformations and dynamics of flexible proteins as evidenced by the increasing number of publications in this area in recent years.<sup>20–24</sup>

Received: September 23, 2014

Revised: June 10, 2015



Nevertheless, recent studies describing the collapse of disordered structures in the absence of solvent have raised questions over the fidelity to which gas-phase methods can capture the solution conformations of these proteins.<sup>25,26</sup> Recent shifts in our understanding of how flexible species escape electrospray droplets during ESI also adds to the complexity of describing the solution to gas-phase transfer of IDPs.<sup>27,28</sup> The present challenge is to understand how to reconcile these recent findings within more established frameworks describing the behavior of disordered proteins in the solvent-free environment. A reevaluation of the solution to gas-phase transfer of IDPs is required. Such studies are also timely due to the recent development and accessibility of ensemble methods, which enable the solution and gas-phase structures of IDPs to be compared more rigorously by considering their conformational space rather than just their net structures.

Here we combine a number of methods to compare the conformations of many different IDPs in solution and gas-phase. SAXS is used in combination with the ensemble optimization method (EOM) to generate a limited number of structures that collectively describe the preferred conformational space of the IDPs in solution. These structures are then compared with those obtained directly in the gas-phase by IMS. We reveal that IDPs sample a 3–5 fold broader conformational space in the gas-phase than in solution. This behavior is replicated by vacuum molecular dynamics (MD) simulations wherein the solution structures spontaneously adopt a range of new charge-dependent conformations, from completely collapsed chains to fully charge extended rods to occupy the same broader conformational space observed by IMS. The genetic algorithm judging optimization of ensembles (GAJOE) is used to validate all findings and confirm the absence of gas-phase structures in solution. Overall our data agree with previous gas-phase studies of IDPs in that ESI-MS captures various qualitative aspects of protein disorder. However, the solution scattering curves of IDPs cannot be reconstructed using restraints derived from IMS, nor can any of the gas-phase models be fitted to the solution scattering behavior of the proteins. The large degree to which MS overrepresents the conformational space of IDPs is arguably the most distinctive feature of the solution to gas-phase transfer of these proteins.

SAXS-derived  $R_g$  distributions contain important information on preferential states or subpopulations within the conformational space of a disordered protein. This ability of SAXS allows us to investigate the accuracy to which ESI charge state distributions represent the conformational space that is occupied by IDPs in solution.<sup>24,29–32</sup> We find dramatically different structural depictions of the IDPs depending on the technique employed to characterize these proteins. In solution the IDP ensembles largely describe unimodal distributions of conformers, whereas gas-phase methods describe a conformational space that is characterized by 3 distinct subpopulations broadly defined as compact (*c*), intermediate (*i*) and extended (*e*) with increasing  $\Omega$  and charge. This discrepancy has not been previously described for IDPs and indicates the capacity of ESI to generate additional subpopulations for these proteins. To investigate this further the charge signatures ( $z_{ave}$ ) of all the IDP subpopulations were obtained and compared to values predicted for different solution and gas-phase states. We find a wide range of  $z_{ave}$  across the different IDP subpopulations providing evidence for a bifurcation in ESI mechanisms which

drives the generation of these subpopulations and the conformational space expansion of the IDPs.

We have applied state of the art ensemble methods to investigate the solution to gas-phase transfer of IDPs. The most striking feature of this process is the large extent to which the conformational space of these proteins becomes expanded in the absence of solvent. These effects originate during ESI which imposes the generation of additional subpopulations that are absent in solution with widely contrasting charge signatures and shapes. The present data suggest that previous definitions of the gas-phase behavior of IDPs are too simplistic. The subpopulations and structures observed for IDPs by ESI-MS cannot be interpreted directly relative to the conformational space, geometries and flexibility of these proteins in solution. Future experiments should be directed to understand how these ionization effects manifest themselves for different proteins in various IDP classes.<sup>33</sup> Clues relating variations in the internal structure and persistence length of the IDPs with their charge signatures and shapes are discussed.

## ■ EXPERIMENTAL SECTION

**Protein Preparation.** In order to cover a spectrum of IDPs, we selected six IDPs, AavLEA1 (Q95V77),<sup>34</sup> human calpastatin domain 1 (P20810; segment: 137–237 residues),<sup>35,36</sup> ERD10 (P42759), ERD14 (P42763),<sup>37</sup> L19 (P0A7K6)<sup>38</sup> and Map2c (P15146–3) and one molten globule protein domain, NCBD domain of human p300 (Q09472, segment: 2038–2142 residues). All proteins were produced via recombinant expression in *E. coli* BL21(DE3) Star expression strain, and purified using protocols described previously (see below for a short description including modifications). All purifications were carried out on an AKTA Avant (GE Healthcare) FPLC system, with prepacked columns provided by GE Healthcare. ERD10 and ERD14 were purified using three-step purification method, first by ion exchange on HiTrap Q FF at pH 9.5 with gradient elution, followed by two gel filtration steps (Superdex 200 and Superdex 75).<sup>37</sup> L19 was purified under denaturing conditions using a HiTrap SP FF column, with an additional gel filtration step with Superdex 75.<sup>38</sup> AavLEA1 was purified using HisTrap HP followed by gel filtration with Superdex 75.<sup>39</sup> Rat Map2c was purified on SP Sepharose and Superdex 200.<sup>34</sup> Human calpastatin domain 1 (corresponding to A137-K277 of human calpastatin, SwissProt entry P20810) was purified in two steps, on HiTrap Q FF and Superdex 75.<sup>35,36</sup> All manipulations were carried out at low temperature with increased protease inhibitor concentrations (1 tablet/50 mL of Roche complete ultra tablet, 0.5 mM phenyl-methyl-sulfonyl-fluoride and 2 mM benzamidine) in order to avoid proteolytic cleavage. The final purity of all the proteins was determined via analytical gel filtration and MS and was found to be ~95%, except for Map2c where the final purity was ~85%. We are grateful to the following scientists for sharing expression vectors with us, rat brain Map2c expression vector was a gift from Andrew Matus (Friedrich Miser Institute, Basel, Switzerland), L19 was a gift from Renée Schroeder (Max Perutz Laboratories, Institute of Microbiology and Genetics, Vienna, Austria), AavLEA1 is a gift from Alan Tunnicliffe (Institute of Biotechnology, University of Cambridge, U.K.), human calpastatin domain 1 is a courtesy of Masatoshi Maki, Nagoya University, Nagoya, Japan and scattering curves together with the protein sample of human calpastatin domain 1 has been provided by Kris Pauwels (Structural Biology Department, Vlaams Instituut voor Biotechnologie (VIB)).

**SAXS Measurement.** SAXS measurements of Calpastatin and L19 were carried out in Soleil on the SWING beamline, while AavLEA1, ERD10, ERD14, NCBD and Map2c were measured in DESY on the PETRAIII beamline. The measurements in Soleil were performed as follows. Monodisperse samples were obtained using HPLC gel filtration upstream to the SAXS measurement capillary in 20 mM TRIS buffer at pH 7.5 as a standard buffer, supplemented with 150 mM NaCl and 5 mM DTT. The data was processed with Foxtrot

(obtained at Soleil) and average curves were exported for further analysis. At DESY the measurements were carried out in batch mode on three different concentrations, where all sample and buffer pairs have been analyzed without chromatographic separation before data acquisition in the same buffer.<sup>40</sup> Samples were under constant flow to ensure removal of radiation damaged proteins. Approximately 70% of the frames were used to produce the final scattering curves. The concentration range of all proteins varied from ~1.5 to 10 mg/mL with an expose time of 1s/frame and each sample measured in triplicate. The data were processed using the ATSAS program package.<sup>41</sup> Each scattering curve went through quality testing, and final curves have been generated by merging low  $q$  regions (~0.075 1/ $\text{\AA}$ ) of low concentration samples with high  $q$  regions (~0.075 1/ $\text{\AA}$ ) of high concentration samples. No difference was observed for scattering curves obtained in 200 mM ammonium acetate used for ESI-MS.

**Ensemble Calculations.** Ensemble methods were used to describe the IDP scattering curves according to the following relationship (eq 1).

$$I(s) = \sum_k v_k I_k(s) \quad (1)$$

Where  $I(s)$  is the summed intensity of scattered light,  $I_k(s)$  is the scattering intensity from the  $k$ th component and  $v_k$  is the volume fraction for that component.<sup>42</sup> Eq 1 states that for a sample of  $n$  independent and randomly oriented particles the intensity of scattered radiation is equivalent to the sum of the form factors of each-and-every particle in solution. Conformational ensembles were calculated using procedures described for EOM and ATSAS program package with modification.<sup>3</sup> Briefly, a pool of random extended conformations were generated using Flexible-meccano v1.1s (pool size = 10.000) with PDB files printed.<sup>43</sup> The conformers were checked for quality and side chains were completed using Sccomp.<sup>44</sup> Theoretical scattering curves were calculated for all models, using CRYSTAL v2.8.2 and ensembles were selected using GAJOE v2.0 three times in parallel for comparison.<sup>41</sup> GAJOE were utilized to describe the data by selecting a subset of structures from a library of 10 000 conformations that collectively minimizes the  $\chi$ -value with the SAXS profiles  $I_{\text{exp}}(s)$  according to the following relationship (eq 2).

$$\chi = \sqrt{\frac{1}{K-1} \sum_{j=1}^K \left[ \frac{\mu I(s_j) - I_{\text{exp}}(s_j)}{\sigma(s_j)} \right]^2} \quad (2)$$

where  $K$  is the number of experimental points,  $\sigma(s_j)$  are the standard deviations and  $\mu$  is a scaling factor.<sup>3</sup> Calculations were also performed by limiting the ensemble size to 1, in order to select conformation that describes the scattering curve best (single representation). Conformational ensembles together with the SAXS scattering curves have been deposited in the PED database: <http://pedb.vib.be/> Map2c (8AAB), ERD10 (9AAB), ERD14 (1AAC), AavLEA1 (3AAC), NCBD (2AAC), calpastatin (7AAB) and L19 (4AAC).<sup>45</sup> Collision cross sections of the ensembles were obtained using the Exact Hard Sphere Scattering (EHSS) method implemented within Mobcal.<sup>46</sup>

**Mass Spectrometry.** The mass spectra of all IDPs were obtained in positive ion mode on a second generation Synapt HDMS (Waters Corp. Manchester, UK) quadrupole-ion trap-IMS instrument fitted with a nanoflow electrospray ionization source (nano-ESI). Samples were infused from nano-ESI needles that were prepared in-house as previously described.<sup>47</sup> Lyophilized proteins were initially dissolved in 200 mM ammonium acetate (~pH 7.0) and desalted using P6 Micro Bio-Spin columns (Bio-Rad, CA, USA) pre-equilibrated with the same buffer. Map2c was desalting by dialysis using a 10 kDa MWCO slide-alysyer (Thermo Fisher Scientific, Cramlington, UK) with 100  $\mu\text{L}$  Map2c dialyzed twice for >5 h against 500 mL of 200 mM ammonium acetate at room temperature. Desalted proteins were diluted with 200 mM ammonium acetate to a working concentration of between 1–10  $\mu\text{M}$ . All spectra were obtained using instrumental conditions at the limit for ion transmission with respective voltages for cone and trap DC bias of 25 and 30 V, and trap/transfer CE of 5 V. Gas-phase

unfolding of the IDPs was performed in argon by sequentially increasing the Trap collision voltage over the range of 5–200 V. The 7+ to 11+ ions of ERD10 were obtained by charge stripping the protein with the addition of 20 mM triethyl amine.

**Ion Mobility Mass Spectrometry.** Collision cross sections ( $\Omega$ ) of the IDPs were obtained by Traveling Wave Ion-mobility Mass Spectrometry (TWIMS) as previously described.<sup>48</sup> The 7+ to 11+, 14+ to 19+ and 15+ to 21+ ions of ubiquitin (Sigma-Aldrich, U6253) cytochrome C (Sigma-Aldrich, 30396) and horse heart myoglobin (Sigma-Aldrich, M1882) respectively were used as calibrants to enable determination of  $\Omega$  from experimental drift time measurements. IMS separation of the ions was performed with a constant wave height of 22 V and a wave velocity of 400 m/s. The drift times of all of the IDPs were then obtained without further modification of these instrumental settings or gas pressures. The drift times of the proteins used for calibration were corrected for mass-dependent flight using methods described previously.<sup>48</sup> The ATDs of each ion were extracted and the drift times converted to  $\Omega$ . The converted ATDs of each IDP were represented as a ListPlot3D using Mathematica V 8.0 (Hanborough, UK). The gas-phase subpopulations ( $c$ ,  $i$  and  $e$ ) of ERD10 were compared with the charge state distribution of the protein by direct extraction of the subpopulations from the protein contour plot. The partial mass spectra of each subpopulation was found to describe a Gaussian function the linear combination of which could describe the charge state distribution of the protein without further modification.

**Molecular Dynamics.** All MD simulations were performed without solvent using the GROMOS96 53a6 united atom force field implemented within GROMACS 4.5.5.<sup>49</sup> The GROMOS96 53a6 force field has been shown recently to capture the behavior of disordered proteins in solution.<sup>50</sup> Three structures were taken from the SAXS ensemble of ERD10 that encompassed the full range of solution geometries. Eight different charge states encompassing the experimental ionization states of the protein were generated by neutralization of negatively charged amino acids assuming initial protonation states of all charged residues at pH 7.0, histidine residues were considered neutral. MD simulations of charge neutral IDP species were also performed on the average singular representation model of each protein. For each structure and ionization state of ERD10 a range of different charge isomers were permuted and production MD simulations performed on charge isomers with the lowest Coulomb energy, as described previously.<sup>51</sup> Our simulation method was broadly based on a previously published methodology for solvent free simulations.<sup>52</sup> First, a steepest descent energy minimization was performed to eliminate unfavorable contacts and steric overlaps. This was directly followed by unrestrained vacuum simulations. Neither periodicity nor cutoffs were used during the calculations. A small integration time step of 1 fs was used for ensuring energy conservation, and bonds to H atoms were constrained using the LINCS algorithm. We used a dielectric constant of  $2\epsilon^\circ$  (where  $\epsilon^\circ$  is the dielectric permittivity of vacuum).

MD simulations were performed at 300 K. For the 7+ to 24+ charge states of ERD10, a process of simulated annealing was also incorporated that involved repeated heating to 800 K for 100 ps followed by gradual cooling back to 300 K over a 100 ps time scale. Similar procedures have been used previously to obtain global minima in gas-phase simulations.<sup>53</sup> Longer cooling times were found to increase the efficiency of structural collapse but had little effect on the degree of structural compaction. For the 30+ to 43+ charge states the process of simulated annealing was omitted as it was found to induce large bending oscillations in the rod-like structures. Example structures for each charge state were selected on the basis of the experimental  $\Omega$  of the ions.  $\Omega$  were obtained from protonated structures using the Exact Hard Sphere Scattering (EHSS) method implemented in MOBCAL.<sup>46</sup> Fully collapsed conformations were obtained from charge neutral species of single representation SAXS structures of each IDP using the simulated annealing process described above.

For the generation of vacuum ERD10 structures for scoring gas-phase structures against SAXS curve (using GAJOE), each charge state was simulated for 200 ns. Simulated annealing was applied in 60 cycles for the 0+, 7+, 10+, 12+, 18+ and 24+ charge states. Each cycle

involved instantaneous heating of the protein to 800 K for 100 ps followed by gradual cooling to 300 K over 100 ps and then a longer equilibration at 300 K for 1.8 ns. Simulated annealing was not applied for the 30+, 36+ and 43+ charge states. Each charge state was simulated 5-times and 110 structures were extracted from each trajectory to generate 550 structures at each ionization state. Structures were extracted from the 300 K equilibration phase for the 0+ to 24+ ions over the course of the simulation. For higher charge states structures were extracted at various time points that spanned the last 100 ns of each trajectory. To confirm no structural conversions on longer time scales, 2  $\mu$ s trajectories were also simulated for the 7+, 24, and 43+ charge states (Figures S5, S6).

**Assessing the Discriminating Power of SAXS/EOM.** A hybrid scattering curve of ERD10 was prepared using a volume-fraction weighted sum of the experimental and the 7+ theoretical curve with 10% contribution of the 7+ ion (collapsed structural model) in the linear combination (eq 1). The ability of SAXS to “see” the low population of compact states among the flexible species was demonstrated visually in the normalized Kratky plots of the hybrid SAXS data (Figure S7). The ability of GAJOE to detect collapsed subpopulations in the hybrid curve was also explored. EOM procedures were utilized against the modified SAXS profile using a conformational library of solution proteins generated by FM doped with 5% fully collapsed gas-phase structures. The ability of the EOM to find the collapsed states is shown clearly in the  $R_g$  intensity distributions (Figure S7). A complete gas-phase library encompassing all charge states and structures was also screened against the experimental SAXS curve of ERD10. The gas-phase library was prepared by extracting MD structures at various time points during the trajectories of the charge neutral, 7+, 10+, 12+, 18+, 24+, 30+, 36+ and 43+ ions. 550 structures were extracted from each trajectory resulting in a library of 4950 different conformations covering the full range of ionization states and  $\Omega$  of ERD10 (see previous section). Each structure was initially evaluated for goodness-of-fit to the experimental SAXS curve of the protein using CRYSTAL. The 4950 gas-phase structures were then combined and screened against the solution scattering curve of ERD10. This procedure confirmed that is was not possible to describe the experimental SAXS curve of ERD10 with any combination of gas-phase states. The gas-phase library was also combined with an equal number of solution states generated by FM to generate a hybrid library which was then used to evaluate the experimental scattering curve of ERD10. On repeated iterations this unbiased procedure excluded the gas-phase structures from optimized ensembles in favor of solution states generated by FM (Figure S7). These experimental controls demonstrate the power of SAXS and EOM modeling procedures to accurately describe the conformational space of a protein in solution as has also been previously defined elsewhere.<sup>3,4,42,54</sup>

#### Linear Deconvolution of IDP Charge State Distributions.

IDP subpopulations were extracted from the charge state distributions of each IDP according to previously published protocols using a limited number of Gaussian functions.<sup>13</sup> All IDP charge state distributions could be explained by  $\leq 3$  subpopulations relating to compact, intermediate and fully extended conformational families. The charge state signature ( $z_{ave}$ ) of each subpopulation was obtained from the centroid ( $\mu$ ) of the fits. Charge state signatures of model structures were obtained from their SASA using VADAR with a 1.4 Å water probe according to published relationships between protein  $z_{ave}$  SASA.<sup>32</sup>

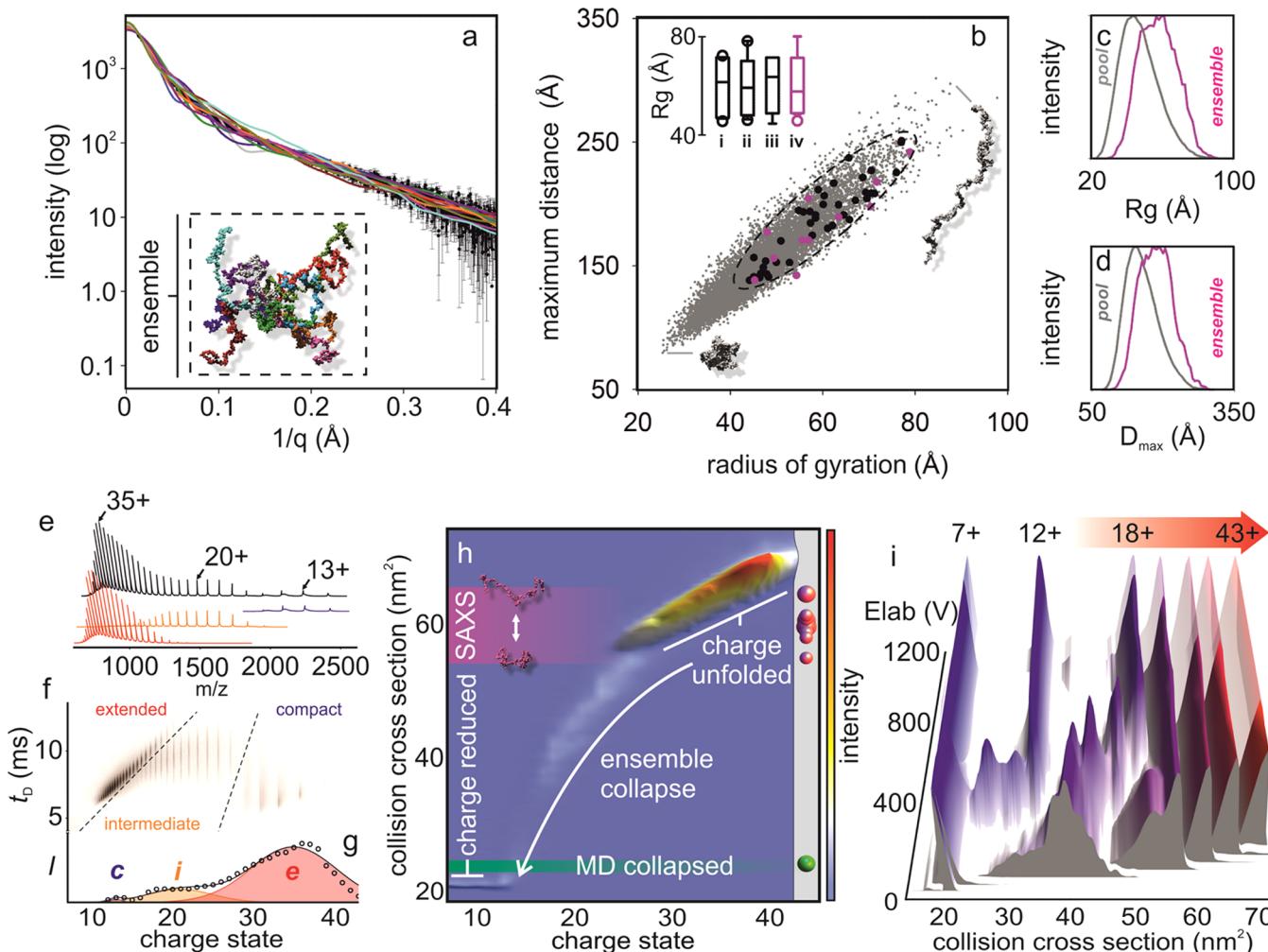
## RESULTS

**Solution and Gas-Phase Conformations of ERD10.** The solution ensemble of the IDP ERD10 was obtained by a combined SAXS-EOM strategy (Table S1). SAXS provides a weight-average description of protein structure, but recent developments in ensemble methods permit the scattering curves of IDPs to be interpreted in terms of protein flexibility. The EOM is the original and most used ensemble method.<sup>3</sup> It utilizes a strategy involving the generation of a large

conformational pool describing the potential space available to a protein from which a subensemble is selected based on the SAXS profile (eq 2). The subensemble is encoded by a limited number of structures which collectively define the conformational space of a protein in solution. Structural pools are generated by Flexible-Meccano (FM) with optimum sizes of ~10 000 different structures to cover a wide range of potential geometries that IDPs sample in realistic Ramachandran space as defined by coil libraries.<sup>43</sup> Theoretical SAXS curves are then generated for each structure and the optimum subensemble defined by a Monte Carlo approach utilized by a genetic algorithm implemented by GAJOE. SAXS-EOM strategies can be used in isolation or in hybrid workflows typically with Nuclear Magnetic Resonance (NMR), which can provide additional localized information complementary to SAXS. SAXS is especially sensitive to ensemble dimensions and has the capacity to identify subpopulations within an ensemble of disordered states with high resolution.<sup>5,54</sup> SAXS is preferential for this present study where defining the solution envelope is more meaningful given the coarse structural output of IMS. The protein ERD10 was carefully selected for initial study as this protein is predicted to be fully unstructured with a limited tendency for local structure.<sup>37</sup>

The SAXS curve of ERD10 was obtained in batch mode at DESY on the PETRAIII beamline and processed using the ATSAS package, FM and Sccomp (Experimental Section).<sup>42</sup> Ensembles were obtained using the procedure outlined for the EOM with 10 000 random conformers generated by FM. A subensemble was then selected by GAJOE based on the SAXS curve of the protein (Experimental Section). The SAXS curve of ERD10 describes a broad range of extended structures defining the conformational space occupied by this protein in solution (Figure 1a). The initial ensemble of ERD10 was encoded by 11 different structures with individual weights of  $n \leq 4$ . As expected, repeated optimizations resulted in the selection of different structures from the conformational pool. Collectively, however, each ensemble described the same overall properties with regard to the ensemble dimensions and yielded comparable high quality fits to the scattering data  $\chi = 0.66$  (Figure 1b). The intensity distributions of the ERD10 subensemble provide additional information relating to the existence of protein subpopulations. The distributions of ERD10 explain monomodal distributions of structures with no evidence for any preferential configurations and an unambiguous shift toward higher more elongated regions of the FM pool (Figure 1c,d). The ERD10 ensemble is described by structures that are scattered evenly around the conformational space of the protein (Figure 1b). This suggests that the maximum occupancy of each state, the fraction of time that ERD10 spends in each conformation, is low further supporting the absence of preferential species for this protein.<sup>55</sup> In solution ERD10 is characterized by a single broad population of structures occupying the more extended regions of the FM library of potential states.

ERD10 was then investigated by native-MS on a second generation Synapt HDMS (Experimental Section). The protein displays the characteristically broadened charge state distribution expected for a disordered protein which spans a large range of ionization states from the 12+ to 43+ charge states (Figure 1e). The gas-phase structures of ERD10 were obtained by Traveling Wave IMS (TWIMS) (Experimental Section). TWIMS rapidly sizes gas-phase ions based on their drift times as they are propelled through a mobility cell pressurized



**Figure 1.** Solution and gas-phase conformations of ERD10: (a) SAXS curve of ERD10 (black) overlaid with those from an optimized ensemble (colored lines) and their corresponding structures (hatched box). (b) FM-generated structural pool of ERD10 represented in  $R_g$  and  $D_{\max}$  conformational space (gray) example structures are shown at the periphery of the library of 10 000 conformations. The conformational space occupied by ERD10 is enclosed by the hatched ring. The black dots represent the sum of structures selected on 3 separate optimizations, and the pink dots are structures from a separate single optimization. The box plots show the spread of  $R_g$  values for subensembles optimized on four separate occasions as shown in the scatter plot. All subensembles optimized with low error  $\chi \sim 0.66$ . (c,d)  $R_g$  and  $D_{\max}$  distributions for ERD10 for the FM pool (gray) and optimized ensemble (taken from ensemble iv, (pink)) showing the monomodal distribution of extended species in solution. (e) Mass spectrum of ERD10 obtained on a Waters second generation Synapt HDMS (black) extracted mass spectra for the extended (red) intermediate (orange) and compact (blue) subpopulations taken directly from (f) are shown. (f) Mobiligram of ERD10 drift times encompassing 3 distinct regions as shown. (g) Linear deconvolution of the ERD10 charge state distribution displaying 3 subpopulations for this protein. (h) Comparison of the gas-phase and solution conformations of ERD10. Features relating to charge unfolded and charge reduced ions are highlighted. The  $\Omega$  of each structure within the solution ensemble (iv) of ERD10 is shown on the gray strip to the right of the figure as a bubble plot with bubble size representing the relative weight of each conformation within the optimized ensemble (pink). The SAXS ensemble is represented by the pink strip and example structures at the periphery are shown. The  $\Omega$  of the fully collapsed charge neutral species of ERD10 obtained by MD is also given (green). (i) Gas-phase unfolding of ERD10 showing  $\Omega$  as a function of the collision voltage are shown for the 7+, 12+, 18+, 24+, 30+, 36+ and 43+ ions, ranging from blue to red. The collision voltage is reported in laboratory frame (Experimental Section).  $\Omega$  and charge share a linear relationship at high energy, as shown. Data could not be obtained in the transparent regions due to loss of ion transmission (7+) or ion fragmentation (18+ to 43+).

with  $N_2$  by a traveling wave. Elongated structures interact more frequently with the  $N_2$  buffer gas and undergo more roll over events against the wave direction than compact shapes resulting in longer drift times for these species. Instrument calibration permits the  $\Omega$  of the ions to be obtained, which is a structural parameter related to their rotationally averaged projected area.<sup>56</sup> ERD10 forms monodisperse ion packets at low and high charge states with the Arrival Time Distributions (ATDs) becoming broader and more polydisperse at intermediate charge (Figure 1f). The charge state distribution and

mobiligram of ERD10 describes 3 distinct subpopulations that can be broadly defined as compact (c), intermediate (i) and extended (e) and which are centered on the 13+, 20+ and 35+ ionization states, respectively (Figure 1g). Near perfect reconstruction of the charge state distribution of ERD10 can be performed by direct extraction of these conformational regions from the mobiligram of ERD10 assuming each to be a Gaussian function (Experimental Section). Native-MS describes the conformational space of ERD10 by a wide range of charge states characterized by 3 distinct subspecies. These subpopu-

lations are not evident from the SAXS curve of the protein in spite of the known ability of SAXS to identify multimodal distributions of proteins with high accuracy.<sup>12,54,57,58</sup>

The  $\Omega$  of ERD10 were obtained and plotted against charge and intensity. The  $\Omega$  of the SAXS ensemble of the protein were also obtained to permit a direct comparison of the conformational space of the protein in both solution and gas-phase environments. The most remarkable feature of this comparison is the great extent to which the conformational space of ERD10 expands in the absence of solvent effects (Figure 1h). The broad range of solution structures defined by their  $R_g$  naturally appear narrower with respect to  $\Omega$  due to the weaker shape dependency of IMS relative to SAXS.<sup>59</sup> The solution ensemble spans structures from  $\sim 55 \text{ nm}^2$  to  $\sim 65 \text{ nm}^2$ , whereas IMS defines gas-phase structures from  $\sim 20 \text{ nm}^2$  to  $\sim 70 \text{ nm}^2$ . Gas-phase structures of ERD10 were then characterized further by employing various strategies of charge reduction and thermal activation by collision induced dissociation (CID). Charge stripping of ERD10 to generate the 7+ to 11+ ions of the protein resulted in no further decrease in  $\Omega$  indicating that the states at  $\sim 20 \text{ nm}^2$  represent fully collapsed gas-phase structures (Figure 1h, Experimental Section). The dominant lobe at high charge (subpopulation e) is characterized by a linear relationship between  $\Omega$  and charge indicative of charge unfolded states.<sup>19</sup> To confirm this assignment ERD10 was subjected to CID to induce Coulomb unfolding of the protein (Experimental Section). For charge states within the range of 12+ to 18+ this process results in ion extension via multiple states to a limiting  $\Omega$  where Coulomb effects are maximized. The linear relationship between  $\Omega$  and charge for ions  $\geq 24+$  is maintained at high energy confirming that the conformations of these species are fully extended commensurate with their net charge (Figure 1i). In the gas-phase ERD10 becomes fully compressed or stretched between limiting values defined by the charge state of the protein. The protein samples more polydisperse structures between these extremes but the  $\Omega$  of these species cannot be defined by the scattering curve of this protein. ERD10 displays all the qualitative features of disorder expected by ESI-MS. However, ESI-MS overrepresents the conformational space of ERD10 significantly and on direct comparison with SAXS the difference is most striking.

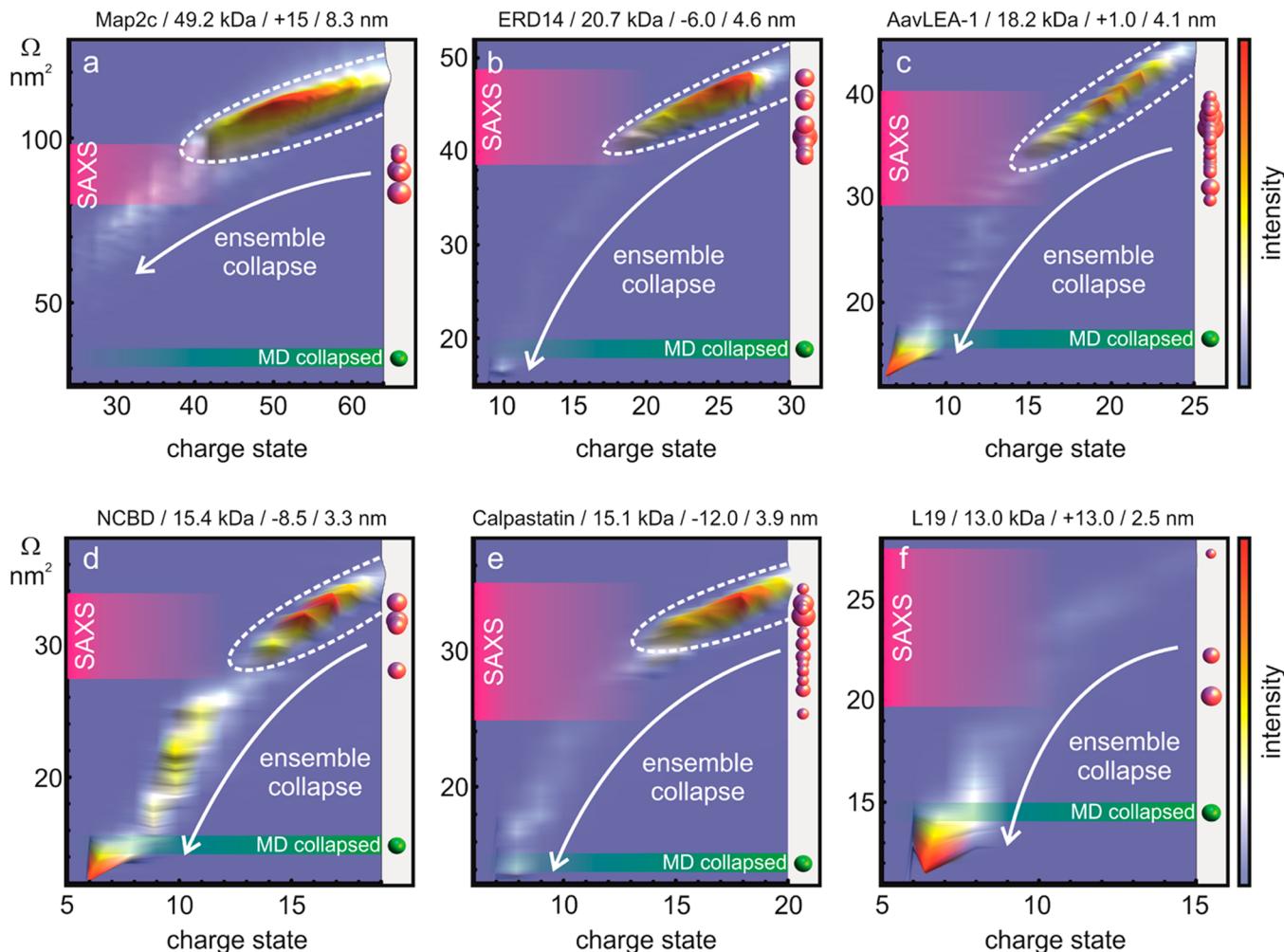
**IDPs Undergo a 3–5 Fold Conformational Space Expansion in the Gas-Phase.** The same approach outlined above for ERD10 was then employed to compare the solution and gas-phase conformation of a range of many different IDPs of different functional classes (Table S1, Figures S1 and S2). In all cases the same general trend observed for ERD10 is upheld involving a significant (3–5 fold) expansion of the preferred conformational regions of the proteins on transfer from solution to gas-phase (Figure 2). Of all the IDPs studied, L19 and NCBD become most compressed by gas-phase effects presumably due to the respective low disorder content (18%) and molten globule property of these proteins (Table S1). For the remaining IDPs the population of different gas-phase species vary significantly. In general the gas-phase structures of the IDPs cannot be easily reconciled by their solution ensembles. AavLEA1 for example populates  $\sim 50\%$  fully compact structures in complete contrast with the solution behavior of the protein where no traces of compact/spherical species are observed. Overall the gas-phase structures of the IDPs are characterized by a conformational space broadened significantly by charge unfolding and structural collapse. As with ERD10 limiting regions at high and low charge are

characterized by monodisperse structures which flank an intermediate region where the proteins display greater structural heterogeneity. However, the  $\Omega$  of these apparently more flexible and structurally diverse gas-phase conformations are too compact to reflect the conformational space of the proteins in solution.

**Vacuum MD Simulations of ERD10.** Vacuum MD simulations were then employed to gain more insight into the behavior of ERD10 in the absence of solvent (Experimental Section, SI). In brief, 3 structures were taken that spanned the range of conformations from SAXS ensemble of the protein. To each structure 8 different charge states were assigned (7+, 10+, 12+, 18+, 24+, 30+, 36+ and 43+) to cover the range of charge states observed by experiment. A charge neutral species was also prepared. Charge assignment was performed by neutralization of negatively charged amino acids assuming initial protonation states of all residues at pH 7.0. For each structure and ionization state of ERD10 a range of different charge isomers were permuted and production MD simulations performed on charge isomers with the lowest Coulomb energy, as described previously.<sup>51</sup>

ERD10 was studied with the GROMOS 53a6 force field implemented within GROMACS 4.5.5 (Experimental Section).<sup>49</sup> During the simulations no significant relationship between the initial solution structure and the trajectories of the protein was observed (Figure S3 and S4). Accordingly the behavior of the average structure of ERD10 is reported. In vacuum ERD10 spontaneously adopts a range of new structures that vary with increasing charge from fully collapsed spheres to elongated rod-like conformations (Figure 3a,b). Overall, the vacuum behavior of ERD10 can be summarized by two principle characteristics in relation to the gas-phase conformations observed by IMS. First, the vacuum rearrangement of ERD10 allows the protein to occupy completely the same broader conformational space observed by IMS, from  $\sim 20 \text{ nm}^2$  to  $\sim 70 \text{ nm}^2$ . A plot of  $\Omega$  against charge for the structures obtained by MS is in close agreement with the similar relationship obtained by experiment. The structural pool generated by FM, which defines the potential space available to the protein in solution prior to optimization, was also found to span  $\Omega$  significantly narrower than those obtained by MD or IMS (Figure 3c). Second, as the net charge of ERD10 increases, the protein exhibits an increased tendency to populate local minima during vacuum collapse. Presumably this behavior is due to charge repulsion in the collapsing chains, as a much smoother potential energy surface is apparent during the vacuum collapse of the charge neutral species of ERD10. Structures trapped in local minima are stable on the  $\mu\text{s}$  time scale but can be freed from local conformations to some degree by simulated annealing (Experimental Section, Figure S5). The tendency to populate local minima during structural collapse of ERD10 could be associated with the structural heterogeneity of gas-phase structures observed for this protein by IMS in the range of the 12+ to 24+ ions. Overall the  $\Omega$  of ERD10 obtained by MD are highly compatible to those obtained by IMS.

The gas-phase structures of ERD10 score poorly against the SAXS curve of the protein (Experimental Section). In particular the more compact conformations at low charge give rise to  $\chi$  values that fall outside of the range reported by CRYSTOL (Figure 3d). The ability of SAXS/EOM to judge the conformational space of a protein was also investigated. These experiments demonstrated the discriminating power of SAXS and the ability of the EOM to see different species in



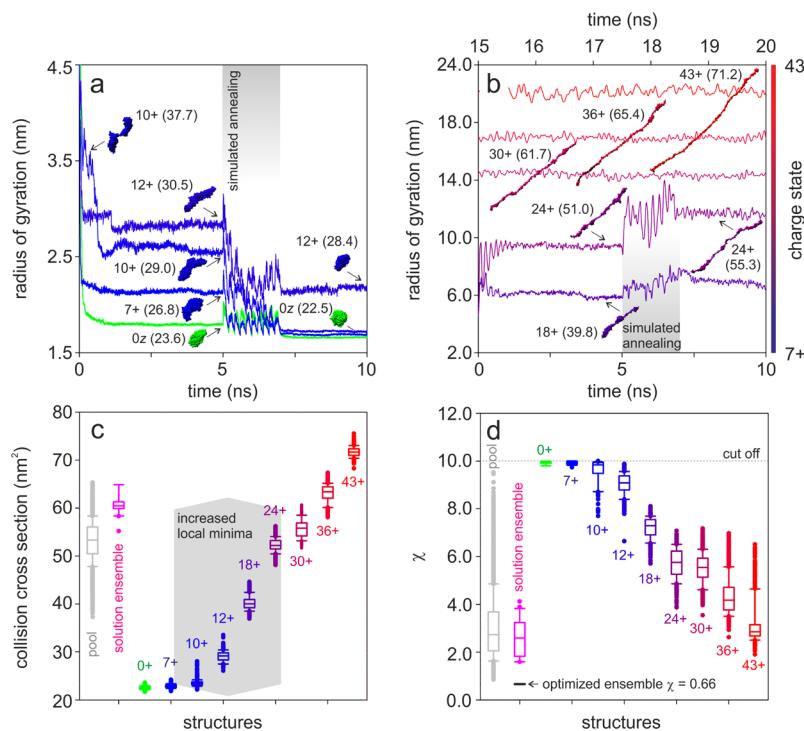
**Figure 2.** Comparison of the solution and gas-phase structures of IDPs: (a–f) Map2c, ERD14, AavLEA1, NCBD, Calpastatin and L19, respectively. Solution structures were obtained using SAXS and the EOM (Experimental Section). Experimental  $\Omega$  were obtained on a Waters second generation Synapt HDMS (Experimental Section). The  $\Omega$  of each structure with the solution ensemble of ERD10 is given on the gray strip to the right of the figure as a bubble plot with bubble size representing the relative weight of each conformation within the optimized ensemble (pink). The conformational space of each IDP defined by the SAXS ensemble is also represented by the pink strip to the left. The  $\Omega$  of a fully collapsed charge neutral MD structure obtained for each protein is given (green). Intense features for charge unfolded conformations are highlighted by a dashed line except for L19 where these conformations were below the limit of detection. The large extent of ensemble collapse is indicated. The respective molecular weights, predicted net charge in solution and weight-average  $R_g$  of each IDP are shown.

solution scattering data (Experimental Section, SI, Figure S7). In summary, SAXS and gas-phase methods describe IDP structures that contrast significantly with each other. Certainly, structural restraints for IDPs that are obtained by gas-phase methods cannot be used to describe the conformations of these proteins in solution.

**ESI Imposes Additional Subpopulations in IDPs.** The  $R_g$  distributions generated by the EOM provide important information regarding preferential states within the conformational space of IDPs.<sup>54</sup> ESI-MS charge state distribution are also routinely used to access this information from which the weights of different subpopulations can be quantified by linear deconvolution. The  $R_g$  distributions of ERD10 derived from SAXS describe the conformational space of this protein as unimodal. This is also supported by the random scattering of structures within the ensemble of ERD10 (Figure 1b). This definition of the behavior of ERD10 in solution contrasts significantly with that described by ESI-MS where *c*, *i* and *e* subpopulations are clearly visualized (Figure 1e–g). Since these gas-phase subpopulations are separated on the basis of their

charge their generation must be related to the ESI process. Otherwise the single distribution of shapes found in solution as defined by SAXS would be mirrored in the charge state distribution of the protein. The conformational changes incurred during the solution to gas-phase transfer of IDPs are apparently coupled to events that transpire during protein ionization.

The relationship between charge and shape is frequently interpreted by the charged residue mechanism (CRM) where proteins remain entombed in electrospray droplets and charge transfer occurs at the last stages of droplet dehydration.<sup>29,30,32</sup> Protein charge signatures ( $z_{\text{ave}}$ ) depend on the number of charges that can be maintained by an ESI droplet of equivalent size and hence the solvent accessible surface area (SASA) of a protein. The  $z_{\text{ave}}$  of proteins that follow the CRM have been defined empirically for globular proteins and their complexes.<sup>29,30,32</sup> These relationships underpin the use of ESI-MS to characterize the conformations of proteins with more open structures, where increasing charge signatures are interpreted by the associated increase in SASA thereby enabling charge to



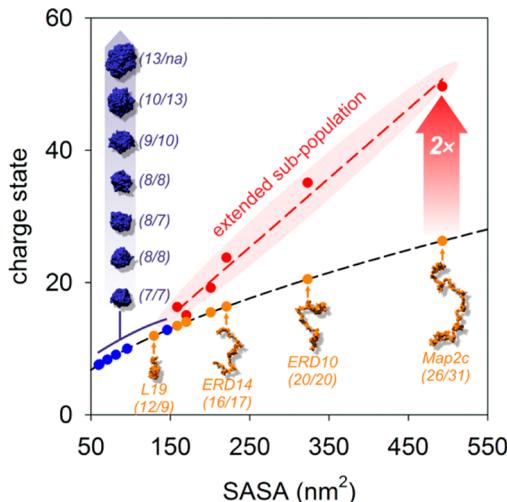
**Figure 3.** Vacuum MD of ERD10 and compatibility with SAXS: (a,b) Vacuum MD trajectories of the 0+, 7+, 10+ and 12+ charge states (a) and the 18+, 24+, 30+, 36+ and 43+ charge states (b) of ERD10. Selected structures along the trajectories are given along with their  $\Omega$  ( $\text{nm}^2$ ). Time points where simulated annealing was applied is highlighted. The time points for the 30+, 36+ and 43+ charge states is given in the upper x-axis. (c)  $\Omega$  of the FM-library (pool), GAJOE-optimized ensemble and MD structures of ERD10 at different charge states. Each charge state contains 550 different states obtained by MD incorporating multiple rounds of simulated annealing to find low energy structures. The region in which the ions/trajactories exhibit an increased tendency to populate local minima in the absence of simulated annealing is highlighted. (d) Compatibility ( $\chi$ ) of the FM-library, GAJOE-optimized ensemble and MD structures of ERD10 with the SAXS curve of the protein. The cutoff at which CRYSTOL can report errors ( $\chi = 10.0$ ) is shown. Differential weighting of structures from the solution ensemble (pink) provides the best fit to the experimental data ( $\chi = 0.66$ ) as reported in the optimized ensemble (black).

be used as a restraint for structural modeling<sup>14,31,24</sup>. More recently the ionization of flexible species has been described by extrusion or the chain ejection mechanism (CEM), the latter of which will be referred to from here on due to its particular relevance to the behavior of proteins.<sup>27,28</sup> In the CEM charge transfer drives the ejection of highly charged proteins from ESI droplets prior to the latter stages of droplet dehydration. Protein ejection confounds canonical charge-structure relations for IDPs as the  $z_{\text{ave}}$  of these proteins is also influenced strongly by variables extraneous to their SASA such as the size and charge of the electrospray droplet from which they emerge. The idea common to either process is that ESI charge state distributions reflect the conformational space that is sampled by a protein in solution. The large discrepancies in the size and features of the conformational space of IDPs in solvent and gas-phase presently reported suggest a more complex ionization behavior for these proteins.

To understand these effects more closely the  $z_{\text{ave}}$  of all IDP subpopulations were extracted by linear deconvolution of their ESI-MS charge state distributions (Experimental Section, Figure S8 and Table S2). As with ERD10 the ESI-MS spectra of the IDPs describe *c*, *i* and *e* subpopulations whereas the SAXS profiles generally define the proteins as unimodal. The  $z_{\text{ave}}$  of each IDP ensemble was then obtained according to accepted charge-structure relationships (Experimental Section). The charging behavior of the IDPs is clearly varied between the different subpopulations (Figure 4). The  $z_{\text{ave}}$  of the intermediate IDP subpopulations are in reasonable agreement

with values predicted for the solution conformations of the proteins. In contrast the  $z_{\text{ave}}$  of the extended IDP subpopulations exceed values predicted for their corresponding solution states by  $\leq 2$ -fold. In some instances ions reach the maximum charge of the proteins including ionization of the carboxamide side chains (Table S1). The  $z_{\text{ave}}$  of the compact IDP subpopulations cannot be explained by the solution structures of the proteins. The IDP geometries associated with these subpopulations represent fully collapsed structures that are not evident from the SAXS curves of the proteins. The experimental  $z_{\text{ave}}$  of the compact IDP subpopulations were then compared to values predicted for their corresponding fully collapsed structures. On comparison the degree of close parity is evident with the experimental and calculated charge signatures falling within a single charge state for the majority of the proteins. The compact IDP subpopulations are not only characterized by fully collapsed shapes absent in solution but also possess the charge state signatures expected for globular proteins of equivalent size.

IDP subpopulations visualized in the charge state distributions of the proteins are clearly not associated with preformed protein conformers in solution as these subpopulations would be visualized in the  $R_g$  distributions obtained by SAXS. The charge signatures of the ESI-MS subpopulations are also highly varied and are difficult to encapsulate by a single ionization process defined by their SASA or solution geometries. The present data suggest that the solution to gas-phase transfer of IDPs is governed by processes that have competing effects on



**Figure 4.** Bifurcated ESI mechanisms in IDPs: Comparison of experimental and calculated charge signatures of IDPs. The experimental charge signatures were obtained by linear deconvolution of the ESI-MS charge state distributions of each IDP (Experimental Section, Figure S8, and Table S2) Theoretical charge signatures were obtained for the average solution structure of each IDP along with their corresponding fully collapsed structure obtained by vacuum MD simulations. No significant difference in predicted charge signatures was observed across the structures within each IDP ensemble. Charge signatures were calculated according to  $z_{\text{ave}} = 0.671 \times \text{SASA}^{0.592}$  dashed line.<sup>32</sup> The structures of the fully collapsed conformations of each IDP (blue) are shown with increasing mass along with their respective predicted and observed charge signatures given in parentheses. Example solution structures are shown for L19, ERD14, ERD10 and Map2c (orange) along with their respective predicted and observed charge state signatures in parentheses. The observed  $z_{\text{ave}}$  of the extended subpopulation of each IDP are plotted against the SASA of their respective solution states (red). Observed charge states of the extended subpopulations increase linearly with SASA according to  $z_{\text{ave}} = 0.105 \times \text{SASA} - 0.821$ ,  $R^2 = 0.99$ .

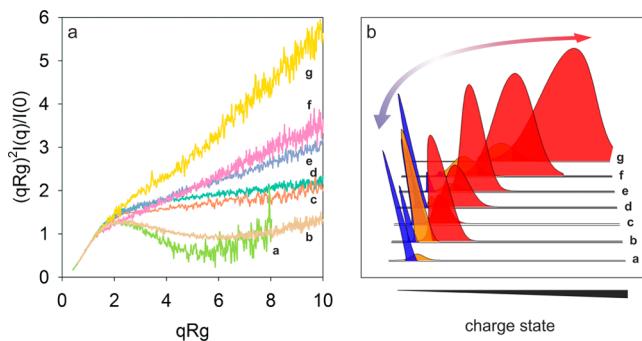
the geometries and conformational space of the IDPs. These processes give rise to additional subpopulations not apparent in solution but which are visualized on the basis of their charge by ESI-MS.

## ■ DISCUSSION

Ensemble methods have been utilized to carry out the most extensive comparison of the solution and gas-phase structures of IDPs conducted to date. SAXS is invaluable in understanding the solution to gas-phase transfer of IDPs as it provides access to information that is compatible with that of ESI-MS and IMS on ensemble subpopulations and dimensions, respectively. This information is encoded by potential structures occupying the IDP ensembles which can be extracted and studied individually in vacuum by MD simulations. SAXS structures also allow the interrogation of IDP  $z_{\text{ave}}$  enabling the formulation of hypotheses relating to the ionization of these proteins. The transferal of IDPs from solution to gas-phase by ESI-MS is associated with overwhelming changes to the structures of these proteins. These effects are broadly defined by a 3–5 fold increase in the conformational space of these proteins and the generation of additional subpopulations during ESI. The present data are not in favor of depictions of the behavior of IDPs where gas-phase restraints are interpreted directly with respect to the solution structures of these proteins.

ESI contributes most significantly to the gas-phase rearrangement of IDP structure. The unimodal spread of IDP configurations observed in solution is reported as 3 distinct subpopulations on the basis of their charge. These gas-phase subpopulations exhibit a range of  $z_{\text{ave}}$  that cannot be easily explained by a single ionization process. Contrary to predictions expected from current extrusion/ejection mechanisms our data suggest that IDPs can remain entrapped in electrospray droplets to follow ionization processes expected for globular proteins according to their SASA. IDPs entrapped in electrospray droplets undergo complete collapse, presumably due to vanishing solvent to give rise to structures with  $z_{\text{ave}}$  identical to those expected for globular proteins of equivalent mass. The intermediate and extended IDP subpopulations identified by ESI exhibit a bifurcation in  $z_{\text{ave}}$  as predicted by the CRM. The extended subpopulations tend toward the maximum charge state possible for the IDPs and exhibit  $z_{\text{ave}}$  exceeding values predicted by the CRM by as much as 2-fold. This extent of protein ionization has not been previously described for IDPs and could represent signatures anticipated for ionization mechanisms according to ejection/extrusion processes. The origins of the intermediate IDP subpopulations are difficult to define but could reflect a range of residual moderately extended structures populated in local minima during protein collapse, as observed for ERD10 during vacuum MD (Figure 3a,b). IDPs with intermediate charge signatures display a characteristic degree of structural heterogeneity not observed for ions at high or low charge, the conformations of which are monodisperse within any given charge state (Figure 1i). It is tempting to relate this property directly with the dynamic structure of the IDPs in solution but the  $\Omega$  of these ions are significantly more compact than predicted from their scattering curves. The apparent gas-phase flexibility of IDPs with intermediate charge could simply reflect frustration in the gas-phase energy surfaces of these ions. This may give rise to multiple states unable to satisfy the mutually exclusive requirements for increasing intramolecular contacts while reducing electrostatic repulsion. We propose that subpopulations apparent in ESI-MS charge state distributions of IDPs reflect an array of different ionization effects rather than preformed states as previously defined for these proteins. These ionization effects fully stretch or compress the IDPs to bring about the conformational space expansion of the proteins. Variations in ESI processes have been recently reported for other flexible systems such as polyethylene glycol (PEG).<sup>60</sup>

ESI-MS does not capture the solution configurations of IDPs. But the IDP subpopulations generated by ESI provide a metric allowing clear differentiation between the IDPs and this must be related to some chemical aspects of the proteins. The IDP Kratky plots show large differences in the internal structure of the proteins (Figure 5a). These plots classify the IDPs with globular aspects to their structure (L19 and NCBD) those that behave like freely jointed chains with low persistence length (Calpastatin and AavLEA1) and IDPs with increasing chain stiffness (ERD10, ERD14 and Map2c). There is a tendency for IDPs with higher persistence length to favor higher charge states and extrusion while more flexible or globular type IDPs tend toward collapse and the CRM (Figure 5). This demonstrates the involvement of additional factors such as internal structure in defining the charge and shapes of IDPs in the gas-phase. A range of different variables such as net charge, size, surface activity, polymer stiffness and shape potentially combine to define the gas-phase properties of IDPs. Future



**Figure 5.** Comparison of IDP Kratky plots and ESI subpopulations: (a) Normalized Kratky plots of the IDPs (a–g) for L19, NCBD, Calpastatin, AavLEA1, ERD10, ERD14 and Map2c, respectively. The bell shaped curves at a  $q$ -value of  $\sim 2$  for L19 (a) and NCBD (b) reflect some globular aspects in protein structure for these IDPs due to the respective low disorder predictions (18%) and the molten globule property of these proteins. Large differences in rise of the curves at high  $q$ -values are seen between the remaining IDPs and can be interpreted as differences in the persistence length of the proteins which increases over 5-fold between Calpastatin (c) and Map2c (g). (b) IDP subpopulations extracted from the ESI-MS charge state distributions by linear deconvolution. The subpopulations are colored blue, orange and red for the compact, intermediate and extended subpopulations, respectively. Note the general relationship between the population of the different ESI-MS species and the internal structure of the proteins according to their Kratky plots.

experiments should focus on understanding these aspects of IDPs in relation to their gas-phase structures. This may bring to light new perspectives on the function of IDPs and electrospray mechanisms in general.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/jacs.5b06027](https://doi.org/10.1021/jacs.5b06027).

Figures S1–S8 and Tables S1–S2. ([PDF](#))

## AUTHOR INFORMATION

### Corresponding Authors

\*[antoniborysik@kcl.ac.uk](mailto:antoniborysik@kcl.ac.uk)

\*[ptompa@vub.ac.be](mailto:ptompa@vub.ac.be)

### Author Contributions

†AJB and DK contributed equally.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Drs. Pierre Lebrun and Cy Jefferies for useful discussions on SAXS and evaluation of the data and Drs. Lorna Smith and Zoe Hall for useful discussion with vacuum MD. We are grateful for the measurement time granted by Soleil synchrotron committee, on the SWING beamline, as well as on the PETRAIII beamline at DESY, and Guy Vandenbussche (Structure and Function of Biological Membranes, Free University Brussels (ULB)) for the MS analyses and Karolien Van Belle for the excellent technical assistance. This work was supported by the Odysseus grant G.0029.12 from the Research Foundation Flanders (FWO) to PT, MG is supported by a VIB International postdoctoral (omics@VIB) fellowship and DK is supported by a FWO postdoctoral fellowship.

## REFERENCES

- (1) Tompa, P. *Nat. Chem. Biol.* **2012**, *8*, 597.
- (2) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *Annu. Rev. Biophys.* **2008**, *37*, 215.
- (3) Bernado, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. *J. Am. Chem. Soc.* **2007**, *129*, 5656.
- (4) Bernado, P.; Svergun, D. I. *Methods Mol. Biol.* **2012**, *896*, 107.
- (5) Sibilie, N.; Bernado, P. *Biochem. Soc. Trans.* **2012**, *40*, 955.
- (6) Tompa, P.; Varadi, M. *Structure* **2014**, *22*, 177.
- (7) Varadi, M.; Kosol, S.; Lebrun, P.; Valentini, E.; Blackledge, M.; Dunker, A. K.; Felli, I. C.; Forman-Kay, J. D.; Kriwacki, R. W.; Pierattelli, R.; Sussman, J.; Svergun, D. I.; Uversky, V. N.; Vendruscolo, M.; Wishart, D.; Wright, P. E.; Tompa, P. *Nucleic Acids Res.* **2014**, *42*, D326.
- (8) Schwalbe, M.; Ozanne, V.; Bibow, S.; Jaremkow, M.; Jaremkow, L.; Gajda, M.; Jensen, M. R.; Biernat, J.; Becker, S.; Mandelkow, E.; Zweckstetter, M.; Blackledge, M. *Structure* **2014**, *22*, 238.
- (9) Mittag, T.; Marsh, J.; Grishaev, A.; Orlicky, S.; Lin, H.; Sicheri, F.; Tyers, M.; Forman-Kay, J. D. *Structure* **2010**, *18*, 494.
- (10) Rubio-Cosials, A.; Sidow, J. F.; Jimenez-Menendez, N.; Fernandez-Millan, P.; Montoya, J.; Jacobs, H. T.; Coll, M.; Bernado, P.; Sola, M. *Nat. Struct. Mol. Biol.* **2011**, *18*, 1281.
- (11) Petoukhov, M. V.; Vicente, J. B.; Crowley, P. B.; Carrondo, M. A.; Teixeira, M.; Svergun, D. I. *Structure* **2008**, *16*, 1428.
- (12) Kim, H. S.; Wilce, M. C.; Yoga, Y. M.; Pendini, N. R.; Gunzburg, M. J.; Cowieson, N. P.; Wilson, G. M.; Williams, B. R.; Gorospe, M.; Wilce, J. A. *Nucleic Acids Res.* **2011**, *39*, 1117.
- (13) Dobo, A.; Kaltashov, I. A. *Anal. Chem.* **2001**, *73*, 4763.
- (14) Frimpong, A. K.; Abzalimov, R. R.; Uversky, V. N.; Kaltashov, I. A. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 714.
- (15) Brocca, S.; Testa, L.; Sobott, F.; Samalikova, M.; Natalello, A.; Papaleo, E.; Lotti, M.; De Gioia, L.; Doglia, S. M.; Alberghina, L.; Grandori, R. *Biophys. J.* **2011**, *100*, 2243.
- (16) Bernstein, S. L.; Liu, D.; Wyttenbach, T.; Bowers, M. T.; Lee, J. C.; Gray, H. B.; Winkler, J. R. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1435.
- (17) Jurneczko, E.; Cruickshank, F.; Porrini, M.; Nikolova, P.; Campuzano, I. D.; Morris, M.; Barran, P. E. *Biochem. Soc. Trans.* **2012**, *40*, 1021.
- (18) Illes-Toth, E.; Dalton, C. F.; Smith, D. P. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1346.
- (19) Shelimov, K. B.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997**, *119*, 2987.
- (20) Jurneczko, E.; Cruickshank, F.; Porrini, M.; Clarke, D. J.; Campuzano, I. D.; Morris, M.; Nikolova, P. V.; Barran, P. E. *Angew. Chem., Int. Ed.* **2013**, *52*, 4370.
- (21) Pacholarz, K. J.; Porrini, M.; Garlish, R. A.; Burnley, R. J.; Taylor, R. J.; Henry, A. J.; Barran, P. E. *Angew. Chem., Int. Ed.* **2014**, *53*, 7765.
- (22) Zhou, M.; Robinson, C. V. *Curr. Opin. Struct. Biol.* **2014**, *28*, 122.
- (23) Beveridge, R.; Covill, S.; Pacholarz, K. J.; Kalapothakis, J. M.; MacPhee, C. E.; Barran, P. E. *Anal. Chem.* **2014**, *86*, 10979.
- (24) D'Urzo, A.; Konijnenberg, A.; Rossetti, G.; Habchi, J.; Li, J.; Carloni, P.; Sobott, F.; Longhi, S.; Grandori, R. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 472.
- (25) Pagel, K.; Natan, E.; Hall, Z.; Fersht, A. R.; Robinson, C. V. *Angew. Chem., Int. Ed.* **2013**, *52*, 361.
- (26) Saikusa, K.; Kuwabara, N.; Kokabu, Y.; Inoue, Y.; Sato, M.; Iwasaki, H.; Shimizu, T.; Ikeguchi, M.; Akashi, S. *Analyst* **2013**, *138*, 1441.
- (27) Konermann, L.; Rodriguez, A. D.; Liu, J. *Anal. Chem.* **2012**, *84*, 6798.
- (28) Chung, J. K.; Consta, S. *J. Phys. Chem. B* **2012**, *116*, 5777.
- (29) de la Mora, J. F. *Anal. Chim. Acta* **2000**, *406*, 93.
- (30) Kaltashov, I. A.; Mohimen, A. *Anal. Chem.* **2005**, *77*, 5370.
- (31) Testa, L.; Brocca, S.; Grandori, R. *Anal. Chem.* **2011**, *83*, 6459.
- (32) Hall, Z.; Robinson, C. V. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1161.

- (33) Tompa, P. *Trends Biochem. Sci.* **2012**, *37*, 509.
- (34) Szasz, C. S.; Alexa, A.; Toth, K.; Rakacs, M.; Langowski, J.; Tompa, P. *Biochemistry* **2011**, *50*, 5834.
- (35) Wendt, A.; Thompson, V. F.; Goll, D. E. *Biol. Chem.* **2004**, *385*, 465.
- (36) Kiss, R.; Kovacs, D.; Tompa, P.; Perczel, A. *Biochemistry* **2008**, *47*, 6936.
- (37) Kovacs, D.; Kalmar, E.; Torok, Z.; Tompa, P. *Plant Physiol.* **2008**, *147*, 381.
- (38) Semrad, K.; Green, R.; Schroeder, R. *RNA* **2004**, *10*, 1855.
- (39) Goyal, K.; Tisi, L.; Basran, A.; Browne, J.; Burnell, A.; Zurdo, J.; Tunnacliffe, A. *J. Biol. Chem.* **2003**, *278*, 12977.
- (40) SWING Beamline, <http://www.synchrotron-soleil.fr/Recherche/LignesLumiere/SWING>, accessed April 17, 2013.
- (41) Petoukhov, M. V.; Franke, D.; Shkumatov, A. V.; Tria, G.; Kikhney, A. G.; Gajda, M.; Gorba, C.; Mertens, H. D. T.; Konarev, P. V.; Svergun, D. I. *J. Appl. Crystallogr.* **2012**, *45*, 342.
- (42) Konarev, P. V.; Petoukhov, M. V.; Volkov, V. V.; Svergun, D. I. *J. Appl. Crystallogr.* **2006**, *39*, 277.
- (43) Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J. R.; Jensen, M. R.; Segard, S.; Bernardo, P.; Charavay, C.; Blackledge, M. *Bioinformatics* **2012**, *28*, 1463.
- (44) Eyal, E.; Najmanovich, R.; McConkey, B. J.; Edelman, M.; Sobolev, V. *J. Comput. Chem.* **2004**, *25*, 712.
- (45) Varadi, M.; Kosol, S.; Lebrun, P.; Valentini, E.; Blackledge, M.; Dunker, A. K.; Felli, I. C.; Forman-Kay, J. D.; Kriwacki, R. W.; Pierattelli, R.; Sussman, J.; Svergun, D. I.; Uversky, V. N.; Vendruscolo, M.; Wishart, D.; Wright, P. E.; Tompa, P. *Nucleic Acids Res.* **2014**, *42*, D326.
- (46) Shvartsburg, A. A.; Jarrold, M. F. *Chem. Phys. Lett.* **1996**, *11*, 86.
- (47) Hernandez, H.; Robinson, C. V. *Nat. Protoc.* **2007**, *2*, 715.
- (48) Ruotolo, B. T.; Benesch, J. L.; Sandercock, A. M.; Hyung, S. J.; Robinson, C. V. *Nat. Protoc.* **2008**, *3*, 1139.
- (49) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. J. *Comput. Chem.* **2005**, *26*, 1701.
- (50) Cino, E. A.; Wong-ekkabut, J.; Karttunen, M.; Choy, W. Y. *PLoS One* **2011**, *6*, e27371.
- (51) Patriksson, A.; Marklund, E.; van der Spoel, D. *Biochemistry* **2007**, *46*, 933.
- (52) Hall, Z.; Politis, A.; Bush, M. F.; Smith, L. J.; Robinson, C. V. *J. Am. Chem. Soc.* **2012**, *134*, 3429.
- (53) Liu, D.; Wyttenbach, T.; Carpenter, C. J.; Bowers, M. T. *J. Am. Chem. Soc.* **2004**, *126*, 3261.
- (54) Tria, G.; Mertens, H. D.; Kachala, M.; Svergun, D. I. *IUCrJ* **2015**, *2*, 207.
- (55) Bertini, I.; Giachetti, A.; Luchinat, C.; Parigi, G.; Petoukhov, M. V.; Pierattelli, R.; Ravera, E.; Svergun, D. I. *J. Am. Chem. Soc.* **2010**, *132*, 13553.
- (56) Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2401.
- (57) Bertini, I.; Fragai, M.; Luchinat, C.; Melikian, M.; Mylonas, E.; Sarti, N.; Svergun, D. I. *J. Biol. Chem.* **2009**, *284*, 12821.
- (58) Gatzeva-Topalova, P. Z.; Warner, L. R.; Pardi, A.; Sousa, M. C. *Structure* **2010**, *18*, 1492.
- (59) Hewitt, D.; Marklund, E.; Scott, D. J.; Robinson, C. V.; Borysik, A. J. *J. Phys. Chem. B* **2014**, *118*, 8489.
- (60) Larriba, C.; Fernandez de la Mora, J.; Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1332.