

A semi-empirical framework for interpreting traveling wave ion mobility arrival time distributions

Sugyan M. Dixit and Brandon T. Ruotolo*

Department of Chemistry, University of Michigan, Ann Arbor, MI 48109

*Address reprint requests to **Brandon T. Ruotolo**, address: 930 N. University Ave, Ann Arbor, MI 48109, phone number: (+1) 734-763-2443, fax number: (+1) 734-615-3718, email address: bruotolo@umich.edu

Abstract

The inherent structural heterogeneity of biomolecules is an important biophysical property that is essential to their function, but is challenging to characterize experimentally. We present a workflow that rapidly and quantitatively assesses the conformational heterogeneity of peptides and proteins in the gas phase using traveling wave ion mobility (TWIM) arrival time distributions (ATDs). We have established a set of semi-empirical equations that model the TWIM ATD peak width and resolution across a wide range of wave amplitudes (V) and wave velocities (v). In addition, a conformational broadening parameter, δ , can be extracted from this analysis that reports on the contribution of conformational heterogeneity to the broadening of TWIM ATD peak width during ion mobility separation. We use this δ value to evaluate the conformational heterogeneity of a set of helical peptides, and our analysis correlates well with previous peak width observations reported for these ions. Furthermore, we use molecular dynamics simulations to independently investigate the general flexibility of these peptides in the gas phase, and generate similar trends found in experimental TWIM data. Finally, we extended our analysis to Avidin, a 64 kDa homotetramer, and quantify the structural heterogeneity of this intact complex using TWIM ATD data as a function of cross-linking. We observe an initial reduction in δ values as a function of cross-linker concentration, demonstrating the sensitivity of our δ value analysis to changes in flexibility of the assembly.

Introduction

The functions of biomolecules are inherently linked to their structures, making their study critical for wide-ranging research efforts in biochemistry and human disease[1]. Ultimately, protein functions depend upon discrete motions on the atomic scale, giving rise to structural ensembles that are responsible for carrying out various cellular processes[2]. As such, the quantitative assessment of structural ensembles is critically important in understanding the mechanistic details of biomolecular function. For example, proteins adapt their structure to many different binding partners, and therefore, can exhibit large conformational heterogeneity[3]. In addition, a significant fraction of the proteins in eukaryotes contain disordered regions that are involved in many important biophysical processes, but are currently insufficiently understood[4]. The prevalence of dynamics and flexibility in our understanding of protein biophysics has stimulated the development of many novel analytical techniques and computational modeling tools aimed at the detailed assessment of proteins in motion[5].

Despite the importance of dynamic motion in protein function, its quantification and characterization has remained a challenge for biophysical measurement techniques for decades. Nuclear magnetic resonance (NMR) spectroscopy is able to probe protein movements on timescales ranging from nanoseconds to milliseconds with atomic resolution[6]. In addition, small angle X-ray scattering (SAXS) measurements have more recently begun to provide information on protein dynamics[7]. In conjunction with experimental data, ensembles of structures have been generated computationally in order to study structural microstates and functional disorder in proteins[8, 9]. Despite technical advancements, however, the experimental techniques described above require pure, monodisperse, high concentration samples, which severely limit the biomolecular ensembles that can be probed. Furthermore, the computational sampling of protein dynamics remains challenging due to the

difficulties in extending simulation times to match those relevant for most biological processes and a general inability to completely account for configurational entropy in such simulations[10].

By virtue of soft ionization techniques that enable the introduction of solvent-free biomolecular structures in the gas phase[11], mass spectrometry (MS) methods such as native MS[12–14], tandem MS in combination with ion activation methods[15–17], hydrogen deuterium exchange (HDX) MS[18, 19], and chemical cross-linking (CXL) MS[20, 21] have been used to study the structure of proteins and protein complexes using small amounts of sample[22]. Ion mobility (IM)-MS is a structural MS method currently undergoing a period of rapid development, capable of separating protein ions according to their orientationally-averaged size on the millisecond timescale[23]. Several types of IM separators have been coupled to MS, each having their own strengths and weaknesses[24, 25]. For instance, drift tube IM (DTIM) works by introducing a time-defined packet of ions into a chamber containing both a weak electric field (E) and an inert buffer gas[26]. Under typical DTIM conditions, ion arrival times are directly proportional to their orientationally-averaged collision cross sections (CCSs), values which serve as coarse-grained structural restraints in biomolecular modeling efforts[27, 28]. DTIM arrival time distributions (ATDs) are well characterized by theory, and have previously been used to assess the structural heterogeneity of biomolecules in the gas phase[29–33]. The characterization of such heterogeneity can report on the ensemble of biomolecular structures in solution, as well as the structural heterogeneity of such systems in the gas phase. Despite this, traveling wave ion mobility (TWIM)[34–36], which uses time-varying electric fields to achieve IM separations, is the most prevalent form of IM-MS for work in the area of structural biology[25]. In contrast to DTIM, assessments of TWIM peak widths in an effort to elucidate the structural heterogeneity of biomolecules, as has been done previously in DTIM measurements, remains challenging due to our currently incomplete understanding of ion transport within TWIM analyzers [37–40].

In this report, we construct a semi-empirical model that describes the widths of TWIM ATDs across a wide range of TWIM parameters, and is capable of predicting ATD peak widths for nominally mono-conformational biomolecules for use in assessing the structural polydispersity of biomolecules generally. By using a group of model peptides known to have rigid structures in the gas phase, we tested our empirical expression and detected a subtle structural transition in the peptides as a function of sequence length. Furthermore, we correlated our IM-MS measurements with molecular dynamics (MD) simulations, observing strong correlations between computed structural ensembles and our experimental TWIM ATDs. Finally, we applied our model to analyze the TWIM ATDs of unmodified and cross-linked Avidin ions detecting shifts in protein complex peak widths in a manner correlated with the attachment of CXL agents. We conclude by projecting the general utility of quantitative TWIM peak width assessments for gas-phase structural biology.

Experimental Section

Chemicals and Materials

Ac-Ala_n-Lys peptides were custom made from Anaspec, CA. The peptides were dissolved in 90% TFA (Fisher Scientific, O4901) in water to achieve a final concentration of 1 mg/ml for nano-ESI analysis. Avidin (Sigma-Aldrich, A9275) was prepared in 200mM ammonium acetate at a final concentration of 50 μ M (Sigma-Aldrich, 09689). BS3 (Thermo Scientific Pierce, PI-21580) was freshly prepared in HEPES (Sigma-Aldrich, H3375) at pH 7.3 before CXL experiments. DL-polyalanine (Sigma-Aldrich, P9003) was used at a final concentration of 1mg/ml in 49.5/49.5/1 ratio of water/acetonitrile (Fisher Scientific, A9984)/acetic acid (Fisher Scientific, A38212).

Chemical Cross-linking

Avidin samples were buffer exchanged into 200mM HEPES, pH 7.3, using micro biospin 6 columns (Bio-rad, CA). BS3 was added to these samples in the following protein to cross-linking reagent

ratios: 1:1, 1:50, 1:150, 1:500, 1:1000, 1:1500, and 1:2000. After incubating the samples for 30 minutes at room temperature, the reaction was quenched by buffer exchanging the cross-linked samples into 200mM ammonium acetate.

TWIM-MS

All data was collected on a Synapt G2 TWIM-MS instrument (Waters, Milford MA). Instrumentation details can be found elsewhere[41]. Briefly, ions were generated using nano ESI, and then pulsed into the TWIM cell with 100 μ s gate pulse width. The TWIM cell is comprised of a stacked ring ion guide (SRIG), where direct current (DC) voltage is applied to two pairs of ring electrodes in a repeating pattern throughout the cell. A series of DC pulses generates a time varying potential, defined by its wave amplitude (**V**) and wave velocity (**v**). The ultimate structure of the resultant waveform is nominally sinusoidal [34–37]. In our experiments, the TWIM separator was operated both at a pressure of 3.5 mbar (200 ml/min and 90 ml/min flow rates for He and N₂, respectively) and 4 mbar (200 ml/min and 100 ml/min flow rates for He and N₂, respectively), for peptide and Avidin ion separations, respectively. TWIM-MS data was acquired at values of **V** ranging from 20 V to 40 V in 2 V increments and values of **v** ranging from 200 m/s to 500 m/s in 20 m/s increments. After the TWIM cell, ions travel to a transfer region that transports the mobility separated ions into the orthogonal acceleration (oa) time of flight mass analyzer (ToF). TWIM ATD is recorded by synchronization of the oa-ToF acquisition with the gated release of ions from trap into the TWIM cell. CCS was calibrated using D,L polyalanine peptides at a **v** value of 520 m/s and **V** values of 20, 25, and 30 V. He CCS values were used to construct calibration function, which was then used to predict He CCS values in our experiments. Three replicate measurements taken at different days were used for calibration in order to obtain uncertainty values as described previously[41].

SIMION Modeling

A Synapt G2 TWIM cell was simulated using SIMION 8.1 (Scientific Instrument Services Inc., Ringoes, NJ, USA)[42]. Potential array (PA) files were created in order to apply potential to 4 electrodes at once and create a 4 repeat pattern that mirrors the implementation of traveling wave dc voltage in TWIM cell[34, 36]. A specific script was written to obtain the voltage and electric field strength in the G2 model for downstream analysis.

MD Simulations

Given the challenges listed above, we have chosen to adopt a targeted MD strategy to aid the interpretation of our TWIM peak width data. Our targeted strategy aims to map the gas-phase conformational landscapes for a series of well-characterized peptides, known to adopt helical structures in the gas phase [32]. We do not aim to produce quantitative depictions of the peptide structural ensemble from these simulations that can be directly compared to TWIM peak width data, but rather capture the qualitative trends observed in our experimental data that are related to the general flexibility of these systems in the gas phase. Furthermore, since our peak width analysis procedures cannot discriminate between broadening modes that result from ensembles of static structures (resulting in Gaussian peak shapes) or slowly interconverting structural families having similar CCS values (resulting in pseudo-Gaussian peak shapes), our MD approach is aimed only at probing the general gas-phase flexibilities of the peptide ions studied here. MD simulations were performed with CHARMM on a workstation with an Intel Xeon processor with eight CPU cores at 2.50 GHz. CHARMM22 force field was employed as it contains the CMAP correction for improved treatment of peptide backbones to achieve more accurate peptide conformations[43]. Helical Ac-Ala_n-Lys peptides with *n* from 6 to 19 were constructed in CHARMM by fixing the phi and psi angles to -47 and -67 degrees, respectively, and placing the charge on the Lys residue. Peptides were energy minimized using a 10 step conjugate gradient, followed by 100 steps of an adopted basis Newton Raphson (ABNR) method minimization *in vacuo*. The models were then equilibrated at 300 K for 50 ps, after which, they were

subjected to a simulated annealing (SA) cycle. Briefly, the system was heated from 300 K to 1000 K in 10 K increments and cooled to 0 K in 10 K decrements, with each temperature step lasting for a minimum of 100 ps. The system was then equilibrated at 0 K for 500 ps. The lowest energy structure from the SA run was then subjected to gradual heating with final temperature of 300 K, 400 K, and 500 K. 100 ns of constant temperature MD simulation was performed at those temperatures saving coordinates every 5 ps, generating 20,000 structures. This procedure was performed for three replicas having different random initial velocities. Our total analysis for each peptide ion, therefore, incorporates a total of nine replicas produced by SA, tracked over three different temperatures. It is important to note that our MD protocol does not provide accurate quantitative data on the energy barriers or interconversion timescales of the structures captured for our analysis. As such, all models generated during the procedures described above were collected and compared to our TWIM data in a qualitative manner. All CHARMM input script files were written in house.

Hierarchical Clustering

A hierarchical clustering method[44, 45] from scipy[46, 47] was used to classify structural families extracted from MD simulations. For each constant temperature run, 1000 structures were selected at regular intervals in 100 ns runs for classification. Pair-wise RMSD values were calculated for all combinations of structures using an in-house script. Pairwise euclidean distance matrix was generated using RMSD matrix using scipy. The resulting distance matrix was then used for hierarchical clustering using average method.

Theoretical CCS Calculations

IMPACT[48] and IMOS[49, 50] were used for CCS calculations for model structures. IMPACT was used for CCS calculations on all the structures resulting from MD simulations. The IMOS diffusive trajectory method, which accounts for diffuse scattering in momentum transfer calculation, was used

with He gas at 300 K with 50,000 total gas molecules. Non-integer partial charges were included in the structure from CHARMM. Overall, IMOS trajectory method calculations were employed on 280 total structures from our 300K constant temperature run, with 20 structures extracted from each peptide system, in order to test and validate the IMPACT results obtained.

Data Analysis

TWIM ATD data was extracted using TWIMExtract[51]. For CHARMM output trajectory files, in-house scripts were written to analyze the trajectories and extract the structures. IM-MS 3D plots were generated using Driftscope (Waters, Milford MA). Data was analyzed using python, numpy, and scipy[46, 47]. Matplotlib[52] was used to generate all the output plots shown in this work. Additional details are provided in the Supporting Information.

Theory: A Semi-empirical Width and Resolution Expression for TWIM

For DTIM separations, Equation 1 is an analytical solution to the general transport expression that describes the ATD of a single conformation ion species:

$$F(z, t) = C \frac{1}{4(\pi Dt)^{1/2}} \left(v_d + \frac{z}{t} \right) \left[1 - \exp \left(-\frac{r_0^2}{4Dt} \right) \right] \exp \left[-\frac{(z - v_d t)^2}{4Dt} \right] \quad (1)$$

where $F(z, t)$ is the function estimating the ATD of an ion, D is diffusion constant, t is arrival time, v_d is drift velocity, z is the position of the ion in traverse direction, r_0 is the radius of the drift tube entrance aperture, and C is a constant that is dependent on the initial formation of ion packet[53]. As discussed above, a similar expression is not currently available for TWIM ATD analyses [37].

Generally, we can define the widths of IM ATDs as a sum of a series of band broadening terms:

$$W = W_D + W_P + W_{SC} + W_{RXN} + W_C \quad (2)$$

where W is the width of an IM ATD, and each subscript in Equation 2 indicates the origin of the broadening factor indicated, where D is diffusion, P is gate pulse width, SC is space charge, RXN is

reaction chemistry, and $\mathbf{C}$ is conformational heterogeneity exhibited in the timescale of IM separation[54–56]. $\mathbf{W}_{sc}$ and $\mathbf{W}_{RXN}$ typically have a negligible impact on $\mathbf{W}$, as ion number densities are kept low and inert neutrals are used for IM separation. As such, practical estimates of $\mathbf{W}$ depend only on ion diffusion, pulse width, and conformational heterogeneity. As expressed previously for DTIM separations, we model TWIM peak widths using[54]:

$$\mathbf{w}^2 = \gamma + \beta t_g^2 + \delta t_{diff}^2 \quad (3)$$

where $\mathbf{w}$ is the experimental TWIM ATD width (full width at half the maximum peak height, or **fwhm**), t_g is gate pulse width, t_{diff} is diffusion limited width, and we define δ as a parameter that describes any non-diffusion broadening incorporated into the total TWIM peak width model to achieve a good fit, referred to below as the conformational broadening parameter. By substituting an expression for diffusion-limited TWIM peak width as defined previously[37] into Equation 3, we obtain:

$$\mathbf{w}^2 = \gamma + \beta t_g^2 + \delta \frac{16kTv\ln 2}{qLKE^2} t^2 \quad (4)$$

where k is Boltzmann's constant, T is the temperature, v is wave velocity, q is the charge of the ion, L is the length of the TWIM cell, K is ion mobility, E is electric field, and t is arrival time of the ion. By simplifying Equation 4 we arrive at:

$$\mathbf{w}^2 = \gamma + \beta t_g^2 + \alpha \frac{v}{KE^2} t^2 \quad (5)$$

where

$$\alpha = \delta \frac{16kTv\ln 2}{qL} \quad (6)$$

In Equations 5 and 6, α , β , and γ are fitting parameters, where α and $(\gamma + \beta t_g^2)$ are the slope and intercept, respectively, obtained from a linear regression between $\mathbf{w}^2$ and vt^2/KE^2 . Similarly, β is the slope obtained from the linear regression between $\mathbf{w}^2$ and t_g^2 as previously determined in DTIM systems[54]. In contrast to DTIM however, a linear relationship does not exist between $\mathbf{w}^2$ and t_g^2 (Figure

S1), and instead exhibit a complex relationship that is also dependent on other factors such as $\mathbf{V}$ and $\mathbf{v}$. In order to simplify downstream expressions of TWIM peak width, we set β equal to 1 in the analyses reported here. In order to validate this approach, we probed β values ranging from 1×10^{-3} to 1×10^3 , but observed no significant improvement in TWIM ATD fit quality (data not shown).

In order to effectively utilize Equation 5 to predict TWIM peak widths, estimates of ion velocities are required to provide ion arrival time values ($\mathbf{t}$). A previous description of TWIM ion transport theory[37] provides the following relationship between ion mobility, the structure of the TW electric field, and overall ion transit time:

$$\mathbf{t} = \frac{Lv}{K^2 \left(\frac{V}{x}\right)^2} + \mathbf{b} \quad (8)$$

This expression, upon rearrangement, becomes:

$$\mathbf{t} = x^2 \frac{Lv}{K^2 V^2} + \mathbf{b} \quad (9)$$

Linear regression analysis of Equation 9 yields a slope of x^2 and an intercept of $\mathbf{b}$ (Figure S3a and Table S2). The $\mathbf{V}/\mathbf{x}$ ratio shown in Equation 8 is equivalent to the effective E an ion experiences during its flight during TWIM separation, and can be used in Equation 7 to estimate TWIM ion arrival times (Figure S3b and Table S3). The $\mathbf{x}$ values shown in Equations 8 and 9 vary as a function of $\mathbf{K}$ (Figure S3c) which necessitates calibration across a range of $\mathbf{V}$ and $\mathbf{v}$ in order to be able to predict TWIM arrival times.

We have further extended these empirical equations in order to predict TWIM resolution. By dividing Equation 5 through on both sides by $\mathbf{t}^2$ we arrive at:

$$\left(\frac{w}{t}\right)^2 = \frac{\gamma + \beta t_g^2}{t^2} + \alpha \frac{v}{KE^2} \quad (10)$$

IM resolution is typically defined as:

$$\mathbf{R} = \frac{t}{w} \quad (11)$$

where $\mathbf{R}$ is the resolution. Substituting $\mathbf{R}$ into Equation 10 we produce:

$$R^{-2} = R_p^{-2} + R_d^{-2} \quad (12)$$

where R_p and R_d are the contributions to IM resolution related to the initial ion pulse width and diffusion, respectively. Using Equations 10 and 12 we can define R_p and R_d individually as:

$$R_p = \frac{t}{(\gamma + \beta t_g^2)^{1/2}} \quad (13)$$

$$R_d = \left(\frac{E^2 K}{\alpha v} \right)^{1/2} \quad (14)$$

Further, we can use Equations 5, 7, and 11 to obtain a complete R expression:

$$R = \frac{Lv}{K^2 E^2 \left(\gamma + \beta t_g^2 + \frac{\alpha L^2 v^3}{K^5 E^5} \right)^{1/2}} \quad (15)$$

Additionally, by setting $\frac{dR}{d\left(\frac{E^2}{v}\right)} = 0$ we find the optimal E^2/v to be:

$$\frac{E^2}{v_{optimal}} = \left[\frac{\alpha L^2}{2K^5(\gamma + \beta t_g^2)} \right]^{1/3} \quad (16)$$

In this work, we test the above empirical sets of equations in order to model the TWIM ATD width and resolution for a range of biomolecular ions. In addition, we similarly evaluate the conformational broadening parameter δ for specific peptides and protein complexes.

Results and Discussion

To test our set of semi-empirical equations describing TWIM peak width, we used a series of Ac-Ala_n-Lys peptides, which have previously been observed to adopt rigid helical structures in the gas phase[57, 58]. Figure 1a and 1b shows TWIM-MS data collected for Ac-Ala_n-Lys peptides contained within our n=19 sample. Within this sample, we observe a distribution of Ac-Ala_n-Lys peptides, with n ranging from 6 to 19. The appearance of n < 19 peptide signals is likely due to the hydrolysis of n=19 peptides under the acidic conditions used to dissolve the original hydrophobic peptide sample, and such conditions mirror those used previously to analyze the gas-phase structures of these sequences[32, 57,

58]. We observe both $[M+H]^+$ and $[M+2H]^{2+}$ peptide ions in our IM-MS data, with the latter group detected having relatively low signal intensities. Due to the significantly larger signal intensity for the singly-charged peptide ions in our dataset, our detailed peak width analysis focuses on these signals exclusively.

Our peptide TWIM ATD data were acquired for a range of V and v settings at 3.5 mbar of pressure in the TWIM cell. ATDs were extracted for individual peptide systems and fitted with a Gaussian function to obtain a centroid arrival time, full width half max (***fwhm***), and resolution values, which were calculated using Equation 11. A contour plot of resolution and ***fwhm*** as a function of V and v is shown in figure 1c and figure 1d, respectively, for Ac-Ala₁₈-Lys¹⁺. In this data, the ***fwhm*** decreases as V increases and v decreases, reaching a maximum at the smallest V and highest v values probed here. On the other hand, the TWIM resolution trend seen in figure 1c does not mirror the ***fwhm*** trend, and instead reaches optimally large values at a V to v ratio of 0.075-0.125 (Figure S4), as observed previously for TWIM analyzers[56]. The resolution values observed in this study differ from those reported previously [56, 59], which discussed TWIM resolution in CCS space. We computed resolution values reported here [60] in drift time space in order to directly link our experimental measurements to the semi-empirical model we devised.

After adjusting the effective E field to yield an accurate estimation of centroid arrival times (Equation 9, Figure S3, and Table S3), we implemented Equations 5 and 15 to fit our experimental dataset in terms of ***fwhm*** and R , respectively. Figure 2a shows a plot of ***fwhm***² vs vt^2/KE^2 for Ac-Ala₁₈-Lys¹⁺ data. Through linear regression, and the application of Equation 5, we extract an α value for these data, an umbrella term that includes conformationally-derived peak broadening, with a strong correlation coefficient ($R^2 = 0.9751$) indicating that our model fits the experimental data well. In addition, we utilize the standard error resulting from our linear regression to estimate the error associated with δ value determined in this way. We compared measured R , R_d , and R_p values with computed resolution R values as a function of E^2/v in Figure 2b. As E^2/v increases, R_d increases whereas R_p decreases, as

expected. R , which is the weighted sum of R_p and R_d from Equation 15 models the functional form observed experimentally. The red dashed line indicates the $E^2/v_{optimal}$ value (from Equation 16) where maximum R is observed is predicted to be observed. A summary of the fitting parameters and $E^2/v_{optimal}$ for all the peptides is shown in Table S4. Figure 2c plots measured and computed ***fw****hm* as a function of E^2/v . Computed ***fw****hm* values agree well with those measured by TWIM, as indicated by the linear regression analysis shown. Specifically, our semi-empirical TWIM width relationship models the ***fw****hm* values for the peptides studied here with an average relative standard deviation of about 5% (Table S5 and Figure S5).

In order to quantitatively evaluate the peak broadening observed in our TWIM peptide data related to conformational polydispersity, we evaluated the conformational broadening parameter, δ , by separating this value from the remainder of the α term found in Equation 6. When the δ term that describes a TWIM peak approaches a value of 1, the width of that peak is diffusion-limited. Any increase in the δ value needed to describe the width of a TWIM ATD is, therefore, evidence of non-diffusional broadening, specifically those related to the ensemble of gas-phase structures occupied by the ion. Figure 3 displays the δ values for Ac-Ala_n-Lys peptides required to fit the TWIM peak widths recorded in our experiments.

We observe δ values near to 1 for Ac-Ala_n-Lys peptides where $n = 6 - 11$, indicating negligible contributions to observed TWIM peak widths from the peptide conformational ensemble. Furthermore, we observe a sharp transition in the magnitude of the δ values required to fit our TWIM peak widths at a peptide length of $n=12$, after which δ values remain above 1.2 for peptides with lengths of $n = 13 - 19$. Some of our TWIM peak widths require δ values slightly below 1 in order to generate accurate fits, likely due to minor inaccuracies in our empirical relationships (Table S5 and Figure S5). We observe a large shift in δ value as the length of the Ac-Ala_n-Lys¹⁺ peptides increases, indicating significant increases in peptide structural heterogeneity for sequences with $n > 11$. This result correlates well with previous

DTIM measurements, where singly-charged Ac-Ala_n-Lys ions were generally observed to be helical and rigid. In addition, DTIM widths for similar, singly-charged polyalanine peptides were reported to scale from diffusion-limited values for short sequences, to values 1.5 times higher for n = 20 peptides[32].

In order to model the ensemble of peptide structures present in our TWIM-MS experiments, we utilized MD simulations to generate a large population of Ac-Ala_n-Lys peptide structures, equilibrated at 300 K, 400 K, and 500 K. At each temperature, we pooled all low energy structures generated, and plotted these as a histogram to create CCS distributions, which were subsequently fit to Gaussian functions for comparison with TWIM datasets (Figure S6). Weighted mean values for each distribution were plotted against the number of alanine residues in the peptide analyzed, as shown in Figure 4a.

To further evaluate the quantitative agreement between our computed structural ensembles and those observed experimentally by TWIM, we evaluated the weighted mean CCS values generated from both IMPACT and IMOS, and observed good agreement with experimental CCS values, regardless of the temperature used in our MD simulations (Figure S7). Importantly, we find that two linear regression models provide the best fit for the experimental data shown in Figure 4a (Figure S8) with a first linear trend covering peptides with n = 6-13 and a second trend covering n = 14-19. Slope values for these two trend lines are 14.9 Å²/Alanine for shorter peptides and 16.5 Å²/Alanine for those that are longer. We observe similar variance in the fit when the two trends described above are restructured so that the first linear regression covers peptides with n=6-12, and the second covers those with n = 13-19 (Figure S8), clearly indicating a shift in CCS increase per-alanine value observed at either n = 12 or 13. Importantly, the transition point recorded for our CCS trend lines occurs at the same peptide lengths where we observe a sharp increase in δ values in Figure 3.

In pursuit of a deeper structural interpretation for our TWIM data, we interrogated our MD results by classifying the resulting structural ensembles in terms of the conformational families produced during our simulations. Hierarchical clustering on a reduced dataset of ~1000 structures extracted from

each constant temperature run was used to classify structural families. Families accounting for greater than 2% of the total population were considered for assessing peptide flexibility (Tables S7 and S8). Figures 4b, 4c, and 4d show our classification results for $n = 6$ peptide models at 300K, 400K, and 500K, respectively, whereas Figures 4e, 4f, and 4g show classifications for $n = 19$ peptides similarly structured with respect to temperature. In general, our simulations reveal broader ranges of ion CCS values for longer peptides, in agreement with our TWIM peak width data. Our cluster analysis indicates that while primarily helical, structural heterogeneity occurs in Ac-Ala_n-Lys peptide ions at their C-termini (Figure S9 and Tables S6, S7, and S8). Notably, MD simulations for shorter peptide sequences, from $n = 6$ -10, contain a structural family where helicity is not maintained throughout the peptide. For example, cluster 1 in 500K dataset for singly-charged Ac-Ala₆-Lys exhibits a more random coil type conformation (Figures 4d and S-9). Despite these structural differences, random coil and helical peptide families have very similar mean CCS values (Tables S6, S7, and S8). For larger peptides, the helical structural families in our MD simulations appear to more completely dominate the structural landscape, with significant variations in the C-termini (Figure 4e, 4f, 4g, and Figure S9). While the widths of our MD-generated ensembles in CCS space do not quantitatively correlate with our experimental peak widths from TWIM data, such agreement is not likely given the unknown temperatures of the ions measured and the relative simplicity of our simulations (Figure S-10)[37, 61, 62]. Overall, however, our MD simulations point to the conformational diversity of helical peptide ion C-termini as the main driver of the increased δ values observed in Figure 3.

The ultimate aim of our TWIM width analysis workflow is the rapid assessment of protein structural heterogeneity. To demonstrate the capabilities of our TWIM method, we measured δ values for Avidin, a 64 kDa homo-tetrameric protein complex, as a function of the concentration of added CXL reagent in solution. We anticipate that CXL reactions will, in general act to constrain the Avidin structure, yielding a population of gas-phase ion structures of reduced structural heterogeneity, and thus

decreased δ values. Figure 5 shows δ values recorded for Avidin as a function the ratio between BS3, a CXL reagent that targets primary amines, and the available Lys residues within the Avidin sequence. Remarkably, we note δ values for Avidin ions between 10-20 times those observed for Ac-Ala_n-Lys peptide ions, indicating a dramatically increased level of structural polydispersity for the protein complex. As the BS3/Lys ratio is increased, we observe a significant decrease in the δ values required to accurately fit our TWIM peak widths, indicative of a restrained population of Avidin ion structures across both charge states probed. We note that different δ values are required to fit different Avidin charges states, and that changes in Avidin TWIM peak width appear non-uniform across those Avidin charge states probed here, indicating that each protein complex charge state is comprised of a unique structural ensemble. We observe minimal δ values BS3/Lys ratios of 1.38 and 13.8 for 15+ and 16+ ions respectively. As the BS3/Lys ratio is increased beyond 13.8, we further observe an increase in δ values for both charge states, indicating an increase in structural polydispersity, likely driven by the prevalence of dead-end reaction products known to dominate under such reaction conditions previously[63–65]. Overall, our Avidin results illustrate the performance of our empirical TWIM peak width analysis approach. The changes in δ we observe match our expectations for CXL modified Avidin complex ions, as well as revealing new quantitative insights into gas-phase protein complex structures.

Conclusions

Dynamic ensembles, rather than static structures, dictate the function of active biomolecules. The TWIM width analysis approach presented here provides a rapid assay of protein structure heterogeneity. We develop and validate a semi-empirical relationship that accurately models TWIM ATD widths and resolutions across a wide range of instrument settings. In addition, we use this workflow to examine the structure of peptide and protein complex model systems. By extracting a conformational broadening parameter, δ , from our empirical formulae, we quantified the conformational broadening for a series of Ac-Ala_n-Lys peptides, with $n = 6 - 19$. Our width analyses allowed us to classify these peptide

ions into two families in terms of their structural heterogeneity, with $n=6-11$ having δ values indicating a diffusion-limited peak width, and $n \geq 12$ exhibiting δ values congruent with more significant conformational broadening. Furthermore, we utilized MD simulations in combination with hierarchical clustering analysis in order to locate the likely source of the structural polydispersity in our TWIM peptide data to the C-terminal regions of the longer sequences studied. Finally, we implemented our TWIM width analysis approach to study the influence of chemical cross-linking on the structural ensemble occupied by the Avidin tetramer, recovering large δ values that decrease sharply upon CXL treatments, followed by increases at when CXL agent excesses drive the formation of dead-end reaction products. Our quantitative TWIM peak width analyses also detected charge state dependent effects, indicating the presence of distinct structural ensembles for Avidin tetramers previously hidden within iso-CCS TWIM features.

We envision that our TWIM peak width analysis approach will enable a broad spectrum of applications in protein structural biology and biophysics. Examples include, but are not necessarily limited to, the study protein aggregation, evaluating the functional ensembles of intrinsically disordered proteins (IDPs), and rapidly monitoring the effect of ligand binding on proteins targets. Generating accurate estimates of biomolecular ensembles remains a challenge for computational biochemistry, and such difficulties extend directly to our ability to accurately model structural distributions *in vacuo* in a manner that corresponds to our TWIM-MS experiments. Furthermore, untangling the influence of charge state on the biomolecular ensembles quantified in our Avidin measurements will likely require the long-term application of charge manipulation, high resolution IM, and tandem IM experiments in combination with sophisticated MD simulations. However, it is clear from the data presented here that our empirical relationships describing TWIM peak width performs well across a wide range of protein analytes, and is readily accessible (within seconds) of sample introduction. Future applications of TWIM,

in combination with other gas-phase structural biology approaches, will undoubtedly lead to an improved understanding of the structural ensembles associated with function biomolecules.

Acknowledgements

We thank Kevin Giles and David Langridge at Waters for providing initial SIMION files to build G2 model and extract voltage and electric field. We further thank Kevin Giles for helpful discussions with TWIM theory. We would also like to thank Efrosini Artikis from the Brooks group at University of Michigan for helpful MD discussions. Finally, we gratefully acknowledge funding from the National Science Foundation Division of Chemistry under Grants 1808541 and 1253384 (with co-funding from the Division of Molecular and Cellular Biosciences) for supporting our efforts in TWIM theory development

References

1. Osadchy, M., Kolodny, R.: Maps of protein structure space reveal a fundamental relationship between protein structure and function. *Proceedings of the National Academy of Sciences*. 108, 12301–12306 (2011). doi:10.1073/pnas.1102727108
2. Sutto, L., Marsili, S., Valencia, A., Gervasio, F.L.: From residue coevolution to protein conformational ensembles and functional dynamics. *Proc Natl Acad Sci U S A*. 112, 13567–13572 (2015). doi:10.1073/pnas.1508584112
3. Boehr, D.D., Nussinov, R., Wright, P.E.: The role of dynamic conformational ensembles in biomolecular recognition. *Nature chemical biology*. 5, 789–796 (2009)
4. van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., Kim, P.M., Kriwacki, R.W., Oldfield, C.J., Pappu, R. V., Tompa, P., Uversky, V.N., Wright, P.E., Babu, M.M.: Classification of intrinsically disordered regions and proteins. *Chem Rev*. 114, 6589–6631 (2014). doi:10.1021/cr400525m
5. Oldfield, C.J., Dunker, A.K.: Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem*. 83, 553–584 (2014). doi:10.1146/annurev-biochem-072711-164947
6. Lange, O.F., Lakomek, N.A., Fares, C., Schroder, G.F., Walter, K.F., Becker, S., Meiler, J., Grubmuller, H., Griesinger, C., de Groot, B.L.: Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science*. 320, 1471–1475 (2008). doi:10.1126/science.1157092
7. Bernado, P., Mylonas, E., Petoukhov, M. V, Blackledge, M., Svergun, D.I.: Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc*. 129, 5656–5664 (2007). doi:10.1021/ja069124n
8. Dedmon, M.M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M., Dobson, C.M.: Mapping

- long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J Am Chem Soc.* 127, 476–477 (2005). doi:10.1021/ja044834j
9. Nakajima, N., Nakamura, H., Kidera, A.: Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides. *Journal of Physical Chemistry B.* 101, 817–824 (1997). doi:DOI 10.1021/jp962142e
 10. van Gunsteren, W.F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Geerke, D.P., Glattli, A., Hunenberger, P.H., Kastenholtz, M.A., Ostenbrink, C., Schenk, M., Trzesniak, D., van der Vegt, N.F.A., Yu, H.B.: Biomolecular modeling: Goals, problems, perspectives. *Angewandte Chemie-International Edition.* 45, 4064–4092 (2006). doi:10.1002/anie.200502655
 11. Kelleher, N.L.: From primary structure to function: Biological insights from large-molecule mass spectra. *Chemistry and Biology.* 7, 37–45 (2000). doi:10.1016/S1074-5521(00)00081-8
 12. Heck, A.J.R.: Native mass spectrometry: a bridge between interactomics and structural biology. *Nature Methods.* 5, 927–933 (2008). doi:10.1038/nmeth.1265
 13. van den Heuvel, R.H., Heck, A.J.R.: Native protein mass spectrometry: from intact oligomers to functional machineries. *Current Opinion in Chemical Biology.* 8, 519–526 (2004)
 14. Mehmood, S., Marcoux, J., Gault, J., Quigley, A., Michaelis, S., Young, S.G., Carpenter, E.P., Robinson, C. V.: Mass spectrometry captures off-target drug binding and provides mechanistic insights into the human metalloprotease ZMPSTE24. *Nature Chemistry.* 8, 1152–1158 (2016). doi:10.1038/nchem.2591
 15. Chan, D.S.-H., Kavanagh, M.E., McLean, K.J., Munro, A.W., Matak-Vinković, D., Coyne, A.G., Abell, C.: Effect of DMSO on Protein Structure and Interactions Assessed by Collision-Induced Dissociation and Unfolding. *Analytical Chemistry.* 89, 9976–9983 (2017).

doi:10.1021/acs.analchem.7b02329

16. Zhang, Z., Browne, S.J., Vachet, R.W.: Exploring salt bridge structures of gas-phase protein ions using multiple stages of electron transfer and collision induced dissociation. *J Am Soc Mass Spectrom.* 25, 604–613 (2014). doi:10.1007/s13361-013-0821-8
17. O’Brien, J.P., Li, W.Z., Zhang, Y., Brodbelt, J.S.: Characterization of Native Protein Complexes Using Ultraviolet Photodissociation Mass Spectrometry. *Journal of the American Chemical Society.* 136, 12920–12928 (2014). doi:10.1021/ja505217w
18. Mistarz, U.H., Brown, J.M., Haselmann, K.F., Rand, K.D.: Probing the Binding Interfaces of Protein Complexes Using Gas-Phase H/D Exchange Mass Spectrometry. *Structure.* 24, 310–318 (2016). doi:10.1016/j.str.2015.11.013
19. Zhang, A., Fang, J., Chou, R.Y., Bondarenko, P. V, Zhang, Z.: Conformational difference in human IgG2 disulfide isoforms revealed by hydrogen/deuterium exchange mass spectrometry. *Biochemistry.* 54, 1956–1962 (2015). doi:10.1021/bi5015216
20. Leitner, A., Faini, M., Stengel, F., Aebersold, R.: Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem Sci.* 41, 20–32 (2016). doi:10.1016/j.tibs.2015.10.008
21. Shi, Y., Fernandez-Martinez, J., Tjioe, E., Pellarin, R., Kim, S.J., Williams, R., Schneidman-Duhovny, D., Sali, A., Rout, M.P., Chait, B.T.: Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol Cell Proteomics.* 13, 2927–2943 (2014). doi:10.1074/mcp.M114.041673
22. Konermann, L., Vahidi, S., Sowole, M.A.: Mass spectrometry methods for studying structure and dynamics of biological macromolecules. *Anal Chem.* 86, 213–232 (2014). doi:10.1021/ac4039306
23. Lanucara, F., Holman, S.W., Gray, C.J., Evers, C.E.: The power of ion mobility-mass spectrometry

- for structural characterization and the study of conformational dynamics. *Nature Chemistry*. 6, 281–294 (2014). doi:10.1038/nchem.1889
24. May, J.C., McLean, J.A.: Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Analytical Chemistry*. 87, 1422–1436 (2015). doi:10.1021/ac504720m
 25. May, J.C., Morris, C.B., McLean, J.A.: Ion Mobility Collision Cross Section Compendium. *Anal Chem*. 89, 1032–1044 (2017). doi:10.1021/acs.analchem.6b04905
 26. Allen, S.J., Giles, K., Gilbert, T., Bush, M.F.: Ion mobility mass spectrometry of peptide, protein, and protein complex ions using a radio-frequency confining drift cell. *The Analyst*. 141, 884–891 (2016). doi:10.1039/C5AN02107C
 27. Politis, A., Park, A.Y., Hall, Z., Ruotolo, B.T., Robinson, C. V.: Integrative modelling coupled with ion mobility mass spectrometry reveals structural features of the clamp loader in complex with single-stranded DNA binding protein. *Journal of Molecular Biology*. 425, 4790–4801 (2013). doi:10.1016/j.jmb.2013.04.006
 28. Eschweiler, J.D., Frank, A.T., Ruotolo, B.T.: Coming to Grips with Ambiguity: Ion Mobility-Mass Spectrometry for Protein Quaternary Structure Assignment. *J Am Soc Mass Spectrom*. (2017). doi:10.1007/s13361-017-1757-1
 29. Shelimov, K.B., Jarrold, M.F.: Conformations, Unfolding, and Refolding of Apomyoglobin in Vacuum: An Activation Barrier for Gas-Phase Protein Folding. *J. Am. Chem. Soc.* 119, 2987–2994 (1997). doi:10.1021/ja962914k
 30. Shelimov, K.B., Clemmer, D.E., Hudgins, R.R., Jarrold, M.F.: Protein structure in Vacuo: Gas-phase conformations of BPTI and cytochrome c. *Journal of the American Chemical Society*. 119, 2240–2248 (1997). doi:10.1021/ja9619059
 31. Shi, H., Pierson, N. a., Valentine, S.J., Clemmer, D.E.: Conformation types of ubiquitin [M+8H]⁸⁺

- ions from water:methanol solutions: Evidence for the N and A states in aqueous solution. *Journal of Physical Chemistry B*. 116, 3344–3352 (2012). doi:10.1021/jp210797x
32. Hudgins, R.R., Mao, Y., Ratner, M.A., Jarrold, M.F.: Conformations of Gly(n)H⁺ and Ala(n)H⁺ peptides in the gas phase. *Biophysical journal*. 76, 1591–7 (1999). doi:10.1016/S0006-3495(99)77318-2
33. Wytttenbach, T., Von Helden, G., Bowers, M.T.: Gas-phase conformation of biological molecules: Bradykinin. *Journal of the American Chemical Society*. 118, 8355–8364 (1996). doi:10.1021/ja9535928
34. Giles, K., Pringle, S.D., Worthington, K.R., Little, D., Wildgoose, J.L., Bateman, R.H.: Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Communications in Mass Spectrometry*. 18, 2401–2414 (2004). doi:10.1002/rcm.1641
35. Pringle, S.D., Giles, K., Wildgoose, J.L., Williams, J.P., Slade, S.E., Thalassinou, K., Bateman, R.H., Bowers, M.T., Scrivens, J.H.: An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry*. 261, 1–12 (2007). doi:10.1016/j.ijms.2006.07.021
36. Giles, K., Williams, J.P., Campuzano, I.: Enhancements in travelling wave ion mobility resolution. *Rapid Communications in Mass Spectrometry*. 25, 1559–1566 (2011). doi:10.1002/rcm.5013
37. Shvartsburg, A.A., Smith, R.D.: Fundamentals of traveling wave ion mobility spectrometry. *Analytical Chemistry*. 80, 9689–99 (2008). doi:10.1021/ac8016295
38. Kune, C., Far, J., De Pauw, E.: Accurate Drift Time Determination by Traveling Wave Ion Mobility Spectrometry: The Concept of the Diffusion Calibration. *Anal Chem*. 88, 11639–11646 (2016). doi:10.1021/acs.analchem.6b03215
39. Mortensen, D.N., Susa, A.C., Williams, E.R.: Collisional Cross-Sections with T-Wave Ion Mobility

- Spectrometry without Experimental Calibration. *J Am Soc Mass Spectrom.* 28, 1282–1292 (2017).
doi:10.1007/s13361-017-1669-0
40. Richardson, K., Langridge, D., Giles, K.: Fundamentals of travelling wave ion mobility revisited: I. Smoothly moving waves. *International Journal of Mass Spectrometry.* 428, 71–80 (2018).
doi:10.1016/J.IJMS.2018.03.007
41. Ruotolo, B.T., Benesch, J.L.P., Sandercock, A.M., Hyung, S.-J., Robinson, C. V: Ion mobility–mass spectrometry analysis of large protein complexes. *Nature Protocols.* 3, 1139–1152 (2008).
doi:10.1038/nprot.2008.78
42. Manura, D., Dahl, D.: SIMION (R) 8.0/8.1 User Manual. Scientific Instrument Services, Ringoes, NJ (2011)
43. Buck, M., Bouguet-Bonnet, S., Pastor, R.W., MacKerell Jr., A.D.: Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme. *Biophys J.* 90, L36-8 (2006). doi:10.1529/biophysj.105.078154
44. Mullner, D.: Modern hierarchical, agglomerative clustering algorithms. eprint arXiv:1109.2378. (2011)
45. Zepeda-Mendoza, M.L., Resendis-Antonio, O.: Hierarchical Agglomerative Clustering. In: Dubitzky, W., Wolkenhauer, O., Cho, K.-H., and Yokota, H. (eds.) *Encyclopedia of Systems Biology.* pp. 886–887. Springer New York, New York, NY (2013)
46. Walt, S. van der, Colbert, S.C., Varoquaux, G.: The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering.* 13, 22–30 (2011).
doi:10.1109/mcse.2011.37
47. Oliphant, T.E.: Python for scientific computing. *Computing in Science and Engineering.* 9, 10–20 (2007). doi:10.1109/MCSE.2007.58

48. Marklund, E.G., Degiacomi, M.T., Robinson, C. V, Baldwin, A.J., Benesch, J.L.: Collision cross sections for structural proteomics. *Structure*. 23, 791–799 (2015). doi:10.1016/j.str.2015.02.010
49. Larriba, C., Hogan, C.J.: Free molecular collision cross section calculation methods for nanoparticles and complex ions with energy accommodation. *Journal of Computational Physics*. 251, 344–363 (2013). doi:http://dx.doi.org/10.1016/j.jcp.2013.05.038
50. Larriba, C., Hogan, C.J.: Ion Mobilities in Diatomic Gases: Measurement versus Prediction with Non-Specular Scattering Models. *The Journal of Physical Chemistry A*. 117, 3887–3901 (2013). doi:10.1021/jp312432z
51. Haynes, S.E., Polasky, D.A., Dixit, S.M., Majmudar, J.D., Neeson, K., Ruotolo, B.T., Martin, B.R.: Variable-Velocity Traveling-Wave Ion Mobility Separation Enhancing Peak Capacity for Data-Independent Acquisition Proteomics. *Analytical Chemistry*. 89, 5669–5672 (2017). doi:10.1021/acs.analchem.7b00112
52. Hunter, J.D.: Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*. 9, 99–104 (2007). doi:10.1109/MCSE.2007.55
53. Mason, E.A., McDaniel, E.W.: *Transport properties of ions in gases*. Wiley, New York (1988)
54. Siems, W.F., Wu, C., Tarver, E.E., Hill, H.H.J., Larsen, P.R., McMinn, D.G.: Measuring the Resolving Power of Ion Mobility Spectrometers. *Analytical Chemistry*. 66, 4195–4201 (1994). doi:10.1021/ac00095a014
55. Wu, C., Siems, W.F., Asbury, G.R., Hill, H.H.: Electrospray ionization high-resolution ion mobility spectrometry - Mass spectrometry. *Analytical Chemistry*. 70, 4929–4938 (1998). doi:DOI 10.1021/ac980414z
56. Zhong, Y., Hyung, S.-J., Ruotolo, B.T.: Characterizing the resolution and accuracy of a second-generation traveling-wave ion mobility separator for biomolecular ions. *The Analyst*. 136, 3534

- (2011). doi:10.1039/c0an00987c
57. Hudgins, R.R., Ratner, M.A., Jarrold, M.F.: Design of helices that are stable in vacuo. *Journal of the American Chemical Society*. 120, 12974–12975 (1998). doi:DOI 10.1021/ja983021q
 58. Hudgins, R.R., Jarrold, M.F.: Helix formation in unsolvated alanine-based peptides: Helical monomers and helical dimers. *Journal of the American Chemical Society*. 121, 3494–3501 (1999). doi:DOI 10.1021/ja983996a
 59. Giles, K., Williams, J.P., Campuzano, I.: Enhancements in travelling wave ion mobility resolution. *Rapid Communications in Mass Spectrometry*. 25, 1559–1566 (2011). doi:10.1002/rcm.5013
 60. Dodds, J.N., May, J.C., McLean, J.A.: Correlating Resolving Power, Resolution, and Collision Cross Section: Unifying Cross-Platform Assessment of Separation Efficiency in Ion Mobility Spectrometry. *Analytical Chemistry*. 89, 12176–12184 (2017). doi:10.1021/acs.analchem.7b02827
 61. Morsa, D., Gabelica, V., De Pauw, E.: Effective temperature of ions in traveling wave ion mobility spectrometry. *Anal Chem*. 83, 5775–5782 (2011). doi:10.1021/ac201509p
 62. Merenbloom, S.I., Flick, T.G., Williams, E.R.: How hot are your ions in TWAVE ion mobility spectrometry? *J Am Soc Mass Spectrom*. 23, 553–562 (2012). doi:10.1007/s13361-011-0313-7
 63. Swaim, C.L., Smith, J.B., Smith, D.L.: Unexpected products from the reaction of the synthetic cross-linker 3,3'-dithiobis(sulfosuccinimidyl propionate), DTSSP with peptides. *Journal of the American Society for Mass Spectrometry*. 15, 736–749 (2004). doi:10.1016/j.jasms.2004.01.011
 64. Clifford-Nunn, B., Showalter, H.D.H., Andrews, P.C.: Quaternary Diamines as Mass Spectrometry Cleavable Crosslinkers for Protein Interactions. *Journal of The American Society for Mass Spectrometry*. 23, 201–212 (2012). doi:10.1007/s13361-011-0288-4
 65. Rozbeský, D., Rosůlek, M., Kukačka, Z., Chmelík, J., Man, P., Novák, P.: Impact of Chemical Cross-

Linking on Protein Structure and Function. *Analytical Chemistry*. 90, 1104–1113 (2018).

doi:10.1021/acs.analchem.7b02863

Figure Legends

Figure 1. a) A mass spectrum recorded for a Ac-Ala_n-Lys peptide where $n = 19$. **b)** A plot of TWIM drift time versus m/z , where ion intensity denoted by color-coded contour scale as indicated, for the same data shown in a. In both **a)** and **b)** we observe both a singly and a doubly charged series for $n = 6 - 19$ of peptides, with the latter series is detected at substantially lower signal intensities when compared to the former. Contour plots that show the influence of **c)** resolution and **d)** *fwhm* as a function of V and v for singly-charged Ac-Ala₁₈-Lys peptides, with values for both figures of merit indicated by the color axis shown.

Figure 2. a) A plot of $fwhm^2$ as a function of vt^2/KE^2 where linear regression analysis yields α as the slope and $(\gamma + \beta t_g^2)$ as the y-intercept, from Equation 5. The correlation coefficient and best fit equation from the linear regression analysis is shown on the plot. **b)** A plot of TWIM resolution as a function of E^2/v . Color coded trends are shown to represent predictions derived from our empirical relationships for diffusion-limited (green) and initial pulse width limited (blue) TWIM resolution. The predicted trend from the complete empirical resolution model is also shown (red solid line), and exhibits a strong correlation to experimentally measured TWIM peak widths (black points). The value for $E^2/v_{optimal}$ was determined using Equation 16, and is indicated on the plot (red, dashed line). **c)** A plot of *fwhm* as a function of E^2/v , where TWIM data points (in black) are compared against the trend predicted from Equation 3 (red, solid line). The inset shows the correlation between TWIM fwhm values from both our empirical model and experiment, exhibiting an excellent level of correlation and a slope of 0.99.

Figure 3. A plot of the conformational broadening parameter (δ) extracted from our TWIM peak fits as a function of number of alanine residues in the peptide ions measured. The error bars shown represent the experimental error, propagated from the slope of linear regression shown in Figure 2a. The dotted horizontal line at $\delta = 1$ indicates the expected δ value for diffusion limited TWIM peak widths. Larger δ values indicate significant conformational broadening.

Figure 4. a) A trend line analysis of Ac-Ala-Lys CCS data plotted as a function of Ala residues contained within the peptide (black dot points). Theoretical CCS values representing the mean of our MD-generated structural ensembles are also shown, computed using the IMPACT (blue diamonds) and IMOS (green triangles) trajectory methods. Two linear regression models were required to fit the experimental CCS data as indicated by the red and blue dashed lines. **b), c)** and **d)** show the clustered CCS distributions for Ac-Ala₆-Lys peptide ions at 300 K, 400 K, and 500 K, respectively. **e), f),** and **g)** show similarly clustered CCS distributions for Ac-Ala₁₉-Lys peptide ions at 300 K, 400 K, and 500 K, respectively. Clusters are numbered as 0 and 1, as referred to in the text. Ensembles of overlaid structures extracted from clusters are also shown below each CCS distribution plot, and are labelled with respect to their cluster of origin.

Figure 5. Bar graph of δ values needed to fit TWIM peak widths recorded for unmodified and BS3 cross-linked Avidin tetramer 15+ (blue) and 16+ (red) ions. We interpret the initial decreases in δ values upon treatment with CXL agents as evidence of rigidified protein structures, whereas increases in δ values observed at larger BS3 excesses are likely due to the prevalence of dead-end cross-links under such conditions, as illustrated in the schematic shown.

Figure 1

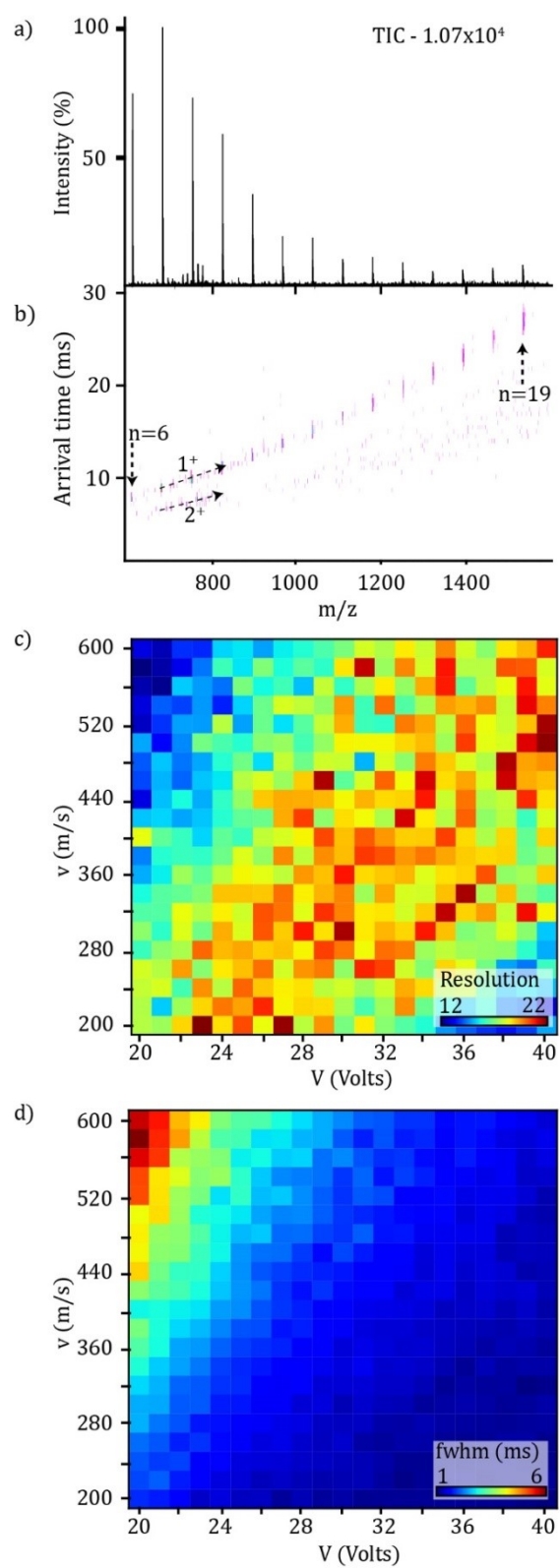


Figure 2

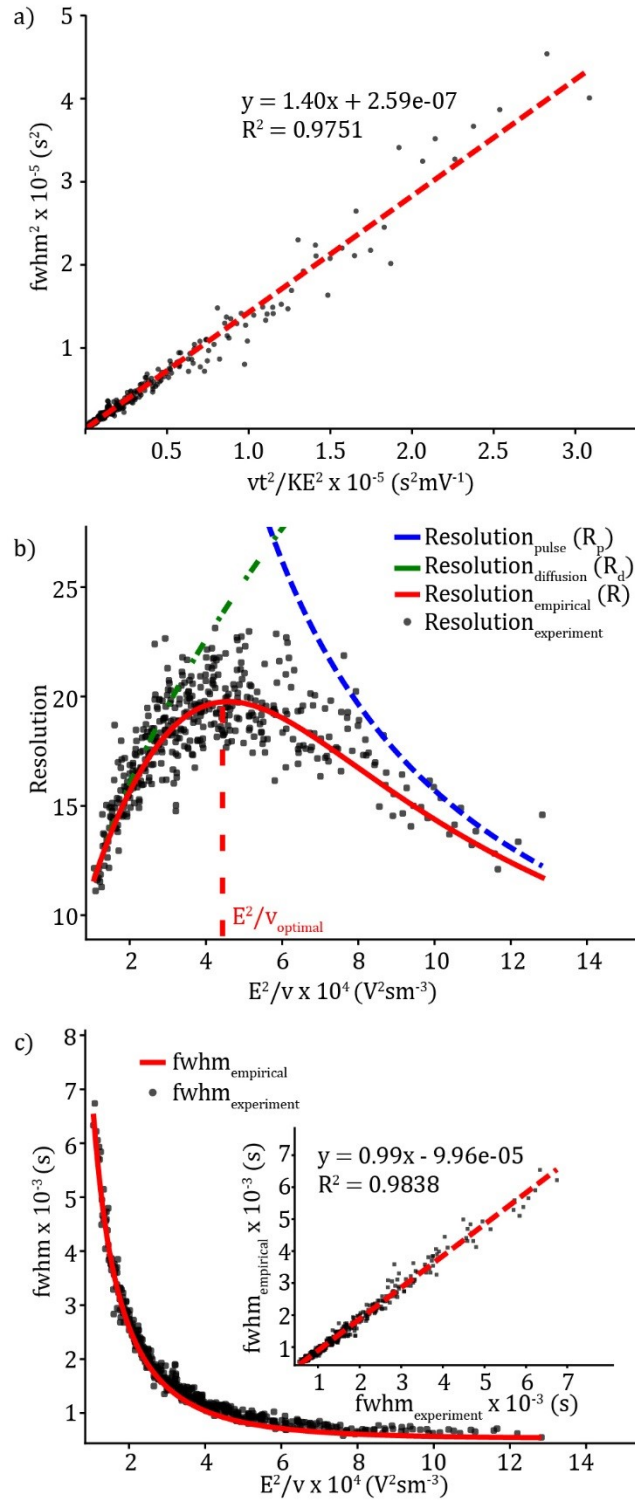


Figure 3

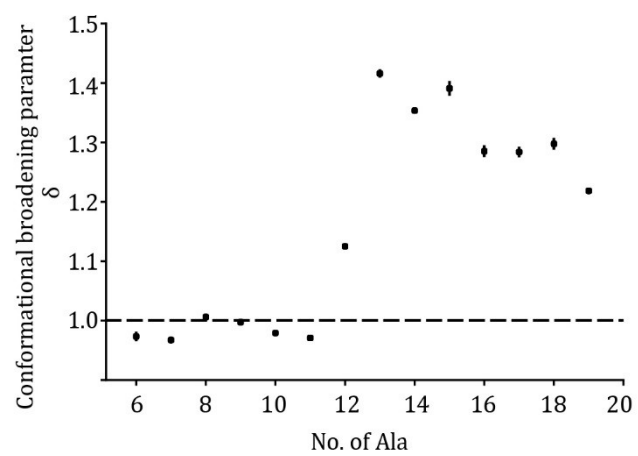


Figure 4

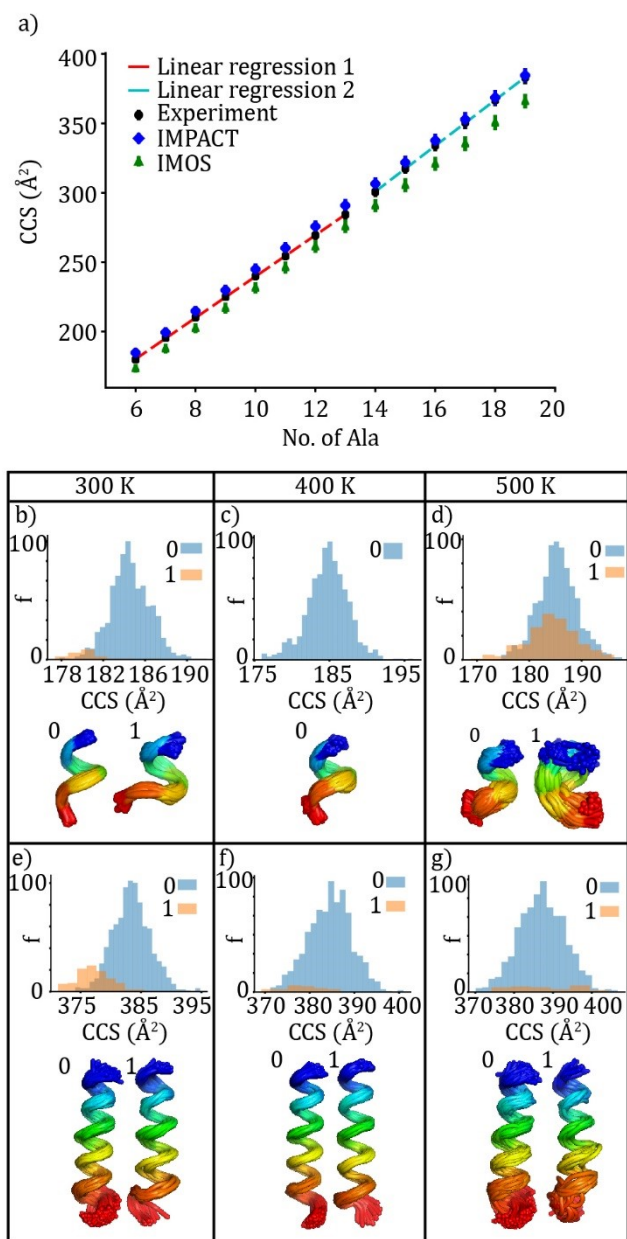


Figure 5

