A Mass-Spectrometry-Based Framework To Define the Extent of Disorder in Proteins

ARTICLE *in* ANALYTICAL CHEMISTRY · OCTOBER 2014

Impact Factor: 5.64 · DOI: 10.1021/ac5027435 · Source: PubMed

6 AUTHORS, INCLUDING:

Rebecca Beveridge
Research Institute of Molecular Pathology
4 PUBLICATIONS **41** CITATIONS
SEE PROFILE

Kamila J Pacholarz
The University of Manchester
6 PUBLICATIONS **56** CITATIONS
SEE PROFILE

Cait Macphee
The University of Edinburgh
74 PUBLICATIONS **3,236** CITATIONS
SEE PROFILE

Perdita Barran
The University of Manchester
110 PUBLICATIONS **1,790** CITATIONS
SEE PROFILE

analytical chemistry

Perspective

pubs.acs.org/ac

A Mass-Spectrometry-Based Framework To Define the Extent of Disorder in Proteins

Rebecca Beveridge,[†] Sam Covill,[‡] Kamila J. Pacholarz,[‡] Jason M. D. Kalapothakis,[‡,§] Cait E. MacPhee,[§] and Perdita E. Barran*[,†]

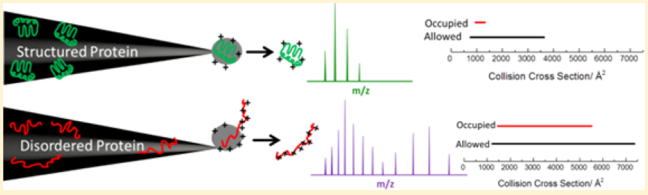
[†]Manchester Institute of Biotechnology, Michael Barber Centre for Collaborative Mass Spectrometry, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom

[‡]School of Chemistry, University of Edinburgh, Joseph Black Building, West Mains Road, Edinburgh EH9 3JJ, United Kingdom

[§]School of Physics and Astronomy, University of Edinburgh, James Clerk Maxwell Building, Mayfield Road, Edinburgh EH9 3JZ, United Kingdom

Supporting Information

ABSTRACT: In the past decade, mass spectrometry (MS) coupled with electrospray ionization (ESI) has been extensively applied to the study of intact proteins and their complexes, often without the requirement of labels. Solvent conditions (for example, pH, ionic strength, and concentration) affect the observed desolvated species; the ease of altering such extrinsic factors renders ESI-MS an appropriate method by which to consider the range of conformational states that proteins may occupy, including natively folded, disordered and amyloid. Rotationally averaged collision cross sections of the ionized forms of proteins, provided by the combination of mass spectrometry and ion mobility (IM-MS), are also instructive in exploring conformational landscapes in the absence of solvent. Here, we ask the following question: "If the only technique you had was ESI-IM-MS, what information would it provide on the structural preferences of an unknown protein?" We have selected 20 different proteins, both monomeric and multimeric, ranging in mass from 2846 Da (melittin) to 150 kDa (Immunoglobulin G), and we consider how they are presented to a mass spectrometer under different solvent conditions. Mass spectrometery allows us to distinguish which of these proteins are structured (melittin, human beta defensin 1, truncated human lymphotactin, Cytochrome C, holo hemoglobin-α, ovalbumin, human transthyretin, avidin, bovine serum albumin, concanavalin, human serum amyloid protein, and Immunoglobulin G) from those that contain at least some regions of disorder (human lymphotactin, N-terminal p53, α-Synuclein, N-terminal MDM2, and p53 DNA binding domain) or denatured due to solvent conditions (ubiquitin, apo hemoglobin-α, apo hemoglobin-β) by considering two experimental parameters: the range of charge states occupied by the protein (Δz) and the range of collision cross sections in which the protein is observed (ΔCCS). We also provide a simple model to predict the difference between the collision cross sections of the most compact and the most extended form of a given protein, based on the volume of the amino acids it contains. We compare these calculated parameters with experimental values. In addition, we consider the occupancy of conformations based on the intensities of ions in the mass spectra. This allows us to qualitatively predict the potential energy landscape of each protein. Our empirical approach to assess order or disorder is shown to be more accurate than the use of charge hydropathy plots, which are frequently used to predict disorder, and could provide an initial route to characterization. Finally, we present an ESI-IM-MS methodology to determine if a given protein is structured or disordered.



The organization of polypeptide chains into three-dimensional (3D) shapes has been investigated for many years. An early view of protein structure, described as the lock and key model, was proposed by Emil Fischer[1,2] in which exact stereochemical recognition was deemed essential for protein−protein or protein−ligand interactions. This purported that every protein possessed *one* energy-minimized structure, which was essential for its biological function. This model agreed with emerging evidence of organized 3D structures from crystallography.[1,3,4] Karush was among the earliest to critique this structure function paradigm,[5] proposing instead a theory of configurational adaptability, which was also supported by Koshland,[6] who called the concept "the induced-fit theory".[1,6]

As the number of protein structures solved by X-ray crystallography increased, it became apparent that the resolution could vary across the unit cell,[1] and this was attributed to localized disorder. The concept that proteins can be functional and yet lack a resolvable structure has gained momentum in recent years. It is now accepted that there is a progression of structure in functional proteins from those that are highly structured, perhaps possessing few flexible regions, to

Received: July 23, 2014
Accepted: October 3, 2014

   dx.doi.org/10.1021/ac5027435 | *Anal. Chem.* XXXX, XXX, XXX−XXX

those that are completely disordered. Here, a protein, or region of a protein, is considered to be disordered if it lacks a stable structure, on the time scale of a nuclear magnetic resonance (NMR) experiment,[7−11] and displays flexible, interconverting, coil-like conformations.[11] The location of a given protein in this structure progression appears to be associated with function; highly structured proteins are found to be involved in biosynthesis and transport,[10] whereas disordered proteins are involved in regulation of transcription and translation, cellular signal transduction, protein phosphorylation, storage of small molecules, and regulation of the self-assembly of large multiprotein complexes such as flagella and ribosomes.[8,11−14] It has been shown that proteins arising from chromosomal translocations are enriched in higher levels of disorder, which allows the fused proteins to avoid removal by cellular surveillance mechanisms, and such fusion proteins are, in turn, associated with disease[9] and implicated in many neurodegenerative conditions.[15,16]

Certain amino acids have been suggested to be more *disorder-inducing* than others; intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) are enriched in P, E, K, S and Q, and depleted in W, Y, F, C, I, L and N, compared to the average folded protein.[7,17,18] This low hydrophobicity is thought to deter the formation of a hydrophobic cluster; the high net charge, in turn, induces electrostatic repulsion, favoring extended conformations, while proline is a known breaker of secondary structural elements.[19]

Different methods must be used to purify and characterize IDPs than traditionally applied to most structured systems. Unfolded or disordered proteins can be aggregation prone, structurally dynamic,[20] and susceptible to proteolysis,[21] and, as already stated, they are incompatible with high resolution in crystallography. In NMR experiments, full assignment of unfolded and partially folded proteins is possible.[22,23] NOEs can be used to characterize the structure and dynamics of proteins, while coupling constraints and amide temperature coefficients can show regions of secondary structure.[22] Regions of disorder are signified in HSQC spectra by low dispersion in the proton dimension, while retaining high dispersion in the ^{15}N dimension, because of the sequence dependence of the nitrogen chemical shift.[24,25] However, NMR is limited by the molecular mass of a protein, it has a low tolerance to paramagnetic ligands and the high protein concentration that is required can lead to aggregation,[26] as well as requiring milligram-level amounts of protein, which may be hard to obtain.

Since the development of electrospray ionization by Fenn and co-workers,[27] by which large proteins can be transferred from solution into the gas phase as multiply charged ions, mass spectrometry has been used to monitor protein structure.[28−31] The technique that transfers proteins as gently as possible from aqueous solutions in the presence of additives that stabilize structures (for example, volatile salts) to the gas phase for analysis is known as native MS. It is fast and sensitive and can simultaneously provide mass and compositional information on several species in one spectrum.[32] Native MS tends to utilize nanoelectrospray ionization (n-ESI) with low sample consumption (flow rates on the order of nanoliters per minute (nL/min)). The n-ESI spray droplets are smaller than those obtained with ESI, and the methodology requires lower spray potentials and source temperatures, which help to maintain native-like interactions, along with collisional cooling as the proteins enter the mass spectrometer, although the latter also applies to standard ESI.[33,34]

A plethora of studies provides compelling evidence that the stoichiometry of protein−protein or protein−ligand complexes can be preserved from solution. This supports the view that aspects of solution topology can be retained following ionization;[35,36] although there is also evidence for collapsed and elongated states, these, of course, are not mutually exclusive. ESI presents proteins to the gas phase with a distribution of protons, and MS separates these because they have different charges. Desolvation gives rise to charge-state distributions which indicates different solution conformations[37,38] each with a somewhat different presentation of chargeable sites during solvent and salt removal. Protons will preferentially remain on or migrate to solvent-accessible sites of high proton affinity, and they will be removed from sites of low proton affinity.[35] It is for this reason that ordered proteins tend to occupy lower charge states; their tertiary fold will include a hydrophobic core, which can remain stable following ionization and, hence, less residues are solvent-accessible. Conversely, disordered proteins are less hydrophobic and often have an excess of charged sites;[35] hence, they exist in solution in extended forms, resulting in more solvent-accessible protonation/deprotonation sites.

A crystallizable form of a protein that will diffract is required for X-ray crystallographic studies and, NMR is unable to resolve multiple conformers in a solution; in contrast, ESI-MS is able to distinguish all adopted conformations of a protein, perhaps with a bias to those that are present with higher charged states.[39] Although seemingly simple information, the extent of charging of monomeric protein ions holds information about its conformation in solution.[26] Through comparison to other biophysical techniques, it is known that solution conditions that preserve structure in a structured protein will tend to show only low charge-state values (high m/z values) with a relatively narrow distribution of such peaks, whereas intrinsically disordered, or denatured, proteins present with a wider range of charge states distributed about the mode.[40,41] This was demonstrated by Kaltashov et al. for the protein myoglobin by fitting Gaussian functions to the envelopes of the charge states as a linear addition of basis functions.[38] They observed that the basis functions (specific to each protein) do not change greatly under different solvent conditions, except in intensity. They concluded that each Gaussian envelope corresponds to one main conformation and the width of these envelopes reflects the degree of heterogeneity of that conformation, with wider envelopes corresponding to greater heterogeneity. It still remains to be seen whether any physical parameters concerning the proteins in solution can be derived from such Gaussian fitting functions, but, at a minimum, it provides an intuitive relationship between charge-state distributions (CSDs) in the gas phase and the conformations present in solution.

Both *intrinsic* order and disorder will influence the CSDs of proteins, as will *induced* order and disorder through changing solvent conditions. Acids and methanol will tend to unfold and denature proteins, allowing increased charging during ionization. Ethanol can induce beta-sheets[42] and its addition to aqueous solutions of α-Synuclein prior to ESI alters the ensuing CSD.[40] The choice of counterions and ionic strength will also have an effect, although many traditional biological buffers are not compatible with MS analysis. A salt used commonly to prepare aqueous solutions for ESI-MS of proteins is ammonium acetate, which, as well as being volatile, has a tendency to

dx.doi.org/10.1021/ac5027435 | *Anal. Chem.* XXXX, XXX, XXX−XXX

produce a more compact CSD, attributed to a "salt-crowding" effect[43] (similar to "salting in").

Ion mobility is complimentary to mass spectrometry, and they are both employed here as the hybrid technique (IM-MS). This provides information on the shape of the protein, in the form of rotationally averaged collision cross sections (CCSs), in addition to m/z ratios.[44] A typical mobility experiment involves pulsing ions into a drift tube filled with a buffer gas, at a known pressure and temperature, across which a weak electric field is applied. As the ions are pulsed into the chamber, they are drawn through by the electric field. They are hindered by collisions with the buffer gas, which slow the ions. Larger ions will collide more frequently with the buffer gas; hence, they will be slowed to a greater degree than more-compact ions. The arrival time of the protein ions will therefore depend on the size of the ion, and the charge present upon it: proteins with a higher charge will be drawn through the cell faster by the electric field. For any given charge state there may be multiple conformers which can be separated by their arrival times. Smaller arrival times correspond to more-compact conformers and larger arrival times correspond to more-extended conformers.

Mass spectrometry and ion mobility experiments can provide us with a wealth of information on protein structure. Whether a protein exists in one structural family or several interconverting transient conformations, IM-MS is a valuable technique that can be utilized as a route to structural characterization, and it is applicable to proteins up to the megadalton (MDa) range.[45] Here, we present a framework derived from mass spectrometry experiments that can be applied to any protein to gain information on the degree of folding and dynamics, as an initial route to characterization. During this process, we also fit parameters to be used during data analysis. Although it is widely assumed that IDPs present a wide CSD and structured proteins present a narrow CSD, and, similarly, that IDPs will present a wide range of CCSs, compared to structured proteins, no comprehensive dataset has yet been tested to validate this assumption. We set out to do this here.

MATERIALS AND METHODS

Materials. The sequences of all proteins used are found in the Supporting Information Table S1, along with the ordering information (where possible). Human Cytochrome C (C3483), bovine β-casein (C6905), bovine serum albumin (A9418), hen egg white Lysozyme (L6876), hen egg white ovalbumin (A7641), egg white avidin (A9275), *Canavalia ensiformis* concanavalin A (C2010), and equine heart myoglobin (M1882) were purchased from Sigma−Aldrich. Human transthyretin was purchased from SCIPAC, U.K., human serum amyloid P component was purchased from CalBioChem Germany. Human lymphotactin residues 1−72 and full-length human lymphotactin were provided as a kind gift from Brian Volkmann, Medical College of Wisconsin;[46] N-terminal-MDM2 was a gift from Ted Hupp, University of Edinburgh; human α-Synuclein was a gift from Tilo Kunath, University of Edinburgh; N-terminal human p53 was a gift from Galina Selivanova, The Karolinska Institute; and Immunoglobulin G was a gift from UCB Pharma.[47] Ammonium acetate, liquid chromatography-mass spectrometry (LC-MS)-grade water, and LC-MS-grade methanol were all purchased from Fisher Scientific.

Mass Spectrometry. Mass spectrometry (MS) experiments were carried out on a Q-ToF Ultima instrument, unless otherwise stated. The details of each experiment are found in previously published work.[48−53] Data were analyzed using modified Masslynx software (version 4.1), Microsoft excel 2010 and Originlab 9.0. Ions were produced by positive nanoelectrospray ionization (Z-spray source) within a spray voltage range of 1.6−1.7 kV, a cone voltage of 45−80 V, and a source temperature of 80 °C. Nanospray tips were prepared in-house from borosilicate glass capillaries (Kwik-Fil, World Precision Instruments, Inc., Sarasota, FL, USA) using a Flaming/Brown micropipette puller (Model P-97, Sutter Instrument Co., Novato, CA, USA). Tips were filled with 10−15 μL of sample using gel loading tips (Eppendorf, Hamburg, Germany).

Ion Mobility. Instrument operation and measurement acquisition methods for each system examined have been described in detail elsewhere.[48−53] Briefly, ion mobility experiments were performed at a helium pressure of ∼3.5−4.0 Torr and a cell temperature of ∼300 K. These were measured carefully during each experiment. Drift times were recorded at six drift voltages between 60 V and 15 V. All mobility data were obtained from plots of arrival time versus drift voltage over this range. Experiments were performed in triplicate on separate days. Nanoelectrospray ionization produces a constant beam of ions that must be trapped to allow discrete ion packets to be pulsed into the drift cell. This is achieved by raising the DC voltage on the top hat lens element directly before the entrance elements to the cell, trapping the ions in the precell hexapole. The trapping DC is then lowered for 40 μs through the use of a pulse generator (Stanford Research Systems, Sunnyvale, CA, USA) allowing a pulse of ions into the drift cell. Ions are focused into the drift cell using a three-element Einzel lens, separated by mobility within the cell and then accelerated through the quadrupole and ToF mass analyzers to a MCP detector. A minimum of 2000 scans was collected for each drift voltage. Data were processed according to published procedures[54] to determine CCSs from measured arrival times over a range of drift voltages.

Fitting Procedures. The data of Figure 2a (structured proteins), presented later in this work, were fitted by a power function $y = ax^b$, with b fixed to a value of 0.5. The rest of the data were subject to linear fitting. For fits to Δz and ΔCCS for structured proteins, the slope was fixed to zero (0).

Charge-Hydropathy Analysis. Predictions of disorder in six proteins were analyzed using four different prediction methods. Protein sequences (found in the Supporting Information) were submitted to the PONDR and DISOPRED Web servers. For PONDR, they were analyzed using PONDR-VLXT, PONDR-VSL2, and PONDR-VL3. Access to PONDR was provided by Molecular Kinetics (Indianapolis, IN, USA; E-mail: main@molecularkinetics.com). [Note that VL-XT is copyrighted by the WSU Research Foundation (1999, all rights reserved). PONDR is copyrighted by Molecular Kinetics (2004, all rights reserved). For DISOPRED, we used DISOPRED3 available from the PSIPRED Protein Sequence Analysis Workbench at University College London (U.K.), http://bioinf.cs.ucl.ac.uk/psipred/.[55]]

We compare the charge and hydropathy indices with experimental data from charge state and collision cross section analysis, respectively.

Theoretical Modeling of Collision Cross Sections. The lower bound of the collision cross section of a protein was calculated by assuming that, when folded, the protein approximates a sphere in shape. This would describe a globular highly compact form of a protein. The value of protein density

dx.doi.org/10.1021/ac5027435 | Anal. Chem. XXXX, XXX, XXX−XXX

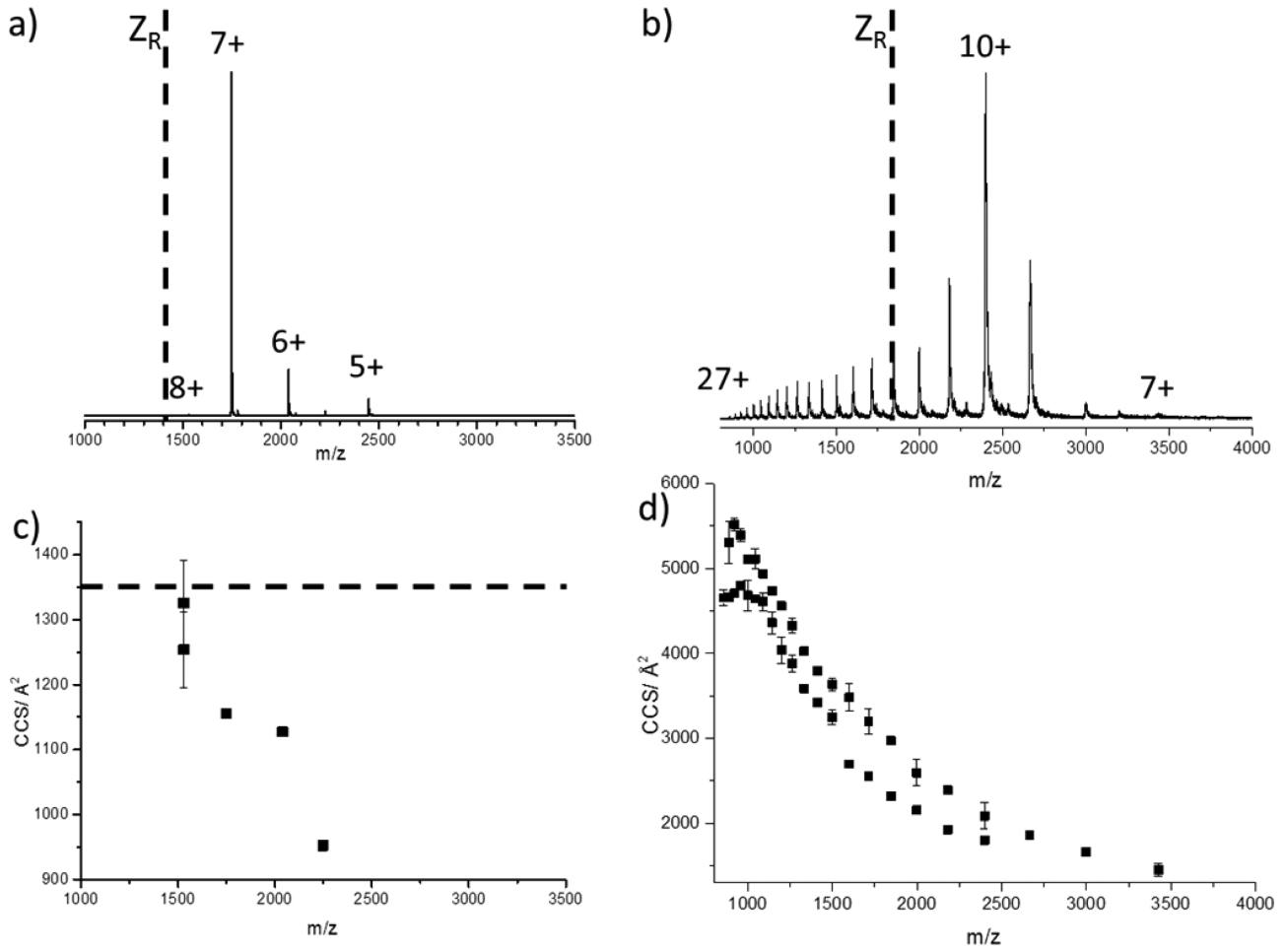


Figure 1. (a) Mass spectrum of Cytochrome C. (b) Mass spectrum of β-casein. Both spectra were recorded under similar conditions: 50 mM ammonium acetate, pH 6.8. (c) The rotationally averaged collision cross sections (CCSs) of each charge state of Cytochrome C (see Table S2 in the Supporting Information); the dashed line represents the CCS of the crystal structure of Cytochrome C, calculated by the trajectory method.[62] (d) The rotationally averaged CCSs of each charge state of β-casein (see Table S3 in the Supporting Information).

used here was ρ = 0.904 Da/Å^3, which has been suggested as being most relevant for small globular proteins.[56]

Using the molecular weight M_w of the protein, the volume of the protein sphere can be calculated via $V = M_w/\rho$. Therefore, the radius of the sphere is given as $r = [3V/(4\pi)]^{1/3}$. The CCS of a sphere of this radius is given by eq 1:

$$\mathrm{CCS}_{\mathrm{lower}}\ (\text{Å}^2) = \pi r^2 = \pi\left(\frac{3V}{4\pi}\right)^{2/3} \tag{1}$$

This gives the geometric size of the sphere of the protein. However, the CCS of the molecules inferred from their electrical mobility in helium always exceeds this value.[57,58] Work by De la Mora on a series of monomeric proteins demonstrates that the experimental CCS is ~1.19 times greater that of the theoretical CCS based on mass.[59] Therefore, we multiply the answer from eq 1 by a scaling factor of 1.19 to predict the smallest possible measured CCS.

Conversely, the upper bound on the CCS of a protein can be calculated knowing that the furthest distance between α-carbons in a protein chain is 3.63 Å.[32] Therefore, for a polypeptide of n residues, the maximum linear dimension is n(3.63) Å.[32] A completely unfolded protein that has been stretched out from end to end can be modeled as a cylinder of length, l (Å) = 3.63n, and radius of such a cylinder is given by the geometric average of the sum of the radii of the amino acids contained in the protein's sequence. The average volume of an amino acid in a protein's sequence is given by eq 2:

$$\bar{V} = \frac{\sum_i^{\text{amino acids}} V_i N_i}{n} \tag{2}$$

Here, the sum is over all amino acids i, V_i is the volume of the ith amino acid, and N_i is the number of amino acids of type i in the protein sequence. The average amino acid radius is then approximated by $r = [V/(\pi h)]^{1/2}$, where h is the height (h = 3.63 Å). This forms the radius of the fully extended protein cylinder. The CCS is then given by the rotationally averaged CCS of this cylinder, since the protein "tumbles" in the drift tube of the apparatus, because of the low electric field.

The rotationally averaged CCS is given by the projection area of the cylinder:

$$\mathrm{CCS}_{\mathrm{upper}}\ (\text{Å}^2) = \text{projection area of cylinder} = \left(\frac{4}{\pi}\right)rl + 2r^2 \tag{3}$$

This figure is then multiplied by the helium scaling factor of 1.19 to give the final answer of the largest possible CCS.

These theoretical values are highly approximate and do not take into consideration disulfide bridges, proline residues, or

dx.doi.org/10.1021/ac5027435 | Anal. Chem. XXXX, XXX, XXX−XXX

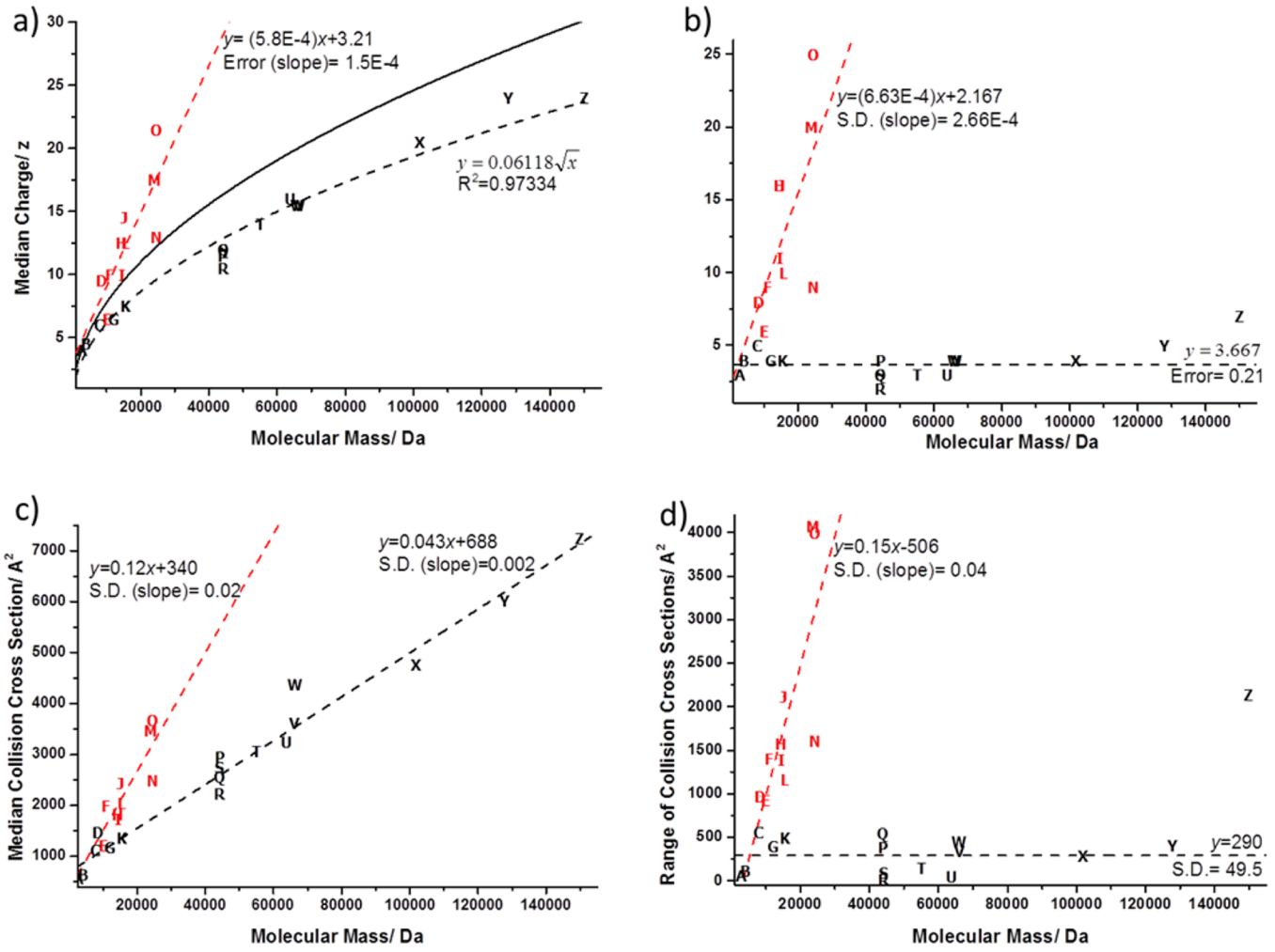


Figure 2. Schematic depiction of how the molecular mass of a protein relates to different mass spectrometry and ion mobility measurements: (a) median charge state, (b) range of charge states, (c) median collision cross section, and (d) range of collision cross sections. Proteins shown in red are known to be unfolded, either due to intrinsic disorder or denaturing conditions. Legend: A, melittin;[50] B, human beta-defensin 2;[83] C, lymphotactin (1−72);[53] D, ubiquitin (denatured);[84] E, lymphotactin;[53] F, N-terminal p53 (see Figure S1 and Table S4 in the Supporting Information); G, Cytochrome C (see Figure 1, as well as Table S2 in the Supporting Information); H, α-synuclein (see Figure S2 and Table S5 in the Supporting Information); I, N-terminal MDM2 (see Figure S3 and Table S6 in the Supporting Information); J, hemoglobin α_{apo};[48] K, hemoglobin α_{holo};[48] L, hemoglobin β_{apo};[48] M, β-casein (see Figure 1, as well as Table S3 in the Supporting Information); N, p53 DNA binding domain;[85] O, p53 DNA binding domain, pH 1.5;[85] P, hen egg white ovalbumin (reduced, conformation **1**) (see Figure S4 and Table S8 in the Supporting Information); Q, hen egg white ovalbumin (reduced, conformation **2**) (see Figure S4 and Table S8 in the Supporting Information); R, hen egg white ovalbumin (intact, conformation **1**) (see Figure S4 and Table S7 in the Supporting Information); S, hen egg white ovalbumin (intact, conformation **2**) (see Figure S4 and Table S7 in the Supporting Information); T, TTR (tetramer) (see Figure S5 and Table S9 in the Supporting Information); U, Avidin (tetramer) (see Figure S6 and Table S10 in the Supporting Information); V, BSA (conformation **1**) (see Figure S7 and Table S11 in the Supporting Information); W, BSA (conformation **2**) (see Figure S7 and Table S11 in the Supporting Information); X, Concanavalin A (tetramer) (see Figure S8 and Table S12 in the Supporting Information); Y, SAP (pentamer) (see Figure S9 and Table S13 in the Supporting Information); Z, Immunoglobulin G4 A[66]. The solid black line in panel (a) refers to the De la Mora relation,[59] whereas the dashed black line refers to the equation $y = 0.06118x^{1/2}$, which is fitted to our data with an R^2 value of 0.97334.[60]

other noncovalent interactions or restrictions, but they do serve as upper and lower boundaries with which to compare experimental data.

Determining the Population of Conformations from Charge State Distribution (CSD) and Collision Cross Section (CCS) Data. For any given CSD, the relative intensities of ions in the mass spectrum have been normalized against the base peak. These data are then plotted against the average CCS value for each charge state to provide the gas-phase conformational occupancy for each protein studied. The response factors of the mass spectrometer for any given charge and conformational state are considered equivalent.

RESULTS

Nano-electrospray Ionization (n-ESI) Ion Mobility–Mass Spectrometry (IM-MS) Signatures for a Structured Protein versus a Disordered Protein. We first illustrate the behavior of two exemplar proteins under native ESI-MS solution conditions where (i) Cytochrome C (12 229 Da) is known to be folded[60] and (ii) β-casein (23 980 Da) is known to be unstructured.[61] Figure 1A shows the mass spectrum of Cytochrome C, which presents a narrow CSD where $\Delta z = 4$, with most of the intensity in the $[M+7H]^{7+}$ ion. This narrow CSD is typical of a structured protein.[38,48] The dashed black line in Figures 1A and 1B, labeled Z_R, corresponds to the

dx.doi.org/10.1021/ac5027435 | Anal. Chem. XXXX, XXX, XXX−XXX

Rayleigh limit[59] which predicts the highest charge that a sphere of a given molecular mass can accommodate (see below). Because all of the charge states for Cytochrome C are lower than the Rayleigh limit, it can be assumed that all of the ionized proteins are roughly spherical in shape, although it is possible that a low charged anisotropic conformation may also fulfill this criterion. From ion mobility experiments, the rotationally averaged CCS for each charge state has been calculated (Figure 1C), and ranges from 952 Å^2 for the $[M+5H]^{5+}$ ion to 1326 Å^2 for the larger $[M+8H]^{8+}$ conformation, giving a ΔCCS of 374 Å^2. The collision cross section calculated from the crystal structure is shown by the dashed black line, which is slightly larger than the cross section populated by the gas-phase molecules. This indicates a compaction of the crystallographic structure in the gas phase, which is to be expected upon the removal of solvent.[62] By contrast, the mass spectrum of β-casein shows a very wide CSD from $[M+7H]^{7+}$ to $[M+27H]^{27+}$; $\Delta z = 21$ (Figure 1B). This appears typical of an unstructured protein.[40] The higher intensities for the lower charge states ($z = 9-12$) suggests that β-casein may exist predominantly in a compact state in solution, possessing less solvent-accessible area. (This point will be revisited later in the paper.) There are 16 peaks which correspond to ions with a charge state above the Rayleigh limit; we assume that these conformations are extended and therefore not spherical in shape. A very wide range of CCSs are found; from 1450 Å^2 for $[M+8H]^{8+}$ to 5519 Å^2 for $[M+26H]^{26+}$, providing a ΔCCS value of 4069 Å^2. Interestingly, there is slight compaction between $[M+26H]^{26+}$ and $[M+27H]^{27+}$. This could either be due to neutralization, by protonation, of two repulsing groups or perhaps due to the very highly charged species aligning in the drift tube. No crystal structure has been solved for the disordered protein β-casein; hence, we cannot reference our CCS measurements to data from an alternative structural technique as for Cyctochrome C.

Can the Median Charge (z) and the Width of the Charge State Distribution (Δz) Distinguish between Structured and Disordered Proteins? As exemplified by the test proteins Cytochrome C and β-casein above (Figure 1), structured proteins appear to occupy a small number of charge states (low Δz) with a low median value of z, whereas intrinsically disordered proteins occupy a wider range of charge states (high Δz) with a higher value of median z. IM-MS analysis was perfomed on 18 other proteins under a variety of solution conditions to test the generality of this phenomenon (see Figure 2). Figure 2A correlates the median charge state presented by a protein (the midpoint between the highest and lowest charge state observed) against the molecular mass. For structured proteins (denoted by black letters in the figure), there is a strong positive correlation between molecular mass and median charge state, which is fitted to a curve where $y = 0.6118x^{1/2}$. This correlation is expected, since the size of a protein increases with the surface area on which it can accommodate protons during ionization. The median charge states range from 4 for melittin (2.8 kDa) to 24 for Immunoglobulin G (150 kDa). The curve for the natively structured proteins initially is steep for proteins below 20 kDa, and then begins to level off at higher molecular weights.

The proteins that are expected to present more extended conformations, because of either intrinsic disorder or denaturing conditions (depicted in Figure 2A by red letters), have a distinctly higher median charge than folded proteins of a similar molecular mass, with just a small amount of overlap in the low M_w region. This dataset was best-fitted with a straight line function corresponding to $y = (5.8 \times 10^{-4})x + 3.21$, with a standard deviation to the slope of 1.5×10^{-4}. The significantly higher median charge for disordered or solvent-disrupted proteins versus molecular mass can be attributed to a higher number of solvent-accessible ionizable sites available for protonation in these proteins, compared with natively folded proteins, which have less ionizable residues available for protonation. The p53 DNA-binding domain studied here is the wild-type protein, rather than the superstable quadruple mutant designed by Nikolova et al.[63]

The black line in Figure 2a is the solution of an empirical relationship derived by De la Mora[59] over the mass range of our test set of proteins, which describes the limit to the maximum charge that a spherical protein can hold (eq 4):[59]

$$Z_R = 0.0778\sqrt{m} \quad (4)$$

where Z_R is the maximum charge and m is the mass of the protein. The median charge states of folded proteins under physiological-like conditions lie under this line, whereas the majority of the unstructured proteins lie above it (see Figure 2A). The De La Mora relationship, which is more commonly used to justify folded/compact low-charge-state species, can also be seen to provide a boundary between predominantly or completely folded/globular protiens and unfolded/extended proteins. Although the de la Mora equation lies substantially above the median charge state of the structured proteins, it correlates well with the maximum charge states of these proteins, as originally described;[59] we find a better fit to our median charge state data with a constant of 0.6118. A similar approach has been previously reported by Testa et al.,[30] although they did not consider any proteins larger than hemoglobin.

Figure 2B plots the number of charge states populated by a protein following n-ESI, (Δz). Here, the results are remarkable. Over the entire molecular mass range investigated, from 2846 Da to 150 kDa, folded proteins are present, with $\Delta z \leq 7$. By contrast, for disordered or denatured proteins, $\Delta z \geq 7$. It is surprising that (i) the charge state range of the natively folded proteins appears to be independent of molecular mass, and, in particular, (ii) even large proteins display a narrow charge state range following n-ESI. We might predict that for large proteins, protonation on one part of the molecule would be independent of protonation sites on other areas. However, Figure 2b demonstrates the following, when characterizing an unknown protein of molecular mass ≤100 kDa, using ESI-MS: (1) if it is present with more than 7 charge states, it can be assumed to have at least some disordered regions, or be denatured under the chosen solution conditions; or (2) if it presents with less than 7 charge states, it is more likely structurally stable. Again, we have fit the data for both the structured proteins and the disordered set; the fitting functions and standard deviations are found on the figure.

The comparison between charge state range and molecular mass was extended into the MDa range. (See Figure S10 in the Supporting Information.) It is apparent that, for protein complexes above 150 kDa, Δz increases slightly with molecular mass. Because Hepatitis B Virus Capsid T = 4 has a value of $\Delta z = 25$,[64] we can extrapolate a straight line fit from the data for Cytochrome C where $\Delta z = 4$, to the data for the 4MDa capsid; this gives a gradient of 5×10^{-6} (data not shown), as opposed to the gradient of zero used in the fit to our data (Figure 2b), showing how weak the correlation between mass and Δz is for ordered proteins. We can only speculate here on the reasons for

dx.doi.org/10.1021/ac5027435 | Anal. Chem. XXXX, XXX, XXX−XXX

the small Δz: it is probable that, during the last stages of desolvation, protonated side chains on the surface of a quasi-spherical protein will seek to pair with any proximal deprotonated acidic groups. Given the lowered dielectric of the vacuum compared with bulk solution, the range of electrostatic attraction between these groups will increase as the solvent is removed. As these groups pair, the charge on each is neutralized. The location of each chargeable group on the surface of the protein or protein complex will ultimately limit the ability of a given side chain to find a partner, and there will be side chains left unpaired. The net difference in the total number of unpaired protonated side chains (for positive ionization mode, as shown here) is apparently small for structured proteins, resulting in the low Δz that we observe. The energy in the pairing interactions will be stronger than that found in a solution salt bridge, and this will result in a caging effect to the surface of the protein. We might anticipate that such an effect would tighten the conformation of the protein compared to that found in bulk solution, and indeed that is what we find (see, for example, Figure 1 above for Cytochrome C and ref 62). This apparent ability of structured proteins to accommodate a discrete range of charges on their surface may be considered to be a partial solution to the Thompson problem.[65]

Can the Median Collision Cross Section (CCS) and the Width of the CCS (ΔCCS) Distinguish between Structured and Disordered Proteins? Plots similar to those in Figures 2a and 2b were constructed using ion mobility data. Again, the best fits to the data and the corresponding standard deviations are shown in the figure. The median CCS displayed by a protein, plotted against the molecular mass (Figure 2c), highlights a positive correlation between molecular mass and median CCS. Unfolded proteins have a tendency to have a higher median CCS, although there is some overlap for low masses as for the median charge (Figure 2a).

Figure 2d shows the range of collision cross sections for each protein (ΔCCS) plotted against molecular mass (M), and it is immediately obvious that this parameter shows very little dependence on molecular mass. With the exception of IgG ("Z" on plot), the presumed folded proteins consistently occupy a very narrow CCS range; the difference between the largest and smallest conformation is generally <550 Å^2, and the ΔCCS values scatter, with respect to M. This result is remarkable. Even SAP, a 128 kDa pentamer with a median CCS of over 6000 Å^2, has a ΔCCS value of just 408 Å^2. Conversely, unfolded proteins occupy an extremely wide range of collision cross sections, indicating conformational ensembles ranging from compact to extended states, as shown previously for β-caesin (Figure 1b). This supports the assertion made above in the CSD analysis, that a large Δz is found along with a large ΔCCS, and that both are signatures of disordered or denatured proteins. The narrow CCS range found for the protein complexes also supports the electrostatic caging effect that we have attributed to the small Δz found for structured proteins.

IgG, "Z" on the plot, has a ΔCCS value of 2140 Å^2, which is larger than that found for the other structured proteins. We attribute this to inherent conformational flexibility between different subunits.[66] Data obtained from static measurements via CryoEM and TEM has also indicated that IgGs are flexible and dynamic, capable of subunit waving and rotation, as well as bending and wagging.[67] The CCS range is actually even larger than shown in Figure2D, since the arrival time distribution (ATD) for each charge state for IgG is very wide, but the CCS shown here is taken from the apex of the ATD.[66]

Hen egg white ovalbumin and bovine serum albumin (BSA) each show two distinct conformations within one charge state (see Tables S7, S8, and S11 in the Supporting Information). Here, the two conformations have been treated separately. Both conformations of BSA are present in $[M+14H]^{14+}$ to $[M+17H]^{17+}$, spanning four charge states (Figure 2, letters "V" and "W"). The smaller conformation has a median cross section of 3615 Å^2 and a range of 345 Å^2 (V), while the larger conformation has a median cross section of 4390 Å^2 and a range of 446 Å^2 (W). Ovalbumin was analyzed in its reduced (P and Q) and oxidized state (R and S), since reduced ovalbumin has been previously reported to possess a similar structure to intact ovalbumin, albeit with significantly reduced conformational stability.[68,69] The conformations of the reduced form have median CCSs of 2565 Å^2 (Q) and 2964 Å^2 (P) and ranges of 679 Å^2 and 382 Å^2, respectively. The larger conformation (P) spans four charge states, while the smaller conformation spans three (Q) (see Figure S4, as well as Tables S7 and S8, in the Supporting Information). The smaller conformation of intact ovalbumin (R) spans charge states $[M+10H]^{10+}$ and $[M+11H]^{11+}$, has a median CCS of 2235 Å^2, and a ΔCCS value of 19 Å^2. The larger conformation (S) spans charge states $[M+11H]^{11+}$ to $[M+13H]^{13+}$, has a median CCS of 2752 Å^2, and a ΔCCS value of 99 Å^2. This agrees with previous work by Tatsumi[68] and Takahashi,[69] who proposed that reduction of the disulfide bond lowers conformational stability; the reduced form has a wider ΔCCS value in both conformations, although the median CCSs are similar to the oxidized form.

The CCS values of the proteins shown in Figures 2c and 2d are taken from the peak of the ATD for each resolvable conformation, or conformational family. The width of the ATDs may also be used to differentiate structured proteins from unstructured proteins (see Figure S11 in the Supporting Information), and this will form the subject of further study.

To summarize, we have shown that structured proteins and complexes below 130 kDa have $\Delta z \leq 5$ and the difference between the largest and smallest CCS is <550 Å^2. Proteins that are unfolded, either due to intrinsic disorder or to denaturing solvent conditions, have $\Delta z \geq 7$ and present with a range of conformations with a difference in CCS of >750 Å^2.

A Comparison of Experimental Findings with Theoretical Predictors of Disorder. Proteins that possess relatively high numbers of charged groups in solution have been linked with a tendency to be disordered. It is common to consider the net charge of the amino acids in the protein at a given solution pH versus their hydropathy as a predictor to distinguish ordered and disordered proteins.[70] One such method, based on the research by Uversky et al.,[70] proposes that a combination of high net charge and low hydrophobicity is an important contributing factor to intrinsic disorder in proteins. Here, we compare 10 of the investigated proteins, 5 structured and 5 disordered, with the charge-hydropathy scale (provided by PONDR, Figure 3A). We also compare the calculated parameters (mean net charge and mean scaled hydrophobicity), separately, with the Δz values that we derive experimentally.

Two proteins were assigned by charge-hydropathy analysis as being disordered; N-terminal p53 and Cytochrome C. N-terminal p53 has also been shown to be devoid of tertiary structure by several biophysical techniques,[71,72] and here we have shown that it presents a wide range of Δz values (Figure

dx.doi.org/10.1021/ac5027435 | Anal. Chem. XXXX, XXX, XXX−XXX

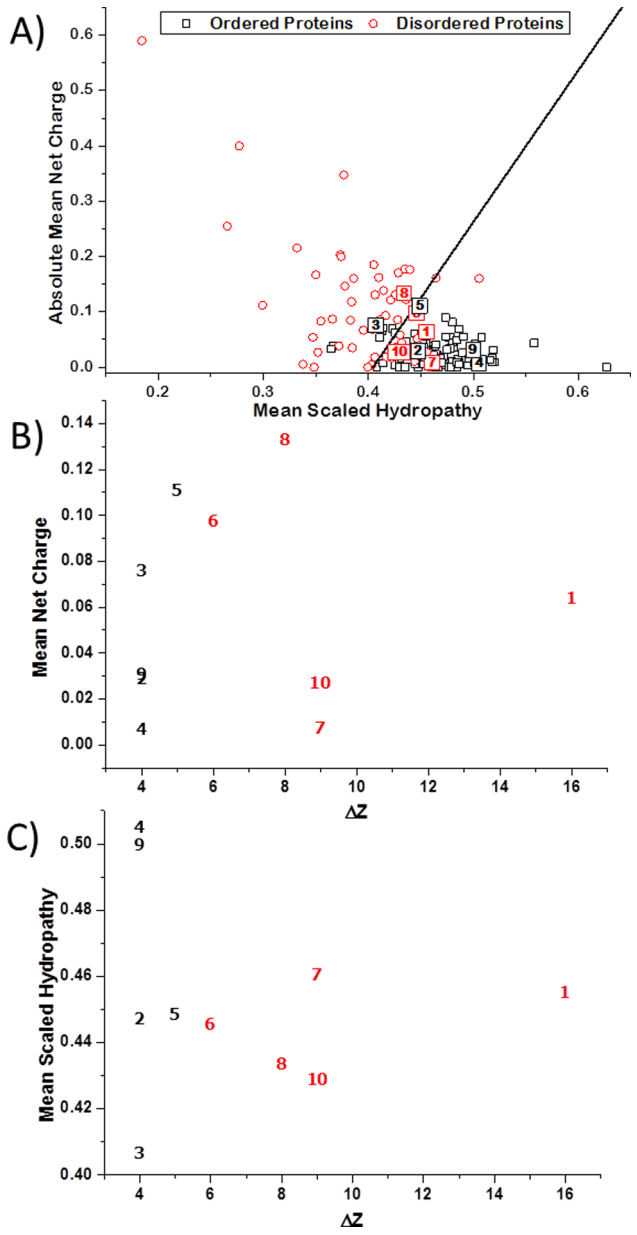


Figure 3. (A) Comparison of the mean net charge and the mean scaled hydrophobicity of 10 of the investigated proteins. Comparisons of the charge state range are made to (B) the mean net charge and (C) the mean scaled hydrophobicity. Legend: (1) α-synuclein, (2) bovine serum albumin, (3) Cytochrome C, (4) hemoglobin-α chain, (5) lymphotactin (1−72), (6) WT lymphotactin, (7) N-terminal MDM2, (8) N-terminal p53, (9) hen egg white ovalbumin, and (10) p53 DNA-binding domain. Image adapted from Uversky et al.[70] (Copyright 2000, Wiley−Blackwell). Proteins shown in red are known to be unfolded, either due to intrinsic disorder or denaturing conditions. Proteins shown in black have been previously shown to be structured in solution.

2). Native Cytochrome C is folded into a compact conformation with a covalently bound heme group, required for the folding of Cytochrome C. Without this heme group the protein loses the majority of its secondary and tertiary structure,[73,74] and this heme group is not taken into account in the charge-hydropathy analysis, which can explain the discrepancy between mass spectrometry and charge-hydropathy results. Lymphotactin WT and lymphotactin (1−72) both sit on the boundary between ordered and disordered. Lymphotactin (1−72) lacks the disordered tail that the WT contains,[75] so a bigger difference would be expected between the two forms of the protein, as is shown by the IM-MS outputs of Δz and ΔCCS (see Figures 3B and 3C).

α-Synuclein, p53 DBD, and NT-MDM2 are all classified by this charge-hydropathy analysis as being ordered. This is inconsistent with our MS results, as well as the perceived wisdom from a range of other biophysical techniques.[42,72,76−78] Bovine serum albumin, hemoglobin-α, and ovalbumin are assigned as structured by charge-hydropathy analysis, and this is in agreement with the IM-MS data.

Figures 3b and 3c both show that there appears to be little differentiation of net charge or hydropathy with respect to whether the protein is folded or intrinsically disordered. Our experimental results, however, illustrate the strength of experiment compared with this theoretical approach; there is clear distinction of folded and disordered proteins, with respect to the experimental parameters Δz which is compared to charge, and ΔCCS which is compared to hydropathy.

We also used two more methods of predicting disorder from the primary sequence that are more sophisticated than the charge-hydropathy analysis, namely, other versions of PONDR (PONDR-VLXT PONDR-VSL2 and PONDR-VL3) and DISOPRED.[55] PONDR uses the primary sequence to evaluate attributes such as the sequence complexity and hydrophobicity, which are then used as inputs to make predictions. The predictors have been trained on sets of disordered and ordered sequences, which allows them to generalize new sequences. There are then several outputs that differ in the method of characterization of the training data, such as X-ray crystallography or NMR.[79] DISOPRED3, on the other hand, was trained on a set of sequences with high-resolution X-ray structures. Disorder is identified by residues that are in the sequence but have coordinates missing from the crystal structure.[55] We carried out PONDR and DISOPRED analysis on a set of six proteins: three structured proteins (Cytochrome C, bovine serum albumin, and ovalbumin) and three disordered proteins (α-synuclein, N-terminal MDM2, and p53 DNA-binding domain). Table 1 shows a summary of the analysis, which is displayed in Figure S11 in the Supporting Information.

α-Synuclein is classified as being over 52% disordered by all of the prediction methods, which fits well with both our analysis and that from other research techniques.[42,80] Disorder predictions of MDM2 range from 26% to 36% disordered, and all predictors display a structured region between amino acids 42 and 111 (Figure S11 in the Supporting Information). p53-DBD is predicted to be disordered to an extent of 11%−42%, which is lower than we would expect from such a wide range of Δz and ΔCCS values. Surprisingly, the prediction values of this IDP are similar to that of BSA (9%−38%), which is characterized as being structured by a variety of techniques. The predictions of the extent of disorder in ovalbumin are very low, as expected, with values ranging from 0.2% to 24%. Cytochrome C has a wide range of predictions, from 8% to 54%, and one reason for these discrepancies could be due to the covalently bound heme group, which is known to stabilize the tertiary structure of the protein but will not be taken into consideration by the information in the primary sequence, as described above. Unlike the rather binary output provided by our ion mobility mass spectrometry analysis (recall Figures 2A−D), where we classify proteins as either ordered or

dx.doi.org/10.1021/ac5027435 | Anal. Chem. XXXX, XXX, XXX−XXX

Table 1. Percentage of Disordered Residues in Six Different Proteins, as Predicted by DISOPRED, PONDR-VL3, PONDR-VSL2, and PONDR VLXT[a]

protein	% disordered DISOPRED	% disordered PONDR-VL3	% disordered PONDR-VSL2	% disordered PONDR VLXT
α-synuclein	56	100	91	52
N-terminal MDM2	31	26	35	36
P53 DNA binding domain	11	37	42	26
Cytochrome C	8	50	54	9
ovalbumin	0.2	5	18	24
bovine serum albumin	9	38	32	18

[a]Data extrapolated from Figure S12 in the Supporting Information.

disordered, these disorder predictors seek to capture the progression of structure.

How Much of the Possible Conformational Space Is Occupied in the Presented Collision Cross Sections? In order to put the ion mobility results into context, and to start to consider the progression of structure from ordered to disordered, we apply a simple model to consider the most extended and most compact possible conformations. The largest and smallest cross sections are calculated based on the volume of the amino acids in the polypeptide, arranged in either a spherical (compact) or cylindrical (extended) configuration. These calculated cross sections are compared to the maximum and minimum CCSs seen in ion mobility experiments (see Figure 4).

This comparison (Figure 4) reveals several interesting features. The ordered proteins (Cytochrome C, myoglobin, and lysozyme), have experimentally measured ΔCCS under native conditions that are significantly narrower than the limits from the model. This can be rationalized by considering that these proteins possess high degrees of tertiary structure under the experimental solution conditions, which prevent the proteins from unfolding. The CCS values measured for these structured proteins have a tendency to be toward the lower end of the allowed CCS range, which implies that they prefer to adopt compact conformations; although Cytochrome C is theoretically able to occupy cross sections ranging from 759 Å^2 to 3632 Å^2, the experimental results show that, under native conditions, the cross sections only range from 952 Å^2 to 1326 Å^2.

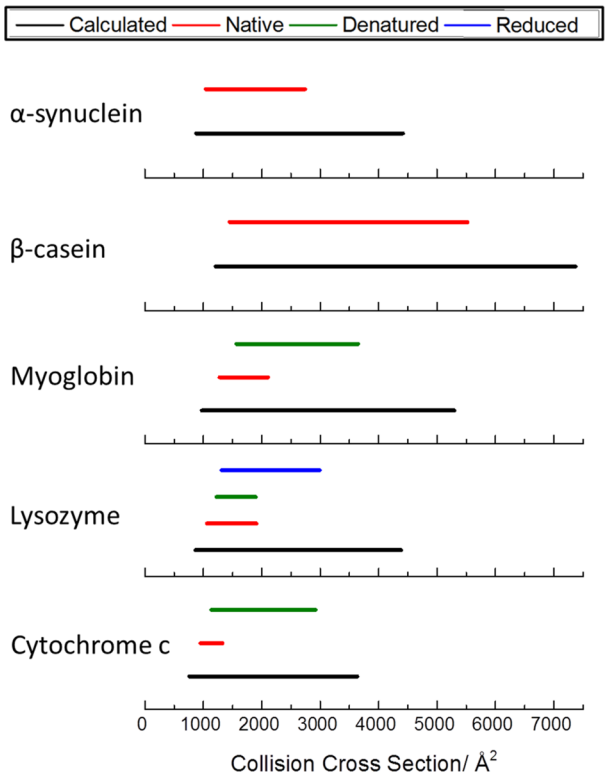


Figure 4. Comparison of the theoretical and experimental collision cross sections: The black bars show the theoretical CCS range. Cytochrome C, lysozyme, and myoglobin are examples of structured proteins. Denatured lysozyme is constrained with disulfide bonds; β-casein and α-synuclein are examples of IDPs. (See Table S14 in the Supporting Information.)

When lysozyme is sprayed from more denaturing conditions (water:methanol:formic acid ratio = 49:50:1), the smallest CCS measured increases from 1016 Å^2 (with ammonium acetate) to 1222 Å^2, although the upper end remains very similar (1906 Å^2, compared to 1895 Å^2). This suggests that the salt provides conditions that stabilize the most compact conformations, but the more hydrophobic solvent environment and decreased pH does not allow lysozyme to present extended conformations (as for Cyctochrome C), because of the fact that it is constrained by four intramolecular disulfide bonds. Under reducing conditions, the experimental upper CCS value increases considerably, to 2989 Å^2, as the four disulfide bridges are cleaved.

Reduced lysozyme, denatured Cytochrome C, and denatured myoglobin show good agreement between their experimentally measured CCSs and their respective theoretically permitted ranges. This can be attributed to solution conditions, which destabilize their tertiary structure, inducing unfolding and, consequently, allowing the protein to adopt a range of conformations, from very compact to very extended.

The disordered proteins, β-casein and α-synuclein, behave as predicted. They explore most of their predicted range of CCS, even when sprayed from high ionic strength solution conditions, because they are intrinsically disordered with little stabilizing secondary or tertiary structure. They present to the gas phase with many different conformations, ranging from compact to extended, unhindered by energetic constraints. They are qualitatively very similar to structured proteins sprayed from denaturing conditions.

Can We Quantify the Relative Population of Conformations from Compact to Extended? To provide more detail on the nature of the observed states of a given protein, and the extent of structure or disorder, the relative intensities of the different charge states (as shown in a CSD) will be more informative than simply considering the width of the CSD alone. Plotting CSD along with CSS data allows us to construct postulated energy landscapes. A protein that populates a wide range of CCSs, with similar intensities of ions across the entire range, will denote a polypeptide chain sampling many different

dx.doi.org/10.1021/ac5027435 | Anal. Chem. XXXX, XXX, XXX−XXX

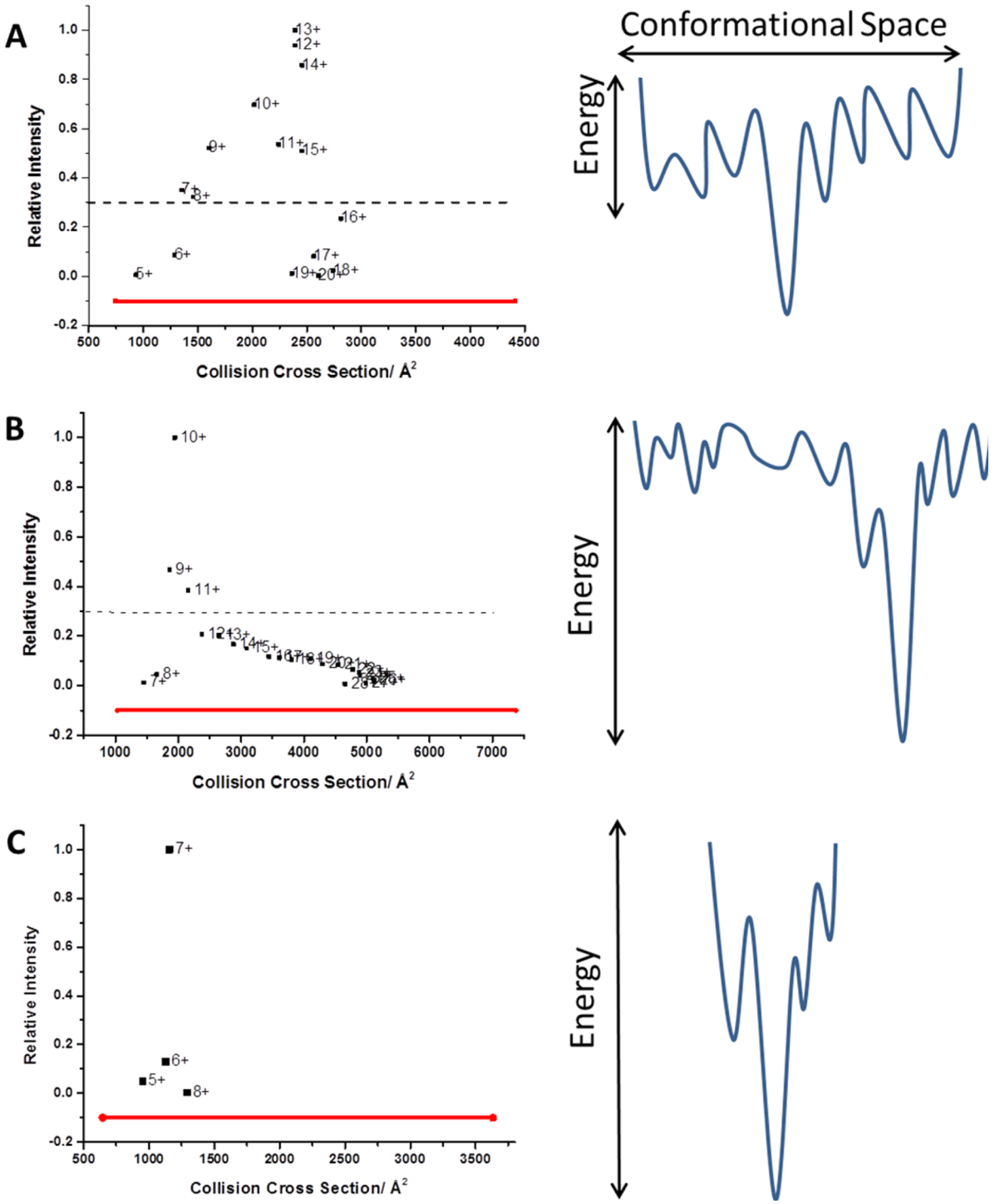


Figure 5. Range of theoretical CCSs occupied by a protein ((A) α-synuclein, (B) β-casein, and (C) Cytochrome C) is shown in red. The relative intensity of each charge state is plotted against its CCS. On the right-hand side, the illustration shows a postulated shape of the folding landscape of each protein.

conformations, with little preference for any particular one. This is exemplified by α-synuclein (Figure 5A), which samples almost the entire range of theoretically possible conformations. Nine charge states ($[M+7H]^{7+}$ to $[M+15H]^{15+}$) out of the 16 ($[M+5H]^{5+}$ to $[M+20H]^{20+}$) have more than 30% intensity of the base peak (shown by the dashed line in Figures 5A and 5B).

It has previously been reported that β-casein contains areas of structure and disorder.[61,81] Livney et al. described the structure as having several definable parts, with others that are more flexible and dynamic, allowing transformation between several energetically favorable protein conformations. This is reflected in the intensities of the different charge states and corresponding CCSs of the protein. The most intense peaks correspond to $[M+9H]^{+}$, $[M+10H]^{10+}$, and $[M+11H]^{11+}$, which are of low charge and low CCS (1862−2157 Å^2), relative to the majority of the peaks. These correspond to the protein in an energetically stable conformation, demonstrated by the troughs in the postulated energy landscape shown in

dx.doi.org/10.1021/ac5027435 | *Anal. Chem.* XXXX, XXX, XXX−XXX

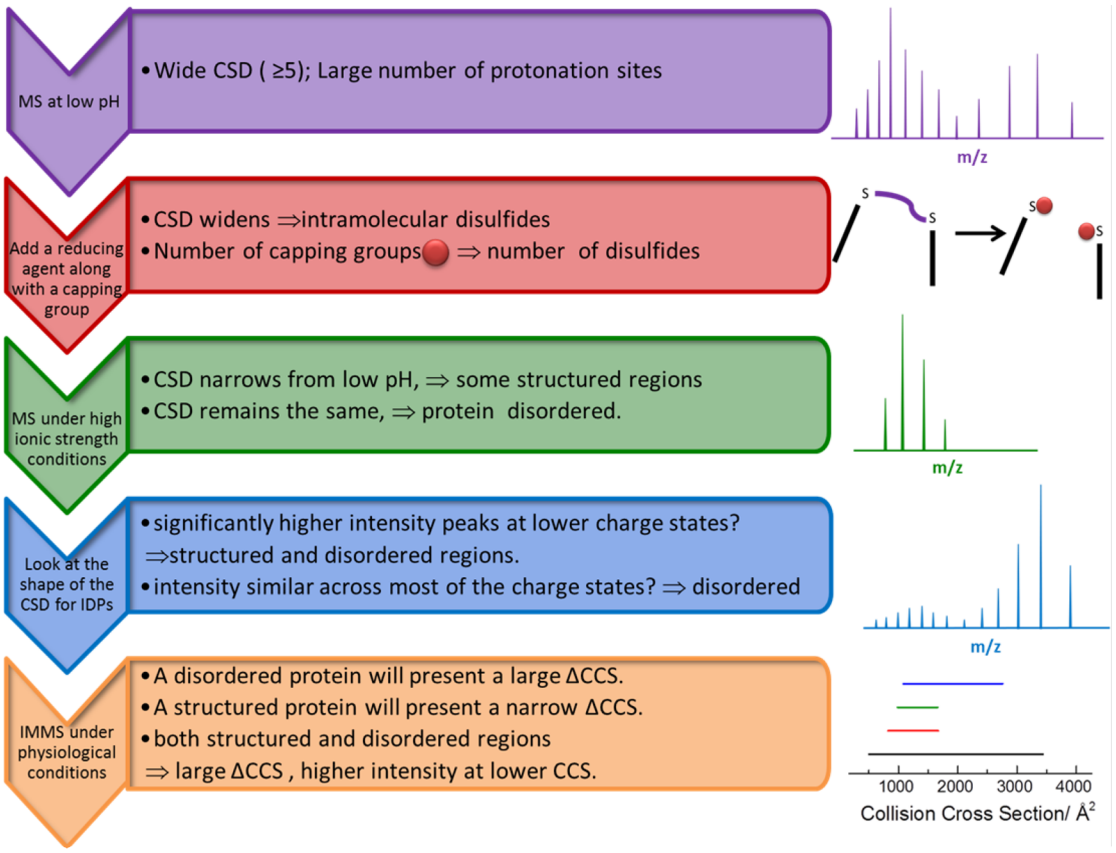


Figure 6. Flowchart displaying the procedure for elucidating structural information on an unknown protein.

Figure 5B. The peaks of higher charge and larger CCS are of much lower intensity, suggesting a dynamic ensemble of transiently populated conformations. Just 3 charge states out of 26 ($[M+9H]^{+}$, $[M+10H]^{10+}$, and $[M+11H]^{11+}$) have more than 30% intensity of the base peak.

A structured protein can be considered to possess a landscape with one low energy minima, in which most of the protein molecules will be situated.[82] The corresponding mass spectrum of a protein with this funnel-shaped landscape will display a narrow charge state distribution of five peaks or less, corresponding to one single conformational family with minimum dynamics. This is demonstrated by Cytochrome C in Figure 5C. Cytochrome C has been reported to possess a narrow folding funnel with one energetically favorable conformational family.[60] This is demonstrated by the narrow CSD, with the dominant species being the $[M+7H]^{7+}$ ion with low CCS values compared to the theoretical CCS values available to the polypeptide chain.

Here, we have demonstrated how the intensities of different charge states of a protein can be used to make structural predictions of proteins with varying extents of disorder.

The Framework and How To Determine the Structural Preferences of an Unknown Protein. The objective of this investigation was to determine what structural information could be obtained about an unknown protein if mass spectrometry was the only method available (see Figure 6). The first step is to perform an ESI-MS experiment at low pH. From this, it is facile to determine the mass of the protein. Through the addition of a reducing agent and a thiol capping group, one is able to determine whether intramolecular disulfide bonds are present within the protein. This is because the protein conformation—and, therefore, the CSD—will be altered if disulfide bonds are broken. The number of capping groups seen in the mass shift will reflect how many free cysteine residues are present. The CSD will normally be wide, because of the protein being denatured. Returning to the oxidized form, the next step would be to conduct an ESI-MS experiment under high ionic strength conditions, and at a range of pHs. If the protein has any regions of structure under these physiological-like conditions, the CSD will narrow. If the CSD remains the same as that under low pH, this indicates an intrinsically disordered protein with no structure to be disrupted by the low pH. Significantly more populated peaks at lower charge states under physiological conditions indicates a protein that has regions of order (compactness) and disorder (extended states), whereas a natively unstructured IDP will display roughly equal intensity across the entire range of charge states (see Figure 6).

Information obtained from mass spectrometry can be complemented by ion mobility experiments, which provide further information on conformational occupancy of the protein, in the form of a rotationally averaged CCS for each charge state, which can be compared to data obtained from other methods. A structured protein will present a narrow range of CCSs, a disordered protein will present a wide range of CCSs, and a protein with both structured and disordered regions will present a large range of CCSs with higher intensity of more-compact conformations.

■ CONCLUSIONS

The objective of this research was to provide a framework for mass spectrometry (MS) experiments that could be performed

dx.doi.org/10.1021/ac5027435 | Anal. Chem. XXXX, XXX, XXX−XXX

as an initial characterization technique to determine the dynamic properties of a protein for which no structural information is known. To observe trends in ion mobility−mass spectrometry (IM-MS) results with respect to structure and disorder, MS and IM-MS results were compared for a set of 20 proteins, spanning a molecular mass range of over 147 kDa with varying extents of secondary tertiary and quaternary structure (see Figures 2A−D). It was noted that, while structured proteins have a median charge lower than the De la Mora/Rayleigh limit, disordered proteins have a higher median charge and, more importantly, display a wider charge state distribution than structured proteins. For proteins with a mass of M <150 kDa, structured proteins present with $z \leq 7$ states or less, whereas disordered proteins will present with $z \geq 7$. The low Δz for structured proteins is only very weakly dependent on mass, and, here, we have suggested that it correlates to a polyelectrostatic view of charge pairing on desolvation.

Comparison of collision cross sections (CCSs) measured by IM-MS show that, although there is little distinction in terms of median CCS between structured and disordered systems, the range of CCSs that a given protein presents reveals its conformational flexibility; structured proteins will have a CCS range of <750 Å^2, whereas disordered systems will have a CCS range of >750 Å^2.

Comparisons were then made between measured CCSs and theoretically calculated CCSs for a set of five proteins, three of which are structured and two of which are disordered (see Figure 3). When analyzed under native-like conditions, the structured proteins display a much narrower CCS range than is *theoretically* available and, moreover, these are at the smaller end of the allowed CCS range. When denatured and, in the case of proteins containing disulfide bonds, reduced, the measured cross-section range increases to almost all of the space that is theoretically available. In the case of IDPs, however, the CCS range is wide, even under native-like conditions, because they are free to access most of their conformations. When the shape of the charge state distribution (CSD) was considered along with the CCS range, it is apparent that IDPs will have a wide CSD with similar intensities across the charge states, structured proteins will have a narrow CSD with most of the intensity in just one or two charge states, and proteins with regions of both structure and disorder, or a tendency to disorder, will have a wide CSD, with most of the intensity in the lower charge states (see Figure 4). From this information, we have built a framework of exemplar experiments, shown in Figure 6. This manuscript allows us to propose a workflow for the use of MS and IM-MS as initial characterization techniques to obtain information about the extent of structure or disorder in unknown proteins. In this initial investigation we have primarily selected systems that are either distinctively structured or unstructured, future work will focus on more conformationally diverse proteins.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Protein sequences, mass spectra, arrival time distributions and tables of experimental collision cross sections obtained from DT IM-MS experiments, and an extension of the Δz versus molecular mass relationship for other proteins can be found in the Supporting Information. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: Perdita.barran@manchester.ac.uk.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The Schools of Chemistry and Physics and Astronomy at the University of Edinburgh, and the School of Chemistry at the University of Manchester are thanked for an award of an Biotechnology and Biological Research Council (BBSRC) case studentship to R.B., and for their continuing support of our research. LGC, Ltd., is thanked for providing R.B. with a case top up. The BBSRC are also thanked for funding the following projects, which, in part, supported the work herein: Nos. BB/L002655/1, BB/L002655/1, and BB/L015048/1. We thank the following for supplying us materials: Brian Volkmann, Medical College of Wisconsin; Ted Hupp, University of Edinburgh; Tilo Kunath, University of Edinburgh; Galina Selivanova, The Karolinska Institute; and UCB Pharma and all Barran group members past and present who have contributed to these studies. We thank Frédéric Cazals (Inria Research Centre, France) for suggesting that structured proteins may "solve" the Thompson problem and Jim Warwicker (MIB, Manchester, U.K.) for helpful discussions on polyelectrostatics.

■ REFERENCES

(1) Serdyuk, I. N. *Mol. Biol. (Moscow, Russ. Fed., Engl. Ed.)* **2007**, *41*, 297−313.
(2) Lemieux, R. U.; Spohr, U. *Adv. Carbohydr. Chem. Biochem.* **1994**, *50*, 1−20.
(3) Kendrew, J. C.; Dickerson, R. E.; Strandberg, B. E.; Hart, R. G.; Davies, D. R.; Phillips, D. C.; Shore, V. C. *Nature* **1960**, *185*, 422−427.
(4) Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. C. T. *Nature* **1960**, *185*, 416−422.
(5) Karush, F. *J. Am. Chem. Soc.* **1950**, *72*, 2705−2713.
(6) Koshland, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 98−104.
(7) Chen, J. *Arch. Biochem. Biophys.* **2012**, *524*, 123−131.
(8) Wright, P. E.; Dyson, H. J. *J. Mol. Biol.* **1999**, *293*, 321−331.
(9) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197−208.
(10) Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L. *Curr. Opin. Struct. Biol.* **2008**, *18*, 756−764.
(11) Tompa, P. *Trends Biochem. Sci.* **2002**, *27*, 527−533.
(12) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradović, Z. *Biochemistry* **2002**, *41*, 6573−6582.
(13) Uversky, V. N. *Protein Sci.* **2002**, *11*, 739−756.
(14) Laity, J. H.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2000**, *295*, 719−727.
(15) Uversky, V. In *Protein Folding and Misfolding: Neurodegenerative Diseases*, Ovádi, J., Orosz, F., Eds.; Springer: Dordrecht, The Netherlands, 2009; pp 21−75.
(16) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. In *Annual Review of Biophysics*; Annual Reviews: Palo Alto, CA, 2008; pp 215−246.
(17) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. R.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C. H.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, M.; Garner, E. C.; Obradovic, Z. *J. Mol. Graph. Modell.* **2001**, *19*, 26−59.
(18) Williams, R. M.; Obradovi, Z.; Mathura, V.; Braun, W.; Garner, E. C.; Young, J.; Takayama, S.; Brown, C. J.; Dunker, A. K. *Pac. Symp. Biocomput. 2001* **2001**, 89−100.
(19) Toniolo, C.; Bonora, G. M.; Mutter, M.; Pillai, V. N. R. *Makromol. Chem.* **1981**, *182*, 2007−2014.
(20) Uversky, V. N. *Curr. Alzheimer Res.* **2008**, *5*, 260−287.

(21) Iakoucheva, L. M.; Kimzey, A. L.; Masselon, C. D.; Bruce, J. E.; Garner, E. C.; Brown, C. J.; Dunker, A. K.; Smith, R. D.; Ackerman, E. J. *Protein Sci.* **2001**, *10*, 560−571.
(22) Zhang, Z. Q.; Smith, D. L. *Protein Sci.* **1993**, *2*, 522−531.
(23) Eliezer, D.; Yao, J.; Dyson, H. J.; Wright, P. E. *Nat. Struct. Biol.* **1998**, *5*, 148−155.
(24) Eliezer, D.; Kutluay, E.; Bussell, R., Jr; Browne, G. *J. Mol. Biol.* **2001**, *307*, 1061−1073.
(25) Glushka, J.; Lee, M.; Coffin, S.; Cowburn, D. *J. Am. Chem. Soc.* **1989**, *111*, 7716−7722.
(26) Kaltashov, I. A.; Mohimen, A. *Anal. Chem.* **2005**, *77*, 5370−5379.
(27) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64−71.
(28) Beveridge, R.; Chappuis, Q.; Macphee, C.; Barran, P. *Analyst* **2013**, *138*, 32−42.
(29) Sharon, M.; Robinson, C. V. In *Annual Review of Biochemistry*; Annual Reviews: Palo Alto, CA, 2007; pp 167−193.
(30) Testa, L.; Brocca, S.; Grandori, R. *Anal. Chem.* **2011**, *83*, 6459−6463.
(31) Keppel, T. R.; Howard, B. A.; Weis, D. D. *Biochemistry* **2011**, *50*, 8722−8732.
(32) Heck, A. J. R. *Nat. Meth* **2008**, *5*, 927−933.
(33) Wilm, M.; Mann, M. *Anal. Chem.* **1996**, *68*, 1−8.
(34) Karas, M.; Bahr, U.; Dülcks, T. *Fresenius J. Anal Chem.* **2000**, *366*, 669−676.
(35) Konermann, L. *J. Phys. Chem. B* **2007**, *111*, 6534−6543.
(36) Hall, Z.; Politis, A.; Robinson, C. V. *Structure* **2012**, *20*, 1596−1609.
(37) Kaltashov, I. A.; Abzalimov, R. R. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1239−1246.
(38) Dobo, A.; Kaltashov, I. A. *Anal. Chem.* **2001**, *73*, 4763−4773.
(39) Gabelica, V.; Galic, N.; Rosu, F.; Houssier, C.; De Pauw, E. *J. Mass Spectrom.* **2003**, *38*, 491−501.
(40) Frimpong, A. K.; Abzatimov, R. R.; Uversky, V. N.; Kaltashov, I. A. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 714−722.
(41) Kuprowski, M. C.; Konermann, L. *Anal. Chem.* **2007**, *79*, 2499−2506.
(42) Uversky, V. N. *J. Biomol. Struct. Dyn.* **2003**, *21*, 211−234.
(43) Sterling, H.; Batchelor, J.; Wemmer, D.; Williams, E. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1045−1049.
(44) Uetrecht, C.; Rose, R. J.; van Duijn, E.; Lorenzen, K.; Heck, A. J. R. *Chem. Soc. Rev.* **2010**, *39*, 1633−1655.
(45) Snijder, J.; Rose, R. J.; Veesler, D.; Johnson, J. E.; Heck, A. J. R. *Angew. Chem., Int. Ed.* **2013**, *52*, 4020−4023.
(46) Kuloğlu, E. S.; McCaslin, D. R.; Kitabwalla, M.; Pauza, C. D.; Markley, J. L.; Volkman, B. F. *Biochemistry* **2001**, *40*, 12486−12496.
(47) 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*, 7765−7769.
(48) Faull, P. A.; Korkeila, K. E.; Kalapothakis, J. M.; Gray, A.; McCullough, B. J.; Barran, P. E. *Int. J. Mass Spectrom.* **2009**, *283*, 140−148.
(49) 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**, 7899−7903 (DOI: 10.1002/ange.201402863).
(50) Florance, H. V.; Stopford, A. P.; Kalapothakis, J. M.; McCullough, B. J.; Bretherick, A.; Barran, P. E. *Analyst* **2011**, *136*, 3446−3452.
(51) Jurneczko, E.; Cruickshank, F.; Porrini, M.; Clarke, D. J.; Campuzano, I. D. G.; Morris, M.; Nikolova, P. V.; Barran, P. E. *Angew. Chem., Int. Ed.* **2013**, *52*, 4370−4374.
(52) McCullough, B. J.; Kalapothakis, J.; Eastwood, H.; Kemper, P.; MacMillan, D.; Taylor, K.; Dorin, J.; Barran, P. E. *Anal. Chem.* **2008**, *80*, 6336−6344.
(53) Harvey, S. R.; Porrini, M.; Konijnenberg, A.; Clarke, D. J.; Tyler, R. C.; Langridge-Smith, P. R. R.; MacPhee, C. E.; Volkman, B. F.; Barran, P. E. *J. Phys. Chem. B* **2014**, in press (DOI: 10.1021/jp504997k).
(54) Kalapothakis, J.; Barran, P. In *Encyclopedia of Biophysics*, Roberts, G. C. K., Ed.; Springer: Berlin, New York, 2013.
(55) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. *J. Mol. Biol.* **2004**, *337*, 635−645.
(56) Fischer, H.; Polikarpov, I.; Craievich, A. F. *Protein Sci.* **2004**, *13*, 2825−2828.
(57) de la Mora, J. F.; de Juan, L.; Eichler, T.; Rosell, J. *TrAC: Trends Anal. Chem.* **1998**, *17*, 328−339.
(58) Tammet, H. *J. Aerosol Sci.* **1995**, *26*, 459−475.
(59) Fernandez de la Mora, J. *Anal. Chim. Acta* **2000**, *406*, 93−104.
(60) Lyubovitsky, J. G.; Gray, H. B.; Winkler, J. R. *J. Am. Chem. Soc.* **2002**, *124*, 5481−5485.
(61) Yousefi, R.; Shchutskaya, Y. Y.; Zimny, J.; Gaudin, J.-C.; Moosavi-Movahedi, A. A.; Muronetz, V. I.; Zuev, Y. F.; Chobert, J.-M.; Haertlé, T. *Biopolymers* **2009**, *91*, 623−632.
(62) Jurneczko, E.; Barran, P. E. *Analyst* **2011**, *136*, 20−28.
(63) Nikolova, P. V.; Henckel, J.; Lane, D. P.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 14675−14680.
(64) Uetrecht, C.; Versluis, C.; Watts, N. R.; Roos, W. H.; Wuite, G. J. L.; Wingfield, P. T.; Steven, A. C.; Heck, A. J. R. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 9216−9220.
(65) Thompson, J. J. *Philos. Mag.* **1904**, *7*, 237−265.
(66) Pacholarz, K. J.; Porrini, M.; Garlish, R. A.; Burnley, R. J.; Taylor, R. J.; Henry, A. J.; Barran, P. E.. Unpublished work, 2014.
(67) Brekke, O. H.; Michaelsen, T. E.; Sandlie, I. *Immunology Today* **1995**, *16*, 85−90.
(68) Tatsumi, E.; Yoshimatsu, D.; Hirose, M. *Biosci., Biotechnol., Biochem.* **1999**, *63*, 1285−1290.
(69) Takahashi, N.; Koseki, T.; Doi, E.; Hirose, M. *J. Biochem.* **1991**, *109*, 846−851.
(70) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. *Proteins: Struct., Funct., Bioinf.* **2000**, *41*, 415−427.
(71) Dawson, R.; Muller, L.; Dehner, A.; Klein, C.; Kessler, H.; Buchner, J. *J. Mol. Biol.* **2003**, *332*, 1131−1141.
(72) Bell, S.; Klein, C.; Muller, L.; Hansen, S.; Buchner, J. *J. Mol. Biol.* **2002**, *322*, 917−927.
(73) Fisher, W. R.; Taniuchi, H.; Anfinsen, C. B. *J. Biol. Chem.* **1973**, *248*, 3188−3195.
(74) Stellwagen, E.; Rysavy, R.; Babul, G. *J. Biol. Chem.* **1972**, *247*, 8074−8077.
(75) Tuinstra, R. L.; Peterson, F. C.; Elgin, E. S.; Pelzek, A. J.; Volkman, B. F. *Biochemistry* **2007**, *46*, 2564−2573.
(76) Worrall, E.; Wawrzynow, B.; Worrall, L.; Walkinshaw, M.; Ball, K.; Hupp, T. *J. Chem. Biol.* **2009**, *2*, 113−129.
(77) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradovic, Z.; Dunker, A. K. *J. Mol. Biol.* **2002**, *323*, 573−584.
(78) Uhrinova, S.; Uhrin, D.; Powers, H.; Watt, K.; Zheleva, D.; Fischer, P.; McInnes, C.; Barlow, P. N. *J. Mol. Biol.* **2005**, *350*, 587−598.
(79) Xue, B.; Dunbrack, R. L.; Williams, R. W.; Dunker, A. K.; Uversky, V. N. *Biochim. Biophys. Acta, Proteins Proteomics* **2010**, *1804*, 996−1010.
(80) Bussell, R.; Eliezer, D. *J. Biol. Chem.* **2001**, *276*, 45996−46003.
(81) Livney, Y. D.; Schwan, A. L.; Dalgleish, D. G. *J. Dairy Sci.* **2004**, *87*, 3638−3647.
(82) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10−19.
(83) Cecco, M. D. *Biophysical Studies to Elucidate Structure−Activity Relationships in Beta Defensins*; Univeristy of Edinburgh: Edinburgh, U.K., 2011.
(84) Li, J. W.; Taraszka, J. A.; Counterman, A. E.; Clemmer, D. E. *Int. J. Mass Spectrom.* **1999**, *185*, 37−47.
(85) Jurneczko, E.; Cruickshank, F.; Porrini, M.; Nikolova, P.; Campuzano, I. D. G.; Morris, M.; Barran, P. E. *Biochem. Soc. Trans.* **2012**, *40*, 1021−U1372.