

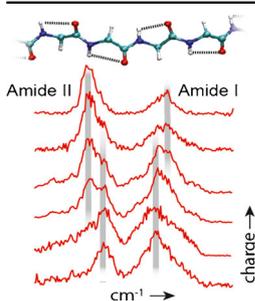
# From Compact to String—The Role of Secondary and Tertiary Structure in Charge-Induced Unzipping of Gas-Phase Proteins

Stephan Warnke,<sup>1</sup> Waldemar Hoffmann,<sup>2</sup> Jongcheol Seo,<sup>1</sup> Erwin De Genst,<sup>3</sup>  
Gert von Helden,<sup>1</sup> Kevin Pagel<sup>1,2</sup>

<sup>1</sup>Fritz Haber Institute of the Max Planck Society, Faradayweg 4-6, 14195, Berlin, Germany

<sup>2</sup>Institute of Chemistry and Biochemistry, Freie Universität Berlin, Takustrasse 3, 14195, Berlin, Germany

<sup>3</sup>Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK



**Abstract.** In the gas phase, protein ions can adopt a broad range of structures, which have been investigated extensively in the past using ion mobility-mass spectrometry (IM-MS)-based methods. Compact ions with low number of charges undergo a Coulomb-driven transition to partially folded species when the charge increases, and finally form extended structures with presumably little or no defined structure when the charge state is high. However, with respect to the secondary structure, IM-MS methods are essentially blind. Infrared (IR) spectroscopy, on the other hand, is sensitive to such structural details and there is increasing evidence that helices as well as  $\beta$ -sheet-like structures can exist in the gas phase, especially for ions in low charge states. Very recently, we showed that also the fully extended form of highly charged protein ions can adopt a distinct type of secondary structure that features a characteristic  $C_5$ -type hydrogen bond pattern. Here we use a combination of IM-MS and IR spectroscopy to further investigate the influence of the initial, native conformation on the formation of these structures. Our results indicate that when intramolecular Coulomb-repulsion is large enough to overcome the stabilization energies of the genuine secondary structure, all proteins, regardless of their sequence or native conformation, form  $C_5$ -type hydrogen bond structures. Furthermore, our results suggest that in highly charged proteins the positioning of charges along the sequence is only marginally influenced by the basicity of individual residues.

**Keywords:** Ion mobility-mass spectrometry, IRMPD spectroscopy, Gas-phase proteins, Secondary structure

Received: 28 July 2016/Revised: 22 October 2016/Accepted: 1 November 2016/Published Online: 5 December 2016

## Introduction

In the gas-phase environment of a mass spectrometer, proteins can adopt a broad range of distinct structures. When electrosprayed from native, buffered solutions, especially large proteins and protein complexes are known to retain features of their condensed-phase conformation, a fact that found broad application in the field of native mass spectrometry (MS) [1–4]. Typically, the resulting ions exhibit rather compact shapes and acquire a relatively low number of charges during the ioniza-

tion process. On the other hand, under denaturing conditions, such as in the presence of organic solvents, proteins tend to adopt at least partially unfolded conformations in solution, which lead to broader distributions of higher charge states in the mass spectrum [5]. Ions on the lower end of this charge state distribution are often found to adopt multiple conformations that can even coexist in the gas phase [6, 7]. Also due to their tendency to dynamically interconvert, they are still structurally not fully understood. In contrast, ions of higher charge states adopt discrete, extended structures. These ions are characterized by a steady increase in their ion mobility (IM) collision cross-section (CCS) with growing number of charges, which is often described as a gradual unzipping of partially helical to extended and largely unfolded structures driven by intramolecular Coulomb repulsion of like charges [6–9]. Early gas-phase infrared multiple photon dissociation (IRMPD)

**Electronic supplementary material** The online version of this article (doi:10.1007/s13361-016-1551-5) contains supplementary material, which is available to authorized users.

Correspondence to: Kevin Pagel; e-mail: kevin.pagel@fu-berlin.de

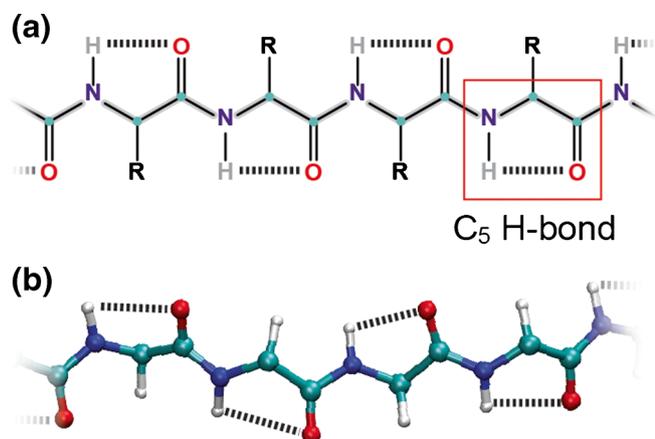
spectroscopy experiments showed that this transition from helix to extended is accompanied by the appearance of a new, intense absorption band at  $1480\text{ cm}^{-1}$ , the origin of which, especially when assuming an increasingly unordered structure, was not clear [10].

Using a combination of ultracold gas-phase IR spectroscopy and quantum chemical calculations, we were recently able to show that the emerging vibrational mode of proteins in high charge states is arising from newly formed structures that, in contrast to what was previously expected, are not at all unordered [11]. Instead, these extended structures are characterized by a regular and highly ordered network of intramolecular  $C_5$ -type hydrogen bonds between adjacent N–H and C=O groups (Figure 1). Using a combination of ion mobility-mass spectrometry (IM-MS) and gas-phase IRMPD spectroscopy [12], we here shed further light on the gas-phase unzipping of highly charged ions. In particular, we investigate the impact of the initial secondary structure and the location of basic amino acid residues within the molecule. Our results indicate that the Coulomb-driven unzipping to a defined secondary structure is a general feature of all highly charged protein ions.

## Experimental

### Synthesis

The peptides Ac-(KA<sub>2</sub>)<sub>7</sub>K-NH<sub>2</sub> (**1**) and Ac-A<sub>14</sub>K<sub>8</sub>-NH<sub>2</sub> (**2**) were synthesized on a 0.05 mmol scale using an automatic peptide synthesizer (Syro XP-1; Multi-SynTech GmbH, Witten, Germany) and a standard Fmoc/tBu strategy using a Fmoc-L-Lys(Boc)-rink amide resin. The final acetylation step was performed three times with a 5 mL mixture of DIPEA (*N,N*-diisopropylethylamine, 10%) and acetic anhydride (10%) in DMF (*N,N*-dimethylformamide). After cleavage from the resin, the peptides were precipitated and washed with diethyl ether and used for analysis without further purification.



**Figure 1.** (a) Chemical structure, and (b) model structure of a protein backbone featuring  $C_5$ -type hydrogen bonds. For clarity, side chains are abbreviated as 'R' in (a) and omitted in (b)

### Materials and Sample Preparation

Ubiquitin and all solvents were purchased from Sigma-Aldrich (Taufkirchen, Germany) and used without further purification. Human recombinant  $\alpha$ -synuclein was expressed in *E. coli* BL21(DE3) using the plasmid pT7-7 for  $\alpha$ -synuclein and purified according to published protocols [13].

For nano-electrospray ionization (nESI), samples were dissolved in H<sub>2</sub>O or MeOH/H<sub>2</sub>O v/v 50:50 and further diluted to the desired concentration of 10–20  $\mu\text{M}$  before they were loaded into in-house prepared Pd–Pt-coated borosilicate capillaries. Voltages of 0.8–1.2 kV were applied. Highly charged species were generated from a solution containing 1% of formic acid. To increase the number of charges even further, the supercharging reagent propylene carbonate (PC) was added to the solutions at a concentration of v/v 15% [9].

### Ion Mobility Spectrometry

The ion mobility method is well established [14, 15]. Experiments were performed on an in-house constructed drift-tube ion mobility-mass spectrometer (IM-MS) similar to one described previously [12, 16]. Briefly, ions are brought into the gas phase using nano-electrospray ionization (nESI) and are transferred into the vacuum of the instrument. Here, they are first collected and then accumulated in the trap region of an electrodynamic ion funnel. Packets of ions are then pulsed into the He-filled drift region ( $\sim 5$  mbar), which they traverse under the influence of a weak electric field ( $\sim 10\text{ V cm}^{-1}$ ). An activation region between ion trap and drift region, approximately 5 mm in length, allows to collisionally activate ions. Typical electric fields used for activation are between 100 and  $200\text{ V cm}^{-1}$ . A second electrodynamic ion funnel subsequent to the drift region collects and guides ions into a high-vacuum region where  $m/z$  selection occurs in a quadrupole mass filter. To determine the overall size of  $m/z$  selected species, the time-dependent ion current after release of the trap can be obtained as an arrival time distribution (ATD). The measured ion drift times can then be converted into angular-averaged collision cross-sections (CCS) as described previously [17, 18]. More details about drift time-CCS conversion are given in the first section of the [Supplementary Material](#).

### IR Spectroscopy

The utilized instrument allows for conformer and  $m/z$  selective IR spectroscopy based on an IRMPD detection scheme. For this, a narrow drift-time window (100  $\mu\text{s}$  in width) can be selected by electrostatic deflection prior to mass selection. Drift time and  $m/z$  selected ions then traverse a quadrupole guide and are irradiated by a 10  $\mu\text{s}$  pulse of IR photons with typical pulse energies between 10 and 40 mJ and a spectral width of 0.5%. The wavenumber-dependent photofragmentation is monitored by means of TOF MS. IR spectra are composed from this IRMPD data by computing the fragmentation yield for each wavelength step. For each species, two independent wavenumber scans, each composed of 20 averages per wavenumber

point, are presented within this work. A detailed description of the underlying procedures can be found in the first section of the [Supplementary Material](#). IR light was generated by the free electron laser facility housed in the Fritz Haber Institute [19].

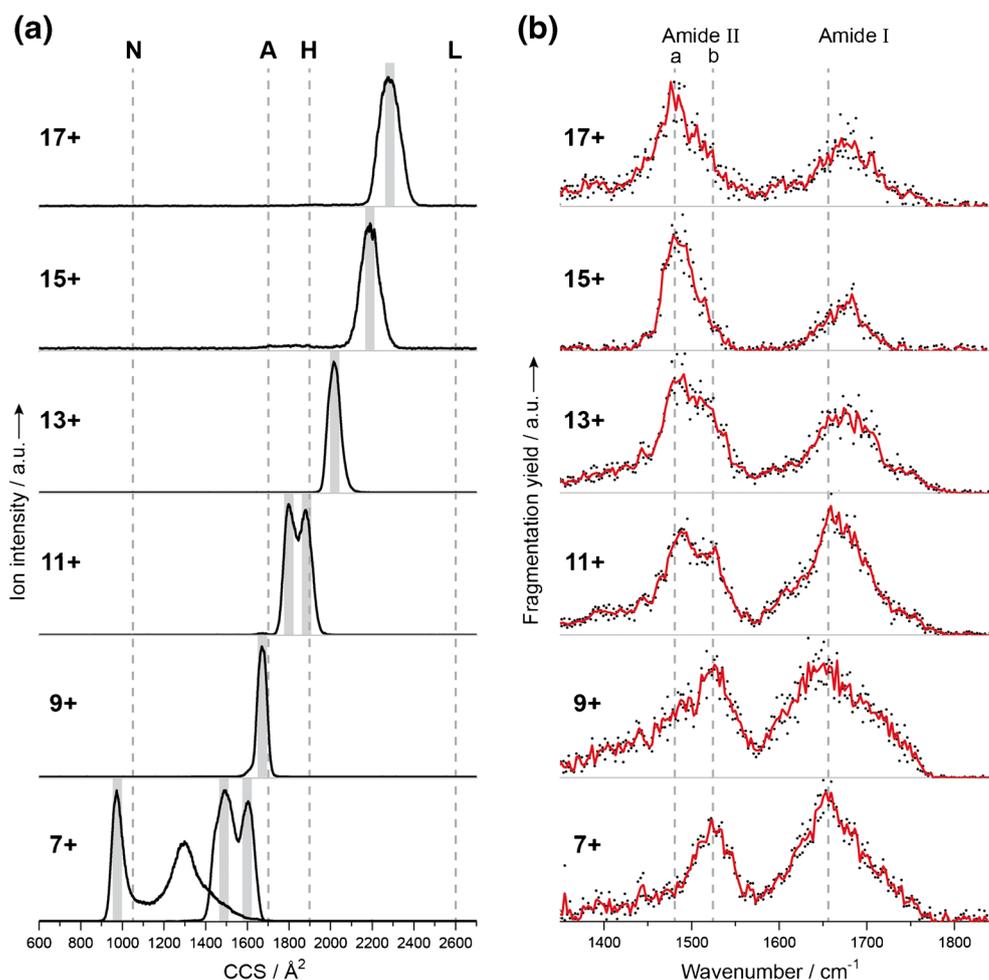
## Results and Discussion

### *Ubiquitin – From Compact to String*

The average charge of an electrosprayed protein ion primarily depends on the solvent conditions. Low charge states can be generated from buffered solutions, whereas higher charge states are typically produced from denaturing solvents. Here we investigate ubiquitin ions in a charge state range from 6+ to 17+. Ions with up to 13 charges (which equals the number of the apparent basic sites Lys, Arg, His, and the N-terminus) were observed from

denaturing solutions containing water/methanol and 1% of formic acid. In order to increase the number of charges on the gas-phase protein ions well above the number of basic sites in the sequence, the supercharging reagent propylene carbonate (PC) was used [9].

The different structures that ubiquitin ions can adopt in the gas phase have been investigated previously using ion mobility-mass spectrometry (IM-MS) and related techniques [7, 9, 20, 21]. Such experiments typically yield arrival time distributions (ATDs) from which the overall size of the ions by means of an angular-averaged collision cross-section (CCS) can be deduced. A subset of the ATDs acquired within this work is shown in Figure 2a, where the drift-time axis is converted to CCS (the full set is shown in Supplementary Figure S1, which can be found in the electronic version of this article). Compact ions with  $CCS < 1100 \text{ \AA}^2$  were observed for charge states 7+ and lower. A second ATD for 7+ ions is



**Figure 2.** (a) Collision cross-section (CCS) distributions of ubiquitin, which represent the conformer distribution of the investigated ions. Theoretical CCSs of the native structure (N), the partially unfolded A state (A), and that of hypothetical fully  $\alpha$ -helical (H) and fully linear (L) conformations are indicated as dashed, vertical lines for comparison [7]. For 7+ ions two traces are shown, resulting from measurements under gentle instrument conditions (compact ions) and under activating conditions (elongated ions), respectively. The population of ions subjected for further IRMPD analysis is indicated as grey bars. (b) Gas-phase IRMPD spectra of different charge states of ubiquitin. Different conformers of the same charge state yield virtually identical IR spectra (one spectrum is shown exemplarily for 7+ and 11+ ions). Data of two independent wavenumber scans is depicted as black dots and the average is presented as a solid red line

shown in Figure 2a, which was acquired while collisionally activating the ions with an activation voltage prior injection into the drift tube. As a result, the abundance of the partially folded ( $1100 \text{ \AA}^2 < \text{CCS} < 1600 \text{ \AA}^2$ ) and extended  $7+$  ions ( $\text{CCS} > 1600 \text{ \AA}^2$ ) [21] was increased significantly. With increasing charge state, the CCS of the unfolded ions increases gradually from  $1600 \text{ \AA}^2$  ( $8+$ ) to  $2280 \text{ \AA}^2$  ( $17+$ ). The theoretical CCS of the folded, native structure (N) and that of the partially unfolded and mostly helical *A* state (A) found in denaturing solutions are indicated as dashed, vertical lines in Figure 2a, together with the CCS values of two hypothetical structures, an extended, fully  $\alpha$ -helical (H) and a fully linear conformation (L) [7]. The gradual increase in size of ions with charge  $>8+$  is an indicator for an unzipping of the already elongated, presumably helical conformations to even more extended structures [9, 21]. In fact, earlier IM-MS studies showed that ions in even higher charge states than the ones investigated here slowly approach the size of the linear form L with a CCS of  $\sim 2600 \text{ \AA}^2$  [7, 9], the upper size limit for ubiquitin ions.

However, IM-MS methods only yield information about the overall size and shape of protein ions, while further structural details, such as the secondary structure, remain elusive. The stretching and bending vibrational modes of backbone C=O and N-H groups, on the other hand, are particularly sensitive to the secondary structure [22]. Therefore, we used infrared (IR) spectroscopy to further investigate the  $m/z$  and drift-time separated ubiquitin ions. For this, ions of a narrow drift-time window (indicated by grey shaded bars in Figure 2a) were irradiated with intense IR light from the free electron laser facility housed at the Fritz Haber Institute [19]. The wavelength-dependent IR multiple photon dissociation (IRMPD) of the protein ions was subsequently analyzed with the TOF mass spectrometer and IR vibrational spectra were finally obtained following the wavelength-dependent dissociation yield of the ions (for details see the [Experimental](#) section and [Supplementary Material](#)).

The gas-phase IR spectra obtained for ubiquitin are shown in Figure 2b. For each charge state, the average of two individual scans (black dots) is represented as a solid red line. In the investigated wavenumber range from  $1350$  to  $1850 \text{ cm}^{-1}$ , two spectral features centered around  $1650$  and  $1500 \text{ cm}^{-1}$  dominate the IR spectra of all investigated charge states. For a protein, the high-energy band can be assigned to C=O stretching modes (amide I) and the low-energy one usually stems from N-H bending vibrations (amide II). Interestingly, nearly identical IR spectra are obtained for different conformers of one and the same charge state such as the three different conformers of  $7+$  ions and the two different conformers of  $11+$  ions that were selected for irradiation as indicated in Figure 2a (for clarity reasons only one spectrum is shown; further details can be found in [Supplementary Figure S2](#)). This implies that the secondary structure in these conformations of the same charge states is rather similar, despite their considerable differences in overall size. A tentative assignment of the secondary structural content is furthermore possible based on the amide I band position;  $\beta$ -sheets are expected to yield strong amide I

signals at  $\sim 1633 \text{ cm}^{-1}$ , whereas a band at  $\sim 1650 \text{ cm}^{-1}$  typically indicates helical or disordered structures, respectively [22, 23]. For ions with  $\sim 11$  charges and less, such a comparison indicates a mostly helical/disordered structure. This agrees well with the predominantly helical structure of the *A* state (A) but does not support the  $\beta$ -sheet containing native solution structure (N), which is similar in size to the most compact ions found for ions with charges below  $7+$  [7, 24]. A recent study suggests that it is possible to transfer native structural elements into the gas phase under native, buffered solution conditions [25]. The here-investigated compact structures of ubiquitin, however, were generated from denaturing solvents containing methanol, which favors the partially unfolded *A* state in solution [26, 27]. These results suggest that a compact gas-phase structure alone does not guarantee a native secondary structure, but can only serve as an indicator for potentially native structural elements.

Similarly striking is the observation that the shape and/or position of the vibrational bands change considerably with increasing charge state of the ions. The amide I band gradually shifts to higher energies by approximately  $20 \text{ cm}^{-1}$  from  $\sim 1660 \text{ cm}^{-1}$  for ions with charge below  $9+$  to  $\sim 1680 \text{ cm}^{-1}$  for  $17+$  ions. This shift in energy is concomitant with a decrease in overall intensity. Absorption bands in the amide II region are affected even more drastically by an increase in charge. In accordance with IR spectra of condensed-phase proteins [22], a band at  $\sim 1520 \text{ cm}^{-1}$  can be observed for ions of charge states below  $9+$  (labeled amide IIb in Figure 2a). However, for ions with charge  $9+$  and higher, a new spectral feature at  $\sim 1480 \text{ cm}^{-1}$  starts to arise (labeled amide IIa). While the intensity of the amide IIb band decreases with increasing charge state, the amide IIa band represents the dominating feature of the IR spectra for ions with  $14$  and more charges.

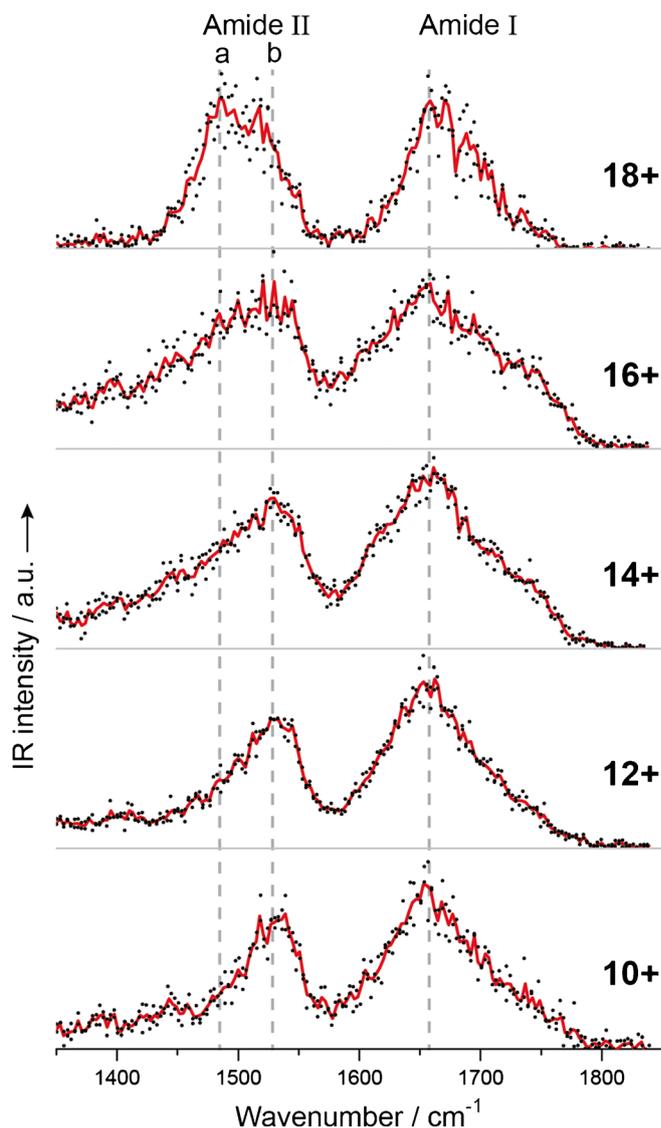
Similar trends were also observed in the gas-phase IR spectra of an earlier IRMPD study on cytochrome *c* [10], and in a more recent study on ubiquitin and cytochrome *c* ions [11]. In the latter, an elaborate experimental setup in which  $m/z$  selected ions are embedded in ultra-cold superfluid helium droplets followed by IR spectroscopic investigation was used. The ions are thereby cooled to  $0.37 \text{ K}$ , which can yield highly resolved IR spectra [28]. Together with quantum chemical calculations and careful consideration of Coulomb energies, it was concluded that the observed spectral evolution originates from an unzipping of helical structures to an extended  $C_5$ -type hydrogen bonding network in which weak hydrogen bonding interactions between adjacent C=O and N-H groups serve as stabilizing elements at high charge states [11]. As a consequence of these relatively weak hydrogen bonds, amide II absorptions are uniformly shifted to the red, giving rise to the observed amide IIa band, whereas amide I vibrations are slightly blue shifted. The particularly high intensities of the amide II band then result from the parallel orientation of the N-H bending transition dipole moments, which can, in contrast to the transition dipoles in a helical structure, add up constructively in this highly regular extended structure.

Compared to the experiments on ultra-cold ions in helium droplets, the here-presented IR spectra were obtained from conformer and  $m/z$  selected room-temperature ions using IRMPD fragments as messengers for photon absorption. The qualitatively similar results suggest that the unzipping of a more complex secondary structure to an extended, linear form is a general effect for all highly charged gas-phase protein ions and not a feature of the helium droplet method [11].

### *$\alpha$ -Synuclein – Does the Condensed-Phase Conformation Make a Difference?*

To test whether the native structure of a protein has an impact on its unzipping behavior at high charge states, we investigated different charge states of  $\alpha$ -synuclein, an intrinsically disordered 140-residue protein.  $\alpha$ -Synuclein is a well-studied protein, which does not adopt a defined conformation under physiological conditions [29]. However, upon interaction with lipid surfaces, it readily assembles into amyloid fibrils, which is the hallmark of Parkinson's disease [30–34]. The predominantly unordered structure of  $\alpha$ -synuclein in solution is also reflected in the gas phase [35]. Here, the CCS of  $\alpha$ -synuclein ions in charge states 10+ to 18+ increase gradually from  $\sim 2450 \text{ \AA}^2$  to  $\sim 2990 \text{ \AA}^2$  (see Supplementary Figure S3), which suggests a disentangling of unordered structure, rather than an unfolding of defined secondary and tertiary structural elements as in the case of ubiquitin. This notion is further supported by the fact that ions with multiple coexisting conformations that largely differ in CCS are generally not observed.

Despite the lack of a native structure, however,  $\alpha$ -synuclein shows similar unzipping behavior as observed for ubiquitin at increasing charge state. Here again, the structural transition is not apparent from the IM-MS data, but very well reflected in the IR spectra. Figure 3 shows a subset of the full data set, which can be found in Supplementary Figure S4. The amide I region changes only marginally over the charge state range from 10+ to 18+. It is centered around  $\sim 1655 \text{ cm}^{-1}$  and may be indicative for disordered and/or helical structure [22]. However, just as in the case of ubiquitin, the amide II range can be divided into the two sub-bands, amide IIa ( $\sim 1486 \text{ cm}^{-1}$ ) and amide IIb ( $\sim 1530 \text{ cm}^{-1}$ ). Also here the amide IIa band is found to be absent for low charge states and increases in intensity with increasing charge and intramolecular Coulomb repulsion. A complete disappearance of the amide IIb band at the highest investigated charge state was not observed in this case. In contrast to the experiments on ubiquitin, however,  $\alpha$ -synuclein was not investigated under supercharging conditions. Therefore, the maximum charge-per-residue ratio of the probed ions may not have been sufficiently large to induce the formation of a completely elongated, linear  $C_5$  arrangement of the peptide backbone. Nonetheless, the general trends in the evolution of the IR spectra of  $\alpha$ -synuclein as a function of charge state are clearly similar to those observed for ubiquitin and cytochrome *c* [10, 11].



**Figure 3.** Gas-phase IRMPD spectra of  $\alpha$ -synuclein ions. Data of two independent wavenumber scans is depicted as black dots and the average is presented as a solid red line. The full set of data including all investigated charge states can be found in Supplementary Figure S4

### *Model Peptides – How Important is the Location of Basic Sites?*

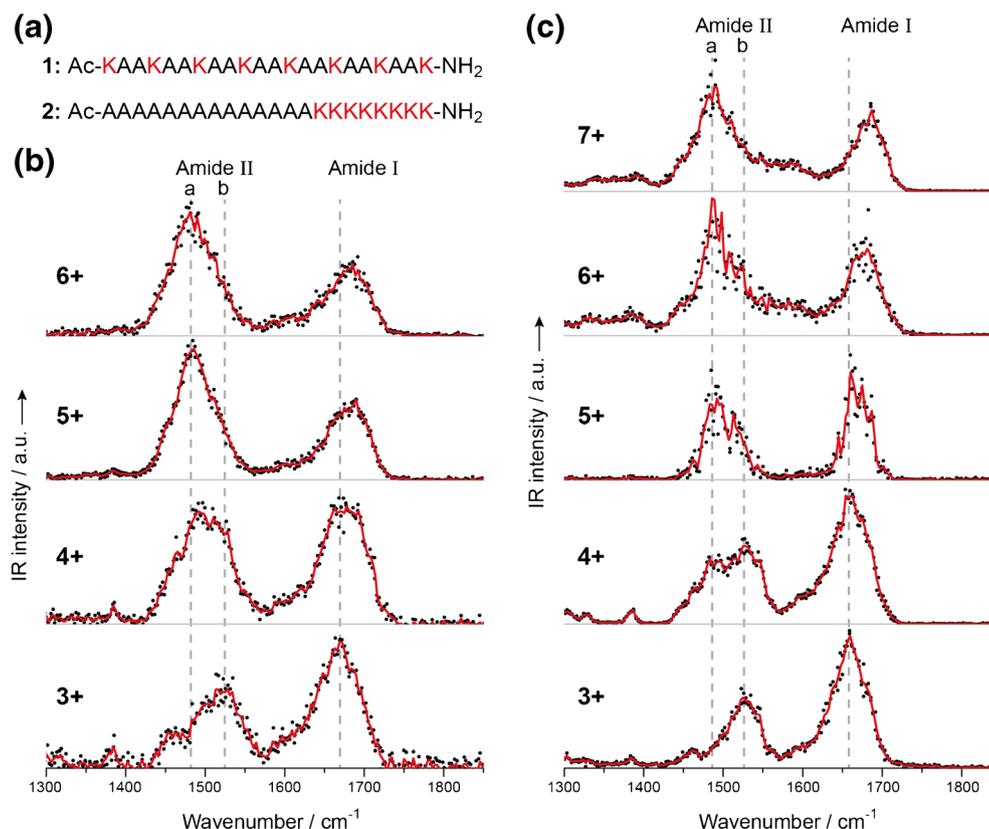
Unzipping of helices or other secondary structure elements will first occur at locations where the charge density leads to a Coulomb energy that is large enough to overcome the stabilization energy of the respective secondary structural motif. The positioning of the charges along the protein sequence, on the other hand, will be influenced by the gas-phase basicity of the constituting amino acid residues. While it may be straightforward to locate possible protonation sites on a protein in low charge states, the situation becomes less clear when the charge density increases and Coulomb repulsion starts to compete with the protonation energy of the most basic sites. To investigate the effect of the sequence position of basic amino acids (Lys,

Arg, His) on the location of charges on the gas-phase ion and its unzipping behavior, we synthesized two model peptides: Ac-(KA<sub>2</sub>)<sub>7</sub>K-NH<sub>2</sub> (**1**) and Ac-A<sub>14</sub>K<sub>8</sub>-NH<sub>2</sub> (**2**). Both consist of the same 22 amino acids, but differ in the order of the residues. While the eight basic Lys residues are distributed evenly in **1**, they are all located exclusively at the C-terminal end in **2**. The N- and C-termini were acetylated and amidated, respectively, to reduce the propensity for protonation at these sites.

Ions in charge states 3+ to 6+ are readily generated for **1** from aqueous solvents. For **2** only 3+ and 4+ ions were observed under the same conditions. Therefore, the supercharging reagent PC was used to generate charge states 5+ to 7+ of peptide **2**. The CCSs of these protonated model peptides increase from just below 480 Å<sup>2</sup> for 3+ ions to 635 Å<sup>2</sup> for **1** and 620 Å<sup>2</sup> for **2** when the charge increases to 6+. The CCS of the 7+ species of **2** was determined to 645 Å<sup>2</sup>. Ions of **2** were generally observed to be slightly smaller than those of **1** in the same charge state (data shown in the Supplementary Figure S5). Multiple, coexisting conformations were observed for neither of the peptides (i.e., all of their ATDs consist of single narrow features, data not shown). However, it is difficult to deduce structural details just based on the sequence and IM-MS data. Earlier work on polyalanine peptides suggests that the N-terminal Ala<sub>14</sub> domain of peptide **2** is likely to be of helical conformation [36, 37], but it is not clear in how far this affects the unzipping of the peptide ions.

The IR spectra of all investigated charge states of **1** and **2** in the amide I and II wavenumber range are shown in Figure 4b and c, respectively. At first glance, it becomes clear that the spectra follow a similar evolution with increasing charge state as observed for the larger proteins investigated here. Concomitant with a decrease in amide I intensity towards higher charge states is the appearance of an amide IIa band and the disappearance of the amide IIb band. The differences between the spectra of **1** and **2**, however, are more subtle. While the amide IIa and IIb bands for both peptides are observed at ~1485 and 1525 cm<sup>-1</sup>, respectively, the amide I band of the lowest charge state of **1** is found to be slightly blue shifted to 1670 cm<sup>-1</sup> with respect to the one of the lowest charge state of **2** at 1660 cm<sup>-1</sup>. The latter may be interpreted as an indicator for helical secondary structure of **2** as often reported for polyalanine peptides that carry C-terminal lysine residues [36, 37]. In the highest observed charge states, however, the amide I bands are observed at 1685 cm<sup>-1</sup> for both peptides.

What are the implications for the peptides' structures and the position of the charges on the peptide sequences? From the observed shifts in band position and relative intensity it is clear that independent of their initial structure, also these model peptides undergo a secondary structural transition to a completely elongated C<sub>5</sub> hydrogen bonded form when Coulomb repulsion is increased. This may be expected for peptide **1** where potentially protonated, basic residues are distributed



**Figure 4.** (a) Amino acid sequence of peptides **1** and **2**; (b) and (c) gas-phase IRMPD spectra of the model peptides **1** and **2** in different charge states. Charge states 5+ to 7+ of **2** were only observed from solvents containing the supercharging reagent PC. Data of two independent wavenumber scans is depicted as black dots and the average is presented as a solid red line

evenly along the sequence, giving rise to evenly distributed charges and a uniform unzipping process over the entire sequence. The observation of almost identical spectral features for **2** is, thus, a little more surprising; it directly suggests that the charges are not exclusively located on the Lys residues but must also be located on the backbone of the polyaniline segment. Otherwise, it could not be explained how the N-terminal, presumably helical rest can unzip to the completely elongated form that leads to the spectral features observed for the highest investigated charge states. This in turn suggests that the residues' individual  $pK_a$  values quickly become insignificant for the distribution of charges; instead, Coulomb repulsion takes over as the dominating factor that governs the charge distribution in highly charged ions.

We can illustrate this by calculating the charge distribution based on the Coulomb potential energies and protonation energies of a given amino acid sequence in two simplified model structures: a fully  $\alpha$ -helical conformation and a fully extended structure. The Coulomb potential energy  $E_C$  of  $N$  charges  $q_i$  separated by distances  $r_{ij}$  ( $i \neq j$ ;  $i, j \in \{1, \dots, N\}$ ) can be calculated as

$$E_C = \frac{k_e}{\epsilon_r} \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{r_{ij}}, \quad (1)$$

where  $k_e$  and  $\epsilon_r$  are Coulomb's constant and the relative permittivity, respectively. Protonation energies  $E_P$ , on the other hand, are estimated from the individual amino acid basicity [38]. At a given structure and protonation state  $z$ , the energy difference  $\Delta E_z = E_C - E_P$  between protonation energy and Coulomb energy must be at a minimum. A Monte Carlo simulated annealing scheme similar to the approach taken by Schnier et al. [38] was applied to calculate the charge distributions on the here-investigated amino acid sequences that lead to such an energy minimum. At equidistant spacing between amino acids, the lengths of the helical ( $H$ ) and the fully extended conformations ( $E$ ) are estimated as  $L_H = 1.5 n_A \text{ \AA}$  ( $n_A$ : number of amino acids) and  $L_E = 3.5 n_A \text{ \AA}$ , respectively. Initially, charges were distributed randomly over the sequence and annealing parameters were chosen such that the same lowest-energy charge distributions were reliably found for repeated calculations.

For the sequence of peptide **1**, the evenly distributed Lys residues primarily accommodate the protons and only higher charge states lead to protonation of the termini or the backbone of the Ala segments. For peptide **2**, we find that the charges in a doubly protonated ion are likely to reside at the first and the last position of the series of eight C-terminal Lys residues. However, independent of helical or extended structure of **2**, a third and a fourth proton will likely be located on the N-terminus (even though capped) and the N-terminal polyaniline chain, respectively. Even higher charge states will lead to protonation of more and more Ala residues. It can therefore be concluded that the actual proton affinity of a site is governed by the Coulomb energy, which exceeds the individual protonation energy of a residue at high charge densities. As a result, protons

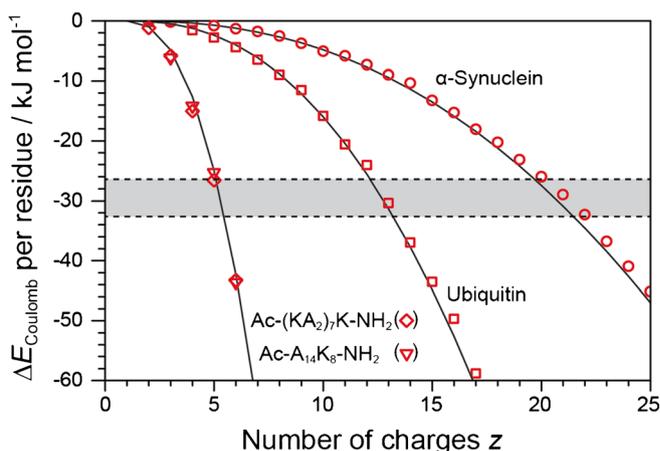
are evenly distributed over the peptide, independent of its amino acid sequence.

### Coulomb Repulsion versus Helix Stabilization Energy

From the calculated charge distributions, the difference in Coulomb energy between a helical (length  $L_H = 1.5 n_A \text{ \AA}$ ) and a fully extended, unzipped structure ( $L_E = 3.5 n_A \text{ \AA}$ ) can now be calculated. Peptides **1** and **2** as well as the well-known amino acid sequences of ubiquitin and  $\alpha$ -synuclein were considered, and the result was divided by the total number of amino acids in the sequence to yield the difference in Coulomb energy  $\Delta E_C$  per amino acid residue as a function of charge state. The results obtained for the four investigated molecules are shown as red symbols in Figure 5. In addition, we considered a much simpler charge distribution model in which the  $N$  equal charges are distributed equidistantly over the length  $L$  of the molecule. The Coulomb potential energy  $E_C$  (1) then becomes

$$E_C^{eq} = \frac{1}{L} \frac{k_e q^2 (N-1)}{\epsilon_r} \sum_{i=1}^N \sum_{j=i+1}^N \frac{1}{j-i}. \quad (2)$$

Using this simple model, the difference in Coulomb energy per residue between helical (length  $L_H$ ) and extended structure (length  $L_E$ ) was calculated and is depicted as black, solid lines in Figure 5. Generally, the obtained values are in good agreement with the result from the more realistic model where individual amino acid basicities are considered. It is important to note that the relative permittivity of vacuum ( $\epsilon_r = 1$ ) was considered for these calculations. The absolute value of  $\epsilon_r$  will depend on the local



**Figure 5.** Difference in Coulomb energy  $\Delta E_C$  per amino acid residue between helical ( $H$ ) and fully extended ( $E$ ) form of the investigated model peptides (22 residues) and proteins ubiquitin (76 residues) and  $\alpha$ -synuclein (140 residues), respectively. Red symbols denote values of  $\Delta E_C$  as calculated by the Monte Carlo simulated annealing approach and solid black lines represent the result of the model considering equidistant charges (according to Equation (2),  $\epsilon_r = 1$ )

environment of the charges, which varies between charges that are buried inside a protein ( $\epsilon_r = 2$ ) [39]) and the predominantly vacuum-exposed charges of the partially or completely unfolded species that are investigated here. An increase in the relative permittivity, however, will lead to reduced values of the intramolecular Coulomb potential energy  $E_C$  (1).

The stabilization energy of an  $\alpha$ -helix ( $H$ ) with respect to a fully extended form ( $E$ ) was previously calculated to be  $-26.4 \text{ kJ mol}^{-1}$  [40] and  $-32.6 \text{ kJ mol}^{-1}$  [41], using different methods of quantum chemical calculations, respectively. These values are indicated as dashed horizontal lines in Figure 5. The extended form of a peptide or protein is more stable than the helical conformation when  $\Delta E_C$  lies below this helix stabilization energy. According to this simple energy consideration, ubiquitin ions in charge states 13+ and above will exhibit a fully extended (i.e., fully unzipped) secondary structure. From the relative intensities of the amide IIa and IIb bands in the experimental IRMPD spectra in Figure 2a, one can tentatively deduce relative abundances of helical and extended structure, respectively, under the assumption of a linear relationship between relative intensities in the experimental spectra and oscillator strength as a first-order approximation. For ubiquitin it follows that the turning point to favor the extended structure occurs around charge state 11+ and the very low intensity in the IIb region indicates virtually no remaining helical structure for charge states above 13+. Similarly, an extended secondary structure of the model peptides is expected for charge states 5+ and above, which is also confirmed by the high relative intensities of the amide IIa band in the respective experimental spectra (Figure 4). According to this energy argument, however, the secondary structure of the protein  $\alpha$ -synuclein will only fully unzip with around 21 or more charges. Therefore, the highest investigated charge state 18+ is only partially extended, giving rise to the similar intensities of amide IIa and IIb bands observed in the experimental spectrum of 18+ ions in Figure 3. Hence, the here-considered simplified electrostatic model based on hypothetical fully helical and fully extended, unzipped structures provides a reasonable explanation for the observed transitions in the IR spectra. On the other hand, the CCSs of all peptide and protein ions in charge states above the turning point for unzipping of secondary structure are observed to further grow with increasing charge. This will be caused by a gradual straightening of possible kinks and bends in the protein backbone as well as by the gradual unwinding of remaining secondary structural elements as the charge increases.

## Conclusions

Ion mobility-mass spectrometry studies suggest an unfolding from compact structures to unfolded conformations when intramolecular Coulomb repulsion exceeds the attractive, structure-defining interactions in gas-phase peptide and protein ions [6, 7, 9, 20, 21]. Further increase in charge leads to a gradual increase in collision cross-section of already extended conformations, which was earlier recognized to be an indicator

for an unzipping of remaining secondary structural elements [6, 9]. In a previous study [11], such a transition in secondary structure of ubiquitin and cytochrome *c* ions was directly observed using a combination of ultracold IR spectroscopy and quantum chemical calculations. It was attributed to an unzipping of helical structures to a completely linear form, which is supported by a  $C_5$  hydrogen bond pattern when Coulomb repulsion exceeds the helix stabilization energy. Using ion-mobility and  $m/z$ -selective IRMPD spectroscopy, we here directly follow the transition in secondary structure as a function of charge state and size for the natively folded protein ubiquitin and the intrinsically disordered protein  $\alpha$ -synuclein. In terms of energy, a simple electrostatic model [11] was shown to provide an explanation for the observed transition.

In addition, two model peptides were designed to probe the influence of the amino acid sequence on the distribution of charges along the peptide chain at high charge states. The experimental results and a Monte Carlo simulated annealing scheme that calculates charge distributions based on Coulomb energy and protonation energy of individual residues provide evidence that the position of basic sites only has a marginal influence on the position of charges at high charge states. Taken together, the here-presented results therefore support the notion that the effect of Coulomb-driven unzipping to a defined secondary structure is general for all highly charged protein ions, independent of their initial structure and amino acid sequence. As such, this highly regular conformation with its characteristic  $C_5$ -type hydrogen bonding pattern can be seen as a universal secondary structure motif that all gas-phase proteins adopt at high charge states.

## Acknowledgments

The authors acknowledge the expert assistance of the FHI free electron laser facility staff, in particular by S. Gewinner and W. Schöllkopf. Beate Koksich is gratefully acknowledged for providing peptide synthesis facilities.

## References

1. Benesch, J.L., Ruotolo, B.T., Simmons, D.A., Robinson, C.V.: Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chem. Rev.* **107**, 3544–3567 (2007)
2. Heck, A.J.R.: Native mass spectrometry: a bridge between interactomics and structural biology. *Nat. Methods* **5**, 927–933 (2008)
3. Konermann, L., Pan, J., Liu, Y.H.: Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **40**, 1224–1234 (2011)
4. Chait, B.T., Cadene, M., Olinares, P.D., Rout, M.P., Shi, Y.: Revealing higher order protein structure using mass spectrometry. *J. Am. Soc. Mass Spectrom.* **27**, 952–965 (2016)
5. Hall, Z., Robinson, C.V.: Do charge state signatures guarantee protein conformations? *J. Am. Soc. Mass Spectrom.* **23**, 1161–1168 (2012)
6. Shelimov, K.B., Clemmer, D.E., Hudgins, R.R., Jarrold, M.F.: Protein structure in vacuo: gas-phase conformations of BPTI and cytochrome *c*. *J. Am. Chem. Soc.* **119**, 2240–2248 (1997)
7. Wyttenbach, T., Bowers, M.T.: Structural stability from solution to the gas phase: native solution structure of ubiquitin survives analysis in a

- solvent-free ion mobility-mass spectrometry environment. *J. Phys. Chem. B* **115**, 12266–12275 (2011)
8. Segev, E., Wyttenbach, T., Bowers, M.T., Gerber, R.B.: Conformational evolution of ubiquitin ions in electrospray mass spectrometry: molecular dynamics simulations at gradually increasing temperatures. *Phys. Chem. Chem. Phys.* **10**, 3077–3082 (2008)
  9. Going, C.C., Williams, E.R.: Supercharging with m-nitrobenzyl alcohol and propylene carbonate: forming highly charged ions with extended, near-linear conformations. *Anal. Chem.* **87**, 3973–3980 (2015)
  10. Oomens, J., Polfer, N., Moore, D.T., van der Meer, L., Marshall, A.G., Eyler, J.R., Meijer, G., von Helden, G.: Charge-state resolved mid-infrared spectroscopy of a gas-phase protein. *Phys. Chem. Chem. Phys.* **7**, 1345 (2005)
  11. González Flórez, A.I., Mucha, E., Ahn, D.S., Gewinner, S., Schöllkopf, W., Pagel, K., von Helden, G.: Charge-induced unzipping of isolated proteins to a defined secondary structure. *Angew. Chem. Int. Ed.* **55**, 3295–3299 (2016)
  12. Warnke, S., Seo, J., Boschmans, J., Sobott, F., Scrivens, J.H., Bleiholder, C., Bowers, M.T., Gewinner, S., Schöllkopf, W., Pagel, K., von Helden, G.: Protomers of benzocaine: solvent and permittivity dependence. *J. Am. Chem. Soc.* **137**, 4236–4242 (2015)
  13. Hoyer, W., Antony, T., Cherny, D., Heim, G., Jovin, T.M., Subramaniam, V.: Dependence of  $\alpha$ -synuclein aggregate morphology on solution conditions. *J. Mol. Biol.* **322**, 383–393 (2002)
  14. Bowers, M.T.: Ion mobility spectrometry: a personal view of its development at UCSB. *Int. J. Mass Spectrom.* **370**, 75–95 (2014)
  15. Wyttenbach, T., Pierson, N.A., Clemmer, D.E., Bowers, M.T.: Ion mobility analysis of molecular dynamics. *Annu. Rev. Phys. Chem.* **65**, 175–196 (2014)
  16. Kemper, P.R., Dupuis, N.F., Bowers, M.T.: A new, higher resolution ion mobility mass spectrometer. *Int. J. Mass Spectrom.* **287**, 46–57 (2009)
  17. Revercomb, H.E., Mason, E.A.: Theory of plasma chromatography/gaseous electrophoresis: review. *Anal. Chem.* **47**, 970–983 (1975)
  18. Mason, E.A., McDaniel, E.W.: Transport properties of ions in gases, pp. 145–159. Wiley, New York (1988)
  19. Schöllkopf, W., Erlebach, W., Gewinner, S., Heyne, G., Junkes, H., Liedke, A., Meijer, G., Platschkowski, V., von Helden, G., Bluem, H., Dowell, D., Lange, R., Rathke, J., Todd, A.M.M., Young, L.M., Lehnert, U., Michel, P., Seidel, W., Wunsch, R., Gottschalk, S.C.: The new IR FEL facility at the Fritz-Haber-Institut in Berlin. Proceedings of FEL 2014, Basel, Switzerland, August 25–29, (2014) WEB04
  20. Koeniger, S.L., Merenbloom, S.I., Sevugarajan, S., Clemmer, D.E.: Transfer of structural elements from compact to extended states in unsolvated ubiquitin. *J. Am. Chem. Soc.* **128**, 11713–11719 (2006)
  21. Valentine, S.J., Counterman, A.E., Clemmer, D.E.: Conformer-dependent proton-transfer reactions of ubiquitin ions. *J. Am. Soc. Mass Spectrom.* **8**, 954–961 (1997)
  22. Barth, A.: Infrared spectroscopy of proteins. *Biochim. Biophys. Acta* **1767**, 1073–1101 (2007)
  23. Jackson, M., Mantsch, H.H.: The use and misuse of FTIR spectroscopy in the determination of protein structure. *Crit. Rev. Biochem. Mol. Biol.* **30**, 95–120 (1995)
  24. Shi, H., Atlasevich, N., Merenbloom, S.I., Clemmer, D.E.: Solution dependence of the collisional activation of ubiquitin  $[M + 7H]^+$  ions. *J. Am. Soc. Mass Spectrom.* **25**, 2000–2008 (2014)
  25. Seo, J., Hoffmann, W., Warnke, S., Bowers, M.T., Pagel, K., von Helden, G.: Retention of native protein structures in the absence of solvent: a coupled ion mobility and spectroscopic study. *Angew. Chem. Int. Ed.* **55**, 14173–14176 (2016)
  26. Pan, Y., Briggs, M.S.: Hydrogen exchange in native and alcohol forms of ubiquitin. *Biochemistry* **31**, 11405–11412 (1992)
  27. Brutscher, B., Bruschweiler, R., Ernst, R.R.: Backbone dynamics and structural characterization of the partially folded A state of ubiquitin by  $^1H$ ,  $^{13}C$ , and  $^{15}N$  nuclear magnetic resonance spectroscopy. *Biochemistry* **36**, 13043–13053 (1997)
  28. González Flórez, A.I., Ahn, D.S., Gewinner, S., Schöllkopf, W., von Helden, G.: IR spectroscopy of protonated leu-enkephalin and its 18-crown-6 complex embedded in helium droplets. *Phys. Chem. Chem. Phys.* **17**, 21902–21911 (2015)
  29. Uversky, V.N.: What does it mean to be natively unfolded? *Eur. J. Biochem.* **269**, 2–12 (2002)
  30. Weinreb, P.H., Zhen, W., Poon, A.W., Conway, K.A., Lansbury Jr., P.T.: NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* **35**, 13709–13715 (1996)
  31. Eliezer, D., Kutluay, E., Bussell Jr., R., Browne, G.: Conformational properties of alpha-synuclein in its free and lipid-associated states. *J. Mol. Biol.* **307**, 1061–1073 (2001)
  32. Uversky, V.N., Li, J., Souillac, P., Millett, I.S., Doniach, S., Jakes, R., Goedert, M., Fink, A.L.: Biophysical properties of the synucleins and their propensities to fibrillate: inhibition of alpha-synuclein assembly by beta- and gamma-synucleins. *J. Biol. Chem.* **277**, 11970–11978 (2002)
  33. Morar, A.S., Olteanu, A., Young, G.B., Pielak, G.J.: Solvent-induced collapse of alpha-synuclein and acid-denatured cytochrome *c*. *Protein Sci.* **10**, 2195–2199 (2001)
  34. Galvagnion, C., Buell, A.K., Meisl, G., Michaels, T.C., Vendruscolo, M., Knowles, T.P., Dobson, C.M.: Lipid vesicles trigger alpha-synuclein aggregation by stimulating primary nucleation. *Nat. Chem. Biol.* **11**, 229–234 (2015)
  35. Phillips, A.S., Gomes, A.F., Kalapothakis, J.M., Gillam, J.E., Gasparavicius, J., Gozzo, F.C., Kunath, T., MacPhee, C., Barran, P.E.: Conformational dynamics of alpha-synuclein: insights from mass spectrometry. *Analyst* **140**, 3070–3081 (2015)
  36. Rossi, M., Blum, V., Kupser, P., von Helden, G., Bierau, F., Pagel, K., Meijer, G., Scheffler, M.: Secondary structure of Ac-Ala<sub>n</sub>-LysH<sup>+</sup> polyalanine peptides (n = 5, 10, 15) in vacuo: helical or not? *J. Phys. Chem. Lett.* **1**, 3465–3470 (2010)
  37. Hudgins, R.R., Ratner, M.A., Jarrold, M.F.: Design of helices that are stable in vacuo. *J. Am. Chem. Soc.* **120**, 12974–12975 (1998)
  38. Schnier, P.D., Gross, D.S., Williams, E.R.: On the maximum charge state and proton transfer reactivity of peptide and protein ions formed by electrospray ionization. *J. Am. Soc. Mass Spectrom.* **6**, 1086–1097 (1995)
  39. Schnier, P.D., Gross, D.S., Williams, E.R.: Electrostatic forces and dielectric polarizability of multiply protonated gas-phase cytochrome *c* ions probed by ion/molecule chemistry. *J. Am. Chem. Soc.* **117**, 6747–6757 (1995)
  40. Tkatchenko, A., Rossi, M., Blum, V., Ireta, J., Scheffler, M.: Unraveling the stability of polypeptide helices: critical role of van der Waals interactions. *Phys. Rev. Lett.* **106**, 118102 (2011)
  41. Hua, S., Xu, L., Li, W., Li, S.: Cooperativity in long alpha- and 3(10)-helical polyalanines: both electrostatic and van der Waals interactions are essential. *J. Phys. Chem. B* **115**, 11462–11469 (2011)