# Analyst

Accepted Manuscript






ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/analyst

# Effects of Charge States, Charge Sites and Side Chain Interactions on Conformational Preferences of a Series of Model Peptide Ions

Chunying Xiao, Lisa M. Pérez, and David H. Russell*

Texas A&M University, Department of Chemistry
College Station, Texas 77843
*russell@chem.tamu.edu

**Abstract:** The effects of charge states, charge sites and side chain interactions on conformational preferences of gas-phase peptide ions are examined by ion mobility-mass spectrometry (IM-MS) and molecular dynamics (MD) simulations. Collision cross sections (CCS) of $[M + 2H]^{2+}$ and $[M + 3H]^{3+}$ ions for a series of model peptides, *viz.* Ac–$(AAKAA)_nY$–$NH_2$ ($AK_n$, n = 3-5) and Ac–$Y(AEAAKA)_nF$–$NH_2$ ($AEK_n$, n = 2-5) are measured by using IM–MS and compared with calculated CCS for candidate ions generated by MD simulations. The results show that charge states, charge sites and intramolecular charge solvation are important determinants of conformer preference for $AK_n$ and $AEK_n$ ions. For $AK_n$ ions, there is a strong preference for helical conformations near the N-terminus and charge-solvated conformations near the C-terminus. For $[AEK_n + 2H]^{2+}$ ions, conformer preferences appear to be driven by charge solvation, whereas $[AEK_n + 3H]^{3+}$ ions favor more extended coil-type conformations.

## Introduction

Peptides and proteins are highly dynamic, sampling many conformations on rapid time scales, and both the dynamics and conformational preferences are sensitive to the local environment[1-4]. Conformational preferences and the dynamics of interconversions among the different conformations are highly dependent on intramolecular interactions, *i.e.,* van der Waals interactions, hydrogen bonding, and electrostatic interactions. The presence of water as well as cations and anions also plays a key role in determining conformational preferences, both in terms of the hydrophobic effects as well as hydrophilic interactions, *viz.* solvent-accessibility to hydrophilic side chains of histidine, lysine, arginine, and interactions (salt-bridges) of these charge sites with oppositely charged aspartate and glutamate ($R\text{-}COO^-$) as well as asparagine and glutamine[5, 6]. The effects of intramolecular hydrophilic interactions on the conformational preferences of peptides/proteins have been extensively studied[7-9] and advances in experimental techniques and molecular dynamics (MD) simulations are providing new approaches to studies of both thermodynamic stability and folding/unfolding kinetics of native and non-native conformations[8, 10]. Although MD simulations are widely used for studies of peptide/protein conformational preferences, simulations only provide candidate conformations that must be evaluated against experimental data, *i.e.,* from circular dichroism (CD)[11], fluorescence and fluorescence resonance energy transfer (FRET)[12, 13], nuclear magnetic resonance (NMR)[14, 15], X-ray diffraction (XRD)[14], and small-angle X-ray scattering (SAXS)[16-18]. There is increasing awareness that mass spectrometry (MS) approaches add new dimensions for understanding peptide/protein structure and extending our understanding of structure/function relationships, especially for studies of systems that are composed of multiple conformations.

MS-based approaches for studies of peptide/protein conformations have evolved to include H/D exchange[19, 20], chemical labeling[21, 22], IR-UV double resonance spectroscopy[23, 24] and tandem MS[25, 26]. IM-MS combined with MD simulation complements the more traditional approaches mentioned above[27] since it provides a direct determination of the conformational heterogeneity of the ion population, and it can be adapted to high throughput workflows for screening complex peptide/protein libraries[25, 28]. Electrospray ionization (ESI) is clearly the ionization method of choice for structural MS since it has been shown to maintain noncovalent interactions[29-32], and even to retain solution phase secondary structure[33]. Recently, Breuker and coworkers studied protein structures in the solution and gas phase, and demonstrated that salt bridges can help to stabilize native/compact structure on short time scales when the proteins are transferred from solution phase to the gas phase[34, 35]. A significant advantage of ESI is the ability to produce mass spectra dominated by multiply-charged ions; however, it is difficult to determine the location of the charge for molecules that contain multiple possible charge sites.

IM-MS is based on the measurement of ion-neutral CCS, which reflects the size and three-dimensional shape of an ion, and provides a direct measurement on population heterogeneity, *viz.* the number of peaks in the IM arrival-time distributions (ATD). *A priori* assignment of IM CCSs is a daunting task, *viz.,* currently there exist no guiding principles for correlating the measured CCS for an ion to specific 3-D shapes. In some cases it is possible to correlate CCS with 3-D shapes derived from other experimental measurements, specifically XRD or NMR[36-39]; however, MD simulations provide a broader sampling of candidate conformations of the ions whose CCS values can be calculated by using MOBCAL[40] (or similar methods)[36, 41]. The "theoretical" CCS obtained using MD simulations and MOBCAL can then be compared to the experimental CCS.

This study employs two series of peptides to illustrate the data mining made possible by integration of IM-MS and MD simulation. The peptide sequences are Ac-$(AAKAA)_n$Y-$NH_2$ (**$AK_n$** series, n = 3–5) and Ac-Y$(AEAAKA)_n$F-$NH_2$ (**$AEK_n$** series, n = 2–5); Ac- indicates acetylation at the N-termini and -$NH_2$ indicates amidation at the C-termini. The N-termini and C-termini of the peptides in each series were protected to reduce the likelihood of collapse of the helix by having charge(s) located at the termini[42]. The model peptides used in this study have been used previously by several groups to probe helical propensities of alanine-containing peptides in solution[43, 44] and in low dielectric gas-phase (solvent-free) environments[45]. Previous CD studies show that in solution, the helical content of both $AK_n$ and $AEK_n$ increases as n increases[43, 44, 46], and the helical content of $AEK_n$ is higher than that of $AK_n$ for peptides of comparable length owing to the stabilization afforded by salt bridges. In prior IM-MS studies of the singly-charged ions for the two peptides, it was found that gas-phase ions of $AK_n$ (n = 3-6) have higher helical content (ca. 60% helical) than $AEK_n$, and helical propensity is the highest when charge is aligned with the helix macrodipole, *i.e.*, the charge is located on or near the C-termini[45]. Despite extensive studies of the $AK_n$ and $AEK_n$ series, the following questions still remain unanswered: (i) how do charge sites and charge states affect conformer preferences; (ii) how do intramolecular interactions involving the polar side chains affect conformer preferences; and (iii) is there any relationship between solution-phase and gas-phase conformations? Here, ESI–IM–MS in conjunction with MD simulations was used to evaluate the conformational preferences of the two model peptide series in an effort to address these questions.

## Experimental Section

**Sample Preparation**. Model peptides $AK_n$ (n = 3 - 5) and $AEK_n$ (n = 2 - 5) were purchased from Shanghai Mocell Biotech Co., Ltd and used without further purification. Solutions of the model

peptides (2 μM) were prepared using 10:90 (v / v) methanol / buffer (buffer: 50mM ammonium acetate, 25mM imidazole; pH was adjusted to 7 with 0.1 M acetic acid).

**ESI–IM–MS**. All mobility profiles were acquired on a Waters Synapt HDMS G2 mass spectrometer (Manchester, U.K.). Ions were produced by nano-ESI with a source temperature of 120 °C and capillary voltage of 1.5 kV, sampling cone voltage of 40 V and extraction cone voltage of 4 V. For ion mobility experiments, the traveling wave ion mobility cell was maintained at 3.01 mbar $N_2$. The travelling wave velocity is 500 m/s and wave height is 20 V for $[AK_n + 3H]^{3+}$; wave velocity and wave height are 300 m/s and 20 V respectively for both $[AK_n + 2H]^{2+}$ and $[AEK_n + 3H]^{3+}$; wave velocity and wave height are 800 m/s and 40 V respectively for $[AEK_n + 2H]^{2+}$ due to the difference in mobility of different peptides.

CCS calibration was performed as previously described by Ruotolo[47] *et al*. Tryptically digested peptides of cytochrome c and myoglobin were used as calibration standards and literature CCS values of doubly-charged peptide ions are cited from the Clemmer group's CCS database[48]. It should be noted that the triply-protonated model peptides are also calibrated using doubly-charged ions because currently there is no standard calibrant database for triply-protonated peptide ions.

To probe the stability of the model peptides, solvents conditions were screened to ensure CCS profiles are not influenced by the solvents (data not shown). Instrumental conditions such as capillary voltage, sampling cone voltage and extraction cone voltage, were also screened to ensure arrival time distribution (ATD) did not change under various conditions.

**Calculation of helical content and number of helical residues**

$$Helical\ content\ (\%) = \left(\frac{\Omega_{obs} - \Omega_{glob}}{\Omega_{helix} - \Omega_{glob}}\right) \times 100 \quad \text{Equation 1}$$

$$number\ of\ helical\ residues = Helical\ content \times total\ residue\ number \quad \text{Equation 2}$$

$\Omega_{obs}$ is the CCS determined by the centroid of the peak profiles. $\Omega_{helix}$ is the calculated CCS for a rigid α-helix cited from literature[45]. $\Omega_{glob}$ is the calculated globular CCS using the standard globular trend line equations:

$$y = 0.1364x + 105.99 \quad \text{For +2 species} \quad \text{Equation 3}$$

$$y = 0.1272x + 134 \quad \text{For +3 species} \quad \text{Equation 4}$$

Here, x is the mass of peptides, and y is the CCS of globular structure.

The standard trend line for doubly-protonated peptide ions (grey in **Figures 1B and 3B,** upper) was drawn based on the literature data of doubly-charged ions[48]. The standard trend line for triply-protonated ions (grey in **Figures 1B and 3B,** lower) was drawn based on standard triply-protonated peptide ions prepared from the miscleavage of trypsin digestion of proteins (ubiquitin, lysozyme (from chicken egg white), aldolase, enolase, avidin (from egg white), albumin (from chicken egg), ribonuclease (bovin pancrease), myoglobin (horse), cytocrome c, BSA (horse)). All proteins were obtained from Sigma-Aldrich and used without further purification.

**Molecular Dynamics (MD) Simulations**. Candidate conformations of $AK_n$ and $AEK_n$ ions were generated by simulated annealing (SA) using AMBER11[49] with the FF99SB force field[50]. For the $AK_n$ and $AEK_n$ peptide series, both the N-termini and C-termini are protected (acetylated and amidated, respectively), thus the lysine side chains, which have high proton affinities, were considered as the charge carrying sites. In cases where the number of lysines exceeded the numbers of protons, all possible proton positions were considered. The α-helical and fully extended conformations were used as starting structures for SA. During one annealing cycle, the

system was heated from 300 K to 1000 K and then cooled down to 300 K over 8400 fs with a time step of 1 fs. After each annealing cycle, the structure energy was minimized. SA was performed for 300 annealing cycles producing 300 minimized structures per simulation. To ensure proper sampling of the conformational space, a second set of simulations was performed starting with the lowest energy structures from the first set of simulations. For $[AEK_4 + 2H]^{2+}$, two sets of simulations did not produce enough sampling, *i.e.*, the candidate structures did not vary much, therefore, a third set of simulations was performed. The highest energy conformation and a randomly selected middle energy conformation with CCSs in the range of $\pm$2% of the experimental value from the first and second set of simulations were chosen as the starting structures of the third set of simulations. CCSs were calculated by the trajectory method in MOBCAL[40]. The secondary structure was calculated using DSSP program[51, 52].

When processing the simulation data, filtering by CCS and RMSD cluster analysis was performed on the candidate conformations. The procedure for making clusters is described in **Figure S1 (**See **SI)**. First, all generated conformations that have CCSs within $\pm$ 2% of the experimental CCS value for the doubly-charged ions form family 1 (in the case of triply-charged ions, $\pm$ 5% of the experimental data form family 1 because the triply-charged ions were calibrated by doubly-charged standard ions which may result in higher error). Structures in family 1 were then clustered based on backbone RMSD (Root-mean-square deviation). Cluster 1 consists of all conformations in family 1 that have an RMSD of the backbone atoms less than a specified value (RMSD cutoff is typically 4-4.5 $Å^2$) when compared to the lowest energy structure in family 1. All of the conformations with an RMSD higher than the specified value make family 2. Cluster 2 contains all structures with an RMSD of the backbone atoms below the cutoff when compared to the lowest energy structure in family 2 and all the conformations with

an RMSD higher than the cutoff make family 3. The procedure was repeated for the remaining structures until all the structures were allocated into a cluster. Therefore, the reference structure for cluster 1 has the lowest energy, the reference structure for cluster 2 has higher energy than that for cluster 1 and the reference structure for the last cluster has the highest energy. It should be noted that due to the different size of peptides, the RMSD cutoff was adjusted to produce a reasonable number of clusters (< 60).

**Results and Discussion**

Biological environments are highly diverse, ranging from solution to lipid membranes, with the latter more closely resembling the gas phase than solution. That is, the dielectric environment of the cell membranes (ε = 2) is similar to that of a vacuum (ε = 1), but it differs greatly from that of an aqueous solution (ε = 80)[53]. Therefore, studies of gas-phase ions offer new approaches for understanding peptide/protein conformer preferences in membrane-like, low dielectric environments different from those in solution[54, 55]. Although gas-phase ions of $[AK_n + H]^+$ and $[AEK_n + H]^+$ have been studied previously[45], their counterparts carrying multiple charges in the gas phase have not been examined using ion mobility.

**1. Charge states and polar side chain interactions are important to conformational preference of peptides.**

**$AK_n$ series. Figure 1A** shows the CCS profiles for $[AK_n + 2H]^{2+}$ (solid line) and $[AK_n + 3H]^{3+}$ ions (dashed line). The CCS profiles for both $[AK_n + 2H]^{2+}$ and $[AK_n + 3H]^{3+}$ ions are narrow and symmetrical, indicative of an ion population that is composed of similar conformers. The only exception is the CCS profile for the $[AK_3 + 2H]^{2+}$ ion, which contains two peaks (denoted A and B). It should be noted that the CCS difference between $[AK_n + 3H]^{3+}$ and $[AK_n + 2H]^{2+}$

decreases from $AK_3$ to $AK_5$ and the CCS for $[AK_5 + 3H]^{3+}$ is even smaller than that for $[AK_5 + 2H]^{2+}$.

Previous studies show that in the gas phase, peptide ions with globular conformations are distributed along a trend line in which the structures lying above the trend line are helical or extended coils[56]. For comparison, the CCS values for $AK_n$ ions are plotted as a function of molecular weight (**Figure 1B)**, and the trend line for the random coil peptide ions is drawn in grey.[48] Note that in both cases, the data points are offset and have different slopes from the expected values of globular peptide ion conformations.

**Figure 1C** shows the calculated helical content (calculated using **Equation 1**, see **Experimental Section**) for the gas-phase $AK_n$ ions (GP, shown in blue), including values for $[AK_n + H]^+$ ions reported previously[45], and the reported values for the solution-phase species (SP, shown in black)[44, 46]. The data clearly show that the helical content for $[AK_n + H]^+$ does not change dramatically and the helical content for $[AK_n + 2H]^{2+}$ ions increases (n = 3 to 5). Yet The helical content for $[AK_n + 3H]^{3+}$ decreases and then increases with n = 3 to 5. By comparison with the helical content in the solution phase (SP, black), the helical content for $[AK_n + 2H]^{2+}$ ($GP^{2+}$) ions is parallel to the trend observed for solution phase ions.

To interpret the experimental data in **Figure 1**, MD simulations (See **Experimental Section: Molecular Dynamics (MD) Simulations**) were used to investigate the following: (i) the effect of charge states and charge sites on conformer preferences, (ii) the effect of intramolecular interactions (charge solvation) on conformer preferences, and (iii) whether extended conformers more closely resembled helices or extended-coils. Representative ribbon structures, which illustrate how polar side chains affect the conformation of peptide ions, of the most populated

clusters (see **Experimental Section** for details) are provided in **Figure 2 (**For more information, see **Figure S2** in **SI**).

**Figure 2A** shows the representative structures for $[AK_n + 2H]^{2+}$ and the K-K interactions are marked with red circles. It should be noted that the K-K interactions include both $K^+$-$K^0$ and $K^0$-$K^0$ interactions ($K^+$ indicates protonated lysines and $K^0$ indicates neutral lysines**.** The charged lysines are marked by "CP" in **Figure 2**. The lysines that are not mentioned in "CP" are neutral). The representative structures for $[AK_n + 2H]^{2+}$ show that (i, i+5) K-K side chain interaction only exists in the representative conformations of $[AK_3 + 2H]^{2+}$ (Peak B) and $[AK_4 + 2H]^{2+}$ and the K-K interaction at the position (i, i+5) inhibits formation of helical conformations. Most of the charged lysines primarily interact with nearby amide linkages to form a charge-solvated structure. Therefore, the conformation preference for $[AK_n + 2H]^{2+}$ is largely related to charge solvation effects. When charge solvation effect is dominant, peptides tend to form a random coil type conformation, *e.g.* $[AK_3 + 2H]^{2+}$. When backbone length increases, charge solvation does not dominate peptide conformation; with high helical propensity of alanine, the representative conformations of $[AK_4 + 2H]^{2+}$ and $[AK_5 + 2H]^{2+}$ ions have high helical propensity.

The representative structures of the most populated clusters for $[AK_n + 3H]^{3+}$ ions are shown in **Figure 2B**. For $[AK_3 + 3H]^{3+}$ and $[AK_4 + 3H]^{3+}$ ions, charge-solvated structures comprise the most populated cluster which indicates that charge solvation effect is the dominant factor to determine the conformations of these two peptide ions. Interestingly, representative structures of the second most populated cluster are extended conformations (include extended coil and helix): the $[AK_3 + 3H]^{3+}$ ion is an extended random coil and $[AK_4 + 3H]^{3+}$ ion appears to be more helical (**Figure S2B**). Moreover, the $[AK_3 + 3H]^{3+}$ ion has more extended conformations (20% of the conformations in family 1 which contains conformations that have CCSs within $\pm$ 5% of the

experimental CCS value for the triply-charged ions (see **MD Simulation in Experimental Section**) than the $[AK_4 + 3H]^{3+}$ ion (14% of the structures in family 1). This could be caused by stronger Coulombic repulsion imparted by the shorter backbone length of the $[AK_3 + 3H]^{3+}$ ion[57]. As backbone length increases, Coulombic repulsion is minimized and charge solvation is not dominant, so the $[AK_5 + 3H]^{3+}$ ion has higher propensity for a helical conformation. The results for the $[AK_n + 3H]^{3+}$ ions indicate that the presence of charges exerts two effects: multiple charges lead to high Coulombic repulsion resulting in peptide ions with an extended conformation, and alternatively more charged lysines also collapse helical structure by breaking backbone H-bonds to form charge-solvated structures. The final conformation adopted by the peptide is the result of the competing effects of Coulombic repulsion and charge solvation.

Based on the simulation results shown in **Figure 2**, the helical content in **Figure 1C** can be understood. The increasing helical content for $[AK_n + 2H]^{2+}$ ions (n = 3 to 5) appears to be a result of higher helical propensity. Yet, $[AK_n + 3H]^{3+}$ ions present a more complicated case. The $[AK_3 + 3H]^{3+}$ ion is very extended due to high Coulombic repulsion. When the backbone becomes longer, Coulombic repulsion diminishes, and charge solvation begins to be more important, so $[AK_4 + 3H]^{3+}$ ion is less extended than $[AK_3 + 3H]^{3+}$. The $[AK_5 + 3H]^{3+}$ ion has higher helical propensity with a much longer backbone and less effect of charge solvation and Coulombic repulsion (**Figure 3B**). Thus, the helical content for $[AK_n + 3H]^{3+}$ ions decreases and then increases from n =3 to 5. Based on the reasoning for the difference in helical content of $[AK_n + 3H]^{3+}$ ion conformations, the larger CCS for $[AK_3 + 3H]^{3+}$ compared to that for $[AK_3 + 2H]^{2+}$ (**Figure 1A**) is due to higher Coulombic repulsion caused by three charges, while the smaller CCS for $[AK_5 + 3H]^{3+}$ compared to that for $[AK_5 + 2H]^{2+}$ is the result of greater charge solvation effect caused by more charges.

**$AEK_n$ series**. $AEK_n$ peptides contain both basic residue lysine and acidic residue glutamic acid, so they are better candidate peptides to study the effect of side chain interactions on peptide conformation. The CCS profiles of $AEK_n$ ions (**Figure 3A**) are less symmetrical and relatively broader. The CCSs for the $AEK_5$ ions are especially broad, composed of multiple peaks that span a range of more than 100 $Å^2$ suggesting the presence of multiple conformers. When compared to the globular peptide trend line (grey, **Figure 3B**), the plot for $[AEK_n + 2H]^{2+}$ (red) is parallel to the trend line, while that for $[AEK_n + 3H]^{3+}$ deviates in slope. This suggests that the structure of $[AEK_n + 3H]^{3+}$ is more extended than that of $[AEK_n + 2H]^{2+}$. **Figure 3C** shows that the calculated helical content for $[AEK_n + H]^{+}$ decreases sharply and then increases slightly (n = 2 to 5). The helical content for $[AEK_n + 2H]^{2+}$ ions increases slightly and then decreases slowly as repeat unit n increases (from 2 to 5) whereas the helical content for $[AEK_n + 3H]^{3+}$ keeps increasing with increase in repeat unit n (from 3 to 5). In comparison to the helical content in the solution phase (SP, black)[43, 46] (**Figure 3C**), only the triply-charged gas-phase ions ($GP^{3+}$, red) display a trend in helical content vs. repeat unit similar to that of solution phase ions.

Representative structures of the most populated clusters for $AEK_n$ are shown in **Figure 4** (For more information, see **Figure S3** in **SI**). For $AEK_n$ ions, when the number of lysine residues (n) exceeds charge state, there are many possible protonation sites that can lead to $AEK_n$ peptides carrying two or three positive charges. Here, two possibilities are considered: 1) $AEK_n^{EN}$ indicates that only lysine residues are protonated and all glutamic acid residues are neutral (EN); 2) $AEK_n^{EC}$ indicates that all lysines are protonated and (n-2) for $[AEK_n^{EC} + 2H]^{2+}$ or (n-3) for $[AEK_n^{EC} + 3H]^{3+}$ glutamic acids are negatively charged (EC).

The representative conformations of the most populated cluster for doubly-charged ions are shown in **Figure 4A**. Side chain – side chain interactions are marked with red circles. The

protonated lysines and deprotonated glutamic acids are marked by "CP". The lysines and glutamic acids that are not listed in "CP" are neutral**.** The representative structures show that only one side chain – side chain interaction is involved in the structures for $[AEK_n^{EN} + 2H]^{2+}$. In the representative structure for $[AEK_2^{EN} + 2H]^{2+}$, the single E-K interaction lies in the region of helix. It appears that the E-K interaction may stabilize the helical structure, however, all other E-K interactions in the representative conformations for $[AEK_n^{EN} + 2H]^{2+}$ ions (**Figure S3A** in **SI**) lie in the random coil region. These results suggest that E-K ion pairs in the position (i, i + 3) are not a perfect helical stabilizer. This conclusion is consistent with previous studies that suggested that (i, i + 3) spacing decreases the helical abundance due to the competition for backbone H-bonds by side-chains[45, 58]. The representative structures of $[AEK_n^{EC} + 2H]^{2+}$ ions (**Figure 4A**) show that when (n-2) glutamic acid residues are deprotonated, the side chain interactions increase greatly. For example, there are six side chain interactions in $[AEK_3^{EC} + 2H]^{2+}$ compared to one in $[AEK_3^{EN} + 2H]^{2+}$. The increased side chain interactions may be caused by a greater number of charged lysine and glutamic acid side chains that can form more stable salt bridges. Yet, these interactions cross-link with each other to make the entire peptide compact.

$[AEK_n^{EN}+3H]^{3+}$ ions (**Figure 4B**) have more side chain – side chain interactions involved in the representative structures compared to $[AEK_n^{EN}+2H]^{2+}$ ions. For example, the $[AEK_3^{EN}+3H]^{3+}$ ion contains two side chain interactions whereas the $[AEK_4^{EN}+3H]^{3+}$ ion contains three. Compared to the compact charge-solvated structure of $[AEK_n^{EN}+2H]^{2+}$ ions, the representative conformations of $[AEK_n^{EN}+3H]^{3+}$ ions are more extended. The simulation results for $AK_n$ (**Figure 2**) suggest that three charges can result in higher Coulombic repulsion than two charges. Therefore, to release the high Coulombic repulsion, under the support of side chain interactions, $[AEK_n^{EN}+3H]^{3+}$ ions form extended coils leading to higher CCS. On the other hand, when one

glutamic acid is deprotonated ($[AEK_4^{EC}+3H]^{3+}$ ion in **Figure 4B**), side chain – side chain interactions do not increase compared to those in ($[AEK_4^{EN}+3H]^{3+}$. The reason should still be Coulombic repulsion which may inhibit cross-linking side chain – side chain interactions, thus favoring an extended coil structure.

The results for $AEK_n$ further suggest that both charge solvation and Coulombic repulsion play an important role in determining peptide conformations. At lower charge state ($[AEK_n + 2H]^{2+}$), Coulombic repulsion is decreased, so charge solvation is the dominant factor influencing structure and the peptide forms a charge-solvated structure. However, at higher charge state, ($[AEK_n + 3H]^{3+}$, n =3, 4), the higher Coulombic repulsion serves to destabilize the collapsed coil and lead to an extended random coil supported by more side chain interactions. With a longer backbone ($AEK_5$), Coulombic repulsion caused by three charges is not strong enough to inhibit cross-linking interaction, therefore, the profile of $[AEK_5 + 3H]^{3+}$ (**Figure 3A**) is broad and includes some compact conformations

From the simulation results, the difference in helical content for $[AEK_n + 2H]^{2+}$ and $[AEK_n + 3H]^{3+}$ (**Figure 3C**) can be explained. $[AEK_n + 2H]^{2+}$ ions prefer charge-solvated structure owing to cross-linking intermolecular interactions, therefore, the calculated helical content for $[AEK_n + 2H]^{2+}$ is low. However, the conformation preference for $[AEK_n + 3H]^{3+}$ ions is extended random coil which results in larger CCS value, so the calculated helical content is higher.

**2. Charge sites preference**.

**Figures 2 and 4** show the charge sites (position) (CP) of the representative structures which suggest that different peptide structures prefer different charge sites. Both $[AK_4 + 2H]^{2+}$ and $[AK_5 + 2H]^{2+}$ ions have higher helical propensity on the N-terminal side with charges located on

the C-terminal side, suggesting that when charges are located on the C-terminal side, peptides prefer structures with high helical propensity. For $[AEK_n+2H]^{2+}$ ions which contain less helical content, the representative conformations show that the positive charges prefer the N-terminal side. However, the individual structures shown in **Figures 2** and **4** are not comprehensive and do not fully represent charge site preference for the peptides. To identify the relationship between charge sites and peptide structures more precisely, two groups of structures were studied and the results are summarized in **Table 1**. Here, net charge distribution (NCD) ($(\sum i)/z$, where "i" denotes the site number of charges in peptides and "z" denotes the charge state, see the caption of **Table 1** for detail) was used to investigate the effect of charge sites on peptide structure[59]. Group A includes all the candidate conformations from MD simulation with CCSs within ±2% of experimental CCS for $[M + 2H]^{2+}$ ions (±5% of experimental value for $[M + 3H]^{3+}$ ions). The structures in Group A with energy of not more than 20 Kcal/mole higher than the energy of the lowest energy structure comprise Group B. Theoretical net charge distribution (TNCD), which is used to evaluate the NCD of Group A and Group B, is calculated based on the premise that charges have the same chance to be located at all possible charge sites. For example, $[AK_5 + 2H]^{2+}$ has ten possible charge sites, $K^3K^8$, $K^3K^{13}$, $K^3K^{18}$, $K^3K^{23}$, $K^8K^{13}$, $K^8K^{18}$, $K^8K^{23}$, $K^{13}K^{18}$, $K^{13}K^{23}$ and $K^{18}K^{23}$. The $(\sum i)/z$ values of four of these charge sites ($K^3K^8$, $K^3K^{13}$, $K^3K^{18}$, $K^8K^{13}$), two ($K^3K^{23}$ and $K^8K^{18}$), and the other four ($K^8K^{23}$, $K^{13}K^{18}$, $K^{13}K^{23}$, $K^{18}K^{23}$) are less than, equal to, and greater than the value of m/2 =13, respectively. Therefore, 40% of TNCD is located at the N-terminal side, 20% is located in the middle and 40% is located at the C-terminal side. The data in **Table 1** show that the NCD in group A for both $AK_n$ and $AEK_n$ ions is close to the TNCD which suggests that for the conformations in Group A, there is no preference as to charge sites and

charges are located on different sites randomly. However, the NCD in Group B varies depending on the identity of the peptides.

In group B, the NCD changes greatly for $AK_n$ ions. For the shorter $[AK_3 + 2H]^{2+}$ ion, group B contains a higher percentage of structures whose net charge is located in the middle of the peptide ion (*i.e.* the charges are located on $K^3K^{13}$, 45% ($AK_3$-A) and 43% ($AK_3$-B)) when compared to the TNCD (33%), since charges that are separated can reduce Coulombic repulsion more efficiently. This further suggests the existence of Coulombic repulsion for short peptide ions with multiple charges. For longer $[AK_n + 2H]^{2+}$ and $[AK_n + 3H]^{3+}$ ions, $AK_4$ and $AK_5$ have a higher percentage of structures with net charge at the C-terminal side in group B compared to the TNCD (**Table 1**). This is because $AK_4$ and $AK_5$ ions have high helical propensity at the N-terminal side (**Figure 3**), and the net charge at the C-terminal side can stabilize the helix macrodipole[59, 60]. Therefore, owing to the preference of net charge site at the C-terminal side and with less side chain – side chain interactions, $AK_4$ and $AK_5$ ions prefer conformations with high helical propensity at the N-terminal side and charge-solvated structure at the C-terminal side. Owing to the restricted conformation space of peptides with high helical propensity, the CCS profiles for $AK_4$ and $AK_5$ ions are narrow and symmetrical (**Figure 1**). On the other hand, for $[AK_3 +2H]^{2+}$ ions, although the middle net charge site ($K^3K^8$) is preferred, N- and C- net charge sites (net charge is located on the N- and C-terminal side) are still competitive, so the possibility of multiple charge sites results in multiple peaks (**Figure 1**).

Unlike $AK_n$ ions, the NCD for $AEK_n$ ions depends on charge states. $[AEK_n + 2H]^{2+}$ ions have a higher percentage of conformations whose net charge lies on the N-terminal side in group B compared to TNCD (**Table 1**). Combined with the charge-solvated structure for $[AEK_n + 2H]^{2+}$ (**Figure 4A**), the result further suggests that N-terminal side net charge does not promote

formation of helical conformations. Conversely, the net charge site in group B is almost the same as the TNCD for $[AEK_n + 3H]^{3+}$ ions. This suggests that if a peptide does not have high helical propensity, even though its conformation is extended, the C-terminal side is not the preferred position for net charges, *viz.* the charge sites are random.

**3. Meaning of model peptides trend lines.**

In **Figures 1B and 3B**, the trend lines for $[AK_n + 2H]^{2+}$ (blue), $[AK_n + 3H]^{3+}$ (blue) and $[AEK_n + 3H]^{3+}$ (red) ions deviate sharply from that of globular conformations, whereas the trend line for $[AEK_n + 2H]^{2+}$ (red) ions deviate only slightly and is parallel to the globular trend line as repeat unit n increases. These trends are different from the trend in helical content with repeat unit n (**Figures 1C and 3C**). To study the meaning of model peptide trend line, the number of helical residues (residues included in the helical region) was calculated using **Equation 2** and plotted in **Figure 5**. Note that the number of helical residues for $[AK_n + 2H]^{2+}$ ions (blue triangle) increases linearly with repeat unit n, whereas these values change very little for the $[AEK_n + 2H]^{2+}$ ions (red diamond, n = 3 – 5) which matches the trend lines for $[AK_n + 2H]^{2+}$ (**Figure 1B**) and $[AEK_n + 2H]^{2+}$ (**Figure 3B**) ions well. A similar relationship is also observed for $[AK_n + 3H]^{3+}$ and $[AEK_n + 3H]^{3+}$. Therefore, the trend lines for model peptide ions in **Figures 1B and 3B** reflect the change in the number of helical residues with repeat unit (molecular weight).

**4. Chemically modified $AK_n$ and $AEK_n$ ions.** The simulation results suggest that for both $AK_n$ and $AEK_n$ ions, charge states, charge sites, and side chain interactions all affect the conformation of peptide ions. Further evidence in support of this general statement can be found in preliminary results for this same series of peptide ions where E and K side chains have been modified, *viz.* methylation of E and K as well as acetylation of K. Selected CCS profiles for the modified

peptides are shown in **Figure S4** (see **SI**). The CCS profiles show that different chemical modifications of charge carrying side chains have different effects on peptide structure, which is mostly triggered by the changes in the side chain interactions and charge sites.

**Figure S4** shows that after methylation of lysine (**K-$(CH_3)_2$**) and glutamic acid residues (**E-$(CH_3)_n$**), CCS increases. Methylation of lysine and glutamic acid does not change the charge carrier position, *i.e*., lysine side chains are still the preferred charge carrier due to their high proton affinity[61]. Hence, alteration of the side chain – backbone/side chain interaction may be the origin of the change in peptide structure. After methylation, the added methyl groups on lysine may inhibit the side chain – backbone and side chain – side chain interactions due to steric effects. As a result, the charge solvation effect decreases. A similar effect was also observed in the case of methylated glutamic acid residues.

After acetylation of the lysine chains (**Figure S2**, **$(K\text{-}Ac)_n$ and $(K\text{-}Ac)_n$&$(E\text{-}CH_3)_n$**), the lysine side chains have no apparent advantage over backbone N and O atoms in terms of proton affinity.[62, 63] All the backbone and side chain carbonyl groups may be charge carriers. Under this condition, the CCS profiles of $AK_n$ and $AEK_n$ ions change drastically. For $AEK_n$ ions, the peak number increases from a single peak in the unmodified peptide to multiple peaks after acetylation. For $AK_n$ ions, although the peak shapes do not change dramatically, the peak width broadens, indicating the presence of multiple conformers. The CCS profiles for the modified peptides clearly show that charge sites and side chain interactions do affect peptide conformation. Differences in charge solvation and Coulombic repulsion, owing to multiple potential sites of protonation, has been shown to have a direct influence on the heterogeneity of peptide ion conformations.

## Conclusion

Here, we examined the effects of charge states, charge sites and polar side chain interactions on peptide ion structure using two model peptide series, $AK_n$ and $AEK_n$. Results obtained from ion mobility and MD simulations can be summarized as follows: 1). Higher charge state leads to increased Coulombic repulsion and if the Coulombic repulsion dominates in the peptide, an extended random coil structure is formed. For example, for $[AEK_n + 3H]^{3+}$ ions, with the support of Coulombic repulsion, cross-linking side chain interactions decrease greatly. As a result, the side chain – side chain interactions help to support an extended random coil. 2). Side chain – side chain interactions at the position (i, i+5) for K-K and (i, i+3) for E-K do not promote formation of a helical structure. Side chain – backbone interaction tends to break backbone H-bonds and leads to a charge-solvated structure. 3). The position of charges is significant to peptide structure: a C-terminal half net charge that can stabilize the helix macrodipole results in a helical structure while an N-terminal half net charge results in a random coil structure. Therefore, considering all the factors affecting peptide structure, preferred conformations of $AK_n$ ions have higher helical propensity at the N-terminal side and charge-solvated structure at the C-terminal side; $[AEK_n + 2H]^{2+}$ ions prefer to be in charge-solvated conformations owing to the cross-linking side chain – side chain and side chain – backbone interactions, while $[AEK_n + 3H]^{3+}$ ions prefer to be in extended random coil conformations owing to the support of side chain – side chain interactions under higher Coulombic repulsion.

Results from the modification of polar side chains of $AK_n$ and $AEK_n$ ions further indicate that side chain interaction and charge sites significantly influence peptide structure. Methylation of lysine and glutamic acid side chains may increase helical propensity owing to the steric effect of bulky methyl groups which reduces charge solvation effects. On the other hand, the possibility of

multiple charge locations leads to more heterogeneous conformations due to differences in charge solvation and Coulombic repulsion.

Lastly, comparison of the helical content of gas-phase ions with different charge state (+1, +2 and +3) to the helical content of solution-phase ions shows that $[AK_n + 2H]^{2+}$ has similar helical content to solution-phase $AK_n$ ions while the helical content of $[AEK_n + 3H]^{3+}$is similar to that of solution-phase $AEK_n$ ions.

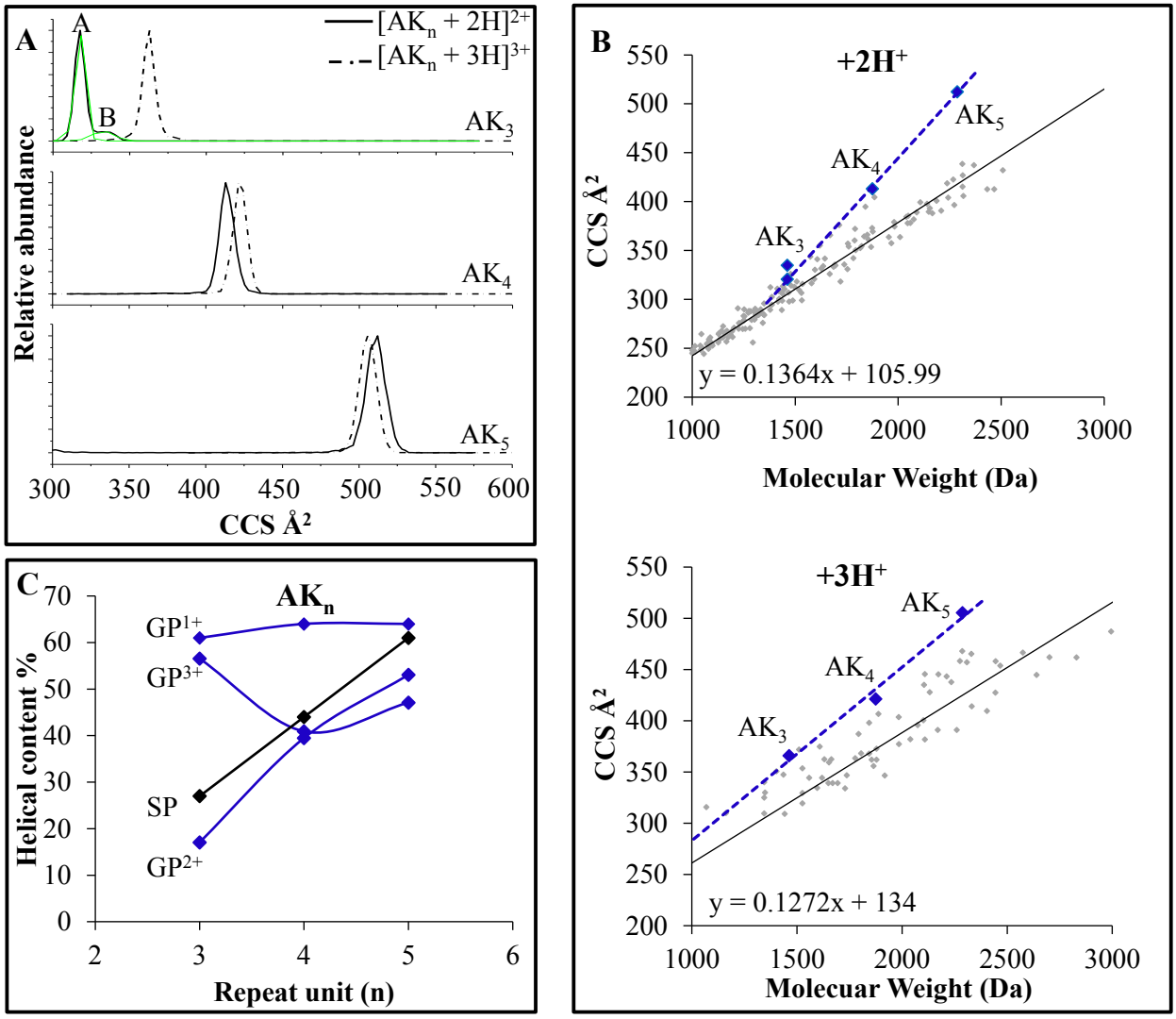



**Figure 1**. CCS profiles and helical content for $AK_n$ **A**. CCS profiles for $AK_n$ ions. Green peaks are the deconvoluted peak profiles for $[AK_3 + 2H]^{2+}$. "A" and "B" are the higher and lower abundance peaks for $[AK_3 + 2H]^{2+}$. **B**. CCS vs. molecular weight of $AK_n$ ions. Grey points are standard peptide data representing CCS for peptide ions that prefer compact globular conformations in the gas-phase. The doubly-charged standard data is cited from the database of the Clemmer group and triply-charged standard data are from digested proteins. **C**. Helical content of $AK_n$ ions in the solution phase (SP, black) and gas-phase (GP, blue).

**Figure 2A: $[AK_n + 2H]^{2+}$**

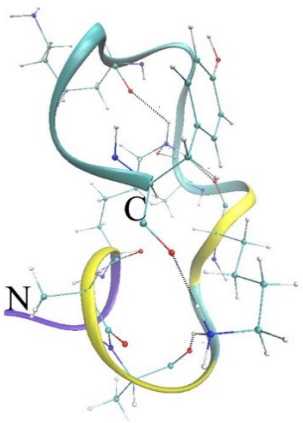
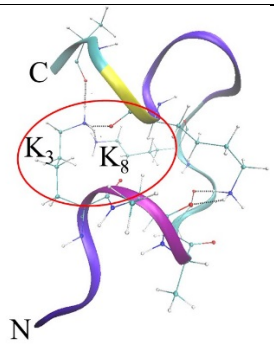
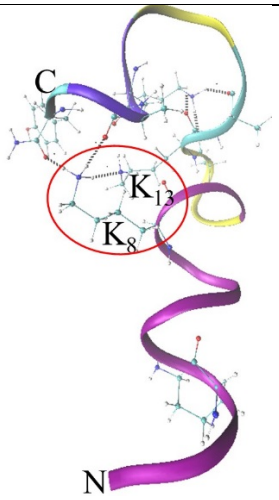
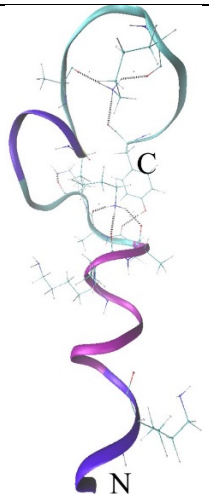
| $AK_3$ (Peak A) Exp. CCS: 321 Å$^2$ | $AK_3$ (Peak B) Exp. CCS: 335 Å$^2$ | $AK_4$ Exp. CCS: 413 Å$^2$ | $AK_5$ Exp. CCS: 512 Å$^2$ |
|---|---|---|---|
|   |   |   |   |
| Cluster 1: 37%<br>CP: $K^3K^8$<br>CCS: 319 ± 3 Å$^2$ | Cluster 2: 28%<br>CP: $K^3K^{13}$<br>CCS: 333 ± 4 Å$^2$ | Cluster 1: 37%<br>CP: $K^8K^{18}$<br>CCS: 416 ± 5 Å$^2$ | Cluster 1: 19%<br>CP: $K^{18}K^{23}$<br>CCS: 503 ± 5 Å$^2$ |

**Figure 2B: $[AK_n + 3H]^{3+}$**

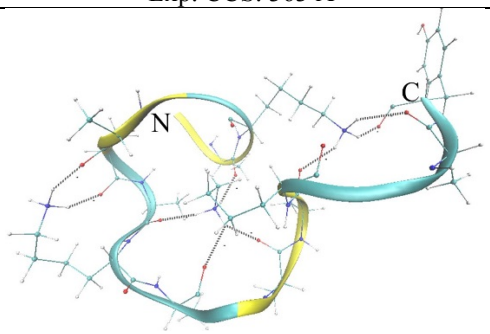
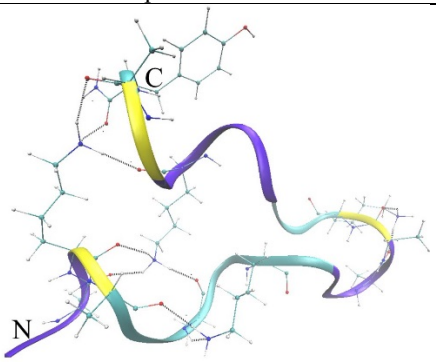
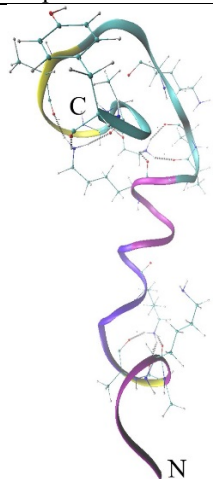
| $AK_3$ Exp. CCS: 363 Å$^2$ | $AK_4$ Exp. CCS: 424 Å$^2$ | $AK_5$ Exp. CCS: 506 Å$^2$ |
|---|---|---|
|   |   |   |
| Cluster 2: 26%<br>CP: $K^3K^8K^{13}$<br>CCS: 351 ± 2 Å$^2$ | Cluster 2: 23%<br>CP: $K^3K^{13}K^{18}$<br>CCS: 424 ± 6 Å$^2$ | Cluster 1: 21%<br>CP: $K^8K^{18}K^{23}$<br>CCS: 484 ± 8 Å$^2$ |

**Figure 2**. Representative ribbon structures of the most populated clusters for $[AK_n+2H]^{2+}$ (**Figure 2A**) and $[AK_n+3H]^{3+}$ (**Figure 2B**) generated by MD simulations. All the conformers shown were generated by VMD. "N" and "C" indicate the N- and C-terminus respectively. "CP" indicates charge sites (position). The superscript on K indicates protonated lysine. For example, $K^3K^8$ indicates that the third and eighth residues, lysines, are protonated. All the polar side chains and the residues that are involved in the interaction with polar side chains are shown. The dashed black lines (---) represent H-bonds. Side chain-side chain interactions are marked with red circles. Colors in the backbones indicate different secondary structures: ■ (Purple): α-helix, ■ (violet): $^3$10-helix, ■ (yellow): turn, ■ (orange): β-sheet, ■ (cyan): random coil; different color on peptide surface indicates different atoms: ■ (cyan): C; ■ (red): O; ■ (blue): N; ■ (grey): H. Charge sites, experimental CCSs and calculated CCSs are also listed.

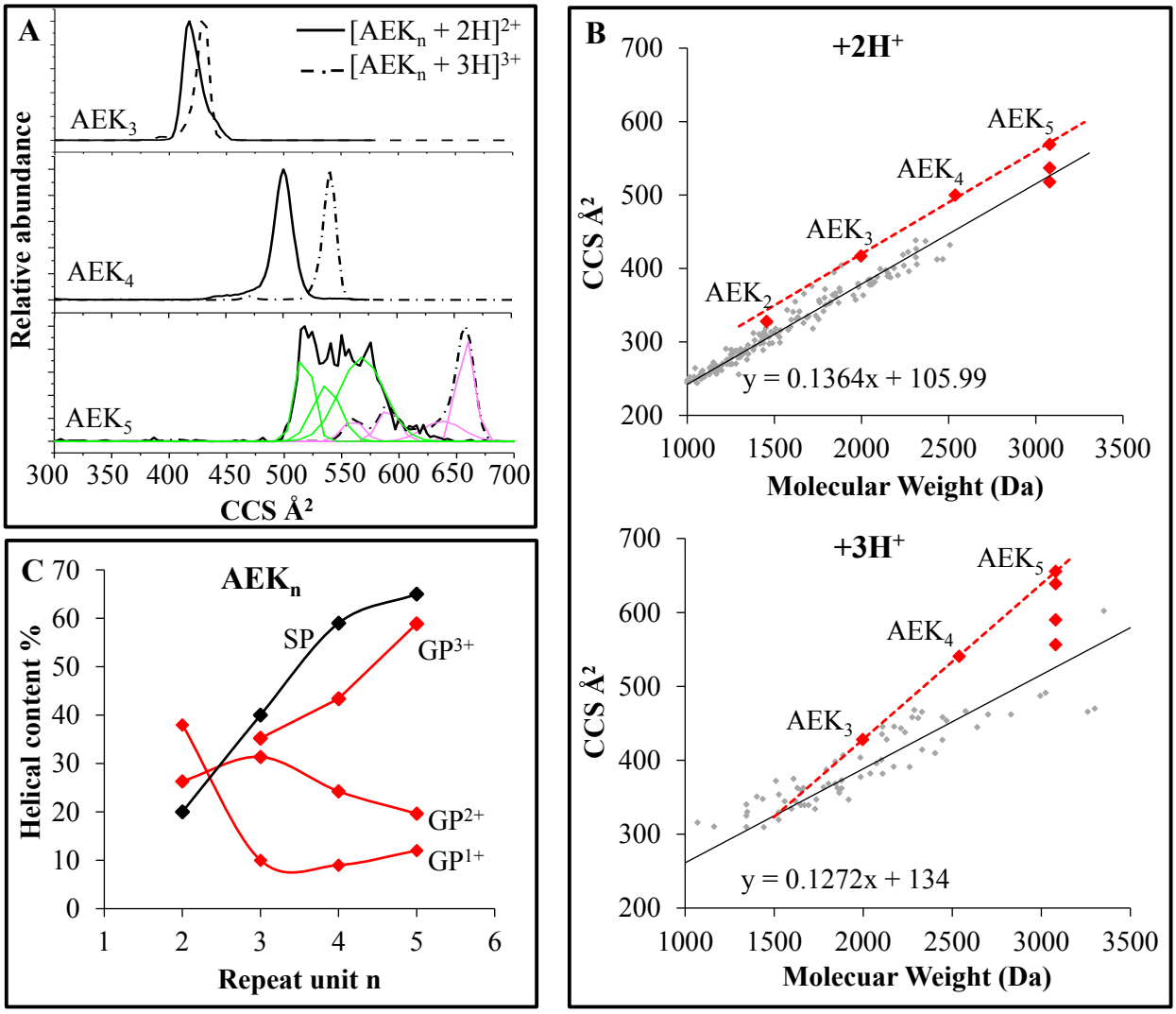



**Figure 3**. CCS profiles and helical content for AEK$_n$. **A**. CCS profiles for AEK$_n$. Green and magenta peaks are the deconvoluted peak profiles for [AEK$_5$ + 2H]$^{2+}$ and [AEK$_5$ + 3H]$^{3+}$ respectively. **B**. CCS vs. molecular weight for AEK$_n$. **C**. Helical content for AEK$_n$ ions in the solution phase (SP, black) and gas phase (GP, red).

**Figure 4A: $[AEK_n + 2H]^{2+}$**

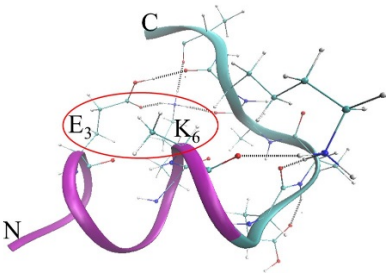
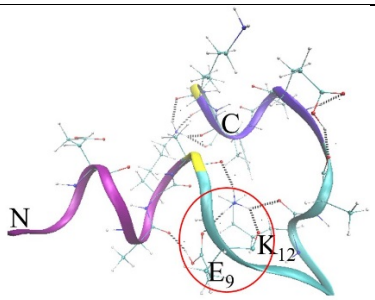
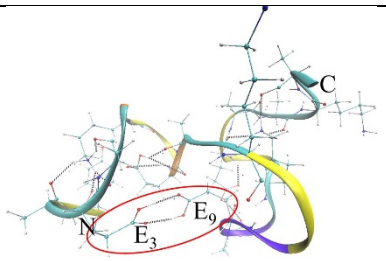
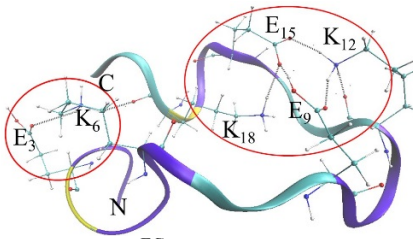
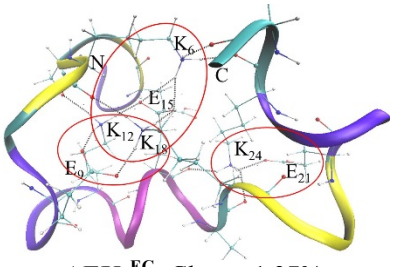
| $AEK_2$ Exp. CCS: 333 Å$^2$ | $AEK_3$ Exp. CCS: 418 Å$^2$ | $AEK_4$ Exp. CCS: 501 Å$^2$ |
|---|---|---|
|  |  |  |
| **$AEK_2^{EN}$**: Cluster 1: 55%<br>CP: $K^6K^{12}$<br>CCS: 335 ± 4 Å$^2$ | **$AEK_3^{EN}$**: Cluster 1 31%<br>CP: $K^6K^{12}$<br>CCS: 419 ± 5 Å$^2$ | **$AEK_4^{EN}$:** Cluster 1 12%<br>CP: $K^6K^{18}$<br>CCS: 493 ± 7 Å$^2$ |
| |  |  |
| | **$AEK_3^{EC}$:** Cluster 1 28%<br>CP: $K^6K^{12}K^{18}E^{15}$<br>CCS: 418 ± 6 Å$^2$ | **$AEK_4^{EC}$:** Cluster 1 37%<br>CP: $K^6K^{12}K^{18}K^{24}E^9E^{15}$<br>CCS: 491 ± 8 Å$^2$ |

**Figure 4B: $[AEK_n + 3H]^{3+}$**

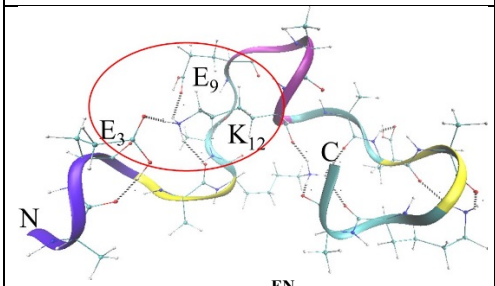
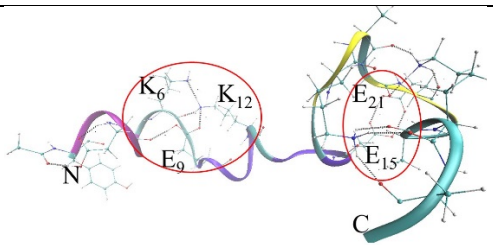
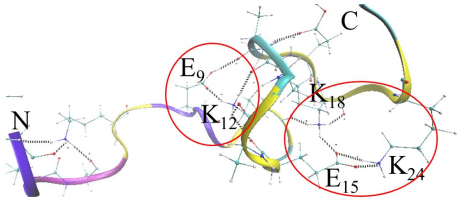
| $AEK_3$ Exp. CCS: 428 Å$^2$ | $AEK_4$ Exp. CCS: 541 Å$^2$ | |
|---|---|---|
|  |  | **$AEK_4^{EN}$**<br>Cluster 1 11%<br>CP: $K^{12}K^{18}K^{24}$<br>CCS: 546 ± 7 Å$^2$ |
| **$AEK_3^{EN}$**<br>Cluster 1 33%<br>CP: $K^6K^{12}K^{18}$<br>CCS: 423 ± 6 Å$^2$ |  | **$AEK_4^{EC}$**<br>Cluster 1 20%<br>CP: $K^6K^{12}K^{18}K^{24}E^{15}$<br>CCS: 557 ± 9 Å$^2$ |

**Figure 4**. Representative ribbon structures of the most populated clusters for $[AEK_n+2H]^{2+}$ (**Figure 4A**) and $[AEK_n+3H]^{3+}$ (**Figure 4B**) generated by molecular dynamics simulations. "$AEK_n^{EN}$" indicates that only lysines are protonated and all the glutamic acids are neutral (EN); "$AEK_n^{EC}$" indicates that all lysines are protonated and some of glutamic acids ((n-2) for doubly-charged ions and (n-3) for triply-charged ions are negatively charged (EC). The superscript of K indicates the residue that is protonated and the superscript of E indicates the residue that is deprotonated (listed by "CP"). The polar side chains and the residues that are involved in the side chain interactions are also represented. The dashed black lines (---) represent H-bonds and the side chain – chain interactions marked with red circles. The color coding is the same as in **Figure 2**.

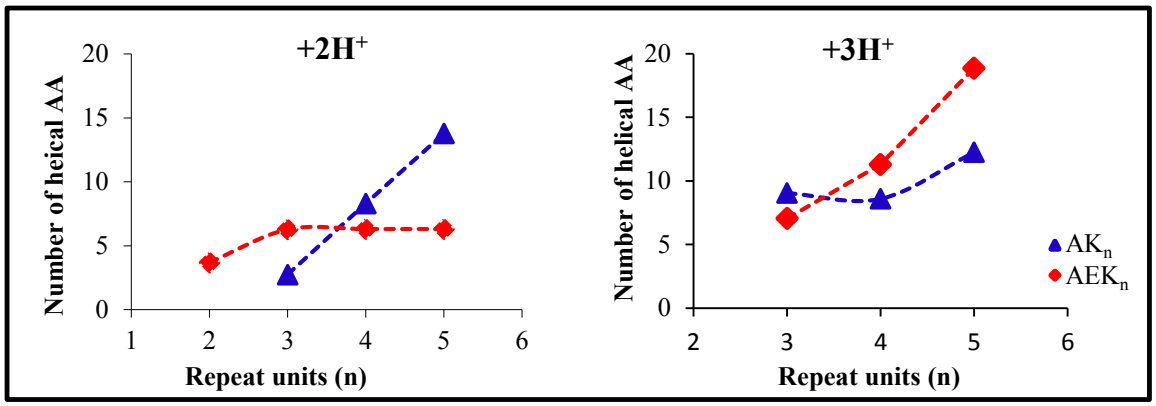



**Figure 5.** The number of helical residues for $AK_n$ (blue triangle) and $AEK_n$ ions (red diamond).

| Theoretical net charge distribution (TNCD) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CS | +2 | | | | | | | +3 | | | |
| Peptides | $AK_3$ | $AK_4$ | $AK_5$ | $AEK_3$ | $AEK_4$ | $AEK_3^{EC}$ | $AEK_4^{EC}$ | $AK_4$ | $AK_5$ | $AEK_4$ | $AEK_4^{EC}$ |
| N (%) | 33 | 33 | 40 | 33 | 33 | / | 17 | 50 | 40 | 25 | 0 |
| M (%) | 33 | 33 | 20 | / | / | / | / | / | 20 | 0 | 25 |
| C (%) | 33 | 33 | 40 | 67 | 67 | 100 | 83 | 50 | 40 | 75 | 75 |
| **Group A (NCD)** | | | | | | | | | | | |
| CS | +2 | | | | | | | +3 | | | |
| Peptides | $AK_3$-A | $AK_3$-B | $AK_4$ | $AK_5$ | $AEK_3$ | $AEK_4$ | $AEK_3^{EC}$ | $AEK_4^{EC}$ | $AK_4$ | $AK_5$ | $AEK_4$ | $AEK_4^{EC}$ |
| N (%) | 37 | 34 | 33 | 32 | 42 | 33 | / | 19 | 57 | 46 | 25 | / |
| M (%) | 33 | 34 | 33 | 19 | / | / | / | / | / | 20 | / | 25 |
| C (%) | 30 | 32 | 34 | 49 | 58 | 67 | 100 | 81 | 43 | 34 | 75 | 75 |
| **Group B (NCD)** | | | | | | | | | | | |
| CS | +2 | | | | | | | +3 | | | |
| Peptides | $AK_3$-A | $AK_3$-B | $AK_4$ | $AK_5$ | $AEK_3$ | $AEK_4$ | $AEK_3^{EC}$ | $AEK_4^{EC}$ | $AK_4$ | $AK_5$ | $AEK_4$ | $AEK_4^{EC}$ |
| N (%) | 26 | 25 | 23 | 7 | 63 | 51 | / | 14 | 28 | 19 | 31 | / |
| M (%) | 45 | 43 | 28 | 17 | / | / | / | / | / | 26 | / | 25 |
| C (%) | 29 | 32 | 49 | 76 | 37 | 49 | 100 | 86 | 72 | 55 | 69 | 75 |

**Table 1. Net charge distribution (NCD) for $AEK_n$ and $AK_n$ ions**. CS is charge state; theoretical net charge distribution (TNCD) is calculated based on the premise that charges have the same chance to be located at all possible charge sites; Group A structures have CCSs that lie in the range of 2% of experimental CCS data for $[M + 2H]^{2+}$ and 5% for $[M + 3H]^{3+}$. Structures in group A with energy not more than 20 kcal/mole higher than the energy of the lowest energy structure comprise group B. The position of net charge is calculated by $(\Sigma i)/z$, where "i" denotes the position number of charges in peptides and "z" denotes the charge state. Middle of peptides is calculated by m/2, where "m" indicates the residue number of the whole peptides. $(\Sigma i)/z < m/2$ indicates the net charge is located on the N-terminal half (**N**) of the peptide, $(\Sigma i)/z = m/2$ indicates the net charge lies near the middle of the peptide (**M**) and $(\Sigma i)/z > m/2$ denotes that the net charge is located on the C-terminal half (**C**).

## Acknowledgements

The authors would like to thank the National Science Foundation (CHE-054587) and the Department of Energy, Basic Energy Sciences (BES DE-FG02-04ER15520) for funding this work.

## References

1. G. Chassaing, O. Convert and S. Lavielle, *European journal of biochemistry / FEBS*, 1986, 154, 77-85.
2. J. R. Macdonald and W. C. Johnson, *Protein Science*, 2001, 10, 1172-1177.
3. Q. Shao, Y. B. Fan, L. J. Yang and Y. Q. Gao, *J Chem Phys*, 2012, 136.
4. R. Beveridge, Q. Chappuis, C. Macphee and P. Barran, *Analyst*, 2013, 138, 32-42.
5. K. R. Shoemaker, P. S. Kim, D. N. Brems, S. Marqusee, E. J. York, I. M. Chaiken, J. M. Stewart and R. L. Baldwin, *Proceedings of the National Academy of Sciences of the United States of America*, 1985, 82, 2349-2353.
6. P. Luo and R. L. Baldwin, *Proceedings of the National Academy of Sciences of the United States of America*, 1999, 96, 4930-4935.
7. J. M. Scholtz, H. Qian, V. H. Robbins and R. L. Baldwin, *Biochemistry*, 1993, 32, 9668-9676.
8. J. S. Smith and J. M. Scholtz, *Biochemistry*, 1998, 37, 33-40.
9. T. Ishimoto, H. Tokiwa, H. Teramae and U. Nagashima, *J Chem Phys*, 2005, 122.
10. M. Karplus and J. A. McCammon, *Nature structural biology*, 2002, 9, 646-652.
11. B. T. Ruotolo and D. H. Russell, *Journal of Physical Chemistry B*, 2004, 108, 15321-15331.
12. B. Schuler, E. A. Lipman, P. J. Steinbach, M. Kumke and W. A. Eaton, *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102, 2754-2759.
13. B. Schuler and W. A. Eaton, *Current opinion in structural biology*, 2008, 18, 16-26.
14. M. Karplus and G. A. Petsko, *Nature*, 1990, 347, 631-639.
15. S. Hwang, Q. Shao, H. Williams, C. Hilty and Y. Q. Gao, *The journal of physical chemistry. B*, 2011, 115, 6653-6660.
16. O. Glatter and O. Kratky, *Small angle x-ray scattering*, Academic Press, London ; New York, 1982.
17. G. L. Hura, A. L. Menon, M. Hammel, R. P. Rambo, F. L. Poole, S. E. Tsutakawa, F. E. Jenney, S. Classen, K. A. Frankel, R. C. Hopkins, S. J. Yang, J. W. Scott, B. D. Dillard, M. W. W. Adams and J. A. Tainer, *Nat Methods*, 2009, 6, 606-U683.
18. L. B. Jensen, K. Mortensen, G. M. Pavan, M. R. Kasimova, D. K. Jensen, V. Gadzhyeva, H. M. Nielsen and C. Foged, *Biomacromolecules*, 2010, 11, 3571-3577.
19. D. Suckau, Y. Shi, S. C. Beu, M. W. Senko, J. P. Quinn, F. M. Wampler, 3rd and F. W. McLafferty, *Proceedings of the National Academy of Sciences of the United States of America*, 1993, 90, 790-793.
20. T. D. Wood, R. A. Chorush, F. M. Wampler, 3rd, D. P. Little, P. B. O'Connor and F. W. McLafferty, *Proceedings of the National Academy of Sciences of the United States of America*, 1995, 92, 2451-2454.
21. J. L. Apuy, X. H. Chen, D. H. Russell, T. O. Baldwin and D. P. Giedroc, *Biochemistry*, 2001, 40, 15164-15175.
22. Y. P. Zhou and R. W. Vachet, *Analytical chemistry*, 2013, 85, 9664-9670.
23. M. Guidi, U. J. Lorenz, G. Papadopoulos, O. V. Boyarkin and T. R. Rizzo, *J Phys Chem A*, 2009, 113, 797-799.

24. J. A. Stearns, C. Seaiby, O. V. Boyarkin and T. R. Rizzo, *Physical chemistry chemical physics : PCCP*, 2009, 11, 125-132.
25. B. T. Ruotolo and C. V. Robinson, *Current opinion in chemical biology*, 2006, 10, 402-408.
26. R. R. Abzalimov, D. A. Kaplan, M. L. Easterling and I. A. Kaltashov, *Journal of the American Society for Mass Spectrometry*, 2009, 20, 1514-1517.
27. T. Wyttenbach, N. A. Pierson, D. E. Clemmer and M. T. Bowers, *Annual review of physical chemistry*, 2014, 65, 175-196.
28. C. S. Creaser, J. R. Griffiths, C. J. Bramwell, S. Noreen, C. A. Hill and C. L. P. Thomas, *Analyst*, 2004, 129, 984-994.
29. K. J. Lightwahl, B. L. Schwartz and R. D. Smith, *Journal of the American Chemical Society*, 1994, 116, 5271-5278.
30. J. A. Loo, *International journal of mass spectrometry*, 2000, 200, 175-186.
31. R. D. Smith, *International journal of mass spectrometry*, 2000, 200, 509-544.
32. J. M. Daniel, S. D. Friess, S. Rajagopalan, S. Wendt and R. Zenobi, *International journal of mass spectrometry*, 2002, 216, 1-27.
33. N. A. Pierson, L. Chen, S. J. Valentine, D. H. Russell and D. E. Clemmer, *Journal of the American Chemical Society*, 2011, 133, 13810-13813.
34. O. S. Skinner, K. Breuker and F. W. McLafferty, *Journal of the American Society for Mass Spectrometry*, 2013, 24, 807-810.
35. J. A. Silveira, K. L. Fort, D. Kim, K. A. Servage, N. A. Pierson, D. E. Clemmer and D. H. Russell, *Journal of the American Chemical Society*, 2013, 135, 19147-19153.
36. C. Bleiholder, T. Wyttenbach and M. T. Bowers, *International journal of mass spectrometry*, 2011, 308, 1-10.
37. S. E. Anderson, C. Bleiholder, E. R. Brocker, P. J. Stang and M. T. Bowers, *International journal of mass spectrometry*, 2012, 330, 78-84.
38. C. Bleiholder, S. Contreras and M. T. Bowers, *International journal of mass spectrometry*, 2013, 354, 275-280.
39. C. Bleiholder, S. Contreras, T. D. Do and M. T. Bowers, *International journal of mass spectrometry*, 2013, 345, 89-96.
40. M. F. Mesleh, J. M. Hunter, A. A. Shvartsburg, G. C. Schatz and M. F. Jarrold, *J Phys Chem-Us*, 1996, 100, 16082-16086.
41. R. V. de Carvalho, D. Lopez-Ferrer, K. S. Guimaraes and R. D. Lins, *Journal of computational chemistry*, 2013, 34, 1707-1718.
42. A. Chakrabartty and R. L. Baldwin, *Advances in protein chemistry*, 1995, 46, 141-176.
43. J. M. Scholtz, H. Qian, E. J. York, J. M. Stewart and R. L. Baldwin, *Biopolymers*, 1991, 31, 1463-1470.
44. C. A. Rohl, J. M. Scholtz, E. J. York, J. M. Stewart and R. L. Baldwin, *Biochemistry*, 1992, 31, 1263-1269.
45. J. R. McLean, J. A. McLean, Z. X. Wu, C. Becker, L. M. Perez, C. N. Pace, J. M. Scholtz and D. H. Russell, *Journal of Physical Chemistry B*, 2010, 114, 809-816.
46. V. Munoz and L. Serrano, *Biopolymers*, 1997, 41, 495-509.
47. B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S. J. Hyung and C. V. Robinson, *Nat Protoc*, 2008, 3, 1139-1152.
48. S. J. Valentine, A. E. Counterman and D. E. Clemmer, *Journal of the American Society for Mass Spectrometry*, 1999, 10, 1188-1211.
49. T. A. D. D.A. Case, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R., R. C. W. Luo, W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg,, I. K. G. Seabra, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell,, H. G. T. Steinbrecher, Q. Cai, X. Ye, J. Wang, M.-J.

Hsieh, G. Cui, D.R. Roe, D.H., M. G. S. Mathews, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and and P. A. Kollman, *AMBER 11, University of California, San Francisco*, 2010.
50. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins*, 2006, 65, 712-725.
51. W. Kabsch and C. Sander, *Biopolymers*, 1983, 22, 2577-2637.
52. R. P. Joosten, T. A. te Beek, E. Krieger, M. L. Hekkelman, R. W. Hooft, R. Schneider, C. Sander and G. Vriend, *Nucleic acids research*, 2011, 39, D411-419.
53. E. K. J. Van Holde, W. C.; Ho, P. S, *Principles of Physical Biochemistry,*Prentice Hall: Upper Saddle River, NJ, 1998.
54. A. M. Klibanov, *Nature*, 2001, 409, 241-246.
55. C. Mattos and D. Ringe, *Current opinion in structural biology*, 2001, 11, 761-764.
56. B. T. Ruotolo, G. F. Verbeck, L. M. Thomson, K. J. Gillig and D. H. Russell, *Journal of the American Chemical Society*, 2002, 124, 4214-4215.
57. C. S. Hoaglund-Hyzer, A. E. Counterman and D. E. Clemmer, *Chemical reviews*, 1999, 99, 3037-3080.
58. R. Sudha, M. Kohtani, G. A. Breaux and M. F. Jarrold, *Journal of the American Chemical Society*, 2004, 126, 2777-2784.
59. A. E. Counterman and D. E. Clemmer, *Journal of the American Chemical Society*, 2001, 123, 1490-1498.
60. R. R. Hudgins, M. A. Ratner and M. F. Jarrold, *Journal of the American Chemical Society*, 1998, 120, 12974-12975.
61. M. F. Bush, J. Oomens and E. R. Williams, *J Phys Chem A*, 2009, 113, 431-438.
62. V. Addario, Y. Z. Guo, I. K. Chu, Y. Ling, G. Ruggerio, C. F. Rodriquez, A. C. Hopkinson and K. W. M. Siu, *International journal of mass spectrometry*, 2002, 219, 101-114.
63. J. Evans, G. Nicol and B. Munson, *Journal of the American Society for Mass Spectrometry*, 2000, 11, 789-796.