MONITORING THE DESOLVATION OF IONS AND CREATE CANDIDATE STRUCTURE FOR THE IONS DETECTED BY ION MOBILITY–MASS SPECTROMETRY STUDY BY MOLECULAR DYNAMICS SIMULATIONS

A Dissertation

By

DOYONG KIM

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee, David H. Russell
Committee Members, Emile A. Schweikert
Christian Hilty
John Scholtz
Head of Department, Simon W. North

August 2016

Major Subject: Chemistry

ABSTRACT

Electrospray ionization ion mobility mass spectrometry (ESI-IM-MS) has emerged in recent years as a tool to separate heterogeneous conformer populations and determine ion-neutral collision cross sections (CCS). However, IM cannot provide specific structural information such as atomic coordinates obtained from X-ray crystallography or nuclear magnetic resonance (NMR) experiments. In this work, molecular dynamics simulations (MDS), and quantum mechanical simulations were used to provide structural information for three families of peptides previously characterized by IM: substance P (SP) and its mutants, phosphopeptides with various phosphorylation locations and mutants, and polyprolines. The simulated structure's theoretical CCS was calculated by using the MOBCAL trajectory method. CCS values from ESI-IM-MS experiments provided boundary conditions to filter simulated structures to within ±3% of experimental CCS values. Additional experimental results, such as electron capture dissociation (ECD) and proline cis/trans conformation results, were also used to filter simulated structures which enhanced further the validity of simulated structure.

The electrospray process was also investigated by utilizing MDS for the ultimate goal to understand the structure evolution of ions from solution to the gas phase. A series of conditioning simulations determined the optimum droplet size to be ~2400 water molecules, and the temperature of desolvation simulation was determined to be 360 K, which is consistent with the previously reported ESI-cryo-IM-MS heated capillary temperature. The net charge of the droplet was determined to be +15. Using these

determined conditions, three desolvation simulations were performed with SP as model. The desolvation results showed evaporation and droplet fission which were consistent with other simulations. A further desolvation study incorporated Cl- in the droplet containing SP^{3+} as counter ions. The structural evolution of the SP^{3+} ion during the desolvation process was monitored via CCS calculations. The calculated CCS of the SP^{3+} ion during the later stages of desolvation was consistent with the previously determined ESI-IM-MS value. This simulation indicates that the droplet desolvation follows the process best described as the combined charge residue-field emission model.

DEDICATION

This dissertation is dedicated to my wife Weonjin Shin and my parents Dr. Hyung Yang Kim and Kyung Mee Kim.

In memory of my grandfather, Kim, Jungcheon.

ACKNOWLEDGEMENT

The graduate career at Texas A&M University with my advisor Dr. David Russell was a great experience. The lessons I learned from Dr. Russell were unforgettable. It was honor for me to participate in the research projects which were coupling simulations and mass spectrometry. The guidance by him was priceless and I would like to express my sincere appreciation.

I would like to thank others whom I worked with during my academic years. Dr. Lisa Perez assisted and answered many questions that I had for my simulations. I really appreciate Dr. Liuxi Chen, who has developed fundamental simulation techniques. I would like to mention former and current members of Russell Lab who have aided me during long periods of graduate school: Dr. Junho Jeon, Nicole Wagner, John Patrick, and Dr. Joshua Silveira. I thank Dr. William Russell for teaching me how to work in LBMS.

Finally, I acknowledge the important financial support provided for this research from the following organizations: The National Science Foundation, the Robert A. Welch Foundation, the United States Department of Energy, the Division of Chemical Sciences and the Department of Chemistry at Texas A&M University. I would also like to thank my committee members, Dr. Emile A. Schweikert, Dr. Christian Hilty, and Dr. Martin Scholtz.

It was a great experience at Texas A&M University for the last six and half years, and I enjoyed the science I did. Investigating the desolvation process of water droplets has been a great interest of mine, and I hope the simulation that I did here at

Texas A&M University can help solve the questions that we have concerning the desolvation process.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

NOMENCLATURE

ESI	Electrospray Ionization
IM	Ion Mobility
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
CCS	Collision Cross Section
MDS	Molecular Dynamics Simulation
SA	Simulated Annealing
TM	Trajectory Method
PA	Projection Approximation
EHSS	Exact Hard Sphere Scattering
PSA	Projection Superposition Approximation
CRM	Charge Residue Model
IEM	Ion Evaporation Model
AMBER	Assisted Model Building with Energy Refinement

CHAPTER I

INTRODUCTION

Determining the structure of biomolecules is one of the most important steps in biochemical research. Traditionally, X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are used to determine the structure of the protein.[1,2] These two techniques are well established and developed. X-ray crystallography in particular can provide biomolecular atomic resolution down to ~0.5 Å.[3,4] According to Research Collaboratory for Structural Bioinformatics Protein Data Bank, ~80% of reported structures were analyzed by using X-ray crystallography, whereas only 16% was determined using NMR.[5] However these two techniques have limitations. X-ray crystallography sample must be crystallized to a perfect quality for analysis, sometimes remarkably difficult task; however, it is still difficult to resolve the exact location of hydrogens even with a perfect crystal.[5] Consequently, most PDB files from X-ray crystallographic information consist of only heavy atoms.[6] NMR does not require the crystallization as X-ray crystallography; however, heterogeneity of biomolecules in solution complicates structure analysis. Also the samples have to be at concentrations where protein aggregation or protein-protein interaction can inadvertently influence the results.[7] ESI-IM-MS is an alternative method of determining protein structure. ESI-IM-MS can analyze even small amount of samples. It is also capable of separating heterogeneous conformational populations and elucidating relative molecular size.[8,9] However, ESI-IM-MS lacks atomic resolution provided by X-ray crystallography and

NMR. Furthermore, the transition from solution to gas phase raises concerns as for the biological relevance of the information gained.

Droplet formation via the electrospray process is widely accepted to consist of evaporation and droplet fission; however, the final stages responsible for producing gas phase ion is still debatable. Currently, there are two widely accepted models for the mechanism of desolvation. First, the charge residue model (CRM) is considered as the most applicable to the desolvation of large molecules, where droplet fission and solvent evaporation occur until only one analyte ion is present in the droplet. Then, solvent evaporation leads to a solvent-free gas phase ion.[10-12] Second, the ion evaporation model (IEM) states that the analyte ion evaporates from the droplet as droplet size is reduced.[13,14] The consensus on IEM is that it is only applicable to small molecule ionization. Recently, direct experimental evidence for CRM was reported by Silveria et al. using ESI-cryo-IM-MS.[15]

ESI-IM-MS can provide biomolecular structural information can be elucidated by ESI-IM-MS via CCS measurements. Electric field applied across IM cell propels ions through buffer gas resulting ion-neutral collisions, thus IM separates ions with different CCS. The larger the ions are, the greater resistance will they encounter due to an increased number of collisions in the drift tube. Interfacing IMS with MS provides the mass to charge ratios allowing for identification of analyte ions. Many research labs have applied ESI-IM-MS to the study of small peptides to large protein complexes.[16-22] However, obtaining detailed structural information from the CCS values alone is challenging. To

determine specific structural information from the CCS measurements, theoretical simulations are necessary.

Molecular dynamic simulation (MDS) is a powerful technique to determine biological structures. MDS technique has proven to be an important tool to understand the biological structure, particularly after 2013 the Nobel Prize was awarded to three theoretical chemists/biologists: Martin Karplus, Michael Levitt, and Arieh Warshel. Since its development, advances in accuracy, speed, and computing power has widened application of MDS. However, simulations still require vast amounts of computational power. Large proteins can be simulated;[21] however, the outcome of those computational results are heavily dependent on the input parameters.[23] Molecular dynamics and quantum mechanics simulations are currently available with many commercial codes.[24,25] These programs simulate structures from which theoretical CCS of the structures can be calculated.

Theoretical CCS can be computed using 4 widely accepted methods: the projection approximation (PA), the exact hard sphere scattering (EHSS), the trajectory method (TM), and the projection superposition approximation (PSA) method. The first three methods can be performed by MOBCAL,[26,27] where the PSA method was developed by Bowers.[28-31] The computational cost is the lowest for the PA but it is not appropriate for large molecules because the method provides artificially lower calculated CCS.[26,32] The EHSS values are more accurate for larger systems, but for the smaller systems the model fails to calculate the correct CCS values because it does not incorporate ion-neutral interaction.[8,26]

The TM works well for any system regardless of the size, but it is the most computationally expensive.[8,32] The PSA method is modification of the PA where the values for the large systems are more accurate while retaining low computational cost. In this work, the TM from MOBCAL was exclusively used to calculate CCS of the simulated structures.

In this research, bimolecular structures will be calculated using MDS in effort to explain the results of ESI-IM -MS experiments. In Chapter III, MDS is utilized to create a candidate library for Substance P and its mutants, phosphopeptides, and poly-proline peptides. CCS values are used to filter the simulated structures. Chapter IV and V will demonstrate the capabilities of MDS to provide new insights to the desolvation process of Substance P where current instrumentation cannot explore.

CHAPTER II

METHODS

Molecular dynamics simulation and simulated annealing molecular dynamics were performed by AMBER 11.[25] Force fields for the standard residues were provided by the AMBER FF99SB all atom force field. Custom force fields for the amino acid residue library, which were not provided by the AMBER FF99SB force field, was made in house by utilizing Gaussian 03[24] and R.E.D. III,[33] charge derivation program. Structures of the custom amino acid residues were optimized by Gaussian 03 using HF method and using 6-31+ basis set.[24] The charge derivation for the custom force field was completed by using resp R.E.D III tools.

Molecular Dynamics Simulation

Starting structures for MDS were built from Xleap, which is included within the AMBER software suite. Each MDS starting structure was set as an extended structure as it was created from Xleap. The MDS simulation starts with the initial minimization of the starting structure. Then it is followed by a slow increase of the system temperature to the pre-set main simulation temperature. During the simulation, the temperature of the system fluctuates around the indicated temperature (see Figure 1 for an example). Typical simulation temperatures are set between 80K to1000K depending on the purpose of the simulation.

Simulated Annealing Molecular Dynamics Simulation (SA-MDS)

Script developed in house by Dr. Lisa Perez was used to perform SA-MDS. The SA-MDS was started with two starting structures, extended and α-helix. The α-helix structure was achieved by altering the backbone dihedral angles of the extended structure by the Xleap program. The difference between molecular dynamics simulation and simulated annealing molecular dynamics simulation is that SA-MDS goes thru temperature cycles whereas the temperature is set to a constant for a typical MDS. In this research, simulated annealing was set to cycle between 300K and 1000K; the temperature cycle is shown in **Figure 1**. The high temperature is necessary to energize the molecule sufficient to overcome potential barriers, resulting in different structures when it is cooled back down to a lower temperature. First tier simulation starts with two starting structure: elongated and helical. The simulation for each calculates a total of 1000 structures. Then the second tier simulation starts with 12 randomly selected structures each from previous simulations to produce an additional 1000 structures for each starting structure. The third tier repeats the previous step by selecting 12 random structure which simulates an additional 1000 structures each. Ultimately, a total of 50,000 candidate structures will be simulated per molecule, which each of the structure is the minimized structure after single heating and cooling cycles. This systematical high temperature annealing cycles can probe the entire potential energy landscape.

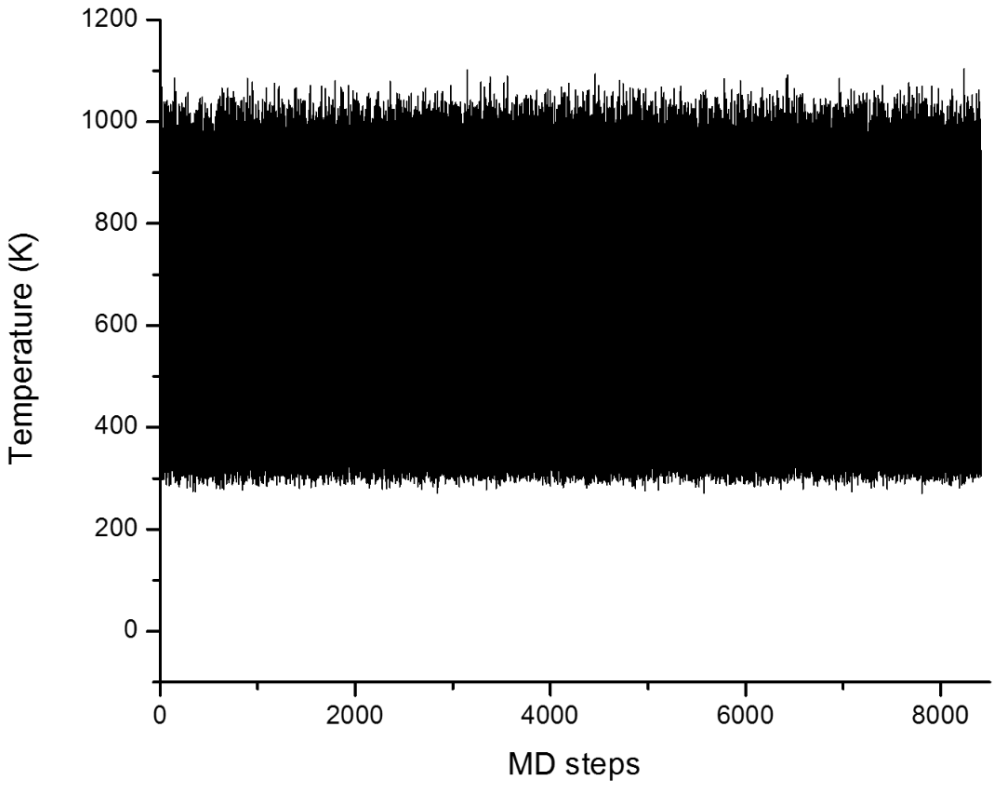

Figure 1. Temperature fluctuation plot for a typical SA-MDS run. This simulation was performed between initial temperature of 300 K and final temperature of 1000 K.

Desolvation Molecular Dynamics Simulation

Peptide solvation was performed using solvateOct TIP3PBOX function from AMBER, which solvates the target peptide using the TIP3P water model. Then the periodic box size is increased to accommodate the evaporation and droplet fission of the parent droplet. After the MDS simulation run, the water molecules evaporated and droplet fission products are removed from the structure PDB file. The distance required for water removal is typically 15Å. The initial temperature setting was 360K until desolvation slows down. Then the temperature is increased to 420K, which speeds up the evaporation rate. If the temperature is not kept constant, the evaporation will cool the system down and eventually freeze the droplet preventing from further evaporation. However when the temperature is kept constant, AMBER maintains the temperature of the system around the prescribed temperature in the input file as shown in **Figure 2**. **Figure 1** the temperature fluctuate around the temperature set in the input files. The temperature is kept constant by artificial collisions to the system so that the heat lost will be compensated for.

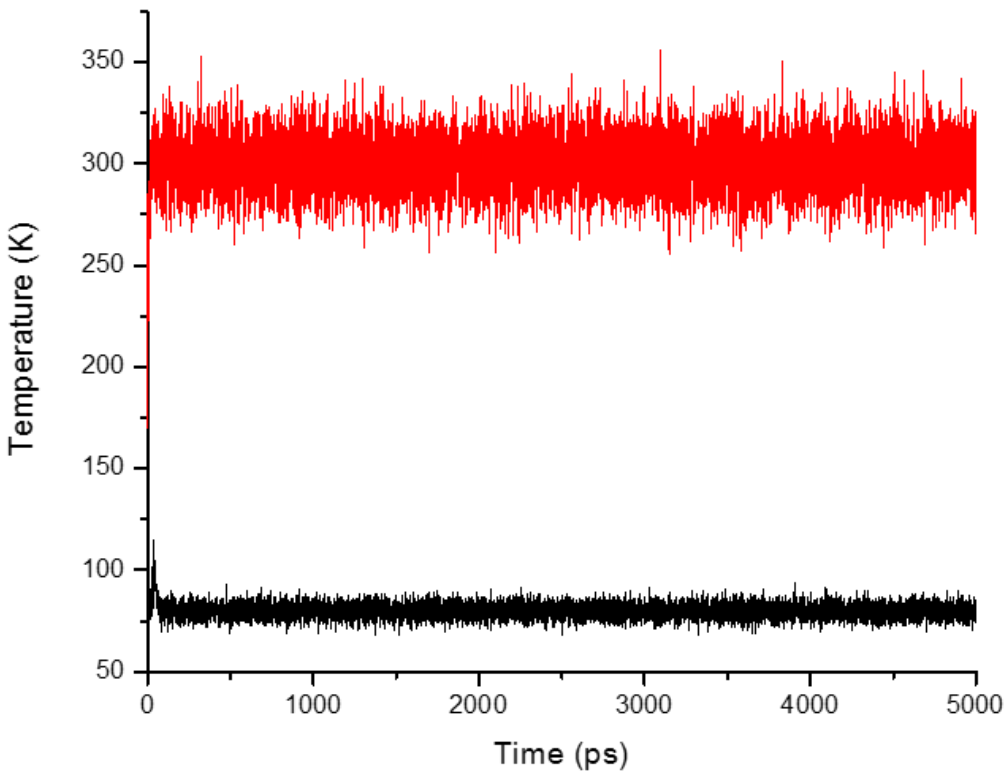

Figure 2. Temperature fluctuation plot for a MDS run set at 300 K (red) and 80 K (black). Note, the temperature does not remain at a constant temperature due to energy flux algorithm, but the temperature is centered on the preset temperature.

Collision Cross Section Calculations

CCS calculation was performed using the MOBCAL[26,27] trajectory method. The trajectory method is computationally more expensive compared to the projection method or elastic hard sphere scattering method. However, the additional computational cost was justified by the higher accuracy obtained from the trajectory method. The experimental CCS measured from drift tube or traveling wave ion mobility was used to filter simulated structures according to their calculated CCS. The filtering boundary was set to ±3% of the experimental CCS. Any structure with a calculated CCS value within the filtering boundary was selected for further analysis. A graphical example of filtering process is shown in **Figure 3**.

Data Analysis

Statistical analysis was performed using the Kclust clustering algorithm provided by AMBER MMTSB tool set. The Kclust algorithm partitions data points into k numbers of clusters where the member of each cluster is centered on the mean; therefore the centroid, or the average of the cluster, can be used as the best representative structure of a particular cluster. The clustering criteria were set to use location of Cα and the rmsd clustering radius was set to 4Å. During the clustering algorithm, many iterations of regrouping is performed, and at the end a new centroid structure is calculated. The clustering can be performed before CCS filtering or after CCS filtering. In this research all of the clustering was performed after the CCS filtering, due to the intense computational resources required by the clustering algorithm. Since the numerous clusters

created during the clustering algorithm, systematically the best fit structures from the top 3 most populated clusters were considered significant.

Distance calculation script was developed in house by calculating the distance of two points in Cartesian coordinates by utilizing distance calculation equation. Structure having the calculated value less than 4.5Å was then selected for further analysis. 4.5 Å is the upper end limit for salt bridge interaction between two opposite charged species.[34] Due to the technical limitation, bash shell does not support decimal values, and the distance was multiplied by 100, then selected any structures having calculated number below 450. This same script was modified to calculate distances between two carbon atoms in the backbone to determine the cis or trans proline isomer by comparing distances.

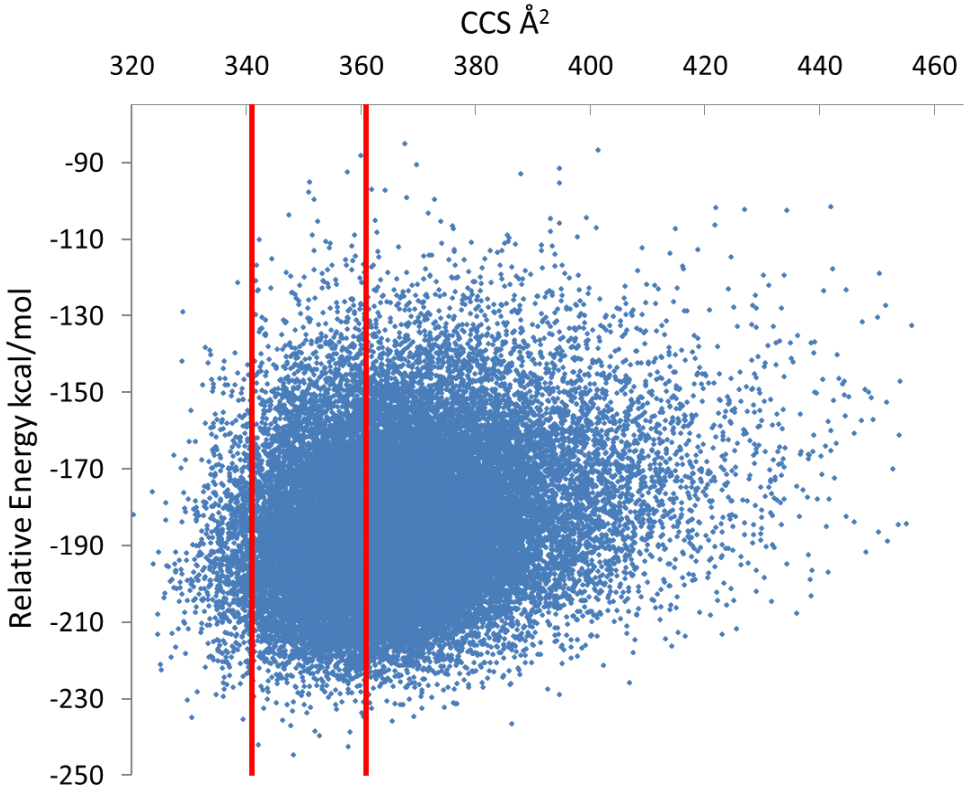

Figure 3. Scatter plot of simulated annealing structure CCS and relative energy of the structure. Only the structure that fall between the red lines, 352 Å ± 3%, are selected and clustered.

CHAPTER III

SIMULATION OF MODEL PEPTIDES

ESI-IM-MS has proven to be very useful for separating heterogeneous conformational populations and determining the CCS of the molecule, from which you can ascertain relative size. However, one fundamental limitation of ESI-IM-MS technique when compared to NMR or X-ray crystallography is that the exact structure cannot be determined. To gain insight on the structure of the molecule from CCS information, a theoretical approach has to be taken. Previous studies have shown that the CCS profile information can be used to filter simulated structures of biomolecules to determine a statistically driven best-fit structure.[35-40] In this chapter three peptides, phosphopeptides, Substance P, and polyproline-13, are simulated by SA-MDS and MDS and filtered by the experimental data including CCS values, electrostatic interactions, and cis-trans proline isomerization.

2+ Phosphopeptide

Phosphoproteins are common in eukaryotes, where up to one third of all proteins are phosphorylated.[41] This makes phosphorylation one of the most common post translational modifications. Previous studies have shown that phosphorylation influences many aspects of a protein or a peptide, such as changing secondary structure, performing

*Reprinted with permission from "Probing the Electron Capture Dissociation Mass Spectrometry of Phosphopeptides with Traveling Wave Ion Mobility Spectrometry and Molecular Dynamics Simulations" Kim, D.; Pai, P.-J.; Creese, A. J.; Jones, A. W.; Russell, D. H.; Cooper, H. J. J. Am. Soc. Mass. Spectrom. 2015, 26, 1004-1013. Copyright 2015 by Springer

on/off signaling, and changing the function of a protein.[42,43] These functions make phosphorylation a biologically important post translational modification. When proper phosphorylation is not conducted, the protein will not function as designed; misphosphorylation can cause many disorders such as Alzheimer's and Parkinson's.[44,45] Hence, researchers have investigated the effects and characteristics of phosphorylation on phosphopeptide and phosphoprotein. Ruotolo et al. and Thalassinos et al. have shown that phosphorylation of some phosphorylated peptides resulted in more compact peptides compared to non-phosphorylated analogues.[46,47] This compaction is thought to be a result of intra-molecular interaction between the phosphate group and basic side chains. In addition, Iakoucheva et al. reported that there are higher probabilities of finding the basic amino acid side chains arginine and lysine near possible phosphorylation sites.[45] Therefore, local intra-molecular interactions, such as salt-bridging, would be more influential to the local and overall structure of the phosphoprotein. To investigate the phosphorylation effect on the structure of the peptides and proteins, ion mobility mass spectrometry (IM-MS) and molecular dynamic simulation are used to generate a library of candidate structures and determine the best representative structure. From these structures, the intra-molecular forces and secondary and tertiary structures can be studied for the effect of phosphorylation.

Previous ESI-ECD-MS studies on a series of model serine-phosphorylated peptides, APLSF**(X)**GSLPKSYVK, where **X =** L or R, have shown evidence of salt-bridge formation between basic residues and the phosphate group by having reduced numbers of backbone cleavage ions between the phosphate group and basic residues, arginine (Arg)

and lysine (Lys). Arginine-containing peptides showed complete absence of backbone fragmentation between the phosphate group and the basic residues leading to a conclusion that arginine promotes formation of salt-bridges. Ion mobility experiments have shown multiple conformations for the model peptides. CCS values for the phosphorylated peptides have revealed that the model peptides are in fact more compact compared to other phosphate group-containing peptides. In this study, AMBER 11 was employed to model the $[M + 2H]^{2+}$ and $[M + 3H]^{3+}$ ions for each peptide. The electrostatic interaction, mainly salt-bridge formation, was identified by the distance between two group distances. Previously defined distance of <4Å by Nussinov et al., was used to find the salt-bridge formation between phosphate group and the basic residues.[34] Because the model peptides contain multiple basic amino acids, the effect of charge site on conformer preferences was included in this study.

As described above, molecular dynamics simulations were performed for each of the phosphopeptides. Candidate structures were generated by simulated annealing and their CCS calculated via the trajectory method in MOBCAL. The structures were subsequently filtered to with ±3% of the experimentally derived CCS, and the filtered structures subjected to *k*-means cluster analysis. To address the question of the nature of any intramolecular non-covalent interactions, for each set of peptides two protonation patterns were considered. For the R6 peptides: (a) protonation of R6, K11 and K15 together with deprotonation of the phospho-group for an overall charge of +2 (for which we might anticipate salt-bridge formation), denoted herein as (R+, K11+, K15+, POx-); and (b) protonation of R6 and K15 with neutral phosphate (ionic hydrogen bonds

anticipated), denoted herein as (R+, K15+, POx0). For the L6 peptides: (a) protonation of the N-terminus, K11 and K15, and deprotonation of the phospho-group (NT+, K11+, K15+, PO_x-), and (b) protonation of K11 and K15 with neutral phosphate (K11+, K15+, PO_x0). (Note that protonation of the N-terminus was not considered when modelling the R6 peptides: N-acetylated versions of the phosphopeptides showed identical ECD fragmentation behaviour to that of the non-acetylated R6 peptides (data not shown), suggesting that protonation of the N-terminus is not involved).

The mobility data for R6pS4 suggests that there are two conformers, one of CCS 337 $Å^2$ and one of 351 $Å^2$.[35] The modelled structures closest to the centroid of the three most populated clusters for conformer with CCS 337 $Å^2$ of R6pS4 (R+, K11+, K15+, POx-) are shown in **Figure 4 A, B** and **C**. **Figure 4 A** comprises ~16% of the filtered structures (total structures 3730); **Figure 4 B** comprises ~13%, and **Figure 4 C** comprises ~10%. The structures were analysed according to the proximity between the basic amino acid side chains and the phosphate group. Ion pairs may be considered salt-bridges if the distance between charged atoms is less than 4 Å.[48] The results of the distance analyses are summarized in **Table 1**. For cluster 1 (**Figure 4 A**), the distance between pS4 and K11 was <4 Å suggesting the presence of a salt-bridge between these residues. This observation correlates well with the ECD data, assuming that the salt-bridge survives the electron capture/dissociation event. The distances between pS4-R6 and pS4K15 were both > 4 Å. For cluster 2 (**Figure 4 B**), all of the protonated residues were >4 Å distance from pS4. For cluster 3 (**Figure 4 C**), pS4-R6 and pS4-K15 were <4 Å suggesting the presence of

salt-bridges. Again, the presence of these salt-bridges would limit the observed ECD fragmentation.

The structures for conformer with CCS 337 $Å^2$ of R6pS4 (R+ K15+ POx0) are shown in **Figure 4 D**, **E** and **F**.[35] The cluster populations are ~17%, ~16% and ~7% of the 6351 total filtered structures respectively. The structures were again analysed according to the proximity between the basic amino acid chains and the phosphate group. For these structures, we are interrogating the presence (or otherwise) of ionic hydrogen bonds rather than salt-bridges, however the limit of 4 Å was retained for consistency. For clusters 1 (**Figure 4 D**) and 3 (**Figure 4 F**), the distance between pS4 and R6 was <4 Å suggesting the presence of an ionic hydrogen bond. In both cases, the distances between pS4 and the lysine residues was >4 Å. Cluster 2 (**Figure 4 E**) had no interaction distances <4 Å. Based on these models, the ECD fragmentation would be predicted to be extensive even if the pS4-R6 ionic hydrogen bond were maintained throughout the ECD event. For this conformer, therefore, the salt-bridge model is the better descriptor of experimental findings.

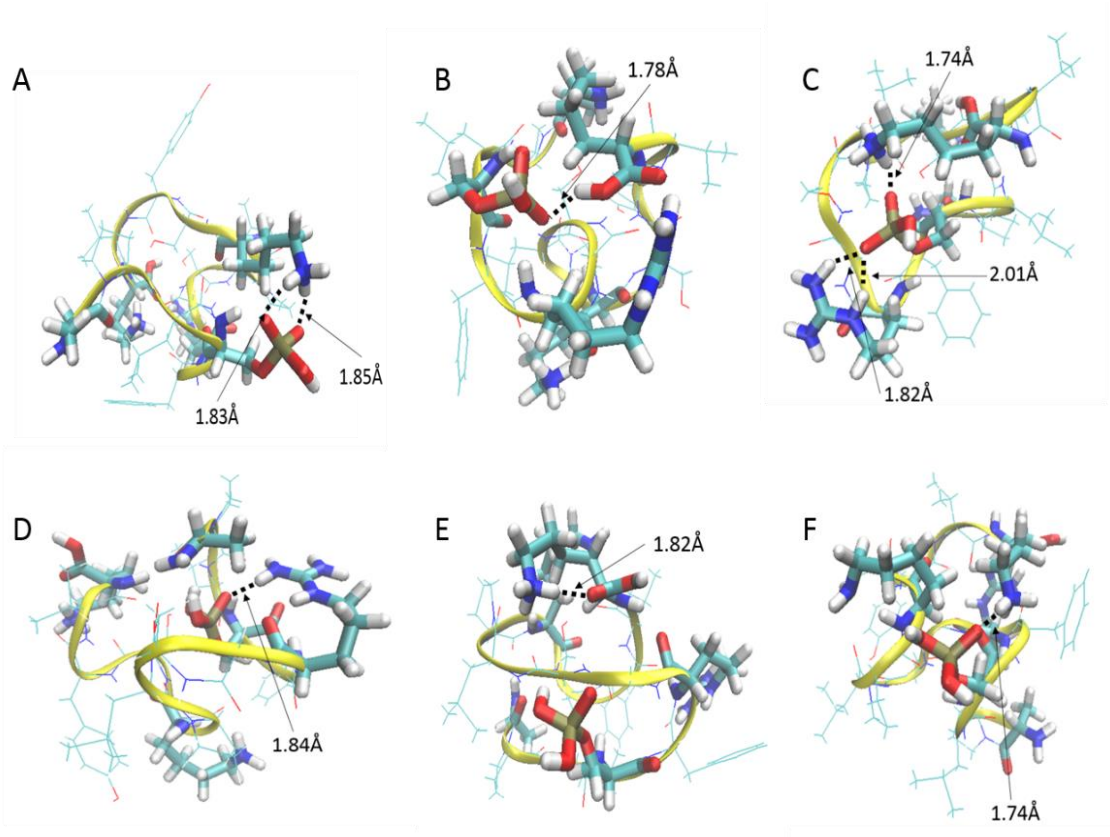

Figure 4. Model structures for the 337 Å2 conformer of R6pS4 (APLpSFRGSLPKSYVK) representative of the three most populated clusters (A, B, C) for (R+, K11+, K15+, POx–) and (D, E, F) (R+, K15+, POx0)

Table 1**.** Distance analysis of model structures of R6pSX and L6pSX peptides. 'Y' indicates distance of < 4Å from phosphoserine.

	R6	**K11**	**K15**		**K11**	**K15**
R6pS4(R+,K11+,K15+, POx-)				**L6pS4**(N+,K11+,K15+, POx-)		
337 Å^2 Structure 1	-	Y	-	336 Å^2 Structure 1	Y	Y
337 Å^2 Structure 2	-	-	-	336 Å^2 Structure 2	-	Y
337 Å^2 Structure 3	Y	-	Y	336 Å^2 Structure 3	-	-
R6pS4 (R+, K15+,POx0)				**L6pS4** (K11+, K15+,POx0)		
337 Å^2 Structure 1	Y	-	-	336 Å^2 Structure 1	-	-
337 Å^2 Structure 2	-	-	-	336 Å^2 Structure 2	Y	-
337 Å^2 Structure 3	Y	-	-	336 Å^2 Structure 3	Y	-
R6pS4(R+,K11+,K15+, POx-)				**L6pS4**(N+,K11+,K15+, POx-)		
351 Å^2 Structure 1	Y	-	-	346 Å^2 Structure 1	Y	-
351 Å^2 Structure 2	-	Y	Y	346 Å^2 Structure 2	-	Y
351 Å^2 Structure 3	-	-	Y	346 Å^2 Structure 3	-	Y
R6pS4 (R+, K15+,POx0)				**L6pS4** (K11+, K15+,POx0)		
351 Å^2 Structure 1	-	Y	-	346 Å^2 Structure 1	-	-
351 Å^2 Structure 2	Y	-	-	346 Å^2 Structure 2	Y	-
351 Å^2 Structure 3	-	-	-	346 Å^2 Structure 3	-	-
R6pS12 (R+,K11+,K15+, POx-)				**L6pS12** (N+,K11+,K15+, POx-)		
340 Å^2 Structure 1	Y	Y	-	372 Å^2 Structure 1	Y	-
340 Å^2 Structure 2	-	Y	Y	372 Å^2 Structure 2		- -

Table 1: continued

	R6	**K11**	**K15**		**K11**	**K15**
340 Å^2 Structure 3	Y	Y	-	372 Å^2 Structure 3	Y	-
R6pS12 (R+, K15+,POx0)				**L6pS12** (K11+, K15+,POx0)		
340 Å^2 Structure 1	-	-	-	372 Å^2 Structure 1	-	-
340 Å^2 Structure 2	-	Y	-	372 Å^2 Structure 2	Y	-
340 Å^2 Structure 3	-	-	Y	372 Å^2 Structure 3	-	-
R6pS12 (R+,K11+,K15+, POx-)						
352 Å^2 Structure 1	-	Y	Y			
352 Å^2 Structure 2	-	Y	Y			
352 Å^2 Structure 3	-	Y	Y			
R6pS12 (R+, K15+,POx0)						
352 Å^2 Structure 1	-	-	Y			
352 Å^2 Structure 2	Y	-	-			
352 Å^2 Structure 3	-	-	Y			
R6pS12 (R+,K11+,K15+, POx-)						
364 Å^2 Structure 1	-	-	Y			
364 Å^2 Structure 2	-	Y	-			
364 Å^2 Structure 3	-	Y	Y			
R6pS12 (R+, K15+,POx0)						
364 Å^2 Structure 1	Y	-	Y			
364 Å^2 Structure 2	Y	-	-			
364 Å^2 Structure 3	Y	-	-			

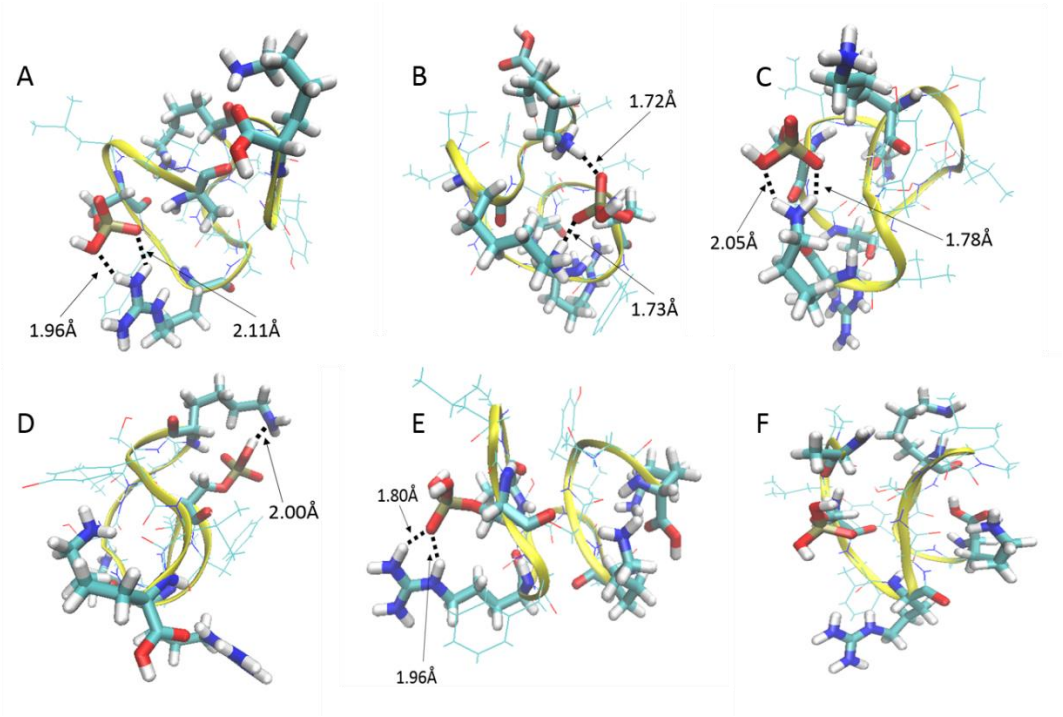

Figure 5. Model structures for the 351 Å2 conformer of R6pS4 (APLpSFRGSLPKSYVK) representative of the three most populated clusters (A, B, C) for (R+, K11+, K15+, POx–) and (D, E, F) (R+, K15+, POx0)

Modelled structures for the second conformer of R6pS4 (351 $Å^2$) are shown in **Figure 5**. The structures for (R6+, K11+, K15+, POx-) are shown in **Figure 5 A, B,** and **C**, with populations ~21%, ~8% and ~7% of the total filtered structures (16758) respectively. Distance analysis revealed that for cluster 1 (**Figure 5 A**) a salt-bridge exists between pS4 and R6 only, and that K11 was buried in the peptide backbone. Based on this, extensive ECD fragmentation would be predicted contrary to what is observed experimentally. The remaining structures show salt-bridges between pS4-K11 and pS4-K15 (**Figure 5 B**) or pS4-K15 (**Figure 5 C**), both of which would lead to reduced ECD fragmentation. The structures for (R6+, K15+, POx0) are shown in **Figure 5 D, E** and **F**, with populations ~10%, ~7% and ~6% respectively. Distance analysis reveals ionic hydrogen bonds between pS4-R6 in structure 2 (**Figure 5 E**) only. The distance between pS4 and K11 in structure 1 is < 4 Å, however, this interaction would be a (non-ionic) hydrogen bond and would be unlikely to survive ECD. The modelled structures for (R6+, K15+, POx0) would lead one to predict extensive ECD, however that is not observed experimentally. Overall, despite the outlier that is (R6+, K11+, K15+, POx-) cluster 1, the salt-bridge structures again appear better descriptors of experimental behaviour.

It is interesting to note that for each of the structures for R6pSX in which the phosphate is negative, interactions between the lysine side chain and phosphate residues are favoured, whereas when the phosphate is neutral, interactions between the arginine side-chain and the phosphate predominate. Arginine is known to confer greater stability to proteins than lysine as a result of the guanidinium group's ability to interact in three directions compared with the single direction available to the lysine's ammonium group

[49,50], allowing arginine to participate in a greater number of electrostatic interactions. It has been suggested that arginine electrostatic interactions in the condensed phase are more stable than those of lysine due to the higher pKa of arginine[50]. Conversely, Jungwirth has shown that, unlike the lysine ammonium group, the diffuse guanidinium group will pair with other guanidinium groups in water[51]. Moet-Ner[52] has shown that ionic hydrogen bond strength depends on the relative proton affinity of the interacting partners. A donor with higher relative proton affinity will transfer a proton less efficiently to the acceptor. The intrinsic PA of arginine is reported to be between 242.8 and 245.2 kcal/mol and that of lysine is between 225.5 and 230.3 kcal/mol[53] although the actual PA will depend on peptide ion structure .

The mobility data for R6pS12 suggests three conformers of CCS 340 $Å^2$, 352 $Å^2$, and 364 $Å^2$.[35] The modelled structures for the three conformers are shown in **Figure 6, 7,** and **8.** Consider the structures for conformer of CCS 340 $Å^2$ (R6+,K11+,K15+,POx-) (**Figure 6 A, B, C**). The populations of the clusters are ~15%, ~10% and ~8% of the 4943 total filtered structures respectively. Distance analysis reveals salt bridges between R6-pS12 and K11-pS12 in the structures shown in **Figure 6 A** and **C**, which correlates very well with the observed ECD patterns. The structure shown in **Figure 6 B** has salt-bridges between K11-pS12 and pS12-K15, for which we might expect some fragmentation central to the peptide sequence but which is not observed. For each of these structures, the phosphate group is participating in multiple electrostatic interactions. Similar behaviour, i.e., multidentate binding, was observed by Breuker and co-workers in their work on structural evolution of cytochrome C ions during electrospray [54]. Distance analysis of the

structures for the same conformer in protonation state (R6+,K15+,POx0) (see **Figure 6 D**, **E**, **F**; populations ~13%, ~8%, and ~8% of the 9308 total filtered structures) suggest (non-ionic) hydrogen bonds between K11 and pS12 in the structure shown in **Figure 6 E**, and an ionic hydrogen bond between pS12-K15 in the structure shown in **Figure 6 F**, none of which correlate with the observed ECD fragmentation. For this conformer, therefore, the salt-bridge structures are the best models to explain the ECD behaviour. Nevertheless, this conformer is the minor of the three and it is likely that conformers with CCS 352 $Å^2$ and 364 $Å^2$ will make a bigger contribution to observed ECD patterns.

For the conformer with CCS 352 $Å^2$ (**Figure 7**), distance analysis of the structures in protonation state (R6+,K11+,K15+,POx-) reveals salt-bridges between K11-pS12 and pS12-K15 in each case. (Cluster populations are ~11%, ~8% and ~8% of 15833 total filtered structures). None showed salt-bridges between R6-pS12. In each structure the arginine protrudes away from the phosphate group, similar to the observations of Tureček and co-workers [55]. Clearly, these models cannot explain observed ECD behaviour. Two out of three structures in protonation state (R6+,K15+,POx0) are also at odds with experimental observations. Those structures shown in **Figure 7 D** and **7 F** show ionic hydrogen bonds between pS12-K15. (Cluster populations are ~15%, ~13% and ~8% of 3595 total structures). For this conformer neither set of model structures accurately describes the ECD. It is worth noting that in this work we have only considered two protonation states and there may be third (or more) which account for this conformer. Nevertheless, as mentioned above, the ECD of doubly-protonated ions of N-acetylated R6pS12 suggests that protonation of the N-terminus does not play a significant role.

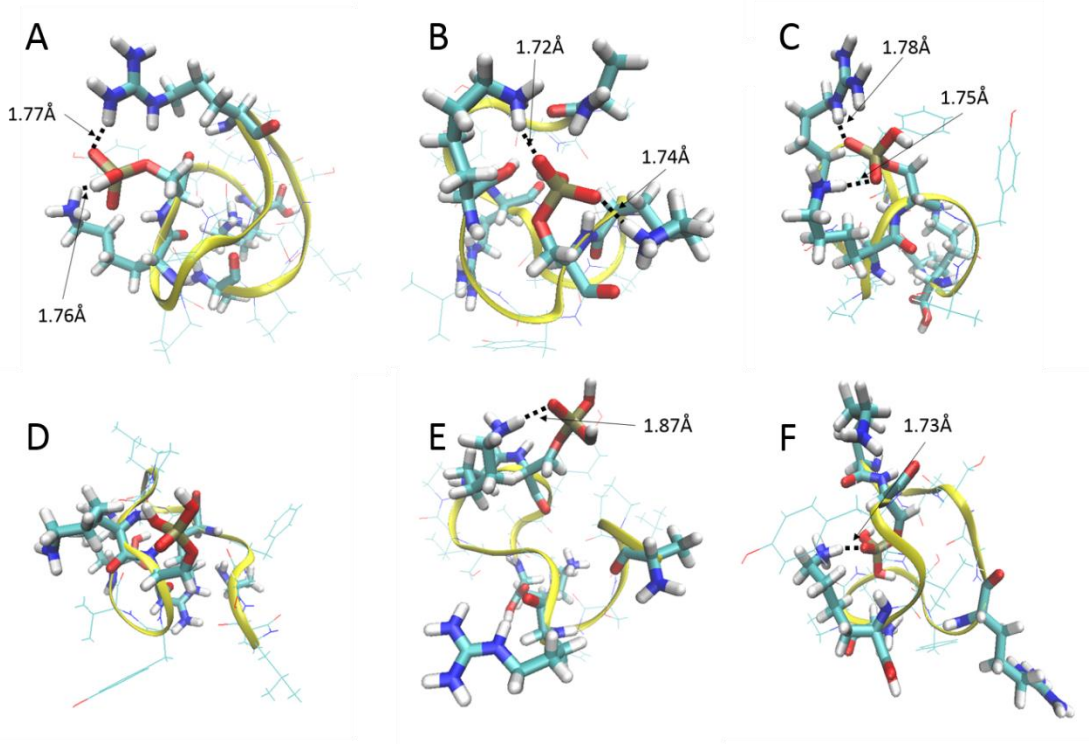

Figure 6. Model structures for the 340 Å2 conformer of R6pS12 (APLSFRGSLPKpSYVK) representative of the three most populated clusters (A, B, C) for (R+, K11+, K15+, POx–) and (D, E, F) (R+, K15+, POx0)

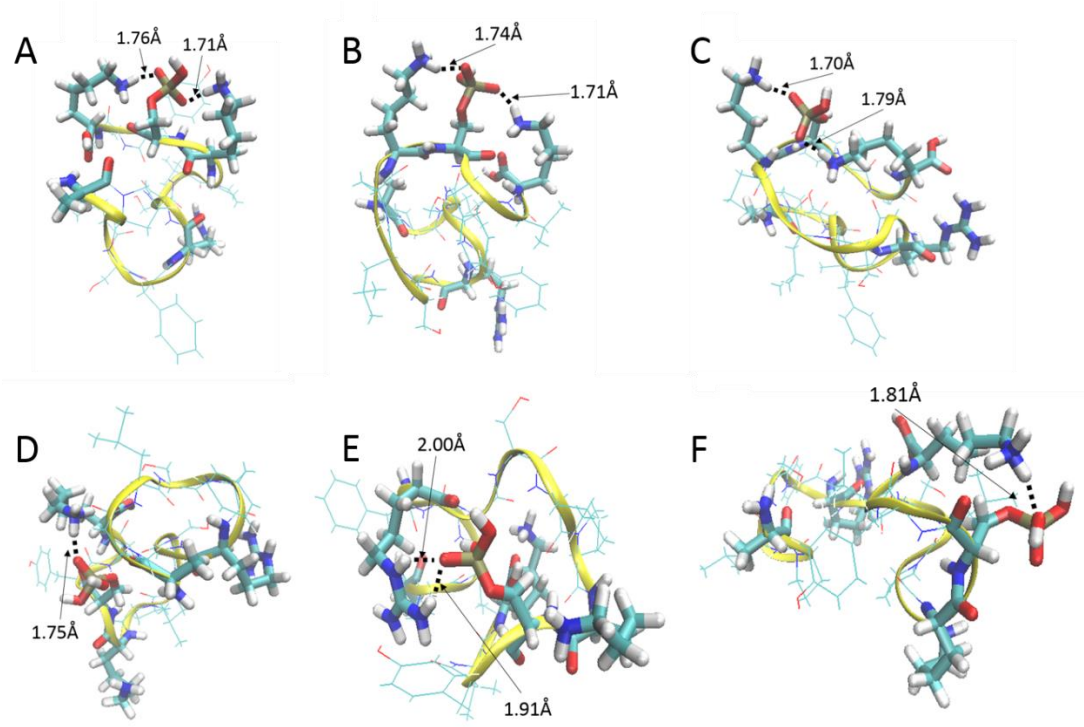

Figure 7. Model structures for the 352 Å2 conformer of R6pS12 (APLSFRGSLPKpSYVK) representative of the three most populated clusters (A, B, C) for (R+, K11+, K15+, POx–) and (D, E, F) (R+, K15+, POx0)

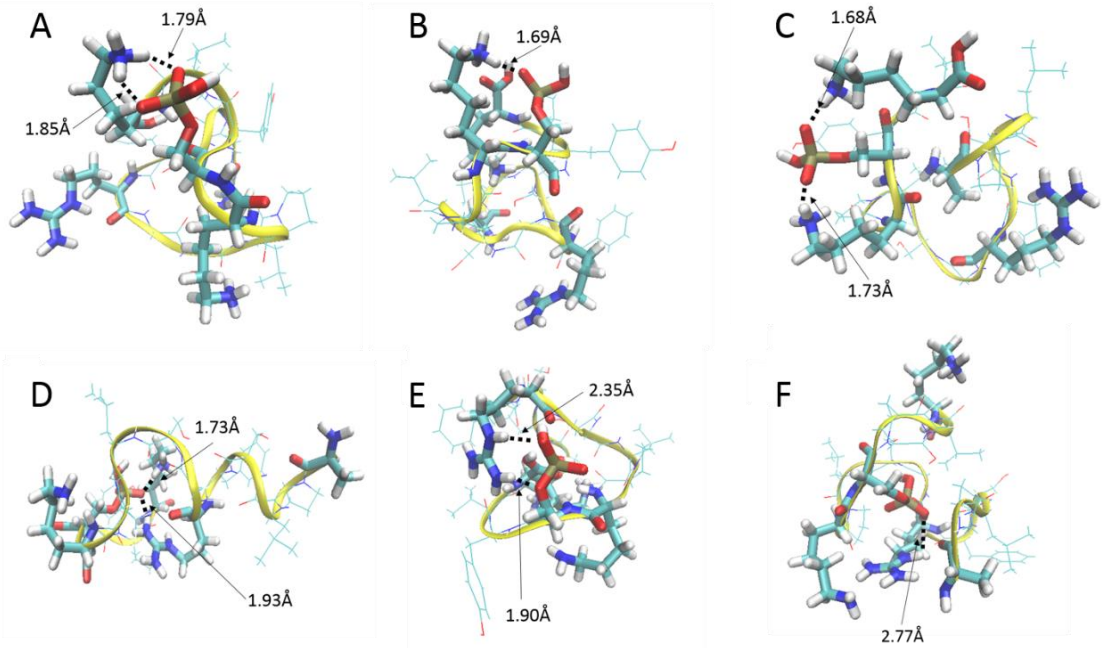

Figure 8. Model structures for the S 364 Å2 conformer of R6pS12 (APLSFRGSLPKpSYVK) representative of the three most populated clusters (A, B, C) for (R+, K11+, K15+, POx–) and (D, E, F) (R+, K15+, POx0)

Distance analysis of the structures for the third conformer (CCS 364 Å^2) (**Figure 8**) in protonation state (R6+,K11+,K15+,POx-; total population of 24583) reveals salt-bridges between pS12-K15 (**Figure 8 A**), population ~10%), K11-pS12 (**Figure 8 B**, population ~7%) or both pS12-K15 and K11-pS12 (**Figure 8 C**, population ~5%). Again in each case, the arginine residue protrudes away from the phosphate group. In contrast, all of the structures (cluster populations ~11%, ~6%, ~5% of total 1249 filtered structures) in protonation state (R6+,K15+,POx0) show ionic hydrogen bonds between R6-pS12 (with an additional ionic hydrogen bond between pS12-K15 in the structure shown in **Figure 8 D**). That is, for this conformer, the structures with neutral phosphate and ionic hydrogen bonds are the most appropriate models for observed ECD fragmentation.

The mobility data for L6pS4 suggests the presence of two conformers of CCS 336 Å^2 and 346 Å^2.[35] The ECD fragmentation observed for this peptide is more extensive than for R6pS4 but less extensive than for L6(unmod), suggesting that some intramolecular interactions involving the phosphate group may be present but are insufficiently strong to completely prevent separation of fragments. The mobility data suggest compaction of structure on phosphorylation. Model structures were generated as described above. The structures for the conformer with CCS 336 Å^2 and protonation state (N+,K11+,K15+,POx-) are shown in **Figure 9 A, B** and **C**. Cluster populations are ~27%, ~15% and ~14% of the total population of 521 filtered structures respectively. Salt-bridges are present in the structures between both pS4-K11 and pS4-K15 (**Figure 9 A**), and pS4-K15 (**Figure 9 B**). That is, over 40% of the structures display salt-bridges. The third structure has no salt-bridges. In the (K11+,K15+,POx0) protonation state (**Figure 9 D, E, F**) (total population

6437 filtered structures), the structures shown in **Figure 9 E** and **Figure 9 F** (~8% and ~8%) show ionic hydrogen bonds between pS4 and K11, and the structure shown in **Figure 9 D** (~13%) shows no interactions between the phosphate and lysine side-chains of < 4 Å.

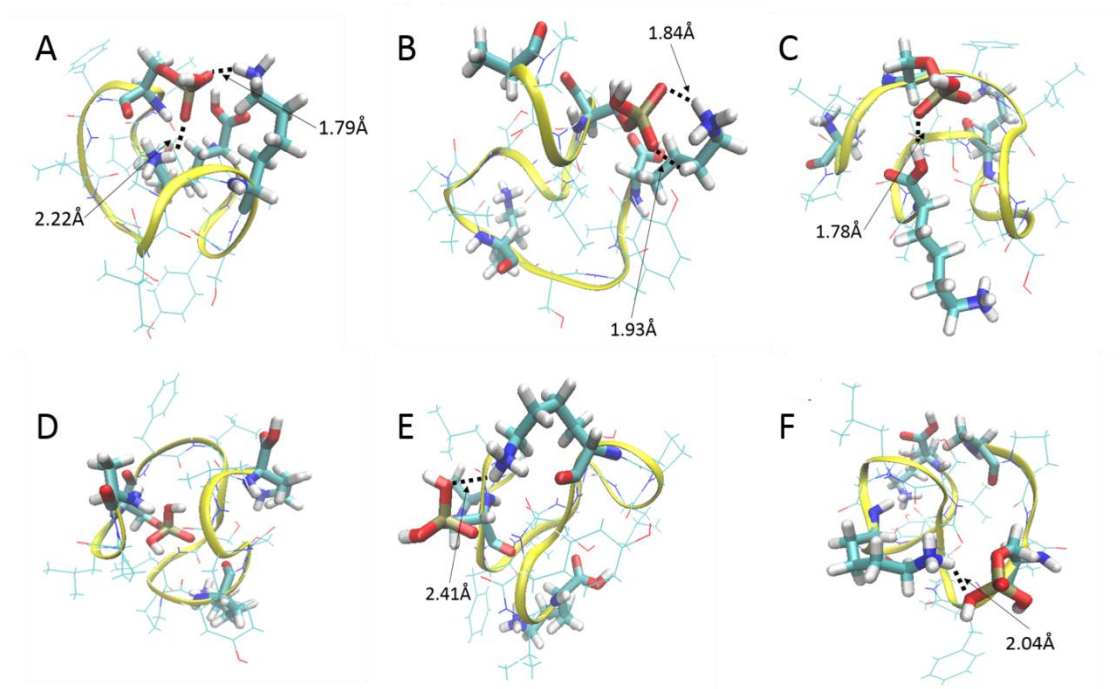

Figure 9. Model structures for the S 336 Å2 conformer of L6pS4 (APLpSFLGSLPKSYVK) representative of the three most populated clusters (A, B, C) for (N+, K11+, K15+, POx–) and (D, E, F) (K11+, K15+, POx0)

Structures for the more abundant conformer (CCS 346 $Å^2$) are shown in **Figure 10**. In protonation state (N+,K11+,K15+,POx-) (total population of 12280 filtered structures), salt-bridges exist between pS4-K11 (**Figure 10 A**, ~9%) and pS4-K15 (**Figure 10 B** & **C**, ~8% and ~6% respectively). In protonation state (K11+,K15+,POx0) (total population 17342 structures), there are no interactions between the phosphate and K11 or K15 in the structures shown in **Figure 10 D**, (~31%) and **Figure 10 F** (~8%) and an ionic hydrogen bond present between pS4 and K11 in the structure shown in **Figure 10 E** (~22%).

The mobility data for L6pS12 suggests one major conformer of CCS 372 $Å^2$ and one minor of CCS 345 $Å^2$.[35] The major conformer, which it is assumed makes the major contribution to ECD, shows no compaction on phosphorylation suggesting no structurally significant intramolecular interactions involving the phosphate group, correlating well with the ECD data. The modelled structures for the conformer with CCS 372 $Å^2$ in both protonation states are shown in **Figure 11**. These structures did not form highly populated clusters (~8%, ~7% and ~5% for (N+,K11+,K15+,POx-) (total population 11759 structures) and ~5%, ~4% and ~4% for (K11+,K15+,POx0) (total population 18230 structures)). For protonation state (N+,K11+,K15+,POx-), salt-bridges were present between K11-pS12 (**Figure 11 A**) and both K11-pS12 and pS12-K11 (**Figure 11 C**). (The interaction indicated in **Figure 11 B** is between pS12 and the N-terminus). For protonation state (K11+,K15+,POx0) ionic hydrogen bonds were present between K11-pS12 (**Figure 11 E**). Any of these model structures could explain the observed ECD fragmentation.

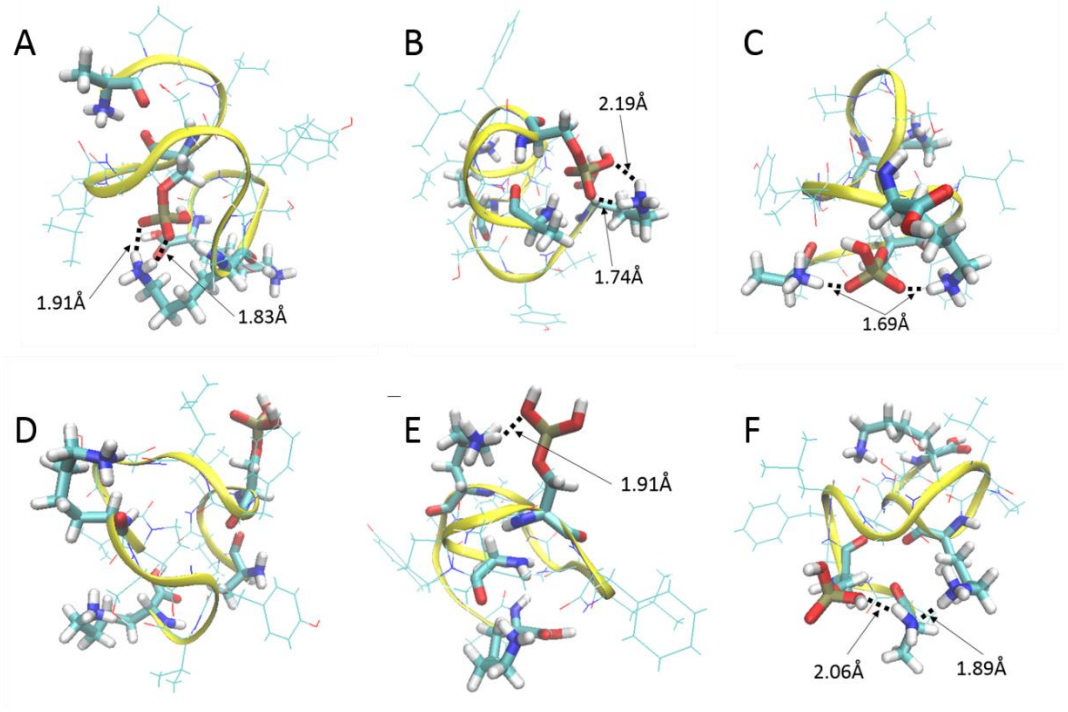

Figure 10. Model structures for the S 346 Å2 conformer of L6pS4 (APLpSFLGSLPKSYVK) representative of the three most populated clusters (A, B, C) for (N+, K11+, K15+, POx–) and (D, E, F) (K11+, K15+, POx0)

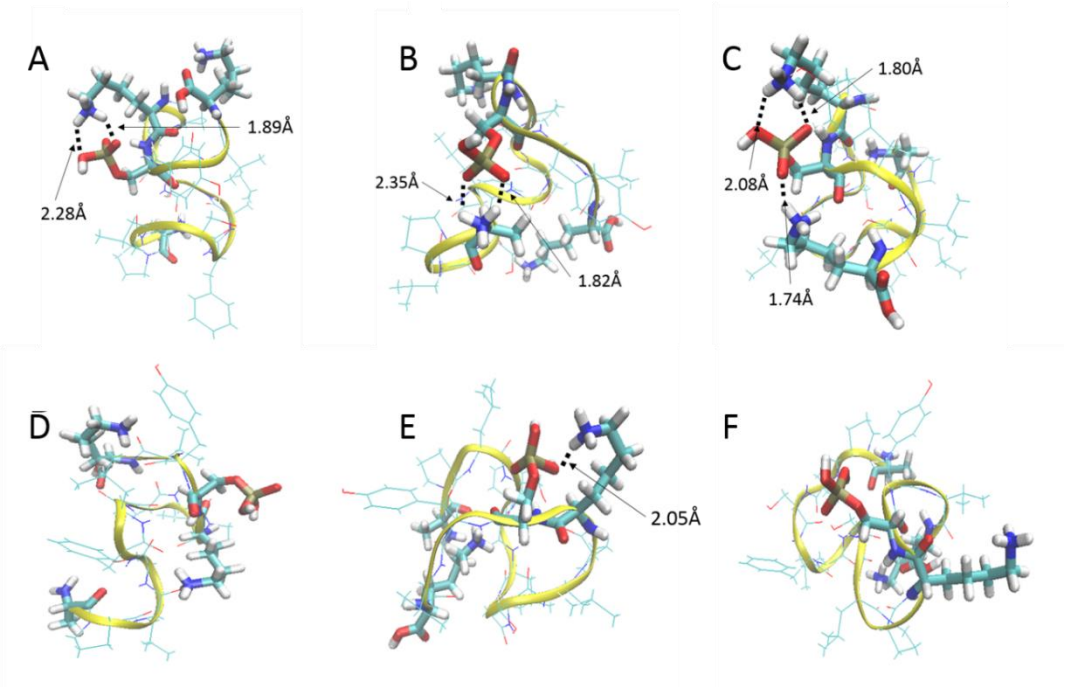

Figure 11. Model structures for the S 372 Å2 conformer of L6pS4 (APLSFLGSLPKpSYVK) representative of the three most populated clusters (A, B, C) for (N+, K11+, K15+, POx–) and (D, E, F) (K11+, K15+, POx0)

In summary, for R6pS4, the model structures in which salt-bridges exist between pS4 (deprotonated) and K11 best fit the ECD data. For R6pS12, the model structures in which ionic hydrogen bonds exist between R6 and pS12 are the best fit. The question arises: why do we see a difference in propensity to form salt-bridges in these isomeric phosphopeptides? Williams et al.[56] have shown that the presence of multiple (adjacent)

arginine residues stabilizes the salt-bridge structures in dipeptide cations. It is possible, therefore, that the presence of arginine at *i+2* from pS stabilises the salt-bridge between pS4 and K11 in R6pS4. Salt-bridge (pS-K) stabilization by arginine would also explain the results for L6pS4: The more extensive fragmentation observed for L6pS4 suggests that any intramolecular interactions are less likely to survive electron capture, i.e., that no salt-bridges exist. Nevertheless, ion mobility data does suggest compaction on phosphorylation of the leucine containing peptide, which may simply reflect the strength of pS-K interactions in comparison to pS-R interactions.

3+ Phosphopeptide

The mobility profile displayed in **Figure 12** shows that the 3+ R series peptides were more elongated than 2+ ions. This elongation can be explained by increased coulombic repulsion due to increased charge state. These R series peptides have a total of four protons on all of the basic residues including N-terminus, and to give a net charge of 3+ the phosphate group is deprotonated. For the R6pS4 mobility profile shown in **Figure 12**, CCS filtered and clustered results of R6pS4 are shown in **Figure 13**. Panel A-C correspond to the best fit structures of the three most populated clusters belonging to the mobility peak of CCS 362 $Å^2$ (12.1%, 11.4%, 8.9% of total 15170), and panel D-E correspond to clusters belonging to the CCS peak of 382 $Å^2$ (11.5%, 9.3%, 5.9% of total 22174). All six best-fit structures show salt-bridge interaction between the phosphate group and a protonated basic residue. However, it is apparent that the number of coordination between the basic residues and the phosphate group decreases as the larger conformation was modeled. The compact conformation 362 $Å^2$ cluster shows that both

Lys and Arg participate in the charge solvation interaction with the phosphate group. However, for the 382 $Å^2$ clusters, all of the structures show that the Arg sidechain is extended from the core of the peptide, while ammonium group of Lys sidechains are coordinating with the phosphate group or the backbone carbonyls. This can be rationalized by the difference in charge characteristics between ammonium group and guanidinium group, where guanidinium group is a delocalized charge group, whereas ammonium group behaves like a point charge.

The mobility data for R6pS8 (**Figure 12**) suggests that there are two conformers of CCS of 370 $Å^2$ and 391 $Å^2$, and CCS filtered and clustered for these conformers are shown in **Figure 14** (A 13.3%, B 9.0%, and C 5.8% of 14259 structures and D 12.1%, 7.0%, and 6.7% of 11563 structures). The best fit structures for the major peak at 370 $Å^2$ shows that all three basic residues are participating in salt-bridge formation shown in **Figure 14** A-C. The structure shown in E exhibits extension of guanidinium ion as the R6pS4 extended structures. However, the two structures for the extended conformer of R6pS8, shown in **Figure 14 D** and **F**, show that the guanidinium ion is coordinating with the phosphate group which was only shown once for the compact conformer. In addition, structure shown in **Figure 14 F**, the Lys sidechain is extended from the core of the peptide.

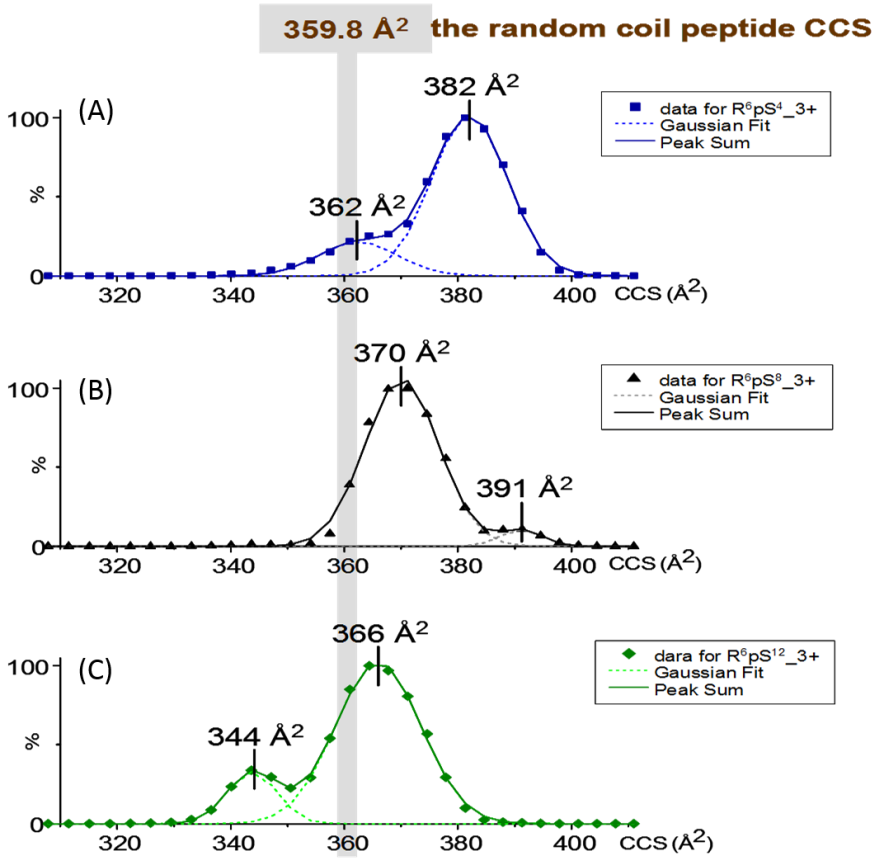

Figure 12. Mobility profile of R6pSX3+ ions are shown above. R6pS4 is shown in (A), R6pS8 is shown in (B), and R6pS12 is shown in (C). Mobility profile shows two peaks for all three peptides.

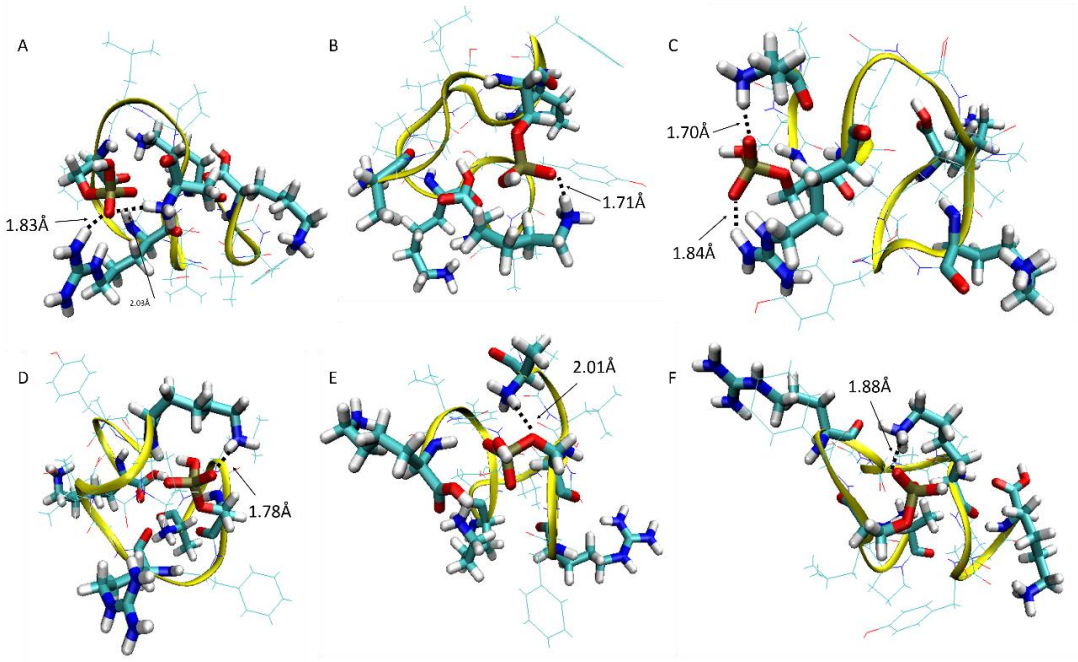

Figure 13. Model structures for the 362 $Å^2$ conformer of $R6pS4^{3+}$ (APLpSFRGSLPKSYVK) representative of the three most populated clusters (A, B, C) and 382 $Å^2$ conformer of $R6pS4^{3+}$ (D, E, F).

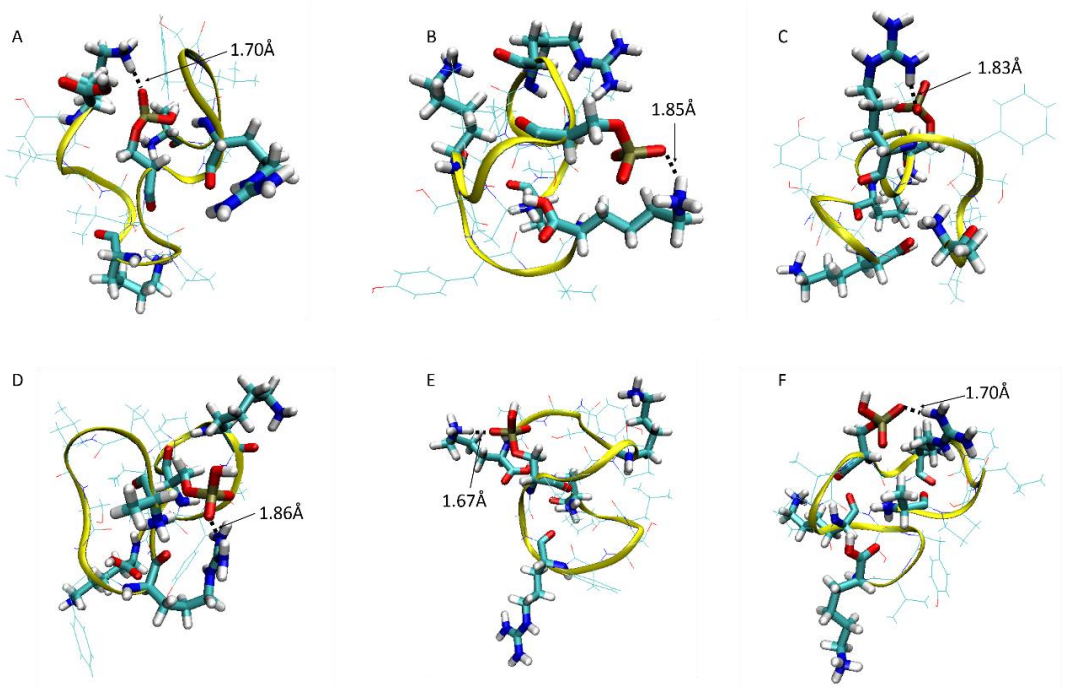

Figure 14. Model structures for the 370 Å2 conformer of R6pS83+ (APLSFRGpSLPKSYVK) representative of the three most populated clusters (A, B, C) and 391 Å2 conformer of R6pS83+ (D, E, F).

The mobility profile data for R6pS12 (**Figure 12**) suggests that there are two conformers, 344 $Å^2$ and 366 $Å^2$. The mobility profile of this peptide shows significant compaction compared to the other two peptides, R6pS4 and R6pS8. The most compact conformer, CCS of 344 $Å^2$, is more compact than the random coil peptides, and its cluster results are shown in **Figure 15** A-C (11.6%, 10.4%, and 10.0% of 1995 structures). The larger conformer, CCS of 366 $Å^2$, clustering result are shown in **Figure 15** D-F (6.5%, 6.1%, and 5.0% of 18461 structures). The smaller conformer shows two coordinations for the phosphate groups with basic residues. However, the larger conformer shows only one coordination of phosphate group for the structures shown in D and E. Although, the phosphate group is located in between two Lys11 and Lys15, the phosphate group shown in structures D and E is coordinating with the guanidinium ion R6. However, for the larger conformation, the phosphate group is not coordinating with the guanidinium ion but rather coordinating with both lysines. From this result, we can postulate that to achieve a compact conformation, the phosphate group has to coordinate with two protonated groups, which includes N-terminus, Lys, and Arg sidechain where the salt-bridge formation ultimately leads to compaction of the peptide.

The L series phosphopeptide simulations were fundamentally different from R series phosphopeptide simulation because L series peptides were simulated with a neutral phosphate group. Having only three sites of protonation, two Lys's and N-terminus, and a deprotonated phosphate group leads net charge of +2. ESI-IM-MS experiment (shown in **Figure 16)** have shown that the L series phosphopeptides were more elongated than the R series. First the L6pS4 3+ mobility profile shows two distinct conformations: the smaller

conformer at 382 $Å^2$ and the larger conformer at 397 $Å^2$. The clustering result for the smaller conformer is shown in **Figure 17** A-C (7.6%, 5.7%, and 4.8% of 19624 structure). The larger conformer clustering result is shown in **Figure 17** D-F (10.8%, 8.3%, and 7.1% of 7380 structures). The clustering results for both peaks show limited interaction between phosphate group and the basic groups and theses structures show extended lysine side chain. Unlike to the R series phosphopeptide, phosphate group also coordinates with the backbone carbonyl as shown in **Figure 17 B**.

The mobility profile for L6pS12 is shown in **Figure 16,** which suggests only one conformer at CCS of 382 $Å^2$. However, the peak is wider than other peaks which suggests that there are an ensemble of conformations leading to the broad peak at 382 $Å^2$. The cluster analysis results are shown in **Figure 18** (6.3%, 5.2%, and 4.8% of 13202 structures). The cluster results show that for all three cases, the K15 ammonium group is extended from the core of the peptide while the N-terminus is coordinating with the neutral phosphate group shown in **Figure 18 A** and **C**. K11 is shown to coordinate with the phosphate group once, shown in **Figure 18 B**.

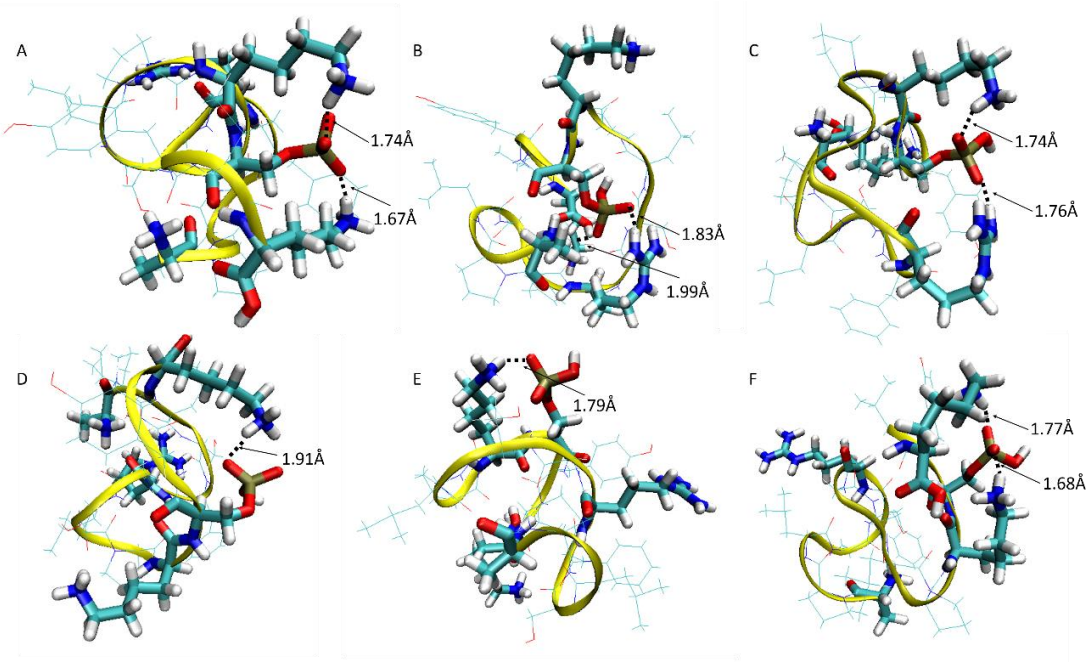

Figure 15. Model structures for the 344 Å2 conformer of R6pS123+ (APLSFRGSLPKpSYVK) representative of the three most populated clusters (A, B, C) and 366 Å2 conformer of R6pS123+ (D, E, F).

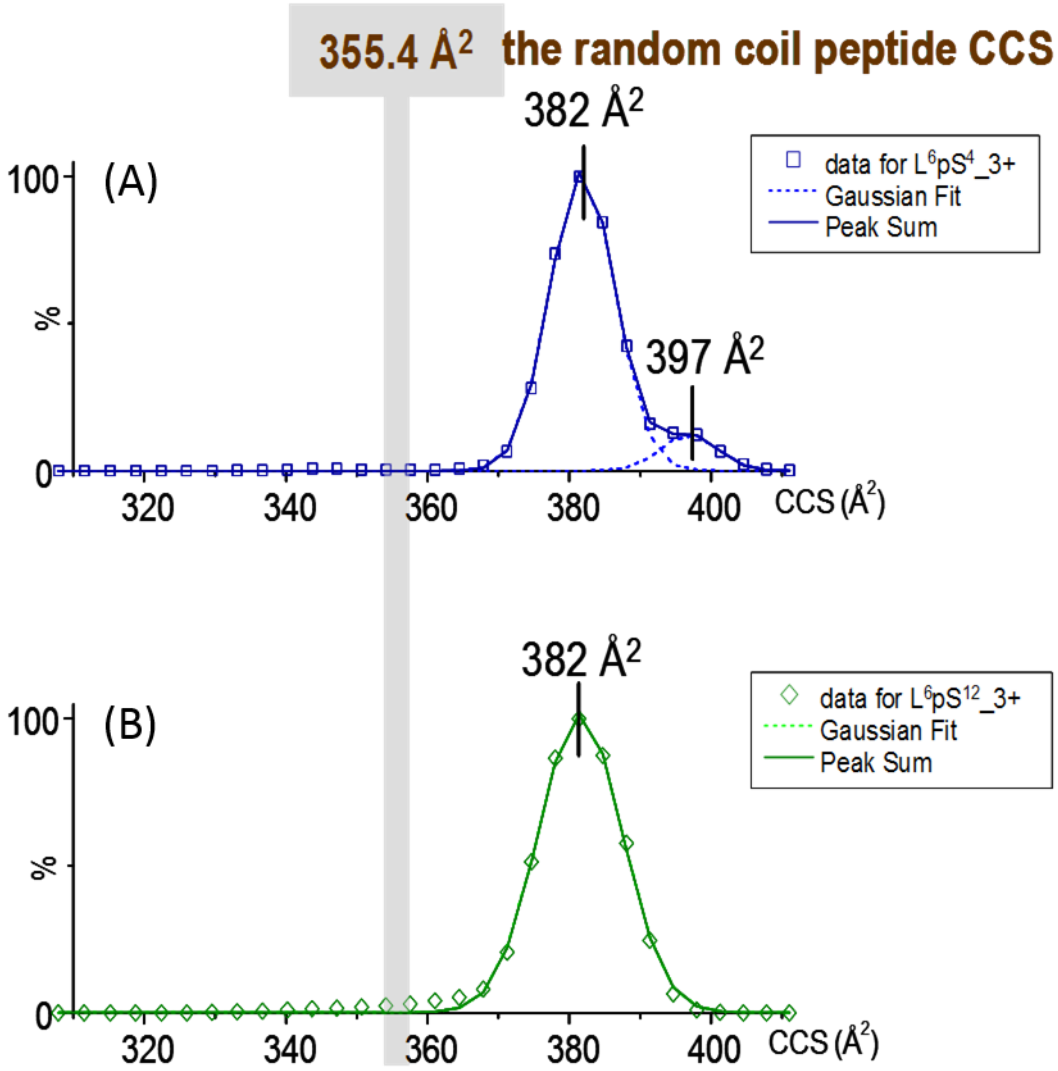

Figure 16. Mobility profile of L6pSX3+ ions are shown above. L6pS4 is shown in (A), and R6pS12 is shown in (B).

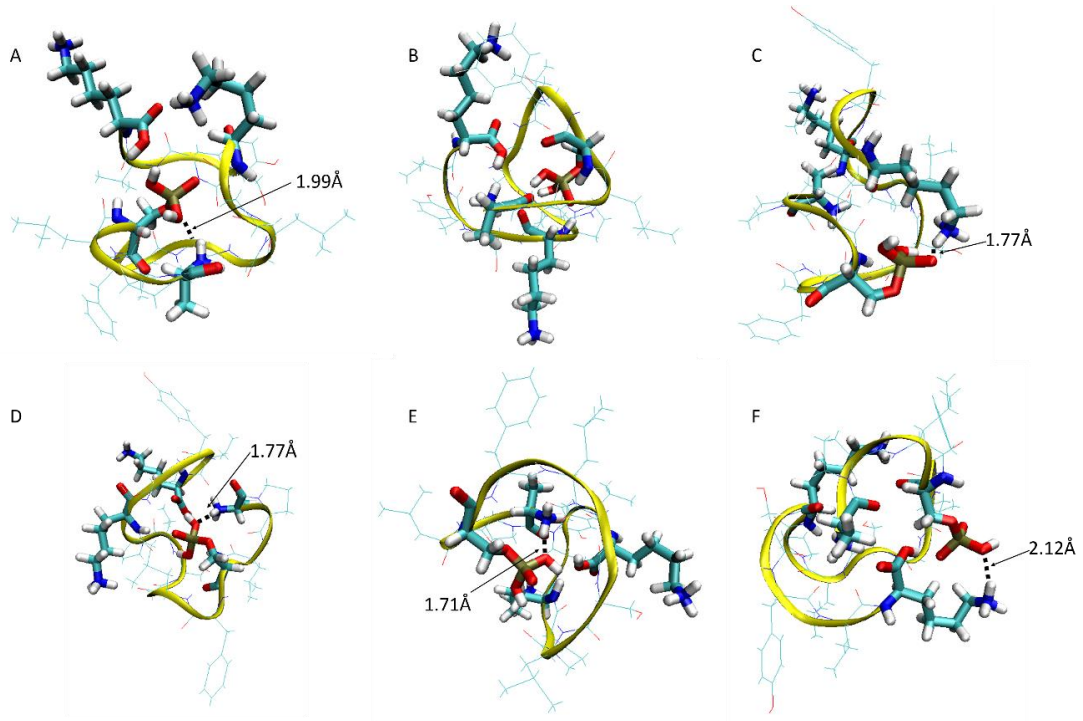

Figure 17. Model structures for the 370 Å2 conformer of L6pS43+ (APLpSFLGSLPKSYVK) representative of the three most populated clusters (A, B, C) and 370 Å2 conformer of L6pS43+ (D, E, F).

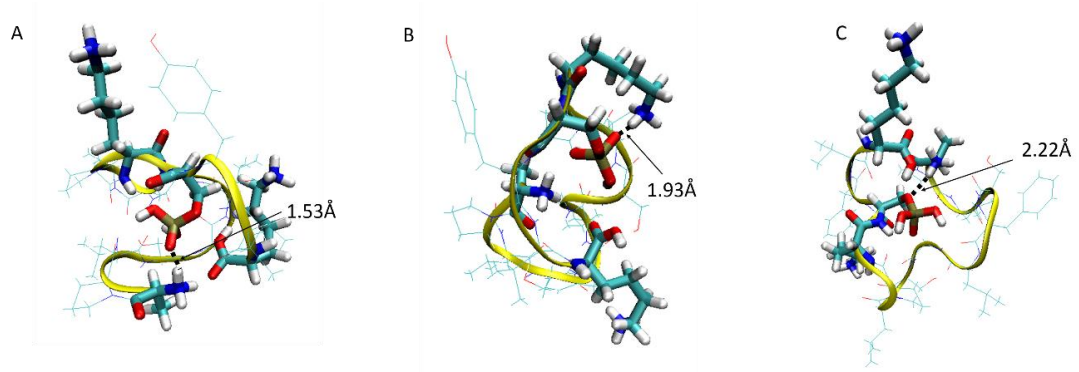

Figure 18. Model structures for the 370 Å2 conformer of L6pS123+ (APLSFLGSLPKpSYVK) representative of the three most populated clusters (A, B, C).

In conclusion, the 3+ phosphopeptide ion structures have shown more elongated structure than the 2+ ions, and this outcome can be explained by coulombic repulsion. However, the IM profile of the R series phosphopeptide has shown that smaller conformations of the series are on the random coil peptide CCS trend-line or even more compact than the random coil peptides. The L series peptide especially has shown dramatic difference because the charge state of the phosphate group have changed. The coordination of arginine varies by the peptide and conformation, where it can coordinate with phosphate or extend radially outward. The preference for Lys-phosphate interactions is attributed to the charge characteristics of lysine and arginine, specifically how Lys ammonium group acts as a point charge, whereas the charge on the guanidinium ion of Arg are delocalized. MDS indicates a strong non-covalent bond between the phosphate

group and protonated basic residue is essential to formation of compact phosphopeptide structures. This result is compared with similar observations for Arg and Lys containing peptides at membrane-aqueous interfaces.

Substance P and Its Mutants

Previous dehydration studies on Substance P (SP) 3+ ions (amino acid sequence RPKPQQFFGLM-NH_2) by Silveria et al. have identified the presence of two conformations, an extended conformation at 368 $Å^2$ and a compact conformation at 316 $Å^2$.[15] The experiment further indicated that the compact conformer ions were produced by the desolvation process; the compact conformer shows a hydration trendline and mild heating induced conversion to the extended conformer. However, the reverse transition was not observed. Initial mutant studies have shown that the stability of the compact conformation drops significantly when both the Q5 and Q6 hydrophilic residues are mutated to alanine resulted peptide almost exclusively exist in extended conformer.[15,57] It was concluded that Q5 and Q6 are involved in charge solvation after the desolvation and charge stabilization.

Furthermore, the cis-trans isomerization of proline residues were implicated in the conformational preference of SP^{3+}. When the prolines were mutated to alanine, the abundance of the compact conformer decreased; therefore, it was concluded that the prolines are also stabilizing factors for preserving compact conformations. The P2A mutant induced a split in the compact conformation mobility peak. From this information, the wild type SP compact conformation structure heterogeneity rises from the proline

conformations. However, when cis trans isomerization was induced to C-terminus via G9P mutation, the mobility peak for the extended conformer was split into two peaks, but the mutation does not affect the stability of the compact conformer. Herein, MDS are used to create a library of candidate structures considering cis and trans isomerization of proline in $[SP+3H]^{3+}$ ions.

The wild type SP and proline mutants' structures were simulated with using SA-MDS and the structures were filtered by CCS, which was experimentally measured, and cis-trans isomerization as described in Chapter II. The ~50000 structures simulated by SA-MDS are filtered by the experimental CCS values reported previously.[15,57] The conformation of proline in the starting structure of the simulation was set to trans conformer. The experiment by Fort et al. has revealed that the mobility peak splitting was due to cis-trans isomerization induced by prolines.[57] Peaks associated with proline mutation are filtered by cis-trans prolines conformation by measuring the distance between two atoms, $C\alpha_{i-1}$ - $C\alpha_{i.}$ The distance between the two Cαs for cis and trans proline isomers were calculated to be ~3.1 Å and ~3.9 Å, respectively. Monitoring these distances allows for monitoring the transition of two isomers as shown in **Figure 19**.

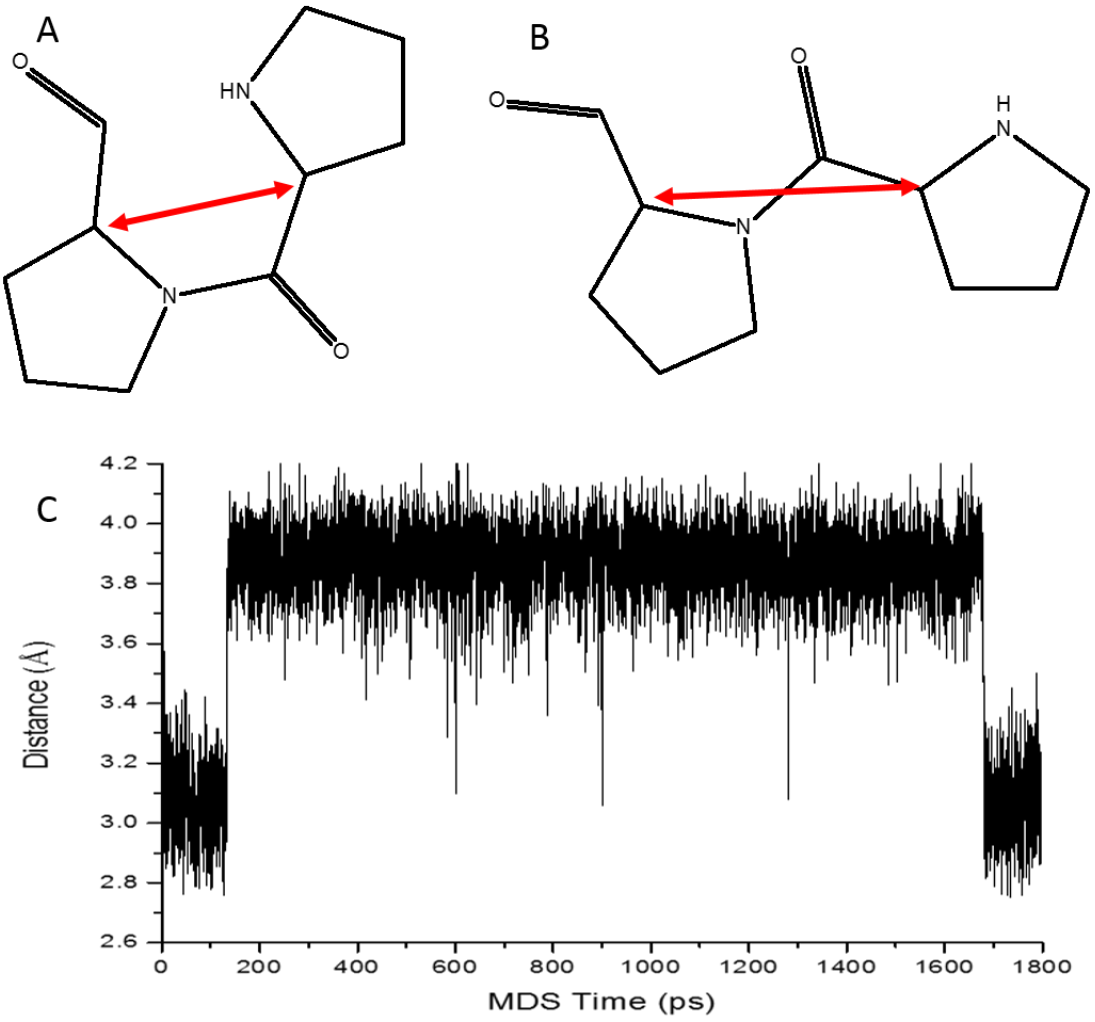

Figure 19. Distance plot and proline diagram between the two alpha carbons of adjacent proline residues, Cαi-1 - Cαi. The distance measured in cis conformation is shown in A), while B) represent the trans conformation. C) is the distance plot in a sample calculation where the distance between two adjacent αCs starts at around 3.1 Å (cis) then extends to 3.9 Å (trans) after ~200 ps. The proline of interest is dynamic as it interconverts between cis and trans isomers.

SP^{3+} wild type simulation results were filtered using experimental CCS values. These structures were further filtered based upon the cis-trans isomerization of Pro 2 and Pro 4. The filtered results are shown in **Table 2**. This results indicate that Pro 2 exists predominantly in the cis conformation (72% of filtered structures) whereas Pro 4 exists mostly in the trans conformation (71% of filtered structures). The best fit structures of the clustering algorithm are displayed in **Figure 20**. **Figure 20 A** shows the structure of all Pro 2/Pro 4 cis conformers of 316 $Å^2$ and **Figure 20** C indicates the structure of all Pro 2/ Pro 4 trans isomers of 316 $Å^2$. The clusters were highly populated compared to the phosphopeptides because the backbone variability was reduced by cis/trans proline separation. Simulation results suggest that the Pro 2 favors cis conformation, while Pro 4 favors trans conformation. All trans conformation denote that the guanidinium ion is extended and not interacting with other residues. Whereas, all cis conformation show that the guanidinium ion is showing intra-molecular charge solvation. The best-fit clusters for the extended conformers are shown in **Figure 20 B** and **D**, and total number of simulated extended conformer is 2289 out of ~50000 simulated structures. However, previous ESI-IM-MS study indicates that the extended conformer is the stable gas phase conformer, but the gas phase SA-MDS does not simulate abundant extended conformer.[15,57] Current simulation results are insufficient to conclude why the cis/trans proline conformation preferences vary between proline residues.

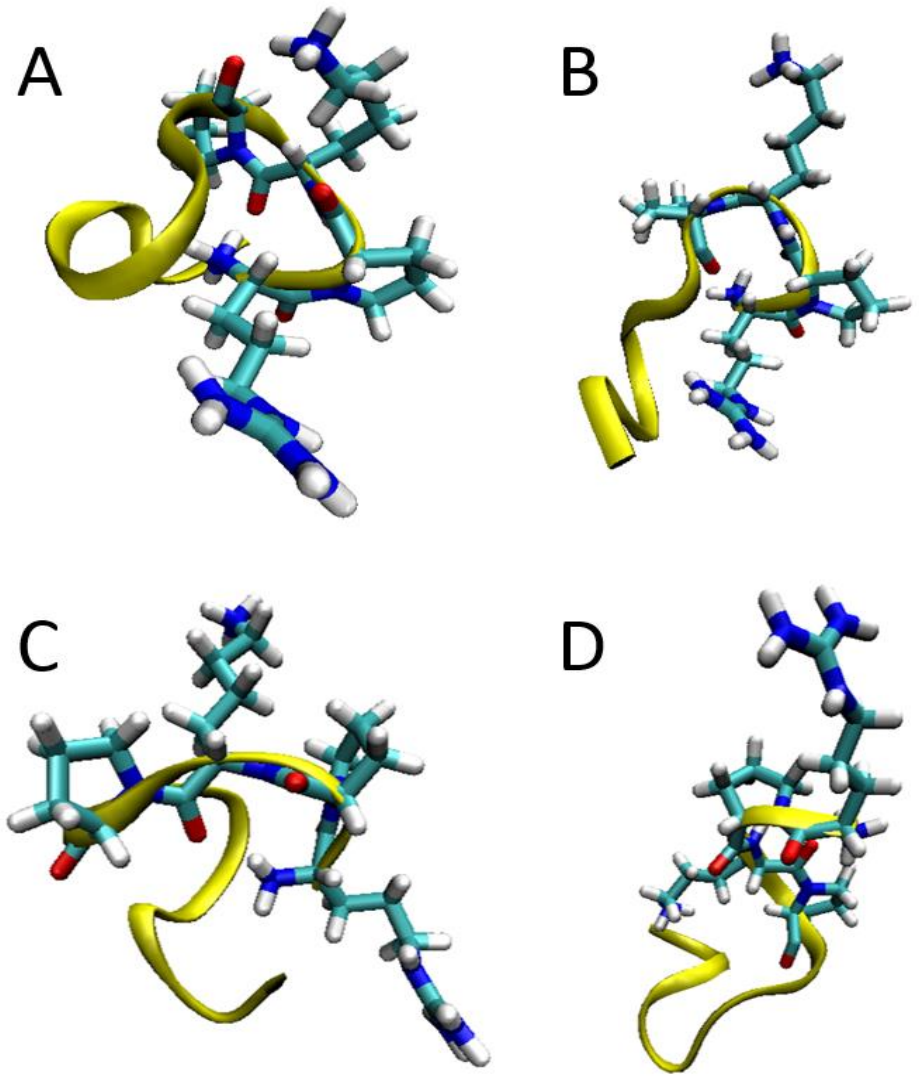

Figure 20. Best-fit structures for the largest population clusters of Substance P: 316 Å^2 all cis conformer (A, 32% of population), 363 Å^2 all cis conformer (B, 77%), 316 Å^2 all trans (C, 79%), and 363 Å^2 all trans (D, 21%). In B, C, and D, the arginine side chain is extended away from the peptide, while in A, it is towards the backbone. The guanidinium ion does not have specific coordination in the best-fit structures.

Table 2. Substance P and mutant SP simulated structure filter results. The structures are filtered by CCS values and then filtered by the proline conformations.

Substance P (RPKPQQFFGLM-NH_2)				G9P		P2A	
316$Å^2$	P2	P4	P2 and P4	352$Å^2$	P9	301$Å^2$	P4
Cis	13668	5548	3315	Cis	2410	cis	4018
trans	5368	13489	3135	Trans	3526	trans	5315
% cis	71.8	29.1	51.4	% cis	40.6	% cis	43.1
363$Å^2$	P2	P4	All	360$Å^2$	P9	312$Å^2$	P4
Cis	1551	663	468	Cis	897	cis	10868
trans	700	1589	505	Trans	1355	trans	13500
%cis	68.9	29.4	48.1	% cis	39.8	% cis	44.6

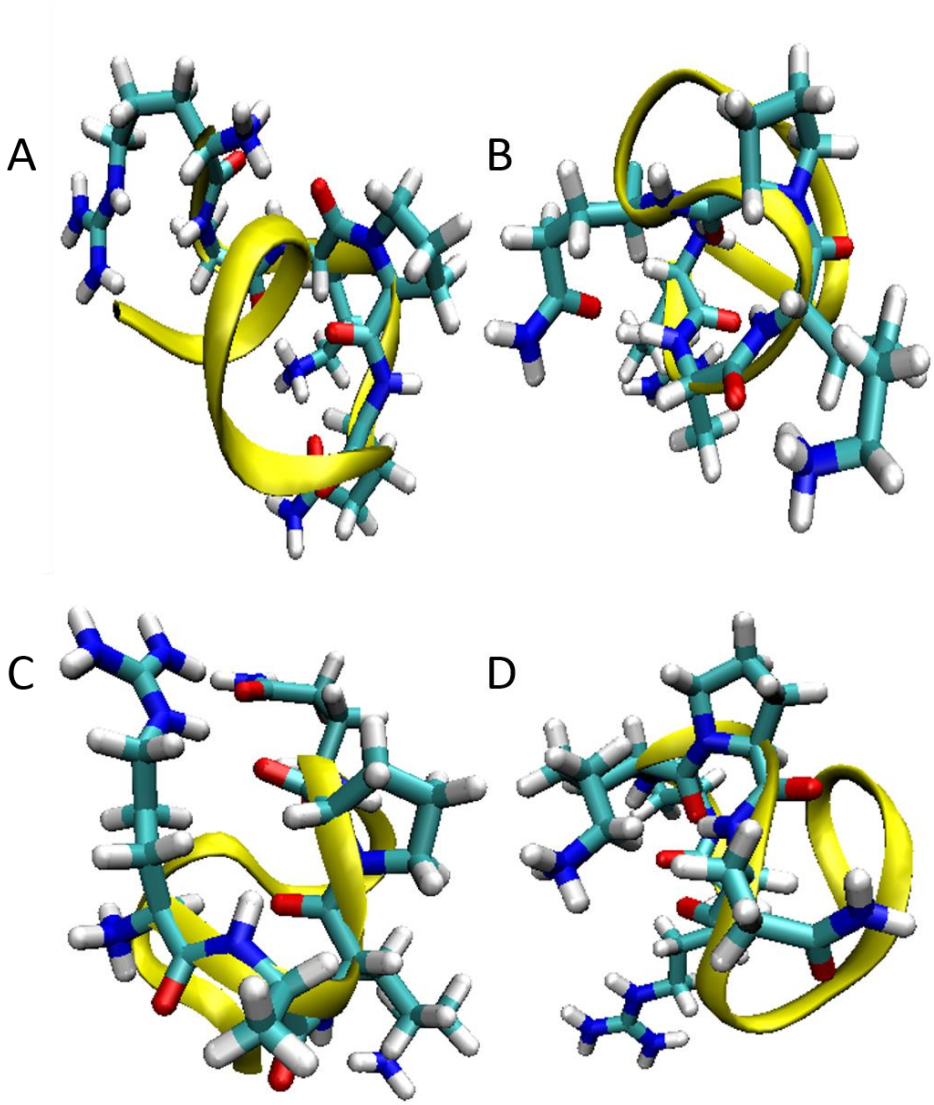

Figure 21. Best-fit structures for the largest population clusters of mutant Substance P, P2A: 301 $Å^2$ cis conformer Pro 4 (A, 11%), 312 $Å^2$ cis conformer Pro 4 (B, 54%), 301$Å^2$ trans conformer Pro 4 (C, 42%), and 312 $Å^2$ trans conformer Pro 4 (D, 21%). Cis proline conformers A and B have guanidinium ions coordinating with the C-terminus; trans prolines C and D have the N-terminus coordinating with the C-terminus.

P2A mutant was simulated because the mutation influenced the compact conformations, and the results are shown in **Figure 21**. The cis conformer of Pro 4 301 $Å^2$ and 312 $Å^2$ are shown in **Figure 21 A** and **B**, respectively. With additional constraints, the cluster relative population was higher than the clustering results when only the CCS values were used to filter the simulation results. Confining the projection of Pro 4 to cis or trans conformer decreased in variation of the backbone structure, which results in increased cluster population. This holds true for the trans isomer filter results for P2A mutants are displayed in **Figure 21 C** (301 $Å^2$ 42%) and D (312 $Å^2$ 21%). The cis isomer preference increased by about 34% compared to the wild type substance P (29% -> 44%).

The mutation studies have only changed the relative abundance of compact conformer or peak splitting of compact conformer. This was due to the mutations disrupted the charge solvation by removing Q or F, or changing N-terminus projection by removing prolines. However, the IM profile suggested that the compact conformer stability was not affected by G9P mutation, but the mutation changed the orientation of the C-terminus leading to splitting of extended conformer. The simulation results for the extended conformers are shown in **Figure 22 A** (352 $Å^2$, 28% of total population), and B (360 $Å^2$, 75% of total population). The best-fit structure shows that the C-terminus is coordinating with the N-terminus, shown in panel A, or the guanidinium ion, displayed in **Figure 22 B**. The CCS filtered structures overlapped significantly because the two mobility peaks are close. Therefore, it is difficult to derive statistical conclusion even though the coordination of the C-terminus is different between the two cases. The trans G9P cluster results are shown in **Figure 22** (**C**, 352 $Å^2$ trans conformer, 36% of total population) (**D**, 363 $Å^2$ trans

conformer, 18% of total population). Both structures shows that, the C-terminus are coordinated with the N-terminus so that the effect of isomerization do not affect the overall structural outcome of the extended conformer.

The SP and mutants simulations have shown that the SA-MDS is capable of producing cis and trans prolines. Analysis program was also modified so that the cis/trans prolines could be identified and sorted according to the proline isomer. When ESI-IM-MS experiment assigns a mobility peak to a specific proline isomer, simulated structure analysis should be only performed on the specific proline isomer. Cis/trans preference was determined and for SP the simulated structures contained up to 72% of cis prolines. However, the study is limited, so more research has to be conducted to determine the factors influence proline isomer since the simulation results are inconclusive.

Polyproline

Proline cis-trans isomerization plays an important role during protein folding. Due to steric hindrance, proline can adopt cis or trans conformation caused by proline's cyclic nature.[58] In nature trans conformation is predominant conformation for non-proline residues and Hilgenfeld and coworkers have determined approximately 5% of the reported prolines were in cis conformation.[59] This rigid characteristic makes proline disrupt secondary structures such as alpha helix or beta sheets.[60] For better understanding of the proline cis-trans isomerization, polyprolines are used as model systems to investigate the isomerization process. Poly-prolines exist in two conformations, poly-Pro I (PPI) helix favored in hydrophobic environments, while poly-Pro II (PPII) helix favored in aqueous

solutions.[61-64] The PPI helix consists of all cis conformation prolines, and it is more compact than the PPII helix, which consists of all trans conformation. Recently, Clemmer and coworkers have investigated poly-proline 13 isomerization by altering solvent conditions and analysis followed by ESI-IM-MS and theoretical calculations.[65,66] Experiments have indicated that poly-proline 13 isomerizes as the peptide solvent condition is changed from water:acetic acid solution to propanol:water:acetic acid solution. In this section the proline isomerization is further investigated by theoretical methods. The cis trans isomerization of prolines are investigated using MDS methods at elevated temperatures > 600 K to monitor the sequence of isomerization and their CCS changes. Doubly charged poly-proline was simulated with a proton placed on the N-terminus, and second proton was placed on the peptide backbone for polypro 13 to make 2+ ion. The simulations were carried out at constant temperature.

The plots of Cα-Cα distances are displayed in **Figure 23** which indicates the cis/trans isomerization of the prolines in gas phase simulation at 800 K. The plot shows that at 800 K isomerization occurs mostly during the last 500 ps of the simulation and the proline conformation transition is static. Two prolines shown in panel I (Pro 10 and Pro 12) are the only prolines that does not isomerize. The panel A (Pro 2), indicate that the proline converts to trans isomer just temporarily, but it predominantly remains in cis conformation. Panel B, J, and K shows isomerization at ~1500 ps and at this time the CCS of polyproline (panel L) shows compaction. However, other proline isomerizations appear that the isomerization does not impact the CCS of the peptide. After 1500 ps the CCS is

converge around 285 $Å^2$ which agrees with polyproline 13 peptide CCS measured by ESI-IM-MS by Clemmer and coworkers.[65]

The temperature effect on the cis/trans isomerization was investigated, and the result of temperature dependent isomerization is shown in **Figure 24** and **25**. Simulation results at 600 K indicate that only one proline isomerization is observed at Pro 5 (**Figure 24 F**). When the temperature increased to 700 K, proline isomerization are observed on two prolines. However, when the temperature is increased to 1000 K (shown in **Figure 25)**, all of the proline except for the protonated proline isomerizes, and most of the prolines are dynamic. From this simulation, we can conclude that the higher the temperature, proline has enough energy to isomerize, but too much energy will make the prolines dynamic. Therefore, subsequent simulations were performed at an elevated temperature of 800 K where there are just enough energy to isomerize prolines in reasonable simulation time and cost.

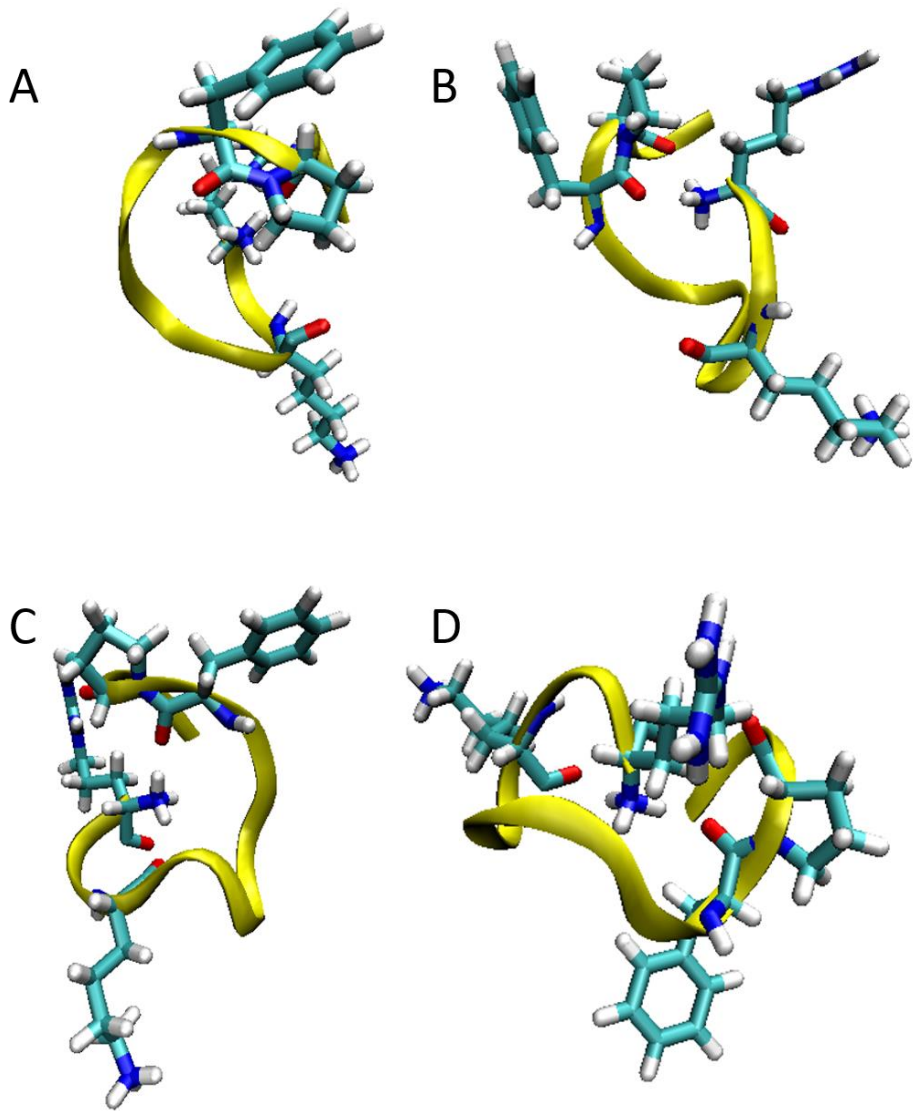

Figure 22. Best-fit structures for the largest population clusters of mutant Substance P, G9P: 352Å2 cis conformer P9 (A, 28%), 360Å2 cis conformer P9 (B, 75%), 352Å2 trans conformer P9 (C, 36%), and 363Å2 trans conformer P9 (D, 18%). The N-terminus coordinated with the C-terminus in three cases (A, C, and D); in B it coordinated with the guanidinium ion.

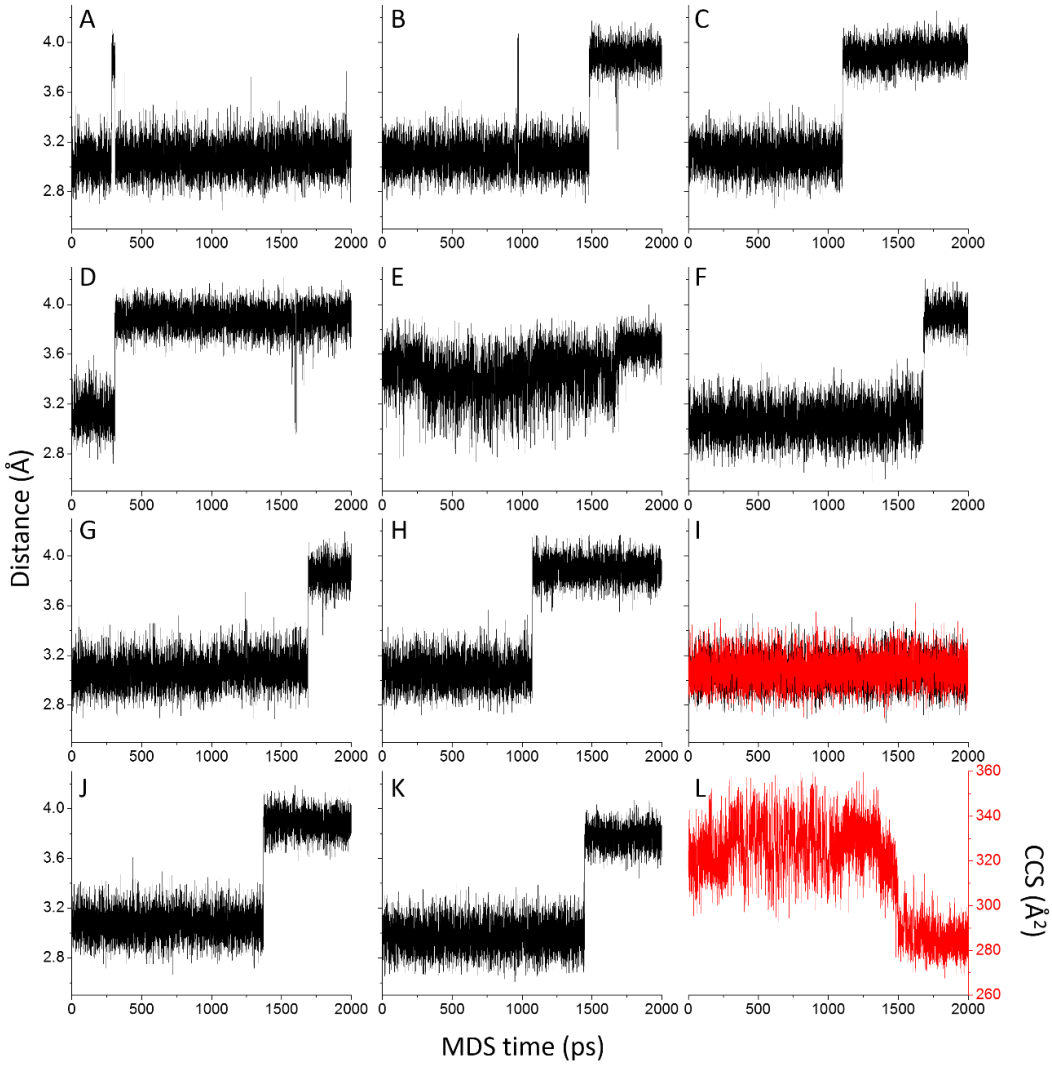

Figure 23. Plot of Cα-Cα distance of polyproline-13 2+ gas phase simulation at 800 K. Protons are on N-terminus and Proline 6. Panel A-H shows the Cα-Cα distance of P1-P2, P2-P3, P3-P4, P4-P5, P5-P6, P6-P7, P7-P8, and P8-P9, respectively. Panel I shows the distance of two Cα-Cα distances of P9-P10 and P11-P12. Panel J and K shows the Cα-Cα distance of P10-P11 and P12-P13, respectively. The CCS of the polyproline 13 peptide is shown at L.

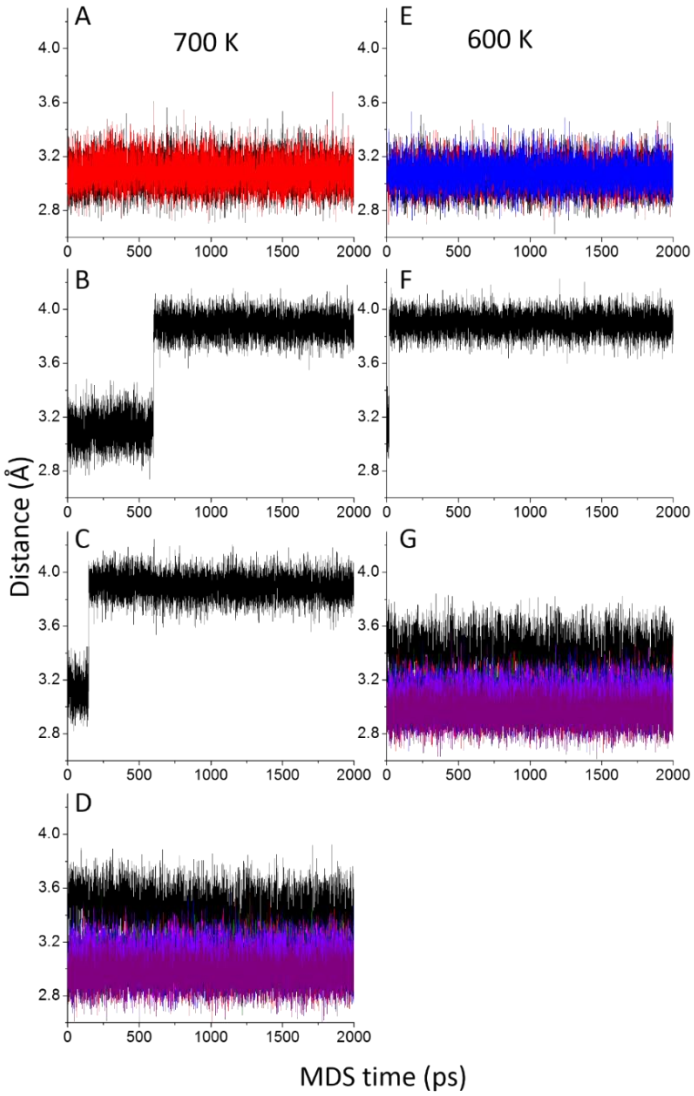

Figure 24. Plot of Cα-Cα distance of polyproline-13 2+ gas phase simulation at 600 K and 700 K. Protons are on N-terminus and Proline 6. Cα-Cα distance from 700 K are shown in A-D and 600 K are shown in E-G. Panel A shows the Cα-Cα distance of P1-P2 and P2-P3. Panel B and C shows the P3-P4 and P4-P5, respectively. Panel D shows the Cα-Cα distance of P5 through P13. For 600 K only one one proline have isomerized and it is shown in panel F, P4-P5.

The two protonation locations were previously suggested by Clemmer and coworkers to be on the N-terminus and amide nitrogen of Pro 6.[66] However, the effect of charge location on polyproline 13 was investigated further by changing the charge location to Pro 3 (**Figure 26)** and to Pro 9 (**Figure 27**). When the proton was on Pro 3 or Pro 9, the total number of prolines converting to trans isomer is greatly reduced compared to the simulations conducted with proton on Pro 6. When the second proton was placed on Pro 3, only two prolines isomerized, Pro 2 and Pro 5 (**Figure 26 A** and **D**). The proline isomerizations were only visible on the N-terminus region when the second proline was placed on Pro 3. When the second proton was placed on Pro 9, total of 5 prolines isomerized and the plots are shown in **Figure 27,** and the isomerization is visible around the center and the C-terminus region of the peptide. The CCS plot of both peptides do not indicate any significant changes in peptide structure, which does not agree with previously reported ESI-IM-MS results.[65] Whereas, all of the prolines were isomerized when the proton was placed on Pro 6. Comparing the results of the three experiments shown in **Figure 25, 26,** and **27,** simulation results are clear that the charge location strongly affects the proline isomerization in gas phase by charge-charge repulsion. The coulombic repulsion can be lost if the distance between charges increases.

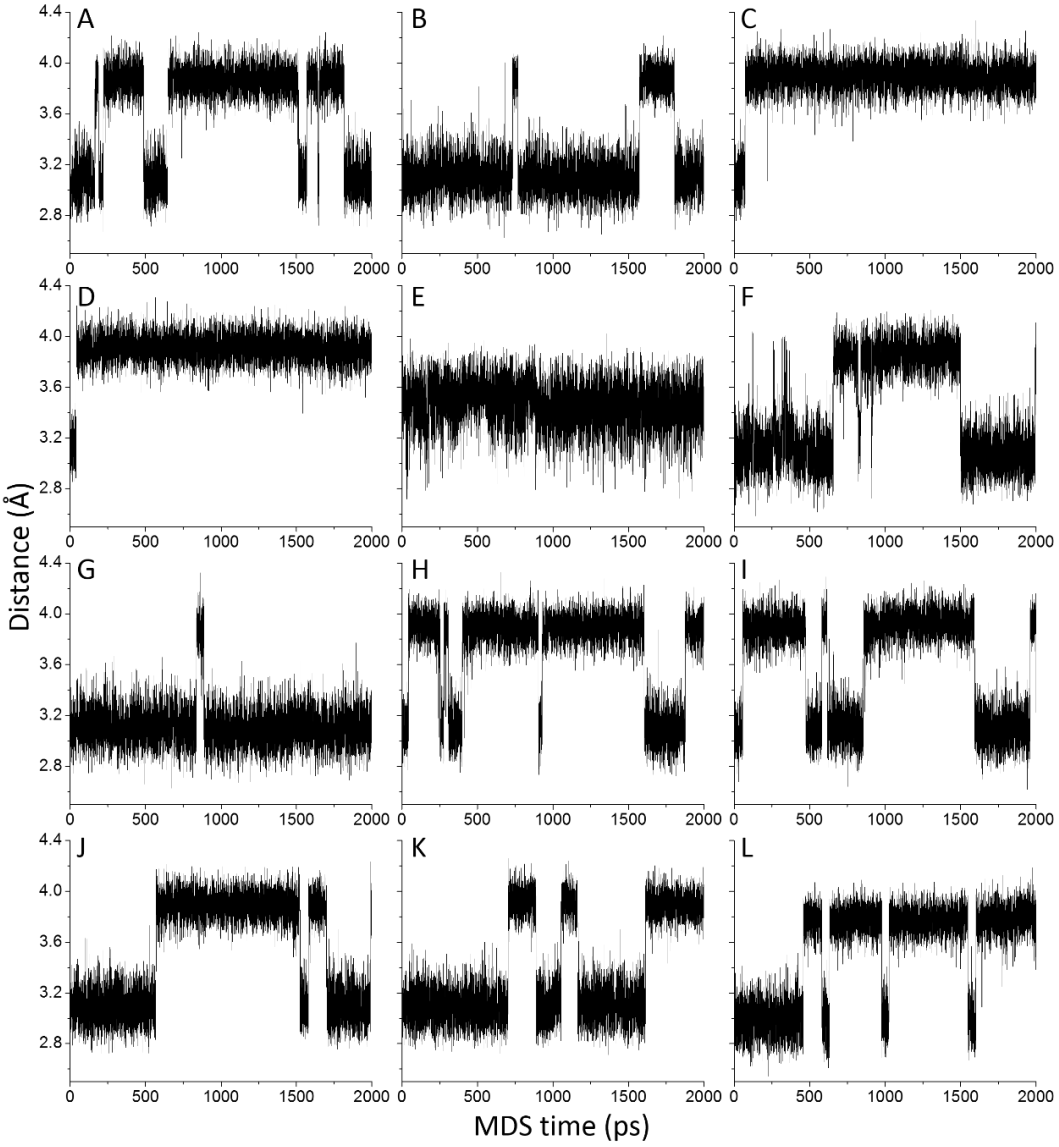

Figure 25. Polyproline 13 2+ gas phase simulation at 1000 K. Protons are on N-terminus and Proline 6. Panel A through L shows Cα-Cα distance of P1-P2, P2-P3, P3-P4, P4-P5, P5-P6, P6-P7, P7-P8, P8-P9, P9-P10, P10-P11, P11-P12, and P12-P13 respectively.

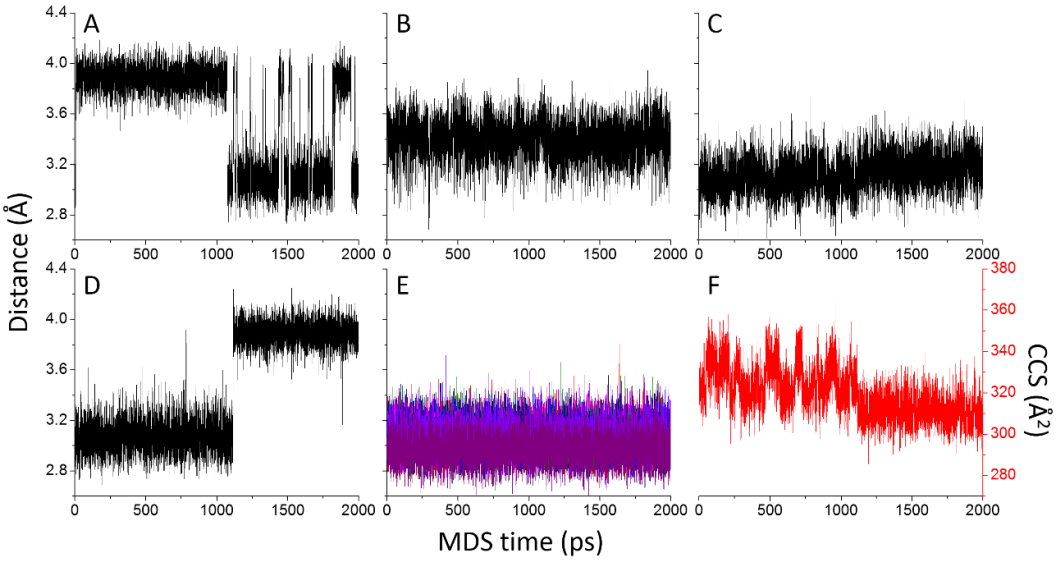

Figure 26. Polyproline 13 2+ gas phase simulation at 800 K. Protons are on N-terminus and Proline 3. Panel A shows Cα-Cα distance of P1-P2, P2-P3 in B, P3-P4 in C, P4-P5in D, and in panel E remaining distances are shown from P5 to P13, which does not isomerize an d locked in cis conformation. The CCS of the polyproline 13 peptide is shown at F.

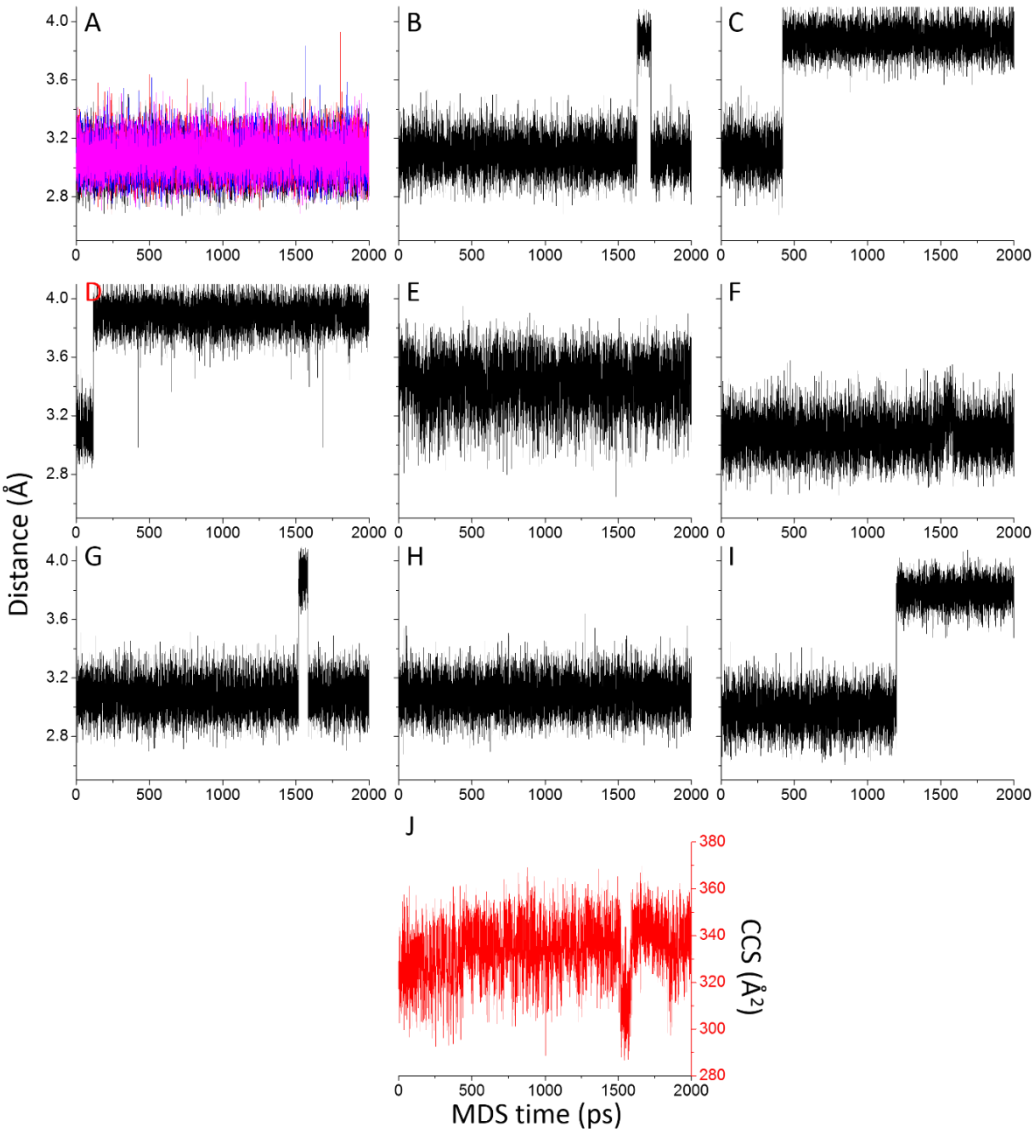

Figure 27. Polyproline 13 2+ gas phase simulation at 800 K. Protons are on N-terminus and Proline 9. Panel A shows 4 Cα-Cα distance of P1-P2, P2-P3, P3-P4, and P4-P5. Panel B shows the Cα-Cα distance of P5-P6, P6-P7 in C, P7-P8 in D, P8-P9 in E, P9-P10 in F, P10- P11 in G, P11-P12 in H, and P12-P13 in I. The CCS of the polyproline 13 peptide is shown at J.

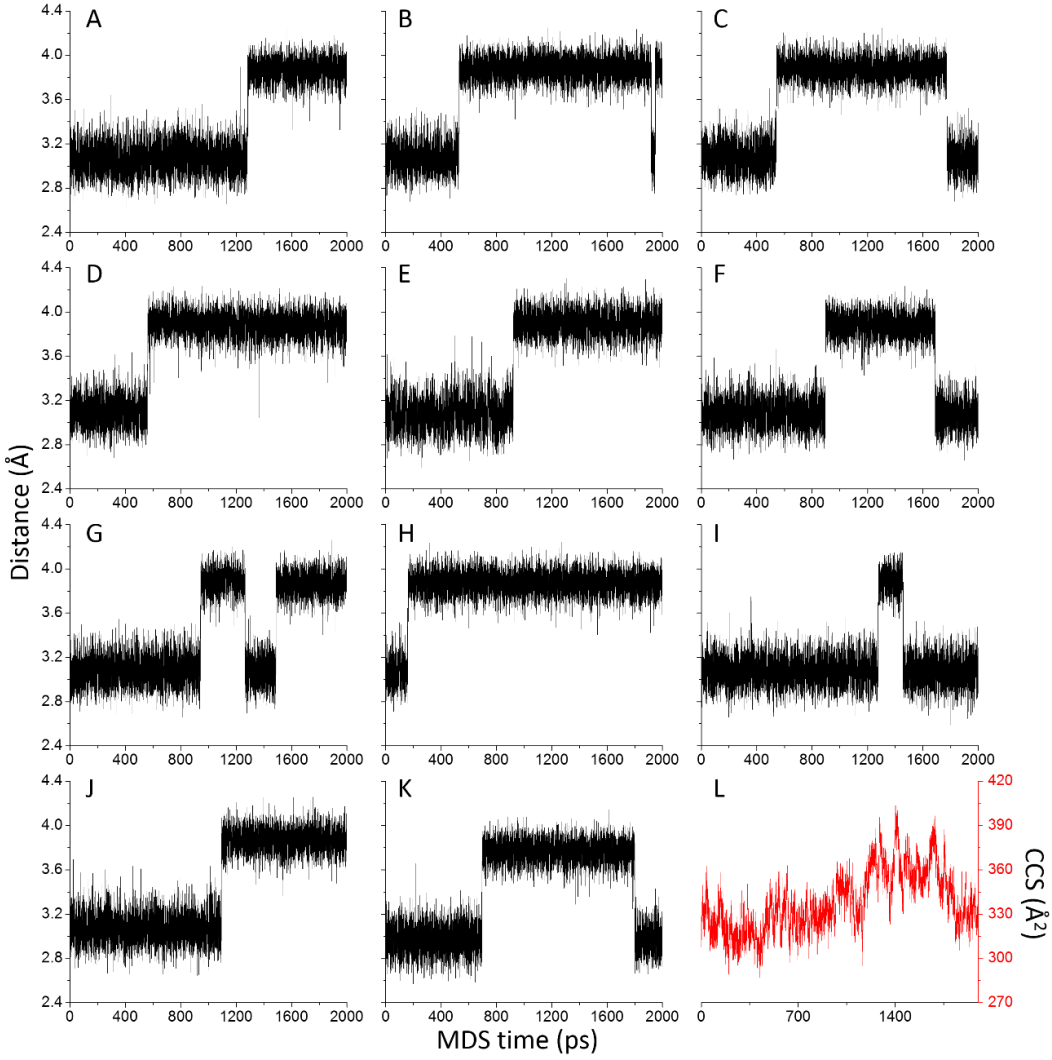

Figure 28. Polyproline 13 2+ solution phase simulation at 1000 K. Panel L shows the CCS of the PP13 peptide. P5-P6 distance is not shown because the second proton is on proline 6 which removes the cis trans isomerization. Panel A-K shows Cα-Cα distance of P1 through P13 except for P5-P6. The CCS of the polyproline 13 peptide is shown at L.

The polyproline 13 cis to trans isomerization was also investigated in solution, and the plot of Cα-Cα distances are shown in **Figure 28**. Unlike the gas phase simulation, the solution phase simulations was carried out at 1000 K. Proline conformation shown in **Figure 28** indicate that most of the prolines stay predominantly in cis conformer and even if the isomerization occurred, most of the prolines converted back to cis conformation. The observed CCS change was also different. The gas phase simulation CCS have shown compact conformation at the end of the simulation, but the solution phase simulation CCS does not display significant compaction. In the gas phase simulation, the trans isomer prolines led to collapse of the peptide structure, but in solution phase, the peptide remained in elongated structure.

Polyproline simulation have successfully demonstrated that the elevated temperature simulation of 800 K and above can readily isomerize prolines. The isomerization of proline isomer will change the peptide structure, and the change is represented by the CCS. Previous experimental structure containing mostly trans proline isomer CCS value agreed with the calculated CCS value from the simulation. The modeling enables to monitor the proline isomerization sequence, and it can distinguish which proline leads to the most change in CCS. The sequence of isomerization observed from this simulation study is different compared to the previous experimentally determined sequence.[65] However, the difference can be attributed to the limitation of the simulation where the bond formation and cleavage is forbidden, which prohibits the proton transfer that might occur. The solvation simulation has suggested that the presence of solvent limits change in peptide structure.

Conclusion

The model peptide simulations have confirmed that the MD simulations can create a candidate library of structures for the peptides characterized by ESI-IM-MS experiments. The result of the simulated structures can be filtered by the experimentally measured CCS values which increases the validity of the filtered structures. It can be followed by a statistical analysis method which groups similar structures into a group and give the most representative structure.

Atom to atom distance calculating algorithm was developed to find structures with strong noncovalent intramolecular interaction which was previously suggested. The same atom to atom distance filter was modified to measure distance between two $C\alpha_i$-$C\alpha_{i-1}$ distances to determine the cis/trans conformation. With the determination of proline conformation, the change in CCS can be correlated to a specific proline isomerization. These studies have indicated that the combination of experimental results such as CCS, distance, and proline cis/trans conformation can be successfully used as simulation boundary conditions, which increase the accuracy of simulation results.

CHAPTER IV

IONS FROM SOLUTION TO THE GAS PHASE: A MOLECULAR DYNAMICS SIMULATION STUDY OF DROPLET DESOLVATION AND STRUCTURE EVOLUTION OF SUBSTANCE P

Native-electrospray ionization (n-ESI) marks a major shift for biological mass spectrometry, leading to a transition from the determination of primary structure to studies of higher order (2^o, 3^o, and even 4^o structure), thereby opening new directions in the field of structural biology. The focus on structure places new demands on understanding how solution-phase structures are influenced as they transition from solution to solvent-free, gas-phase ionic species. In addition, to understand how the solvent-free ions (i.e., in cases where solvent-free ions retain their "native" conformations) are formed with sufficient internal energy to undergo rearrangement reactions to form different solvent-free conformations as exemplified by the question "…for how long, under what conditions, and to what extent, can solution structure be retained without solvent?"[67,68] Chen et al. have recently presented evidence that ubiquitin and metallothionein-2A ions formed using "native-state" ESI are indeed formed as native or "native-like" conformers; however, this study clearly illustrates that great care must be taken in order to preserve the nascent ion population, *i.e.*, post-ionization collisional heating can provide sufficient energy to drive unfolding to form a conformationally heterogeneous ion population.[38] In fact, the heterogeneous ion population observed in their study is quite similar to results obtained by collisional activation previously reported by Clemmer[69] as well as solvent denaturation reported by Bowers.[70]

Fenn's 1989 introduction of ESI with large biomolecules have inspired much experimental and theoretical research aimed at understanding the underlying mechanism of ion formation from water (and mixed organic/water) droplets.[10,71-74] These studies focused on understanding how a charged droplet is produced from the ESI emitter and the mechanism of ionization from these liquid droplets. The initially formed liquid droplets are quite large, thus the molecule residing therein experiences a bulk or bulk-like solvent environment, but the droplet size is rapidly reduced by solvent evaporation and droplet fissions to ultimately yield solvent-free, ions in the gas phase.[10,72,75-77] Results from MD simulations are consistent with this general model of the transition from a large, bulk-like droplet to the solvent-free ion,[78] thus the generally held ion formation mechanisms have converged to two models: (i) the charge residue model (CRM),[10,12] which describes the process as small droplets containing the analyte ion evaporating to produce a population of solvent-free ions; and (ii) the ion evaporation model (IEM),[10,13] which assumes the surface-charge density of a droplet is sufficiently high to eject an ion that resides near or on the surface of the droplet. Gross et al. combined the CRM and IEM ionization mechanisms as the charged-residue/field-emission model.[79] Recently, to investigate the evolution of a biomolecule during the desolvation process, the stepwise rehydration of a small peptide was investigated by Bowers et al.[80] and the stepwise dehydration of small peptides was investigated by Silveira et al.[15] Molecular dynamics simulations have also been used to investigate the desolvation process, where proton, hydronium, Na^+, Ca^{2+}, and Cl^- ions were simulated individually or in some combinations.[81-86]

An often-overlooked detail of ESI, especially important for ion structure studies, is the temperature of the ion population. Early in the development of ESI, Fales showed that hydrated ions of polar molecules are formed in high abundances when electrosprayed from aqueous solutions.[87] Beauchamp and coworkers showed that evaporation of water from such hydrated molecules produced "cold ions", a process that is similar to forming "freeze-dried" samples, at an estimated temperature of ~150 K.[88] We recently reported a similar strategy for preparing hydrated peptide ions followed by cryogenic (80 K)-ion mobility-mass spectrometry (cryo-IM-MS) for investigating the structure of hydrated biomolecules. In this experiment, hydrated biomolecule ions, i.e., $[M + nH]^{n+} \cdot \mathbf{x}(H_2O)$, are observed where x = 0 – 100.[15] This experimental approach has unique potential for addressing the question related to the structural changes that occur in the final stages of solvent evaporation. For example, cryo-IM-MS studies on substance P (SP) showed that two conformations of $[SP + 3H]^{3+}$ (SP^{3+}) ions are observed as solvent-free, gas phase ions. The compact conformer ion is formed by dehydration, via CRM, but the extended conformer ion is a gas-phase equilibrium state formed only after dehydration and heating. A similar structural change that occurs late in the dehydration reactions of dialkylammonium cations was recently reported.[89,90]

Anions have been shown to have significant effects on the product ions formed by ESI,[91-94] but the underlying mechanism(s) by which such effects arise have not been extensively studied, either by simulations or experiment. Although Chait et al. reported on the effects of counter-anions on positive ESI ion yields, most notably charge-reduction reactions,[95] possible effects of counter-anions in droplet evaporation and the formation of

positively-charged solvent-free ions have been largely ignored. Previous MD simulations of ESI have treated the droplet only in terms of bulk water containing positively-charged ions;[75,83,96] however, droplets having net positive charge may also contain negatively-charged counter ions. Konnerman recently reported results from simulations that describe the evolution of an ESI droplet containing a protein (ubiquitin, cytochrome C or holo-myoglobin) and Na^+ and Cl^- ions with the focus on protein structure and charge state;[97] however, in these studies, the detailed intra/intermolecular interactions involving the small ions and the protein were not probed in depth. Here, we use MD simulations (MDS) combined with ion mobility spectrometry to better understand how the presence of anions affects conformational preferences of biomolecule ions. SP was chosen for this study because previous studies have shown that this peptide is unstructured in water and adopts a helical structure in a low dielectric environment;[98] however, our previous ESI studies showed that two conformers with very different CCS were detected in the gas phase.[15,57] This information is then used to correlate conformational changes observed during the final stages of droplet evaporation to the SP^{3+} observed by cryo-IM-MS experiments.

Experimental

Simulations were conducted on a 176-core SGI Altix 450 cluster at the Texas A&M University Chemistry Department Laboratory for Molecular Simulation. The AMBER 11 molecular dynamics package and AMBER ff99SB force fields were used in this study. GAUSSIAN 03 and RED III[33] were used to create custom residues that were not supported in AMBER. Hydronium ion parameters were obtained from results of

Baaden et al.[99] CCS values for simulated species were calculated using the MOBCAL trajectory method.[27,26]

A total of 12 MD simulations were performed. Four replicate MD simulations for the peptide, hydronium ion and chloride anion-containing water droplet were performed by placing the extended solution phase conformer of SP^{3+} in the center of the droplet and randomly placing hydronium and chloride ions near the peptide. An additional four simulations were performed with exactly the same peptide structure, but varying with new ion locations within the droplet. The final four simulations were performed starting with a droplet containing randomly placed Cl^- and H_3O^+ ions and a compact conformer of SP with a CCS that is similar to the experimentally determined value of 316 $Å^2$. The charge sites on the peptide were fixed at the N-terminus and the arginine and lysine side chains, respectively. The starting system was designated to have 15 net positive charges, consisting of 22 hydronium ions, 3 charges on the peptide, and 10 chloride ions in 2348 water molecules. The Rayleigh limit is ~13 for a droplet consisting of approximately 2400 water molecules, as calculated using the surface tension of the TIP3P water model.[100] The nonbonded cutoff was set to 8 Å, and the simulation box size was set to ~230 Å. The TIP3P[101] water model was used and Rayleigh limit was calculated by the surface tension of the TIP3P model. The model have its limitation of its own such as, $H3O^+$ is treated as a static unit,[102] non-polarizable,[103] and surface tension is lower than water.[100] Temperature was set to 360 K using a Langevin thermostat, which represents the heated capillary temperature used in our previous study; because the rate of evaporation slowed dramatically after ~3 ns, the temperature was increased to 420 K.

Melittin from Honey Bee Venom and SP (both from Sigma-Aldrich, St. Louis, MO, USA) were used without further purification. Each peptide was diluted to a concentration of 10 μM with either 18 MΩ water or water/0.1% hydrochloric acid (~12.6 mM, Fisher Scientific, Pittsburg, PA, USA). A Waters Synapt G2 HDMS instrument (Manchester, UK) was used for all ESI-IM-MS experiments. The instrument conditions used were as follows: sample cone 10 V, extraction cone 1 V, trap bias 25 V, helium cell flow rate 200 mL/min, IMS nitrogen flow rate 50 mL/min.

Results and Discussion

MD simulations are frequently used to investigate biomolecule desolvation reactions that occur during ESI, but correlating these results to experimental data has proved to be exceedingly difficult. We have previously shown that cryo-IM-MS provides a means for direct measurement of the effects of hydration on the size/shape of biomolecule ions, where hydration ranges from bulk-like water to remnant water molecules interacting with the most hydrophilic functional groups, and finally to the solvent-free, gas-phase ion.[15,57,90] Recent results from cryo-IM-MS experiments provide the first experimental data that show that dehydration gives rise to specific conformational preferences. In the specific case of the amphipathic peptide SP, the transition from solution to the gas phase occurs via a series of dehydration steps that can be traced to a single, compact peptide ion conformer.[15,90] The results from this study provide a novel approach for correlating predictions from MD simulations to experimental results. This paper will first discuss the changing environment of the droplet through fission events, small water cluster ejection, and evaporation, which will be followed by evaluating the specific

intramolecular interactions responsible for the peptide ion's structure as the size and ion concentrations of the droplet change. The paper will finish with experimental results supporting simulation findings.

Dynamics of Water Droplet Evaporation with Ions: Temperature Optimization

Znamenskiy et al.[81] simulated protonated water droplets at 300 K and reported droplet protrusions and fission. In order to reproduce and optimize simulation temperature, the desolvation process was simulated at four different temperatures of 240 K, 300 K, 360 K and 420 K. Constant temperature was maintained by Langervin dynamics. Constant temperature simulation was necessary for two reasons. First, previous simulations of constant energy reported droplet froze as water evaporated.[78,82] Second, previous cryo-IM-MS experiments,[104] detected protonated water clusters generated from a solution containing HCl at a heated capillary temperature of ~360 K. The hydronium ions were initially randomly placed near the center of the water droplet consisting of ~2400 water molecules.

At all temperatures droplet fission was observed. The hydronium ions migrated to the surface of the droplet regardless of the temperature settings, which leads to local elongation of the droplet followed by droplet fission. This observation is reminiscent of a result previously demonstrated in a study by Marginean et al.[75] which reported droplet elongation and droplet fission. Water evaporation were observed in all temperatures except for 240K, and the droplet was more dynamic at higher temperatures. As the cryo-IM-MS experiments had used desolvation temperatures ~360K and the simulation at 360K

have successfully reproduced previous simulations, 360K was used for folling simulation, unless noted otherwise.

Charge Optimization

Further simulations were performed to study the effect of the number of charges in a water droplet. Scenarios with 10, 15 and 21 hydronium ions in a droplet containing ~2400 water molecules were simulated. The calculated Rayleigh limit of a TIP3P water droplet containing 2400 water molecules is ~13. A snapshot of each simulation after 700 ps of droplet evolution is shown in **Figure 29.** When the total charge within the droplet does not exceed the Rayleigh limit (**Figure 29 A** and **B**), the products of droplet fission product contain more than 150 water molecules, with few exceptions. However, when a droplet containing 21 hydronium ions was simulated, after 900 ps the majority of the hydronium ions were observed in water clusters consisting of less than 100 water molecules, resulting from rapid and sequential droplet fission.

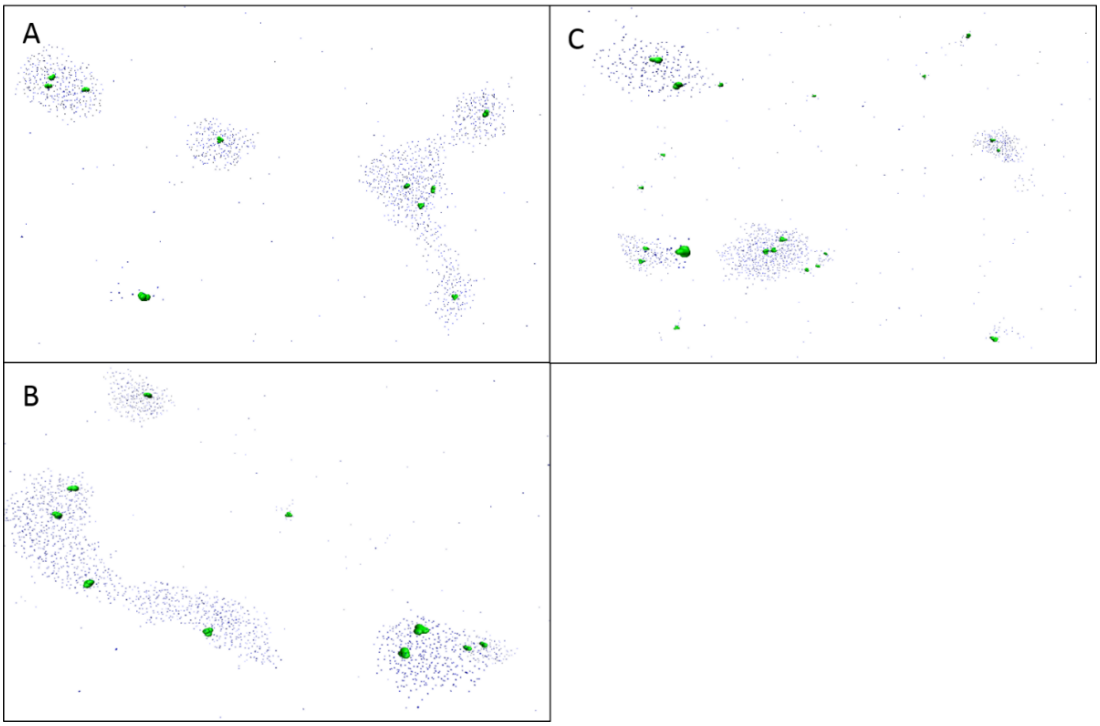

Figure 29. Desolvation of hydronium ions from water droplet after 700 ps at 360K. The simulation starts with hydronium ions (green Licorice) randomly distributed in the center of the water droplet (blue with only oxygen displayed for clarity) and ends after 700 ps (computing constraint). The evaporation of individual water molecules and droplet fission products are visible in all three simulations: A (10 hydronium ions and 2396 water molecules), B (15 hydronium ions and 2383 water molecules) and C (21 hydronium ions and 2376 water molecules); the Rayleigh limit at ~2400 water molecules is ~13 charges. The simulation shows fission resulting from water droplet elongation which is caused by hydronium ion migration from the droplet center to the surface.

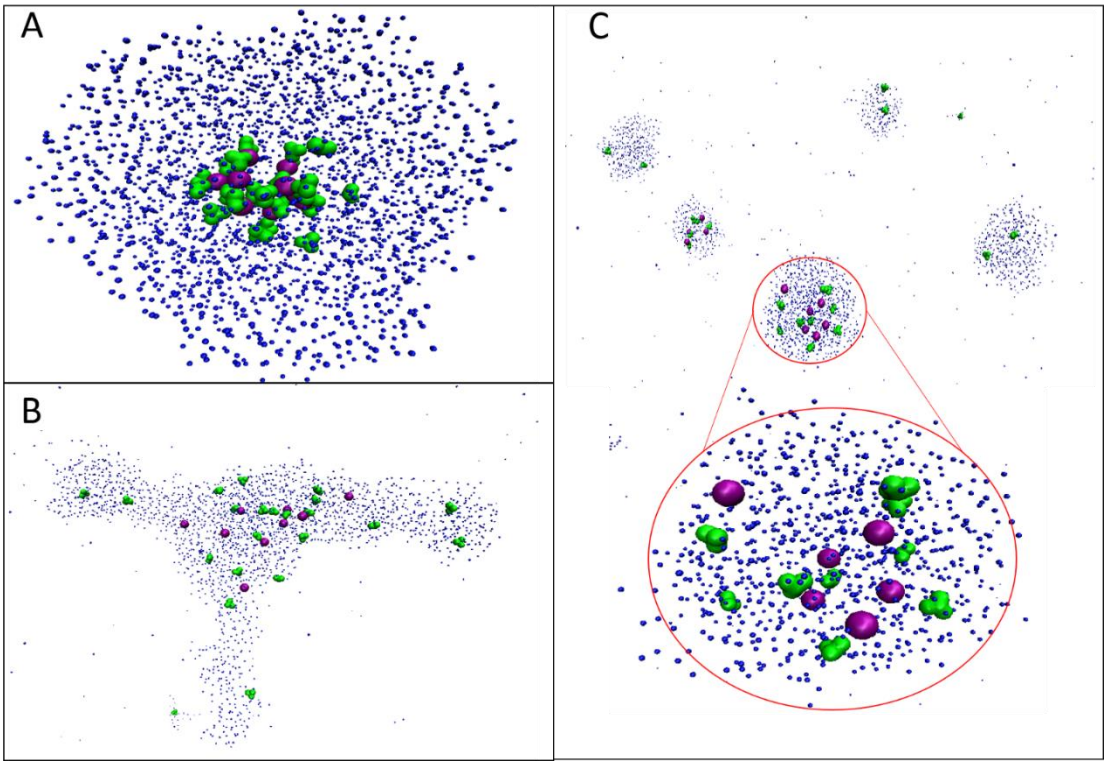

Figure 30. The effect of chloride ions on hydronium ion desolvation. The simulation starts with A 25 hydronium ions (green) and 10 chloride ions (purple spheres) randomly distributed at the center of a 2373 water molecule (blue dots) droplet at 360K. Images acquired at various time points (B 340 ps, C 700 ps) demonstrate the reduced desolvation rate of the hydronium ions with Cl- present.

Chloride Effect Study

Thus far, these simulations have consisted of only water and hydronium ions. In solution, all positively charged species would have a counter ion; consequently, an ESI droplet would contain a mixture of positive and negative charges. Therefore, chloride ions were added as counter ions, while the droplet was set to have a net charge of +15 by having 25 hydronium ions and 10 chloride ions (**Figure 30**). The incorporation of chloride ions have reduced the rate of desolvation compared to simulations containing only hydronium ions. This result is expected, as addition of chloride ions aids in water molecule network as the hydronium chloride ion pair will organize water molecules according to water molecules' dipole moment.

Evolution of the ESI Droplet from MD Simulations.

All replicate simulations showed a similar early desolvation process; **Figure 31** shows snapshots from a single simulation illustrating the observed events in the first 2 ns of desolvation. Initially, water evaporation and water/hydronium cluster ejection dominate the desolvation process (**Figures 31 A** and **B**); however, the hydronium and chloride ions are also migrating to the droplet surface resulting in deformation of the droplet and formation of extended lobes. Formation of theses lobes results in droplet fission events that carry away hydronium and chloride ions (**Figure 31 C**). After these fission events, the droplet regains a more spherical composition (**Figure 31 D**). Then, water evaporation once more dominates the desolvation process. Upon again approaching the Rayleigh limit, an additional droplet fission event is observed in the simulation (**Figure 31 E**). With the

exception of one simulation (see **Figure 32 C**), chloride and hydronium were not ejected from the droplet after droplet fission. Water evaporation continued until all water molecules were removed from the SP^{3+} ion, resulting in increasing concentrations of hydronium and chloride. Finally a desolvated SP^{3+} ion was observed with both hydronium and chloride adducts. These findings are consistent with previous desolvation simulations[86,97,105] confirming that desolvation of this peptide occurs via the combined charge residue-field emission model proposed by Gross et al.[79] Furthermore, this consistency validates the simulation conditions used in this study.

Evolution of SP^{3+} Conformations

Figure 32 shows the evolution of SP^{3+} ion structure, via CCS, as a function of time and numbers of remaining water molecules for select representative simulation. Despite identical starting conditions **Figures 32 A** and **B** show that the folding pathway of SP differ, but result in a similar final compact conformer CCS. The starting conditions of the results shown in **Figure 32 C** differ only in Cl^- and H_3O^+ positions. Interestingly the CCS of this simulation does not produce any compact conformer CCS. All other replicate MD simulation CCS trajectories are provided in **Figure 33**. Overall, SP^{3+} ions rapidly equilibrate to similar extended conformations regardless of initial conformation, which is consistent with the extended conformers observed in bulk-water in a previous NMR study[98], and the final conformation upon complete desolvation is less than 330 $Å^2$ for many of the simulations, which is within 4% of the experimentally determined value of 316 $Å^2$.

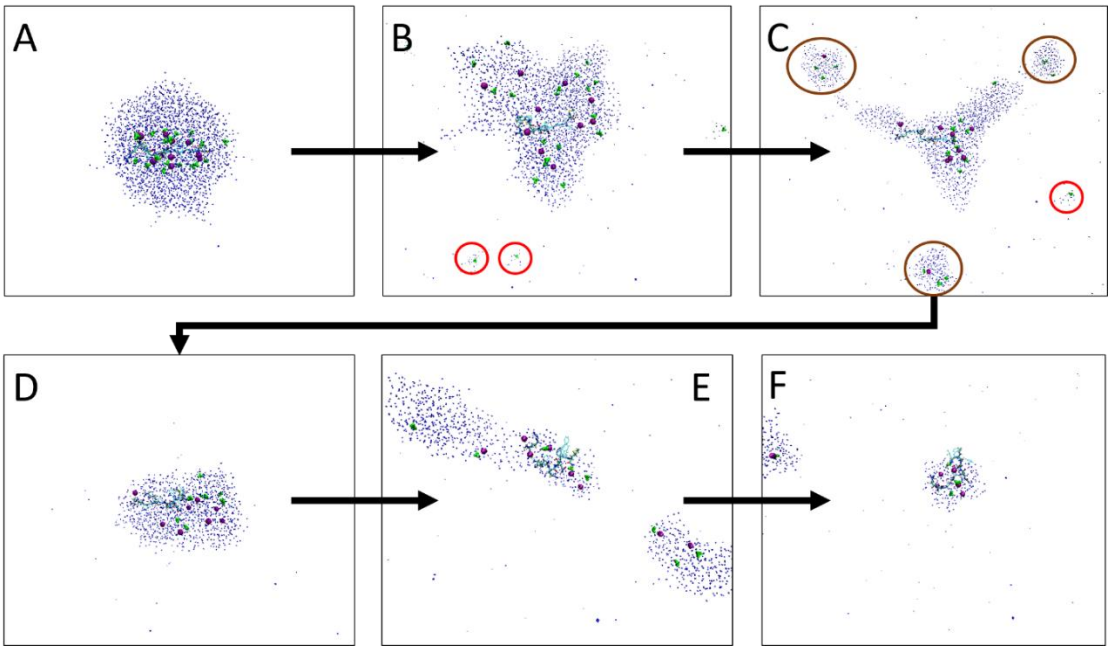

Figure 31. Snapshots of the early stages of the desolvation process from a single simulation shown after (A) 0 ps, (B) ~400 ps, (C) ~560 ps, (D) ~825 ps, (E) ~1500 ps, and (F) ~1550 ps. Red circle indicated emitted hydrated ion cluster while brown circle indicate droplet fission product.

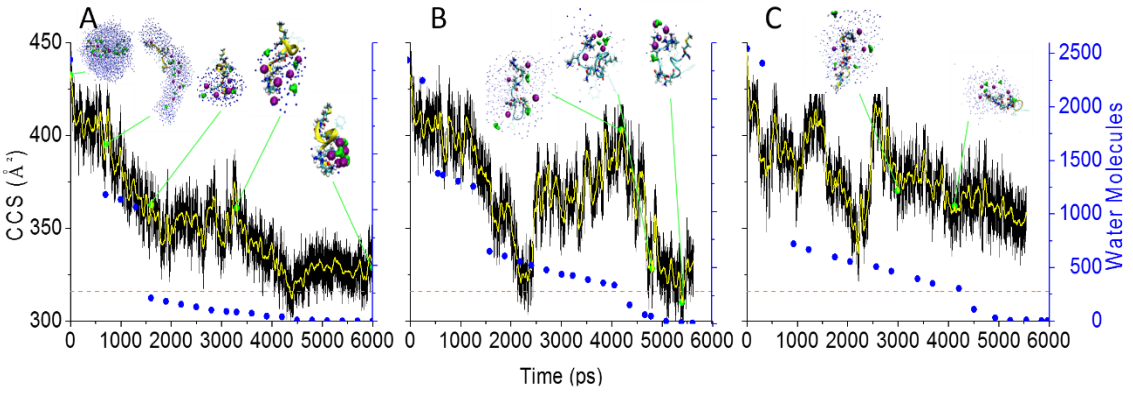

Figure 32. Three plots of simulated SP3+ (RPKPQQFFGLM-NH2) ion CCS and number of water molecules vs. time. The black line represents the simulated CCS of SP^{3+}, the yellow line represents the smoothed CCS of SP^{3+}, the orange dashed line indicates the experimentally determined CCS of SP^{3+} ($316A^2$), and blue circles indicate the number of water molecules in the droplet at a given time. A and B start with same starting structures, with extended SP and same coordinates for chloride and hydronium, while C starts with extended SP but having different hydronium and chloride coordinates. From 0 ps to 4200ps the temperature was set to 360 K and from 4200 ps temperature was set to 420 K to expedite the evaporation of water molecules. Simulation started with 22 hydronium ions and 10 chloride ions. For structure inserts, blue dots represent water molecules, purple spheres represent chloride ions, and hydronium ions are shown in green.

The CCSs for SP^{3+} ions show two abrupt changes during the dehydration process. First is at ~560 ps when vacuum exposure of the hydrophobic region of the peptide results in a coiling of this region, (**Figure 32 A**). As water droplet size is reduced, the hydrophilic N-terminus of SP^{3+} remained buried within the droplet, while the hydrophobic C-terminus region of the peptide remains exposed to vacuum after ~560 ps, resulting in coiling of the hydrophobic region. As further evaporation occurs, the CCS of SP^{3+} ion steadily decreases until ~1800 ps. Subsequent changes in the CCS are the result of evaporation during the transition from ~50 water molecules to fewer than 10, which corresponds to the final steps associated with forming the solvent-free, gas-phase ion from the hydrated complex. The observed changes in conformation at this stage of the dehydration process are not surprising considering that this would correspond to transitions from water stabilizing the hydrophilic side chains to stabilization of these charged groups by intramolecular charge solvation. First collapse of the SP structure during 500~1500ps is due to two factors. First, reduction of the droplet decreases about 80% which results in smaller domain for the peptide solvation. Second, the peptide migrates to the surface of the droplet as the hydrophobic C-terminus coils on the surface of the water droplet. However the second reduction in CCS during 3-4.5 ns and gradual expansion in CCS in last 4.5-6ns are still unexplained.

We previously reported that intramolecular charge solvation involving the charge-carrying N-terminus, arginine (R1), lysine (K3) and the glutamines (Q5 and Q6) are essential for preserving the "kinetically-trapped", compact SP conformer.[3,15] To give some insight on why the CCS reduction is visible when the droplet contains ~100 water

molecules, more in depth analysis of the intermolecular interaction has to be investigated around this time. A more detailed analysis of the reduction in CCS is illustrated in **Figure 33**, which contains distance plots for candidate conformers (**Figure 33 A)** corresponding to the final stages of water and chloride ion loss leading to the final SP conformer. **Figure 33 A** shows the distance between the R1 guanidinium group carbon ($R1C^{\zeta}$) and the Q6 side-chain terminal carbon ($Q6C^{\delta}$). Stepwise fluctuations in distances from one colored region to another colored region indicate stepwise transformation in the structure. The k-Clust statistical analysis algorithm was performed on SP^{3+} structures, which indicated that colored regions have similar structures. **Figure 33 B** shows the best-fit conformers from clustering analysis for each colored region. The purple structure is representative of SP^{3+} when the distance between $R1C^{\zeta}$ and $Q6C^{\delta}$ is less than 9 Å, where a chloride ion and a water molecule are found to be bridging the protons on the guanidinium ion and the Q6 side chain amide group. When the chloride ion is displaced by a water molecule, as shown in **Figure 33 B**, the distance between R1 and Q6 is > 9 Å. It appears that SP^{3+} interconverts between the two structures represented by purple and red until ~3800 ps, at which time an irreversible transition occurs that leads to the group of structures represented in green. Chloride ions that were previously located near the guanidinium group migrate within the SP^{3+} and water network to K3 and the N-terminus, which reflects preference for the point charge of K3 over the diffuse charge of R1. The last transition occurs at ~4000 ps, leading to a structure represented in the blue region where hydrogen bonds form between side-chains on R1 and Q6 ($R1C^{\zeta}$ -$Q6C^{\delta}$ distance is <5Å). The structures in the blue region have the smallest CCS compared to the other three regions. At ~400 ps after the formation

of intramolecular charge solvation, the calculated CCS of bare SP^{3+} ion observed at 4400 ps (**Figure 33 C**) approaches the experimentally measured CCS of 316 $Å^2$;[15] however, for the remainder of the simulation the CCS fluctuates around 325 $Å^2$. It appears that the increase in CCS is due to changes in the position of the hydronium and chloride ions. At ~4400 ps, hydronium and chloride ions are distributed around the charge sites of the peptide and the hydrophilic residues, but as the simulation progresses, these ions relocate to the N-terminus and K3 ammonium group, and disrupt the coiling of the peptide backbone. **Figure 33 D** shows the final conformer of the simulation that illustrates the elongation, or loss of intramolecular interactions, that still favors the compact conformer. At this stage two water molecules, three hydronium ions, and five chloride ions reside near the SP^{3+} ions.

Replicate simulations have revealed that the interactions shown in **Figure 33** has to be formed to create the compact conformer of SP. The overall peptide structure decreases when the guanidinium ion forms intramolecular interaction with core of the peptide especially Q's and the c-terminus. However, some replicate simulations did not show the intramolecular interaction with arginine and glutamine. Replicate simulation shown in **Figure 32 C** and **Figure 34 D** do not show this interaction and does not produce any structure below 330 $Å^2$. In addition, the hydronium and chloride remain scattered around the peptide as the peptide is elongated. The structure of the peptide is more influenced by the location of the peptide in the droplet, HCl ion location, and the evolution of droplet shape associated with the droplet fission.

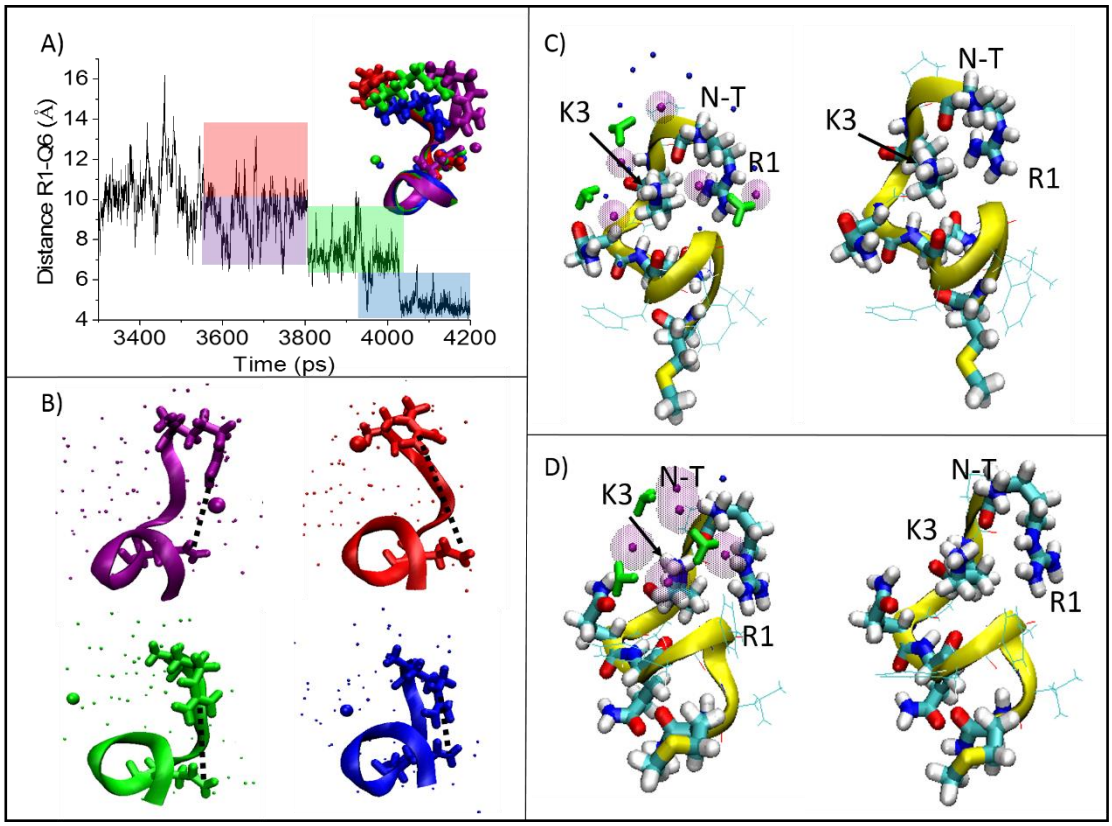

Figure 33. (A) Distance plot between R1Cζ and Q6Cδ during the time period 3300 ps to 4200 ps, and (B) representative structures for the color-coded regions for the trajectory shown in (A). A single Cl^- ion is shown as the large sphere, and the other Cl^- ions are not shown because they do not appear to participate in this interaction; water molecules are shown as small dots. The distance between two carbons is labeled with black dashed lines. Structures are color coded with respect to the colored regions in panel (A). Structural differences between smallest SP^{3+} ion at 4393 ps (CCS = 312 $Å^2$) and larger CCS at 6000 ps (CCS = 328 $Å^2$) is shown in panel (C). The SP^{3+} ion conformer shown in panel (A) has a more compact N-terminal region when compared to the conformer shown in panel (D).

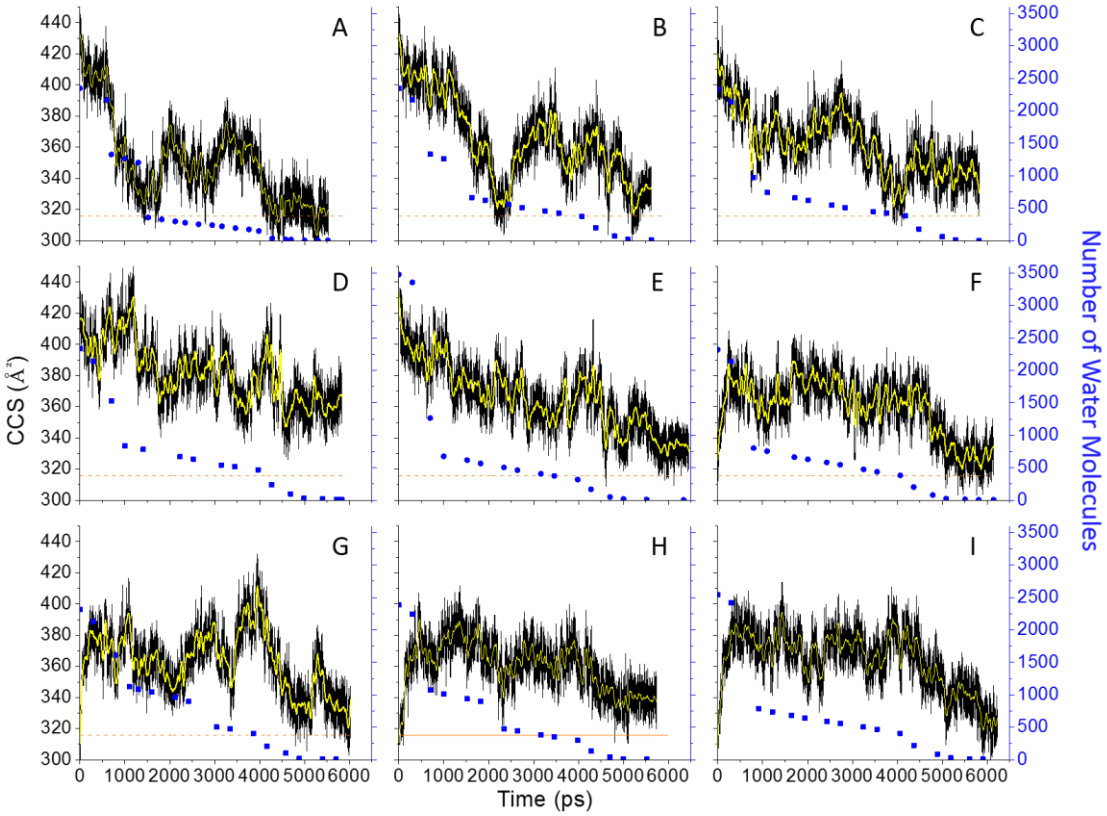

Figure 34. Nine replicate simulation plot is shown. 8 of the replicate simulations indicate that SP CCS decreases approximately 330Å^2 which is 4% of the ESI-IM-MS experimental error for compact conformer. A and B are same replicates as Figure 1 A and B. C, D, and E are the replicates starting with extended SP structure while chloride and hydronium ions are placed randomly. F-I are the replicates starting with compact SP structure while chloride and hydronium ions are placed randomly.

Experimental Results Support the Retention of Chloride Adducts after Desolvation

Although the simulations suggest that interactions between hydrated SP ions and Cl^- ions are retained until the final stage of ESI droplet evaporation, these adduct ions are not observed in the ESI mass spectrum. There is, however, direct evidence that the presence of Cl^- influences the ESI process. For example, the mass spectrum of SP sprayed from a 0.1% HCl solution contains abundant SP^{3+} and SP^{2+} ions (**Figure 35 A**), and the increased abundance (relative to that for ESI from water) of the SP^{2+} ions is a direct result of a Cl^- ion induced charge reduction reaction. Chait et al.[95] observed similar effects from acidic solutions, and Cole et al.[106] and Mancini[107] have suggested possible mechanisms by which charge reduction reactions occur, including loss of individual chloride ions, neutral HCl, and/or $H^+(H_2O)_nCl^-$ clusters. We interpret these results as evidence that interactions of SP with Cl^- ions are weak, thus the reactions leading to loss of HCl and/or $H^+(H_2O)_nCl^-$ are favored. We expanded this study by measuring the effects of Cl^- ions on the ESI product ion yields of the amphipathic peptide melittin. For melittin we observed both products of charge reduction reactions as well as peptide-Cl^- adduct ions. (**Figure 35 B**). We attribute the differences between SP and melittin to the higher number of hydrophilic amino acid side chains, which increases the overall number of charge carriers as well as anion binding sites.

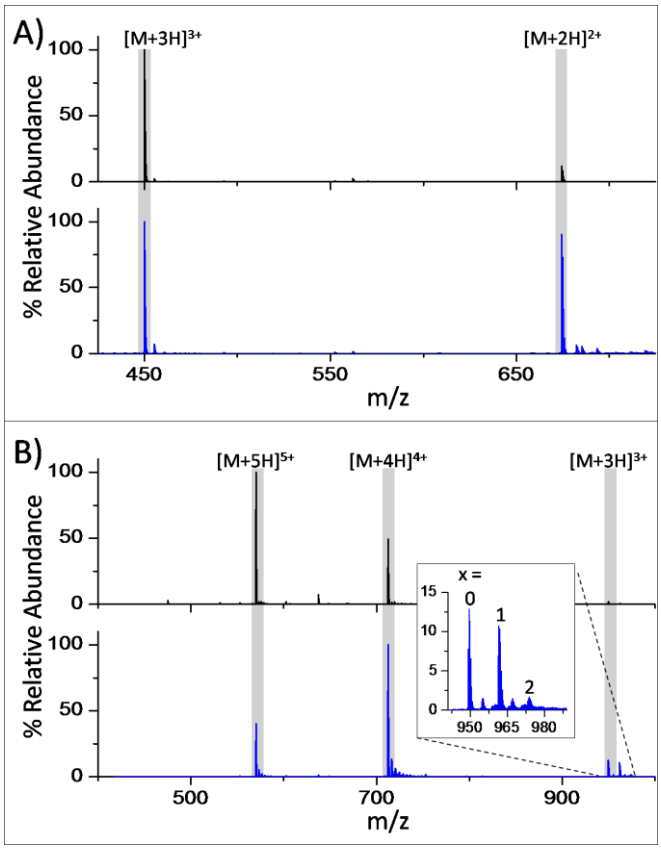

Figure 35. (A) The ESI-mass spectra of SP resulting from water (top) and water/0.1% HCl (bottom) showing the charge reduction induced by the presence of Cl. (B) The ESI-mass spectra of melittin (GIGAVLKVLTTGLPALISWIKRKRQQ-NH_2) resulting from water (top) and water/ 0.1% HCl (bottom). Chloride adducted melittin is observed with a total charge of 3+ when HCl is added to the sample solution (inset).

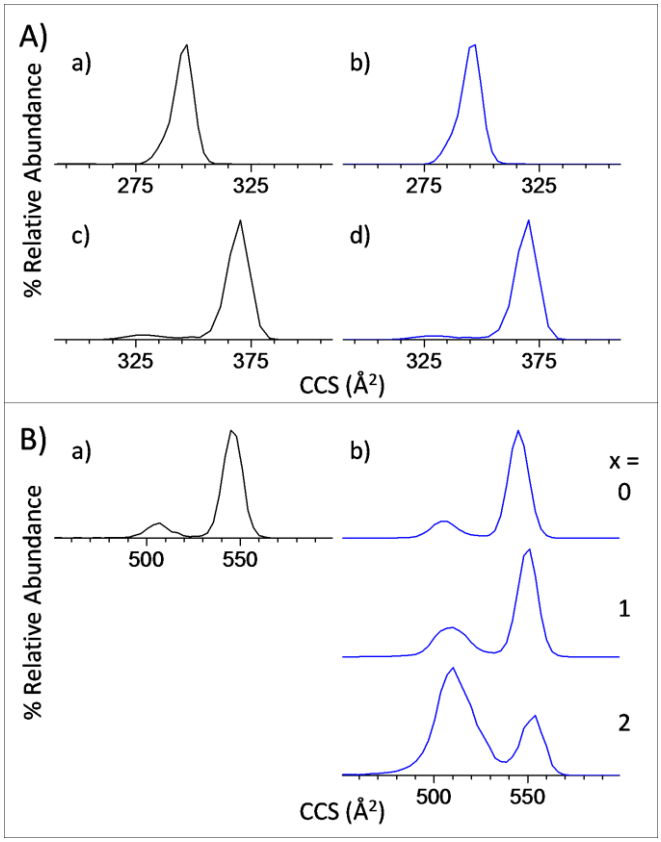

Figure 36. (A) CCS profiles observed for SP2+ [SP + 2H]2+ (a,b) and SP3+ (c,d) ions electrosprayed from water (a,c) and water/0.1% HCl (b,d). (B) CCS profiles observed for melittin $[M + 3H]^{3+}$ ions electrosprayed from water (a) and $[M + nH + xCl]^{3+}$ ions electrosprayed from water/0.1% HCl (b). Each $[M + nH + xCl]^{3+}$ ion in (b) is labeled for the number of chloride adducts (0-2).

We are most interested in how the presence of Cl^- ions affects peptide ion conformational preferences using IM-MS. The CCS profiles for SP and melittin electrosprayed from water and water/0.1% HCl solutions indicate similar conformational preferences (**Figure 36**). It is interesting to note, however, that higher abundances of compact conformers are observed for the melittin-Cl^- adduct ions (**Figure 36 B**). Melittin has been the subject of several ion mobility studies.[108-111] Barran et al. reported a single CCS value of 544 $Å^2$ for the $[M + 3H]^{3+}$ ion.[110] More recently, May and McLean have shown that the $[M + 3H]^{3+}$ ions are composed of three conformer families, compact (CCS ~410 $Å^2$), intermediate (CCS ~485 $Å^2$), and extended (CCS ~515 $Å^2$) conformers.[111] The observed CCS profiles for $[M + 3H]^{3+}$ ions (**Figure 36 B**) indicate two distinct conformer populations, similar to those reported.[110,111] It is interesting to note the small increase in the abundance of the compact conformer for the $[M + 4H + 1Cl]^{3+}$ ion; however, a large increase in the abundance of the compact conformer is observed for the $[M + 5H + 2Cl]^{3+}$ ions. It is generally agreed that the extended conformer of melittin is the gas phase equilibrium structure having a high degree of α-helicity.[110,111] Thus, the compact conformer is best described as a collapsed coil; specifically, Cl^- facilitates and stabilizes intramolecular interactions resulting in a compact conformation, which is similar to that observed with SP.

Conclusions

MD simulations of the dehydration of ESI-formed droplets containing SP^{3+} ion, chloride, and hydronium ions reveal new insights about peptide ion conformational changes in the final stages of dehydration. The overall mechanism by which ions are

produced is best described by combining the charge residue model and field emission model as suggested by Gross et al.[79] Solvent evaporation, droplet fission and ejection of small hydrated ions, *i.e.*, hydrated hydronium ions, occurs throughout the transitions from a large, bulk-like water droplet to formation of the solvent-free ion.[79] Early in the droplet lifetime hydronium ions migrate to the droplet surface and the surface charging results in emission of hydrated hydronium ions and distortion of the droplet resulting in droplet fission. Conformational change of the SP^{3+} ion is also observed. When fully immersed in the droplet, SP^{3+} retained an extended conformation, but as the droplet shrinks, the hydrophobic C-terminus migrates to the droplet surface and eventually it folds into a compact conformer as droplet size decreases. Replicate simulations also indicate that arginine interaction with the core of the peptide is essential to the creation of compact conformer during desolvation process. Without the arginine interaction, the peptide remains elongated. These compact conformer simulation results are in good agreement with previously reported solvent-free CCS values obtained by IM-MS experiments. It also appears that chloride ions promote formation of intramolecular charge solvation of the SP^{3+} ion. The changes in SP^{3+} ion CCS reveal that the structure of the SP^{3+} ion changes owing to changes in the droplet volume.

MD simulations also show that the peptide-cation/anion cluster survives until the end of the desolvation process, where the SP^{3+} ion clusters with hydronium and chloride ions. ESI-IM-MS experiments also show that charge reduction involving loss of chloride ions (as HCl) for SP^{3+}, and chloride adduct favoring compact conformation was observed with the larger peptide melittin. Overall, this agreement in simulations and measurements

can be understood microscopically as favored electrostatic interaction that helps stabilize the biopolymer during the final stages of solvent removal.

CHAPTER V

CONCLUSION

AMBER Molecular dynamics simulation package was used to develop a systematical approach to produce a library of candidate structures, and find the best fit candidate structures for the gas phase ions characterized by ESI-IM-MS experiments. With experimental results such as CCS, ECD, and solution phase NMR structures can be used as simulation boundaries condition to start the simulation and filter the simulated structures. The filtered structures can better describe the experiment results. This combination of theoretical and experimental techniques can provide structural information of biomolecules faster and easier than compared to X-ray crystallography and NMR. This research also probed extensively into electrospray desolvation mechanism, where experimental techniques cannot provide conclusive experimental evidence. Previous simulation studies performed by various groups focused mainly with water droplets containing mainly cations. This research included not only water molecules and hydronium ions but also large biomolecules and Cl^- counter ions. The simulation started with an extended structure solvated by a water droplet, and the structure evolution of SP^{3+} ions was monitored as the surrounding environment changed. At the later stages of the desolvation simulation, the CCS of SP^{3+} ions have approached the previously reported kinetically trapped compact conformer CCS value of 316 $Å^2$. The desolvation process of the entire droplet was consistent with the combined charge residue-field emission model. In addition, for the SP^{3+} ion alone, the simulation outcome resemble that the electrospray desolvation mechanism is similar to salting out mechanism of biomolecules.

CHAPTER VI

FUTURE DIRECTIONS

Simulation method described in previous chapters can be implemented to different classes of biomolecules. Retro-SP and SP Q to A mutant desolvation simulation studies are rational future topic for desolvation simulation. Studies by Servage et al. and Fort et al. have reported ESI-cryo-IM-MS results for dehydration of SP Q to A mutant and retro-SP IM results which can be used as desolvation simulation boundary conditions.[57,90] The study have indicated that compact conformer which has been observed from wild type SP is not observed but rather extended conformer is detected for Q-mutants and intermediate conformer is detected for retro-SP. Desolvation simulation should be able to indicate which charge solvation interaction was lost and ultimately resulting more extended conformers. Next step from SP mutant, non-covalent ubiquitin dimer desolvation can be simulated on the larger system. Recently, Wagner et al. have reported formation non-covalent dimer of ubiquitin in the presence of chloride.[94] Using the CCS values and the charge state information from the experiment the desolvation simulation can be performed. The computational time will be much greater since the ubiquitin is about 6 times larger than SP and the solvation shell has to be larger to accommodate two ubiquitins. Desolvation simulation with various starting structure will be able to identify which interactions are responsible for the formation of non-covalent di-ubiqutin. As with the original SP desolvation simulation study, each model desolvation simulation will be replicated minimum of 12 times with various starting structures and hydronium and chloride locations.

In Recent years, "native" or "native-like," mass spectrometry has been paying attention from mass spectrometry community. The native mass spectrometry studies the structures of protein and/or protein complexes starting from a physiological conditions.[21,112,113] Typically a volatile buffer such as ammonium acetate buffer is used to keep the protein folded while the volatile ammonium acetate evaporates quickly during the desolvation process. From this, next step in the desolvation simulation replacing HCl to ammonium acetate in the simulation has to be considered. The typical buffer concentration used in the native mass spectrometry experiments range from 50mM to 200mM.[113,114] Therefore, the desolvation simulation should be performed at various concentrations of buffer because the concentration can affect the desolvation process. The ammonium acetate simulation will not have the same results compared to the results from HCl because the binding strength and coordination of the anions will be different. For better understanding of the native mass spectrometry, desolvation simulation containing ammonium acetate is essential to understand which interaction is preserved and made by ammonium acetate.

Desolvation simulation methodology will be needed to improve for further desolvation simulation studies. First, the starting structure of the simulation has to be improved. The solution phase structure of biomolecules are provided in PDB Database such as RCSB PDB, but some of the biomolecules have only crystal structures. For those cases, the starting structures should be optimized first in bulk solvent simulation. After the optimization, the system should be submitted for desolvation simulation. Second, the water droplet size has to increase for larger biomolecules. The increased droplet size will

eventually lead to increase in the simulation box. The simulation will take much longer time with increased volume, droplet size, and biomolecules. Third, the simulation temperature has to be reconsidered. Currently the desolvation simulation was performed at 360K, but later the temperature is increased to 420 K for enhanced evaporation as the smaller droplet is slower to evaporate. Currently, the increased temperature does not appear to affect the desolvation process, but future desolvation simulations can be affected by the increased temperature. Furthermore, the droplet temperature will not be at the heated capillary (desolvation region) temperature of ~360 K due to evaporative cooling. Therefore, desolvation temperature of ~300K can be used to test and compare the simulation outcome with 360 K and determine the better desolvation temperature. Forth, the water evaporation cut off have to be reevaluated. Currently, evaporated water molecule removal is removed by observation, but the water removal should be performed by a systematical approach in the future studies. Fifth, the model of water should be changed to TIP4 or TIP5 water model as the TIP3 water model is older model compared to TIP4 and TIP5. It should be noted that the scripts have to change when the water model change. Sixth, mixed solvent system has to be developed. Organic solvents are used to denature the protein, and to model those experiments mixed solvent system has to be developed. Comparison between mixed solvent and "native" state simulation results can determine how the organic solvents interact with the biomolecules at the later stages of desolvation.

REFERENCES

(1) Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A. *Quarterly Reviews of Biophysics* **2007**, *40*, 191-285.

(2) Uversky, V. N. *Protein Sci.* **2002**, *11*, 739-756.

(3) Jelsch, C.; Teeter, M. M.; Lamzin, V.; Pichon-Pesme, V.; Blessing, R. H.; Lecomte, C. *Proc. Natl. Acad. Sci.* **2000**, *97*, 3171-3176.

(4) Schmidt, A.; Teeter, M.; Weckert, E.; Lamzin, V. S. *ACTA CRYSTALLOGR. F* **2011**, *67*, 424-428.

(5) Martz, E. *RCSB Proten Data Bank*, *http://www.rcsb.org/pdb/static.do?p=general_information/about_pdb/nature_of_3d_structural_data.html#Contents*, Jan. 20 2016.

(6) Adamski, C. J.; Cardenas, A. M.; Brown, N. G.; Horton, L. B.; Sankaran, B.; Prasad, B. V.; Gilbert, H. F.; Palzkill, T. *Biochemistry* **2015**, *54*, 447-457.

(7) Anglister, J.; Grzesiek, S.; Ren, H.; Klee, C. B.; Bax, A. *J. Biomol. NMR* **1993**, *3*, 121-126.

(8) Lapthorn, C.; Pullen, F.; Chowdhry, B. Z. *Mass Spectrom. Rev.* **2013**, *32*, 43-71.

(9) Baştuğ, T.; Kuyucak, S. *Biophys. J.* **2003**, *84*, 2871-2882.

(10) Kebarle, P.; Verkerk, U. H. *Mass Spectrom. Rev.* **2009**, *28*, 898-917.

(11) Meyer, T.; Gabelica, V.; Grubmüller, H.; Orozco, M. *WIREs Comput. Mol. Sci.* **2013**, *3*, 408-425.

(12) Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. *J. Chem. Phys.* **1968**, *49*, 2240-2249.

(13) Iribarne, J. V.; Thomson, B. A. *J. Chem. Phys.* **1976**, *64*, 2287-2294.

(14) Gamero-Castaño, M.; Mora, J. F. d. l. *J. Mass Spectrom.* **2000**, *35*, 790-803.

(15) Silveira, J. A.; Fort, K. L.; Kim, D.; Servage, K. A.; Pierson, N. A.; Clemmer, D. E.; Russell, D. H. *J. Am. Chem. Soc.* **2013**, *135*, 19147-19153.

(16) Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E. *Nature Chem.* **2014**, *6*, 281-294.

(17) Jarrold, M. F. *Annu. Rev. Phys. Chem.* **2000**, *51*, 179-207.

(18) Wyttenbach, T.; Bowers, M. In *Modern Mass Spectrometry*; Schalley, C., Ed.; Springer Berlin Heidelberg: 2003; Vol. 225, p 207-232.

(19) Wyttenbach, T.; Pierson, N. A.; Clemmer, D. E.; Bowers, M. T. *Annu. Rev. Phys. Chem.* **2014**, *65*, 175-196.

(20) Konijnenberg, A.; Butterer, A.; Sobott, F. *Biochim. Biophys. Acta* **2013**, *1834*, 1239-1256.

(21) Uetrecht, C.; Rose, R. J.; van Duijn, E.; Lorenzen, K.; Heck, A. J. R. *Chem. Soc. Rev.* **2010**, *39*, 1633-1655.

(22) Harvey, S. R.; MacPhee, C. E.; Barran, P. E. *Methods* **2011**, *54*, 454-461.

(23) Lee, D.; Redfern, O.; Orengo, C. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 995-1005.

(24) M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople; Gaussian, Inc.: Wallingford, CT, USA, 2004.

(25) D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R.; Luo, R. C. W., W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg,; G. Seabra, I. K., K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell,; T. Steinbrecher, H. G., Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H.; Mathews, M. G. S., C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and; Kollman, P. A. *University of California, San Francisco.* **2010**.

(26) Shvartsburg, A. A.; Jarrold, M. F. *Chem. Phys. Lett.* **1996**, *261*, 86-91.

(27) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. *J. Phys. Chem.* **1996**, *100*, 16082-16086.

(28) Bleiholder, C.; Wyttenbach, T.; Bowers, M. T. *Int. J. Mass spectrom.* **2011**, *308*, 1-10.

(29) Anderson, S. E.; Bleiholder, C.; Brocker, E. R.; Stang, P. J.; Bowers, M. T. *Int. J. Mass spectrom.* **2012**, *330–332*, 78-84.

(30) Bleiholder, C.; Contreras, S.; Do, T. D.; Bowers, M. T. *Int. J. Mass spectrom.* **2013**, *345–347*, 89-96.

(31) Bleiholder, C.; Contreras, S.; Bowers, M. T. *Int. J. Mass spectrom.* **2013**, *354–355*, 275-280.

(32) D'Atri, V.; Porrini, M.; Rosu, F.; Gabelica, V. *J. Mass Spectrom.* **2015**, *50*, 711-726.

(33) F.-Y. Dupradeau, A. P., T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski & P. Cieplak *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821-7839.

(34) Kumar, S.; Nussinov, R. *ChemBioChem* **2002**, *3*, 604-617.

(35) Kim, D.; Pai, P.-J.; Creese, A.; Jones, A.; Russell, D.; Cooper, H. *J. Am. Soc. Mass. Spectrom.* **2015**, *26*, 1004-1013.

(36) Pierson, N. A.; Chen, L.; Valentine, S. J.; Russell, D. H.; Clemmer, D. E. *J. Am. Chem. Soc.* **2011**, *133*, 13810-13813.

(37) Pierson, N. A.; Chen, L.; Russell, D. H.; Clemmer, D. E. *J. Am. Chem. Soc.* **2013**, *135*, 3186-3192.

(38) Chen, S.-H.; Russell, D. *J. Am. Soc. Mass. Spectrom.* **2015**, *26*, 1433-1443.

(39) Chen, L.; Gao, Y. Q.; Russell, D. H. *J. Phys. Chem. A* **2012**, *116*, 689-696.

(40) Tao, L.; Dahl, D.; Pérez, L.; Russell, D. *J. Am. Soc. Mass. Spectrom.* **2009**, *20*, 1593-1602.

(41) Schmidt, S. R.; Schweikart, F.; Andersson, M. E. *J. Chromatogr. B* **2007**, *849*, 154-162.

(42) Gharbi-Benarous, J.; Bertho, G.; Evrard-Todeschi, N.; Coadou, G.; Megy, S.; Delaunay, T.; Benarous, R.; Girault, J. P. *Biochemistry* **2004**, *43*, 14555-14565.

(43) Lee, N. Y.; Koland, J. G. *Protein Sci.* **2005**, *14*, 2793-2803.

(44) Lau, L.-F.; Schachter, J. B. In *ACAG* 2004; Vol. 16, p 153-175.

(45) Iakoucheva, L. M.; Radivojac, P.; Brown, C. J.; O'Connor, T. R.; Sikes, J. G.; Obradovic, Z.; Dunker, A. K. *Nucleic Acids Res.* **2004**, *32*, 1037-1049.

(46) Thalassinos, K.; Grabenauer, M.; Slade, S. E.; Hilton, G. R.; Bowers, M. T.; Scrivens, J. H. *Anal. Chem.* **2009**, *81*, 248-254.

(47) Ruotolo, B. T.; Verbeck; Thomson, L. M.; Woods, A. S.; Gillig, K. J.; Russell, D. H. *J. Proteome Res.* **2002**, *1*, 303-306.

(48) Kumar, S.; Nussinov, R. *ChemBioChem* **2002**, *3*, 604-617.

(49) Donald, J. E.; Kulp, D. W.; DeGrado, W. F. *Proteins* **2011**, *79*, 898-915.

(50) Sokalingam, S.; Raghunathan, G.; Soundrarajan, N.; Lee, S.-G. *PLOS One* **2012**, *7*, e40410.

(51) Kubickova, A.; Krizek, T.; Coufal, P.; Wernersson, E.; Heyda, J.; Jungwirth *J. Phys. Chem. Lett.* **2011**, *2*, 1387-1389.

(52) Meot-Ner (Mautner), M. *Chem. Rev.* **2005**, *105*, 213-284.

(53) Green, M. K.; Lebrilla, C. B. *Mass Spec. Rev.* **1997**, *16*, 53-71.

(54) Steinberg, M. Z.; Elber, R.; McLafferty, F. W.; Gerber, R. B.; Breuker, K. *ChemBioChem* **2008**, *9*, 2417-2423.

(55) Moss, C. L.; Chung, T. W.; Wyer, J. A.; Nielsen, S. B.; Hvelpund, P.; Turecek, F. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 731-751.

(56) Prell, J. S.; O'Brien, J. T.; Steill, J. D.; Oomens, J.; Williams, E. R. *J. Am. Chem. Soc.* **2009**, *131*, 11442-11449.

(57) Fort, K. L.; Silveira, J. A.; Pierson, N. A.; Servage, K. A.; Clemmer, D. E.; Russell, D. H. *J. Phys. Chem. B* **2014**, *118*, 14336-14344.

(58) Brant, D. A.; Schimmel, P. R. *Proc. Natl. Acad. Sci.* **1967**, *58*, 428-435.

(59) Jabs, A.; Weiss, M. S.; Hilgenfeld, R. *J. Mol. Biol.* **1999**, *286*, 291-304.

(60) Fischer, G. *Chem. Soc. Rev.* **2000**, *29*, 119-127.

(61) Schimmel, P. R.; Flory, P. J. *Proc. Natl. Acad. Sci.* **1967**, *58*, 52-59.

(62) Traub, W.; Shmueli, U. *Nature* **1963**, *198*, 1165-1166.

(63) Swenson, C. A.; Formanek, R. *J. Phys. Chem.* **1967**, *71*, 4073-4077.

(64) Forsythe, K. H.; Hopfinger, A. J. *Macromolecules* **1973**, *6*, 423-437.

(65) Shi, L.; Holliday, A. E.; Shi, H.; Zhu, F.; Ewing, M. A.; Russell, D. H.; Clemmer, D. E. *J. Am. Chem. Soc.* **2014**, *136*, 12702-12711.

(66) Counterman, A. E.; Clemmer, D. E. *J. Phys. Chem. B* **2004**, *108*, 4885-4898.

(67) Breuker, K.; McLafferty, F. W. *Proc. Natl. Acad. Sci.* **2008**, *105*, 18145-18152.

(68) Loo, J. A. *Int. J. Mass spectrom.* **2000**, *200*, 175-186.

(69) Shi, H.; Atlasevich, N.; Merenbloom, S.; Clemmer, D. *J. Am. Soc. Mass. Spectrom.* **2014**, *25*, 2000-2008.

(70) Wyttenbach, T.; Bowers, M. T. *J. Phys. Chem. B* **2011**, *115*, 12266-12275.

(71) Hamdy, O.; Julian, R. *J. Am. Soc. Mass. Spectrom.* **2012**, *23*, 1-6.

(72) Cech, N. B.; Enke, C. G. *Mass Spectrom. Rev.* **2001**, *20*, 362-387.

(73) Gomez, A.; Tang, K. *Phys. Fluids* **1994**, *6*, 404-414.

(74) Fenn, J.; Mann, M.; Meng, C.; Wong, S.; Whitehouse, C. *Science* **1989**, *246*, 64-71.

(75) Marginean, I.; Znamenskiy, V.; Vertes, A. *J. Phys. Chem. B* **2006**, *110*, 6397-6404.

(76) Liuni, P.; Wilson, D. J. *Expert Rev. Proteomics* **2011**, *8*, 197-209.

(77) Mehmood, S.; Allison, T. M.; Robinson, C. V. *Annu. Rev. Phys. Chem.* **2015**, *66*, 453-474.

(78) Steinberg, M. Z.; Breuker, K.; Elber, R.; Gerber, R. B. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4690-4697.

(79) Hogan, C. J.; Carroll, J. A.; Rohrs, H. W.; Biswas, P.; Gross, M. L. *Anal. Chem.* **2009**, *81*, 369-377.

(80) Wyttenbach, T.; Liu, D.; Bowers, M. T. *Int. J. Mass spectrom.* **2005**, *240*, 221-232.

(81) Znamenskiy, V.; Marginean, I.; Vertes, A. *J. Phys. Chem. A* **2003**, *107*, 7406-7412.

(82) Konermann, L.; McAllister, R. G.; Metwally, H. *J. Phys. Chem. B* **2014**, *118*, 12025-12033.

(83) Ahadi, E.; Konermann, L. *J. Phys. Chem. B* **2009**, *113*, 7071-7080.

(84) Consta, S.; Mainer, K. R.; Novak, W. *J. Chem. Phys.* **2003**, *119*, 10125-10132.

(85) Ahadi, E.; Konermann, L. *J. Am. Chem. Soc.* **2010**, *132*, 11270-11277.

(86) Sharawy, M.; Consta, S. *J. Chem. Phys.* **2014**, *141*, 104321.

(87) König, S.; Fales, H. M. *J. Am. Soc. Mass. Spectrom.* **1998**, *9*, 814-822.

(88) Lee, S.-W.; Freivogel, P.; Schindler, T.; Beauchamp, J. L. *J. Am. Chem. Soc.* **1998**, *120*, 11758-11765.

(89) Servage, K. A.; Fort, K. L.; Silveira, J. A.; Shi, L.; Clemmer, D. E.; Russell, D. H. *J. Am. Chem. Soc.* **2015**, *137*, 8916-8919.

(90) Servage, K. A.; Silveira, J. A.; Fort, K. L.; Russell, D. H. *J. Phys. Chem. B* **2015**, *119*, 4693-4698.

(91) Chang, T. M.; Berden, G.; Oomens, J.; Williams, E. R. *Int. J. Mass spectrom.* **2015**, *377*, 440-447.

(92) Merenbloom, S.; Flick, T.; Daly, M.; Williams, E. *J. Am. Soc. Mass. Spectrom.* **2011**, *22*, 1978-1990.

(93) Han, L.; Hyung, S.-J.; Mayers, J. J. S.; Ruotolo, B. T. *J. Am. Chem. Soc.* **2011**, *133*, 11358-11367.

(94) Wagner, N. D.; Kim, D.; Russell, D. H. *Anal. Chem.* **2016**, *88*, 5934-5940.

(95) Mirza, U. A.; Chait, B. T. *Anal. Chem.* **1994**, *66*, 2898-2904.

(96) Daub, C. D.; Cann, N. M. *Anal. Chem.* **2011**, *83*, 8372-8376.

(97) McAllister, R. G.; Metwally, H.; Sun, Y.; Konermann, L. *J. Am. Chem. Soc.* **2015**, *137*, 12667-12676.

(98) Gayen, A.; Goswami, S. K.; Mukhopadhyay, C. *BBA-Biomembranes* **2011**, *1808*, 127-139.

(99) Baaden, M.; Burgard, M.; Wipff, G. *J. Phys. Chem. B* **2001**, *105*, 11131-11141.

(100) Vega, C.; de Miguel, E. *J. Chem. Phys.* **2007**, *126*, 154707.

(101) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.

(102) Peng, Y.; Swanson, J. M. J.; Kang, S.-g.; Zhou, R.; Voth, G. A. *J. Phys. Chem. B* **2015**, *119*, 9212-9218.

(103) Jungwirth, P.; Tobias, D. J. *Chem. Rev.* **2006**, *106*, 1259-1281.

(104) Silveira, J. A.; Servage, K. A.; Gamage, C. M.; Russell, D. H. *J. Phys. Chem. A* **2013**, *117*, 953-961.

(105) Malevanets, A.; Consta, S. *J. Chem. Phys.* **2013**, *138*, 184312.

(106) Liu, X.; Cole, R. *J. Am. Soc. Mass. Spectrom.* **2011**, *22*, 2125-2136.

(107) Mancini, J. S.; Bowman, J. M. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6222-6226.

(108) Figueroa, I. D.; Torres, O.; Russell, D. H. *Anal. Chem.* **1998**, *70*, 4527-4533.

(109) Figueroa, I.; Russell, D. *J. Am. Soc. Mass. Spectrom.* **1999**, *10*, 719-731.

(110) Florance, H. V.; Stopford, A. P.; Kalapothakis, J. M.; McCullough, B. J.; Bretherick, A.; Barran, P. E. *Analyst* **2011**, *136*, 3446-3452.

(111) May, J. C.; McLean, J. A. *PROTEOMICS* **2015**, *15*, 2862-2871.

(112) Zhang, H.; Cui, W.; Gross, M. L.; Blankenship, R. E. *FEBS Lett.* **2013**, *587*, 1012-1020.

(113) Benesch, J. L. P.; Ruotolo, B. T.; Simmons, D. A.; Robinson, C. V. *Chem. Rev.* **2007**, *107*, 3544-3567.

(114) van Duijn, E. *J. Am. Soc. Mass. Spectrom.* **2010**, *21*, 971-978.