2017

Gas-Phase Protein Structure Characterization with Ion Mobility Mass Spectrometry and Molecular Dynamics Simulations

Samaneh Ghassabi Kondalaji

Gas-Phase Protein Structure Characterization with Ion Mobility Mass Spectrometry and Molecular Dynamics Simulations

Samaneh Ghassabi Kondalaji

Dissertation submitted
to the Eberly College of Arts and Sciences
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Chemistry

Stephen J. Valentine, Ph.D., Committee Chairperson
Peter M. Gannett, Ph.D.
John B. Mertz, Ph.D.
Glen P. Jackson, Ph.D.
Lisa A. Holland, Ph.D.

C. Eugene Bennett Dept. of Chemistry

Morgantown, West Virginia
2017

Keywords: electrospray ionization, ion mobility-mass spectrometry, gas-phase hydrogen-deuterium exchange, molecular dynamics simulation, collision cross section, conformational space sampling

ABSTRACT

Gas-Phase Protein Structure Characterization with Ion Mobility Mass Spectrometry and Molecular Dynamics Simulations

Samaneh Ghassabi Kondalaji

Over the last few decades, the widespread application of soft ionization techniques such as electrospray ionization (ESI) coupled with mass spectrometry (MS) facilitated the characterization of large biomolecules in the gas phase. Under suitable conditions and during the short timescale of the desolvation process associated with ESI (10^{-12}-10^{2} s), the solvent-free protein ions preserve significant portions of their native solution-phase structure and are presumed to be kinetically protected from thermodynamic destabilization such as unfolding-refolding processes occurring in the gas phase that could yield an “inside-out” structure.

The combination of ESI with ion mobility spectrometry- mass spectrometry (IMS-MS) provides information regarding higher-order structure of proteins and their complexes. IMS-MS has the ability to resolve and probe conformers of protein ions of a particular charge state based on their mobility through an inert buffer gas in drift tube under the influence of a constant electric field. An ion’s mobility is dependent upon its charge and shape and can be used to determine the orientationally-averaged collision cross section (CCS) of the ion with the buffer gas. Comparison of experimentally measured CCS values with theoretical CCS numbers obtained from a series of computer-assisted molecular dynamics simulations (MD) which explore the conformational space of protein ions provides insight into candidate structures for each conformer.

Although, CCS values provide a rough estimate of the overall shape of different molecules, IMS-MS alone cannot distinguish three dimensional structures of a series of conformers that represent the same mobility and thus would arrive at the same time at the exit region of the drift tube. The combination of IMS-MS with hydrogen-deuterium exchange (HDX) experiments as a gas-phase chemical probe of structure provides the opportunity to distinguish among these conformer types.

In this work, the gas-phase ion conformers of a model peptide have been studied through CCS measurements, HDX behavior analysis and extensive (MD) simulations.

Initially, an advanced protocol is introduced in order to achieve an impartial sampling of phase space targeting both higher-energy and more thermodynamically-stable structures. The gas-phase transport properties of the ion conformers - including their dynamics at experimental temperatures- have been monitored, and, combined with an optimized clustering and data mining method, accurate CCS determination has been accomplished. These data provided the first criterion to filter through a substantial pool of conformations in order to obtain a series of candidate structures (CCS matched) with significant structural variation.

A hydrogen accessibility scoring (HAS)-number of effective collisions (NEC) model is applied to the candidate structures obtained from MD simulations. The HAS-NEC model produced hypothetical, per-residue deuterium uptake values. This information then provided the overall structural contribution from each *in-silico* structure leading to the best match to experimental results. The comparison of predicted and experimentally observed isotopic envelopes of various mass spectral fragment ions supported the accuracy of the model. With these results, the hypothetical HDX data were employed as a second dimension to narrow the sampled phase space and, together with the accurate CCS values, 13 nominal conformers with specific population contributions to the gas-phase ions were selected.

In the final installment of this work, extensive simulations of the ESI process were performed to monitor the behavior of the peptide ion, charge carriers and the droplets involved in the ionization process. The results provided a series of structures that match the nominal conformers obtained through CCS calculations and the HAS-NEC model. This method validation confirmed the accuracy of the HAS-NEC model in successfully predicting the representative gas-phase structures on a computationally-affordable timescale.

Dedication

To my beloved husband, colleague and co-author, *Mahdiar*

For your love, encouragement and understanding of my ideas

To my parents, *Latifeh* and *Hossein*

For your unconditional love and support, from thousands of kilometers away

Acknowledgments

I would like to express my sincerest gratitude to each and every individual who has served as a mentor for my scientific and academic path. Dr. Stephen Valentine for your guidance, encouragement and incredible insight. You believed in me and worked with my strengths and weaknesses. I am proud and grateful to have you as my academic mentor.

I would like to thank the members of my graduate research committee, Dr. Peter Gannett, Dr. Blake Mertz, Dr. Lisa Holland and Dr. Glen Jackson for your invaluable support and mentorship. For always taking the time to answer my questions and directing me in my scientific and personal path. You taught me how to be a scientist and for that you are always to be admired.

I would also like to thank Dr. Justin Legleiter and Dr. Harry Finklea for always offering productive discussions during my graduate career. Your knowledge and support gave me direction for my future studies and for that I am grateful.

I would like to give a special thanks to Chitrak Gupta for the time, energy and commitment to assist in developing a part of the script utilized in the work presented in Chapter 4.

To my past and present lab mates, Dr. Jim Arndt, Dr. Greg Donohoe, Dr. Megan Maurer, Mahdiar Khakinejad, Hossein Maleki, Kushani Attanayake, Sandra Majuta. I am delighted that I had the pleasure to work with you in the Valentine research group and I wish each and every one of you success and countless accomplishments throughout your personal lives and scientific careers.

Table of Contents

List of Figures

List of Symbols and Abbreviations

ACE	Acetyl cap
AMBER	Assisted Model Building with Energy Refinement
BIRD	Blackbody Infrared Radiative Dissociation
CCS	Collison Cross Section
CD	Circular Dichroism
CEM	Chain Ejection Model
CID	Collision-Induced Dissociation
CRM	Charge Residue Model
E	Electric field
ECD	Electron Capture Dissociation
EHSS	Exact Hard Sphere Scattering
ESI	Electrospray Ionization
ETD	Electron Transfer Dissociation
F	Ion Funnel
FT-ICR	Fourier-Transform Ion Cyclotron Resonance
G	Ion Gate
GAMESS	General Atomic and Molecular Electronic Structure System
HAS	Hydrogen Accessibility Scoring
HDX	Hydrogen-Deuterium Exchange
HF	Hartree–Fock
H_R	Helicity
$H_{R(PA)}$	Population-averaged Helicity
IMS-MS	Ion Mobility-Mass Spectrometry
IEM	Ion Evaporation Model
IA	Ion Activation
K	Mobility

k_B	Boltzmann constant
L	Length
LIT	Linear Ion Trap
m/z	Mass to charge
MALDI	Matrix-Assisted Laser Desorption Ionization
mB	Mass of the Buffer gas
MD	Molecular Dynamics
MEP	Molecular Electrostatic Potential
m_I	Mass of ion
MMTSB	Multiscale Modeling Tools for Structural Biology
MOI	Moment of Inertia
MS	Mass Spectrometry
MS/MS	Tandem mass spectrometry
N	Number density
NEC	Number of Effective Collision
NMR	Nuclear Magnetic Resonance
NNLR	Non-Negative Linear Regression
NNLS	Non-Negative Least Squares
NP	Normalized Population
NPT	isothermal-isobaric ensemble
P	Pressure
PA	Projection Approximation
PES	Potential Energy Surface
PME	Particle mesh Ewald
P_s	Structural Population
r	Spherical radius
R	Residues involved in a helix
R.E.D.	RESP ESP charge Derive

R_{eff}	Effective Radius
R_{min}	Minimum allowed Radius
RMSD	Root Mean Square Deviation
S	Hydrogen Score value
SA	Simulated Annealing
STP	Standard Temperature and Pressure
T	Temperature
T_a	Ambient Temperature
t_D	Drift time
T_{eff}	Effective Temperature
TM	Trajectory Method
TWSS	Total Within-cluster Sum of Squares
VMD	Visual Molecular Dynamics
z	Charge
Z_r	Rayleigh charge
α	Asymmetry parameter
γ	Surface tension
ΔE	Potential Energy difference
ε_0	Vacuum permittivity
ρ	Probability function
u_d	Drift velocity
Ω	Collison cross section
Ω^*	Accurate collision cross section
Ω_{total}	Weighted-average collision cross section

1. Introduction: Mass Spectrometry (MS) and Protein Structure Studies

1.1. Electrospray Ionization*, "Wings for Molecular Elephants"* [1]

The advent of electrospray ionization (ESI) as a soft ionization technique along with matrix-assisted laser desorption ionization (MALDI) provided an opportunity to accomplish mass and structural determination of large biological molecules including peptides and proteins [1-3]. Using a variety of fragmentation techniques such as Electron transfer dissociation (ETD) or collision-induced dissociation (CID), the primary structures of proteins were confirmed [4]; however, due to the undeviating relationship between structures and functions of protein biomolecules, shortly later a question arose regarding the characterization of higher-order structure of proteins and their complexes. That is, does the transition from the solution phase to the gas phase environment during the ESI process and disruption in the weak non-covalent interactions between solvent and biopolymer dramatically alter protein structure? Or, at some level, can the process preserve the native protein conformation? A study by Smith, *et al.,* in 1994 on multimeric protein complexes illustrated that utilizing biologically relevant pH values and operating under mild mass spectrometer interface conditions such as ambient temperature, the gas-phase ions generate lower charge state values – high mass-to-charge ratio (*m/z*) – with less diversity in their charge state distribution [5]. These observations were interpreted as the generation of protein complex ions that retained their intact quaternary structure with conformational constraints that explained the uniformity in charge state values and compactness characterized by the decrease in value of the maximum charge on the protein due to coulombic

repulsion as dictated by the Rayleigh limit [5, 6]. Secondary and tertiary structural elucidation studies performed by ion mobility mass spectrometry suggested that the ESI-generated anhydrous ions preserve significant elements of their solution phase structure and, with these observations, the concept of native electrospray was introduced [7-9]. Further studies even proposed that because of the short timescale of the ESI process, the protein ions do not resemble the equilibrated conformations in the gas phase; rather, they are higher-energy, metastable structures that represent the solution-phase conformations [9-11].

Currently the intact structures of biomolecules can be investigated by a variety of methods performed on mass spectrometry instrumentation. IMS-MS, CID, Blackbody Infrared Radiative Dissociation (BIRD) and mass spectroscopy represent physical probe techniques [12-15] In contrast, hydrogen-deuterium exchange (HDX) and proton transfer reactions are chemical reactivity probes that provide insight into different aspects of higher-order structure of proteins and their complexes [13, 14, 16].

1.1.1. Mechanism of Gas-phase Ion Production during the ESI process

Upon the ESI process, the applied high electric field leads to polarization of the solvent at a capillary tip (ESI needle) and forces the charged particles to move under the influence of the field [17, 18]. In positive ion mode analysis, as this separation of opposite charges occurs, the enrichment of positively charged ions takes place at the solvent meniscus and subsequently leads to its distortion into a cone-shape geometry known as the Taylor cone [19]. This causes a shift in the surface area which is ultimately opposed by the surface tension of the liquid. This

change in surface area along with the accumulation of positive charges leads to the emission of a fine jet emerging from the cone toward the counter electrode located at the entrance of the mass spectrometer [17, 18]. Since the jet is carrying an excess of positive charge, the repulsion between the ions eventually destabilizes the jet into a spray of small droplets. During the flight of these droplets toward the opposite electrode, the evaporation of the solute molecules decreases the radius of the droplets. The maximum number of net charges (Z_r) that can be present in a droplet without destabilizing it can be determined through the Rayleigh limit [6] as given by:

$$z_r = \frac{8\pi}{e}\sqrt{\varepsilon_0 \gamma r^3} \qquad Eq.1-1.$$

In Equation 1-1, r is the radius of the spherical droplet, ε_0 is the vacuum permittivity with a value of 8.83 $\times 10^{-12}$ $C^2.J^{-1}$, γ is the surface tension of the solvent at the surface of the droplet at a specific temperature, and e is the elementary charge.

As the droplet shrinks to a specific radius, the coulombic repulsion between the ions overcomes the surface tension of the liquid and causes a fission event and a formation of smaller droplets [18]. The newly-generated droplets undergo repeated cycles of shrinkage-fission events [18, 20]. These cycles lead to the formation of progeny nanodroplets that eventually release the solvent-free ions at the final stages of the ESI process [18, 20].

Despite the enormous application of ESI for biological sample studies, the exact mechanism by which the transition of the analyte ions from the solution-phase to the gas-phase environment occurs has been the subject of debate for

over a decade. However, the two most probable mechanisms – Dole's Charge Residue Model (CRM) [21] and the Ion Evaporation Model (IEM)- [22] have been suggested to explain the behavior of the analyte in droplets at the late stages of ESI. In CRM, the evaporation-fission events repeatedly occur to a point where the removal of all water molecules leaves a desolvated, analyte ion behind [21, 23]. It has been proposed that the CRM takes place in nanodroplets carrying a same-sized (large) analyte molecule such as native, globular proteins [24]. In contrast, the IEM presents a scenario for direct release of a small, charged analyte ion from the surface of the nanodroplet due to the high electric field [22, 25]. The ejection of the analyte - carrying some of the droplet charges- decreases the charge density and therefore, the electric repulsion on the surface of the droplet [25]. Recently, a third mechanism –the Chain Ejection Model (CEM)- [26] has been proposed by Konermann, *et al.,* through extensive molecular dynamics (MD) simulations with Na^+ cations as the charge carrier species. The CEM has been employed to illustrate the release of an unfolded protein with exposed hydrophobic residues which would be located inside the hydrophobic core in the case of folded, native proteins [26]. During the ESI process, the interactions between the polar solvent molecules (such as water) and these hydrophobic residues are unfavorable. The unfolded protein therefore resides on the surface of the droplet exposing such residues to the surrounding nonpolar environment [26]. The protein thus is "pulled" by one terminus and ultimately released into the gas phase [26]. A summary of these three mechanisms is illustrated in Figure 1-1.

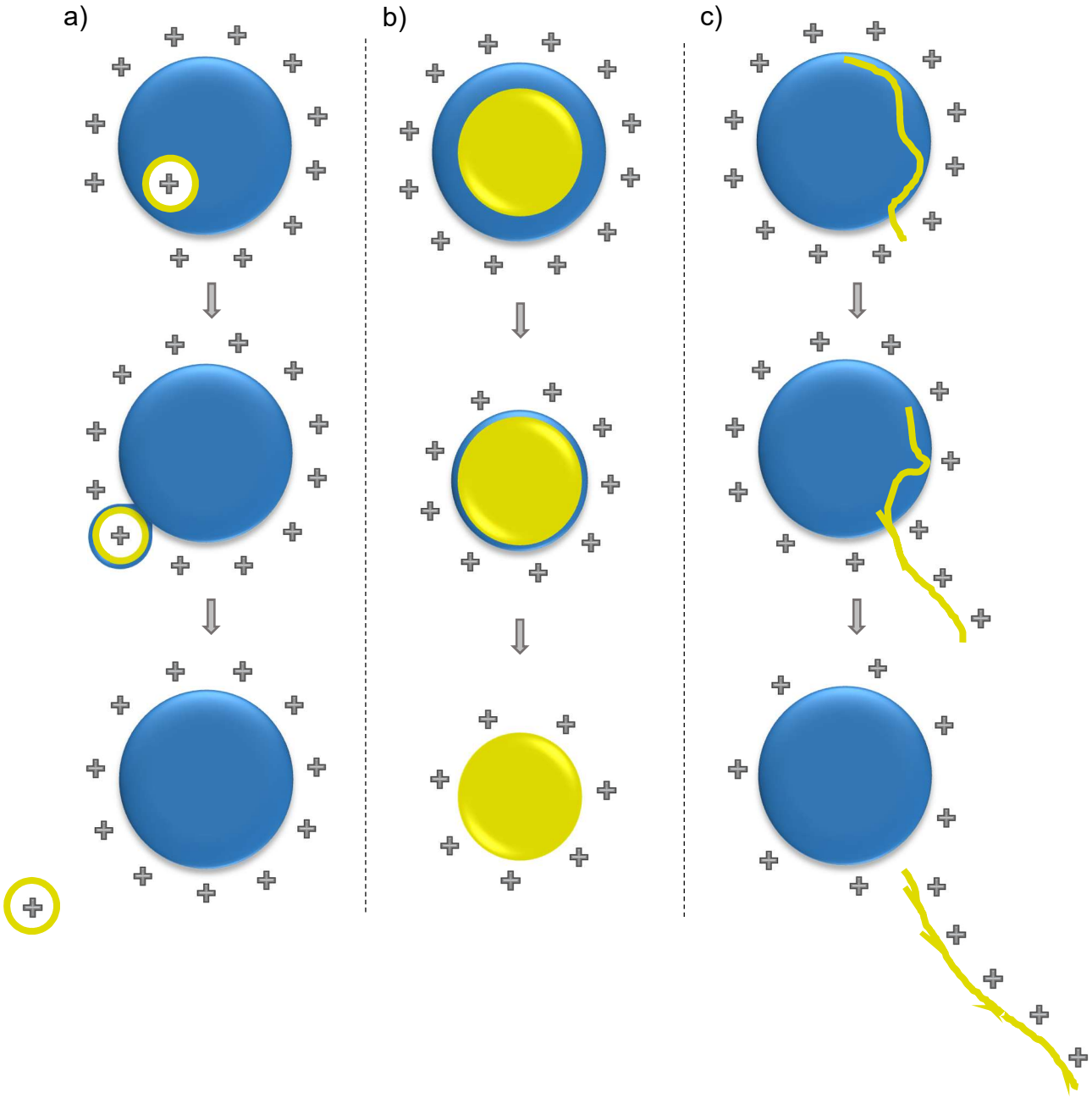

Figure 1-1. Summary of proposed ESI mechanisms. *a)* Ejection of a small analyte ion via IEM, *b)* Evaporation (and fission) event and formation of desolvated native protein by the CRM, *c)* Release of an unfolded protein through the CEM [26].

1.2. Ion Mobility Spectrometry- Mass Spectrometry (IMS-MS)

An IMS-MS instrument traditionally consists of an ionization source, an ion trapping device, a shutter grid to generate a pulse of ions, a linear drift tube of stacked cylindrical lenses separated by electrically non-conductive materials, a buffer gas inlet at the end of the drift tube to introduce a neutral buffer gas such as He and a mass analyzer (Figure 1-2) [27].

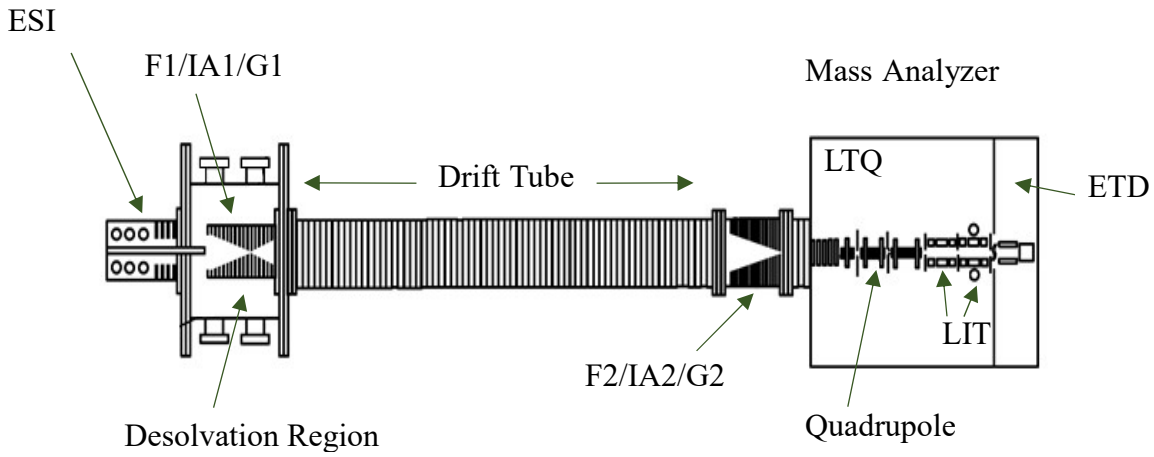

Figure 1-2. Schematic diagram of an ESI-IMS-MS instrument [123]. The buffer gas inlet is positioned near the end of drift tube. Ion funnel (F), ion activation (IA), and ion gate (G) regions are labeled. Also shown is the linear ion trap mass spectrometer and the ETD ion source region.

A potential gradient is applied to lenses to generate a uniform electric field (E) along the drift tube axis with length L to transfer the pulse of ions to the mass analyzer [27, 28]. The ions within the pulse are separated based on their mobility (K) which is directly proportional to the drift velocity of the ions (v_d) [29]. v_d is governed by the magnetite of the applied electric field and momentum transferred

to buffer gas molecules during the non-reactive collision events and ultimately dictates the overall ion drift time (t_D) [29]. The collisional frequency depends on the shape and size of an ion conformer, and consequently its mobility is distinguished from other conformers having the same charge. Using the kinetic theory, ion mobility can be inversely correlated to collision cross section (CCS) values [29]. The CCS (Ω) of a particular ion can be experimentally calculated by measuring t_D using Equation 1-2 [29]:

$$\Omega = \frac{(18\pi)^{1/2}}{16} \cdot \frac{ze}{(k_B T)^{1/2}} \cdot \left[\frac{1}{m_I} + \frac{1}{m_B}\right]^{1/2} \cdot \frac{t_D E}{L} \cdot \frac{760}{P} \cdot \frac{T}{273.2} \cdot \frac{1}{N} \qquad Eq. 1-2 .$$

In Equation 1-2, *N, T, P* and m_B correspond to the neutral number density (at STP), the buffer gas temperature and pressure, and the mass of the buffer gas, respectively. m_I and *ze* represent the mass and charge of analyte ions.

Although the concept of theoretical CCS calculations for rigid molecules such as benzene and naphthalene through a shadow projection-approximation method (PA) was introduced by Mack in 1925 [30], accurate measurements of the arrival time distributions of monoatomic carbon clusters and their mobilities by Jarrold *et al* and their comparison with theoretical mobility values provided an opportunity for the first time to obtain precise structural assignments of various isomers [31]. Later, with the development of the Exact Hard Sphere Scattering (EHSS) [32] and the trajectory method (TM) [33] computational approaches, the Mobcal software was generated [34]. Experimental CCS values for biomolecule ions could then be readily related to energy-minimized, *in-silico* structures that were produced using a variety of conformational sampling techniques [14, 35, 36].

1.3. Gas-phase Hydrogen-Deuterium Exchange (HDX) in IMS-MS

Early gas-phase HDX-MS experiments performed by McLafferty and co-workers suggested the presence of co-existing conformers within each charge state of cytochrome *c* [37]. This was accomplished through the measurement of HDX kinetics using a Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometer. The work of Clemmer and coworkers provided the means to measure the maximum number of exchangble hydrogens for select cytochrome *c* conformers by monitoring the shift in *m/z* values; here, conformers passing through the drift tube were also subjected to gas-phase HDX [16]. In this study, the electrosprayed ions were exposed to a partial pressure of D_2O gas introduced into the buffer gas in the drift tube. The short time scale for exchange in an IMS-MS instrument contrasts with that of FT-IRC-MS measurements in that the latter have a higher possibility for structural disruption (unfolding and refolding processes). Such transformations could expose a larger portion of the protein's surface to the reactive environment of the D_2O gas resulting in relatively lower exchange levels [16]. However the assumption of complete surface access leading to full exchange could not explain the relatively low exchange levels observed for very elongated conformations (large CCS values) [16].

The exchange between hydrogen and deuterium in the gas phase using D_2O, occurs through a "relay" mechanism (Figure 1-3) which requires the generation of a hydrogen-bonded intermediate; a D_2O molecule interacts at two positions on the protein exhibiting a spatial proximity [38]. Subsequently, the deuterium atom will exchange with heteroatom hydrogens on side chains and the protein backbone

(labile hydrogens) falling within a reasonable distance [39, 40]. This mechanism was invoked by Valentine, *et al.,* for computer-generated structures of cytochrome *c* [39]. MD simulations of the structures in which hydrogen accessibility to charge sites was considered, provided an explanation for the observation of increased exchange for compact structures at higher temperatures in contrast to elongated conformers essentially reversing the pattern observed at lower temperatures.

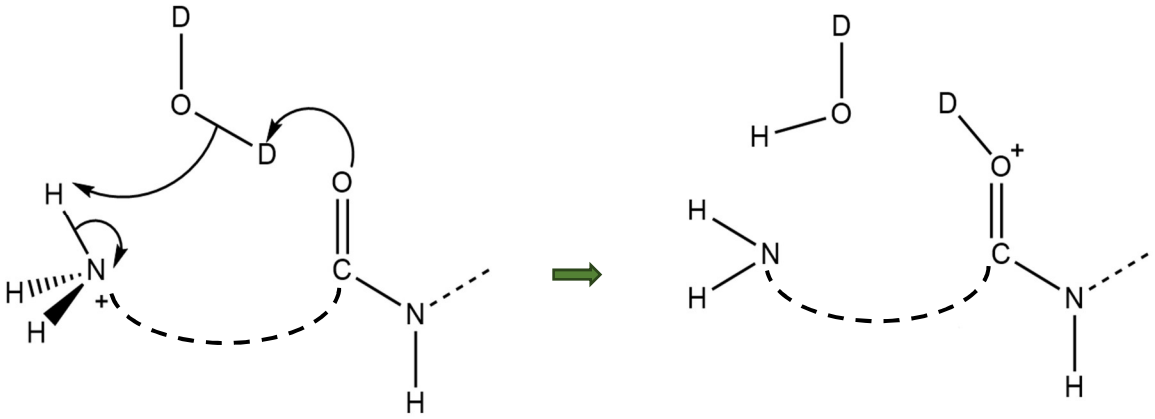

Figure 1-3. The Relay mechanism for the HDX reaction between D_2O and the protein [38]. Formation of the hydrogen-bonded complex facilitates the transfer of the proton from the charge site to the D_2O molecule and a deuterium to an oxygen atom on the amide group having a lower gas-phase basicity. The arrows represent the direction of electron transfer within the complex.

The application of tandem mass spectrometry (MS/MS) with HDX-IMS experiments provided the opportunity for the precise determination of amino acid residue deuterium uptake values [40, 41]. Recently, Valentine and co-workers applied an algorithm referred to as HAS (Hydrogen Accessibility Scoring) on computationally-sampled structures of a synthetic, model peptide which

considered spatial threshold requirements for the exchange process; the approach was shown to predict the deuterium incorporation that best matches the experimentally-observed, per-residue data [40].

1.4. Conformational Space Sampling and Representative Gas-phase Structure

CCS measurements can provide insight into the overall shapes and sizes of ions in the gas phase [29]. Using the drift tube of an IMS-MS instrument, CCS values can be determined with high accuracy from ion t_D values [29]. This requires the determination of measurable parameters such as buffer gas temperature and pressure [29]. Further, conformation space sampling techniques and theoretical CCS calculations using the Mobcal software produces a distribution of conformers of varying potential energy and CCS (Figure 1-4) [14, 34-36]. Filtering the *in-silico* structures with CCS matches exhibiting the confidence levels of ±4% provides candidate structures. The conformer distribution can be further narrowed by selection of the lowest energy structures (Figure 1-4) [14].

There are two major shortcomings to this approach. First, even for the short time scales of the simulations, the selection of the candidate structures from potential energy-CCS distributions is biased toward the lowest-energy *in-vacuo* structure with matching CCS. However, under mild conditions, the electrosprayed ions may represent kinetically-trapped, higher-energy structures. Second, even after extensive conformational space sampling simulations, there is no guarantee that the sampled conformers with matching CCS are representative of all possible

structures of the same size. Thus there is a need for a second selection step for the nominal structures to validate the quality of the sampling techniques.

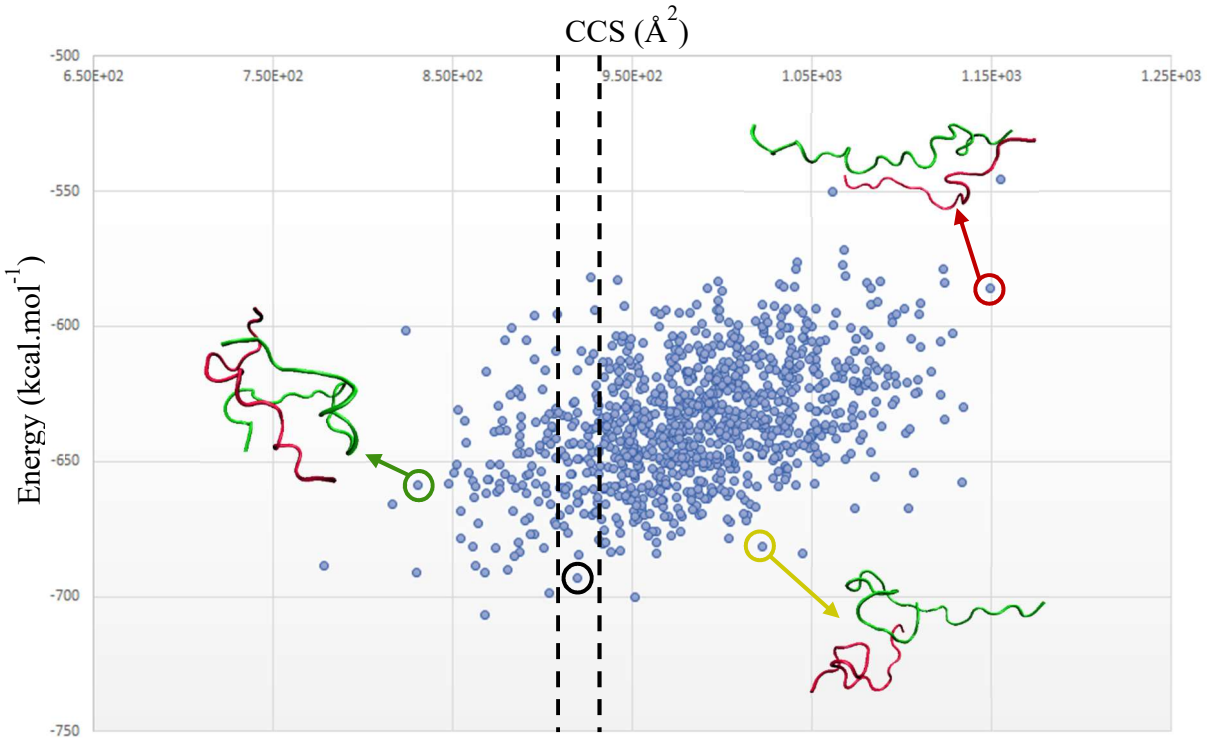

Figure 1-4. The energy-CCS distribution for computationally-generated insulin dimers. Each blue dot represent an *in-silico* structure obtained from conformational space sampling techniques. The dashed lines restrict the CCS region within the confidence level (see text for details). The most stable structure within this range is tagged by black circle.

Gas-phase HDX coupled with ion fragmentation techniques has found application in the selection of *in-silico* structures via the HAS algorithm [40]. The approach provides per-residue uptake data that not only validates the presence of experimentally observed conformer types within the computer-generated structures, but also adds a second dimension to filter through all of the structures

in the matching CCS interval regardless of their associated potential energy values [40].

1.4.1. Simulated Annealing (SA)

The SA approach has been widely used to sample the conformational space for peptides and small proteins [42-44]. The main concept for a SA approach is surmounting the high-energy barriers separating local minima along a potential energy surface (PES) by providing sufficient kinetic energy at higher temperatures. Next the system is equilibrated and gradually cooled to generate a structure characteristic of protein dynamics at lower temperatures followed by a shift to the nearest minimized configuration located in the same potential well via a minimization step. During the cooling process and at a non-zero temperature (*T*), the Metropolis algorithm [45] is applied on a random conformational change and the fate of the new structure is determined through the calculation of the difference in potential energy function before and after this change (ΔE). If the ΔE value is negative, the simulation will accept the new configuration and move downhill along the PES toward lower energy structures; however, if this value is positive, a probability function (ρ in Equation 1-3) related to a Boltzmann distribution will dictate the direction of movement to prevent trapping the structure in a high-energy potential well [45, 46].

$$\rho(\Delta E) = e^{\frac{-\Delta E}{k_B T}} \qquad Eq. 1-3$$

In Equation 1-3, k_B represents the Boltzmann constant. Due to the random nature of displacements in the metropolis algorithm and the use of a probability function,

repeated cycles of the SA approach (Figure 1-5) can be employed to generate a pool of structures. The energy-minimized structures at the end of each cycle are selected and subjected to further analysis.

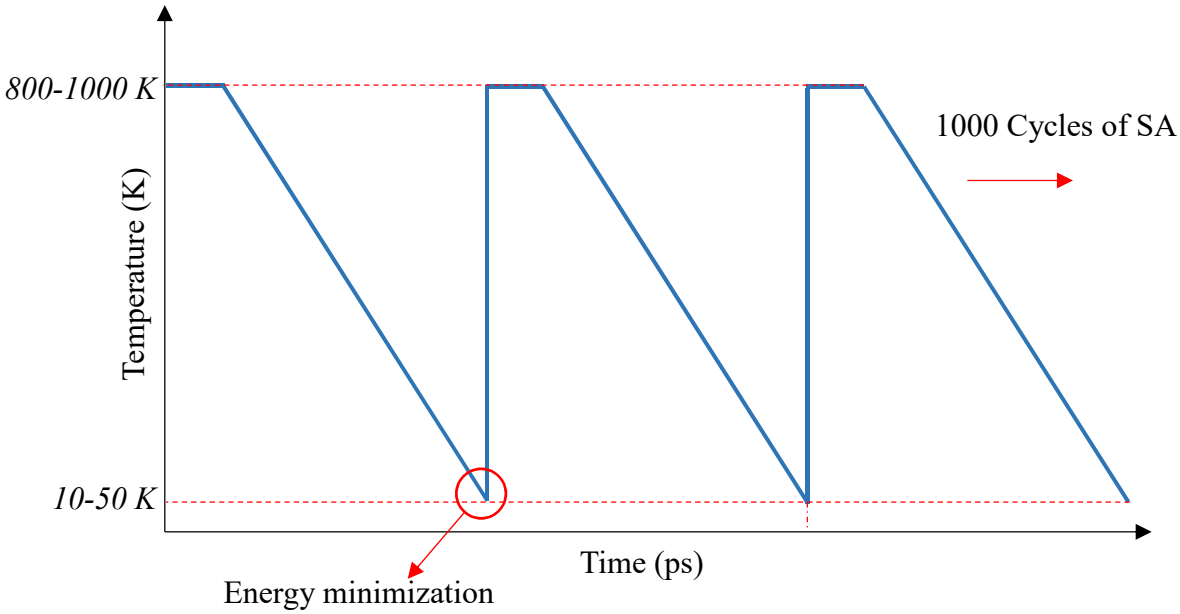

Figure 1-5. 1000 cycles of SA. The energy-minimized, annealed structure (red circle) from the end of each cycle is subjected to the same temperature profile at the next cycle and energy minimized structures are collected for future CCS calculation.

1.5. References

1. Fenn, J.B., *Electrospray wings for molecular elephants (Nobel lecture).* Angewandte Chemie-International Edition, 2003. **42**(33): p. 3871-3894.
2. Fenn, J.B., et al., *ELECTROSPRAY IONIZATION FOR MASS-SPECTROMETRY OF LARGE BIOMOLECULES.* Science, 1989. **246**(4926): p. 64-71.
3. Tanaka, K., et al., *Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry.* Rapid Communications in Mass Spectrometry, 1988. **2**(8): p. 151-153.
4. Shevchenko, A., et al., *Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels.* Analytical Chemistry, 1996. **68**(5): p. 850-858.
5. Lightwahl, K.J., B.L. Schwartz, and R.D. Smith, *OBSERVATION OF THE NONCOVALENT QUATERNARY ASSOCIATIONS OF PROTEINS BY ELECTROSPRAY-IONIZATION MASS-SPECTROMETRY.* Journal of the American Chemical Society, 1994. **116**(12): p. 5271-5278.
6. Rayleigh, L., *XX. On the equilibrium of liquid conducting masses charged with electricity.* Philosophical Magazine Series 5, 1882. **14**(87): p. 184-186.
7. Hudgins, R.R., J. Woenckhaus, and M.F. Jarrold, *High resolution ion mobility measurements for gas phase proteins: correlation between solution phase and gas phase conformations.* International Journal of Mass Spectrometry, 1997. **165**: p. 497-507.
8. Ruotolo, B.T. and C.V. Robinson, *Aspects of native proteins are retained in vacuum.* Current Opinion in Chemical Biology, 2006. **10**(5): p. 402-408.
9. Wyttenbach, T. and M.T. Bowers, *Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-Free Ion Mobility-Mass Spectrometry Environment.* Journal of Physical Chemistry B, 2011. **115**(42): p. 12266-12275.
10. Wyttenbach, T., et al., *The Effect of Calcium Ions and Peptide Ligands on the Relative Stabilities of the Calmodulin Dumbbell and Compact Structures.* Journal of Physical Chemistry B, 2010. **114**(1): p. 437-447.
11. Baumketner, A., et al., *Amyloid beta-protein monomer structure: A computational and experimental study.* Protein Science, 2006. **15**(3): p. 420-428.
12. Polfer, N.C. and J. Oomens, *Reaction products in mass spectrometry elucidated with infrared spectroscopy.* Physical Chemistry Chemical Physics, 2007. **9**(29): p. 3804-3817.
13. Gucinski, A.C., et al., *Separation and Identification of Structural Isomers by Quadrupole Collision-Induced Dissociation-Hydrogen/Deuterium Exchange-Infrared Multiphoton Dissociation (QCID-HDX-IRMPD).* Journal of the American Society for Mass Spectrometry, 2010. **21**(8): p. 1329-1338.
14. Clemmer, D.E. and M.F. Jarrold, *Ion mobility measurements and their applications to clusters and biomolecules.* Journal of Mass Spectrometry, 1997. **32**(6): p. 577-592.
15. Schnier, P.D., et al., *Blackbody infrared radiative dissociation of bradykinin and its analogues: Energetics, dynamics, and evidence for salt-bridge structures in the gas phase.* Journal of the American Chemical Society, 1996. **118**(30): p. 7178-7189.
16. Valentine, S.J. and D.E. Clemmer, *H/D exchange levels of shape-resolved cytochrome c conformers in the gas phase.* Journal of the American Chemical Society, 1997. **119**(15): p. 3558-3566.
17. Cech, N.B. and C.G. Enke, *Practical implications of some recent studies in electrospray ionization fundamentals.* Mass Spectrometry Reviews, 2001. **20**(6): p. 362-387.

18. Kebarle, P. and U.H. Verkerk, *ELECTROSPRAY: FROM IONS IN SOLUTION TO IONS IN THE GAS PHASE, WHAT WE KNOW NOW.* Mass Spectrometry Reviews, 2009. **28**(6): p. 898-917.
19. Taylor, G.I. and A.D. McEwan, *The stability of a horizontal fluid interface in a vertical electric field.* Journal of Fluid Mechanics, 2006. **22**(1): p. 1-15.
20. Grimm, R.L. and J.L. Beauchamp, *Evaporation and Discharge Dynamics of Highly Charged Multicomponent Droplets Generated by Electrospray Ionization.* The Journal of Physical Chemistry A, 2010. **114**(3): p. 1411-1419.
21. Dole, M., et al., *Molecular Beams of Macroions.* The Journal of Chemical Physics, 1968. **49**(5): p. 2240-2249.
22. Thomson, B.A. and J.V. Iribarne, *Field induced ion evaporation from liquid surfaces at atmospheric pressure.* The Journal of Chemical Physics, 1979. **71**(11): p. 4451-4463.
23. Schmelzeisen-Redeker, G., L. Bütfering, and F.W. Röllgen, *Desolvation of ions and molecules in thermospray mass spectrometry.* International Journal of Mass Spectrometry and Ion Processes, 1989. **90**(2): p. 139-150.
24. Fernandez de la Mora, J., *Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism.* Analytica Chimica Acta, 2000. **406**(1): p. 93-104.
25. Fenn, J.B., *Ion formation from charged droplets: Roles of geometry, energy, and time.* Journal of the American Society for Mass Spectrometry, 1993. **4**(7): p. 524-535.
26. Konermann, L., et al., *Unraveling the Mechanism of Electrospray Ionization.* Analytical Chemistry, 2013. **85**(1): p. 2-9.
27. Donohoe, G.C., et al., *A New Ion Mobility-Linear Ion Trap Instrument for Complex Mixture Analysis.* Analytical Chemistry, 2014. **86**(16): p. 8121-8128.
28. Hoaglund, C.S., et al., *Three-dimensional ion mobility TOFMS analysis of electrosprayed biomolecules.* Analytical Chemistry, 1998. **70**(11): p. 2236-2242.
29. Mason, E.A. and E.W. Mcdaniel, *Transport properties of ions in gases*. 1988: John Wiley & sons.
30. Mack, E., *Average cross-sectional areas of molecules by gaseous diffusion methods.* Journal of the American Chemical Society, 1925. **47**: p. 2468-2482.
31. Vonhelden, G., et al., *CARBON CLUSTER CATIONS WITH UP TO 84 ATOMS - STRUCTURES, FORMATION MECHANISM, AND REACTIVITY.* Journal of Physical Chemistry, 1993. **97**(31): p. 8182-8192.
32. Shvartsburg, A.A. and M.F. Jarrold, *An exact hard-spheres scattering model for the mobilities of polyatomic ions.* Chemical Physics Letters, 1996. **261**(1-2): p. 86-91.
33. Mesleh, M.F., et al., *Structural information from ion mobility measurements: Effects of the long-range potential.* Journal of Physical Chemistry, 1996. **100**(40): p. 16082-16086.
34. Jarrold, M.F. *Mobcal, A program to calculate mobilities*. 2/10/2014 2/18/2016]; Available from: http://www.indiana.edu/~nano/software.html.
35. Wyttenbach, T., C. Bleiholder, and M.T. Bowers, *Factors Contributing to the Collision Cross Section of Polyatomic Ions in the Kilodalton to Gigadalton Range: Application to Ion Mobility Measurements.* Analytical Chemistry, 2013. **85**(4): p. 2191-2199.
36. Fernandez-Lima, F.A., et al., *On the Structure Elucidation Using Ion Mobility Spectrometry and Molecular Dynamics.* Journal of Physical Chemistry A, 2009. **113**(29): p. 8221-8234.
37. Suckau, D., et al., *COEXISTING STABLE CONFORMATIONS OF GASEOUS PROTEIN IONS.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(3): p. 790-793.

38. Campbell, S., et al., *Deuterium exchange reactions as a probe of biomolecule structure. Fundamental studies of cas phase H/D exchange reactions of protonated glycine oligomers with D2O, CD3OD, CD3CO2D, and ND3.* Journal of the American Chemical Society, 1995. **117**(51): p. 12840-12854.
39. Valentine, S.J. and D.E. Clemmer, *Temperature-dependent H/D exchange of compact and elongated cytochrome c ions in the gas phase.* Journal of the American Society for Mass Spectrometry, 2002. **13**(5): p. 506-517.
40. Khakinejad, M., et al., *Combining Ion Mobility Spectrometry with Hydrogen-Deuterium Exchange and Top-Down MS for Peptide Ion Structure Analysis.* Journal of the American Society for Mass Spectrometry, 2014. **25**(12): p. 2103-2115.
41. Rand, K.D., et al., *Site-Specific Analysis of Gas-Phase Hydrogen/Deuterium Exchange of Peptides and Proteins by Electron Transfer Dissociation.* Analytical Chemistry, 2012. **84**(4): p. 1931-1940.
42. Chou, K.C. and L. Carlacci, *SIMULATED ANNEALING APPROACH TO THE STUDY OF PROTEIN STRUCTURES.* Protein Engineering, 1991. **4**(6): p. 661-667.
43. Bruccoleri, R.E. and M. Karplus, *CONFORMATIONAL SAMPLING USING HIGH-TEMPERATURE MOLECULAR-DYNAMICS.* Biopolymers, 1990. **29**(14): p. 1847-1862.
44. Baysal, C. and H. Meirovitch, *Efficiency of simulated annealing for peptides with increasing geometrical restrictions.* Journal of Computational Chemistry, 1999. **20**(15): p. 1659-1670.
45. Metropolis, N., et al., *EQUATION OF STATE CALCULATIONS BY FAST COMPUTING MACHINES.* Journal of Chemical Physics, 1953. **21**(6): p. 1087-1092.
46. Wilson, S.R., et al., *APPLICATIONS OF SIMULATED ANNEALING TO THE CONFORMATIONAL-ANALYSIS OF FLEXIBLE MOLECULES.* Journal of Computational Chemistry, 1991. **12**(3): p. 342-349.

2. Comprehensive Peptide Ion Structure Studies Using Ion Mobility Techniques: Part 1. An Advanced Protocol for Molecular Dynamics Simulations and Collision Cross Section Calculation

Reprinted with permission from *Journal of The American Society for Mass Spectrometry:* Comprehensive peptide ion structure studies using ion mobility techniques: Part 1. An advanced protocol for molecular dynamics simulations and collision cross section calculation. Ghassabi Kondalaji, S., Khakinejad, M., Tafreshian, A., Valentine, S., Journal of The American Society for Mass Spectrometry, 2017: p. 1-13.

2.1. Introduction

The study of peptide and protein ion structure in the gas phase offers the opportunity of characterizing these systems in the absence of a complicating, explicitly defined environment, where the dynamic nature of hydrogen bonding [1, 2] and the relatively large number of discreet interacting species make these studies challenging [3, 4]. Gas-phase studies, therefore, provide a means to examine intrinsic (intramolecular) interactions in protein and peptide ion conformers [5-9]. As mentioned above (1-1), the tools and methods used for these purposes have been categorized as physical and chemical probes [5]. As their names imply, a physical probe provides data related to a physical property of an ion (e.g., ion size), while a chemical probe provides information about ion reactivity leading to structural inference [5].

CCS values obtained from IMS measurements can be used as a physical probe of gas-phase ion size [10]. MD simulations can be coupled with CCS measurements to provide a powerful tool for structure studies [11-13]. In this

approach, CCS calculations for *in-silico* structures are compared to experimental values to infer structural information. In early work, *in-silico* structures were largely used for relative size comparisons with protein ions [14]. In these studies, CCS calculations were performed for protein ion structures such as a native, a partially folded, an α-helix (unfolded), and a fully extended structure to effectively ascertain the relative degree of compactness for cytochrome *c* ions. More recently, NMR structures were used as inputs in MD simulations to optimize and produce a nominal structure for native-like, gas-phase ions of ubiquitin [15, 16]. Another approach associated with this method included the production of numerous, random *in-silico* structures to sample conformational space where the structures with matching CCS values could be selected as candidate structures [6, 7, 17]. Traditionally the most stable structure of these was accepted as the most representative of the gas-phase structures (1-4). Such an approach is valid insofar as the structure produced from ESI and subsequent gas-phase ion transfer steps represents a more thermodynamically stable ion conformer rather than a higher-energy, kinetically-trapped state.

Several studies demonstrate the effect of solution conditions (and their solution structure) on the resulting gas-phase conformers [15, 18-25]. One issue is the degree to which such species resemble solution structures and their relative stability in the gas-phase. In studies of peptide ions, collisional activation of ion conformer populations suggest a large portion are kinetically-trapped species [18]. Separate studies have suggested that low charge state protein ions can persist as solution-like structures for extended periods of time in the gas phase [15].

Therefore, gas-phase studies may provide information about antecedent solution states. However, because CCS measurements provide limited structural information related to the shape of the ions and the fact that the computational techniques mentioned above are often directed toward obtaining the most stable gas-phase structures, difficulties arise with regard to structural assessments.

Recently we have proposed the application of gas-phase HDX with IMS and MS/MS to begin to address shortcomings associated with ion conformer selection [26-28]. In this approach, gas-phase ions undergo reactions with D_2O reagent gas during the mobility separation. Subsequently, using ETD fragmentation data, the number of deuteriums incorporated within each residue can be determined. Using the HAS algorithm and an effective collision model, a hypothetical deuterium uptake pattern for each *in-silico* structure can be generated to serve as an additional criterion for structure elucidation. In the second installment of this work, the HDX mechanism and gas-phase ion structure dynamics are utilized to improve the accuracy of the HAS algorithm. A goal of the work reported here is to develop a method to produce accurate structures as well as a means to assess their relative populations. To improve the accuracy of structural determinations, the study first focuses on enhanced conformational space sampling techniques and then CCS calculation methods are examined. It is confirmed that truly comprehensive sampling of conformational space for these types of studies is essential. Here, extensive MD simulations have been performed to produce combined pools of more thermodynamically-stable and higher-energy structures.

The trajectory method [12, 29] implemented in the Mobcal software suite [30] is widely used to calculate theoretical CCS values for *in-silico* ion structures. Although the procedure appears to be relatively straightforward, it can actually be quite complicated. This is especially true for peptide ions where the conformational energy barriers along the energy landscape are significantly smaller than for proteins. That is, peptide ions are highly flexible and can adopt a number of diverse energetically-available structures [31]. Because of this, no single *in-silico* structure is representative of the experimentally observed ion population. For this reason such structures are often referred to as “conformer type” [26]. Therefore, to obtain an accurate structural representation, the dynamic nature of a conformer type should be considered. The CCS exhibited by conformer types can be calculated along a MD trajectory run by averaging values for all structure frames. That said, such a process is prohibitively computationally intensive and approximation methods are highly desirable. To address this challenge, different data mining approaches have been tested and benchmarked to extract the most representative structures from a MD trajectory in order to propose an efficient protocol to calculate accurate CCS values exhibited by a conformer type.

Because the degree of similarity between solution- and gas-phase structures may be answered by studying gas-phase conformer establishment [8, 15, 16, 22, 32-36], the work presented in the discussion below is preliminary yet foundational in nature. That said, recent CCS measurements of partially-dehydrated, electrosprayed ions reveal that the gas-phase structure adaptation can depend on intrinsic factors such as the nature of basic residues and the interaction of other

residues in the stabilization of structure [6, 22]. Therefore, although these studies provide valuable insight regarding gas-phase structure establishment, a statistical comparison of solution- and gas-phase structural types can provide additional insight into the actual resemblance of such species for a variety of biomolecular ions. In a series of manuscripts the application of IMS-HDX-MS/MS coupled with MD simulations for the accurate elucidation of structures (and their populations) in the gas-phase will be presented and, subsequently, the gas-phase structures and solution structures (from extensive MD simulations guided by CD spectroscopy data) will be compared statistically.

2.2. Experimental

2.2.1. Sample Preparation

The model peptide Acetyl-PAAAAKAAAAKAAAAKAAAAK (> 90% purity) was synthesized by Genscript (Piscataway, NJ, USA). Peptide stock solutions were prepared by dissolving 1 mg of the model peptide (without further purification) in 1.0 mL Mili-Q water. ESI solutions were prepared by 1:10 dilution of stock solution with 100 mM solution of ammonium acetate in water. ESI solutions were infused (300 $nL \cdot min^{-1}$) into the hybrid IMS-MS instrument [28, 37] through a pulled-tip capillary biased at 2200 V relative to the instrument entrance orifice.

2.2.2. IMS-MS Measurements

The instrument used for these experiments has previously been described in detail [28, 37]. Briefly, a home-built drift tube coupled to a linear ion trap (LIT) mass spectrometer (LTQ Velos, ThermoScientifc, San Jose, CA, USA) was employed.

A dual ion gating system was utilized to provide the time delay between the release of ions from the ion trap and the selection of ions of a given mobility. The delay time between the first and second ion gates was adjusted to mobility select the ions. Delay times between the two gates were scanned to obtain drift time distributions (0.1 ms increments). For each time increment mass spectra were collected for 0.5 minutes. The LIT *m*/*z* scan range of 400-1,000 was utilized. For MS/MS measurements, an ETD reaction time of 200 ms was employed.

2.2.3. MD Simulations

To perform the *in-vacuo* MD simulations, an initial extended structure of a $[M+3H]^{3+}$ ion was generated using the AMBER12 [38] molecular dynamics package. The non-polarizable all-atom Amber ff12SB force field was employed for structural parameterization. To derive the undefined force field parameters for the COOH-terminal Lysine residue carrying a distinct formal charge in the gas phase, two structures of this single amino acid with ψ and φ dihedral angle values matching those in α-helical and extended structures were generated. Quantum mechanics geometry optimizations for both conformations were carried out at HF/6-31G(d) theory level using the GAMESS software package [39, 40]. The optimized structures were subjected to multi-orientation molecular electrostatic potential (MEP) computations and charge fittings using the R.E.D. server development [39-44] to obtain the empirical force field parameters for this specific residue.

The extended initial structure of the triply-charged peptide ions with charge arrangements of K(6)-K(11)-K(21) and K(6)-K(16)-K(21) were energy minimized

using 2500 steps of steepest descent algorithm followed by 2500 steps of conjugate gradient algorithm. Cyclic SA (Figure 1-5) was employed for conformational space sampling of the energy-minimized structure [17, 28, 45]. During SA runs (1-fs time step), the Berendsen temperature coupling algorithm [46] was used. No long-range cutoffs for non-bonded interactions were considered. The temperature of the *in-vacuo* system was dramatically increased to 1,000 K over 8 ps with a heat bath coupling time constant of 0.2 ps, trans-omega dihedral restraints on the entire peptide backbone, and chirality restraints on all chiral centers. The heated structure was subjected to dynamics at constant temperature and gradually cooled to lower temperatures (10 K) over various designated total SA timescales (10, 20, 40, 100, 400 and 1200 ps) using the coupling time constant of 4.0 ps and was subsequently energy minimized to generate structures at 0 K (the annealed structures). This annealed structure served as the starting point for the next heating-cooling cycle. Of these timescales, 40 and 1200 ps SA runs were used to generate a more diverse pool of higher-energy and more thermodynamically stable structures as candidate conformers. The increased conformer diversity is described in the Results and Discussion section.

After 1,000 cycles of simulation, all resulting annealed structures (end of each SA run) were gradually heated to 300 K over a 100-ps timescale using the Berendsen temperature coupling algorithm with a coupling time constant of 1.0 ps and equilibrated. The final heated-equilibrated structures were subjected to 5 ns production MD *in vacuo* with a 2-fs timestep. The simulations were carried out without the non-bonded cutoffs for long-range interactions. The temperature of the

system was maintained at 300 K using Langevin dynamics with a collision frequency of 1 ps^{-1} [47, 48]. 5,000 structures were sampled from each MD trajectory. The structures with the lowest potential energy in each MD trajectory were extracted to serve as the reference Cartesian coordinate set in mass-weighted, root-mean-square deviation (RMSD) calculations. Backbone-only RMSD values for all structures within a single trajectory were obtained using the *ptraj* [49] module implemented in the AMBER12 software package [38]. RMSD-oriented structures were subjected to a fixed-radius k-means clustering algorithm using the MMTSB toolset [50]. By applying an in-house script to this algorithm, the selected radius values for each trajectory were altered in a fashion such that the number of generated clusters was 50 (±10). Using the Mobcal software and trajectory method (TM) [30], CCS calculations were performed on the structures with the lowest RMSD relative to the mathematically-generated centroids for each resulting cluster.

The high annealing temperature of 1,000 K enhances the sampling process by overcoming high-energy barriers along the PES thus increasing the accessible conformational space for the sampling process [51, 52]. However, such a process facilitates unwanted cis-trans transitions and chirality inversions along the peptide backbone [51-53]. Therefore chirality and trans-omega dihedral restraints are normally applied to prevent such transitions. Notably, the presence of structural restraints may limit the sampled configurations to a narrower portion of conformational space [52]. Although restraints were employed for this study, careful examination showed that such an action did not affect the conformational

space sampling capabilities of SA. Method development and benchmarking processes were only conducted on K(6)-K(11)-K(21) peptide ions. The summary of the MD simulations is presented in Figure 2-1.

2.2.3.1. Restraints on High Temperature SA

To analyze the effect of the applied restraints, 1,000 cycles of 40-ps SA runs have been performed on the peptide ion with the charge arrangement of K(6)-K(11)-K(21) without any chirality and cis-omega dihedral restraints. The resulting structures have been subjected to helicity and pairwise RMSD calculations. The comparison between these data and the results obtained from restrained SA runs with the same cooling period (40-ps), suggests the same range (within ~3%) of conformational space coverage for both simulations. Additionally, the helicity comparison shows a high degree of similarity between the NP values for both SA methods. Because the presence of omega dihedral angle values deviating from 180° (ideal trans peptide bond) can dramatically affect the structure assignments in non-restrained conformers, the values generated from the STRIDE algorithm [54] can only be considered reliable in providing an overall picture of structural diversity in non-restrained simulations and cannot be used in determining the secondary structure for the peptide ions with cis-trans transitions along the backbone.

The overall number of peptide bonds with a cis configuration is determined using the Cispeptide [55] plugin implemented in the VMD software package [56] with an omega dihedral angle threshold of 85°. The data demonstrate that the during the first multiple cycles of simulations, all the sampled structures at 0 K have the omega dihedral value corresponding to a trans peptide bond; however as the

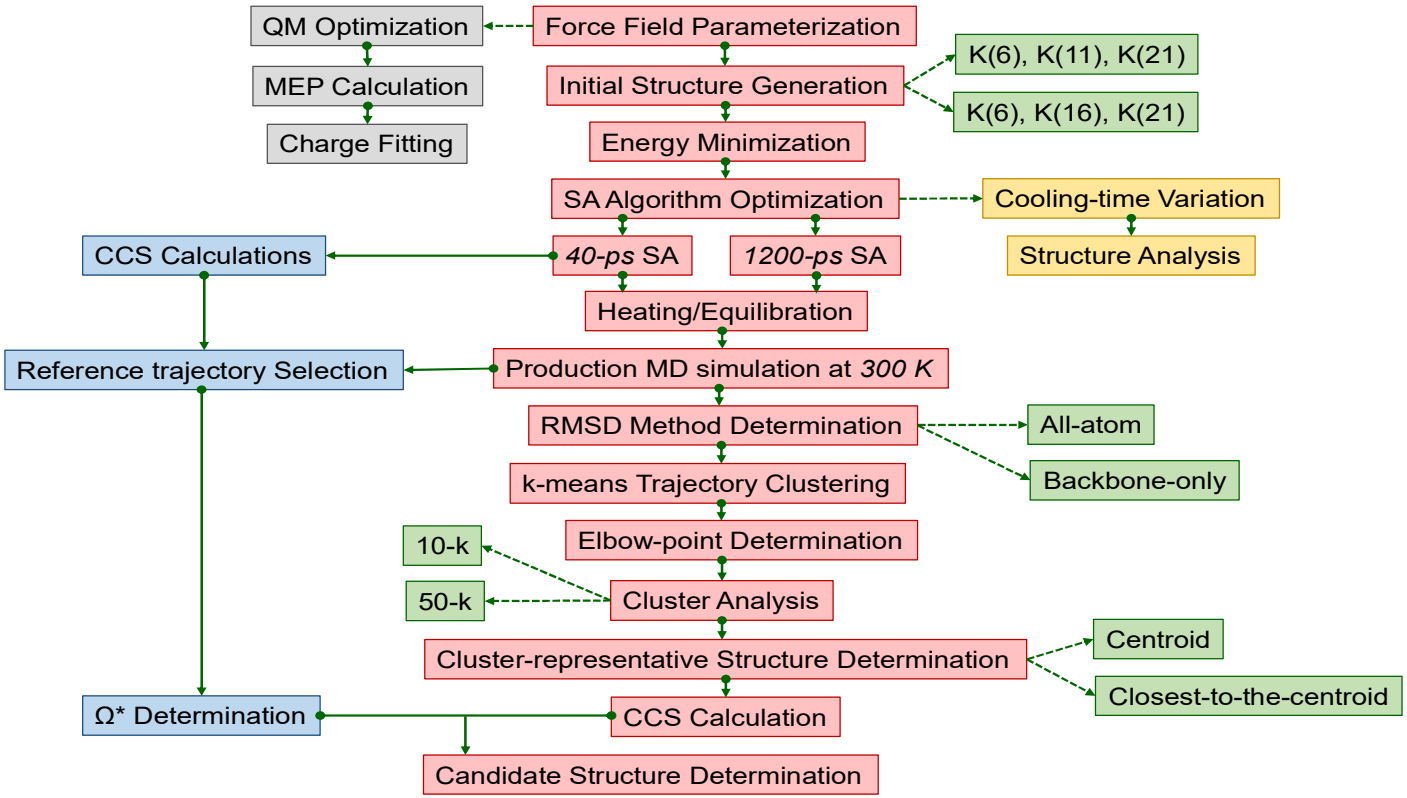

Figure 2-1. Summary of Molecular Dynamics (MD) simulations. The pink boxes represent the main workflow from initial structure generation to selection of candidate structures through a series of extensive cluster analysis and CCS calculations. The grey rectangles show the process applied in calculating the force field parameters for the c-terminal, positively-charged lysine residues through a series of geometry optimizations and charge determination steps. The blue rectangles show the CCS calculations and linear selection of 10 reference structures from the entire CCS range of annealed structures, followed by CCS determinations of their exhibited trajectories (5,000 frames sampled per trajectory) and comparison of these accurate CCS values (Ω*) with CCS measurements obtained through benchmarking procedures (pink rectangles). The analysis of SA algorithms and the related optimizations are shown by the rectangles highlighted in yellow. The filled arrows represent the direction for performing processes. The green rectangles demonstrate the different selections for the original step (pink-rectangle).

cycles proceed, the number of peptide bonds in a cis configuration increases as as the majority of annealed structures have all their omega dihedral values transition to the cis configuration. The presence of a cis peptide bond in the annealed structures decreases the candidate sample pool volume; only structures with trans peptide bonds are acceptable and therefore the number of acceptable annealed structures decreases. For the purposes of this study, the cis-trans restraints have been applied over all simulations to create a large pool of structures. It is worth mentioning that for this particular system, no chirality inversions have been observed in non-restrained simulations, however, due to such occurrences in previous SA runs on different protein systems, chirality restraints with a force constant value of 10 $kcal.mol^{-1}.rad^{-2}$ have also been applied to all SA analyses.

2.2.3.2. Secondary Structure Analysis

To pinpoint the secondary structure elements, the STRIDE algorithm was employed [54]. This algorithm considers the positional-dependent hydrogen bond energy as well as the statistically-derived amino acid propensity to predict the secondary structure. The algorithm identified 3_{10}-helix, α-helix, π-helix, turn and random coil as the existing elements for the sampled structures. The helicity (H_R) for a structure was determined as the ratio of the overall number of amino acid residues (without considering the acetyl cap) existing in any of the three possible helices (R) to the total number of amino acid residues (21). This generated a range of values from 0 (where none of the amino acid residues participates in a helix) to 1 (where all the backbone atoms exist in a helical structure). Normalized population

values (*NP*) for helicity of each SA run with 1,000 annealed structures are presented for the range of 0 to 21 within-helix, *R* residues. Since the formation of a helix requires at least *i* to *i*+3 interactions, very low normalized population values are observed for one and two amino acid residues. Total helicity for 1,000 annealed structures was calculated according to Equation 2-1:

$$TotalHelicity = \sum_{R=0}^{21} H_R \times NP \qquad Eq. 2-1$$

2.2.3.3. Selection of an Appropriate Temperature

MD simulations at constant temperature can provide an approximation of protein ion structural fluctuations which can affect their gas-phase transport properties [31, 57] and hydrogen-deuterium exchange reactivity [58]. The degree of the dynamics with regard to peptide structure highly depends on the energy of the ion and the accessible energy barriers. Therefore it is necessary to perform production simulations at a temperature matching the ion energy levels in the drift tube. Ion collisions with the neutral gas transform a portion of the drift velocity into a random velocity component. As a result, the internal energy of the ions increases, and thus their effective temperature is higher than the ambient temperature. The temperature of ions in the low-field limit regime can be estimated as [59]:

$$\frac{3}{2}k_B T_{eff} = \frac{3}{2}k_B T_a + \frac{1}{2}m_B {v_d}^2 \qquad Eq. 2-2$$

in which k_B is Boltzmann's constant, m_B, v_d, T_a , and T_{eff} are neutral buffer gas mass, ion drift velocity, ambient temperature, and the effective temperature of the

ions, respectively. Using a drift velocity of 100 $m \cdot s^{-1}$ and ambient temperature of 293 K, the effective temperature can be estimated to be 300 K. Therefore, for the simulations to better match the experimental conditions, all the annealed structures have been heated to the higher temperature of 300 K and equilibrated. To sample the peptide ion dynamics at 300 K, the heated-equilibrated structures have been subjected to 5 ns production MD simulations at this temperature. The heating, equilibration and production MD simulations are performed for both charge arrangements of triply-charged protein ions and on structures sampled from the 40-ps and 1200-ps SA runs. Overall, two charge arrangements and two simulated annealing procedures were utilized to produce a pool of 4,000 annealed structures. The trajectories have been sampled to generate 5,000 frames for each initial heated-equilibrated conformation leading to a significantly large pool (5,000×1,000) of structures for a single charge arrangement and a specific SA run.

2.2.4. Cluster Analysis

2.2.4.1. Data Mining and Clustering Algorithm

k-means clustering is one of the simplest (and therefore most applied) methods in data mining to divide *n* vectors of *p* dimensions into *k* distinct clusters. In this method, *k* centroids are selected for the vectors to similar groups and the distances from their corresponding centroids are minimized. That is, the following function is minimized:

$$\sum_{i=1}^{k} \sum_{j \in C(i)} (x_j - \mu_i)^2 \qquad Eq. 2-3$$

In Equation 2-3, x_j is a p×1 vector, *C(i)* is the i^{th} cluster and μ_i is the corresponding centroid vector (the term $(x_j - \mu_i)^2$ represents the same value as $RMSD_{ij}$ in section 2.2.4.2). Moreover, *C* and μ represent the set of clusters and their centroids (i= 1,2,…,k), respectively. Given a large number of clusters, this is a NP-hard problem which means that it is computationally difficult to find an optimal global solution. However, Hartigan, *et al.,* [60] introduced an algorithm that provides optimal local solutions which guarantee that there is no single movement from one cluster to the other that would reduce the objective function above. This algorithm can be implemented in the R [61] software environment using the *kmeans* function in the *stat* package.

It is worth noting that if the members of each cluster are similar enough to their centroids, these *k* centroids can be assumed to be the representative of their associated clusters and used in a subsequent analysis [62].

Note that for the k-means algorithm introduced above, the number of clusters, *k*, is assumed to be set which is the main shortcoming of the k-means algorithm. However, there are a variety of methods to tackle this issue including the Elbow-point method which can be utilized to specify an appropriate number of clusters [63]. In this method, the *k* value is iterated from 1 to a reasonably large number and, for each iteration, the k-means algorithm would be applied and the total within-cluster sum of squares can be computed. A descending trend is expected when plotting the total within-cluster sum of squares (*TWSS*) against *k*, because as *k* increases the members of each cluster more closely approximate the centroid and consequently the TWSS value reduces. However, from a reference frame defined

as the Elbow-point, the objective function does not change dramatically and because minimizing the number of clusters would decrease the computation time for the purposes, the Elbow point is selected as the suitable number of clusters. This algorithm can be implemented in R [61] software using the function *fviz_nbclust()* in the *factoextra* package.

2.2.4.2. Elbow-point Determination

All 4,000 trajectories obtained from MD simulations were subjected to clustering using the *cluster.pl* utility implemented in the MMTSB toolset [50]. This program applies hierarchical (*jclust* script) as well as partitional k-means (*kclust* script) clustering methods; however, due to the significant number of structures, the relatively faster k-means clustering algorithm has been used to perform the analysis. One drawback in selecting the k-means algorithm, is that the *kclust* script requires a “fixed-radius” parameter as the cluster threshold value and the number of generated clusters (*k*) cannot be determined by the user. Thus, a script - developed in house - has been employed to change the radius values for the *kclust* script in order to obtain a desired number of clusters.

To perform cluster analysis for a single trajectory, a series of 5,000 both backbone-only and all-atom RMSD-oriented structures have been subjected to the radius modifying script. For a particular radius value, the respective RMSD values are calculated and the similarity between the structures is determined. The total within-cluster sum of squares (*TWSS*) values are calculated according to Equation 2-4:

$$TWSS = \sum_{i=1}^{k} \sum_{j=1}^{N_i} RMSD_{ij}^2 \qquad Eq. 2-4$$

where $RMSD_{ij}$ is the RMSD of the j^{th} member of cluster i relative to the centroid of the corresponding cluster, N_i is the number of members present in the i^{th} cluster, and k is the number of generated clusters for a particular radius value. The *TWSS* value represents the distance of cluster members from their corresponding centroid. As the script modifies the radius values by increments of 0.2 Å, the new *TWSS* values are calculated. A plot of *TWSS* values versus the different *k* values becomes relatively constant at a value known as an elbow-point which provides the optimal value for the number of generated clusters. The elbow-points have been calculated for all trajectories and the maximum value between all calculated elbow-points results in the optimal *k* value of 50 clusters (Figure 2-2). Based on the behavior of the *TWSS* function at the elbow-point, dividing the trajectories into 50 clusters should result in the greatest mutual similarity among cluster members within the shortest computation time. Therefore, the *k* value of 50 (±10) and the corresponding generated clusters is expected to provide the highest accuracy; however, to examine the possibility for obtaining a desirable accuracy at lower computation times, a *k* value of 10 (±4) which corresponds to the area below the elbow-points for all TWSS plots, has been utilized to generate a different class of clusters. In the present manuscript, the terms 50-*k* and 10-*k* will be used to address *k* values of 50 (±10) and 10 (±4), respectively. In addition to this, the accuracy of the results has been investigated for a third class of analysis where the *k* value is selected to be 1 (no clustering). For the 10 reference trajectories, the accuracy can

be determined by comparison of CCS values of weighted cluster-representative structures with the accurate CCS values (Ω^*) corresponding to the average CCS value for all 5,000 structures in a MD production run. Therefore the selection of the cluster-representative structure is a priority.

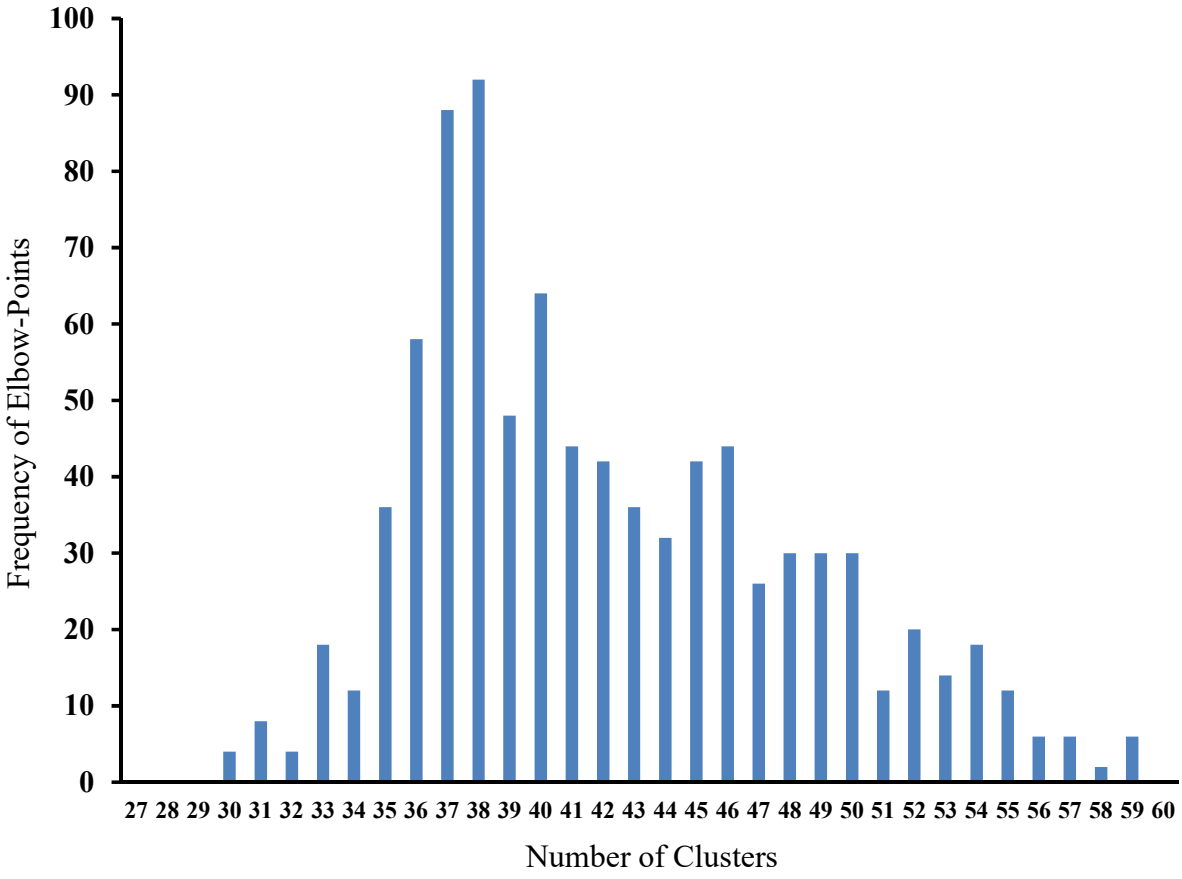

Figure 2-2. Histogram representation of elbow-point occurrences for all production MD simulations at 300 K.

2.3. Results and Discussion

2.3.1. Arrival Time Distribution of $[M+3H]^{3+}$ and $[M+4H]^{4+}$ Ions

Upon electrospraying the model peptide, quadruply-, triply- and doubly-protonated ions of the peptide are produced. Figure 2-3 shows the arrival time distribution of these ions generated by integrating values over a narrow *m/z* range across all t_D values.

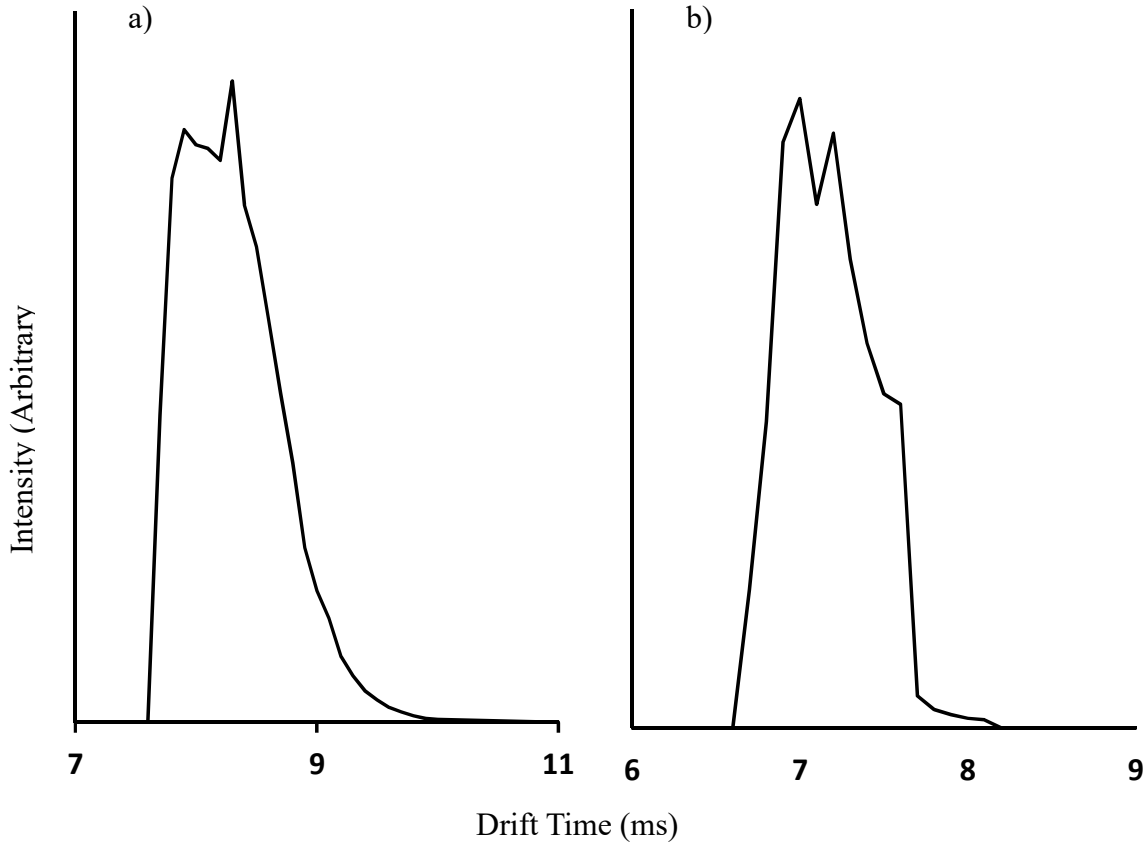

Figure 2-3. Drift time distribution of $[M+3H]^{3+}$ (panel a) and $[M+4H]^{4+}$ (panel b) ions of the model peptide Acetyl-PAAAAKAAAAKAAAAKAAAAK. The distribution is obtained by integrating intensities for all m/z values for these ions across the drift time range.

Drift time distribution for triply protonated ions is depicted at panel a of Figure 2-3. Two well resolved peak at 7.9 ms and 8.3 ms are indicative of a compact

(Ω=417 Å^2) and more diffuse (Ω=438 Å^2) structure types. $[M+4H]^{4+}$ ions, exhibit two unresolved conformers at 7 ms (Ω=492 Å^2) and 7.2 ms (Ω=506 Å^2) as the most prevalent structural types, while an unresolved shoulder at 7.6 ms (Ω=534 Å^2) is an indicator of a more diffuse structure.

2.3.2. Peptide Ion Collision Cross Sections.

Figure 2-4 shows the three-dimensional (3D), raised-relief plot for model peptide ions indicating the abundances of ions at specific t_D and m/z values. Upon electrospraying the model peptide, quadruply-, triply-, and doubly-protonated peptide ions are produced. Quadruply-protonated ions are observed as a small feature at m/z of ~453. These ions exhibit three different conformer types with CCS values of 492, 506 and 534 Å^2 (Figure 2-3). Doubly-charged ions produce a wide, unresolved distribution with CCS values spanning a range of 300-400 Å^2. The most dominant feature in the spectrum corresponds to triply-charged ions. The shape and width of the peak associated with these ions is consistent with the presence of at least three conformer types. The dominant conformer type is the most compact conformer with a CCS value of 417 Å^2. The second most abundant conformer type is slightly more diffuse (Ω = 438 Å^2). The third conformer type represents an unresolved shoulder in the ion's drift time distribution with a CCS value of 464 Å^2. The most compact and slightly more diffuse conformers of triply-charged ions were selected for further study because of their higher overall intensity as well as increased conformer type purity upon mobility selection.

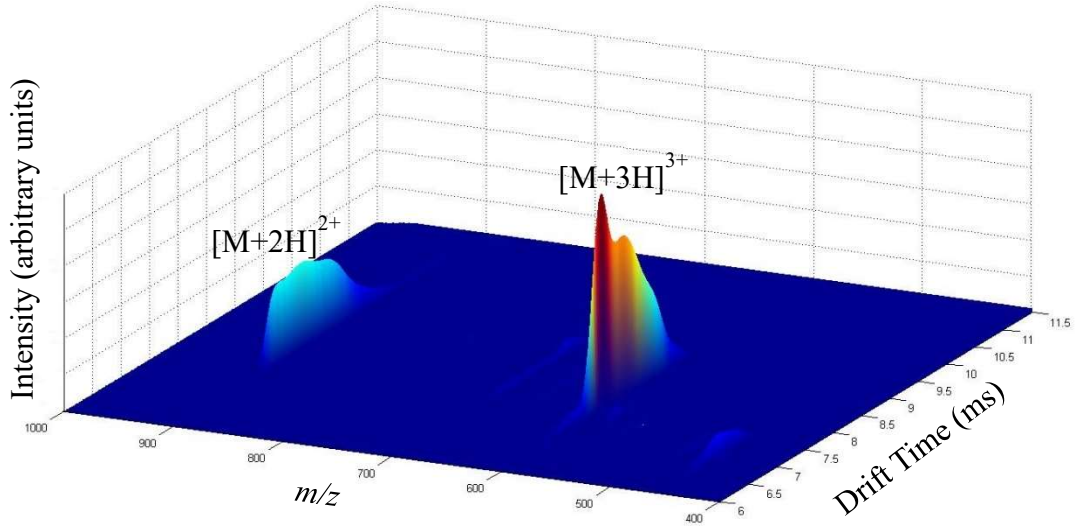

Figure 2-4. Three-dimensional (t_D, *m/z*, intensity) raised-relief plot produced by electrospraying a solution of the model peptide Actyl-PAAAAKAAAAKAAAAKAAAAK. Dominant ions in the distribution are labeled.

2.3.3. Conformational Space Sampling

Several studies suggest that over the short timescale of the ESI process, the protein ions do not resemble equilibrated conformations in the gas phase; rather, they are higher-energy, meta-stable structures presenting similarities to the solution phase conformations [6, 23, 33]. In a single SA cycle, increasing the time of the cooling process leads to the generation of structures that are thermodynamically more stable [64]. Therefore lengthened cooling steps even in multiple cycles of SA simulations (Figure 1-5) will direct the resulting conformational space toward the selection of lower-energy configurations. The primary goal of a cyclic SA approach in these studies is not to determine the global minimum as in many classical SA simulations [53, 65] but to sample a pool of

structures from both thermodynamically-stable conformers as well as higher-energy configurations. Here, the conformational sampling quality is evaluated by varying the designated cooling timescale (see Experimental section) for multiple, 1,000-cycle SA runs. Figure 2-5 panel a shows examples of the ion temperature as a function of the SA times for the $[M+3H]^{3+}$ ion with charge arrangement of K(6)-K(11)-K(21). The resulting annealed structures sampled at the end of each simulation are subjected to additional cycles of heating-cooling steps. The potential energy values of the 1,000 resulting minimized conformers (annealed structures) are averaged for each run. The slower cooling algorithms generated more thermodynamically stable structures as expected (Figure 2-5 b).

A secondary structure analysis has been performed for all 1,000 annealed structures generated from multiple SA runs in order to determine the conformational space available to the protein backbone. Overall, slow-cooling simulations focused the conformational space toward the formation of more helical structures compared to fast-cooling SA runs which leads to the formation of more random structures. Figures 2-5 c and 2-5 d show an increase in the number of amino acid residues that are involved in helical structure.

Considering the number of residues involved in a helix (R) as presented in Figure 2-5 c, the formation of a series of conformers having random structure encompassing at most 13 amino acid residues results from the 40-ps simulations (i.e., the NP is zero for $R \geq 13$), while for the 1200-ps SA runs up to 16 residues can exist in helical structures (i.e., the NP is zero for $R \geq 16$).

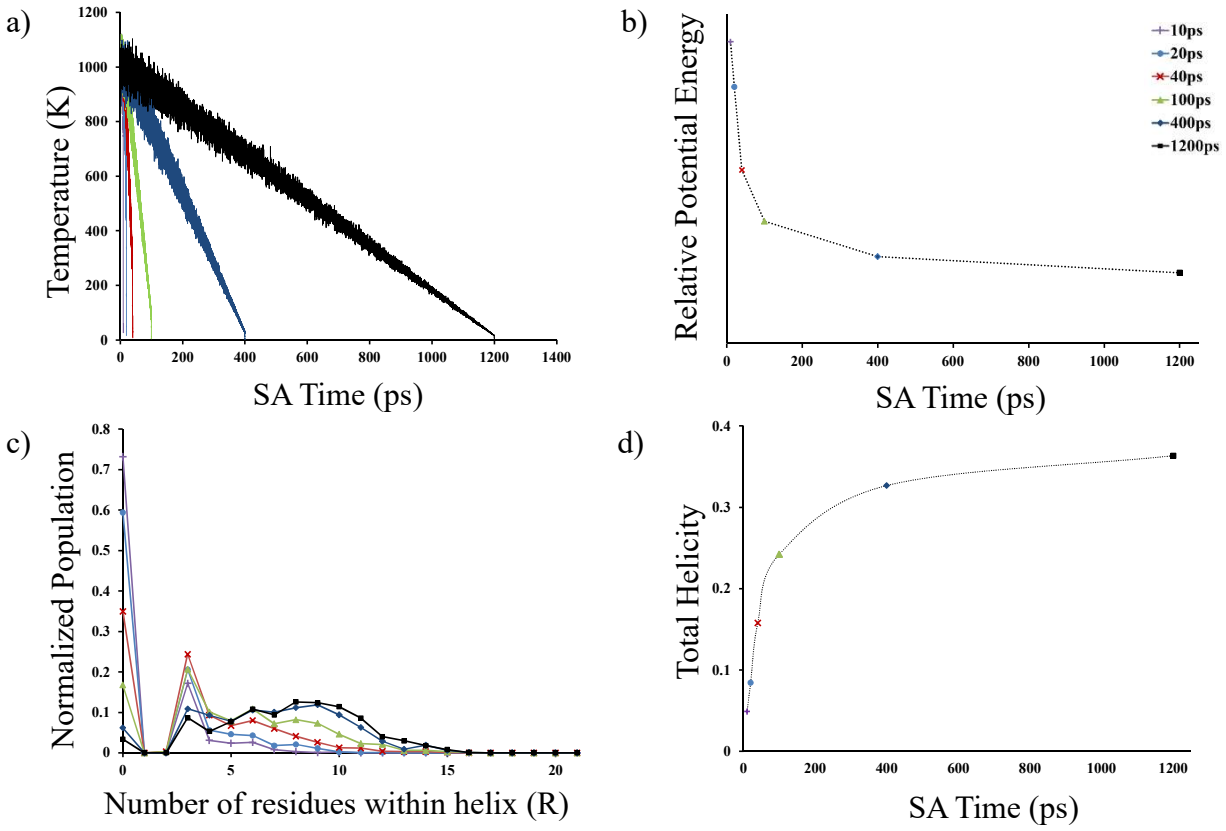

Figure 2-5. The effect of cooling time variation on sampled structures at the end of simulated annealing (SA) runs. Each color represent a specific simulation time ranging from 10 ps to 1200 ps. Panel a shows the change in system temperature as a function of time for six different simulations having the same annealing time period (8 ps). The associated energies of 1,000 structures obtained at the end of the SA cycles are averaged for each run and illustrated in panel b. Panel c shows the secondary structure analysis resulting in a helicity number represented by the within-helix residue count for each sampled structure that has been normalized for the population of 1,000 structures for a single SA run. The total helicity values for the representative SA runs are shown in panel d.

This observation raises the issue of whether or not the conformational space sampled at 40 ps is more limited than that encountered by the peptide ion using cycles of 1200 ps. To estimate the relative diversity of sampled structures, mass-

weighted RMSD calculations were performed pairwise for all 1,000 annealed structures. The average minimum and maximum pairwise RMSD values (data not shown) reveal a slightly larger (~10%) difference in RMSD values for the 40-ps runs compared with the 1200-ps simulations.

Because the ESI process can produce a number of conformers existing as a wide range of conformer types and energies [6, 23, 33], to produce an *in-silico* pool of structures that could contain a subset of species that are representative of the mixture, the 40-ps SA and the 1200-ps SA run were utilized, which is expected to produce a wider range of conformer type and energy than either alone.

2.3.4. Post-SA Clustering

For the purposes of data reduction and to decrease the computation time, the annealed conformers could be clustered and representative structures used for each cluster; for example, if 5 annealed structures exist in each cluster, selection of a single representative conformer decreases the computation time by a factor of ~5. However, a cluster analysis for 1,000 annealed structures obtained from 40-ps SA runs shows that such data reduction techniques can lead to a biased selection of sampled structures and eliminate a variety of conformation types such as structures with 9, 12 and 14 amino acids residues (of 21 total) existing in one of the three helical forms (Figure 2-6). The blue line in Figure 2-6 is representative of data without cluster analysis while the red line shows the trace for the output for the data reduction procedure. Because such an approach can lead to unwanted elimination of some structure types, no data reduction has been carried out. Therefore, all 4,000 annealed structures (2 SA run times, 2 protonation site

arrangements, and 1,000 sampled structures from each cyclic SA run) have been subjected to subsequent analysis without initial clustering fallowing the SA runs.

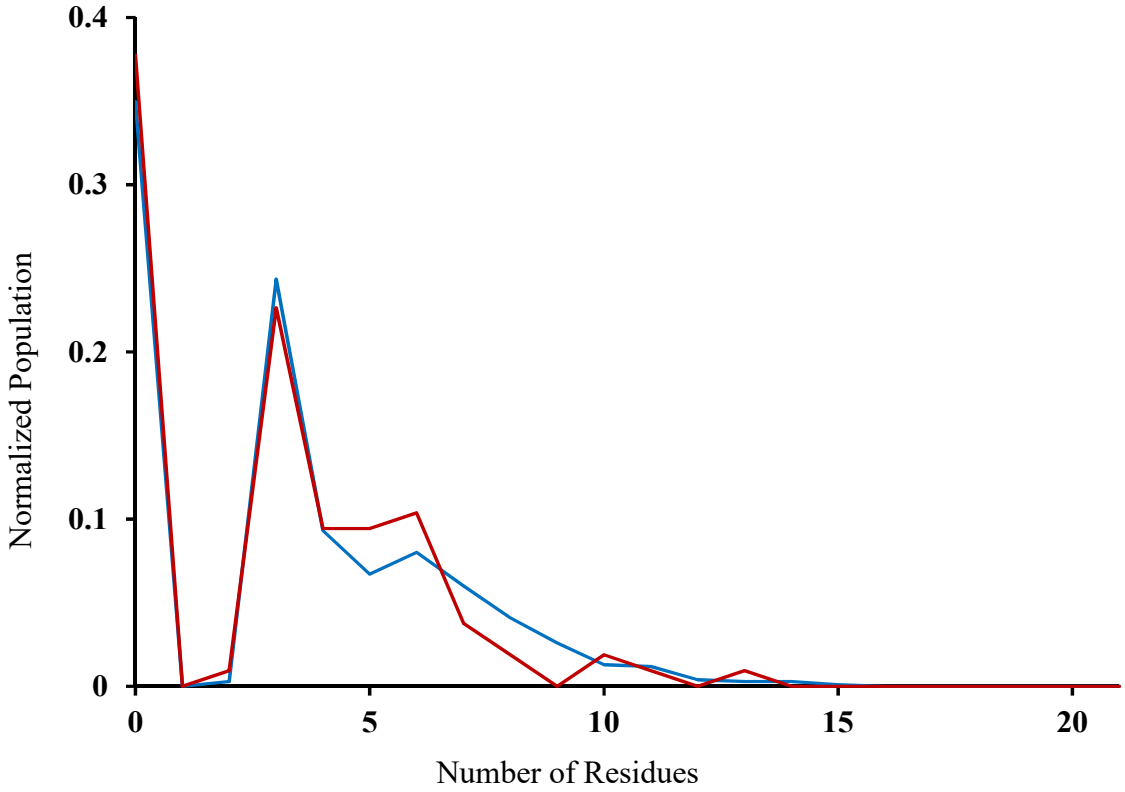

Figure 2-6. Normalized populations of structure types with a specific number of residues in helical structure (R) are depicted. Blue traces shows the value exhibited by all structures while red lines illustrates the values after data reduction with cluster analysis.

2.3.5. Final Temperature in SA Cycles

It can be argued that terminating the cooling step during the SA procedure at 300 K rather than 0 K and subsequently sampling a minimized conformer is more efficient by eliminating the further heating and equilibration steps. In order to

benchmark a conformational space sampling technique, it has been necessary to continue the cooling steps to lower temperatures. The limited energy available at lower temperature (e.g., 10 K) leads to less structural fluctuation and the sampled conformations do not deviate from energetically available structures at such temperatures; therefore, a single structure can be used to benchmark the conformational space sampling methods. Notably, these structures would offer different starting point along the PES and do not represent the entire trajectory with regard to CCS calculations as discussed below. Therefore, the goodness of conformation space sampling can be qualitatively assessed by the diversity of these structures.

Although the production MD was used to simulate conformer fluctuations, it is noted that the timescale (5 ns) is significantly smaller than that of the measurement (~9 ms). Considering that the MD trajectory could sample a localized area of the PES, the fact that some structural transformations may not occur on the short simulation timescale is, to some degree, mitigated by the SA approach where multiple structures are submitted to production MD. That is, a goal of the more extensive SA is to bypass energy barriers that may be associated with such structural transformations. Additionally, the method of filtering candidate structures (CCS and, in the future, HDX matching) utilizes a linear combination of structures; thus in a sense, the filtering has the potential to stitch together the available conformational space. That said, even with this extensive approach, incomplete representation of structural fluctuations cannot be ruled out.

2.3.6. CCS Calculation

The trajectory method (TM) [29] calculation (using the Mobcal [30] software) has been widely utilized to calculate hypothetical CCS values for *in-silico* structures. For a single *in-silico* structure this process is straightforward; however, thousands of captured frames are associated with the production MD simulations. A valid solution is to calculate the CCS values for all structures within a single MD trajectory and average the obtained results (Ω* above); however, this is computationally expensive (if not impossible). For example, in the case of these experiments, CCS values for 20 million *in-silico* structures would have to be calculated with such an approach. Sampling and calculating CCS values for representative structures can provide a remedy for this problem. That is, the average of CCS values for several sampled frames can be nearly identical to the average of CCS values for the entire trajectory. Here, different methods of sampling have been utilized to reduce the calculation cost. To evaluate these methods and obtain reference CCS values for comparisons, the CCS calculations (trajectory method) were performed on the annealed structures. After plotting the potential energy-CCS distribution for these structures [only K(6),K(11),K(21) charge arrangements and 40-ps SA runs], the entire resulting CCS range was divided equally into 10 regions and the most stable annealed structure in each region was selected as a reference structure. To obtain the accurate CCS values for the conformer types at 300 K originating from these 10 reference structures, the Mobcal [30] software was used and the accurate CCS values (Ω^*) were calculated by averaging the CCS values of all the conformers obtained from the corresponding production MD simulations (reference trajectories). The accuracy

of the sampling and data mining methods was then evaluated by comparing the outcome of these methods to Ω^* values for the 10 reference trajectories.

To test the data mining, several techniques including the method of clustering, the number of generated clusters and selection of the centroid or the closest structure to the centroid as the representative structure for each cluster have been employed and benchmarked to find the most efficient lemma. For all of these approaches, CCS values for a conformer type (the CCS representative of the entire trajectory) have been calculated as:

$$\Omega_{total} = \sum_{1}^{k} \Omega_i \times \frac{N_i}{N_t} \qquad Eq. 2-5$$

In Equation 2-5, N_i , N_t are the number of frames in the i^{th} cluster (cluster members) and total number of MD frames (5,000 for these studies), respectively. The value Ω_i is the CCS of the representative structures for the corresponding cluster, while Ω_{total} is the calculated CCS value exhibited by a conformer type. The value k, represents the number of generated clusters for each trajectory.

2.3.7. The Centroid as a Representative Structure

Per-atom averaging over the Cartesian coordinates of multiple structures in a trajectory, or a portion of it, leads to the formation of a geometry referred to as a centroid. This mathematically-generated geometry can represent the structures that have been utilized in its genesis; therefore, it can be implied that in order to reduce the computation time in the CCS calculations, the centroid can be used to describe the gas-phase transport of the corresponding cluster components.

Because increased clustering of the frames in a MD trajectory positions increasingly similar conformations in a particular cluster, it may be expected to enhance the accuracy of this approach. Therefore, the effect of the number of clusters (*k* value) on the geometry of the centroid and its CCS value was investigated. Two structures with the lowest mutual RMSD value were selected from a series of random clusters generated after cluster analysis with *k* values of 1 (no clustering), 10 and 50 using the same production MD trajectory. The new centroids were created for each pair of structures (1st and 2nd structures in Figure 2-7). The CCS calculation using the Mobcal [30] software was performed. This procedure was repeated for various trajectories. Comparison between the calculated CCS values of the structure pairs with the CCS of their corresponding centroid illustrates that the selection of the centroid as the representative structure provides CCS values that are smaller than that of the initial pair (Figure 2-7). In addition, as the initial *k* value and the mutual similarity between each pair increases, the deviation of the CCS of the centroid from the CCS values of the 1st and 2nd structures decreases. The change in this deviation is significantly higher between the 10 to 50 cluster sets (b and c in Figure 2-7) than the 1 to 10 sets which is in agreement with the cluster behavior at the elbow-point suggested by the *TWSS* plot. It is worth mentioning that the Mobcal [30] software was incapable of calculating the TM-CCS values for a majority of the centroids resulting from the no clustering analysis most likely due to significant disruption in the geometry of the generated centroids (Figure 2-7). That is, the centroid geometry tends to improve as the number of clusters increases from 1 to 10 to 50.

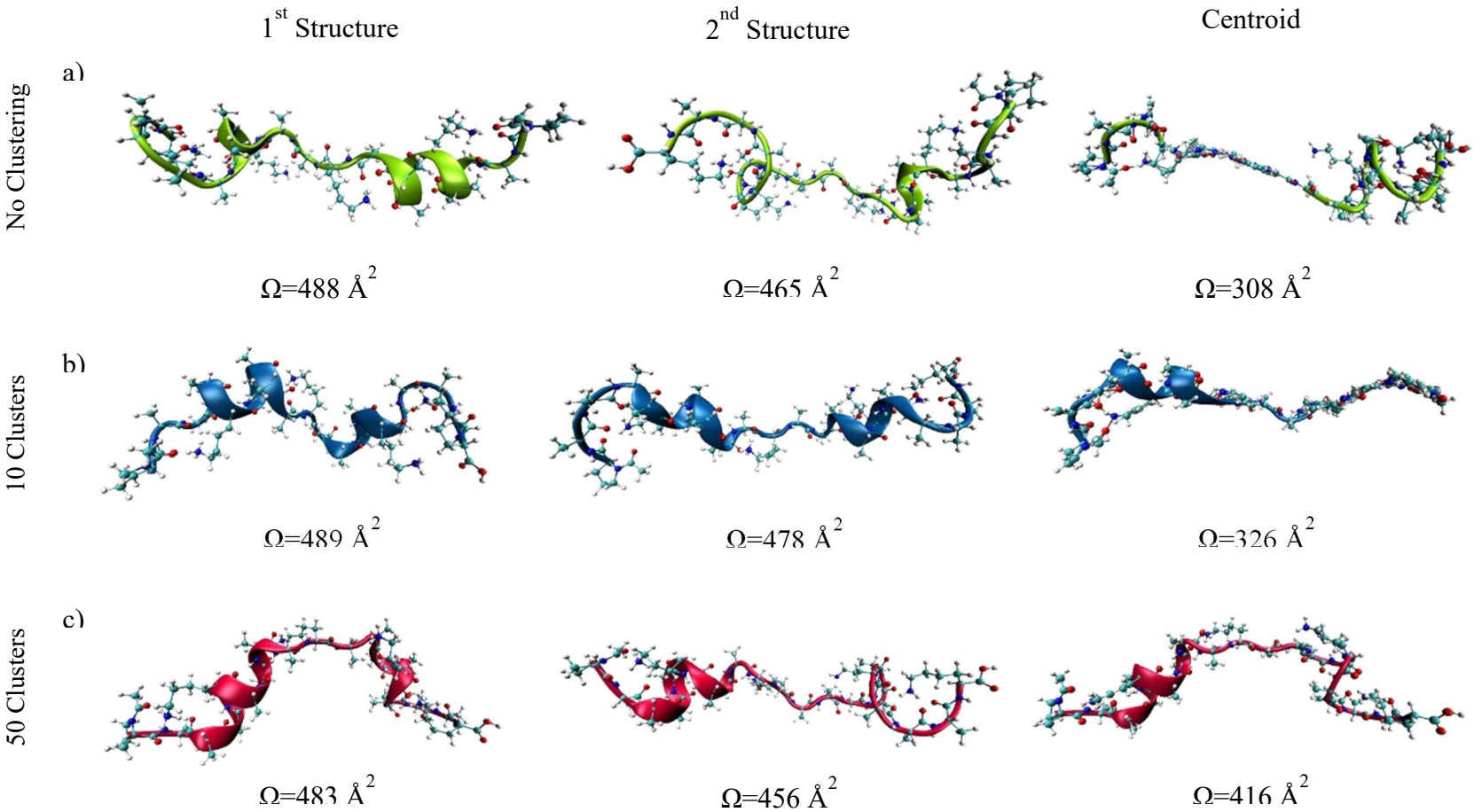

Figure 2-7. Schematics showing challenges with the use of centroid geometries for CCS calculations. Each row compares the calculated CCS values for the centroid and the original structures. The two structures on the left of each row are the conformers with the maximum RMSD within a single cluster in a production MD run. Geometries on the right are the calculated centroid of the conformers on the left. No cluster analysis is performed for the top row (RMSD= 12.57 Å), and for the middle and bottom rows, the maximum number of clusters is confined to 10 (RMSD= 5.18 Å) and 50 (RMSD= 2.66 Å), respectively.

Figure 2-8 shows the average (over 5,000 structures) potential energy associated with collision cross sections for structures corresponding to 1,000 separate production MD trajectories. The CCS values (Ω_{total}) have been calculated using clustering of the 5,000 frames of each trajectory according to Equation 2-5. In general, no significant difference is observed for different alignment methods (top versus bottom in Figure 2-8) performed before cluster analyses.

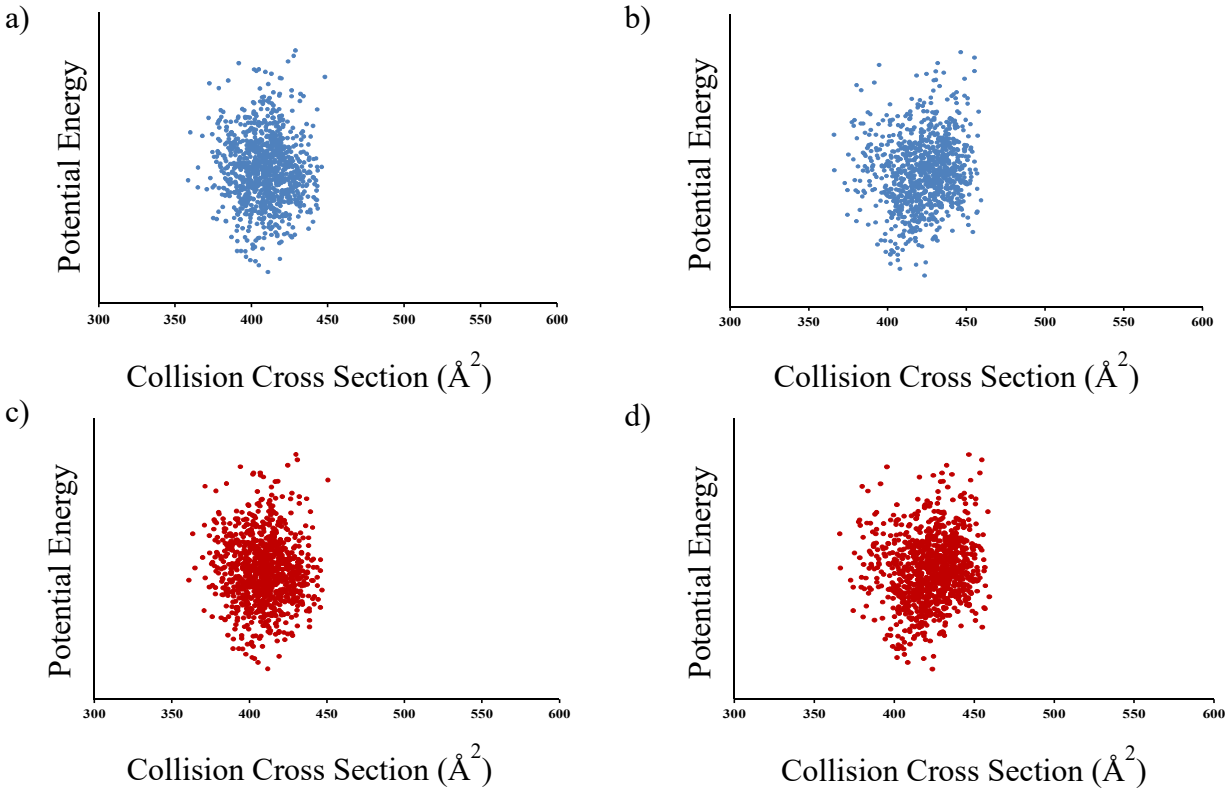

Figure 2-8. CCS values versus potential energy for the $[M+3H]^{3+}$ ions. The centroid geometry has been used to calculate the CCS value. For the top panels, prior to cluster analysis structures are backbone-only RMSD oriented while for the bottom panel orientation is performed according to all-atom criteria. Left and right panels show the results for 10-*k* and 50-*k* cluster analyses, respectively.

Although the majority of the resulting CCS values matches the experimental results for the two more compact conformer types of $[M+3H]^{3+}$ ions, the overall distribution (Figure 2-8- all panels) is a relatively narrow CCS range and does not correlate to the more elongated conformer type. This can be indicative of inaccuracy in the CCS calculations. Notably, the width of the distribution for 50-*k* clusters is slightly larger than the 10-*k* centroid sets with an overall shift toward higher CCS values. This suggests that increasing the number of clusters results in increased accuracy in CCS determinations; however, the accuracy gain with this approach is expected to be limited as further increase in *k* value beyond 50 (the elbow-point) would not significantly affect the cluster members in terms of similarity. It would, however, increase the computation time.

Because the centroid is a mathematically generated three-dimensional arrangement of the atoms based on averaging the Cartesian coordinates of the multiple structures in the corresponding cluster, in a sense, it lacks any realistic physical meaning. This is observed in the centroid geometries illustrated in Figure 2-7. Also as demonstrated in Figures 2-7 and 2-8, this lack of physical meaning in centroid geometry is a source of significant error in CCS calculation. To address this issue, different strategies can be applied. Although these centroids can be reconstructed to a more realistic conformation via a variety of geometry optimization techniques, resulting structures merely follow the energy level criteria which does not guarantee the preservation of the memory of the trajectories from which they have been sampled. It may be argued that this loss in the trajectory information through geometry optimization and structural reconstruction can lead

to inaccuracy in the intended trajectory representation. A remedy to these problems is the use of the closest structures to the centroid as the cluster-representative conformation.

2.3.8. Closest Structure to a Centroid as the Representative Species

In a cluster, the structure with the lowest RMSD relative to the centroid displays the highest degree of similarity to the centroid among all cluster members without having the attendant structural disruption of the centroid. To investigate the behavior of these more realistic structures and the accuracy of this approach, CCS calculations have been performed for all the closest structures to the centroids obtained from cluster analysis (k=1, 10 and 50). Subsequently, the TM-CCS values of the corresponding trajectories (Ω_{total}) were calculated using Equation 2-5. Since the pre-clustering method of structure alignment (all-atom and backbone-only RMSD-orientation) does not affect the resulting CCS values (Figure 2-8), the CCS calculations were only performed for the backbone-only RMSD-oriented trajectories.

To benchmark the accuracy of different CCS determination approaches, the accurate CCS values (Ω^*) for the 10 reference structures are compared with Ω_{total} values obtained with these methods. Figure 2-9 shows the Ω_{total} values after cluster analysis with k value of 1 and the accurate Ω^* values. A comparison between these two data sets illustrates that geometry averaging over the trajectories without subsequent clustering leads to inaccuracy in CCS determination; not a sufficient number of structures is sampled. The error associated with these Ω_{total} values is 5.9 % on average with a maximum value of 15 %.

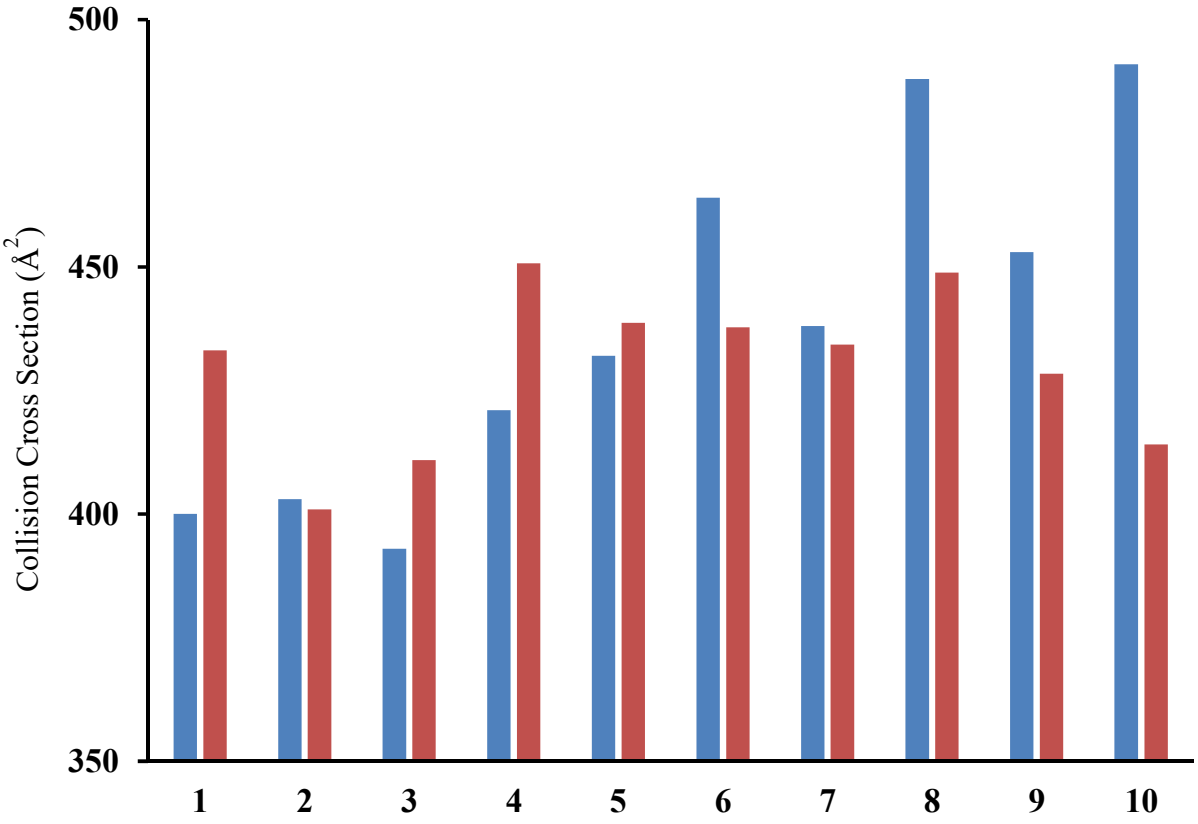

Figure 2-9. Bar graphs showing the calculated CCS values (Ω_{total}) for 10 selected reference structures versus their accurate values (Ω*). Red bars show the CCS value calculated for these 10 structure using the closest structure to the centroid as the representative structure and no cluster analysis. Blue bars are representative of the accurate CCS values (Ω*) for these reference conformers.

The comparison of the no clustering method introduces the question of how clustering affects the overall accuracy. The Ω_{total} values have also been determined using the centroids generated after 10-*k* and 50-*k* cluster analysis. Plotting the average potential energy of the trajectories versus their calculated Ω_{total} values

(Figure 2-10) reveals a high degree of similarity in energy-CCS distributions obtained from the 10-*k* and 50-*k* cluster analysis.

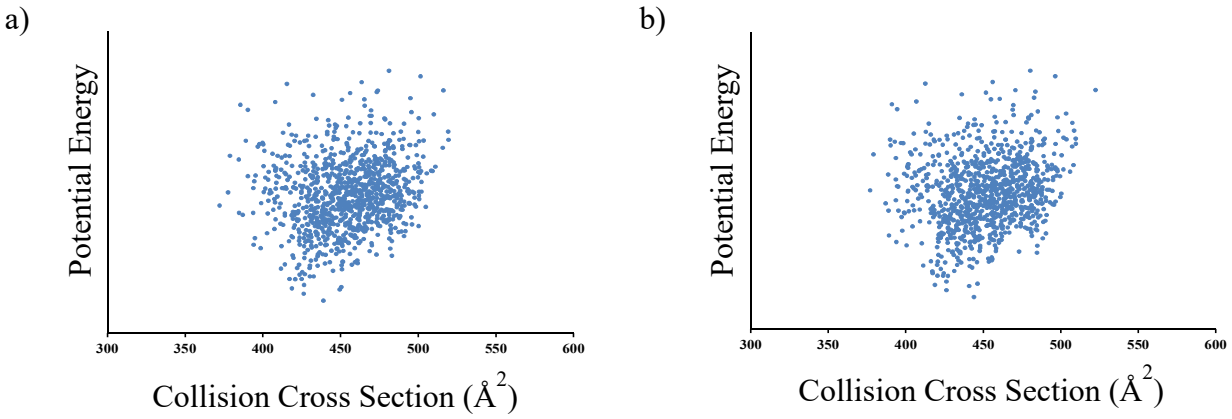

Figure 2-10. CCS values calculated for structure types versus the average potential energy. Structures resulting from production MD runs have been backbone-only RMSD aligned prior to the clustering. The left panel shows the results for the 10-*k* clusters, while the right panel provides results for the 50-*k* clusters.

For direct comparisons among the accuracy achieved from the various cluster analysis methods, the Ω_{total} values obtained from different clustering methods for 10 reference trajectories were compared with their corresponding Ω* values. Figure 2-11 shows these comparisons as well as the CCS values for each of these 10 reference structures. As mentioned above, these 0 K (annealed) structures do not convey information about peptide ion dynamics and therefore do not accurately capture the transport properties of the conformers at 300 K as indicated by the CCS value differences shown in Figure 2-11.

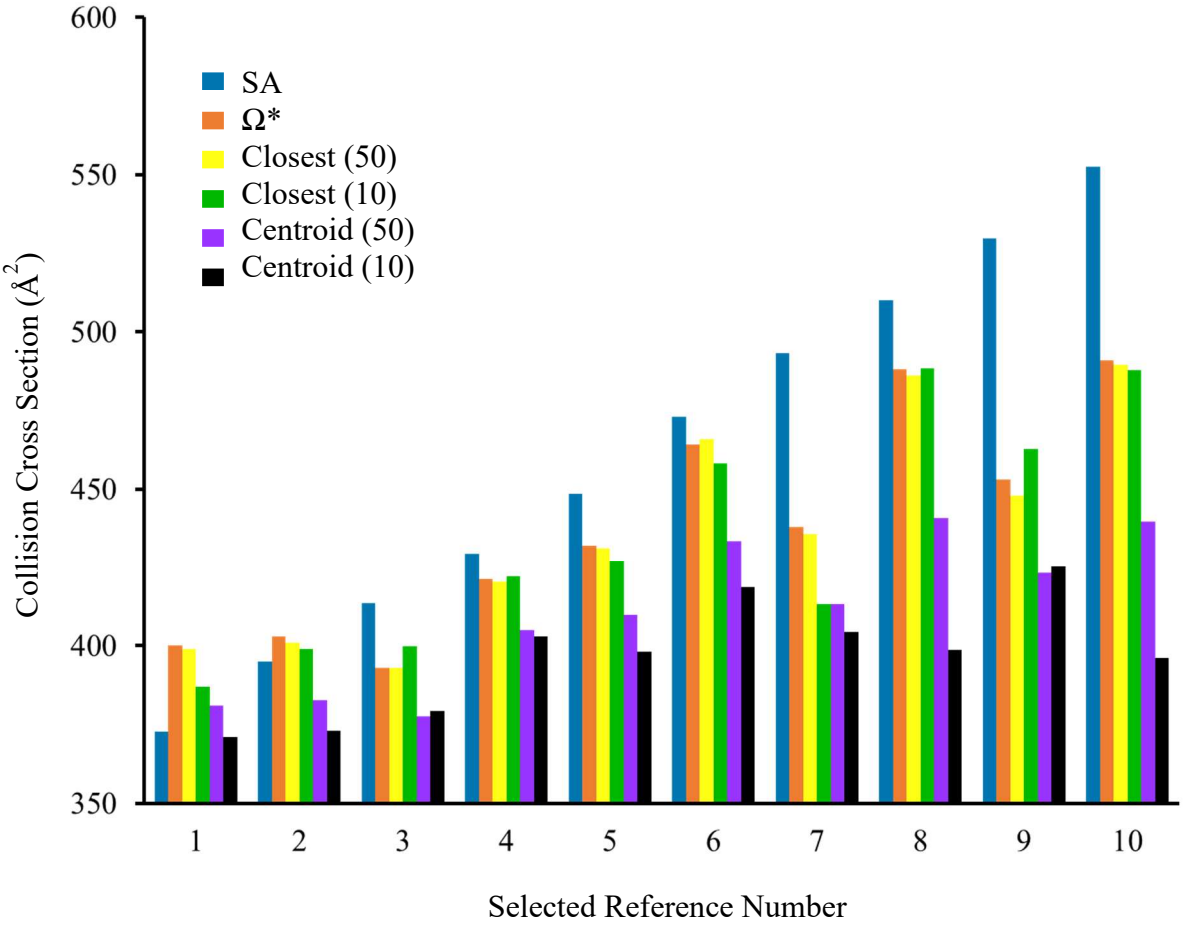

Figure 2-11. Bar graph showing the comparison of calculated CCS values for 10 selected reference trajectories (or annealed structures) obtained from different clustering approaches. The legend illustrates the bar colors corresponding to the CCS values for the 10 annealed structures from SA runs (SA); the accurate reference CCS values for all 5,000 conformers within 10 reference MD trajectories (Ω*); the Ω_{total} values for the 50-*k* clustering method using the closest structure to the centroid [closest (50)]; the Ω_{total} values for a *k*=10 clustering method using the closest structure to the centroid [closest (10)]; the Ω_{total} values for a *k*=50 clustering method using centroids [Centroid (50)]; and the Ω_{total} values for 10-*k* clustering method using the centroids [Centroid (10)].

A useful comparison for this discussion is that of the Ω_{total} values calculated after cluster analysis with *k* values of 10 and 50 while using the centroids as the representative structures. In comparison with Ω*, the Ω_{total} values for the former clustering method lead to a 9.5 % error on average with a maximum value of 23 %. These values decrease to 6.3 % and 10 %, respectively, for the latter cluster analysis (*k*=50). Such large error values support the claim of the inadequacy of centroids to represent the clusters and, by extension, the overall trajectory CCS values.

The Ω_{total} values of the 10-*k* and 50-*k* cluster analyses obtained while utilizing the closest structures to the centroids as the cluster-representative members are also shown in Figure 2-11. Overall, the 10-*k* results exhibit a larger deviation from Ω*. The average error associated with this analysis is 1.7% with a maximum value of 5.6 %. For the same dataset, as the number of clusters generated increases to 50, the error decreases to 0.39 % and 0.97 % as the average and maximum values, respectively.

Overall, the selection of the closest structure to the centroid as the cluster-representative conformer rather than the centroid itself, not only improves the overall accuracy but also captures the change in accuracy between the 10-*k* and 50-*k* cluster analyses (Figure 2-11). Thus the approach can improve the computation time by utilizing a cluster analysis with lower *k* values depending on an acceptable error threshold. Considering the error threshold of 1% to 2% in our experimental analyses, the results generated with the 50-*k*, closest-structure-to-

the-centroid clustering method will be employed in future studies as described in the Chapter 3.

2.3.9. Candidate structures

Experimental CCS values can be calculated from drift time values for dataset features [10]. A time increment of 0.1 ms has been utilized to scan the drift time selection for these ions; therefore, a maximum error of 0.05 ms can be associated with the arrival time measurements. Here it is noted that this represents an upper limit. The drift time measurements are checked daily for accuracy by comparison to $[M+3H]^{3+}$ and $[M+2H]^{2+}$ angiotensin I and bradykinin ions, respectively. Collision cross sections agree with reported values to <±1 %. Notably, for triplicate analyses of these ions, scaled drift times (accounting for buffer gas pressure) were not observed to change (i.e., <0.1 ms). For the $[M+3H]^{3+}$ peptide ions three different conformer types are observed. The conformer type having the greatest intensity was chosen for the gas-phase HDX measurements. Therefore, the focus here is on the compact conformer type (Ω= 417$\pm$3 Å^2) and the slightly more diffuse conformer type (Ω= 438$\pm$3 Å^2).

Ω_{total} values have been determined for four distinct sets of simulations including two different charge arrangements and two different conformational space sampling methods of 40-ps and 1200-ps SA runs (see Experimental section). These calculated values for the corresponding trajectories have been compared with experimental CCS values to obtain a match for the compact and more elongated ion conformations (Figure 2-4). The production MD trajectories resulting in matching CCS values have been selected as the candidate-structure-

originating trajectories. For the compact conformer type, 63 (out of 4,000) trajectories fall within the CCS range for compact ions, and 261 trajectories result in Ω_{total} values that match the CCS value of the more diffuse conformer type. Figure 2-12 shows two structures for each conformer type with different protonation sites. Notably, these conformations are the annealed structures at 0 K that provided trajectories with matching CCS from dynamics at 300 K. Two candidate structures with matching CCS values for the compact conformer contain protonation sites of K(6), K(11), and K(21). The first structure exhibits α-helical secondary structure extending from the K(6) to the K(16) residue, while the second structure exhibits primarily a random coil conformer type. For the compact conformer type having the protonation site of K(6), K(16), and K(21), differences in structures are also observed (Figure 2-12). One displays a turn in the center while the second has two turns at the N-terminus and the C-terminus. For the more diffuse conformer types, Figure 2-12 shows examples in which one structure exhibits more helical nature while the second contains more random coil characteristics for both charge configurations. Notably, all other candidate structures exhibit a wide degree of structural difference. Overall, structures with significant α-helical character exhibit greater stability for this model peptide. This may be expected based on the high α-helix propensity of a polyalanine peptide and the ability to preserve such structures into the gas phase [9, 66].

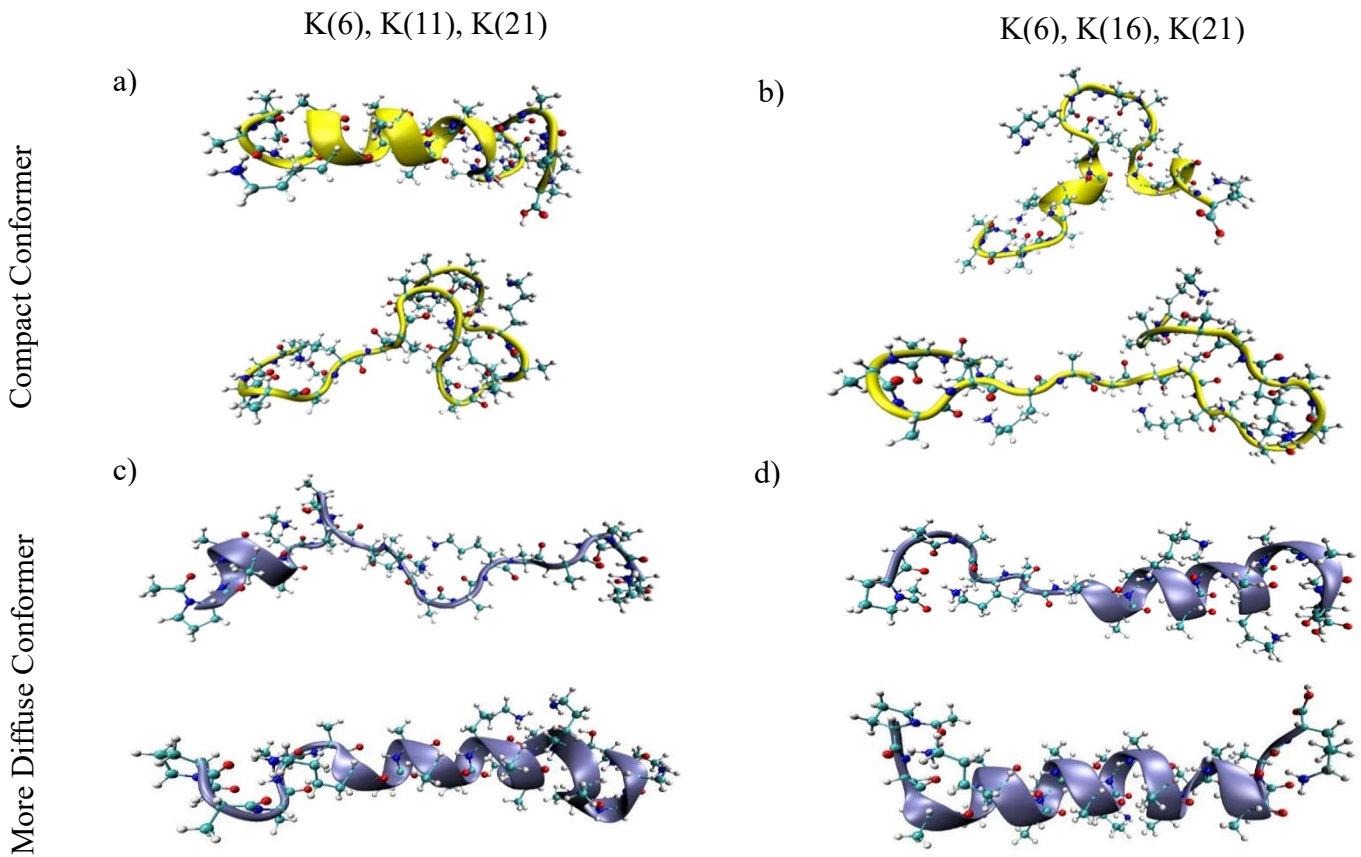

Figure 2-12. Several *in-silico* structures (annealed) with matching CCS values to experimentally determined CCS values for compact (top panels) and diffuse $[M+3H]^{3+}$ ions (bottom panels) of the model peptide Acetyl-PAAAAKAAAAKAAAAKAAAAK. The left panels show structures with protonation sites of K(6), K(11), K(21), and the right panels show those with protonation sites of K(6), K(16), K(21). The structure were obtained from the higher-energy and thermodynamically-stable pool of *in-silico* data. The CCS comparisons were performed between experimentally-obtained CCS values and the calculated Ω_{total} values for trajectories at 300 K and originating from the annealed structures.

A major aim of this study has been to find the best structure (or structures) with significant similarity to the gas-phase conformer types. Figure 2-12 demonstrates that another criterion for filtering candidate structures is necessary. It may be argued that potential energy values can be utilized to find the most stable gas-phase structure and thus be presented as the best match. This argument would be correct if the ions presented equilibrated gas-phase structures, and, as mentioned above, some studies provide evidence contradicting this assumption [6, 18, 22]. In addition to this challenge, the reliability of the comparison to CCS values alone cannot be evaluated. That is, the method results with an *in-silico* structure with no further provision to check the accuracy of the match. It can also be argued that the structures sampled by the MD procedure may not contain species resembling the actual gas-phase conformer type produced in the experiment; that is, there is no guarantee that conformational space sampling produces all types of structures.

To begin addressing these issues, gas-phase HDX has been introduced as a second criterion for structural characterization [26-28, 67, 68]. In Chapter 3, comparisons of the experimental deuterium uptake values and the hypothetical values (obtained from a modified algorithm assessing the accessibility of hydrogens on candidate structures) are used to improve conformer selection and to assign relative population values along a MD trajectory.

2.3.10. Matching Solution and Gas-phase Structures

Figure 2-13 shows the presence of structures with a high degree of similarity to solution structures.

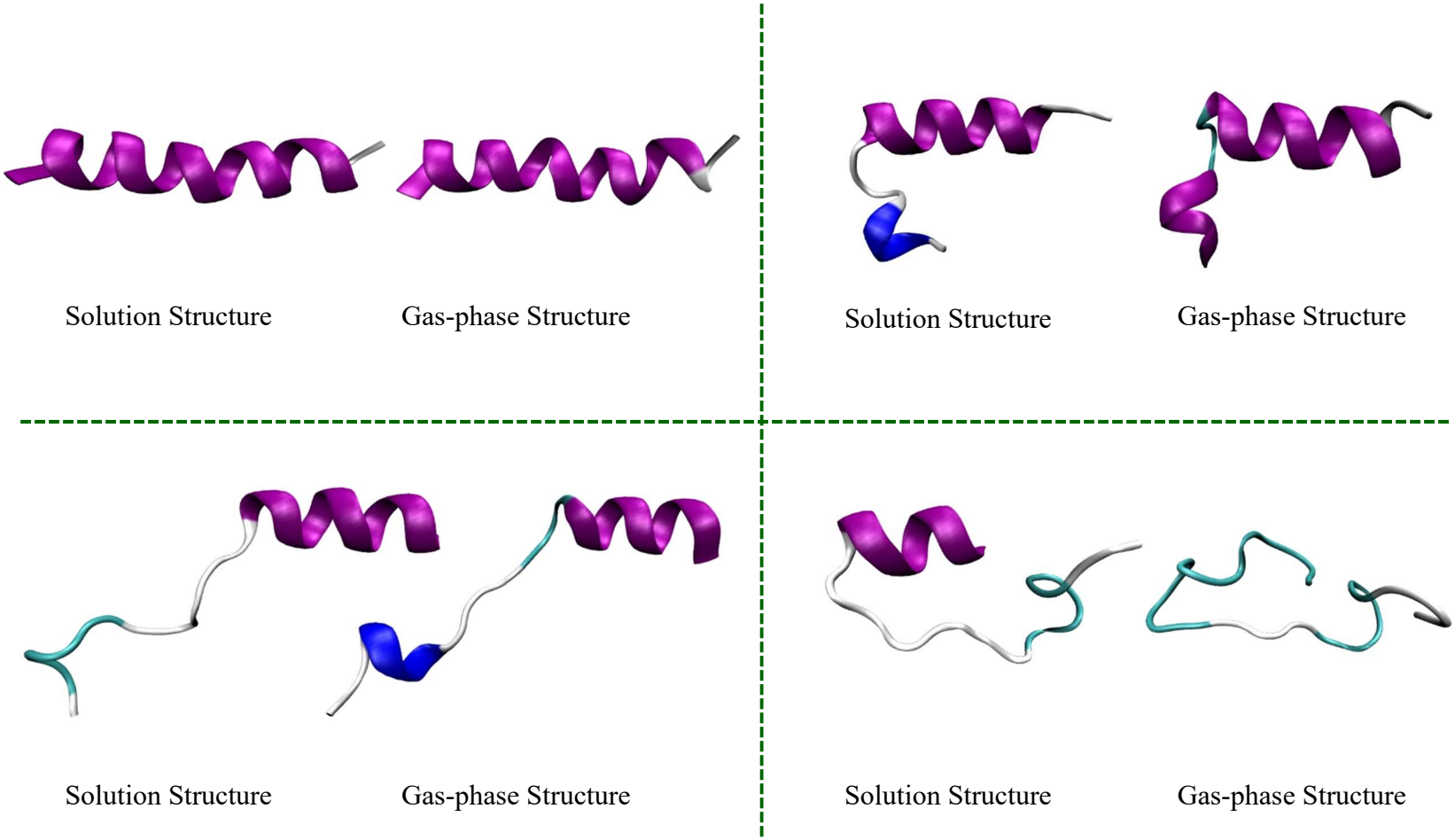

Figure 2-13. Comparison between solution and gas-phase structures for 4 different conformer types. The solution structures were obtained after clustering the trajectory obtained from 0.5 μs production MD in explicit solvent and in the NPT ensemble of 300 K. The gas-phase structures with the lowest RMSD values relative to these 4 conformer types are illustrated.

The latter have been obtained from MD simulations in explicit water that have been cross checked with CD spectroscopy data (Figure 2-14).

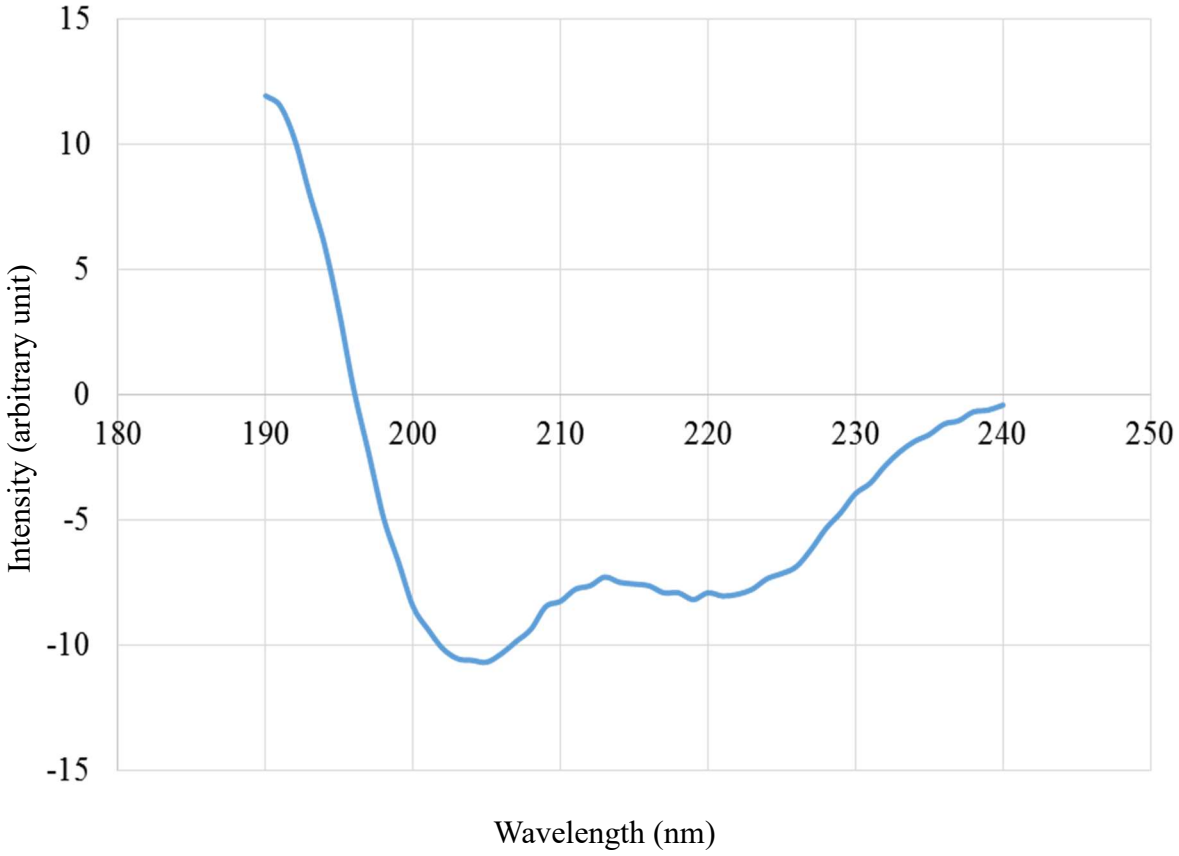

Figure 2-14. CD spectrum of the synthetic model peptide Acetyl-PAAAAKAAAAKAAAAKAAAAK.

To generate the solution structures using the Amber ff12SB force field, an initial, fully-helical structure of the model peptide with amino acid residues exhibiting charge state values corresponding to those expected at pH=7 in water

solvent was generated. The structure was neutralized with Cl^- as the counter ion and solvated using truncated octahedral box of TIP3P water and relaxed [69]. During a 2-ns timescale and under the canonical ensemble in the NVT condition, the system was gradually heated to 300 K using Langevin dynamics with a collision frequency of 1.0 ps^{-1} [48, 70]. This was followed by a 40 ns equilibration at isothermal–isobaric ensemble (NPT) with an average target pressure of 1 atm. The obtained structure was subjected to 0.5 μs production MD in the NPT ensemble with a periodic boundary condition. The non-bonded cutoff for calculating the long-range electrostatic and van der Waals interactions was set to 12 Å. All the simulations were performed using the AMBER12 software package [38].

After removal of water molecules, the generated trajectory was subjected to dihedral cluster analysis with a mask on ψ and φ dihedral angles passing through 10 central amino acid residues. This lead to formation of 4 clusters. The most similar conformer to the centroid of each cluster was selected as a solution-phase representative structure shown in Figure 2-13.

Using backbone-only RMSD calculations and the *cpptraj* module [49] implemented in the AMBER12 software package [38], the closest structure to the centroids obtained from production MD simulations in the gas-phase and cluster analysis with k=50 were compared with the newly-generated representative solution-phase structures. The most similar conformations (lowest RMSD values) were extracted as the solution-like gas-phase structures.

2.4. Conclusion

CCS measurements coupled with MD simulations can serve as a tool for biomolecule structure investigations. Using this approach, a model peptide is employed to develop an advanced protocol for MD simulations. Extensive parameter optimization and method validation are utilized to perform comprehensive sampling of higher-energy and more thermodynamically-stable structures which exhibited an extended range of structural variety. Therefore, the sampled structures accounted for a wide range of structure types that may be produced by electrospray ionization (Figure 2-13).

MD simulations at 300 K have been carried out to monitor gas-phase ion dynamics. CCS calculations reveal that no single structure is representative of the variety of conformers accessed by structural fluctuations. However, careful clustering and data mining can be used to obtain accurate CCS values. Although, as shown here, the data mining with clustering that employs the closest structure to the centroid as a representative structure provides an answer to this problem, the number of clusters plays an important role in the accuracy of the calculated CCS values. Here, it is proposed to use cluster numbers associated with the elbow-point criterion for improved accuracy. CCS values calculated using the optimized parameters exhibit < 0.1 % error compared to the accurate reference (Ω^*) values.

Experimental CCS values are used to filter the sampled structures and result in more than 300 nominal structures. A wide variety of structures, many with completely different secondary structural aspects, meet the CCS filtering criterion.

From the experimental data, there is no preference for selection among these structure or to assign relative population numbers for these species. Gas-phase HDX is proposed to serve as a second criterion to begin addressing this problem. Upcoming work (Chapter 3) will show the application of HDX filtering for assigning the populations of structures from different structural types. Hydrogen accessibility modeling will be utilized to produce a hypothetical deuterium uptake pattern for each *in-silico* structure from which a population number can be allotted to structure types.

2.5. References

1. Marx, D., et al., *The nature of the hydrated excess proton in water.* Nature, 1999. **397**(6720): p. 601-604.
2. Shin, J.W., et al., *Infrared signature of structures associated with the H+(H2O)(n) (n=6 to 27) clusters.* Science, 2004. **304**(5674): p. 1137-1140.
3. McAllister, R.G. and L. Konermann, *Challenges in the Interpretation of Protein H/D Exchange Data: A Molecular Dynamics Simulation Perspective.* Biochemistry, 2015. **54**(16): p. 2683-2692.
4. Skinner, J.J., et al., *Protein hydrogen exchange: Testing current models.* Protein Science, 2012. **21**(7): p. 987-995.
5. Hoaglund-Hyzer, C.S., A.E. Counterman, and D.E. Clemmer, *Anhydrous protein ions.* Chemical Reviews, 1999. **99**(10): p. 3037-3079.
6. Fort, K.L., et al., *From Solution to the Gas Phase: Factors That Influence Kinetic Trapping of Substance P in the Gas Phase.* Journal of Physical Chemistry B, 2014. **118**(49): p. 14336-14344.
7. Pepin, R., et al., *Toward a Rational Design of Highly Folded Peptide Cation Conformations. 3D Gas-Phase Ion Structures and Ion Mobility Characterization.* Journal of The American Society for Mass Spectrometry, 2016: p. 1-14.
8. Fernandez-Lima, F.A., et al., *On the Structure Elucidation Using Ion Mobility Spectrometry and Molecular Dynamics.* Journal of Physical Chemistry A, 2009. **113**(29): p. 8221-8234.
9. Kohtani, M., et al., *Proton transfer-induced conformational changes and melting in designed peptides in the gas phase.* Journal of the American Chemical Society, 2006. **128**(22): p. 7193-7197.
10. Mason, E.A. and E.W. Mcdaniel, *Transport properties of ions in gases*. 1988: John Wiley & sons.
11. Clemmer, D.E., R.R. Hudgins, and M.F. Jarrold, *NAKED PROTEIN CONFORMATIONS - CYTOCHROME-C IN THE GAS-PHASE.* Journal of the American Chemical Society, 1995. **117**(40): p. 10141-10142.
12. Shvartsburg, A.A. and M.F. Jarrold, *An exact hard-spheres scattering model for the mobilities of polyatomic ions.* Chemical Physics Letters, 1996. **261**(1-2): p. 86-91.
13. Shvartsburg, A.A., G.C. Schatz, and M.F. Jarrold, *Mobilities of carbon cluster ions: Critical importance of the molecular attractive potential.* Journal of Chemical Physics, 1998. **108**(6): p. 2416-2423.
14. Shelimov, K.B., et al., *Protein structure in vacuo: Gas-phase confirmations of BPTI and cytochrome c.* Journal of the American Chemical Society, 1997. **119**(9): p. 2240-2248.
15. Wyttenbach, T. and M.T. Bowers, *Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-Free Ion Mobility-Mass Spectrometry Environment.* Journal of Physical Chemistry B, 2011. **115**(42): p. 12266-12275.
16. Chen, S.-H. and D.H. Russell, *How Closely Related Are Conformations of Protein Ions Sampled by IM-MS to Native Solution Structures?* Journal of the American Society for Mass Spectrometry, 2015. **26**(9): p. 1433-1443.
17. Gidden, J. and M.T. Bowers, *Gas-phase conformational and energetic properties of deprotonated dinucleotides.* European Physical Journal D, 2002. **20**(3): p. 409-419.

18. Pierson, N.A., S.J. Valentine, and D.E. Clemmer, *Evidence for a Quasi-Equilibrium Distribution of States for Bradykinin M+3H (3+) Ions in the Gas Phase.* Journal of Physical Chemistry B, 2010. **114**(23): p. 7777-7783.
19. Heck, A.J.R. and R.H.H. van den Heuvel, *Investigation of intact protein complexes by mass spectrometry.* Mass Spectrometry Reviews, 2004. **23**(5): p. 368-389.
20. Heck, A.J.R., *Native mass spectrometry: a bridge between interactomics and structural biology.* Nature Methods, 2008. **5**(11): p. 927-933.
21. Ruotolo, B.T., et al., *Evidence for macromolecular protein rings in the absence of bulk water.* Science, 2005. **310**(5754): p. 1658-1661.
22. Silveira, J.A., et al., *From Solution to the Gas Phase: Stepwise Dehydration and Kinetic Trapping of Substance P Reveals the Origin of Peptide Conformations.* Journal of the American Chemical Society, 2013. **135**(51): p. 19147-19153.
23. Pierson, N.A., et al., *Number of Solution States of Bradykinin from Ion Mobility and Mass Spectrometry Measurements.* Journal of the American Chemical Society, 2011. **133**(35): p. 13810-13813.
24. Hudgins, R.R., J. Woenckhaus, and M.F. Jarrold, *High resolution ion mobility measurements for gas phase proteins: correlation between solution phase and gas phase conformations.* International Journal of Mass Spectrometry, 1997. **165**: p. 497-507.
25. Shi, H.L. and D.E. Clemmer, *Evidence for Two New Solution States of Ubiquitin by IMS-MS Analysis.* Journal of Physical Chemistry B, 2014. **118**(13): p. 3498-3506.
26. Khakinejad, M., et al., *Gas-Phase Hydrogen-Deuterium Exchange Labeling of Select Peptide Ion Conformer Types: a Per-Residue Kinetics Analysis.* Journal of The American Society for Mass Spectrometry, 2015. **26**(7): p. 1115-1127.
27. Khakinejad, M., et al., *Ion Mobility Spectrometry-Hydrogen Deuterium Exchange Mass Spectrometry of Anions: Part 2. Assessing Charge Site Location and Isotope Scrambling.* Journal of the American Society for Mass Spectrometry, 2016. **27**(3): p. 451-461.
28. Khakinejad, M., et al., *Combining Ion Mobility Spectrometry with Hydrogen-Deuterium Exchange and Top-Down MS for Peptide Ion Structure Analysis.* Journal of the American Society for Mass Spectrometry, 2014. **25**(12): p. 2103-2115.
29. Mesleh, M.F., et al., *Structural information from ion mobility measurements: Effects of the long-range potential.* Journal of Physical Chemistry, 1996. **100**(40): p. 16082-16086.
30. Jarrold, M.F. *Mobcal, A program to calculate mobilities*. 2/10/2014 2/18/2016]; Available from: http://www.indiana.edu/~nano/software.html.
31. Wyttenbach, T., et al., *Ion Mobility Analysis of Molecular Dynamics*, in *Annual Review of Physical Chemistry, Vol 65*, M.A. Johnson and T.J. Martinez, Editors. 2014. p. 175-196.
32. Breuker, K. and F.W. McLafferty, *Stepwise evolution of protein native structure with electrospray into the gas phase, 10(-12) to 10(2) S.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(47): p. 18145-18152.
33. Konermann, L., et al., *Unraveling the Mechanism of Electrospray Ionization.* Analytical Chemistry, 2013. **85**(1): p. 2-9.
34. Konermann, L., R.G. McAllister, and H. Metwally, *Molecular Dynamics Simulations of the Electrospray Process: Formation of NaCI Clusters via the Charged Residue Mechanism.* Journal of Physical Chemistry B, 2014. **118**(41): p. 12025-12033.
35. Yue, X.F., S. Vahidi, and L. Konermann, *Insights into the Mechanism of Protein Electrospray Ionization From Salt Adduction Measurements.* Journal of the American Society for Mass Spectrometry, 2014. **25**(8): p. 1322-1331.
36. Hogan, C.J., et al., *Combined Charged Residue-Field Emission Model of Macromolecular Electrospray Ionization.* Analytical Chemistry, 2009. **81**(1): p. 369-377.

37. Donohoe, G.C., et al., *A New Ion Mobility-Linear Ion Trap Instrument for Complex Mixture Analysis.* Analytical Chemistry, 2014. **86**(16): p. 8121-8128.
38. Case, D.A., Darden T.A.; Cheatham, T E. III,; Simmerling, C.L.;, et al. *Amber12, University of California, San Francisco*. 2012.
39. Schmidt, M.W., et al., *GENERAL ATOMIC AND MOLECULAR ELECTRONIC-STRUCTURE SYSTEM.* Journal of Computational Chemistry, 1993. **14**(11): p. 1347-1363.
40. Gordon, M.S. and M.W. Schmidt, *Advances in electronic structure theory: GAMESS a decade later.* Theory and Applications of Computational Chemistry: The First Forty Years, 2005: p. 1167-1189.
41. Vanquelef, E., et al., *RED Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments.* Nucleic Acids Research, 2011. **39**: p. W511-W517.
42. Wang, F., et al., *RED Python: Object oriented programming for Amber force fields.* Abstracts of Papers of the American Chemical Society, 2014. **247**: p. 1.
43. Dupradeau, F.-Y., et al., *The R.ED. tools: advances in RESP and ESP charge derivation and force field library building.* Physical Chemistry Chemical Physics, 2010. **12**(28): p. 7821-7839.
44. Bayly, C.I., et al., *A WELL-BEHAVED ELECTROSTATIC POTENTIAL BASED METHOD USING CHARGE RESTRAINTS FOR DERIVING ATOMIC CHARGES - THE RESP MODEL.* Journal of Physical Chemistry, 1993. **97**(40): p. 10269-10280.
45. Wyttenbach, T., G. von Helden, and M.T. Bowers, *Gas-Phase Conformation of Biological Molecules: Bradykinin.* Journal of the American Chemical Society, 1996. **118**(35): p. 8355-8364.
46. Berendsen, H.J.C., et al., *MOLECULAR-DYNAMICS WITH COUPLING TO AN EXTERNAL BATH.* Journal of Chemical Physics, 1984. **81**(8): p. 3684-3690.
47. Wu, X. and B.R. Brooks, *Self-guided Langevin dynamics simulation method.* Chemical Physics Letters, 2003. **381**(3–4): p. 512-518.
48. Uberuaga, B.P., M. Anghel, and A.F. Voter, *Synchronization of trajectories in canonical molecular-dynamics simulations: Observation, explanation, and exploitation.* Journal of Chemical Physics, 2004. **120**(14): p. 6363-6374.
49. Roe, D.R. and T.E. Cheatham, *PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data.* Journal of Chemical Theory and Computation, 2013. **9**(7): p. 3084-3095.
50. Feig, M., J. Karanicolas, and C.L. Brooks, *MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology.* Journal of Molecular Graphics & Modelling, 2004. **22**(5): p. 377-395.
51. Bruccoleri, R.E. and M. Karplus, *CONFORMATIONAL SAMPLING USING HIGH-TEMPERATURE MOLECULAR-DYNAMICS.* Biopolymers, 1990. **29**(14): p. 1847-1862.
52. Levy, Y. and O.M. Becker, *Energy landscapes of conformationally constrained peptides.* The Journal of Chemical Physics, 2001. **114**(2): p. 993-1009.
53. Baysal, C. and H. Meirovitch, *Efficiency of simulated annealing for peptides with increasing geometrical restrictions.* Journal of Computational Chemistry, 1999. **20**(15): p. 1659-1670.
54. Frishman, D. and P. Argos, *Knowledge-based protein secondary structure assignment.* Proteins-Structure Function and Genetics, 1995. **23**(4): p. 566-579.
55. Schreiner, E., et al., *Stereochemical errors and their implications for molecular dynamics simulations.* Bmc Bioinformatics, 2011. **12**: p. 9.
56. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal of Molecular Graphics & Modelling, 1996. **14**(1): p. 33-38.

57. Von Helden, G., T. Wyttenbach, and M.T. Bowers, *INCLUSION OF A MALDI ION-SOURCE IN THE ION CHROMATOGRAPHY TECHNIQUE - CONFORMATIONAL INFORMATION ON POLYMER AND BIOMOLECULAR IONS.* International Journal of Mass Spectrometry, 1995. **146**: p. 349-364.
58. Valentine, S.J. and D.E. Clemmer, *Temperature-dependent H/D exchange of compact and elongated cytochrome c ions in the gas phase.* Journal of the American Society for Mass Spectrometry, 2002. **13**(5): p. 506-517.
59. Viehland, L.A. and E.A. Mason, *GASEOUS ION MOBILITY IN ELECTRIC-FIELDS OF ARBITRARY STRENGTH.* Annals of Physics, 1975. **91**(2): p. 499-533.
60. Hartigan, J.A. and M.A. Wong, *Algorithm AS 136: A K-Means Clustering Algorithm.* Journal of the Royal Statistical Society. Series C (Applied Statistics), 1979. **28**(1): p. 100-108.
61. Team, R.C., *R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.* 2013.
62. Faber, V., *Clustering and the continuous k-means algorithm.* Los Alamos Science, 1994. **22**(138144.21).
63. Kodinariya, T.M. and P.R. Makwana, *Review on determining number of Cluster in K-Means Clustering.* International Journal, 2013. **1**(6): p. 90-95.
64. Chou, K.C. and L. Carlacci, *SIMULATED ANNEALING APPROACH TO THE STUDY OF PROTEIN STRUCTURES.* Protein Engineering, 1991. **4**(6): p. 661-667.
65. Wilson, S.R., et al., *APPLICATIONS OF SIMULATED ANNEALING TO THE CONFORMATIONAL-ANALYSIS OF FLEXIBLE MOLECULES.* Journal of Computational Chemistry, 1991. **12**(3): p. 342-349.
66. Counterman, A.E. and D.E. Clemmer, *Large anhydrous polyalanine ions: Evidence for extended helices and onset of a more compact state.* Journal of the American Chemical Society, 2001. **123**(7): p. 1490-1498.
67. Donohoe, G.C., M. Khakinejad, and S.J. Valentine, *Ion Mobility Spectrometry-Hydrogen Deuterium Exchange Mass Spectrometry of Anions: Part 1. Peptides to Proteins.* Journal of the American Society for Mass Spectrometry, 2015. **26**(4): p. 564-576.
68. Khakinejad, M., et al., *Ion Mobility Spectrometry-Hydrogen Deuterium Exchange Mass Spectrometry of Anions: Part 3. Estimating Surface Area Exposure by Deuterium Uptake.* Journal of the American Society for Mass Spectrometry, 2016. **27**(3): p. 462-473.
69. Jorgensen, W.L., et al., *COMPARISON OF SIMPLE POTENTIAL FUNCTIONS FOR SIMULATING LIQUID WATER.* Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
70. Wu, X.W. and B.R. Brooks, *Self-guided Langevin dynamics simulation method.* Chemical Physics Letters, 2003. **381**(3-4): p. 512-518.

3. Comprehensive Peptide Ion Structure Studies Using Ion Mobility Techniques: Part 2. Gas-phase Hydrogen/Deuterium Exchange for Ion Population Estimation

Reprinted with permission from *Journal of The American Society for Mass Spectrometry:* Comprehensive gas-phase peptide ion structure studies using ion mobility techniques: Part 2. Gas-phase hydrogen/deuterium exchange for ion population estimation. Khakinejad, M., Ghassabi Kondalaji, S., Tafreshian, A., Valentine, S., Journal of The American Society for Mass Spectrometry, 2017. **28**(5): p. 960-970.

3.1. Introduction

Hydrogen deuterium exchange in the gas phase has been comprehensively studied [1-4] and utilized for protein, peptide and biomolecule ion structure investigations [5-14]. In one pioneering study, D_2O as the exchange reagent was introduced into a Fourier transform ion cyclotron resonance (FTI-CR) mass spectrometer cell. Changes in the isotopic envelope of the gas-phase ions and the number of incorporated deuteriums indicated the existence of co-existing, stable gas-phase structures [5]. Later the same methodology was utilized to investigate gas-phase structural transformations of cytochrome *c* ions [6]. A comprehensive mechanistic investigation utilized different HDX reagents [1]. Quantum mechanics calculations revealed the importance of the reagent gas-phase basicity upon the exchange process, resulting in proposal of a relay mechanism for HDX with D_2O reagent.

The gas-phase HDX behavior of select ion conformers was reported shortly later in which ion injection into a drift tube was used to favor elongated and

compact cytochrome *c* ions [7]. Differences in HDX levels were attributed to the potential for sampling different conformations by the different instruments as well as the timescale of the ICR measurement which could be sensitive to longer timescale protein ion fluctuations. That is, the long timescale of HDX in a FT-ICR cell may provide sufficient time for an ion to widely sample conformational space [15] and adopt different conformers affecting the HDX behavior of the ions [9]. In separate experiments, HDX measurements in a FT-ICR instrument were used to confirm a proposed structural type for $[M+H]^+$ bradykinin ions [2]. Here, molecular dynamics (MD) simulations were utilized to model the HDX behavior of an ensemble of structures of the salt-bridge conformer type [16]; the model was based upon the relay mechanism proposed for HDX with D_2O reagent gas. IMS and HDX measurements also examined rate constants and deuterium uptake values at different buffer gas temperatures for cytochrome c ions [3]. The experiments showed that, although the reaction with D_2O is slower at elevated temperature, maximum deuterium uptake values are higher compared with those at lower temperatures. The lower reaction rate at higher temperature was attributed to decreased formation of the D_2O-protein complex while increased HDX levels indicated greater accessibility to more distal sites due to molecular fluctuations. Indeed, MD simulations at 300 and 500 K in which an exchange threshold distance between labile hydrogens and protonation sites was utilized, suggested increased accessibility at higher temperatures. That is, the increase in uptake value at elevated temperature was explained by higher structural flexibility [3].

The advent of non-ergodic ion dissociation methods like electron capture dissociation (ECD) [17] and electron transfer dissociation (ETD) [18] provided the opportunity to investigate deuterium uptake by individual amino acid residues [8, 19-22] in the absence of isotopic scrambling [23, 24]. Recently, the first, collision cross section measurements coupled with gas-phase HDX and tandem mass spectrometry (MS/MS) were utilized to determine per-residue deuterium uptake values of select ion conformer types of a peptide [8]. A simple model based on heteroatom hydrogen site accessibility (distance) to the charge site and the original site of deuterium incorporation was introduced to consider *in-silico* structures as viable candidates to represent ion conformer type. Patterns in per-residue HDX levels served as a second criterion for structure elucidation in addition to CCS filtering. Uptake values using this model could not completely describe the HDX behavior of the peptide ions in part because of saturation effects at higher partial pressures of D_2O. To address this problem, in separate studies, the contribution of each residue to the HDX rate for select conformer types was measured; using the same distance model, theoretical contributions to rate better approximated measured contributions for *in-silico* structures intended to represent conformer type [20]. Because measuring the contributions to rate constants at the individual amino acid residue level is an extremely time-intensive process and difficult to obtain for larger ions, an effective collision model was proposed to describe deuterium uptake levels at different partial pressures of D_2O [25, 26] for select ion conformers.

Ion structure dynamics can significantly affect the overall drift time distribution [27, 28] as well as the HDX behavior of peptides and proteins [2, 3]. Therefor improvement to the accessibility model would account for such fluctuations. Here we utilize production MD to mimic structural fluctuations of each *in-silico* conformer in order to obtain a more realistic depiction of its reactivity with D_2O. Using two cutoff distances, the accessibility of each exchange site is determined. Additionally, the relative exposed surface area of carbonyls and charge sites is implemented in a theoretical hydrogen accessibility scoring (HAS) method. Using HAS scoring coupled with a number of effective collisions (NEC) model, the deuterium uptake by individual residues is predicted for $[M+3H]^{3+}$ ions of the model peptide acetyl-PAAAAKAAAAKAAAAKAAAAK. Overall, multiple *in-silico* structures are required to explain the experimental HDX uptake patterns and the ETD spectra of the model peptide. Non-negative linear regression (NNLR) is applied to the theoretical uptake values to obtain population levels for proposed structures. In a subsequent report, comparisons of solution- (from CD spectroscopy and MD simulations) and gas-phase structures will be performed. Finally, pathways from solution to gas-phase structures will be investigated to gain insight into gas-phase ion conformer establishment as well as operative ESI mechanisms (Chapter 4).

3.2. Experimental

3.2.1. Sample Preparation

The model peptide Acetyl-PAAAAKAAAAKAAAAKAAAAK (>90% purity) was purchased (Genscript, Piscataway, NJ, USA) and used without further purification.

1 mg/mL stock solutions were prepared by dissolving 1 mg of the model peptide in Mili-Q water. ESI solutions were prepared by performing a 1:10 dilution of the stock solution with a 100 mM solution of ammonium acetate in water. ESI solution were infused (0.5 $\mu L \cdot min^{-1}$) through a pulled-tipped capillary biased 2200 V relative to the hybrid ion mobility spectrometry-mass spectrometry (IMS-MS) instrument entrance [8, 29].

3.2.2. IMS-MS Measurements

A detailed description of drift time measurements is presented in the first manuscript associated with this work (Section 2-2-2) [30]. A home-built, dual gate IMS device coupled to a linear ion trap (LIT) mass spectrometer (LTQ Velos, ThermoScientifc, San Jose, CA, USA) was employed [8, 29]. Delay times between IMS gates were scanned (100 μs time increments) to obtain drift time resolved mass spectra from which drift time distributions could be extracted (Figure 2-4). 400 to 1000 was used as the linear ion trap (LIT) mass-to-charge (*m*/z) range. For each drift time window, mass spectra were recorded for 0.5 minutes.

3.2.3. Peptide Ion Dissociation by ETD

Mobility-selected conformations were subjected to ETD to determine deuterium incorporation at specific amino acid residues. The source filament was set at 90 °C and the reagent (Fluoranthrene) chamber pressure was increased to 20×10^{-5} Torr using N_2 gas. Mobility selected precursor peptide ions were isolated by *m*/*z*. The ion injection time was maintained at 200 ms (5 microscans) for these

ETD measurements. Per-residue uptake values were obtained by subtracting the adjacent *z*-ion total uptake values.

3.2.4. Per-residue Deuterium Uptake Calculations

The deuterium content of each residue was calculated utilizing the z ions. Deuterium uptake values for each z ion were obtained by subtracting the average mass of the ion in the absence of the HDX reagent gas from that recorded for the ion with the addition of the reagent gas in the drift tube. The amount of deuterium incorporated within each residue was determined by deducting the deuterium uptake value of adjacent, lower *m*/*z* z-ions. It is worth mentioning that ETD did not produce z_1 and z_2 ions, therefore the values for these ions were calculated using the total deuterium uptake value for the peptide ions and the c_{20} and c_{19} ions, respectively. The uptake values for the proline residue (ideally obtained from c_1 ions would include the subsequent alanine backbone amide hydrogen), were calculated using the z_{20} ion and the total deuterium uptake value for the peptide ion.

3.2.5. Molecular Dynamics (MD) Simulations

The processes of *in-silico* structure generation and *in-vacuo* MD simulations have been discussed in detail in the prior related work (Section 2-2-3) [30]. Briefly, after calculating the undefined force field parameters using the R.E.D server development [31-36], the AMBER ff12SB force field was employed to generate the extended initial structures of $[M+3H]^{3+}$ peptide ions with charge arrangements of K(6)-K(11)-K(22) and K(6)-K(16)-K(22). Using the AMBER12 [37] package, the

initial structures were energy minimized and subjected to two distinct cyclic simulated annealing (SA) runs – 40-ps and 1200-ps SA algorithms – to produce a pool of annealed structures. In each cycle of the SA runs, the minimized structure was heated to 1000 K, equilibrated and gradually cooled to a lower temperature of 10 K using variant temperature coupling time constants. This was followed by a rapid, minimization-like cooling step to sample the structure at 0 K. After 1000 heating-cooling cycles, each optimized structure was heated and equilibrated at 300 K and subjected to 5 ns of MD simulations at the same temperature with no non-bonded cutoffs for long-range interactions. The resulting production MD trajectories were clustered using a modified k-means algorithm to generate 50($\pm$10) clusters as described in the previous related work [30]. The most similar conformations (nearest structures) to the mathematically-generated centroids of the clusters were selected for trajectory-method [38, 39] (TM) CCS calculations using the Mobcal [40] software. The exhibited CCS values were employed to generate the weighted-average CCS value *(Ω_{total})* representing the CCS of the complete trajectory.

3.2.6. Hydrogen Accessibility Scoring (HAS)

To assign a hypothetical per-residue deuterium uptake pattern to each *in-silico* structure type, an in-house script (Khakinejad, M.) was employed. MD simulations suggest that peptide ions exhibit considerable flexibility during their transit time in the drift tube, therefore an accurate model should account for ion structure fluctuation. The production MD described above is utilized to simulate these fluctuations. For each candidate structure the hydrogen accessibility score

(HAS) is calculated as the sum of the scores for each frame in the production MD according to Equation 3-1:

$$S_{total_j} = \sum_{i=1}^{n} S_{frame_{j,i}} \qquad Eq. 3-1$$

In Equation 3-1, S_{total_j} is the score value for the *j*th labile hydrogen on each candidate structure while $S_{frame_{j,i}}$ is the hydrogen score value for the *j*th labile hydrogen on the *i*th frame of the production MD, and *n* is the number of MD frames (5000).

Individual $S_{frame_{j,i}}$ values are computed in a manner that is similar to that described previously. The scoring approach considers the HDX mechanism for D_2O reagent gas in which a proton from a charge site transfers to the D_2O molecule with concomitant deuteron transfer to a less basic site such as a backbone carbonyl oxygen (Figure 1-3) on the peptide ion [1, 2]. Subsequently this deuterium is transferred to the exchange site. According to the HDX mechanism, the scoring strategy can be divide into two steps. In the first step, the relative propensity for deuterium incorporation on the carbonyls is determined. For the second step, the likelihood of deuteron transfer to adjacent exchange sites is considered. Each of these steps is described briefly below with regard to the generation of $S_{frame_{j,i}}$.

Based on the HDX mechanism, it is essential that the carbonyl group be located within a suitable distance for an extended period of time [3]. Therefore a threshold distance between a carbonyl oxygen and a charge site can be estimated for which exchange will occur (carbonyl-charge cutoff distance). Additionally, the

charge site and carbonyl group are required to be on the surface of the ion. The reaction cross section is increased in proportion to exposed surface area of the charge site and the carbonyl oxygen; that is the greater the surface area exposure, the greater is the number of collisions with D_2O molecules [2]. Therefore, each carbonyl group on a structure can be scored according to Equation 3-2 as:

$$S_{carbonyl_k} = \sum_{l=1}^{m} Surface_{carbonyl_k} \times Surface_{charge_l} \qquad Eq. 3-2$$

In Equation 3-2, m is the total number of charges, while $S_{carbonyl_k}$ and $Surface_{carbonyl_k}$ are the carbonyl score and the normalized surface area (scaled to the maximum possible surface area) for the k_{th} carbonyl of a structure. The term $Surface_{charg\ _l}$ is the scaled surface area of the l_{th} protonation site. To apply the cutoff distance criterion, any carbonyl more distant than the threshold value is scored as zero. Several threshold distances were examined and a detailed discussion of distance optimization results is provided in the Results and Discussion section (3-3). Notably, independent of whether the carbonyl or the charge site is buried in the peptide ion, the carbonyl score is zero.

For single ion structures, each labile hydrogen that is located within the threshold distance of an oxygen of a scored carbonyl group is scored according to Equation 3-3 as:

$$S_{frame_{j,i}} = \sum_{k=1}^{q} S_{carbonyl_k} \qquad Eq. 3-3$$

In Equation 3-3, q represents the number of carbonyls for a single structure. As mentioned above, a summation of each frame score over the entire MD

trajectory yields the overall hydrogen score for an *in-silico* candidate structure (Equation 3-1).

3.2.7. Number of Effective Collisions (NEC) Model

Under the experimental conditions for this study, depending on the partial pressure of D_2O, peptide ions typically experience thousands of collisions with D_2O gas. Measured rate constants indicate that only a very small portion of these collisions lead to reaction (here termed an effective collision). At lower partial pressures of D_2O corresponding to a relatively small number of labile hydrogens that undergo exchange, on average, each reaction leads to deuterium uptake. At higher partial pressures where larger numbers of deuterium are incorporated, the likelihood of exchanging deuterium with deuterium is not negligible; that is, not all of the effective collisions result in deuterium uptake (saturation conditions). Therefore, at very low partial pressures of D_2O gas, the NEC is directly proportional to deuterium uptake values; the NEC at higher partial pressures of D_2O is not. Because the number of ion-D_2O collisions increases linearly with D_2O partial pressure and the ratio of the total number of collisions to the NEC is constant, the NEC increases proportionally with partial pressure of D_2O according to Equation 3-4:

$$n_{effective} = \frac{p_{high}}{p_{Low}} \times n_{effective_{low}} \qquad Eq. 3-4$$

In Equation 3-4, $n_{effective}$ and $n_{effective_{low}}$ are the NEC values at high- and low-partial pressures of D_2O, respectively. $\frac{p_{high}}{p_{Low}}$ is the ratio of partial pressures of

D_2O at different leak valve settings resulting in saturation (p_{high}) and non-saturation (p_{low}) reaction conditions. Using Equation 3-4, the NEC can be calculated for different D_2O partial pressures even those under which saturation conditions may apply.

The HAS scores for particular hydrogens can be normalized for the candidate structures. A relative hydrogen score (S_{R_j}) value for each conformer type can be determined according to Equation 3-5 as:

$$S_{R_j} = \frac{S_{total_j}}{\sum_{j=1}^{n} S_{total_j}} \qquad Eq. 3-5$$

S_{R_j} reveals the contribution of each labile hydrogen to the number of HDX events occurring for the entire peptide. In Equation 3-5, *n* is the total number of labile hydrogen on the peptide ion. With this consideration it is possible to generate Equation 3-6.

$$n_{effective_j} = n_{effective} \times \quad S_{R_j} \qquad Eq. 3-6$$

that is the NEC experienced by a labile hydrogen ($n_{effective_j}$) can be estimated. It is then necessary to convert the NEC to deuterium uptake value. With $n_{effective_j}$ the behavior of an uptake site with regard to a specific number of collisions is simulated for 1000 ions. In the simulation, first an array of 1000×1 is utilized to model a population of 1000 ions containing one exchange site. All the elements in the array are set to zero and for each effective collision a random element is changed to 1. Therefore, the summation of all elements in the array is equivalent to the deuterium uptake for the 1000 ions. It is noteworthy that the uptake value

determined as a function of NEC produces a trend comparable to a pseudo-first order kinetics plot as may be expected (Figure 3-1) [5]. Using the various $n_{effective_j}$ values.

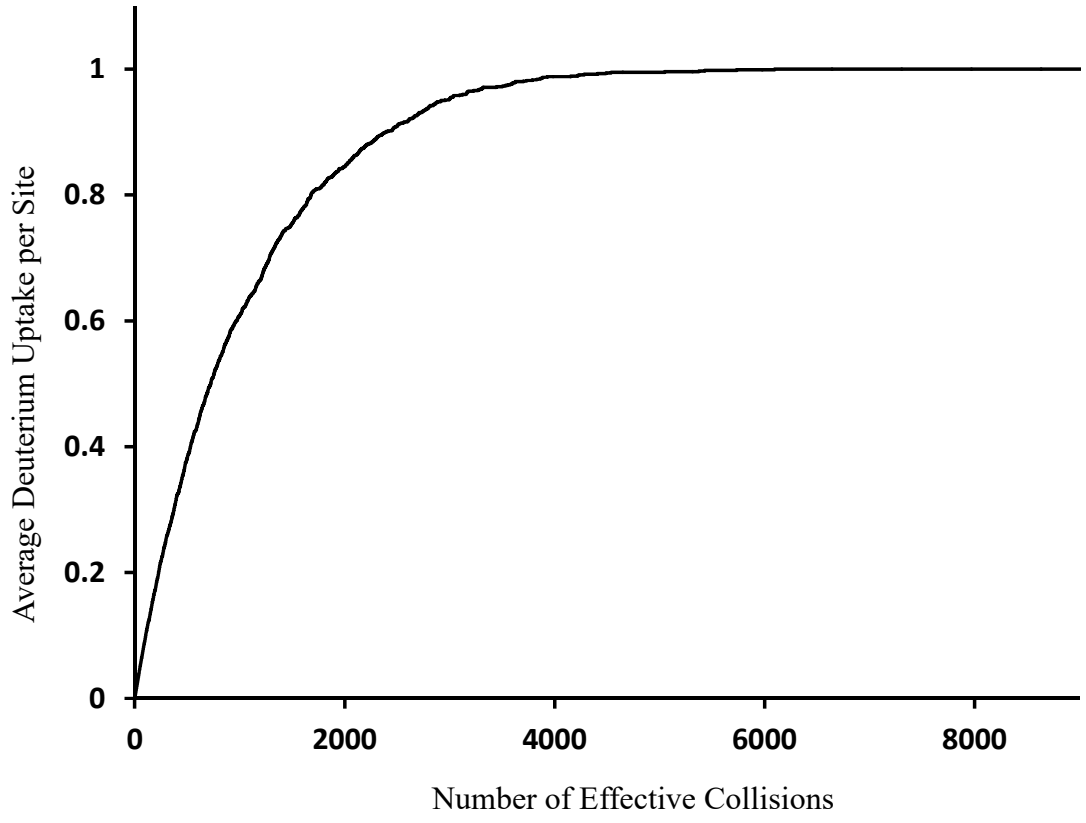

Figure 3-1. Estimated deuterium uptake values versus NEC values for one exchange site within a population of 1000 ions.

3.3. Results and Discussion

3.3.1 Non-negative Linear Regression (NNLR)

Consider a collection of n observations of a response variable assigned as the dependent variable and represented by y with the corresponding features assigned as independent or explanatory variables denoted as $(x_1, x_2, \dots, x_p)$. Linear regression analysis is an effort to characterize the relationship between the response and the independent variables in order to achieve a linear mathematical model that best describes y values in terms of x values. More specifically, the following linear regression model is exploited to characterize our response variables as a linear combination of the explanatory variables:

$$Y = \mathrm{X}\beta + \varepsilon \qquad Eq.3-7$$

where Y is an $n \times 1$ vector of responses, X represents an $n \times p$ matrix of independent variables and $\boldsymbol{\varepsilon}$ denotes an $n \times 1$ vector of a normal distribution of random noise with a mean zero and variance σ. In this model,

$$\beta = (\beta_1, \beta_2, \dots, \beta_p)^T \qquad Eq.3-8$$

indicates a $p \times 1$ vector of coefficients which are allowed to be any real number in general. A solution for β's, denoted as $\hat{\beta}$, can be found by minimizing the sum of squared errors for the n observations according to:

$$\min_{\beta} \sum_{j=1}^{n} \left\| y_j - X^j \beta \right\|_2^2 \qquad Eq.3-9$$

where y_j is the jth response, X^j represents the jth row of matrix X, and $\left|\left| y_j - X^j \beta \right|\right|_2^2$, denotes the Euclidean norm of a vector. The procedure for finding a solution to the equation above is known widely as the method of Least Squares and, as long as X is a full rank matrix, it is shown to provide a unique solution, $\hat{\beta}$ according to:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \qquad Eq. 3-10$$

where the symbols (^{T}) and ($^{-1}$) denote the transpose and inverse of a matrix, respectively. That said, for certain situations the response must be represented as a linear combination of independent variables with all of the coefficients being nonnegative. That is, the objective is to solve the following system of equations simultaneously:

$$\min_\beta \sum_{j=1}^{n} \left\| y_j - X^j \beta \right\|_2^2 \qquad Eq. 3-11$$

$$\beta_i \geq 0, \qquad i = 1,2, \dots, p$$

In this system, all β's are restricted to nonnegative values by imposing p non-negativity constraints. This problem requires a more complicated approach compared with Least Squares to obtain the coefficient estimates and the solution cannot be obtained in a closed-form formula. To solve this problem, Lawson, *et al.,* [41] proposed the algorithm Non-Negative Least Squares (NNLS) which is an active set method. This algorithm can be applied in R [42] using the nnls function in the nnls package which provides the nonnegative coefficient estimates, $\widehat{\beta_1}, \widehat{\beta_2}, \dots, \widehat{\beta_p}$.

3.3.2. Peptide Ion Collision Cross Sections

Electrospraying the model peptide Acetyl-PAAAAKAAAAKAAAAKAAAAK produces a series of $[M+2H]^{2+}$, $[M+3H]^{3+}$, and $[M+4H]^{4+}$ ions as shown in Figure 2-4. $[M+4H]^{4+}$ ions are shown as a small feature in the figure at *m*/*z* of ~453 where three major conformer types are evident. Two conformer types exhibit similar

intensities having collision cross sections of 492 and 506 Å^2 while the third, less-abundant and most-elongated conformer has a cross section of 534 Å^2 (respective drift times are t_D =7, 7.2 and 7.6 ms). In Figure 2-4, the $[M+2H]^{2+}$ ions are depicted as a broad (unresolved) feature spanning a CCS range of 300 to 400 Å^2. As in the previous study (Chapter 2), the $[M+3H]^{3+}$ ions were chosen for HDX experiments. These ions can be categorized into three conformer types. The most abundant (and most compact) conformer type for these ions exhibits a CCS value of 417 Å^2 (t_D =7.9 ms). The second most prevalent ions comprise a, more diffuse (fairly resolved) conformer type (Ω = 438 Å^2, t_D =8.3 ms). The third feature for the $[M+3H]^{3+}$ ions, the most diffuse conformer type, appears as a broad, unresolved shoulder with a CCS of ~464 Å^2(t_D =8.7 ms). The two most abundant conformer types of these ions were selected for characterization by gas-phase HDX.

3.3.3. Peptide Ion Structure Studies Using IMS-HDX-MS/MS Coupled with MD Simulations

The MD approach used in the current study is based on the production of many random *in-silico* structure types and subsequent filtering of the structures using the experimental CCS values and the HDX behavior of the peptide ions. From the previous study (Chapter 2) [30], 4000 *in-silico* structures were produced using a simulated annealing approach and the ion structure dynamics were characterized with MD simulations. CCS values were calculated for the candidate structures and these conformer types were filtered using experimental CCS values. In all, 63 and 261 conformers exhibited MD trajectories with matching CCS values to the compact and more diffuse conformer types of the $[M+3H]^{3+}$ ions,

respectively. These candidate structures comprised an ensemble consisting of a number of significantly different structural types.

Previous experiments have shown that per-residue HDX values can provide information about the relative distances of heteroatom sites to charge sites and deuterium incorporation sites, the surface accessibility of residues, and the charge site configuration of the peptide ions [8, 20, 25, 26]. In this manuscript, HAS scoring coupled with a NEC model is implemented for each candidate structure to produce its theoretical per-residue uptake pattern. Subsequently NNLR analysis is utilized to assign a population for these conformer types. The predicted results for deuterium uptake pattern and ETD mass spectra are compared with experimental data.

3.3.4. Carbonyl-charge Site and Carbonyl-hydrogen Cutoff Distance

Considering that the first and second steps of gas-phase HDX involve hydrogen bonding, it is important to consider optimal atomic interaction distances for the HAS model (see the Experimental section). Because the length of a hydrogen bond in the gas-phase can range from ~2 to 4 Å [43, 44], no precise interaction distance can be elucidated for proposed reaction geometries. In the first step of the exchange process (original incorporation of a deuteron) a threshold distance of 4-8 Å can be estimated, while for the subsequent transfer to an exchange site, a distance range of 2-5 Å can be proposed to represent the decreased flexibility of the backbone groups involved (if the deuteron receptor is a lysine residue it can be extended to 8 Å). The theoretical uptake behavior of peptide ions can be compared with experimental results to optimize threshold

distances. Several different cutoff distances were applied to the model and distances <5 Å and/or <3 Å for carbonyl-charge and carbonyl-hydrogen interactions, respectively, lead to very low total uptake values, while the respective distances that are >7 Å and/or >4 Å yield very high deuterium uptake values. As an example, the experimental data show that residues A(14) and A(15) have relatively high amounts of incorporated deuterium suggesting threshold distances that are ≥6 Å and ≥3.5 Å for the first and second steps, respectively. The best match between theoretical and experimental uptake values resulted from using threshold values of 6.5 Å and 3.5 Å for the respective interactions and these values have been used for the comparisons discussed herein.

3.3.5. HDX reagent partial pressure and reagent reactivity

For the purposes of this study, it is desirable to set the experimental partial pressure of D_2O at a value which produces the maximum contrast between *in-silico* structures. Figure 3-2 shows the deuterium uptake behavior for two exchange sites each having significantly different hypothetical uptake sites (S_{R_j}= 0.1 and 0.05) as a function of $n_{effective}$ for the peptide ion. As may be expected, a NEC of 5×10^3 for 1000 ions, the ion population with the more accessible hydrogen (>S_{R_j}) incorporates two fold more deuteriums than the population with the less accessible hydrogen. As the NEC increases, the deuterium uptake difference between these two ion populations shrinks and, at 2×10^4 effective collisions, the uptake ratio is about 2:3 (low:high S_{R_j} values). Finally, at 1.2×10^5

effective collisions, both ion populations incorporate a similar number of deuteriums.

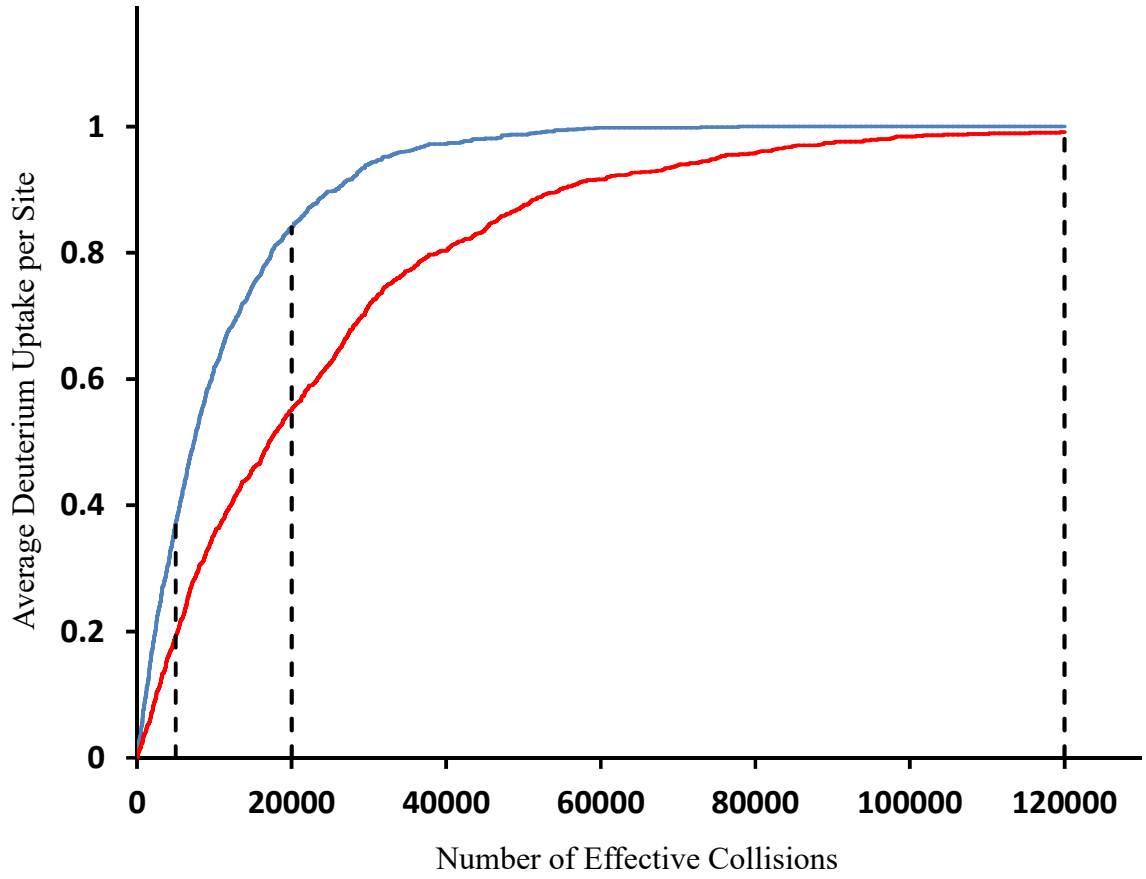

Figure 3-2. Depiction of the HDX behavior of a relatively slow- and fast-exchanging labile hydrogen. The NEC model is utilized to compare the deuterium uptake values of the two sites having relative hydrogen scores of S_{R_1}= 0.1 (blue trace) and S_{R_2}= 0.05 (red trace) within ion populations (1000) over a wide range of NEC values.

This situation is comparable with what may be expected for more highly reactive reagent (e.g. ND_3) where accessible and semi-accessible sites exhibits complete uptake, even at very low partial pressure of HDX reagent. That said,

although the work here does not preclude the use of ND_3, for the timescales of our measurements, a vanishingly small amount of ND_3 would be required (partial pressures not manageable with our apparatus) to obtain a similar level of discrimination. In these studies, D_2O is utilized as HDX reagent, because of its increased contrasting ability with regard to exchange sites. Overall, the deuterium uptake difference initially increases and reaches a maximum value and then decreases as the partial pressure of D_2O increases.

It is instructive to note that the HDX uptake pattern can only serve as an ion structure filtering criterion if the contrast between separate *in-silico* structures is significantly greater than the experimental error. Notably, the experimental error in deuterium uptake values tends to be constant at different partial pressures of D_2O; therefore, at lower partial pressures of D_2O, most protein and peptide ion systems exhibit larger relative errors.

To determine the optimal pressure of D_2O for HDX experiments which accounts for both factors above, the relative standard deviations of deuterium uptake at different NEC values for all candidate structures were calculated. Figure 3-3 shows the results of this calculation for one of the hydrogens on the K(6) residue. As the NEC increases, the contrast in deuterium uptake of *in-silico* structures increases then decreases. Here it is noted that for this hydrogen, a NEC value of 1×10^4 is the last point of the linear range in the kinetics plot (e.g., Figure 3-1). Based on a number of experiments [8, 20], a constant error of 0.1 (in deuterium uptake value) was assumed to calculate the relative experimental error values. Figure 3-3 shows the value of the confidence interval (99%) for the

experimental results as a function of NEC. The maximum difference between the two values occurs at a NEC value of ~20000 collisions corresponding to a partial pressure of ~0.03 torr for D_2O. This partial pressure value was chosen for the remaining studies.

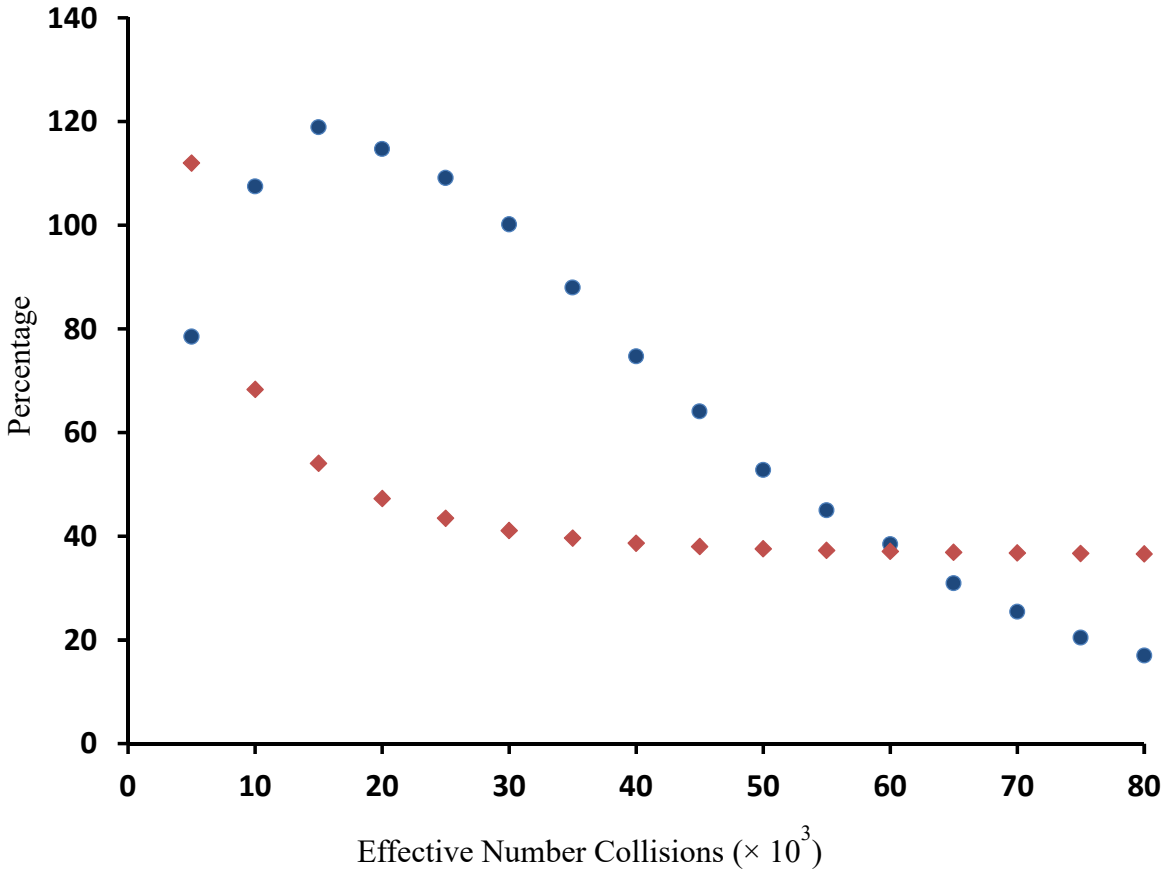

Figure 3-3. Comparison of the relative experimental error and the standard deviation in modeled deuterium uptake exhibited by *in-silico* structures as a function of the NEC (generated using the HAS algorithm – see Experimental section). Red diamonds represent the expected relative experimental error (confidence level >99%) for an exchange site on the K(6) side chain. Blue circles, show the coefficient of variation in the modeled deuterium uptake value of such an exchange site for 61 candidate structures of the compact $[M+3H]^{3+}$ peptide ion conformer.

3.3.6. Experimental and Predicted Deuterium Uptake Values

All the candidate structures with matching CCS were subjected to HAS scoring and subsequently considered by the NEC model, to estimate the deuterium uptake for each exchange site (see Experimental section). With these values for the 63 and 261 candidate structures for the respective more compact and elongated structures, non-negative linear regression was utilized to estimate the contribution of each structure to the overall per-residue HDX pattern. Briefly, the function (which is implemented in the R software suit [45]) determines a combination of theoretical results which best reconstructs the experimental result (least deviation).

Figure 3-4 compares the experimental and theoretical per-residue deuterium uptake values for the more compact conformer (Ω = 417 Å^2) of the $[M+3H]^{3+}$ ions. As demonstrated in Figure 3-4, predicted values for all of the residues are in good agreement with experimental values with the exception of A(20). Of note is the fact that there was no *c*- or *z*-ion for this residue to directly calculate deuterium uptake. Rather, the deuterium incorporation value was calculated using a combination of *c*- and *z*-ions. The total root mean square deviation (RMSD) of the predicted results from the experimental results is ~3.3%.

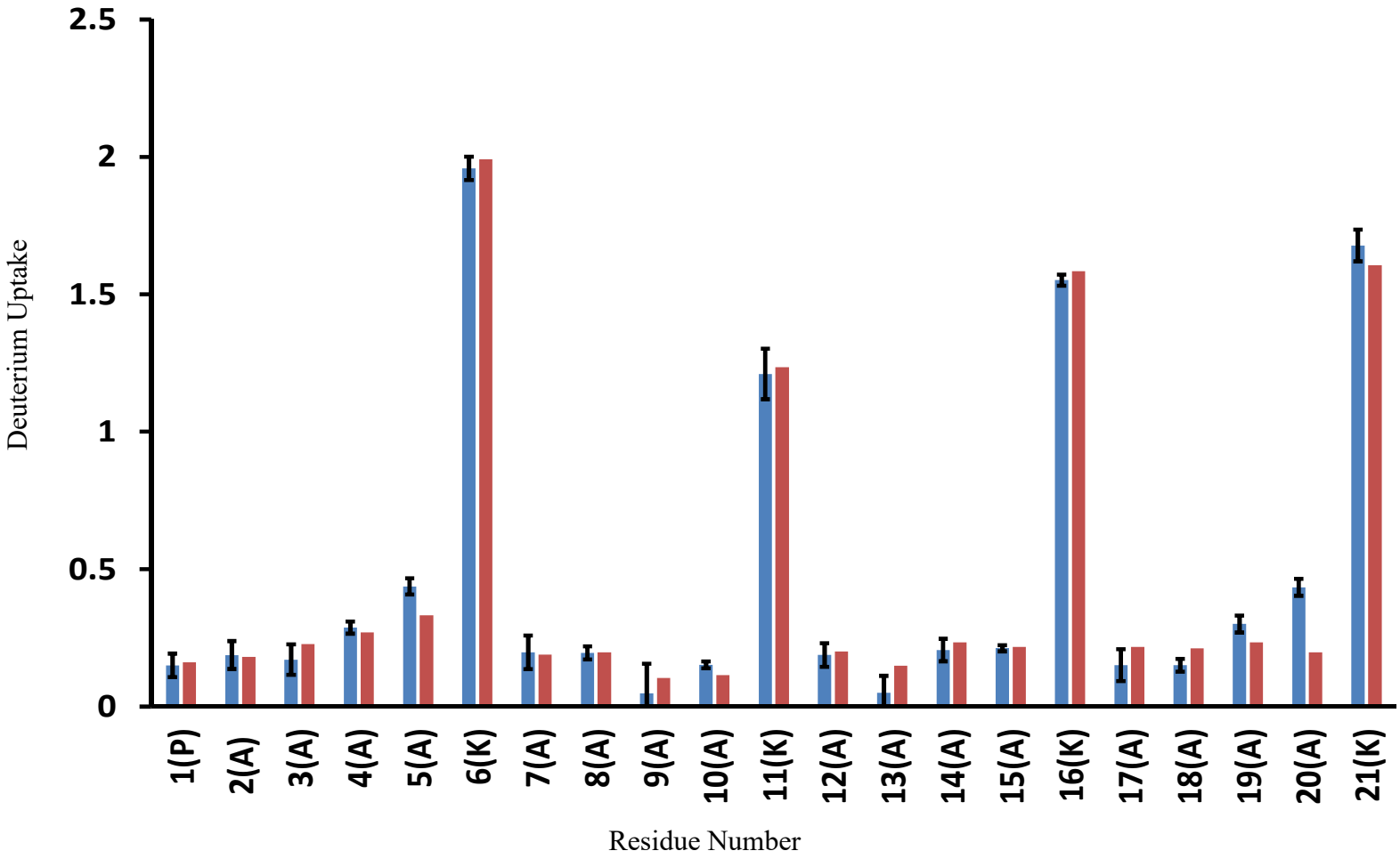

Figure 3-4. Experimental per-residue deuterium content versus modeled deuterium uptake values for the compact $[M+3H]^{3+}$ peptide ions. The blue bars (left) show the experimental deuterium uptake by each residue, while the red bars (right) are the hypothetical values estimated by HAS scoring and the NEC model.

The comparison of experimental and theoretical uptake values for the more diffuse conformer (Ω = 438 $Å^2$) of the $[M+3H]^{3+}$ ions is shown in Figure 3-5. The predicted deuterium uptake values exhibit ~3.1% RMSD compared to the experimental results. As with the more compact conformer, the uptake value for A(20) was calculated using a series of *c*- and *z*-ions. The error associated with this calculation cannot be estimated but may explain the difference between the experimental and predicted uptake values.

Although the HAS-NEC model can approximate the experimental results to within <4% error for both conformer types, the model is based on several assumptions and like any other model these are associated with some error. Additionally, the long experimental timescale (up to ~9 ms) in comparison to the production MD (5 ns) can lead to a lack of conformational space sampling [30] which can also be a source of error. On average, experimental per-residue deuterium uptake values exhibit an average error of <10%. Because, experimental deuterium uptake values and theoretical deuterium uptake values are utilized in the NNLR function to produce a coefficient matrix (containing population values), the final model result is expected to have an average error of <11%. It is instructive to note that the theoretical per-residue deuterium uptake values for *in-silico* structures at 20000 NEC produce, on average, >50% contrast between structures. Thus, the filtering with HDX results is statistically meaningful (confidence interval >99%).

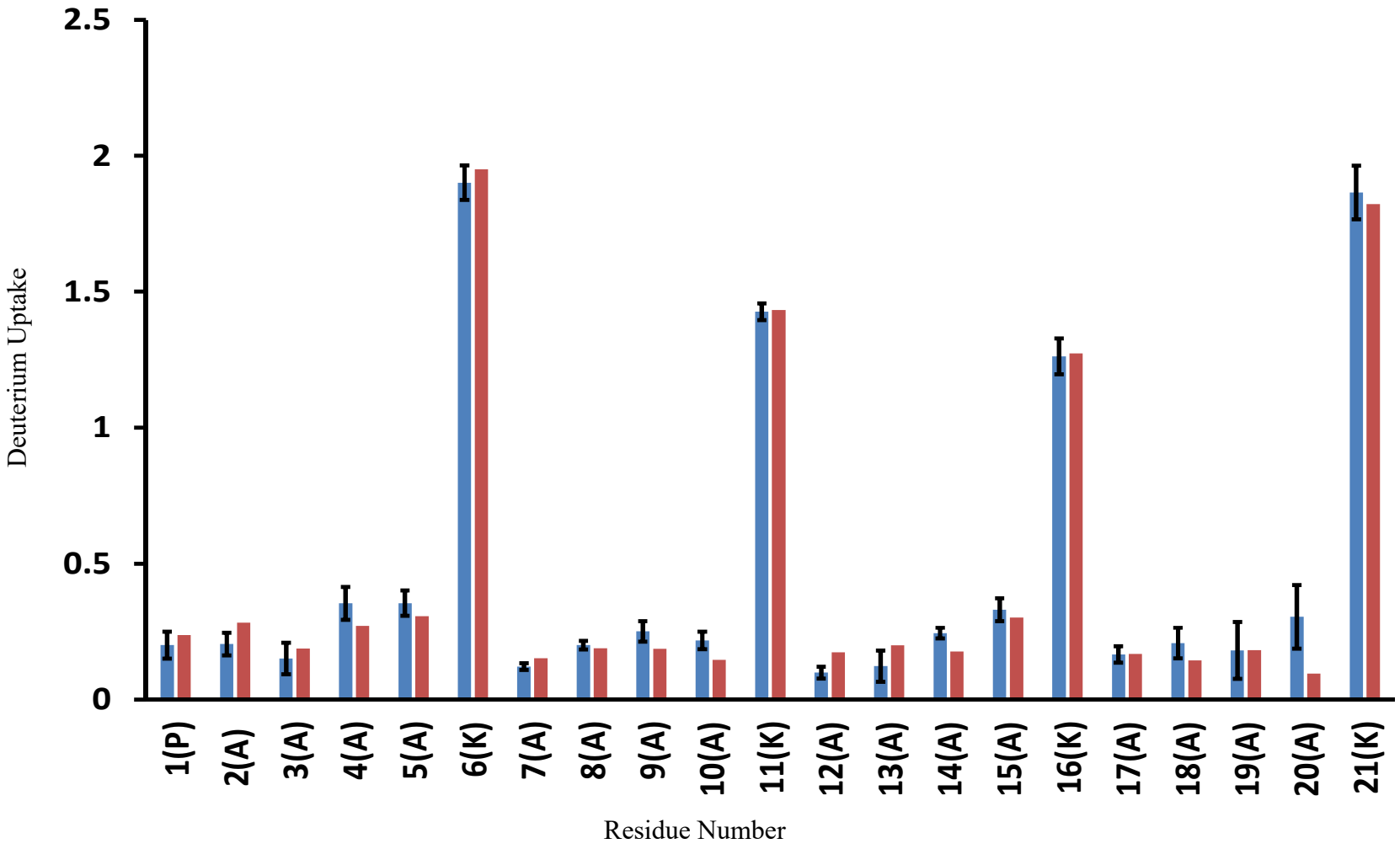

Figure 3-5. Comparison of experimental and theoretical per-residue deuterium content patterns for the more diffuse $[M+3H]^{3+}$ peptide ions. Blue (left) and red (right) bars represent the experimental and theoretical per-residue deuterium uptake values, respectively. Hypothetical values (red bars) are estimated by HAS scoring and the NEC model.

3.3.7. ETD Spectral Construction Using HAS Scoring

Although the comparisons of deuterium uptake values for individual amino acid residues can be used to assess the quality of the fit, this data lacks the isotopic distribution information afforded by the ETD mass spectrum. That is, the ETD mass spectra contain isotopic distribution patterns which can be also be compared with theoretical patterns to evaluate the quality of the fit. Previous studies have shown that isotopic distribution data can be examined to consider the possibility for the existence of multiple gas-phase conformer types of similar mobility that could contribute to the overall HDX uptake values [20]. As an example, a large number of conformer types with similar CCS values but distinct HDX reactivity could dramatically affect the observed isotopic distribution. Figure 3-6 provides an illustration of such an effect using three hypothetical isotopic distributions for a fragment ion. Under saturation conditions, and under complete exchange, the isotopic distribution would be narrow (similar to that for the ions in the absence of D_2O albeit shifted to higher m/z values). Under conditions in which ions of similar mobility and relatively similar HDX propensities would produce a broadened isotopic distribution as shown in Figure 3-6. In the case that ions of similar mobility exhibit very different HDX behavior, multiplet isotopic distributions may be observed (also depicted in Figure 3-6).

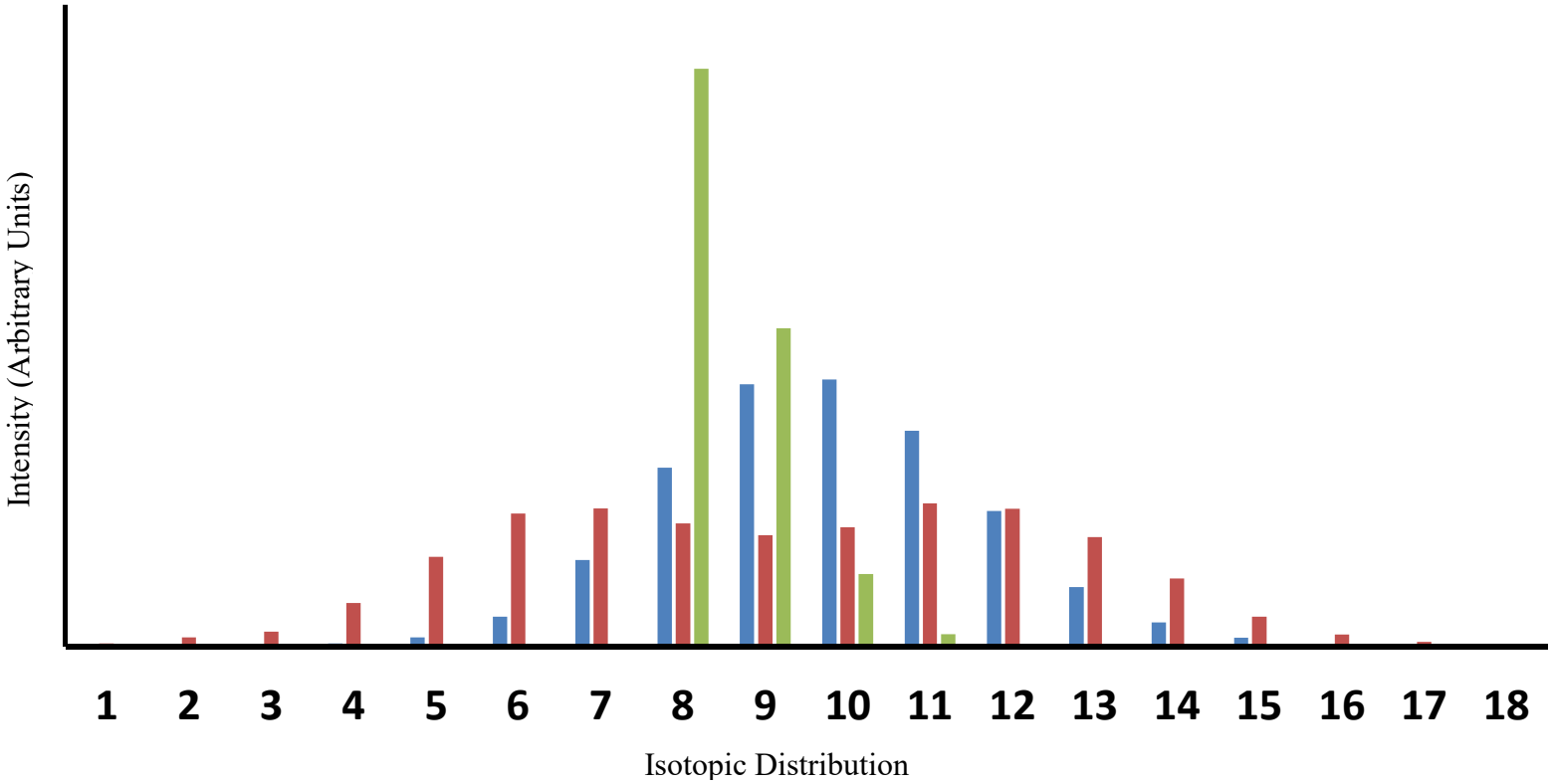

Figure 3-6. Examples of isotopic envelopes for scenarios in which a different number of structural types may be present. Green bars show the proposed isotopic distribution for a single conformer type under HDX saturation conditions. The presence of multiple conformer types (similar mobilites) can lead to broadening (blue bars) or splitting (red bars) of the isotopic envelope (see the description of these scenarios above). The isotopic distribution is intended to represent an uptake level such as that exhibited by the c_{19} ions.

Comparison of theoretically predicted and experimental ETD spectra can provide a degree of estimation for the accuracy of the methodology. Indeed it may be argued that the isotopic distribution be used rather than the per-residue deuterium uptake values for more accurate conformer type selection using the gas-phase HDX data. That said, such a procedure is associated with some significant limitations and obstacles to full implementation. For example, a precise prediction of the isotopic envelope would require per-residue HDX kinetics data, which is experimentally laborious and the modeling for this procedure is extremely time intensive [20]. Here, the HAS scoring coupled with a slight variation of the NEC model (see Experimental Section) presents an uptake value for each ion. Using a Monte Carlo approach, a two-dimensional array ($1000 \times j$ where j is the number of exchange sites on a given fragment ion) is populated with exchange events. In this manner, the distribution of deuterium on the exchange sites for 1000 ions can be estimated and subsequently the isotopic distribution for the best-fit composition of ion populations can be generated. Figure 3-7 compares the predicted and experimental isotopic distribution for c_{10} and c_{19} fragment ions generated from the compact and more diffuse $[M+3H]^{3+}$ precursor ions. Overall, the model generally captures the width of the isotopic distribution for both ions. Additionally the model predicts trends in isotopologue abundance very well; the overall RMSD ranges from 4.5% to 7.5%. It is important to note that we have previously demonstrated that single conformer types do not adequately model the width of the isotopic distribution for ETD fragment ions [20].

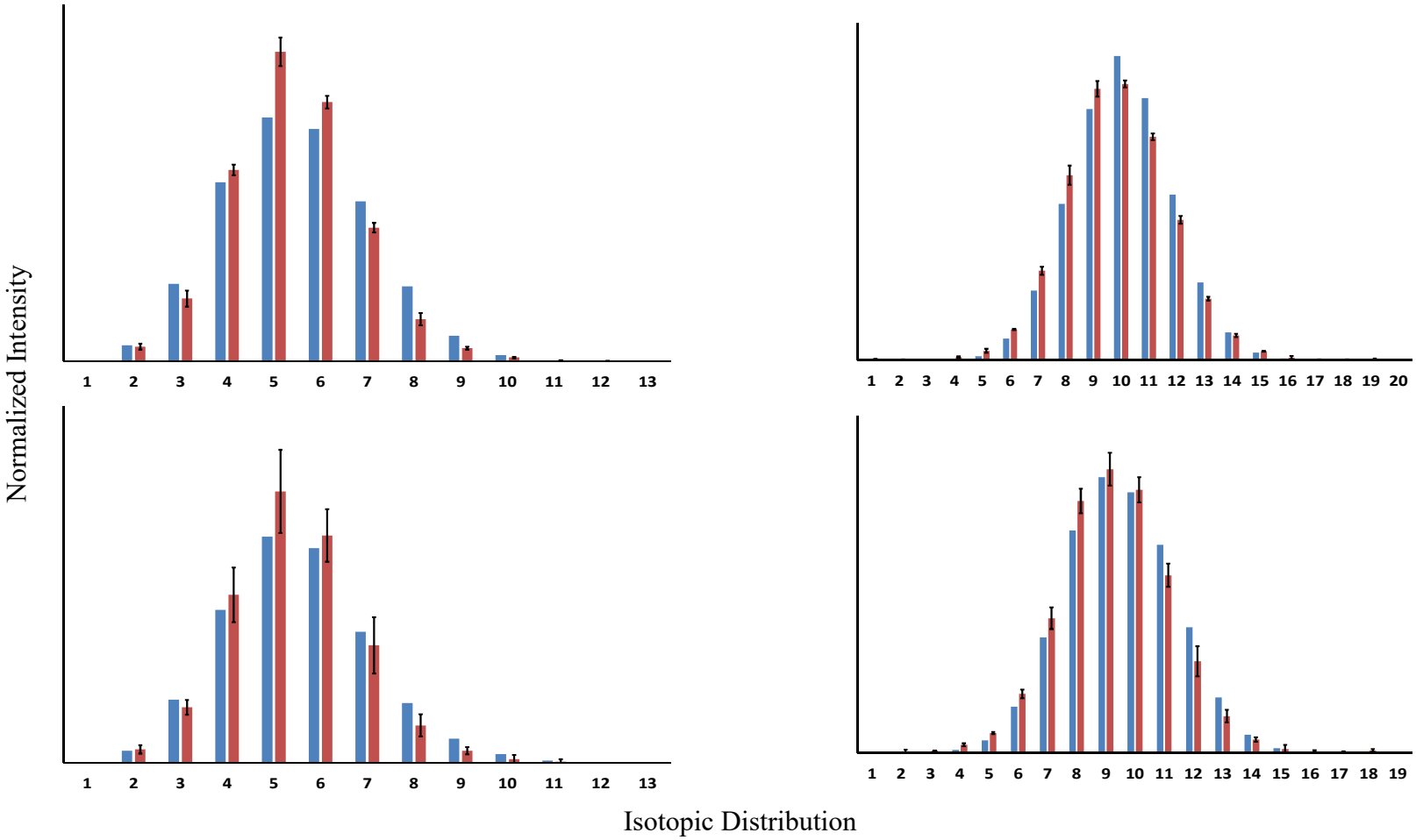

Figure 3-7. A depiction of the theoretical isotopic envelopes versus experimental results for two fragment ions originating from labeled peptide ions (gas-phase HDX). Blue bars (left) represent the predicted values while the red bars (right) show the experimental intensity value of each isotope. The top panels show results for the compact conformer and the bottom panels represent the more diffuse structural type. Panels on the left and right show results for c_{10} and c_{19} ions, respectively.

In general, they produce distributions that are significantly smaller in overall width and shape. The agreement in isotopic distributions shown in Figure 3-7 further confirms a degree of conformer heterogeneity for these peptide ions.

3.3.7.1. Structural Elucidation Using Isotopic Envelope Profiles

The deuterium content value is an average value and is not related to the isotopic distributions of different ions. That is, similar deuterium content profiles could arise from any number of isotopic distribution patterns as long as the average uptake values are the same. Additionally, separate studies have shown that the isotopic envelope contains structural information [5, 6, 20]. Figure 3-7 provides hypothetical scenarios which help clarify this. Consider that at very high NEC values all accessible hydrogens are exchanged by reaction with D_2O. An ion from a single conformer under such conditions (complete saturation) would exhibit the same isotopic envelope profile as the no HDX conditions. This scenario is represented by the green bars in Figure 3-7. The existence of multiple co-existing conformers could lead to different profiles. In one example, different conformers of the same mobility that may uptake similar amounts of deuterium (albeit slightly different levels) result in differences in mass leading to a broadening of the isotopic envelop. Depending on the conformer populations and relative uptake values of these conformers, the isotopic envelope shape would vary (the blue bars in Figure 3-7 provide one example). In contrast, multiple conformers of similar size but containing primarily two significant differences in accessible hydrogens may exhibit overall uptake values which can split the isotopic envelope as shown by the red bars in Figure 3-7.

3.3.8. Gas-phase Ion Structures

Using HDX as a second criterion for ion structure filtering in combination with NNLR analysis the number of candidate structures were decreased to 6 and 7 species (nominal structures) for the compact and more diffuse conformer types respectively. Figures 3-8 and 3-9 show the nominal structures estimated to be the most abundant for the compact and more diffuse conformers of the $[M+3H]^{3+}$ peptide ions, respectively. The lower abundance nominal structures are shown in Figures 3-10 and 3-11 for the compact and more diffuse conformer types, respectively. For comparison, the annealed structures (optimized structure resulting from the SA procedure) are shown along with the gas-phase structures (nearest structures to the centroid of the gas-phase MD trajectory).

The proposed, most abundant gas-phase structures for the diffuse conformer type can be compared according to their degree of helicity. ~57% of the nominal structures (#1 with structural population (P_s)=39%, #3 P_s=16%, #5 P_s =2%) exhibit a high degree of helicity (>40%), while the remaining nominal structures (~43% of all structures) show relatively random structures. Similar to prior predictions [26], the majority of the ion population (including structures: #1, #3, #5, #6) contains the charge site arrangement of K(6), K(16), and K(21). Results for the more diffuse conformer type exhibit a greater variety in nominal structures (Figures 3-9 and 3-11) but, in aggregate, less helicity is observed than the more compact ions. Structures #2, #3, #6, #5 have a charge site configuration that is comprised of the K(6), K(11), and K(21) residue

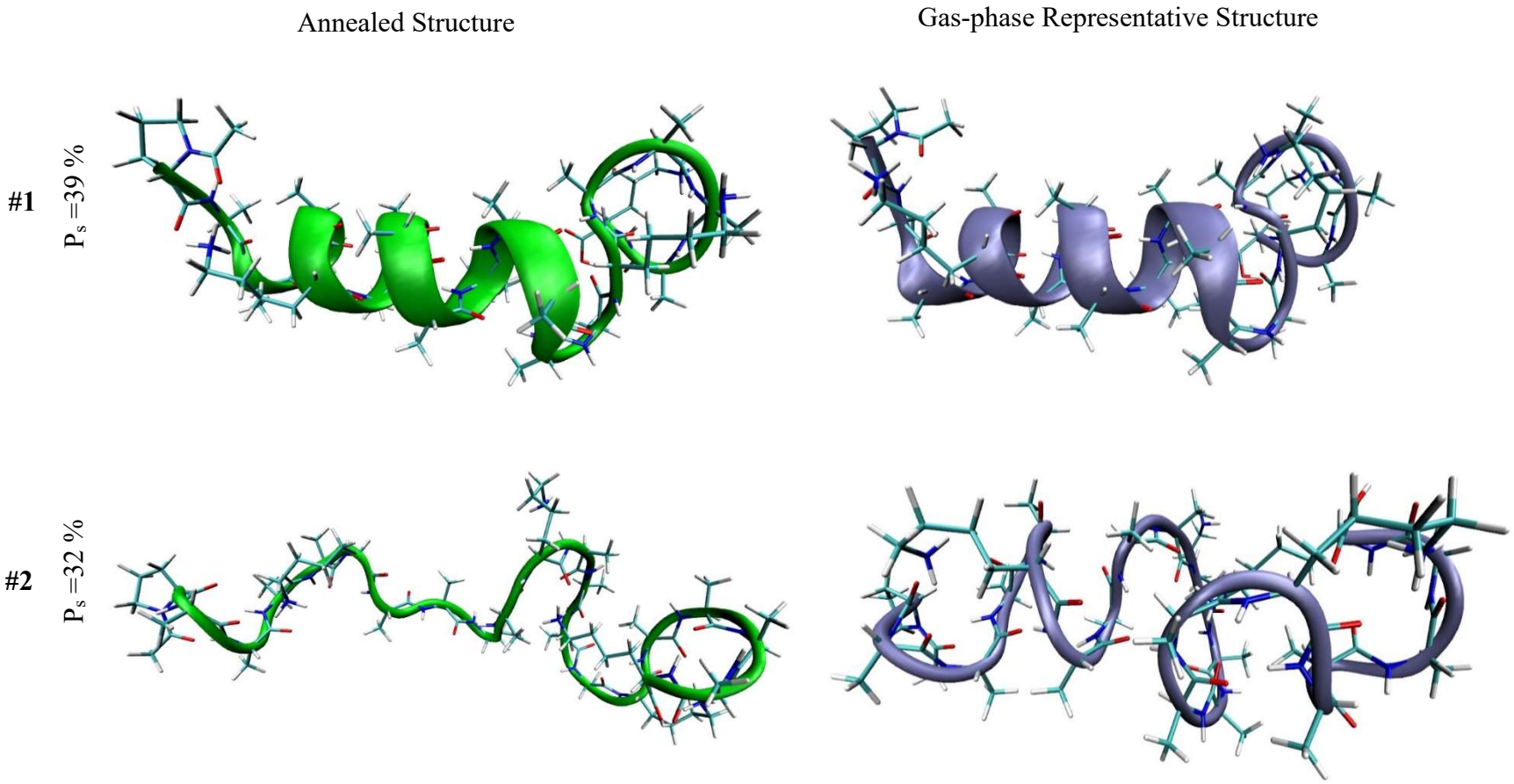

Figure 3-8. The most prevalent structures (nominal structures) matching the criteria for compact $[M+3H]^{3+}$ peptide ions. The panels labeled as "Annealed structure" show the conformers resulting from SA while the "Gas-phase structure" is obtained by finding the nearest structure to the centroid of the production MD. Proposed population values (P_s) and a structure number (#) are also assigned to each structure and labeled in the figure.

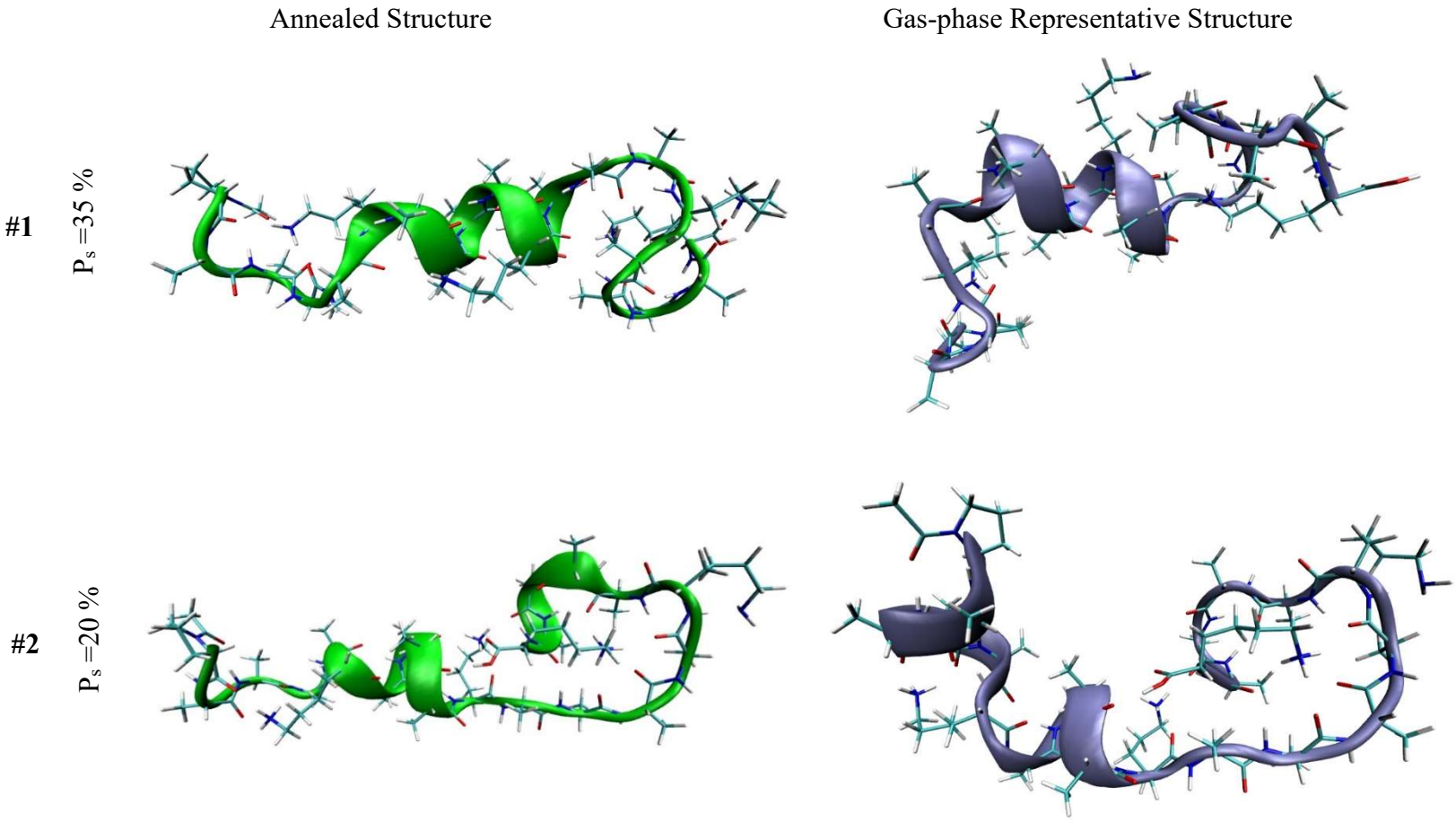

Figure 3-9. The most abundant nominal structures representing the more diffuse $[M+3H]^{3+}$ peptide ion conformer. Annealed structures are shown in the left panels and the closest structures to the centroids of productions MD are depicted in the right panels. Relative population (P_s) and structure numbers (#) are also provided.

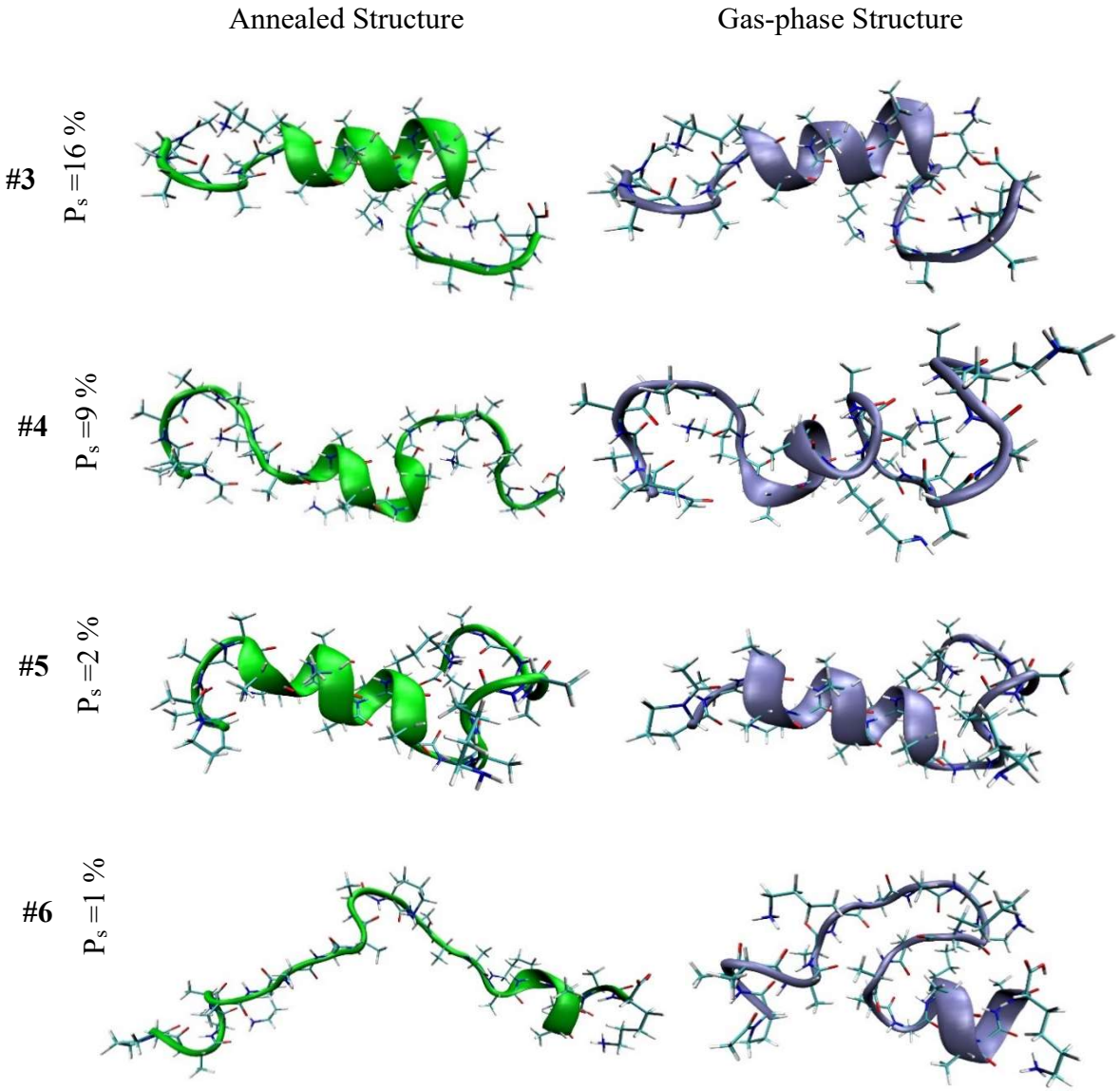

Figure 3-10. Less abundant nominal structures for the compact conformer type of the $[M+3H]^{3+}$ peptide ions. The annealed structure panel depicts the simulated annealing structure which leads to the nominal structure having CCS and HDX matching trajectories. The nearest structure to the trajectory centroid is used to visually represent gas-phase structure. The trajectory is used to represent conformational fluctuations. The assigned population is denoted by the "P_s" parameter.

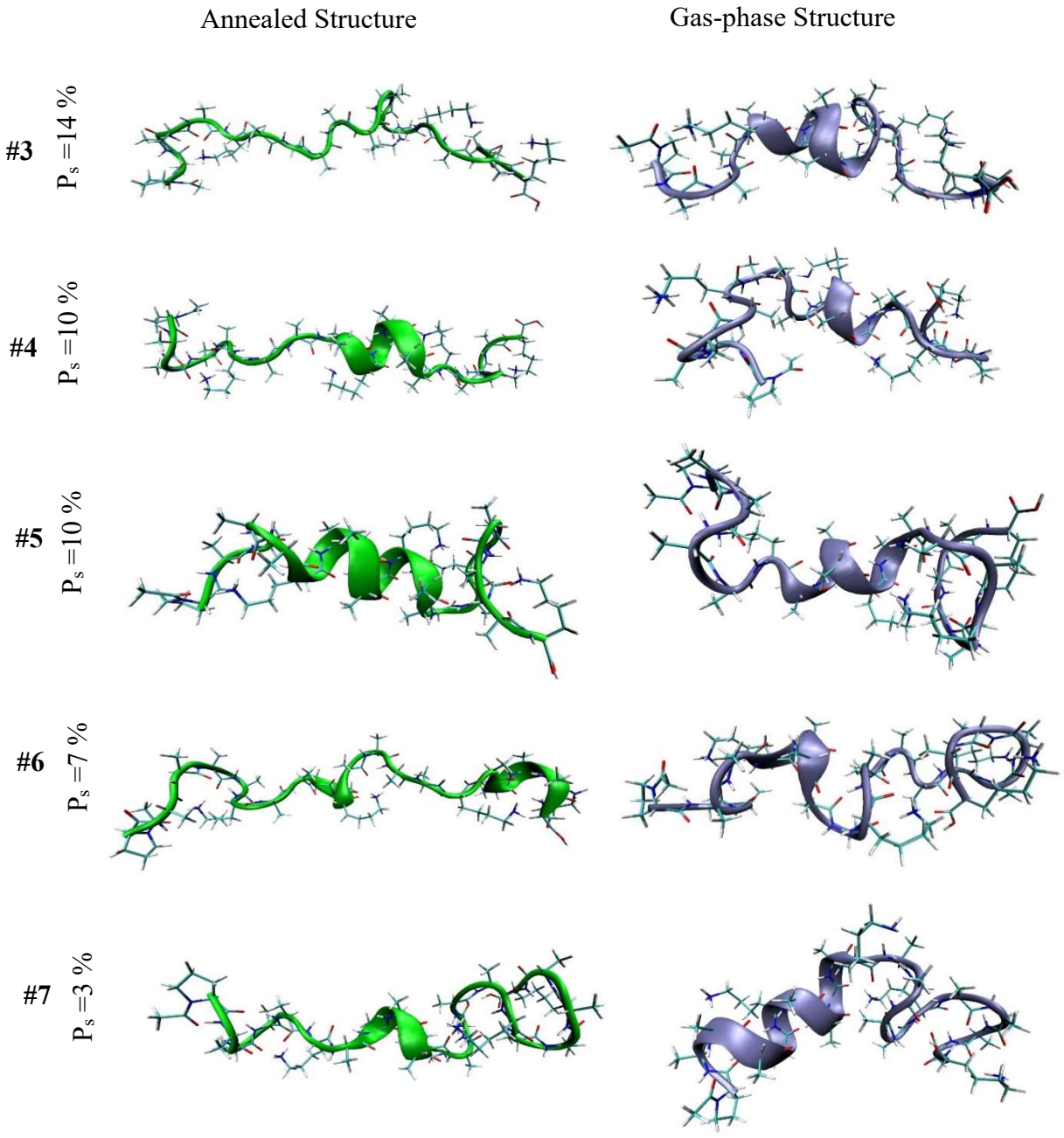

Figure 3-11. Depiction of less abundant nominal structures of the more diffuse triply-protonated ions. Results from simulated annealing and the nearest structure to the centroid of MD production are depicted in each panel for each nominal structure. The assigned populations and structure numbers are shown at the left for each panel.

For the compact ions, structures #1, #3, #5 display a high degree of similarity. Additionally, structures #1 and 5# have a similar appearance. The degree of structural similarity here may suggest a sampling of the same area of conformation space. As discussed in the first installment of this work [30], the production MD simulation time is not comparable to the drift time of the ions. Simulated annealing can be exploited to overcome barriers on the potential energy surface and therefore address the short time scale of production MD. These structures point to the same area of conformation space and thus may not be independent of one another at 300K. Therefore, the relatively short timescale of the MD production runs (in comparison to drift time) and the lack of conformation space sampling causes these apparent differences. To clarify, one can assume two trajectories point to the same area of conformation space available for an ion for an extended period of time. Additionally it is understood that NNLR is a linear combination of contributions as illustrated in Equation 3-12:

$$(V_1 + V_2)a = (V_1)a + (V_2)a \qquad Eq.3-12$$

In equation 3-12, ***a*** is coefficient matrix (containing the ion population), and V_1 and V_2 are the HDX results for trajectory 1 and trajectory 2, respectively. Therefore, although V_1 and V_2 are presented to the function as separate entities, they may point to the same trajectory at different time points. These structures may not necessary be different conformer types; rather, they sample different portions of the same trajectory.

3.3.9. Ion Segment HDX Behavior Analysis for $[M+4H]^{4+}$ Ions

The study of quadruply-charged ions provides the unique opportunity to assess the HAS-NEC model for another ion structure type (i.e., diffuse and less flexible ion conformers). $[M+4H]^{4+}$ peptide ions exhibit three partially-resolved conformer types with CCS values of Ω = 492, 506 and 534 $Å^2$ as shown in Figure 3-12. These ions produce very low intensities such that the ETD spectra produced upon drift selection for each conformer type is not of sufficient quality to investigate the HDX behavior of these ion conformers individually. Therefore a wide drift time selection was applied encompassing all three conformer types in order to study the behavior of diffuse and less flexible ion conformers (see below).

Interestingly, for these more elongated $[M+4H]^{4+}$ peptide ions, lower uptake values (~4.5 deuteriums) are observed compared with the $[M+3H]^{3+}$ peptide ions (~10 deuteriums). This difference in uptake has been reported previously for peptides ions [8, 20]. These earlier studies show that lower levels of exchange may be attributed to differences in HDX kinetics and site accessibility. That said, this lower level of deuterium incorporation limits the application of per-residue deuterium uptake studies for these ions. Considering a hypothetical 21-residue peptide with an uptake value of 4.5 deuteriums, on average, the deuterium contribution from each residue would be ~0.2 deuteriums. The experimental error in obtaining per-residue deuterium uptake information is typically ~0.1 deuteriums [8, 20].

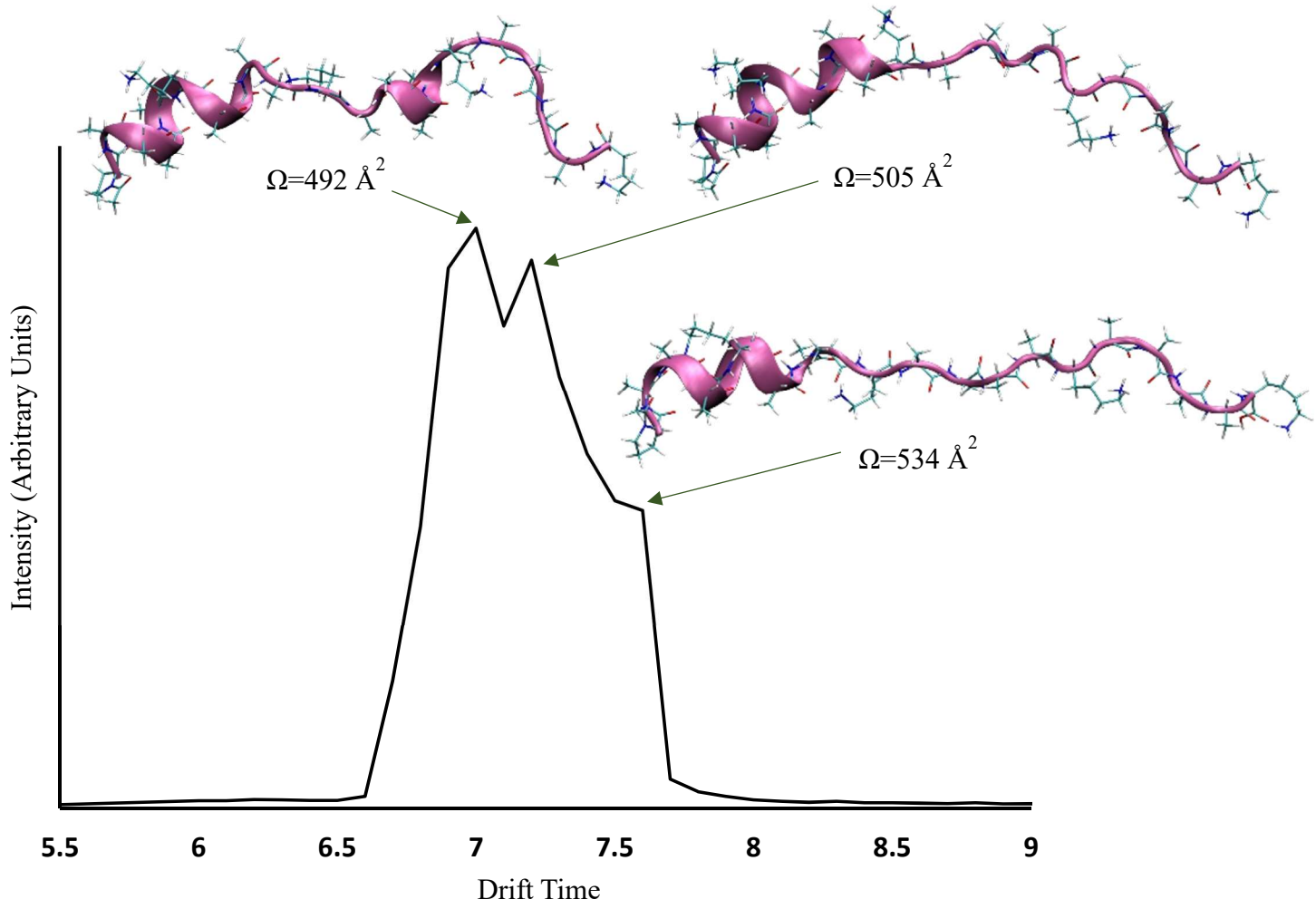

Figure 3-12. Drift time distribution of the $[M+4H]^{4+}$ peptide ions. Quadruply-charged ions of the model peptide exhibit three different conformer types of increasing elongation as demonstrated by the peak apexes and the low-mobility shoulder. *In-silico* structures with matching CCS values and HDX behavior are shown in pink.

It may thus be argued that for species exhibiting limited exchange and having considerable numbers of amino acid residues, the per-residue deuterium uptake value would not be meaningful. This would limit the applicability of the approach. Here, another strategy which divides the peptides into several segments has been utilized to study the HDX behavior of such ions. For the $[M+4H]^{4+}$ peptide ions, four different segments, namely Acetyl-PAAAAK, AAAAK, AAAAK and AAAAK, are used. HAS-NEC is then utilized to propose theoretical uptake values for each segment. Here it is noted that such an approach can not only address situations in which low levels of per-residue deuterium incorporation are encountered but also those in which incomplete fragmentation may be observed. Notably, both problems would be relevant in larger protein ion analyses and so the study of the quadruply-charged ions presents a proof-of-principle structure characterization.

The candidate structures of $[M+4H]^{4+}$ ions were subjected to the HAS-NEC model. Figure 3-13 compares the experimental and predicted uptake values for nominal structures generated from MD simulations [30]. Gas-phase structures representing the different conformers types of $[M+4H]^{4+}$ ions filtered by the HAS-NEC model are shown in Figure 3-12. It is worthwhile to mention that for the quadruply-charged ions, the elongation of the ion due to excessive charge density would not allow preservation of solution-like structures; therefore, these structures are discussed from a reference point of gas-phase structure stabilization. All three conformer types display a degree of helicity within the first peptide ion segment (Acetyl-PAAAAK) where the charge located at K(6) may stabilize the helix [46]. Significant coulomb repulsion and the presence of a positive charge at N-terminal

locations of subsequent segments do not allow these portions to adopt specific secondary structure. Overall, they are relatively rigid structures of varying degree of elongation.

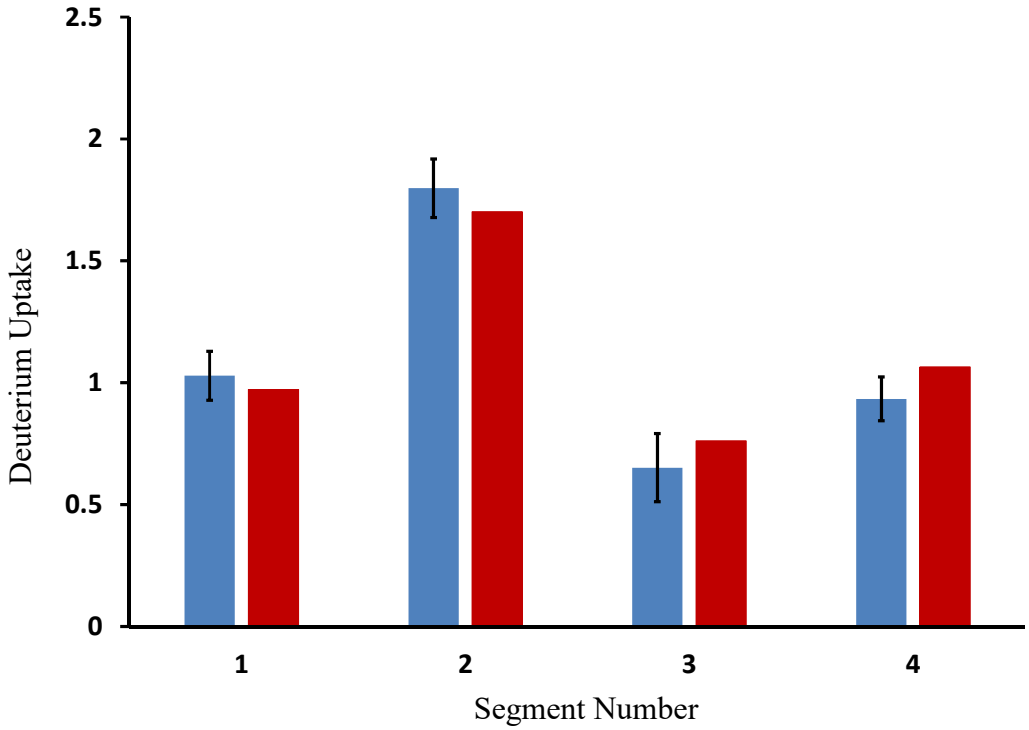

Figure 3-13. Comparison of experimental deuterium uptake and theoretical values for segments of the quadruply-charged peptide ions. The blue bars show the experimental deuterium uptake value while the red bars represent the predicted deuterium uptake values.

As with the $[M+3H]^{3+}$ ions, application of the HAS-NEC model can also capture the exchange behavior of the quadruply-charged ions. Having established the types of conformers contributing to the HDX behavior, the total uptake

deuterium uptake pattern of conformer types within the drift selection can be expressed as:

$$V = \alpha_1 V_1 + \alpha_2 V_2 + \alpha_3 V_3 \qquad Eq. 3-13$$

In which V represents a vector of the experimental deuterium uptake for the 4 segments. V_i represents the hypothetical deuterium uptake for the i^{th} conformer (Figure 3-12) from HAS-NEC. The coefficients α_i are calculated using peak intensity for each conformer type comprising the mobility selection. Figure 3-13 shows the comparison of the experimental and theoretical values from this treatment. The high degree of agreement between theory and experiment is promising for the segment-wise application of the HAS-NEC model to larger protein systems in which incomplete fragmentation patterns are encountered.

3.3.10. Effective Temperature of Ions in the Presence of D_2O

The effective temperature of the ions can affect the HDX kinetics [3]. The introduction of D_2O to the drift tube can affect the ion velocity as well as neutral gas average mass, therefore it is worthwhile to consider a change in the effective temperature of the ions. Under low-field conditions this value can be estimated by Equation 2-2. Figure 3-14 shows the effective temperature of the ions at different partial pressures of D_2O (added upon the 2.600 Torr of He). Although, the effective temperature decreases as the partial pressure of D_2O increases, the miniscule change in ion temperature would not affect the HDX behavior [3]. Consequently it can be argued that over an extended D_2O partial pressure range the change in the

ratio of the number of collisions to number of effective collisions would be negligible.

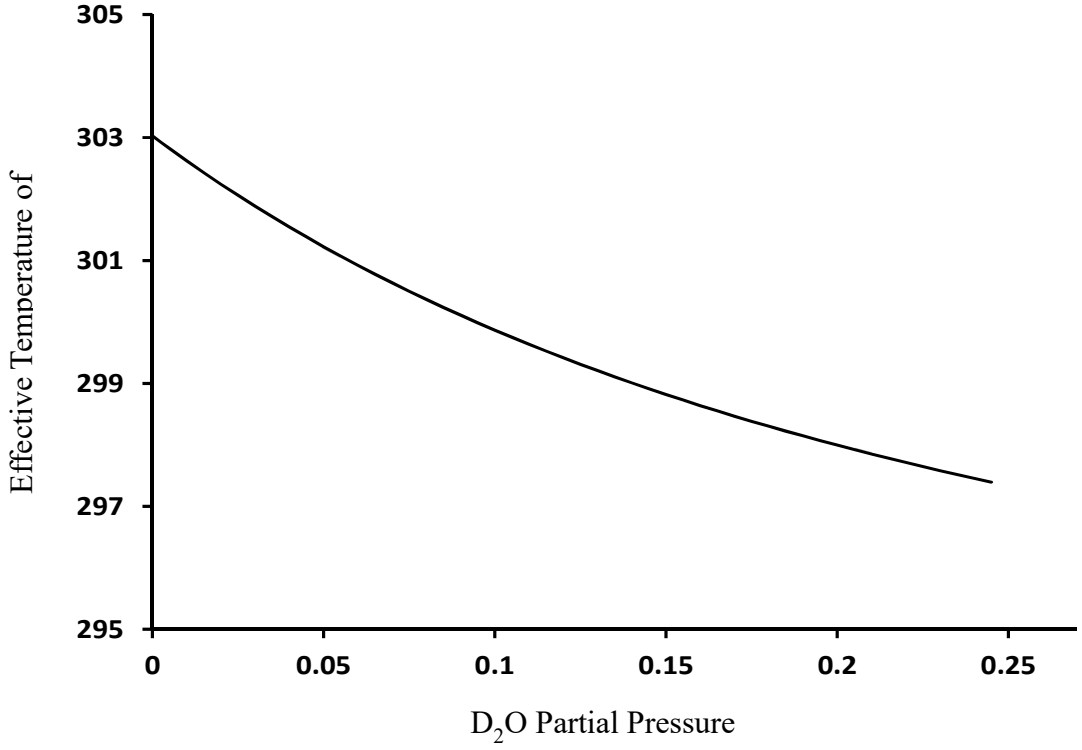

Figure 3-14. Calculated effective temperature of $[M3H]^{3+}$ ions at different partial pressures of D_2O.

3.3.11. ETD Spectra of Compact Conformer Types of $[M+3H]^{3+}$ Ions

Figure 3-15 shows the ETD spectrum collected after drift selection of compact, triply-protonated ions of the model peptide. The insets show expanded *m*/*z* regions encompassing representative *z* ions to show the isotopic distribution, the ion signal quality, and the lack of rearrangements in the production of fragment ions during the ETD process.

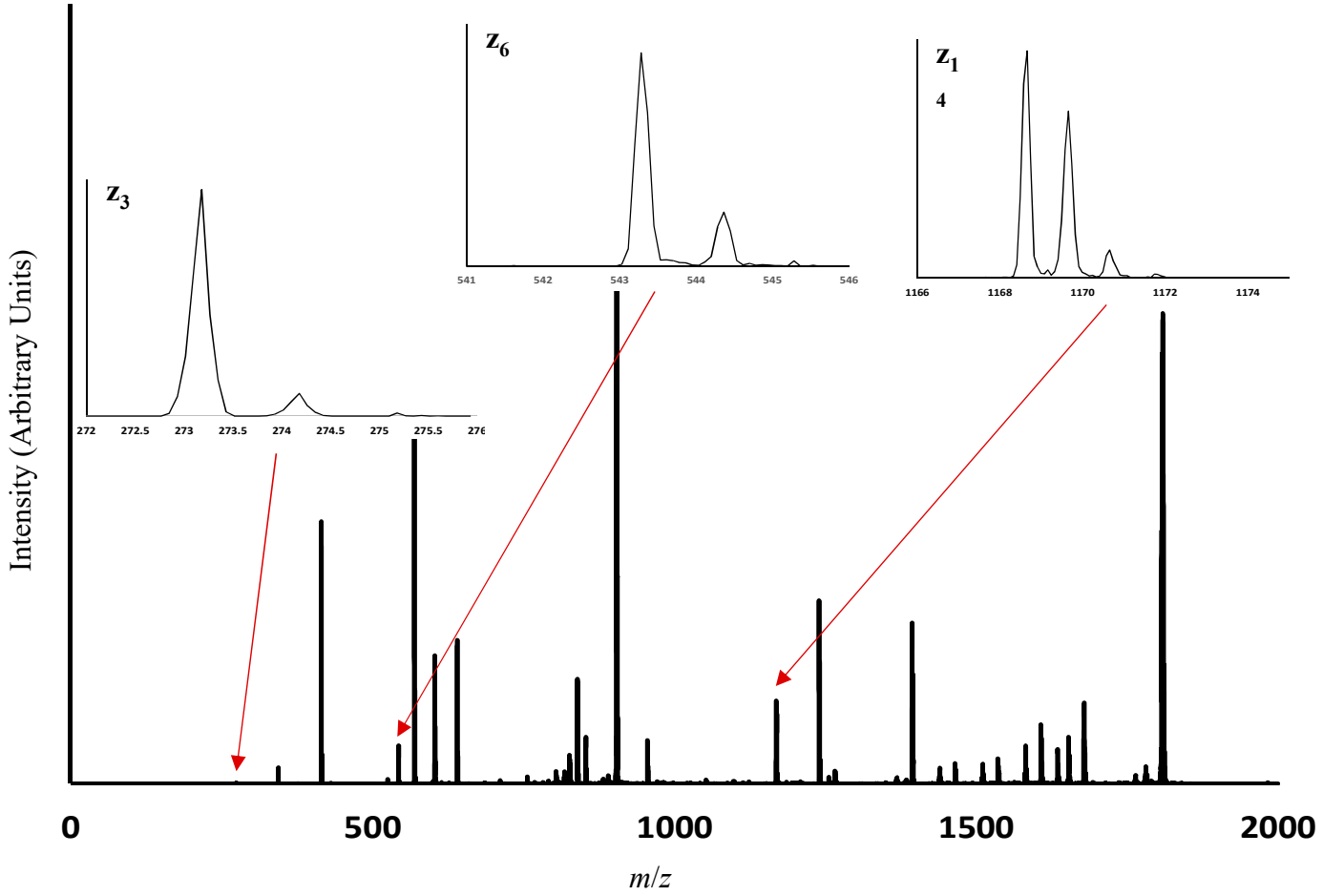

Figure 3-15. ETD spectrum of compact conformers of $[M+3H]^{3+}$ peptide ions. The insets show expanded m/z regions for several z ions to depict the isotopic distributions produced upon fragmentation.

3.3.12. HDX behavior comparison for the compact and more diffuse conformers of $[M+3H]^{3+}$ peptide ions

Although the triply-charged ions exhibit relatively similar collision cross sections, there are notable differences in deuterium content for specific amino acid residues after the precursor ions undergo gas-phase HDX. Figure 3-16 shows the per-residue deuterium content for the different ion conformer types of the triply-charged ions. Significant differences in deuterium content are observed for the A5, A7, A9, A10, A12, A15, K11, K16, and K21 residues.

3.4. Conclusion

In the first installment of this work (Chapter 2), the simulated annealing approach was utilized to sample conformation space for the model peptide acetyl-PAAAAKAAAAKAAAAKAAAAK. Production MD at constant temperature was applied on the sampled structures to provide a more realistic conception of the *in-silico* structures at 300K. Additionally a novel procedure was proposed to calculated CCS values for the *in-silico* trajectories. More than 300 candidate structures with matching CCS values to the experimentally measured CCS values for more compact and diffuse conformer types of $[M+3H]^{3+}$ peptide ions were extracted from a pool of 4000 structures. Here, the application of gas-phase HDX measurements is presented for further structure elucidation. A new algorithm based on the mechanism of gas-phase HDX with D_2O reagent [1, 2] is proposed to yield a theoretical relative site-specific reactivity for *in-silico* structures (HAS scoring). A Monte Carlo approach is applied to model the HDX behavior of an exchange site as a function of the number of effective collisions (NEC).

Subsequently HAS-NEC is applied for all candidate structures (matching CCS) to produce a hypothetical deuterium uptake pattern for each structure. Non-negative linear regression is then exploited to solve a best fit reconstruction of the experimental results. In this manner, not only is the number of candidate structures reduced by a factor of 20, but also a relative population is assigned to the nominal structures. To assay the accuracy of the model, the isotopic envelopes of 4 fragment (*c*-) ions were generated and compared to experimental distributions.

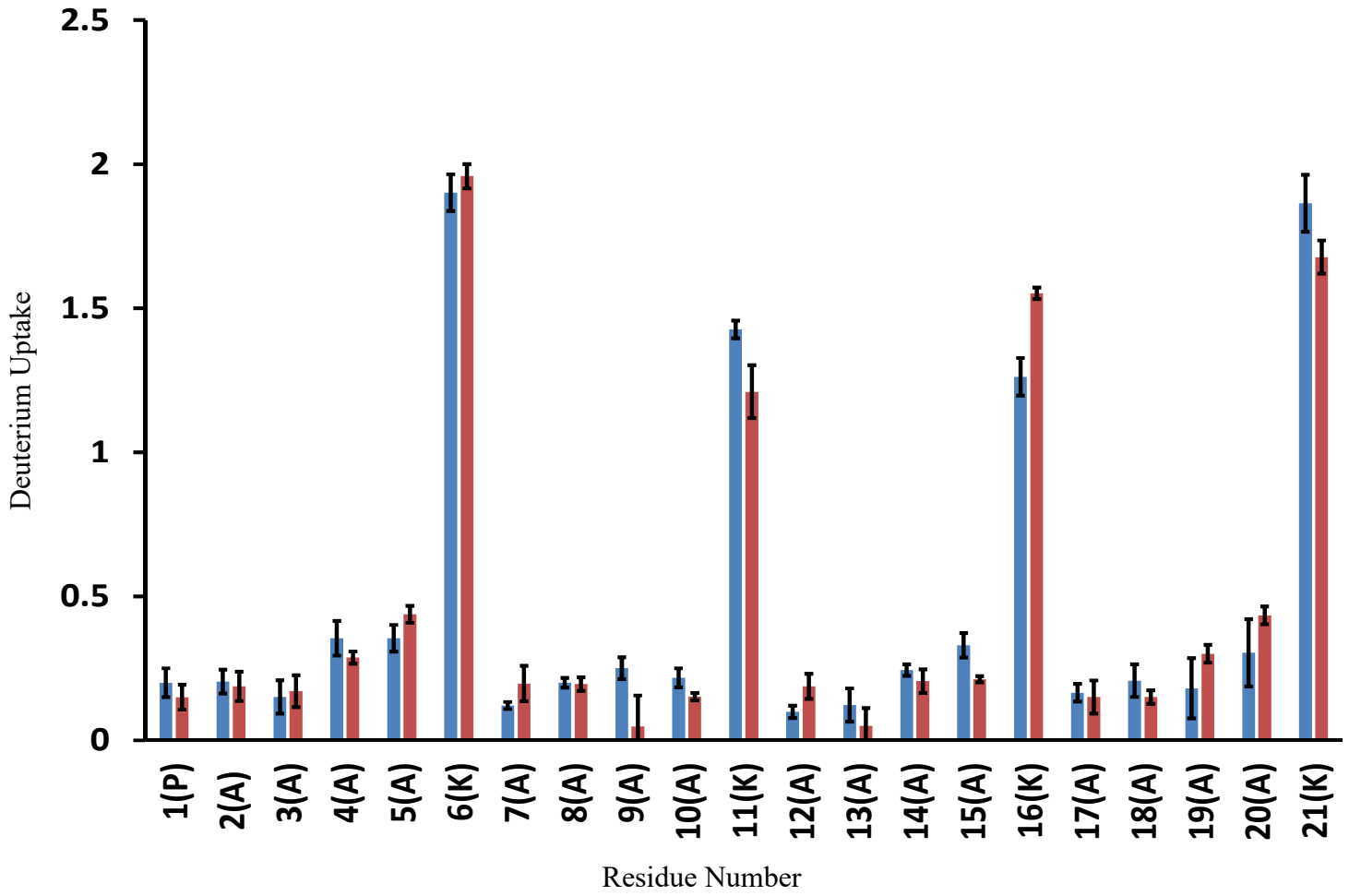

Figure 3-16. Per-residue deuterium uptake comparison of compact and more diffuse $[M+3H]^{3+}$ peptide ions. Blue (left) and red (right) bars show the per-residue deuterium uptake pattern for more diffuse and more compact conformers of triply-protonated ions of the model peptide, respectively.

3.5. References

1. Campbell, S., et al., *Deuterium exchange reactions as a probe of biomolecule structure. Fundamental studies of cas phase H/D exchange reactions of protonated glycine oligomers with D2O, CD3OD, CD3CO2D, and ND3.* Journal of the American Chemical Society, 1995. **117**(51): p. 12840-12854.
2. Wyttenbach, T. and M.T. Bowers, *Gas phase conformations of biological molecules: The hydrogen/deuterium exchange mechanism.* Journal of the American Society for Mass Spectrometry, 1999. **10**(1): p. 9-14.
3. Valentine, S.J. and D.E. Clemmer, *Temperature-dependent H/D exchange of compact and elongated cytochrome c ions in the gas phase.* Journal of the American Society for Mass Spectrometry, 2002. **13**(5): p. 506-517.
4. Donohoe, G.C., M. Khakinejad, and S.J. Valentine, *Ion Mobility Spectrometry-Hydrogen Deuterium Exchange Mass Spectrometry of Anions: Part 1. Peptides to Proteins.* Journal of the American Society for Mass Spectrometry, 2015. **26**(4): p. 564-576.
5. Suckau, D., et al., *COEXISTING STABLE CONFORMATIONS OF GASEOUS PROTEIN IONS.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(3): p. 790-793.
6. Wood, T.D., et al., *GAS-PHASE FOLDING AND UNFOLDING OF CYTOCHROME-C CATIONS.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(7): p. 2451-2454.
7. Valentine, S.J. and D.E. Clemmer, *H/D exchange levels of shape-resolved cytochrome c conformers in the gas phase.* Journal of the American Chemical Society, 1997. **119**(15): p. 3558-3566.
8. Khakinejad, M., et al., *Combining Ion Mobility Spectrometry with Hydrogen-Deuterium Exchange and Top-Down MS for Peptide Ion Structure Analysis.* Journal of the American Society for Mass Spectrometry, 2014. **25**(12): p. 2103-2115.
9. Bohrer, B.C., N. Atlasevich, and D.E. Clemmer, *Transitions between Elongated Conformations of Ubiquitin M+11H (11+) Enhance Hydrogen/Deuterium Exchange.* Journal of Physical Chemistry B, 2011. **115**(15): p. 4509-4515.
10. Fattahi, A., B. Zekavat, and T. Solouki, *H/D Exchange Kinetics: Experimental Evidence for Formation of Different b Fragment Ion Conformers/Isomers During the Gas-Phase Peptide Sequencing.* Journal of the American Society for Mass Spectrometry, 2010. **21**(3): p. 358-369.
11. Rand, K.D., et al., *Site-Specific Analysis of Gas-Phase Hydrogen/Deuterium Exchange of Peptides and Proteins by Electron Transfer Dissociation.* Analytical Chemistry, 2012. **84**(4): p. 1931-1940.
12. Reid, G.E., R.J. Simpson, and R.A.J. O'Hair, *Probing the fragmentation reactions of protonated glycine oligomers via multistage mass spectrometry and gas phase ion molecule hydrogen/deuterium exchange.* International Journal of Mass Spectrometry, 1999. **191**: p. 209-230.
13. Felix, T., M. Reyzer, and J. Brodbelt, *Hydrogen/deuterium exchange of nucleoside analogs in a quadrupole ion trap mass spectrometer.* International Journal of Mass Spectrometry, 1999. **191**: p. 161-170.
14. Gucinski, A.C., et al., *Separation and Identification of Structural Isomers by Quadrupole Collision-Induced Dissociation-Hydrogen/Deuterium Exchange-Infrared Multiphoton*

Dissociation (QCID-HDX-IRMPD). Journal of the American Society for Mass Spectrometry, 2010. **21**(8): p. 1329-1338.
15. Wyttenbach, T. and M.T. Bowers, *Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-Free Ion Mobility-Mass Spectrometry Environment.* Journal of Physical Chemistry B, 2011. **115**(42): p. 12266-12275.
16. Schnier, P.D., et al., *Blackbody infrared radiative dissociation of bradykinin and its analogues: Energetics, dynamics, and evidence for salt-bridge structures in the gas phase.* Journal of the American Chemical Society, 1996. **118**(30): p. 7178-7189.
17. Zubarev, R.A., N.L. Kelleher, and F.W. McLafferty, *Electron capture dissociation of multiply charged protein cations. A nonergodic process.* Journal of the American Chemical Society, 1998. **120**(13): p. 3265-3266.
18. Syka, J.E.P., et al., *Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(26): p. 9528-9533.
19. Rand, K.D., et al., *ETD in a Traveling Wave Ion Guide at Tuned Z-Spray Ion Source Conditions Allows for Site-Specific Hydrogen/Deuterium Exchange Measurements.* Journal of the American Society for Mass Spectrometry, 2011. **22**(10): p. 1784-1793.
20. Khakinejad, M., et al., *Gas-Phase Hydrogen-Deuterium Exchange Labeling of Select Peptide Ion Conformer Types: a Per-Residue Kinetics Analysis.* Journal of The American Society for Mass Spectrometry, 2015. **26**(7): p. 1115-1127.
21. Pan, J.X., et al., *Structural Interrogation of Electrosprayed Peptide Ions by Gas-Phase H/D Exchange and Electron Capture Dissociation Mass Spectrometry.* Analytical Chemistry, 2012. **84**(1): p. 373-378.
22. Mistarz, U.H., et al., *Simple Setup for Gas-Phase H/D Exchange Mass Spectrometry Coupled to Electron Transfer Dissociation and Ion Mobility for Analysis of Polypeptide Structure on a Liquid Chromatographic Time Scale.* Analytical Chemistry, 2014. **86**(23): p. 11868-11876.
23. Rand, K.D., et al., *Electron capture dissociation proceeds with a low degree of intramolecular migration of peptide amide hydrogens.* Journal of the American Chemical Society, 2008. **130**(4): p. 1341-1349.
24. Zehl, M., et al., *Electron Transfer Dissociation Facilitates the Measurement of Deuterium Incorporation into Selectively Labeled Peptides with Single Residue Resolution.* Journal of the American Chemical Society, 2008. **130**(51): p. 17453-17459.
25. Khakinejad, M., et al., *Ion Mobility Spectrometry-Hydrogen Deuterium Exchange Mass Spectrometry of Anions: Part 3. Estimating Surface Area Exposure by Deuterium Uptake.* Journal of the American Society for Mass Spectrometry, 2016. **27**(3): p. 462-473.
26. Khakinejad, M., et al., *Ion Mobility Spectrometry-Hydrogen Deuterium Exchange Mass Spectrometry of Anions: Part 2. Assessing Charge Site Location and Isotope Scrambling.* Journal of the American Society for Mass Spectrometry, 2016. **27**(3): p. 451-461.
27. Wyttenbach, T., et al., *Ion Mobility Analysis of Molecular Dynamics*, in *Annual Review of Physical Chemistry, Vol 65*, M.A. Johnson and T.J. Martinez, Editors. 2014. p. 175-196.
28. Lee, S., et al., *Determination of Cross Sections by Overtone Mobility Spectrometry: Evidence for Loss of Unstable Structures at Higher Overtones.* Journal of Physical Chemistry B, 2010. **114**(38): p. 12406-12415.
29. Donohoe, G.C., et al., *A New Ion Mobility-Linear Ion Trap Instrument for Complex Mixture Analysis.* Analytical Chemistry, 2014. **86**(16): p. 8121-8128.

30. Ghassabi Kondalaji, S., et al., *Comprehensive Peptide Ion Structure Studies Using Ion Mobility Techniques: Part 1. An Advanced Protocol for Molecular Dynamics Simulations and Collision Cross-Section Calculation.* Journal of The American Society for Mass Spectrometry, 2017: p. 1-13.
31. Schmidt, M.W., et al., *GENERAL ATOMIC AND MOLECULAR ELECTRONIC-STRUCTURE SYSTEM.* Journal of Computational Chemistry, 1993. **14**(11): p. 1347-1363.
32. Gordon, M.S. and M.W. Schmidt, *Advances in electronic structure theory: GAMESS a decade later.* Theory and Applications of Computational Chemistry: The First Forty Years, 2005: p. 1167-1189.
33. Vanquelef, E., et al., *RED Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments.* Nucleic Acids Research, 2011. **39**: p. W511-W517.
34. Wang, F., et al., *RED Python: Object oriented programming for Amber force fields.* Abstracts of Papers of the American Chemical Society, 2014. **247**: p. 1.
35. Dupradeau, F.Y., et al., *The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building.* Phys Chem Chem Phys, 2010. **12**(28): p. 7821-39.
36. Bayly, C.I., et al., *A WELL-BEHAVED ELECTROSTATIC POTENTIAL BASED METHOD USING CHARGE RESTRAINTS FOR DERIVING ATOMIC CHARGES - THE RESP MODEL.* Journal of Physical Chemistry, 1993. **97**(40): p. 10269-10280.
37. Case, D.A., Darden T.A.; Cheatham, T E. III,; Simmerling, C.L.;, et al. *Amber12, University of California, San Francisco*. 2012.
38. Mesleh, M.F., et al., *Structural information from ion mobility measurements: Effects of the long-range potential.* Journal of Physical Chemistry, 1996. **100**(40): p. 16082-16086.
39. Shvartsburg, A.A. and M.F. Jarrold, *An exact hard-spheres scattering model for the mobilities of polyatomic ions.* Chemical Physics Letters, 1996. **261**(1-2): p. 86-91.
40. *Mobcal software*. 1997 [cited 2015 5/4/2015]; Available from: http://www.indiana.edu/~nano/software.html.
41. Lawson, C.L. and R.J. Hanson, *Solving least squares problems*. Vol. 15. 1995: SIAM.
42. Team, R.C., *R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.* 2013.
43. Legon, A.C. and D.J. Millen, *DIRECTIONAL CHARACTER, STRENGTH, AND NATURE OF THE HYDROGEN-BOND IN GAS-PHASE DIMERS.* Accounts of Chemical Research, 1987. **20**(1): p. 39-46.
44. Legon, A.C., *Prereactive complexes of dihalogens XY with Lewis bases B in the gas phase: A systematic case for the halogen analogue B - XY of the hydrogen bond B - HX.* Angewandte Chemie-International Edition, 1999. **38**(18): p. 2687-2714.
45. Team", R.C., *R: A Language and Environment for Statistical Computing.* 2015.
46. Kohtani, M., et al., *Proton transfer-induced conformational changes and melting in designed peptides in the gas phase.* Journal of the American Chemical Society, 2006. **128**(22): p. 7193-7197.

4. Comprehensive Peptide Ion Structure Studies Using Ion Mobility Techniques: Part 3. Relating Solution-phase to Gas-phase Structures

4.1. Introduction

The advent of electrospray ionization (ESI) and its application in biomolecular studies not only brought a Nobel Prize to the area of mass spectrometry, but also introduced the possibilities for protein structure characterization [1-3]. The combination of this soft ionization technique with a variety of fragmentation techniques including electron transfer dissociation (ETD) and collision-induced dissociation (CID) facilitated the determination of the primary structures of proteins [4]; however, the undeviating relationship between the higher-order structures of proteins and their biological functions provides an impetus for the analysis of protein behavior in transition from the solution-phase to the gas-phase and vacuum environment of the mass spectrometer.

Early work on multimeric protein complexes sprayed from solutions with biologically relevant pH values and under mild mass spectrometer interface conditions illustrated that the protein complex ions retained their intact quaternary structure during this transition [5]. Later, ion mobility mass spectrometry-mass spectrometry (IMS-MS) provided insight into the secondary and tertiary structures of biomolecules [6, 7]. These protein ions did not exhibit the “equilibrated” gas-phase structures; rather, they were suggested to resemble higher-energy, solution-phase-like conformations [7-9]. These observations introduced the concept of native electrospray and supported the idea that a significant portion of elements of non-covalent interactions governing solution-phase structures of the

proteins are preserved upon spraying [7, 10]. Therefore, ESI-MS has been utilized for over a decade in protein structure elucidation.

Studies shows that introducing high voltages (~2-3 kV) to accomplish the ESI process, leads to the movement of ions under the influence of the electric filed [11, 12]. This ultimately causes a distortion in the solvent at the capillary tip and the formation of a Taylor cone which is dominated by ions exhibiting the polarity corresponding to the field [13]. This change in the surface area of the solution as well as the single-polarity charge accumulation leads to emission of a fine jet and eventually parent droplets [11, 12]. As the droplets fly toward the mass spectrometer, the evaporation of solvent molecules shrinks the droplet to a point where the repulsion between the same-charge ions on the surface overcomes the surface tension of the droplet [12]. This is the point where fission occurs with the generation of progeny droplets to stabilize the droplets [12, 14]. The maximum number of net charges (Z_r) that can be tolerated by the droplet before a fission event is determined by the Rayleigh limit (Equation 1-1) that is directly proportional to the temperature-dependent surface tension value of the solvent constructing the last hydration shell of the droplet [15].

Ultimately, the repeated shrinkage-fission cycles guide the late-stage progeny droplets to a final steps of the ESI process where the release of the dry analyte ion occurs [12, 14].

Despite the extensive application of ESI in protein structure elucidation, there is a controversy with regard to the determination of the exact mechanism involved in the formation of anhydrous ions at the final stage of the ESI process. Two

proposed mechanisms - Dole's Charge Residue Model (CRM) [16] and the Ion Evaporation Model (IEM) [17]- are normally used to explain the underlying process. CRM illustrates the scenario where evaporation and shrinkage of the droplet having a similar size to the carried, native-like (globular) protein ion continues to complete dryness of the analyte ion [16, 18, 19]. In contrast, the IEM describes the ejection of a small analyte ion directly into the gas-phase environment. Indeed, the IEM governs the release of charge carrier ions such as Na^+ and H_3O^+ present on the surface of the droplet during the fission events thereby reducing the electrostatic repulsion by direct removal of an ion [17, 20-23]. It is worth mentioning that these small charge carriers leave the droplet in the form of small ion-solvent clusters [20, 21, 23-25].

Konermann, *et al.,* proposed a third mechanism – the Chain Ejection Model (CEM) which illustrates the release of an unfolded protein or polymer [23]. According to the CEM, the exposure of the hydrophobic surfaces - which would be hidden inside the protein core in a folded configuration- to the non-polar, gas-phase environment upon the ESI process, results in the ejection of the extended analyte from the surface of the droplet [23].

Molecular dynamics (MD) simulations have been widely used to accomplish structural and mechanistic insight into the behavior of the solvent, charge carrier species as well as the protein structure inside the droplet [21, 23, 26]. In work described here, extensive MD simulations are employed to examine the structure establishment of the gas-phase protein ions by monitoring the fate of the model peptide (Acetyl-PAAAAKAAAAKAAAAKAAAAK) during the transition from the

progeny droplets in the late stages of the ESI process to dryness and further sampling the protein's dynamics in the gas phase. Furthermore, these ESI-generated protein ion conformations are compared with the filtered structures selected from CCS and deuterium uptake behavior criteria (Part 1 and Part 2 installment of this work) as a means to validate the structure selection and HDX algorithm development introduced in Chapters 2 and 3. This validation empowers the algorithms for accurate "conformer type" determination by significantly reducing the computation time otherwise required to obtain such structures through simulating the multiple shrinkage-fission cycles of the ESI process.

4.2. Experimental

4.2.1. Solution-phase representative Structures

The non-polarizable all-atom Amber ff14SB [27] force field was employed to generate the initial structure of the model peptide (Acetyl-PAAAAKAAAAKAAAAKAAAAK) in a fully-helical configuration. All the Lysine residues as well as the C-terminal carbonyl group carried the distinguished charge state values representing those expected at pH=7 in water. The peptide was solvated in a truncated octahedral water box using the rigid 3-point TIP3P water model [28]. The system was neutralized by addition of 3 Cl^- ions and energy minimized. Using the Periodic boundary conditions, the structure was subjected to equilibration for 5 ns at isothermal-isobaric (NPT) ensemble with an average target pressure of 1 atm. The temperature was regulated by a Langevin thermostat [29, 30] with a collision frequency of 1 ps^{-1}. Bond lengths containing the hydrogen atoms were constrained using the SHAKE algorithm [31]. The particle mesh Ewald

(PME) method [32] and a cut-off value of 12 Å were employed to handle the long-range electrostatic interactions. A subsequent 0.5 μs production MD resulted in a trajectory which was sampled for further analyses. The Amber 16 [33] molecular dynamics package and the GPU-accelerated *pmemd* implementation were employed to perform the MD simulations.

After removal of counter ions and water molecules, the *cpptraj* module [34] was employed to perform dihedral cluster analysis with a mask on the dihedral angles passing through 10 central amino acid residues. The closest structures to the mathematically-generated centroids were subjected to a secondary structure analysis using the STRIDE [35] algorithm. This provided 4 major structure types which were selected as the solution-representative conformations.

4.2.2. The ESI Process Simulation

4.2.2.1. Droplet Generation

The fully helical conformer from the selected closest-to-the-centroid structures was modified in order to generate gas-phase $[M+3H]^{3+}$ peptide ions with charge state arrangements of K(6)-K(11)-K(21) and K(6)-K(16)-K(21). The secondary structure elements and dihedral angle values were preserved. The empirical force field parameters for the C-terminal Lysine residue terminating the peptide with a carboxylic acid group were calculated using the R.E.D. server [36-41]. Briefly, two different ACE-capped initial structures of this residue were subjected to geometry optimization at HF/6-31G(d) theory level fallowed by molecular electrostatic potential (MEP) computation and charge fitting steps.

The peptide ion was solvated in an octahedral TIP3P water box [28]. By applying an in-house script, the system was carved into a spherical droplet with a 2.5 nm distance from the center of mass of the peptide as the droplet diameter.

The force field parameters for hydronium ion (H_3O^+) were generated by modifying the Baaden [42] parameters for this ion to be compatible with Amber 16 [33]. Using the surface tension value of 52.3 $mJ.m^{-2}$ [43] for TIP3P water model at 300 K representing the experimental effective temperature, the Rayleigh limit was determined as a maximum of 13 charges which leads to an upper limit value of 10 H_3O^+ ions for the droplet. Therefore droplets with 0, 1, 4 and 10 H_30^+ ions located at random initial positions inside the droplets were generated. The number of water molecules ranged from 1814 to 1840 molecules.

4.2.2.2. Droplet Shrinkage

After extensive (>5000) minimization steps, the relaxed droplet was equilibrated at 300 K for 50 ps and subsequently subjected to a desolvation method similar to the algorithm introduced by Konermann, *et al.,* [26] described as "The trajectory stitching". Briefly, in order to regulate the temperature during the process and prevent evaporation cooling, every 150 ps the simulation was stopped and new, initial random velocities from a Maxwellian distribution at 300 K were assigned. In addition to this, the evaporated molecules present at distances larger than 6.5 nm from the center of mass of the droplet were removed to decrease the computational cost. The simulation then restarted and this cycle continued until all hydronium ions were ejected from the system and all water molecules evaporated. This lead to an overall simulation times ranging from 600 to 950 ns.

The simulations were performed in vacuum using the GPU-accelerated *pmemd* implementation in Amber 16 [33]. No long-range cutoff values for handling the electrostatics and van der Waals interactions were considered. The temperature during each 150-ps run was maintained at 300 K using Langevin dynamics [29, 30] with a collision frequency of 1 ps^{-1}. Four replicas were generated for each system leading to an overall 32 runs (two different charge arrangements for peptides in droplets with overall charges of +3, +4, +7 and +13).

4.2.3. Post-ESI Gas-phase Structures

After complete desolvation, the remaining peptide ions were subjected to production MD simulations at 300 K for 5 ns with no cutoffs for the long-range interactions. The sampled "post-ESI" trajectories were subjected to a k-means clustering algorithm using the *kclust* module in the MMTSB toolset [44]. In combination with an in-house script [45], the number of generated clusters was optimized according to an elbow-point. This provided the value of 50 (±10) clusters per trajectory of 5,000 sampled structures. Collison cross section (CCS) calculations were performed for the resulting closest structures to the centroids [45] using the Mobcal [46] software suite and using trajectory method (TM) [47].

The pairwise all-atom RMSD calculations were performed between all structures sampled from the 32 post-ESI trajectories and the 300 trajectories containing the gas-phase candidate structures (with matching charge state arrangements) obtained at the first implementation of this work (Chapter 2) [45] using an in-house script and the *cpptraj* [34] module implemented in Amber 16 [33]. Briefly, in order to obtain these 300 trajectories, 4,000 cycles of simulated

annealing (SA) were performed on peptides with K(6)-K(11)-K(21) and K(6)-K(16)-K(21) charge arrangements. The pool of thermodynamically-stable as well as higher-energy "annealed structures" were heated to 300 K, equilibrated and subjected to production MD simulation at the same temperature. k-means cluster analysis was performed on the resulting trajectories. CCS calculations on the closest structure to the obtained centroids provided a series of trajectories with matching CCS values to experimental values. These 300 trajectories will be referred to as "the reference trajectories". For more detail see the first implementation of this method as described in Chapter 2.

4.2.4. Peptide Ions in Bulk Solution

To achieve a comparison between the behavior of the peptide ions in a nano-scale droplet and bulk solution of the same solvent, the two different charge arrangements of the $[M+3H]^{3+}$ fully helical peptide ions used in the ESI simulations, were hydrated in a truncated octahedral water box (12 Å) of TIP3P [28] without the present of any counterions. These systems were energy minimized. Using the periodic boundary conditions, the relaxed systems were subjected to a NVT-ensemble heating simulation to obtain a final temperature of 300 K using a Langevin thermostat with varying collision frequencies ranging from 1.0 to 0.1 ns^{-1} over the course of the simulation. This was followed by an equilibration step for 1 ns with a cutoff value of 15 Å for long-range interactions at NPT conditions. The production MD simulations were further performed by continuing the equilibration conditions for 100 ns. After removal of all the water molecules, secondary structure analysis using the STRIDE [35] algorithm was performed on the dry peptide ions.

4.3. Results and Discussion

4.3.1. The fate of the Water Droplet

The evaporation of the last, single water molecules at a close distance to and interacting with charged lysine residues in the final stages of the evaporation process was computationally expensive. Additionally, these last remaining water molecules act as adducts and do not alter the conformation of the peptide ions [21, 48]. Therefore, in the final stages of the simulation, the last 3 remaining water molecules (one for each charged lysine residue) were eliminated manually in 3 out of 4 repetitions for each charge arrangement of K(6)-K(11)-K(21) and K(6)-K(16)-K(21).

Calculation of backbone-only RMSD values on protein structure during the shrinkage simulations -obtained for the other replica (complete evaporation to dryness)- validates this methodology by showing a plateau-like behavior for the RMSD plot at the final stages of evaporation containing limited numbers of water molecules to complete dryness (Figure 4-1). The *cpptraj* module [34] was used to obtain RMSD values (Figure 4-1) for the stripped peptide ions relative to the initial fully-helical, pre-equilibrated structures at t=0 of the corresponding ESI simulation.

The droplet geometry is a parameter that needs to be considered in order to validate the application of the sphere-specific Rayleigh equation (Equation 1-1) over the course of the evaporation.

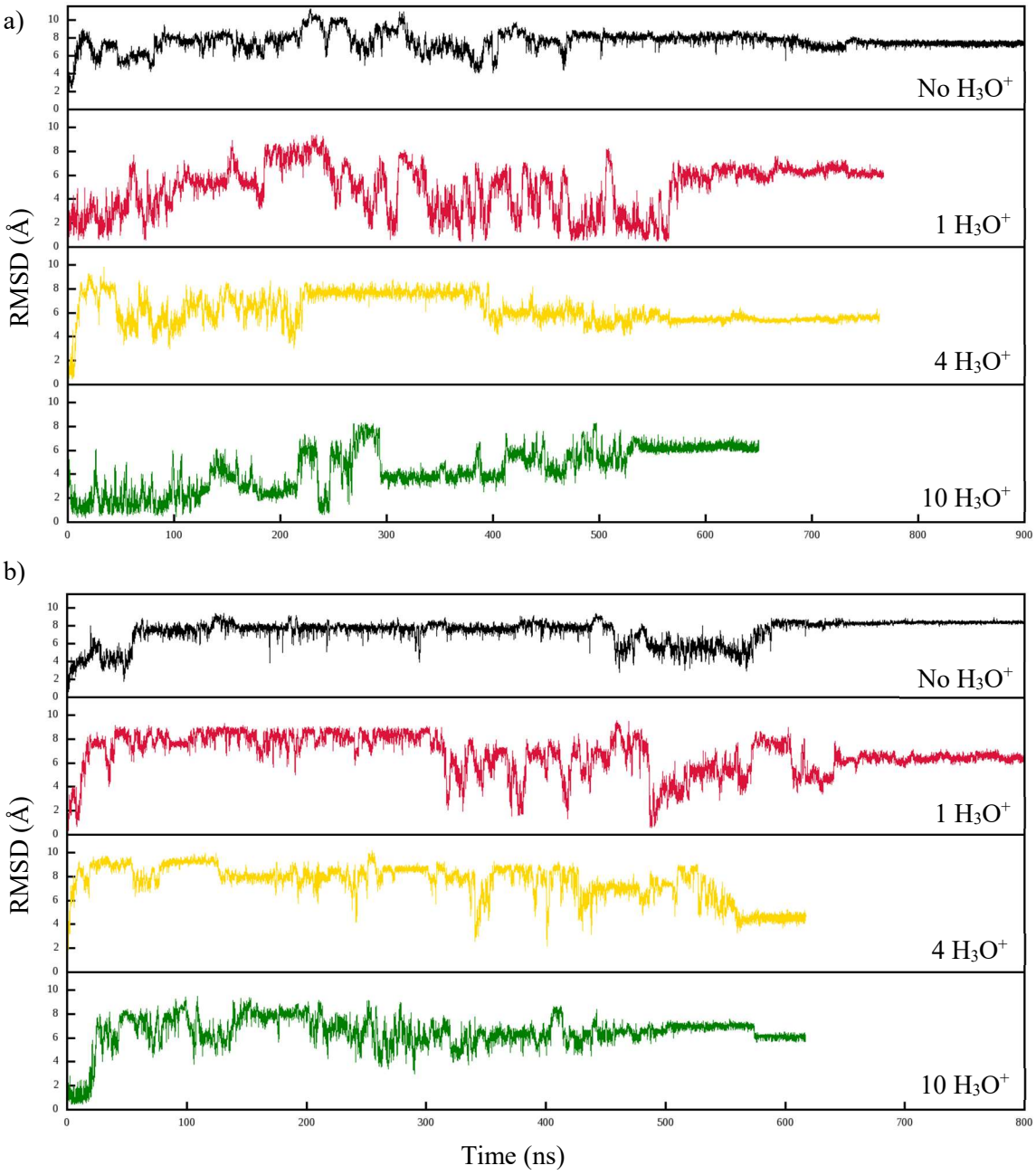

Figure 4-1. Backbone-only RMSD values of peptide ions during the ESI process relative to the initial fully-helical structures. Panel a represents data for the K(6)-K(11)-K(21) charge arrangement while the panel b shows the data obtained from the K(6)-K(16)-K(21) charge arrangement. Each color represents a droplet with a distinguishing total charge value at t=0 of the simulation. The simulations were performed to complete dryness.

To evaluate the droplet geometry, an asymmetry parameter (α) [49, 50] (Equation 4-1) as well as I_1/I_3, I_1/I_2 and I_2/I_3 ratios were introduced through calculating the principal moments of inertia (MOI) - I_1, I_2 and I_3 ($I_1 > I_2 > I_3$)– for the droplet during the ESI simulation. Only water molecules were considered as the shape defining molecules.

$$\propto = \frac{2I_1 - I_2 - I_3}{I_1 + I_2 + I_3} \qquad Equation\ 4-1$$

α values below 0.1 (slight deviation from sphere to an ellipsoidal geometry) - as the threshold limit here- and the MOI ratios of 1 correspond to a spherical geometry.

Figure 4-2 shows the I_1/I_3 ratios of the droplets in series of evaporation-to-dryness simulations. The results show a steady trend at a value of 1 with slight variation from 1 for the first several hundred nanoseconds of the simulations. In addition to slight distortion in the droplet geometry, these initial variations also correspond to the frames sampled from the trajectory before applying the 6.5 nm removal cutoff at the end of each 150-ps cycle. During these 150-ps simulations, the evaporated water molecules were allowed to diffuse from the surface of the droplet until they reach a distance of 6.5 nm from the center of mass of the droplet and therefore were considered in the MOI calculations.

As shown in Figures 4-2 and 4-3, during the last stages of evaporation, the I_1/I_3 ratios exhibit a significant variation from a value of 1 representing the perfect spherical geometry.

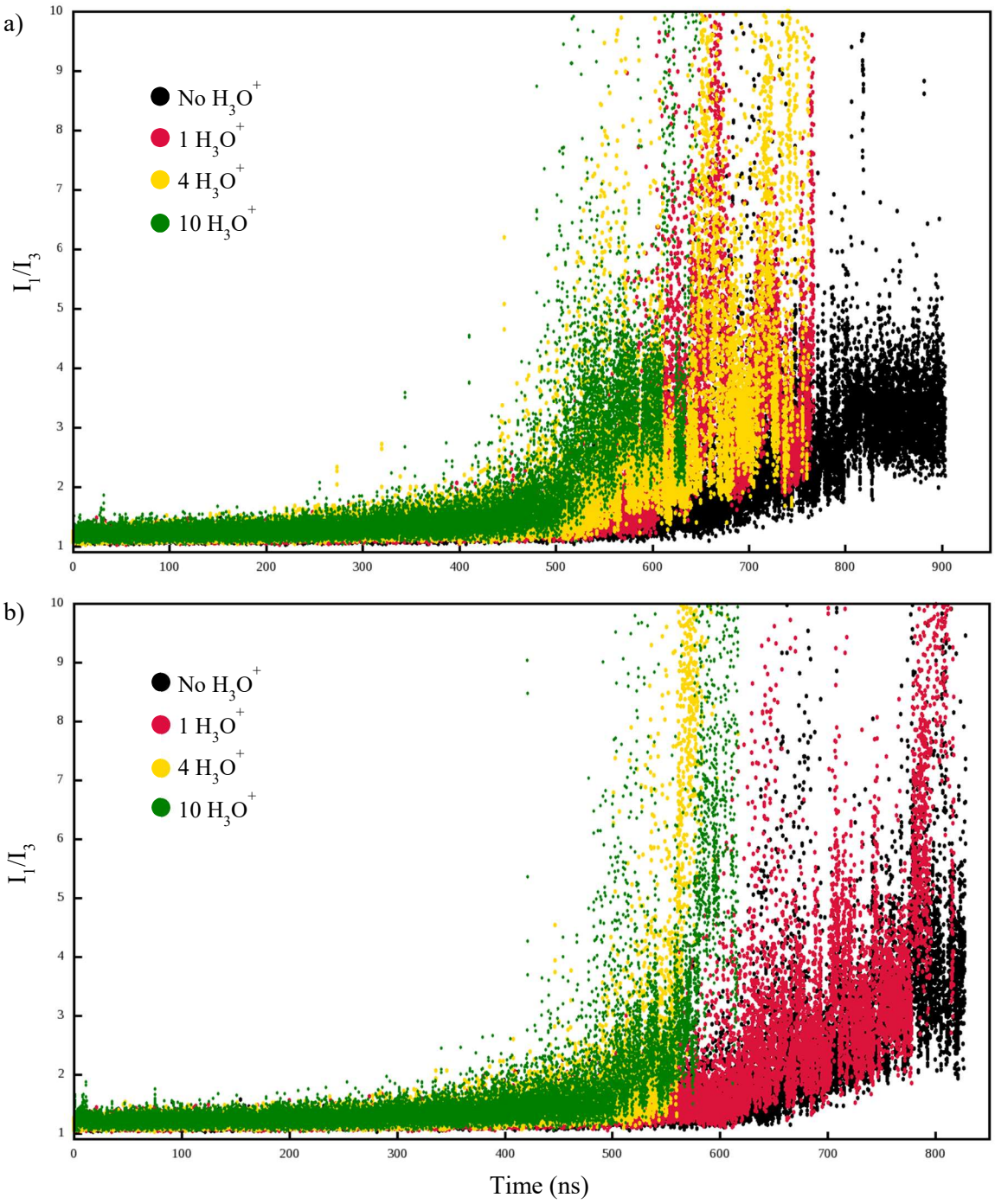

Figure 4-2. The I_1/I_3 ratios of the droplets during the evaporation to dryness. Each color represents a particular charge state value for the initial droplet. Panels a and b represent the MOI ratio for droplets containing K(6)-K(11)-K(21) and K(6)-K(16)-K(21) charge arrangements, respectively.

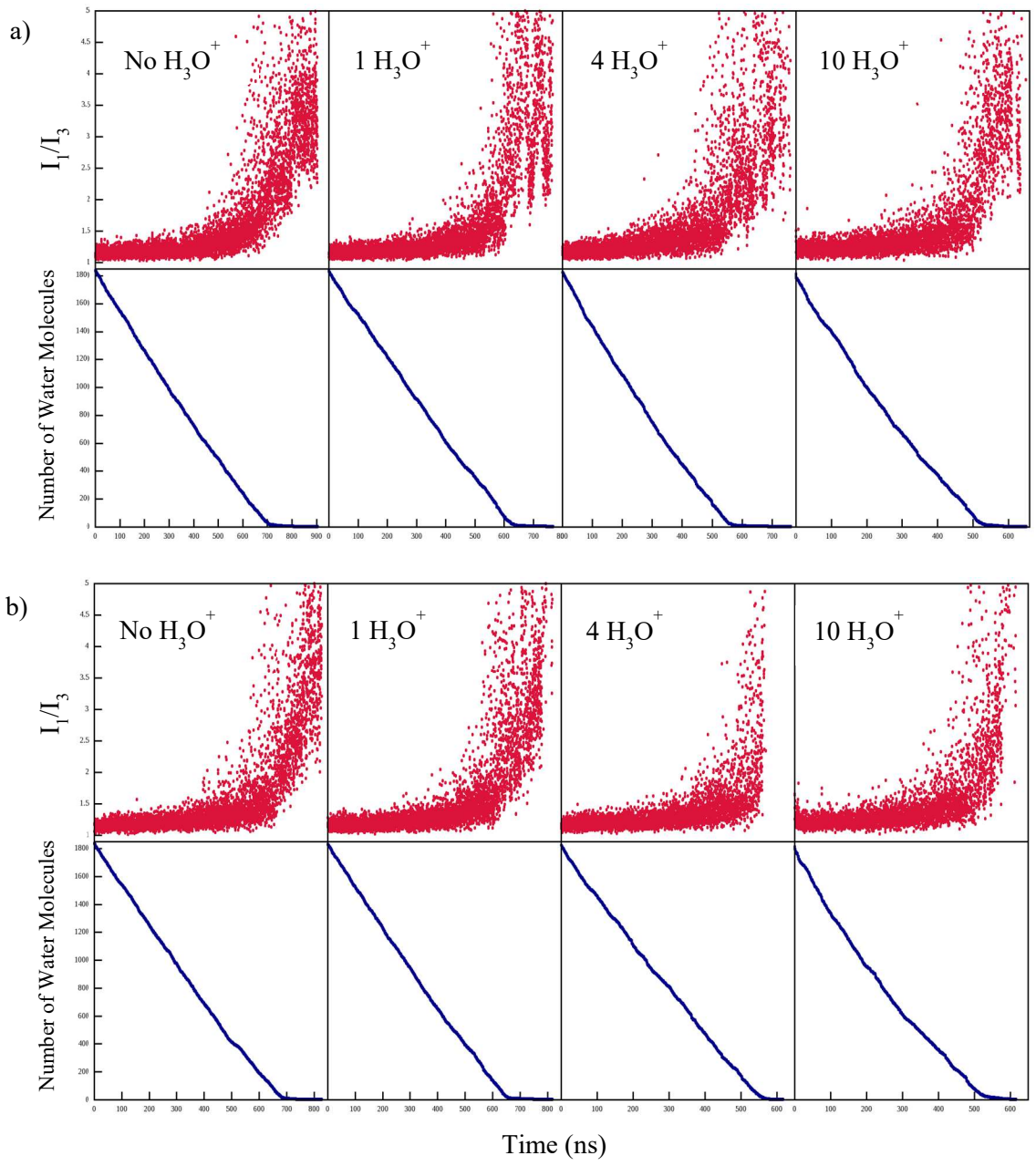

Figure 4-3. Comparison of the number of remaining water molecules to the ratio of the moment of inertia (MOI) over the course of the ESI simulation. Panel a represents the data for the K(6)-K(11)-K(21) charge arrangement while panel b shows the same parameters for the K(6)-K(16)-K(21) charge arrangement. Significant variation in the MOI ratio is observed near the dry state.

Visualizing the droplets at such late stages as well as monitoring the number of remaining water molecules (Figure 4-3) reveals the presence of a very limited number of water molecules on the surface of the droplets. In such cases, the remaining system was not considered as a droplet and the Rayleigh equation was not applied in this regime. However, since the ejection of all H_3O^+ ions occurs at stages long before this state, the system asymmetry did not introduce any error in Rayleigh limit monitoring.

Figure 4-4 presents the changes in the value of the asymmetry parameter (α) as the droplet recedes to leave an anhydrous peptide ion. The same behavior for the MOI ratios is also observed for α values. At the beginning of the simulations, this value shows an acceptable agreement with the value of 0.1 as the threshold limit with deviations at some points throughout the trajectory. As the simulation proceeds, the value of α dramatically increases (as high as 0.5) corresponding to the region of the trajectory where the number of remaining water molecules is significantly limited.

In conclusion, due to the values obtained for α and MOI ratios, the droplet would be considered spherical in the beginning of the simulations as required for further analysis.

It is worth mentioning, as the number of H_3O^+ ions increases, the droplet evaporation occurs at considerably higher rates (Figure 4-3).

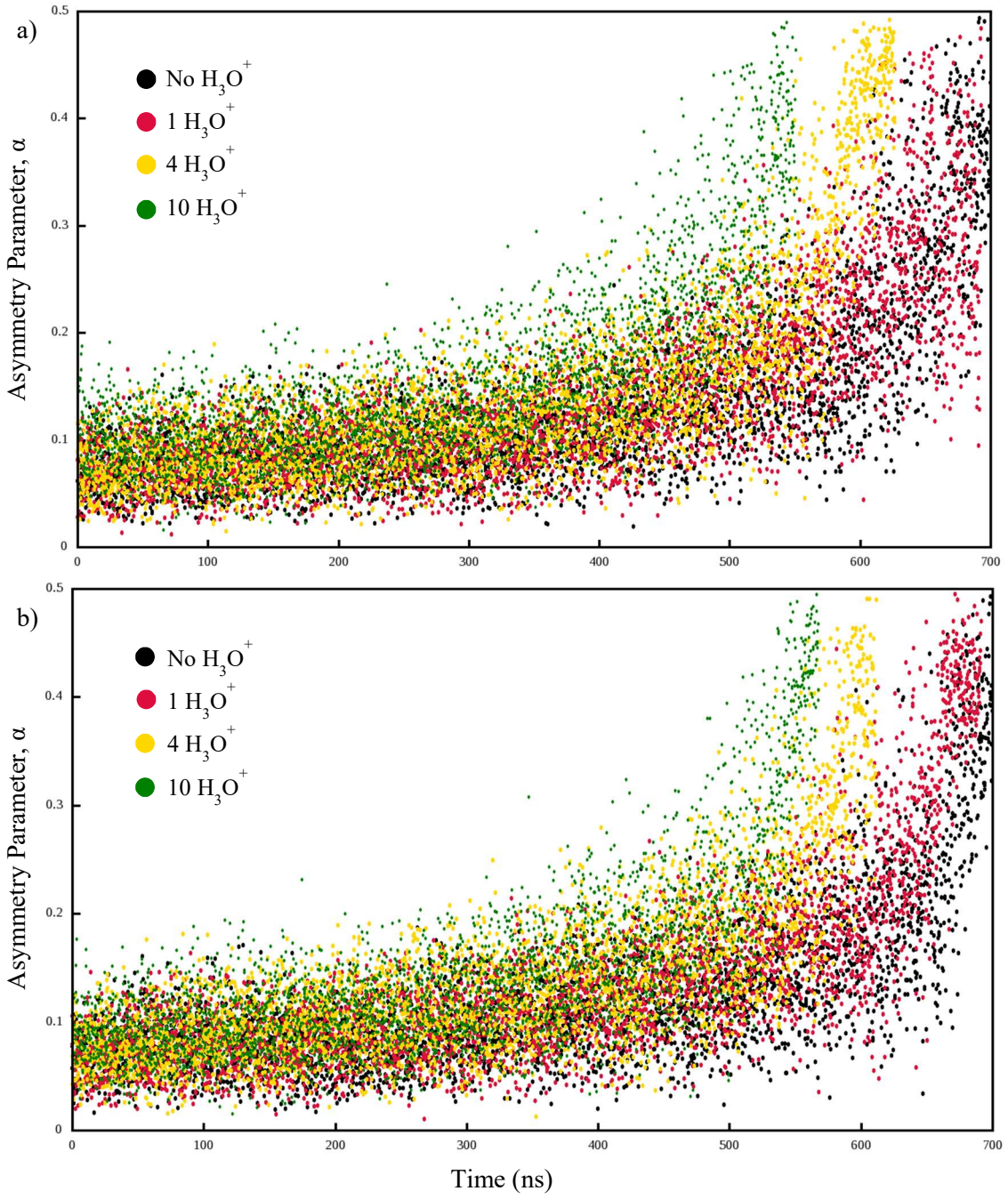

Figure 4-4. The asymmetry parameter (α) of the droplets during the process of evaporation to dryness. Each color represents a particular charge state value for the initial droplet. Panels a and b represent the asymmetry values for droplets containing K(6)-K(11)-K(21) and K(6)-K(16)-K(21) charge arrangements, respectively. The results correspond to separate replicates than those depicted in Figure 4-2 and 4-3.

4.3.2. The Fate of the Hydronium Ion(s)

The majority of the ESI-representative MD simulations start from initial droplets carrying the number of charges matching the maximum allowed value determined by the Rayleigh limit (+13 in this work); however, studies on water droplets carrying a variety of cations and anions (Na^+, NH_4^+, Cl^-, *etc.*) suggest that the utilization of the Rayleigh limit in calculating this upper value for droplets with diameters on the nanometer scale leads to an overestimation in the number of tolerable charges [24, 51]. This effect is profound for droplets carrying predominantly cations [24]. In addition to this, previous studies showed that the surface tension value of the simulated water models is dependent on the size of the droplet [52]. Therefore, in this work, the initial droplets were generated in a fashion to exhibit a variety of total charge state values equal to (+13) and below the Rayleigh limit (+3, +4, +7).

Droplet geometry is related to the Rayleigh limit through the radius of the droplet. In order to calculate the radius of the droplet, the difference in maximum and minimum Cartesian coordinates of the water molecules in the spherical droplet was determined as ΔX, ΔY and ΔZ values for each sampled frame during the ESI simulations. Due to small asymmetries observed for the droplet in the initial few hundred nanosecond of the simulation (see Figure 4-2 and 4-3), the minimum number among the ΔX, ΔY and ΔZ values for each frame was considered as the effective diameter and from that the effective radius (R_{eff}) of the droplet was determined.

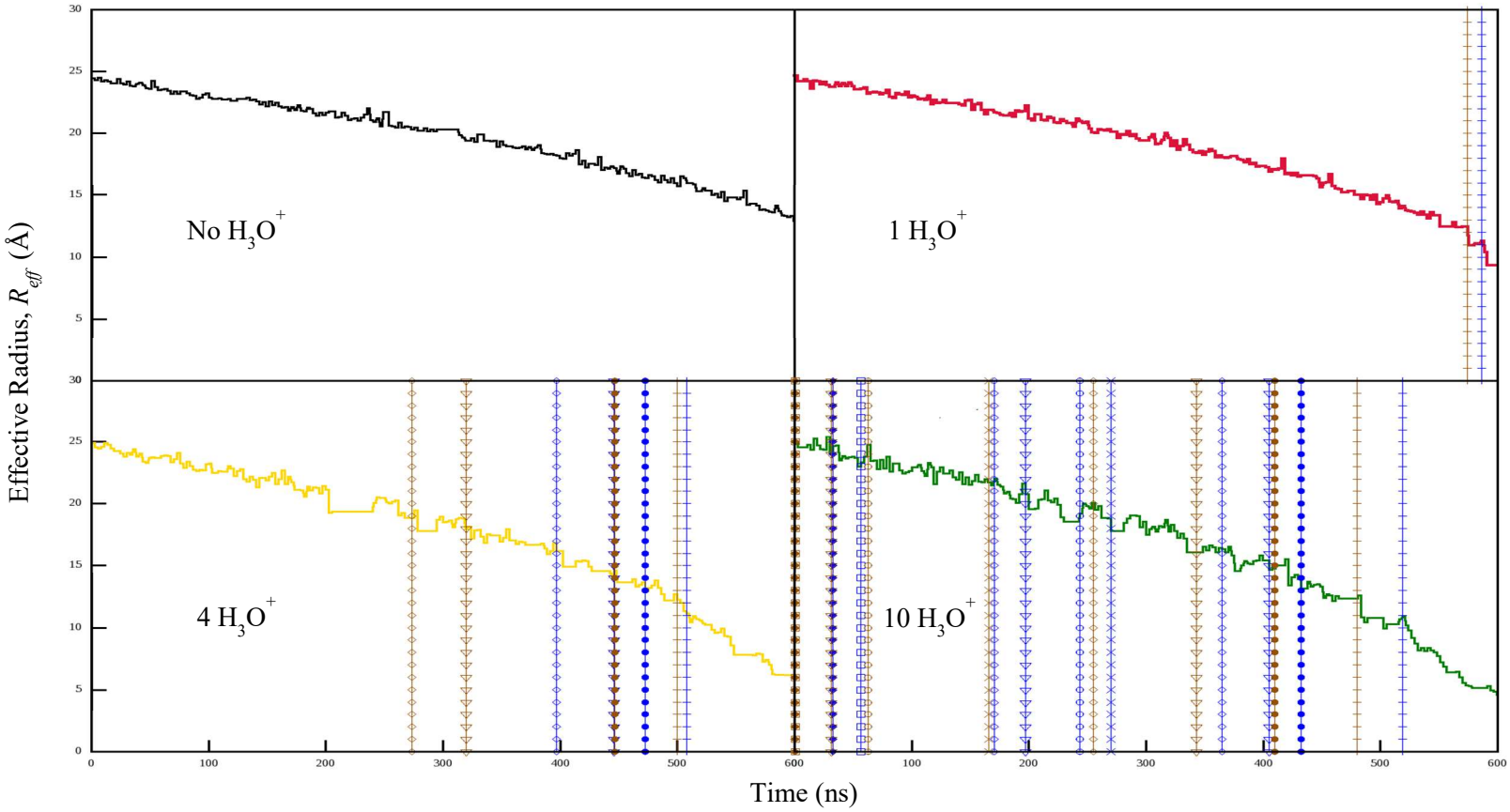

Figure 4-5. The effective radius of the droplets for ESI-representative trajectories obtained for the K(6)-K(11)-K(21) charge arrangement and the four separate droplet charge states of +3, +4,+7 and +13 (0, 1,4 and 10 hydronium ions). The blue vertical lines represent the minimum Rayleigh radius (R_{min}) while the brown vertical lines show the effective radius at which ejection of hydronium ions takes place. The R_{min} value and the related effective radius at the actual ejection points are shown with the same point style.

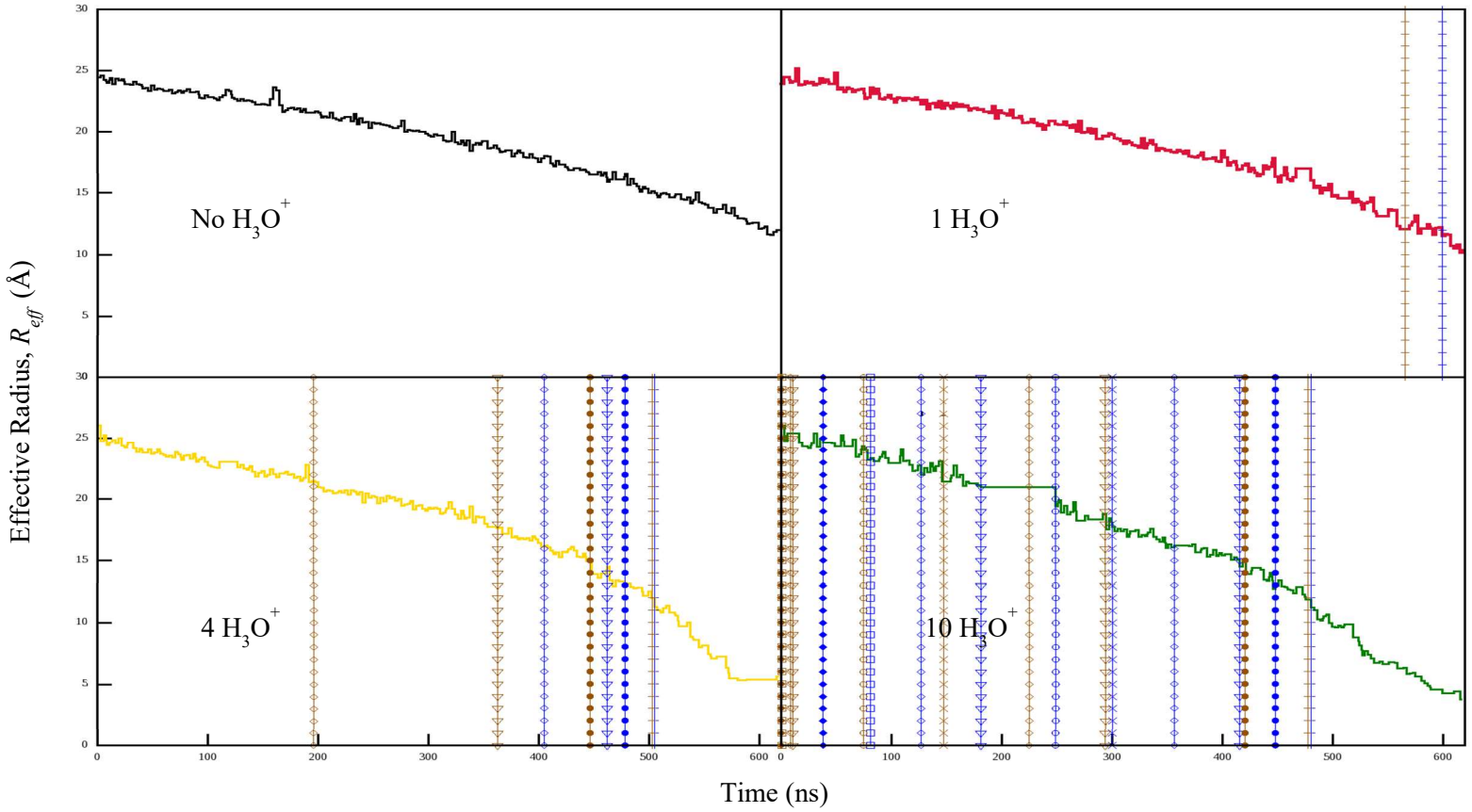

Figure 4-6. The effective radius of the droplets for ESI-representative trajectories obtained for the K(6)-K(16)-K(21) charge arrangement and the four separate droplet charge states of +3, +4,+7 and +13 (0, 1,4 and 10 hydronium ions). The blue vertical lines represent the minimum Rayleigh radius (R_{min}) while the brown vertical lines show the effective radius at which ejection of hydronium ions takes place. The R_{min} value and the related effective radius at the actual ejection points are shown with the same point style.

The Rayleigh equation was solved to obtain the minimum allowed radius (R_{min}) value for a particular droplet charge state ranging from +13 to +3. The R_{min} value is the threshold for a droplet as any radius value below that will require a decrease in the maximum allowed charge state value by direct ejection of the charge carrier species.

The R_{eff} values at which the ejection of the H_3O^+ ions occurred, were extracted from the ESI-representative trajectories. Comparison between the R_{eff} and R_{min} value of the same charge state (+13 to +3) revealed that the ejection of the charge carrier cations happens before the R_{eff} shrinks to the corresponding R_{min} values; that is, the ejection of the H_3O^+ ions takes place at time points before the calculated Rayleigh limit in every case. These results are illustrated in Figures 4-5 and 4-6 for both peptide charge arrangements and droplets of +3 to +13 charge states. These two figures show that in droplets generated at the Rayleigh limit (+10 H_3O^+ and total charge value of +13), the initial ejection of the charge carriers occurs at the very beginning of the simulations and possibly reaches the actual Rayleigh limit corresponding to the water model, droplet size and the surface tension at that particular time points.

This suggested that the generation of an ESI-representative simulation may require an initial droplet with a radius much larger than R_{min} for a particular charge state. This strategy however will increase the computational time. One way to overcome this issue is to start the simulation with a total charge state value below the Rayleigh limit. In such cases the droplet evaporation takes place and the ion ejection occurs at the accurate Rayleigh limit (or below) corresponding to the

simulation parameters. This behavior is observed for the droplets of +7 and +4 charge states shown in Figures 4-5 and 4-6; the ejection of the first hydronium ion takes place after 200 ns and >500 ns, respectively.

A radial distribution analysis of hydronium ions with respect to the center of mass of the droplet throughout the trajectories shows that these charge carriers prefer to locate at a radius near that of the surface of the droplets where the electrostatic repulsion between such same-polarity ions -that is competing with the hydration free energy- is at a minimum. The presence of these ions would disrupt the hydrogen bond network of the solvent at the surface. This effect can be dramatic for ions such as Na^+ which are not capable of forming hydrogen bonds with water molecules. Therefore, one major benefit of initiating ESI-representative simulations at lower charge state values is that there will be less interruption in such hydrogen bond networks and consequently the surface tension values would be higher and thus closer to the values computed for the water models. It appears that the presence of a counter ion such as Cl^- inside the droplet can also affect the location of the hydronium ions through a “shielding” effect, therefore improving the surface tension values [21].

Figure 4-7 shows snapshots of a ESI-representative trajectory obtained from the droplet containing four H_3O^+ ions and peptide ions with charge arrangement of K(6)-K(16)-K(21). As the simulation proceeds, the water molecules evaporate from the surface of the droplet that leads to a decrease in overall droplet size (see Figures 4-5 and 4-6).

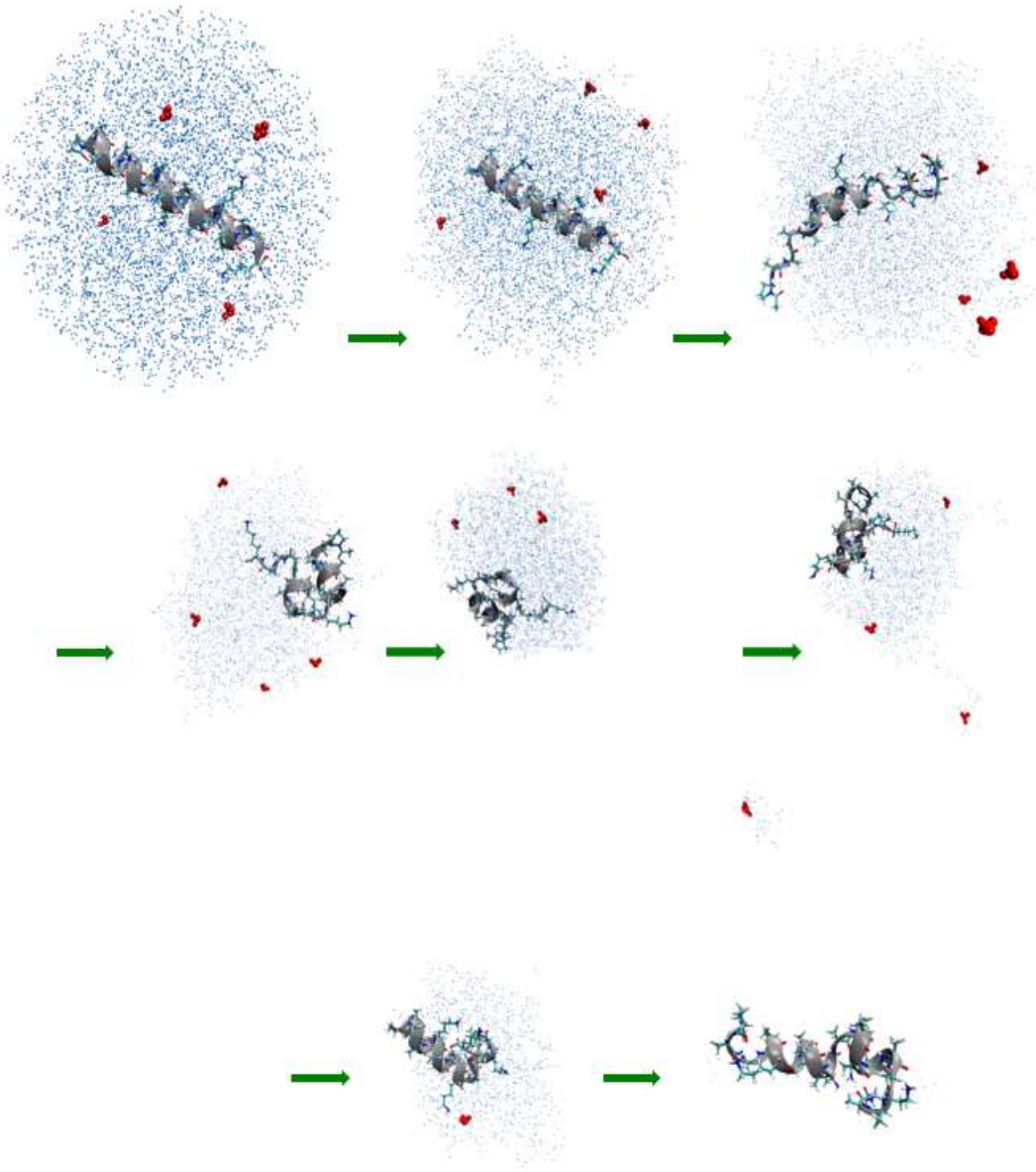

Figure 4-7. Diagram illustrating snapshots of an ESI-representative trajectory with four H_3O^+ ions at the beginning of the simulation. The Hydronium ions are shown by red balls. The peptide secondary structure is represented in grey. The green arrows shows the direction of trajectory sampling. Blue dots represent the water molecules.

At a particular radius, the repulsion between the charge carriers overcomes the surface tension value of the molecules residing at the surface of the droplet. At

this time point, the droplet deforms and generates a cone-shaped tail directing the H_3O^+ ion toward the vacuum environment. The H_3O^+ eventually leaves the droplet as a cluster of water molecules. This behavior is observed in a variety of other simulations that investigate the ESI process [21, 23, 26].

4.3.3. The Fate of the Peptide Ion

Radial distribution analysis of the protein relative to the center of mass of all the solvent molecules in the droplet reveals that the protein moves toward the surface of the droplet as the simulation proceeds. As shown in Figure 4-7, the protein exposes its hydrophobic alanine residues to the non-polar vacuum environment while solvating the highly polar, charged lysine residues in the droplet core. This is also in agreement with previous studies monitoring the behavior of small unfolded peptides in a nanodroplet [23]. However, as the numbers of H_3O^+ species in the initial droplet increase, the translocation of the peptide toward the surface of the droplet occurs over shorter timescales. Therefore, it could be argued that the repulsion between the triply-charged peptide ions and the hydronium ions can also have an influence in locating the peptide at the surface of the droplet and as discussed below, could lead to conformational heterogeneity in the gas phase.

The solution-phase MD simulation of the model peptide carrying the charge state value representing a pH=7 (Section 4.2.1) sampled a series of possibilities for the peptide structure and revealed a rapid dynamics between fully helical, partially helical and random structures. Although the nanodroplet lacks the exact same properties as the bulk solution (such as surface tension values), it is large enough to allow structural fluctuations for the peptide ions. Such dynamics are the

origin of the deviations observed in the RMSD plots at Figure 4-1. Therefore, the initial peptide structures used in the ESI-representative simulations were fully helical as the most abundant structure observed for the peptide in bulk solution. Notably, peptide molecules lack the hydrophobic interactions present at the core of folded globular proteins suggested to preserve elements of secondary, tertiary and quaternary structure. Therefore, it may be expected that peptides produce higher fluctuations in secondary structure. Such wide fluctuations necessitate the performance of the ESI-representative simulation at higher numbers of repetition in order to capture as many as possible variations in final structures for such simulations. In this work, limited somewhat by computation power, 32 replicates were generated to sample these structures.

The division of replicates into four different charge carrier numbers (0, 1, 4 and 10) was utilized to investigate the effect of such numbers of these ions on the behavior of the droplet as well as the peptide ion structure. The high electrostatic repulsion among these ions for droplets exhibiting the Rayleigh limit at the beginning of the simulation (10 H_3O^+-simulations with total a of +13 charge state value for the droplet), can direct the peptide ion toward a point in conformational space where the structure is trapped. The generation of nanodroplets with lower initial charge state values (+7, +4, +3 with 4, 1 and 0 H_3O^+ ions, respectively) can decrease the repulsion and providing more tolerance for peptide structural fluctuations.

The behavior of the peptide in the nanodroplet was further investigated by comparison of the peptide structure in the droplet with the results obtained from a

trajectory representing an “infinite radius” droplet. To achieve this, the $[M+3H]^{3+}$ peptide ions with two separate charge arrangements similar to those in the nanodroplets were solvated and subjected to dynamics (see Section 4.2.4). These trajectories will be referred to as “infinite-radius trajectories” in this chapter. This trajectory can provide a better understanding of the effect of the vacuum environment surrounding the nanodroplet in the ESI-representative simulations.

Secondary structure analysis using the STRIDE [35] algorithm was performed to achieve more insight into structural details of the peptide ions. These calculations were separately performed on 4 sections of the peptide ions. Amino acid residues 2 through the first lysine residue (K(6)), residues 8 to residue 12 (K(11)), residues 13 to residue 17 (K(16)) and residue 18 through the last lysine residue at the C-terminal (K(21)). The Acetyl cap is ignored for such calculations. These analyses were applied on various (seven) trajectories: infinite-radius trajectories, post-ESI gas-phase production trajectories and five radius-scanning trajectories obtained from ESI-representative trajectories containing sampled frames extending the point at which the droplet radius evolves from 25 Å to 20 Å, 20 Å to 15 Å, 15 Å to 10 Å, 10 Å to 5 Å and, finally, 5 Å to the dryness.

The helicity (H_R) for a structure (see Section 2.2.3.2 for more details) was calculated for each peptide section. H_R values were employed to generate a population-averaged Helicity ($H_{R(PA)}$) value according to Equation 4-2:

$$H_{R(PA)} = \frac{\sum_{n=1}^{N} H_R \times n}{N} \qquad Equation\ 4-2$$

where N represents the total number of frames in the trajectory and n is the number of frames exhibiting a particular H_R value. The $H_{R(PA)}$ results are shown in Figures

4-8 through 4-15 for two charge arrangements of K(6)-K(11)-K(21) and K(6)-K(16)-K(21) and four different droplets of +3, +4, +7 and +13 charge state. RMSD relative to fully helical structures for corresponding infinite-radius trajectory and ESI-representative simulations (2.5 nm to dryness) are also illustrated.

A glance at the ESI-representative trajectories on Figures 4-8 to 4-15 as well as the RMSD plots on Figures 4-1 and 4-2 reveals that the peptide undergoes structural fluctuations inside the nanodroplet. This behavior is in agreement with infinite-radius trajectories where protein is surrounded by a large network of water molecules. Therefore, the selection of initial 2.5 nm radius for the droplet was large enough to allow possible walks in the conformational space of the peptide.

However, as the number of water molecules decreases to limited values at the end of the ESI-representative simulations, the structure exhibits a “frozen behavior” and significantly restricted fluctuations (plateau regions in RMSD plots at the end of the ESI simulations). The mixture entropy effects as well as the solution-phase interactions between the water and peptide that assist the random walks on potential energy surface (PES) are absent in the vacuum environment. Then, the energy barriers between the possible solution-phase structures of a peptide can now be high enough for the peptide to overcome in the gas phase. Therefore, it is possible that this frozen peptide exhibiting a specific element of the secondary structure at dryness stages will not be significantly altered by further post-ESI gas-phase production simulations.

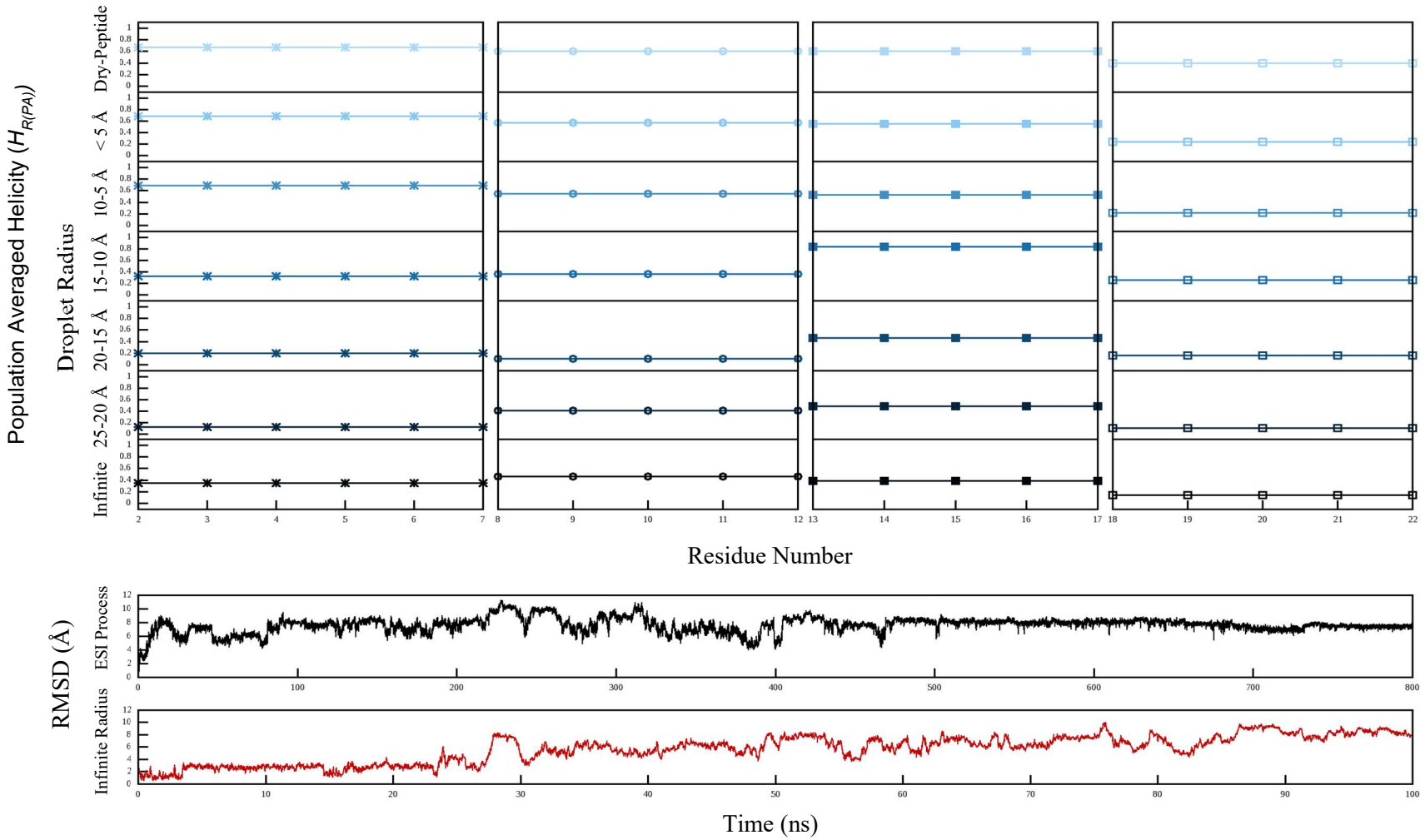

Figure 4-8. The top panel presents the $H_{R(PA)}$ results. The 25 to < 5 Å plots are obtained from the ESI-representative simulations carrying zero H_30^+ ions. The bottom panels represent the RMSD values for the ESI-Representative trajectory with a +3 total charge state (zero H_30^+ ions) (black, top) and infinite radius trajectory (red, bottom). All results are obtained for calculated for the K(6)-K(11)-K(21) charge arrangement simulations.

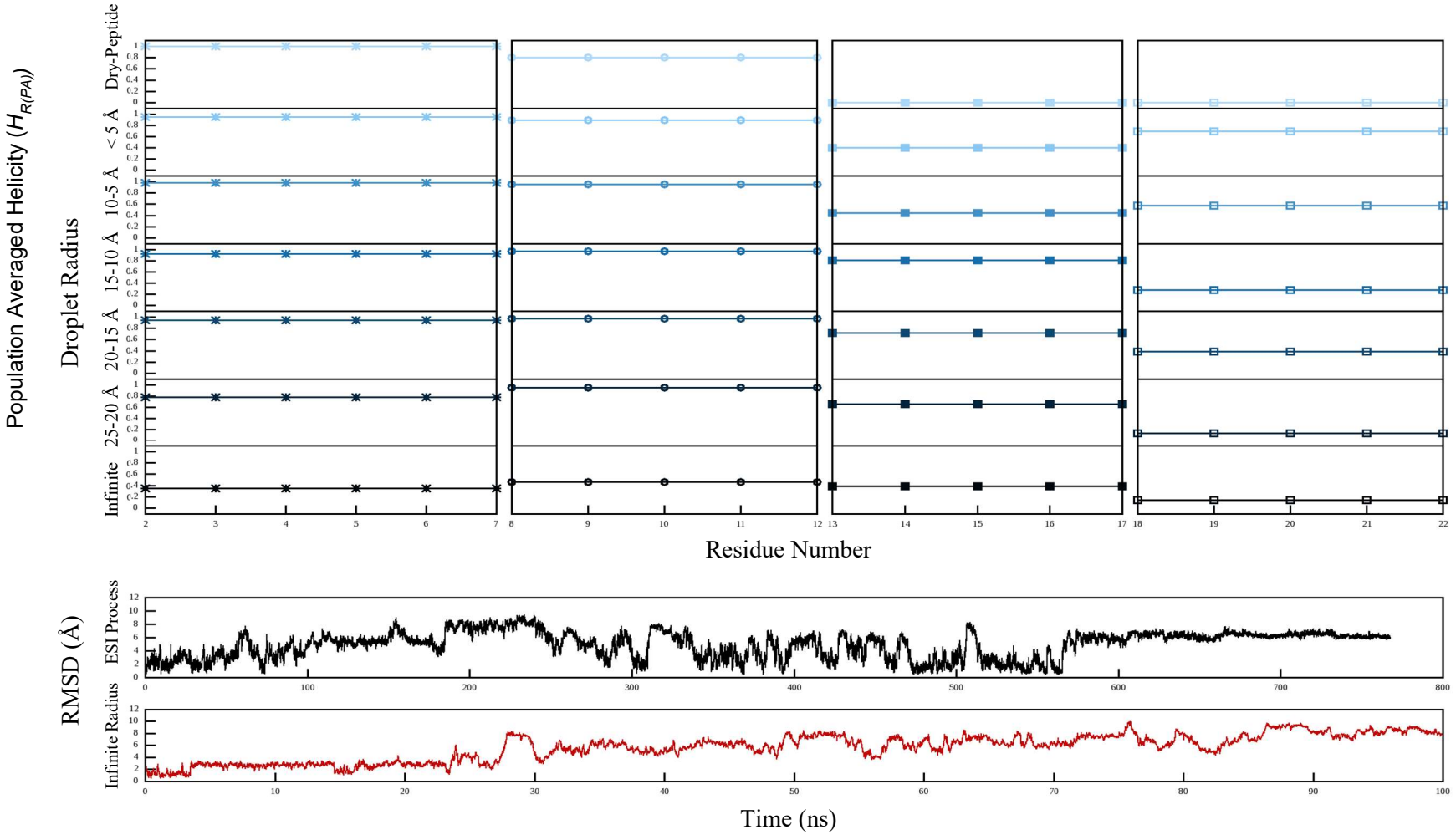

Figure 4-9. The top panel presents the $H_{R(PA)}$ results. The 25 to < 5 Å plots are obtained from the ESI-representative simulations carrying one H_30^+ ions. The bottom panels represent the RMSD values for the ESI-Representative trajectory with a +4 total charge state (one H_30^+ ions) (black, top) and infinite radius trajectory (red, bottom). All results are obtained for calculated for the K(6)-K(11)-K(21) charge arrangement simulations.

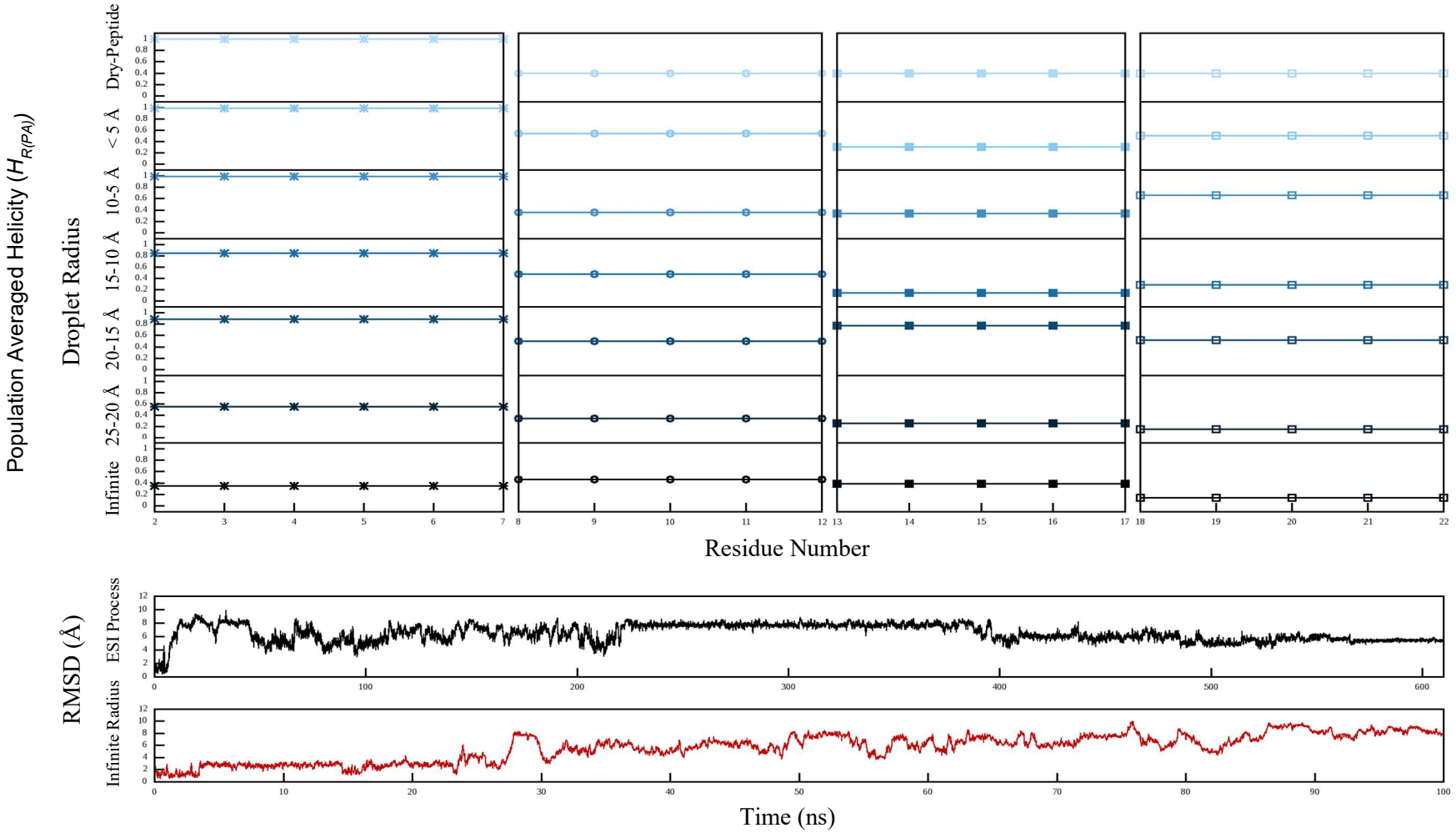

Figure 4-10. The top panel presents the $H_{R(PA)}$ results. The 25 to < 5 Å plots are obtained from the ESI-representative simulations carrying four H_30^+ ions. The bottom panels represent the RMSD values for the ESI-Representative trajectory with a +7 total charge state (four H_30^+ ions) (black, top) and infinite radius trajectory (red, bottom). All results are obtained for calculated for the K(6)-K(11)-K(21) charge arrangement simulations.

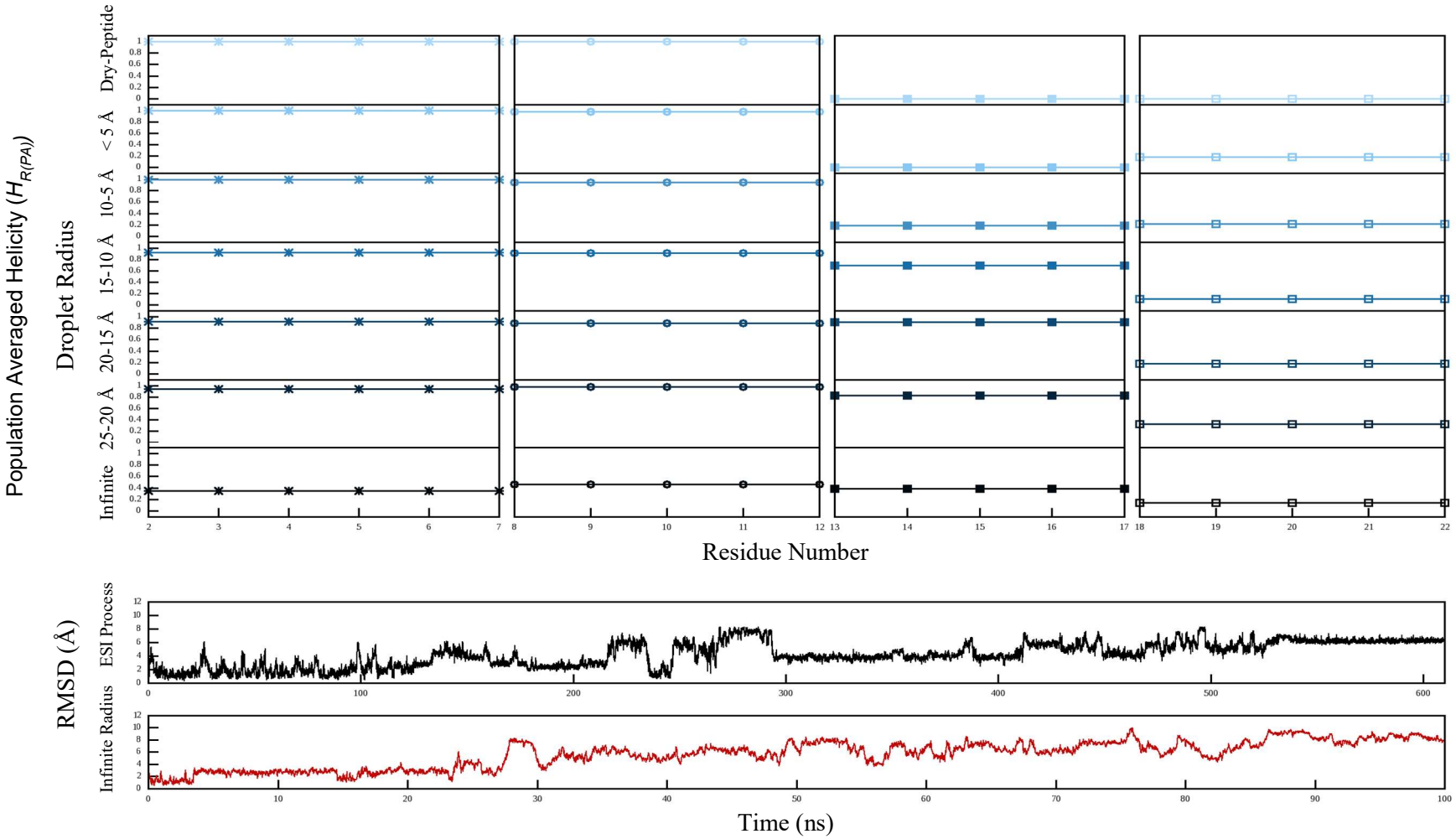

Figure 4-11. The top panel presents the $H_{R(PA)}$ results. The 25 to < 5 Å plots are obtained from the ESI-representative simulations carrying ten H_30^+ ions. The bottom panels represent the RMSD values for the ESI-Representative trajectory with a +13 total charge state (ten H_30^+ ions) (black, top) and infinite radius trajectory (red, bottom). All results are obtained for calculated for the K(6)-K(11)-K(21) charge arrangement simulations.

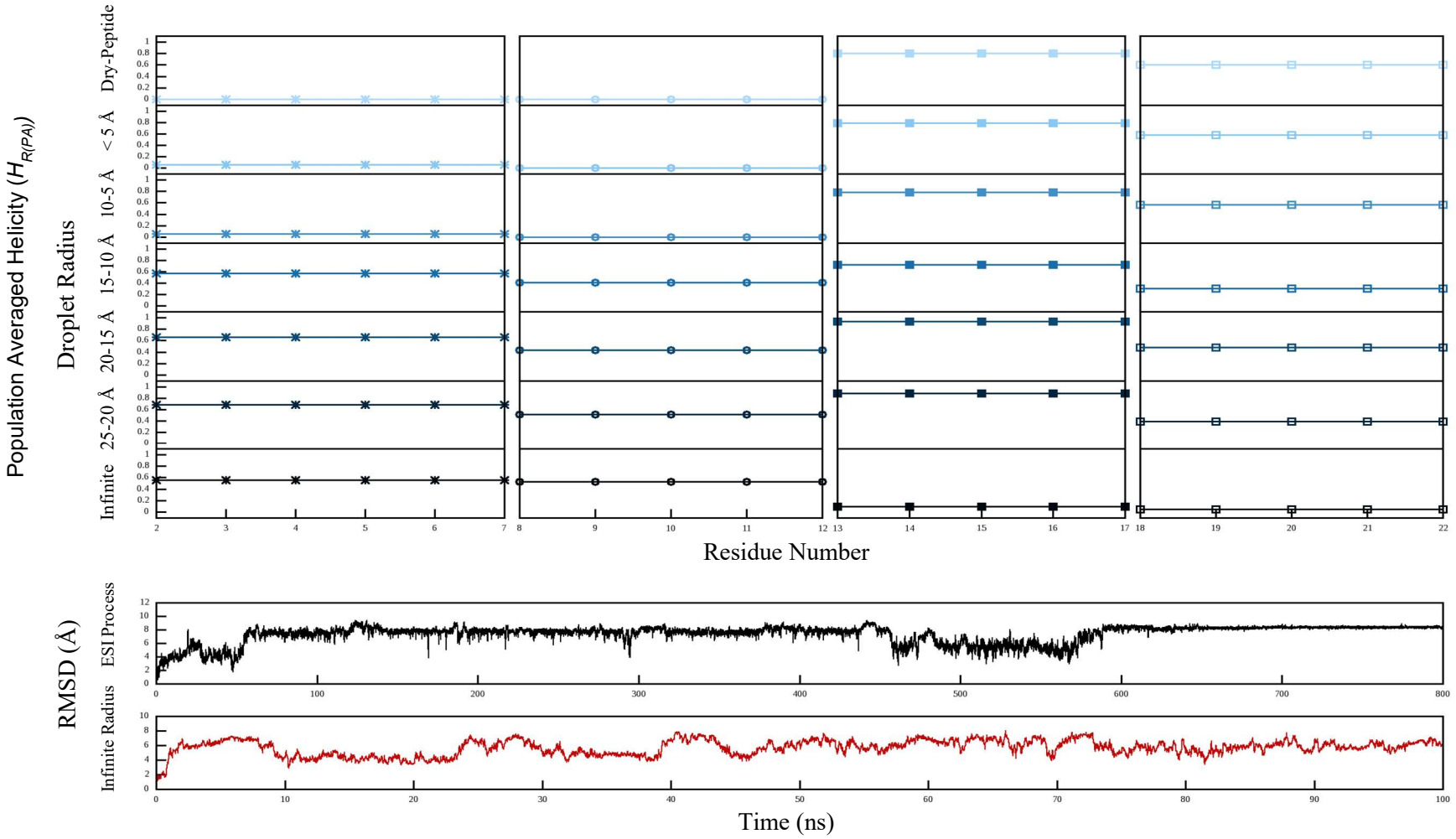

Figure 4-12. The top panel presents the $H_{R(PA)}$ results. The 25 to < 5 Å plots are obtained from the ESI-representative simulations carrying zero H_30^+ ions. The bottom panels represent the RMSD values for the ESI-Representative trajectory with a +3 total charge state (zero H_30^+ ions) (black, top) and infinite radius trajectory (red, bottom). All results are obtained for calculated for the K(6)-K(16)-K(21) charge arrangement simulations.

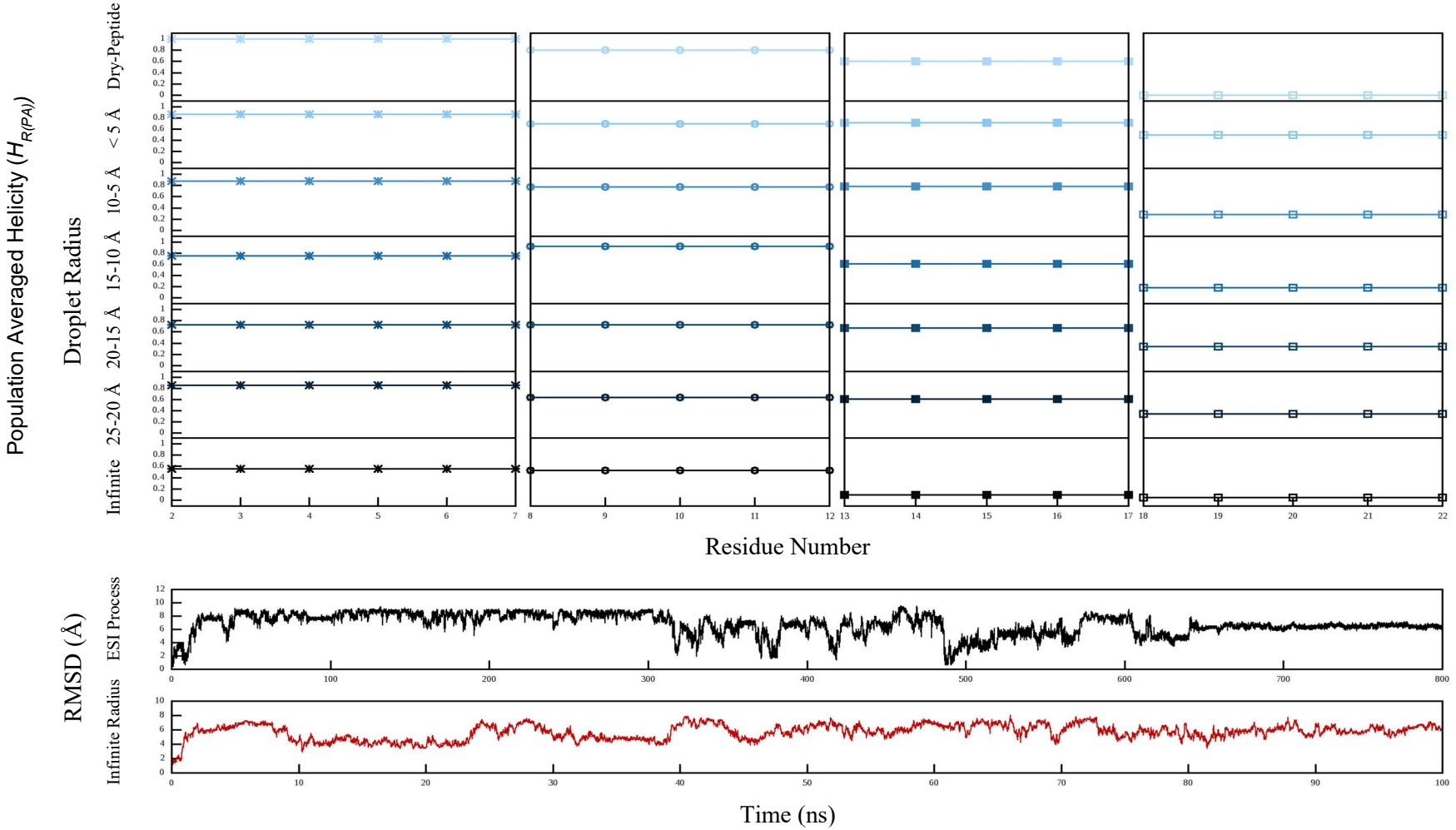

Figure 4-13. The top panel presents the $H_{R(PA)}$ results. The 25 to < 5 Å plots are obtained from the ESI-representative simulations carrying one H_30^+ ions. The bottom panels represent the RMSD values for the ESI-Representative trajectory with a +4 total charge state (one H_30^+ ions) (black, top) and infinite radius trajectory (red, bottom). All results are obtained for calculated for the K(6)-K(16)-K(21) charge arrangement simulations.

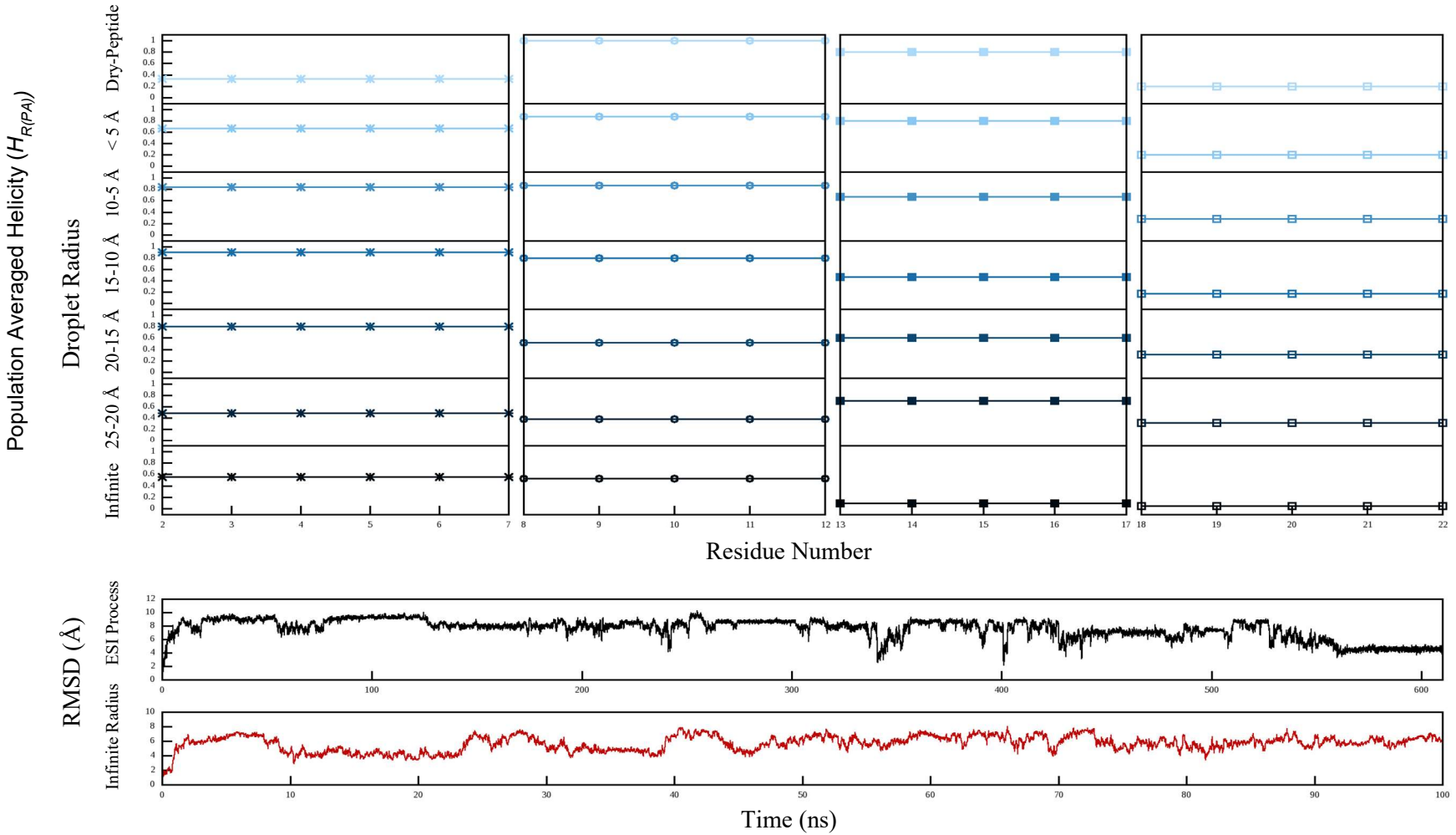

Figure 4-14. The top panel presents the $H_{R(PA)}$ results. The 25 to < 5 Å plots are obtained from the ESI-representative simulations carrying four H_30^+ ions. The bottom panels represent the RMSD values for the ESI-Representative trajectory with a +7 total charge state (four H_30^+ ions) (black, top) and infinite radius trajectory (red, bottom). All results are obtained for calculated for the K(6)-K(16)-K(21) charge arrangement simulations.

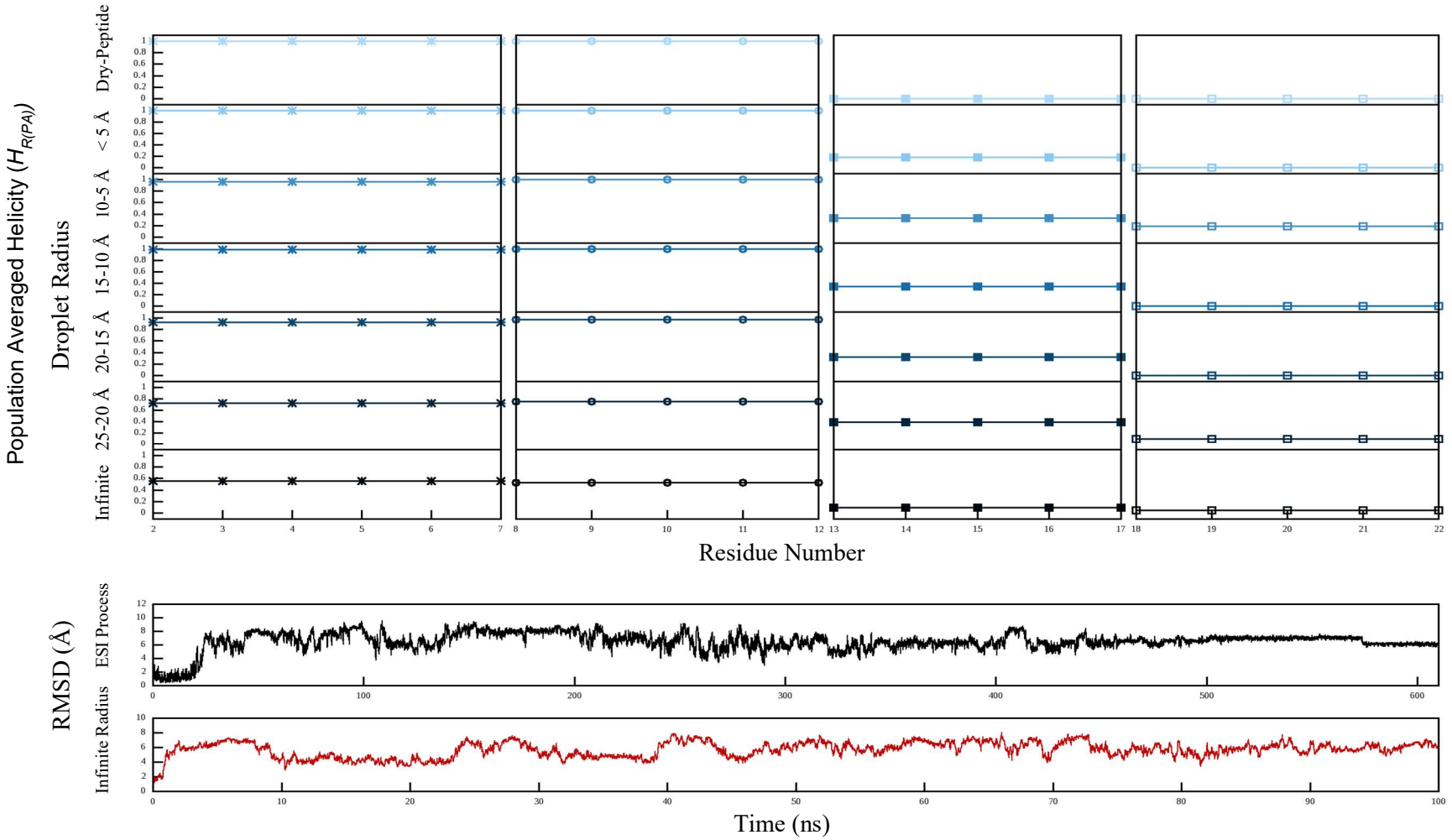

Figure 4-15. The top panel presents the $H_{R(PA)}$ results. The 25 to < 5 Å plots are obtained from the ESI-representative simulations carrying ten H_30^+ ions. The bottom panels represent the RMSD values for the ESI-Representative trajectory with a +13 total charge state (ten H_30^+ ions) (black, top) and infinite radius trajectory (red, bottom). All results are obtained for calculated for the K(6)-K(16)-K(21) charge arrangement simulations.

Figure 4-12 is an example of such behavior where the same secondary structures for all 4 sections of the peptide are observed for the peptide in the droplet with radius of < 5 Å as well as the dry peptide in post-ESI production simulations. Again, this freezing behavior necessitates the replicate simulation for the ESI process in the study of biomolecular structure establishment.

This conclusion does not undermine the requirement of further post–ESI simulation as cases may occur where the energy barriers are sufficiently low such that, the peptide can exhibit changes in elements of the secondary structure where it can be captured by further dynamics in vacuum. Such behavior is observed in Figures 4-9, 4-13 and 4-14 where the helicity values in post-ESI production MD simulations are different than those calculated for the last stages (< 5 Å droplet radius) of the ESI-representative simulations.

The stability of the helical structure for the model peptide is dependent on two effects; the intramolecular hydrogen-bond strength and charge-induced helix disruption and stabilization. In the solution-phase environment, the constant competition between the formation of hydrogen bonds between the water molecules and the peptide can weaken the hydrogen bonds within the helical structure of the peptide. This competition does not exist in the solvent-free vacuum environment. Therefore, as the number of surrounding water molecules decreases to produce a dry peptide, due to this lack of competing forces, the strength of the hydrogen bonds within the helix increases. In contrast, as the number of water molecules surrounding the peptide decreases, the shielding effect of the media decreases. Therefore the helix will face higher repulsion from the positively-

charged residues (lysine here) located N-terminally to the dipole moment of the helix [53]. As the evaporation in the ESI-representative simulations occurs, the population-averaged helicity values can initially increase -as a results of increase in the intramolecular hydrogen bond strength – fallowed by a decrease in this value due to disruption induced by charged residues (see Figure 4-14 residues 8 to 12).

In addition to these phenomena, the presence of two charged residues with the same polarity before and after the helix and the corresponding repulsion between the, will disrupt the helical structure [53]. In the case of the model peptide here, the propensity of the proline residue located at the N-terminus to initiate a helical structure can act as the third factor governing the secondary structure of the peptide. Downstream to this residue toward the C-terminus of the peptide, the proline's helix initiation effect decreases. Figures 4-13 and 4-15 are two examples of such behavior. The population–averaged helicity factor constantly decreases from the first section (residues 2-7) toward the last segment at the C-terminal end (residues 18 to 22).

Figures 4-9 and 4-13 show an example of the effect of such opposing forces on the resultant secondary structures by the comparison between the population averaged helicity parameter for two different charge arrangements in trajectories of the same total droplet charge state (+4 with 1 H_3O^+ ion at the beginning of the simulations). As observed in these Figures, although the two peptides exhibit similar trends in their $H_{R(PA)}$ values for the first, second and the last segments of the primary sequence, the third segments (residues 13 to 17) exhibit a different behavior. The $H_{R(PA)}$ values obtained from the post-ESI production simulations

generally decrease from the N-terminus to the C-terminus as the distance from the proline residue decreases. However, this decrease is significantly dramatic in Figure 4-9 for the residues 13 to 17 in comparison with the same plot in Figure 4-13. The presence of a charged lysine residue (K(11)) located N-terminally to the third segment disrupts the helix formation while this effect is not observed until the last segment for the peptide with the K(6)-K(16)-K(21) charge arrangement (Figure 4-13) that lacks a positive charge for the K(11) residue.

4.3.4. Reference vs Post-ESI Trajectory comparison

By applying the experimental CCS values as the first filtering dimension, 300 structures – produced by two different simulated annealing algorithms- were selected as the "candidate" structures (see Chapter 2 for more details). Deuterium uptake values were employed as the second criterion to yield 13 "conformer types" as the gas-phase representative species among the originated 300 candidate structures of the model peptide (Chapter 3). As mentioned previously (Section 4.1), one of the aims of this study is to make a comparison between the ESI-generated structures and these conformer types in order to validate the filtering methodology (Chapters 2 and 3). To achieve this, all-atom pairwise RMSD calculations between all the frames containing the 300 reference trajectories (see Section 4.4.2) and the post-ESI production simulations of the same peptide charge arrangements were performed (32×5,000 ESI-generated frames with 300×5,000 candidate-containing frames). The most similar structures (lowest RMSD value) from the 32 ESI-representative replicate trajectories to the candidate configurations were extracted. These data were compared with the 13 CCS/HDX-filtered conformer

types. None of the peptide ions from the droplets of +3 and +4 charge states (16 replicates) matched these 13 conformers. This also supports the idea that the ESI-representative simulations require an initial droplet with close-to-the-Rayleigh-limit charge states. The droplets with +13 charge states (8 replicates) only had two matching structures while the ESI-representative simulations exhibiting a system with +7 charge state (8 replicates) provided five structural matches. This limited number of matches for +13 (equal to calculate Rayleigh limit) trajectories can be possibly explained through high electrostatic interactions resulted from the charge carries in the droplet that may direct the structure toward a position on PES where the structure gets trapped and cannot overcome the energy barriers after complete ejection of charge carrier hydronium ions and further production runs at 300 K. Figure 4-16 illustrates two examples of these matching structures with their relative RMSD values.

This significant agreement between the results obtained from two different methods -CCS/HDX filtering algorithm and ESI-representative simulations – validates the methodology applied in former analysis and provides an opportunity to replace the CCS/HDX filtering algorithm with extensive, computationally expensive ESI simulations. The possibility of achieving high sampling rates through conformational space sampling techniques such as simulated annealing (4,000 annealed structures here) empowers this method in providing insight into the gas-phase structures of the model peptide observed inside the drift tube of an IMS-MS device.

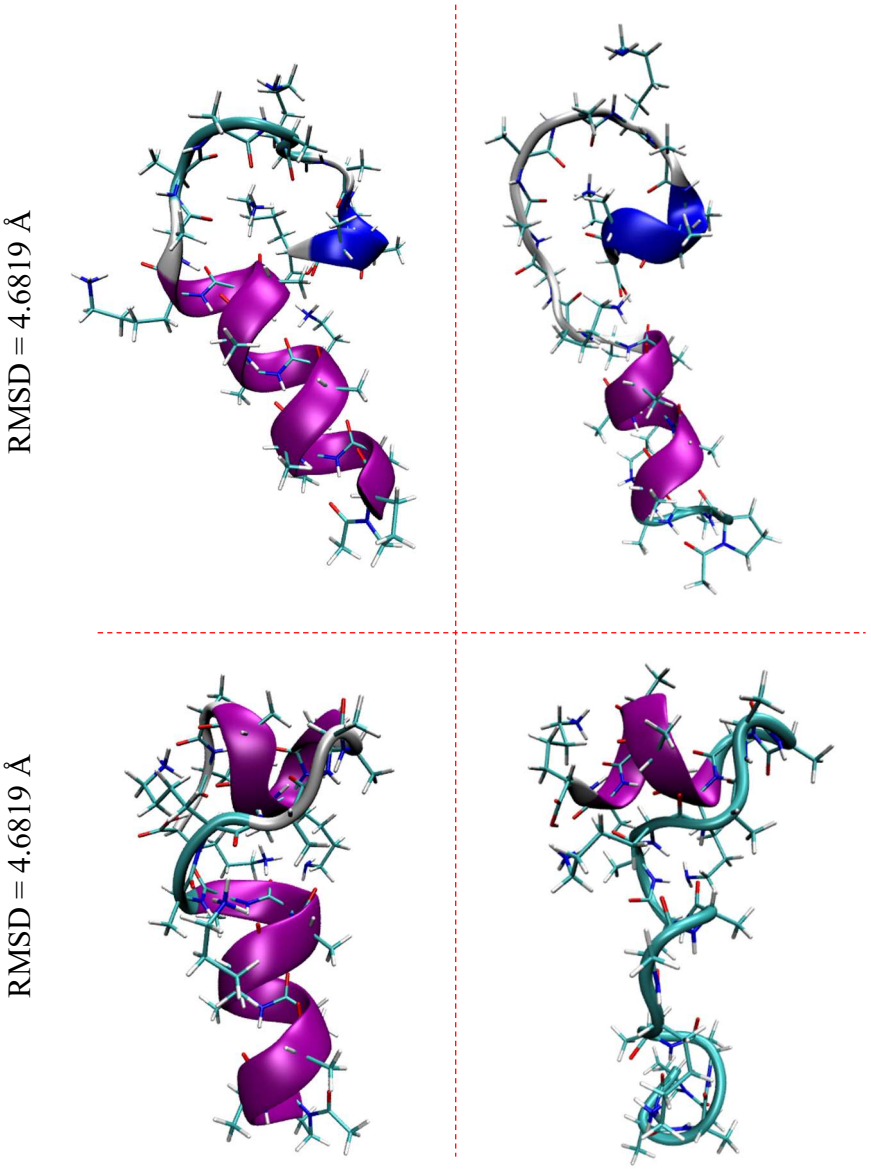

Figure 4-16. Gas-phase conformer types obtained from the CCS/HDX method (right) and ESI-representative simulations (left) with lowest relative RMSD values. The configurations at the top represent a match for a +13 nanodroplet while the bottom structures originate from a +7 nanodroplet.

4.4. Conclusion

Molecular dynamics simulations can be employed to provide insight into the behavior observed in the electrospray ionization process. In this work, several replicates of *in-silico* nanodroplets that simulate the progeny droplets at the last stages of the ESI process were generated and subjected to extensive MD simulations to provide anhydrous peptide ions. Certain elements of the droplet behavior were investigated; including: droplet radius; ejection of charge carries such as H_3O^+ through the ion evaporation model; limitations of the Rayleigh limit approximation for nanoscale droplet simulations; and, the number of these charge carriers. It is proposed here that the ESI-representative simulations require the analysis to start with a number of charge carriers that is below the calculated Rayleigh limit for the water model for a specific droplet radius and simulation temperature. Comparatively, starting the simulations with a limited number of charge carriers does not provide a realistic picture of the ESI process due to the significantly higher concentration of these species at the last stages of the ESI process.

Extensive secondary structure analyses of the model peptide in systems with varying numbers of surrounding solvent molecules were performed. Two opposing effects -intramolecular hydrogen bond strength and charge-induced helix disruption– can be used to describe the observed elements of secondary structure. It is worth mentioning that the deviation from the trend expected in these systems can result from structural “freezing” along the PES due to the lack of sufficient energy to overcome the structural barriers. In addition to these phenomena, the

effect of the proline residue with regard to its propensity to propagate a helical structure within the overall peptide structure was discussed.

As a final note, a comparison between the ESI-generated structures and the conformations obtained through CCS calculations and deuterium uptake levels provided a series of highly congruent structures. These results strongly support the methodology applied in Chapters 2 and 3 and validate the developed algorithms involved in CCS determinations and HDX scoring. It is proposed that the extensive, computationally expensive ESI-representative simulations can be replaced with filtering expansive conformational space structures using gas-phase HDX experiments in a drift tube to obtain gas-phase representative structures. Additionally, the latter method lacks the uncertainties concerning the ESI simulations such as obtaining an accurate surface tension value for the water model at specific droplet radii to determine the number of charge carrier species. That said, there is a growing need for improving these simulations as they provide valuable insight into the mechanism of ESI process, the behavior of charge carriers and even obtained mass spectra which otherwise would be difficult –if not impossible- to describe.

4.5. References

1. Fenn, J.B., *Electrospray wings for molecular elephants (Nobel lecture).* Angewandte Chemie-International Edition, 2003. **42**(33): p. 3871-3894.
2. Fenn, J.B., et al., *ELECTROSPRAY IONIZATION FOR MASS-SPECTROMETRY OF LARGE BIOMOLECULES.* Science, 1989. **246**(4926): p. 64-71.
3. Tanaka, K., et al., *Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry.* Rapid Communications in Mass Spectrometry, 1988. **2**(8): p. 151-153.
4. Shevchenko, A., et al., *Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels.* Analytical Chemistry, 1996. **68**(5): p. 850-858.
5. Lightwahl, K.J., B.L. Schwartz, and R.D. Smith, *OBSERVATION OF THE NONCOVALENT QUATERNARY ASSOCIATIONS OF PROTEINS BY ELECTROSPRAY-IONIZATION MASS-SPECTROMETRY.* Journal of the American Chemical Society, 1994. **116**(12): p. 5271-5278.
6. Hudgins, R.R., J. Woenckhaus, and M.F. Jarrold, *High resolution ion mobility measurements for gas phase proteins: correlation between solution phase and gas phase conformations.* International Journal of Mass Spectrometry, 1997. **165**: p. 497-507.
7. Wyttenbach, T. and M.T. Bowers, *Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-Free Ion Mobility-Mass Spectrometry Environment.* Journal of Physical Chemistry B, 2011. **115**(42): p. 12266-12275.
8. Wyttenbach, T., et al., *The Effect of Calcium Ions and Peptide Ligands on the Relative Stabilities of the Calmodulin Dumbbell and Compact Structures.* Journal of Physical Chemistry B, 2010. **114**(1): p. 437-447.
9. Baumketner, A., et al., *Amyloid beta-protein monomer structure: A computational and experimental study.* Protein Science, 2006. **15**(3): p. 420-428.
10. Ruotolo, B.T. and C.V. Robinson, *Aspects of native proteins are retained in vacuum.* Current Opinion in Chemical Biology, 2006. **10**(5): p. 402-408.
11. Cech, N.B. and C.G. Enke, *Practical implications of some recent studies in electrospray ionization fundamentals.* Mass Spectrometry Reviews, 2001. **20**(6): p. 362-387.
12. Kebarle, P. and U.H. Verkerk, *ELECTROSPRAY: FROM IONS IN SOLUTION TO IONS IN THE GAS PHASE, WHAT WE KNOW NOW.* Mass Spectrometry Reviews, 2009. **28**(6): p. 898-917.
13. Taylor, G.I. and A.D. McEwan, *The stability of a horizontal fluid interface in a vertical electric field.* Journal of Fluid Mechanics, 2006. **22**(1): p. 1-15.
14. Grimm, R.L. and J.L. Beauchamp, *Evaporation and Discharge Dynamics of Highly Charged Multicomponent Droplets Generated by Electrospray Ionization.* The Journal of Physical Chemistry A, 2010. **114**(3): p. 1411-1419.
15. Rayleigh, L., *XX. On the equilibrium of liquid conducting masses charged with electricity.* Philosophical Magazine Series 5, 1882. **14**(87): p. 184-186.
16. Dole, M., et al., *Molecular Beams of Macroions.* The Journal of Chemical Physics, 1968. **49**(5): p. 2240-2249.
17. Thomson, B.A. and J.V. Iribarne, *Field induced ion evaporation from liquid surfaces at atmospheric pressure.* The Journal of Chemical Physics, 1979. **71**(11): p. 4451-4463.
18. Schmelzeisen-Redeker, G., L. Bütfering, and F.W. Röllgen, *Desolvation of ions and molecules in thermospray mass spectrometry.* International Journal of Mass Spectrometry and Ion Processes, 1989. **90**(2): p. 139-150.

19. Fernandez de la Mora, J., *Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism.* Analytica Chimica Acta, 2000. **406**(1): p. 93-104.
20. Metwally, H., et al., *Mechanism of Protein Supercharging by Sulfolane and m-Nitrobenzyl Alcohol: Molecular Dynamics Simulations of the Electrospray Process.* Analytical Chemistry, 2016. **88**(10): p. 5345-5354.
21. Kim, D., et al., *Ions from Solution to the Gas Phase: A Molecular Dynamics Simulation of the Structural Evolution of Substance P during Desolvation of Charged Nanodroplets Generated by Electrospray Ionization.* Journal of the American Chemical Society, 2017. **139**(8): p. 2981-2988.
22. Hogan, C.J., et al., *Combined Charged Residue-Field Emission Model of Macromolecular Electrospray Ionization.* Analytical Chemistry, 2009. **81**(1): p. 369-377.
23. Konermann, L., et al., *Unraveling the Mechanism of Electrospray Ionization.* Analytical Chemistry, 2013. **85**(1): p. 2-9.
24. Caleman, C. and D. van der Spoel, *Evaporation from water clusters containing singly charged ions.* Physical Chemistry Chemical Physics, 2007. **9**(37): p. 5105-5111.
25. Znamenskiy, V., I. Marginean, and A. Vertes, *Solvated Ion Evaporation from Charged Water Nanodroplets.* The Journal of Physical Chemistry A, 2003. **107**(38): p. 7406-7412.
26. Konermann, L., R.G. McAllister, and H. Metwally, *Molecular Dynamics Simulations of the Electrospray Process: Formation of NaCI Clusters via the Charged Residue Mechanism.* Journal of Physical Chemistry B, 2014. **118**(41): p. 12025-12033.
27. Maier, J.A., et al., *ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB.* Journal of Chemical Theory and Computation, 2015. **11**(8): p. 3696-3713.
28. Jorgensen, W.L., et al., *COMPARISON OF SIMPLE POTENTIAL FUNCTIONS FOR SIMULATING LIQUID WATER.* Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
29. Grest, G.S. and K. Kremer, *MOLECULAR-DYNAMICS SIMULATION FOR POLYMERS IN THE PRESENCE OF A HEAT BATH.* Physical Review A, 1986. **33**(5): p. 3628-3631.
30. Wu, X. and B.R. Brooks, *Self-guided Langevin dynamics simulation method.* Chemical Physics Letters, 2003. **381**(3–4): p. 512-518.
31. Ryckaert, J.-P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* Journal of Computational Physics, 1977. **23**(3): p. 327-341.
32. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems.* The Journal of Chemical Physics, 1993. **98**(12): p. 10089-10092.
33. D.A. Case, R.M.B., D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. and T.L. Lin, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao , P.A. Kollman, *AMBER 2016.* 2016, University of California, San Francisco.
34. Roe, D.R. and T.E. Cheatham, *PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data.* Journal of Chemical Theory and Computation, 2013. **9**(7): p. 3084-3095.
35. Frishman, D. and P. Argos, *Knowledge-based protein secondary structure assignment.* Proteins-Structure Function and Genetics, 1995. **23**(4): p. 566-579.
36. Schmidt, M.W., et al., *GENERAL ATOMIC AND MOLECULAR ELECTRONIC-STRUCTURE SYSTEM.* Journal of Computational Chemistry, 1993. **14**(11): p. 1347-1363.

37. Gordon, M.S. and M.W. Schmidt, *Advances in electronic structure theory: GAMESS a decade later.* Theory and Applications of Computational Chemistry: The First Forty Years, 2005: p. 1167-1189.
38. Vanquelef, E., et al., *RED Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments.* Nucleic Acids Research, 2011. **39**: p. W511-W517.
39. Wang, F., et al., *RED Python: Object oriented programming for Amber force fields.* Abstracts of Papers of the American Chemical Society, 2014. **247**: p. 1.
40. Dupradeau, F.-Y., et al., *The R.ED. tools: advances in RESP and ESP charge derivation and force field library building.* Physical Chemistry Chemical Physics, 2010. **12**(28): p. 7821-7839.
41. Bayly, C.I., et al., *A WELL-BEHAVED ELECTROSTATIC POTENTIAL BASED METHOD USING CHARGE RESTRAINTS FOR DERIVING ATOMIC CHARGES - THE RESP MODEL.* Journal of Physical Chemistry, 1993. **97**(40): p. 10269-10280.
42. Baaden, M., M. Burgard, and G. Wipff, *TBP at the Water−Oil Interface: The Effect of TBP Concentration and Water Acidity Investigated by Molecular Dynamics Simulations.* The Journal of Physical Chemistry B, 2001. **105**(45): p. 11131-11141.
43. Vega, C. and E. de Miguel, *Surface tension of the most popular models of water by using the test-area simulation method.* The Journal of Chemical Physics, 2007. **126**(15): p. 154707.
44. Feig, M., J. Karanicolas, and C.L. Brooks, *MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology.* Journal of Molecular Graphics & Modelling, 2004. **22**(5): p. 377-395.
45. Ghassabi Kondalaji, S., et al., *Comprehensive Peptide Ion Structure Studies Using Ion Mobility Techniques: Part 1. An Advanced Protocol for Molecular Dynamics Simulations and Collision Cross-Section Calculation.* Journal of The American Society for Mass Spectrometry, 2017: p. 1-13.
46. Jarrold, M.F. *Mobcal, A program to calculate mobilities*. 2/10/2014 2/18/2016]; Available from: http://www.indiana.edu/~nano/software.html.
47. Mesleh, M.F., et al., *Structural information from ion mobility measurements: Effects of the long-range potential.* Journal of Physical Chemistry, 1996. **100**(40): p. 16082-16086.
48. Konermann, L., A.D. Rodriguez, and J. Liu, *On the Formation of Highly Charged Gaseous Ions from Unfolded Proteins by Electrospray Ionization.* Analytical Chemistry, 2012. **84**(15): p. 6798-6804.
49. Tieleman, D.P., D. van der Spoel, and H.J.C. Berendsen, *Molecular Dynamics Simulations of Dodecylphosphocholine Micelles at Three Different Aggregate Sizes: Micellar Structure and Chain Relaxation.* The Journal of Physical Chemistry B, 2000. **104**(27): p. 6380-6388.
50. Abel, S., F.-Y. Dupradeau, and M. Marchi, *Molecular Dynamics Simulations of a Characteristic DPC Micelle in Water.* Journal of Chemical Theory and Computation, 2012. **8**(11): p. 4610-4623.
51. Marginean, I., P. Nemes, and A. Vertes, *Order-chaos-order transitions in electrosprays: The electrified dripping faucet.* Physical Review Letters, 2006. **97**(6): p. 4.
52. Zakharov, V.V., E.N. Brodskaya, and A. Laaksonen, *Surface properties of water clusters: a molecular dynamics study.* Molecular Physics, 1998. **95**(2): p. 203-209.
53. Kohtani, M., et al., *Proton transfer-induced conformational changes and melting in designed peptides in the gas phase.* Journal of the American Chemical Society, 2006. **128**(22): p. 7193-7197.

Gas-Phase Protein Structure Characterization with Ion Mobility Mass Spectrometry and Molecular Dynamics Simulations

Samaneh Ghassabi Kondalaji

Dissertation submitted to the Eberly College of Arts and Sciences
at West Virginia University in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy in Chemistry

C. Eugene Bennett Department of Chemistry

APPROVAL OF THE EXAMINING COMMITTEE

Stephen J. Valentine, Ph.D.,
Chair

Peter M. Gannett, Ph.D.

John B. Mertz, Ph.D.

Glen P. Jackson, Ph.D.

Lisa A. Holland, Ph.D.

7-7-17
Date