Electronic Thesis and Dissertation Repository

1-21-2019 3:00 PM

Protein Conformations in the Gas Phase Probed by Mass Spectrometry and Molecular Dynamics Simulations

Maryam Bakhtiari
The University of Western Ontario

Supervisor
Konermann, Lars
The University of Western Ontario

Graduate Program in Chemistry
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science
© Maryam Bakhtiari 2019

Abstract

Mass Spectrometry (MS) has been revolutionized by the ability to produce intact gaseous protein ions by electrospray ionization (ESI). The question to what extent these ions retain solution-like conformations under "native" ESI conditions remains a matter of debate. Traditional high-resolution structure determination techniques only report on proteins in the condensed phase. For this reason, MD simulations play an important role in exploring the behavior of gas phase proteins. In this research, mobile and non-mobile proton MD simulations along with mass spectrometry data at 300 K in both positive and negative ion mode indicated that native-like conformations were largely retained. Surface titratable side chains were found to adopt orientations that were less extended than in crystals and in solution (with the radius of gyration [R_g] values 3-5% lower than for the X-ray coordinates), causing the gaseous protein to be somewhat more compact than in the condensed phase. Calculated collision cross sections of these MD structures were in good agreement with experimental data.

Keywords: mass spectrometry (MS), electrospray ionization (ESI), molecular dynamics (MD) simulations, ion mobility mass spectrometry (IM-MS), gas phase simulation, mobile proton algorithm

Acknowledgments

Table of Contents

List of Symbols and Abbreviations

CCS – collision cross section

CEM – chain ejection model

CHARMM – chemistry at Harvard molecular mechanics

CRM – charged residue model

Da – Dalton

DC – direct current

DFT – density functional theory

ESI – electrospray ionization

IEM – ion ejection model

IM-MS – ion mobility mass spectrometry

MALDI – matrix-assisted laser desorption/ionization

MD – molecular dynamics

MS – mass spectrometry

NMR – nuclear magnetic resonance

OPLS/AA – optimized potential for liquid Simulation – all-atom

PDB – protein data bank

RF – radio frequency

TOF – time-of-flight

TWIMS – traveling-wave ion mobility spectrometry

Chapter 1

1. Introduction

1.1. Proteins

Biopolymers can be divided into three sub-groups: polynucleotides, polysaccharides, and peptides/proteins. Proteins participate in every process that takes place in a cell. For example, they play a key role in cellular defense, transport, signaling, and as enzymes that catalyze all steps of the cellular metabolism. Proteins are linear chains of amino acids (usually these comprise the twenty "canonical" amino acids). These amino acids can combine differently to form antibodies, transporters, enzymes, and so forth. All of these residues are α-amino acids, which means a carboxylic acid group and an amino group are connected to the same carbon atom (α-carbon). The difference among amino acids comes from the various side chains (-R groups) which differ in charge, structure, size and their solubility in water. The α-carbon in 19 amino acids (except glycine) is connected to four different functional groups, giving rise to two different stereoisomers called D and L (Figure 1.1. (A)). All amino acids found in proteins are L-stereoisomers. Amino acids are linked to one another via peptide bonds (-CO-NH-). The linear sequence of amino acids represents the primary structure or the backbone of the protein (Figure 1.1. (B)). Hydrogen bonds between backbone CO and NH groups give rise to secondary structure such as turns, α-helices, and β-sheets. The final three-dimentional arrangement of the atoms in a protein represents the tertiary structure. While secondary structure, is often mediated by local interactions (particularly the backbone H-bonds in α-helices), tertiary structure tends to involve longer range contacts. Quaternary structure refers to the three-dimensional arrangement of proteins that have two or more subunits.
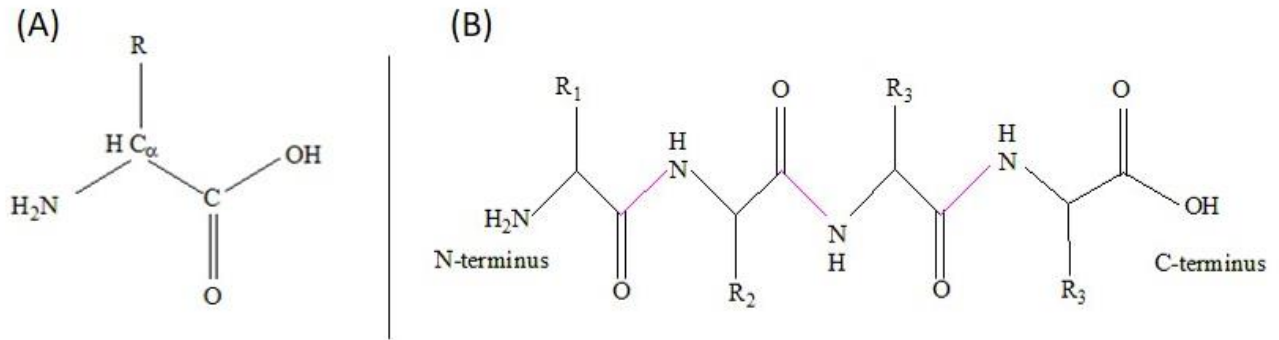
Figure 1.1. (A) Structure of an amino acid, **(B)** A polypeptide.

The fully folded, biologically active structure of a protein is called the "native" state. This native structure contains many intra-chain contacts, and it corresponds to the lowest free energy conformation that is kinetically accessible during folding. Based on their polarity, size, and shape, amino acids can be divided into different groups. Amino acids with basic side chains (lysine, arginine) and those that have acidic side chains (aspartic acid and glutamic acid) are potential charge carriers. These residues have a net charge 1+ or 1-, respectively, at pH 7. Histidine is weakly basic, i.e., it tends to be protonated only at pH < 7. Proteins can be classified according to their overall structure. "Globular proteins" are folded into a spherical shape. "Fibrous proteins" are made of long strands or sheets. Their tertiary structure is pretty simple and includes large secondary structures. They provide support and protection and assist to form the shape of vertebrates, while globular proteins have a variety of secondary structures and participate in enzymes and regulators' structures. "Membrane proteins" are another broad group of proteins which have an amphiphilic nature and accordingly a poor solubility in water. Based on some statistics, they include about 33.3 percent of proteins. A number of membrane proteins are covalently attached to lipids to hold the proteins to the membranes. "Intrinsically disordered proteins (IDPs)" are the last group of proteins which lack a stable structure even under native conditions. The focus of this work is on globular proteins, many of which are commonly found in nature.

The compact structure of globular protein is stabilized by various types of weak interactions (see below). As an example of a typical globular protein, myoglobin is so tightly packed that its interior can only accommodate four water molecules. Due to hydrophobic effect, a dense core is formed at the center of globular proteins, while titratable (acidic/basic) side chains and polar residues are in contact with the solvent.[1-4]

1.2. Noncovalent Interactions

Secondary, tertiary, and quaternary structure, are mediated by noncovalent interactions. These contacts are weaker than covalent bonds and can classified as follows:

1. Hydrogen bonds. A hydrogen donor group such as OH or NH and a hydrogen acceptor with a lone pair of electrons such as: NH_2-R are needed to establish a hydrogen bond. For example, backbone NH hydrogens (donor) and backbone C=O groups give rise to the secondary structure.

2. Hydrophobic interactions are formed among nonpolar residues which tend to avoid water; as a result they come together and form the "hydrophobic core" in the protein. These interactions along with other forces give rise to the tertiary structure.

3. In ionic (electrostatic) interactions positive and negative charges interact with each other through Coulombic contacts (see details below). The hydrophobic effect and electrostatic interactions are the two main forces which contribute the most to the stabilization of the native structure.

4. Van der Waals interactions (including London dispersion forces), are defined as dipole-mediated attractive contacts involving two uncharged moieties that are in close contact with one another.[2-3]

1.3. Electrostatic Interactions and the Formation of Salt Bridges

With the exception of the hydrophobic effect, all of intramolecular and intermolecular contacts are among amino acids are governed by "electrostatic interactions", as expressed in Coulomb's Law:

$$E = \frac{q_1 q_2}{4\pi\varepsilon_\circ \varepsilon r} \tag{1.1}$$

q_1 and q_2 are charges that are separated by the distance r, ε_o refers to the permittivity of vacuum, and ε is the dielectric constant of the environment.[2] Electrostatic contacts represent a long-range interaction, and they play an important role to the protein's structure and free energy.[5-6] Salt bridges (also referred to as ion pairs) are important for recognition, thermostability, and allosteric regulation. Aspartic acid (Asp), glutamic acid (Glu), and C-terminus are negatively charged, while lysine (Lys), arginine (Arg), and the N-terminus represent the positively charged in salt bridges.[7-10] Typical salt bridges have heavy atom distances of around < 4 Å. In addition to the electrostatic contribution, salt

bridges in proteins are usually reinforced by hydrogen bonds (e.g. in the case of R-NH3$^+$ $^-$OOC-R).[7-9, 11]

1.4. Methods for Studying Protein Structure and Conformation

Numerous techniques can be used to study the structure and function of proteins. All of them have their unique strengths and weaknesses.

1.4.1. X-ray Crystallography

X-ray crystallography reveals the three-dimensional structure from a crystallized sample. Crystal growth starts with very concentrated samples which then undergo slow dehydration. An ordered crystal is formed by alignment of the molecules in three-dimensional space (3D), forming a unit cell. This unit cell is the smallest repeating unit from which the entire crystal is assembled. The adjacent unit cells are being translated along the cell axes, giving rise to crystal planes of atoms. These planes that are formed by the electron clouds that surround individual nuclei. Interaction of these electrons with X-ray photons produces a diffraction pattern that contains information regarding the structure and the size of the unit cell. Fourier transform analysis of these diffraction pattern then reveal the 3D electron density in the crystal, from which the atomic structure of the protein can be determined.[2, 12-13]

For small molecules, crystallography is relatively simple, but the enormous size of protein renders the situation more challenging. The "hanging drop method" is frequently used to grow protein crystals. In this method, a droplet of protein solution is exposed to a buffer gas in a closed container. If the experimental conditions are not conducive to formation of large crystals it may be possible to use seeds to produce large crystals. The crystal is then exposed to an X-ray beam and the diffraction patterns are recorded using a CCD camera or arrays of scintillation counters. Since crystals are often damaged by heating or free radicals (from X-ray exposure), a number of individual crystals are often required to give a complete structure.[2]

X-ray crystallography is still considered the gold standard for atomically resolved structural data. The RCSB Protein Data Bank (PDB) is the largest resource for the three-dimensional biomolecular structural data.[14] Crystallography can also reveal the structures of protein-DNA, protein-ligand, and protein-protein complexes.[15] Other methods can barely compete with X-ray crystallography in these areas.[16]

1.4.2. Nuclear Magnetic Resonance (NMR) Spectroscopy

One of the issues with the X-ray crystallography is that some proteins and complexes might not crystallize at all, or they might not crystallize in their biologically active conformations.[17]

Nuclear magnetic resonance (NMR) spectroscopy has emerged as an alternative to X-ray crystallography as it interrogates proteins in their physiological solution phase conformation. NMR methods can provide information about protein structure and dynamics that cannot be obtained by X-ray crystallography. NMR spectroscopy measures the absorbance of radio frequency (RF) radiation of ^1H, ^{13}C, and ^{15}N that are exposed to a strong magnetic field.[16] Every nucleus can give rise to resonance at a specific frequency. Differences in the surrounding chemical environment result in unique shielding and local magnetic fields, translating into slightly different resonance frequencies. In the case of proteins, the naturally abundant proton (^1H) with the nuclear spin $I = \frac{1}{2}$, ^{13}C, and ^{15}N nuclei are the most commonly used atoms for NMR-based protein structure determinations. An NMR spectrum shows peaks that reflect the different resonances related to each atom in the molecule.[2] The resonance frequencies are reported as "chemical shift" (δ, reported in part per million - ppm) relative to a suitable standard.[2, 16]

Felix Bloch and Edward Mills Purcell were awarded 1952 Nobel Prize in Physics for developing new methods to measure the nuclear magnetic precision for the large scale systems.[18-19] In 2002, Kurt Wüthrich won part of the Nobel Prize in Chemistry for developing NMR spectroscopy for biological systems in solution.[20]

1.4.3. Cryo-Electron Microscopy (Cryo-EM)

Cryo-Electron Microscopy (Cryo-EM) is an imaging technique that can visualize large molecules at low resolution. Although all images represent 2D projections, 3D structural data can be reconstructed from the 2D data if they are recorded from many different angles. Along with X-ray crystallography, cryo-EM can provide structural information of large assemblies. The wavelength range of the electrons used in cryo-EM is 0.0015–0.040 Å. The scattering of these photons is more efficient than in the case of X-ray photons. The biggest concern in cryo-EM is the vacuum environment in the microscope. The samples have to undergo rapid freezing at liquid nitrogen temperature to not to let the solvent to sublimate prior to data acquisition. Cryo-EM can be applied to characterize biological molecules within the cell and it is becoming a main stream technology to study protein assemblies at

a resolution of a few Angstroms. When integrated with X-ray crystallography and NMR, cryo-EM can be used to attain atomically resolved information on the structure and dynamics of biological complexes.[2, 21]

1.5.   Mass Spectrometry (MS)

Mass spectrometry (MS) has revolutionized life sciences. Rapid improvements in this area take place on an ongoing basis, highlighted by the 2002 Chemistry Nobel Prize to John Fenn.[1, 22] MS can decipher protein complexes and interaction networks.[23] MS makes use of a variety of soft ionization techniques that allow the conversion of large biomolecules into gaseous ions without fragmentation.[24] Figure 1.2. represents the schematic layout of a mass spectrometer.[2]
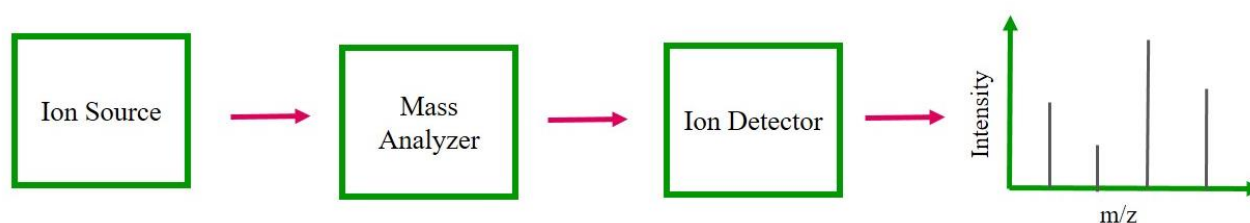
Figure 1.2. Schematics of a mass spectrometer

As a first step in an MS experiment the analyte has to be ionized and transferred into the gas phase. To study proteins, one needs to use soft ionization techniques such as Matrix-Assisted Laser Desorption / Ionization (MALDI) and Electrospray Ionization (ESI). Both of them can convert proteins into gaseous ions without fragmentation This is in contrast to traditional techniques such as Electron Impact (EI), where covalent bonds are broken. In MALDI, a protein is embedded into a solid matrix and exposure to a laser pulse culminates in the formation of gaseous ions. In ESI, solution phase analytes turn into gas phase ions.[25] Electrospray ionization (ESI) is the most commonly used soft ionization technique that can be applied even to proteins with very large molecular weights.[2] In this thesis, ESI-MS is used extensively, which is why it will be discussed in some more detail below.

1.5.1. Electrospray Ionization (ESI)

ESI-MS was introduced in the 1980s by Fenn et al.[22] ESI can ionize inorganic ions, polymers and proteins, all the way to large noncovalent complexes.[26] ESI takes place at atmospheric pressure. The process starts with analyte solution that is infused into a narrow metal capillary (~100 μm) to which an electric potential of several kV has been applied. Usually, infusion rates are in the range of 1-100 μL min[-1]. At the capillary tip, solution is distorted into a Taylor cone which ejects a plume of droplets. Excess charge on these droplets is due to the presence of cations, including protons that are produced inside the capillary because of charge balancing reactions such as $2H_2O \rightarrow O_2 + 4e^- + 4H^+$. In other words, the ESI source acts as an electrochemical cell. The charge density on the shrinking droplets increases until surface tension γ is balanced by Columbic repulsion. At this so-called Rayleigh limit the number of charges z_R is given by[25]:

$$z_R = \frac{8\pi}{e}\sqrt{\varepsilon_0 \gamma r^3} \tag{1.2}$$

where r is the droplet radius, ε_0 is the vacuum permittivity as mentioned earlier, γ is the surface tension and e is the elementary charge. By repeating evaporation and fission, ESI droplets with the radius of a few nanometers are ultimately generated. Gaseous analyte ions are released from these nanodroplets, and the ions are directed via a sampling orifice towards the mass analyzer where they will be counted and detected. Figure 1.3. represents the schematics of ESI source in positive ion mode.[25]
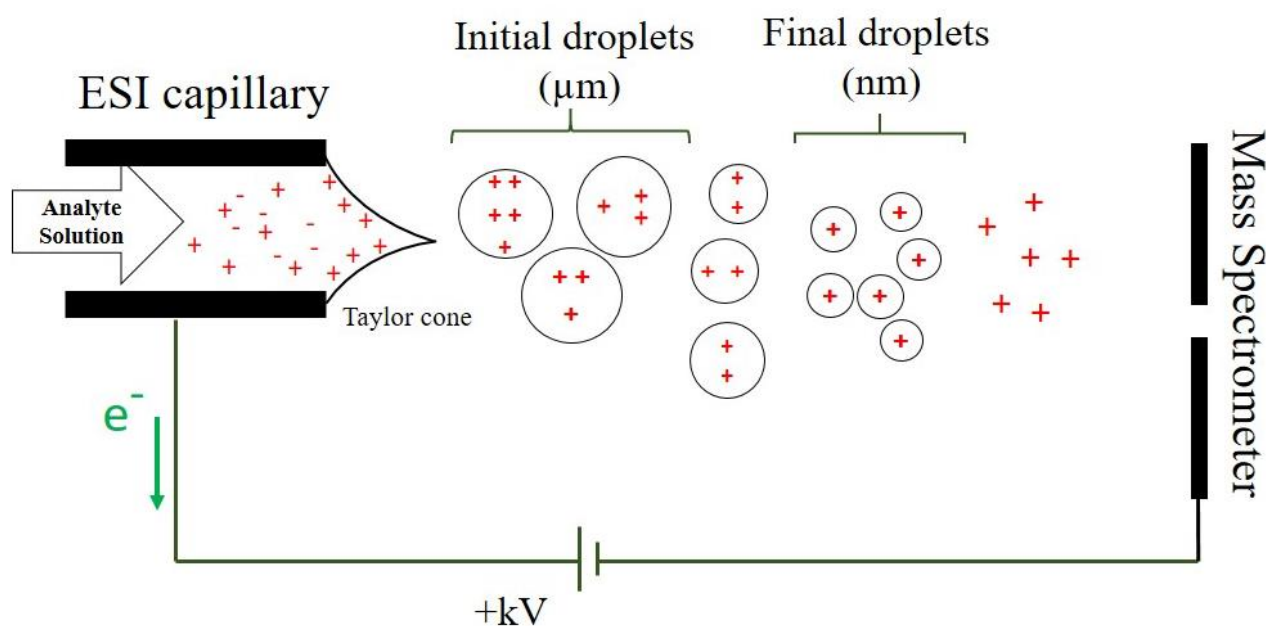
Figure 1.3. ESI source in positive ion mode

There are different mechanisms to be considered for the formation of gaseous ions from nanodroplets. Rayleigh-charged nanodroplets are surrounded by a radial electric field that is high enough to cause small solvated ions to eject from the droplet surface; this is probably how small ions such as gaseous Na^+ are formed. This mechanism is called ion evaporation model (IEM). In the charge residue model (CRM) firstly the solvent evaporates and then compact analytes (such as folded globular proteins) accept the remaining droplet charge at the point of analyte release into the gas phase. The chain ejection model (CEM) applies to partially nonpolar polymer chains such as unfolded proteins. In this case it is envisioned that the unfolded chains are extruded from the droplet surface in a stepwise sequential manner. Ions produced in positive ion mode are multiply charged, i.e., they have the composition $[M + zH]^{z+}$. Correspondingly, $[M - zH]^{z-}$ ions are produced in negative ion mode. The three mechanisms are being shown in Figure 1.4.[25].
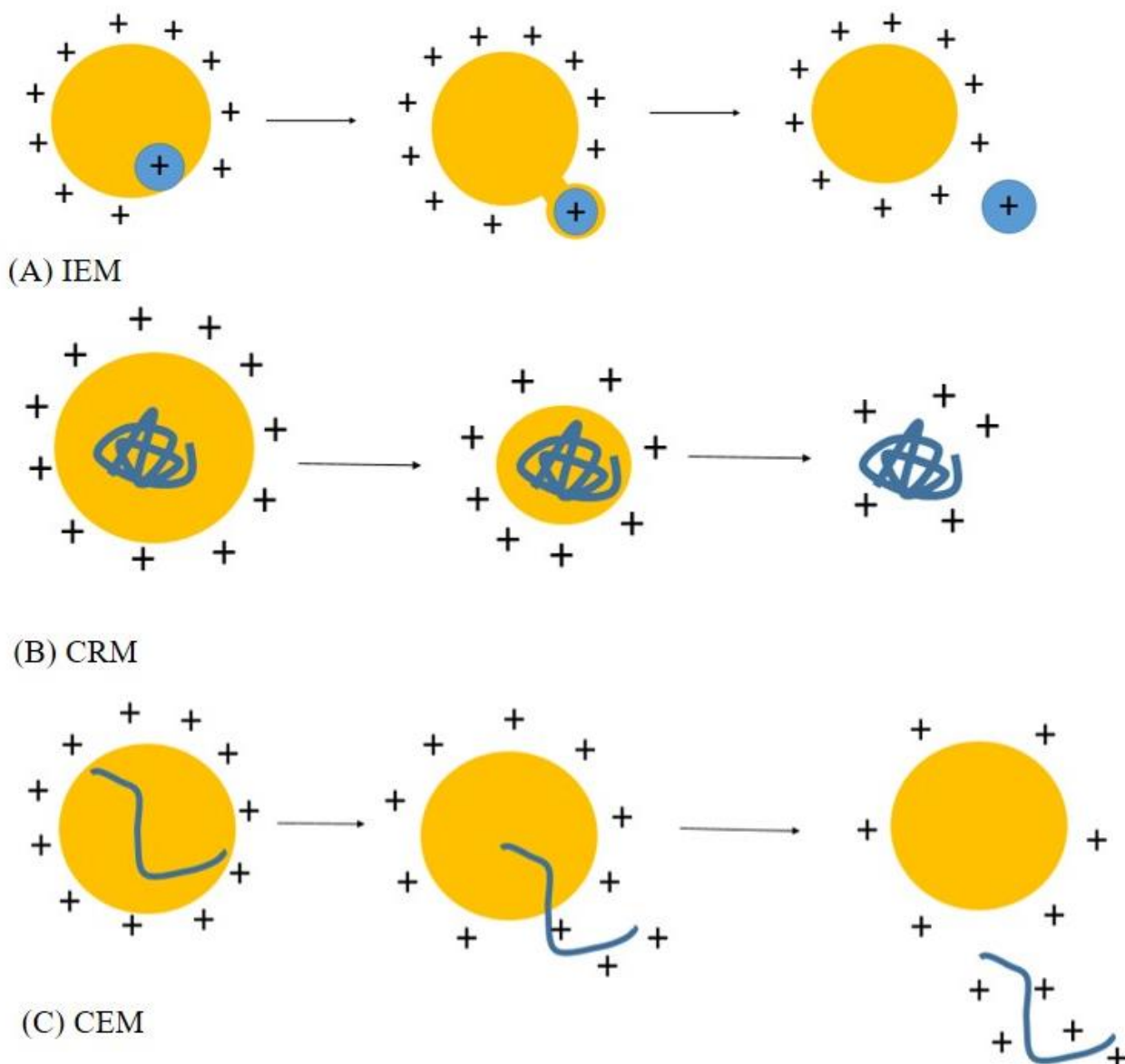
Figure 1.4. Suggested mechanisms for the formation of nanodroplets in ESI-MS; (A) IEM, (B) CRM, and (C) CEM.

1.5.2. Mass Analyzers

To separate gaseous ions based on their m/z, a mass analyzer is required. Various types of mass analyzers are available; these include Orbitraps, Fourier transform ion cyclotron (FT-ICR), quadrupoles, ion traps, and time-of-flight (TOF) analyzers. Since TOF is used in this work as a mass analyzer, this particular technique will be discussed in some more detail.

1.5.2.1 Time-of-Flight (TOF) Mass Analyzers

TOF mass analyzer was established over fifty years ago called by the name of time dispersion mass spectrometer" and "ion velocitron". The components of a TOF include a 1-2 meter long field-free "flight tube"; ions with different masses are accelerated by an electric pusher pulse into this tube before they reach the detector. The ions gain the same amount of potential energy. If we assume there are two ions m_1 and m_2 with the same charge z produced at the same time, but the mass arrangement of $m_1 < m_2$, m_1 will reach the detector earlier because it is lighter than m_2. A TOF mass analyzer determines the m/z of an ion. The ions' potential energy produced by the pusher is converted to kinetic energy (equation 1.3):

$$E_{pot} = E_{kin} \tag{1.3}$$

$$zeU = \frac{1}{2}mv^2 \tag{1.4}$$

where U is the voltage pulse, m is the mass of the ion, v is the velocity of the ion, and z stands for the charge state of the ion. Since $v = \frac{d}{t}$, equation 1.4 can be represented as:

$$zeU = \frac{1}{2}m.\left(\frac{l}{t_d}\right)^2 \tag{1.5}$$

$$t_d = \sqrt{\frac{l^2}{2eUz.m^{-1}}} \tag{1.6}$$

where l is the length of the flight tube, t is the time the ions spent to reach the detector, and $\frac{l}{\sqrt{2eU}}$ is constant. Thus, equation 1.6 can be rewritten as:

$$t_d = l.\left(\frac{m}{z}.\frac{1}{2eU}\right)^{1/2} \tag{1.7}$$

As it is seen from equation 1.7, the flight time of an ion only depends on its *m/z*. Modern TOFs also employ an electrostatic mirror (reflectron) that produces a V- or W-shaped ion trajectory and reflects the ions toward the detector. If two ions have identical mass and charge, but different velocities, the faster ion will penetrate deeper in the reflectron and it will reach the detector later. In this way the reflectron increases the resolution of the TOF mass analyzer.[2, 27] Figure 1.5. represents the schematics of the mass spectrometer used in this work.
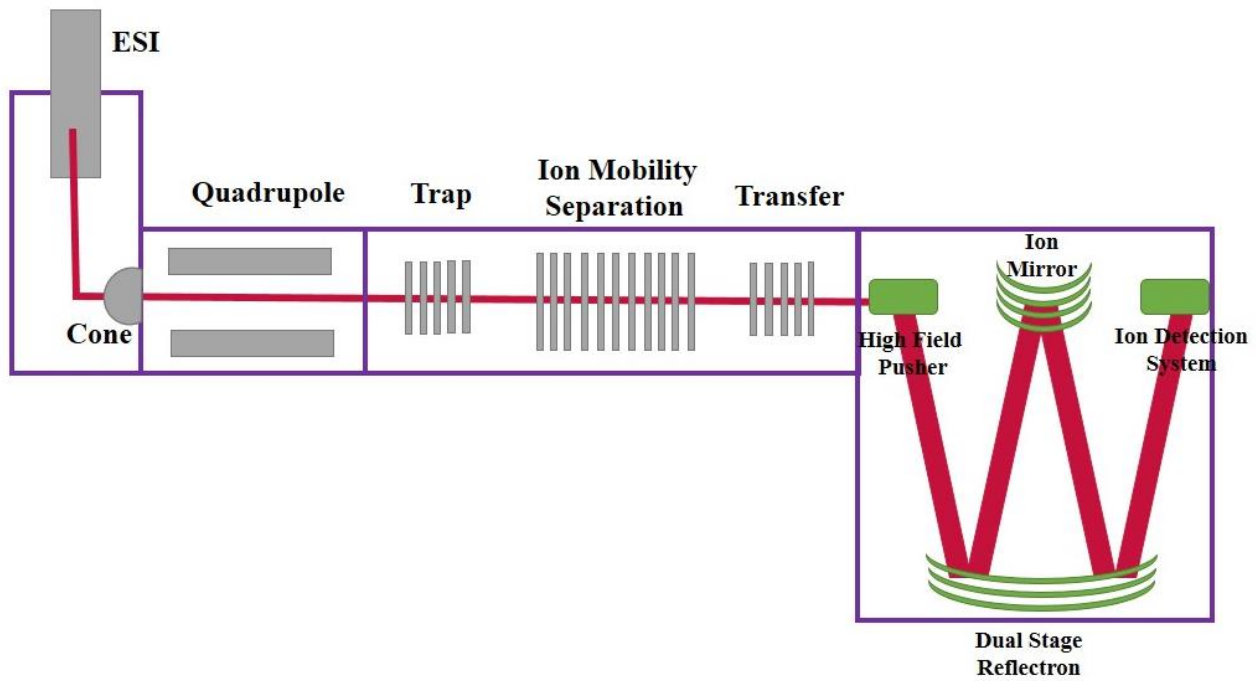
Figure 1.5. Schematic layout of mass spectrometer used in this work (ESI-MS, Q-TOF produced by WATERS). The main elements of Q-TOF are T-Wave ion guide, quadrupole, Tri-wave, and the time-of-flight detector.

1.5.3. Ion Mobility-Mass Spectrometry (IM-MS)

One way to explore the structures of gaseous proteins is ion mobility-mass spectrometry (IM-MS).[28] In IM-MS, multiply charged ions that are generated during electrospray become separated as a result of their different mobilities. Compact ions have larger mobilities than extended ones. In traditional IM-MS ions travel through a drift tube under the influence of a low electric field and in the presence of a buffer gas that provides friction.[29-30] The arrival time is a combination of the time the ions stay in the drift tube (i.e., the drift time t_d) and the duration required for the ion to pass other sections of the instrument before arriving to the detector. The mobility values can be used to derive gas-phase collision cross section (CCS) by the aid of Mason-Schamp equation 1.5:[31-32]

$$\Omega = \frac{3z_e}{16N}\left[\frac{2\pi}{\mu k_B T}\right]^{0.5}\frac{1}{K_0} \tag{1.5}$$

N is the background gas number density, z_e defines the ionic charge, μ is the reduced mass of the ion–neutral pair, k_B is Boltzmann's constant, T stands for the gas temperature. K_0 is the reduced mobility at standard conditions that is defined by:

$$K_0 = \frac{L}{t_d E}\frac{P}{760}\frac{273.2}{T} \tag{1.6}$$

where P is the pressure of the buffer gas, L defines the length of the drift tube, and E is the electric field.[33]

The drift time t_d will be shorter for small ions and for ions that carry more charges. Collision cross section (CCS) is proportional to t_d (as explained in more detail below).

1.5.3.1. Travelling Wave Ion Mobility-Mass Spectrometry (TWIM-MS)

The preceding considerations refer to traditional drift tube IM-MS. However, many commercial instruments (including the one used for the current work) employ travelling wave voltages with nitrogen as buffer gas. A TWIMS device contains a series of ring electrodes that constitute a traveling wave ion guide (TWIGs). Opposite radio frequency (RF) potentials are applied to the electrodes to prepare a radial potential barrier which traps the ions in the radial direction. In addition, "waves" of direct current voltage (DC) are superimposed onto the RF voltage, in order to propel the ions through the TWIMS device in the axial direction. The DC voltage maxima jump from ring to ring, thereby "peristaltically" pushing the ions along. Ions with large CCSs experience occasional "rollover" events, i.e., they will fall behind the wave crest, and then get swept along with the subsequent crest. In this way, ions with different mobilities migrate through the TWIMS at different velocities which enables their separation[34]. Figure 1.6. shows the schematics of a TWIMS separator[35-36].

Empirically, it has been found that t_d in TWIM-MS can be described as[35-36]:

$$\Omega = z \cdot F \cdot t_d{}^B \tag{1.7}$$

where F and B are constant that must be determined experimentally.
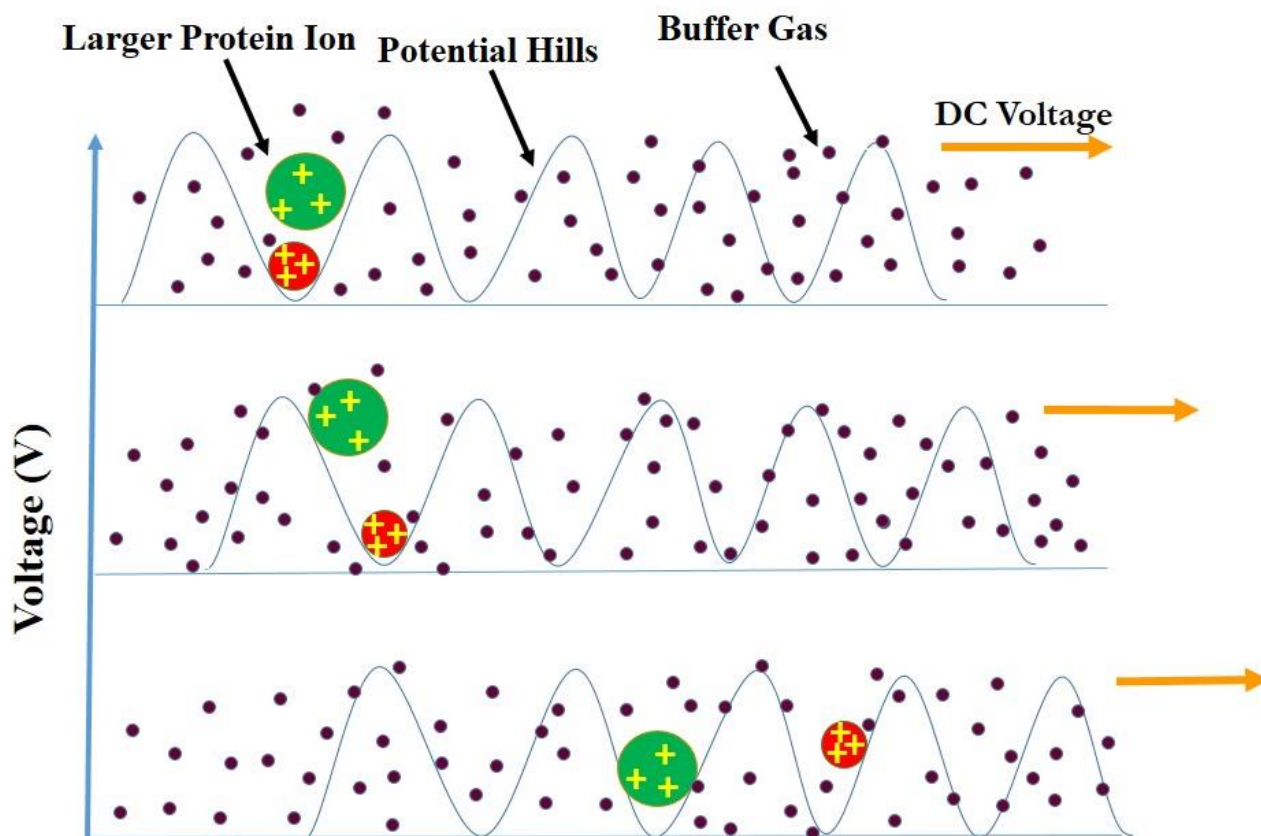
Figure 1.6. Travelling Wave IMS Separator which is formed by RF and DC voltage

1.5.4. Collision cross Section (CCS)

The CCS is related to the "size" of the analyte ion and it is one of the parameters that determines t_d.[28,] [33] Electrostatic repulsion among excess positive charges in a multiply charged protein ion can trigger unfolding. As a result, the CCS of a protein often increases with the number of charges.[37] The particular drift gas used can also affect the CCS (e.g., CCS(He) < CCS(N_2)).[38]

Experimentally, the CCS can be determined from t_d. It is also possible to estimate CCS values on the basis of protein X-ray structures or computational models. Jarrold and Shvartsburg developed MOBCAL which is widely used to calculate the theoretical protein CCSs.[39-42] The three CCS calculation methods available are referred to as projection approximation (PA)[30, 43], exact hard sphere scattering (EHSS)[30, 40, 44], and trajectory method (TM).[30, 41] The TM represents the most reliable approach; it considers long-range interactions and adjacent collisions between the buffer gas and the ion.[30] For the EHSS model, the atoms in an ion are considered as hard spheres[45] and the momentum transfer cross section related to scattering angles among the gas atom that arrives and the one that

leaves is averaged. This model ignores the effects of long-rage potential between the buffer gas and the ion but considers scattering processes. For PA calculations the atoms are also considered as hard spheres[45] and the geometric projection areas over all directions are being averaged; however, this method disregards the long-range interactions and the scattering between ion and the buffer gas.[30]

In 2017, a TM-based CCS calculator called Collidoscope was introduced by Prell's group.[46] This approach takes advantage of parallel processing and optimized trajectory sampling and allows for both He and N_2 as the collision gas. It also contains a charge placement algorithm to define the foreseeable charge site configurations for protonated protein ions which exists in an input geometry in pdb file format. Collidoscope calculates the ion's CCS by modeling the orientationally-averaged momentum transfer term. This method is much faster than the TM implementation in MOBCAL, and therefore Collidoscope has been adopted by numerous research groups (including the current work, all the CCS values in this work are measured by this method).

1.5.5. Structure of Gaseous Protein Ions Produced by Native ESI-MS

"Native" MS refers to ESI experiments conducted under non-denaturing conditions and where collisional heating is kept at a minimum.[10, 23, 26, 47] Although the term native MS was first coined in 2004[48], it describes approaches that were introduced much earlier.[23, 49-51] In native IM-MS, mobility values are measured in a drift tube, traveling wave ion guide, or trapped ion mobility instrumentation in the low-field limit.[46] This approach has been applied to both protein-protein and protein-ligand complexes.[23, 52-56]

Heck defines this term "native ESI-MS" by referring to conditions under which protein complexes retain their quaternary structure.[57] Also, Native ESI-MS tries to preserve the solution (tertiary) structure during protein transfer into the gas phase.[23] Native ESI-MS has become a key method for studying protein-protein and protein-ligand interactions and it is turning into a powerful structural biology technique.[58-60] Native ESI-MS can disclose protein binding stoichiometry, composition, dynamics, and their arrangement of the subunits. The technique only requires very small amounts of sample.[58]

Despite the undeniable successes of native ESI-MS, it still remains controversial to what extent electrosprayed proteins and protein complexes retain a solution-like conformation in the gas phase. Breuker et al. suggested that globular proteins such as cytochrome c and ubiquitin undergo a temporal structural evolution after ESI that includes side-chain collapse, protein unfolding, followed by refolding into multiple gaseous structures in 10^{-12} to 10^2 s.[61-62] On the other hand, some studies

have shown that ubiquitin ions formed by ESI are tightly folded and have collision cross sections similar to the native state,[26] supporting the views of Heck and Robinson that gas-phase ions retain solution phase structure. [63-64] Wyttenbach believes that ubiquitin preserves solution-like structure in the gas phase during ESI; using ESI-IM-MS, they studied ubiquitin under native conditions in different charge states. The results show that the native state of ubiquitin 7+ survives under thermal (300 K) conditions for more than 100 ms after the solvent completely leaves the multiply charged ion. Sharp peaks around CCSs of 1000 $Å^2$ were observed in IM-MS experiments for z = |6 – 8|. These results are also close to the CCS obtained by molecular dynamics simulations through EHSS approach; reporting 1060 $Å^2$ for z = |6 – 8| which they call the "native state charge states".[26]

Recently, a paper published by Julian's lab[6] discussed the possibility of preserving native structures in the gas phase by salt bridges. Julian and co-workers developed a photoelectron transfer dissociation (PETD) technique that can be applied to detect zwitterionic pairs in peptides and proteins. Using a modified linear quadrupole ion trap (to accept laser light from a 266 nm Nd:YAG laser) and Orbitrap mass spectrometer, *ab initio* calculation, and MD simulations, they studied ubiquitin and a number of peptides. The results of this research disclose that salt bridges (or the zwitterionic pairs as discussed earlier) are frequently found in large gaseous peptides and proteins, and the number of these salt bridges increases with protein size.[6]

1.6.   Molecular Dynamics (MD) Simulations

Molecular dynamics (MD) simulations[65] can provide insights into protein motions and structures from an atomistic perspective.[66] This approach is particularly important for gas phase investigations, because traditional high-resolution structure determination methods (X-ray crystallography, NMR spectroscopy, cryo-EM) only report on protein behavior in the condensed state. MD simulations rely on iterative integrations of Newton's Equations[25].

Newton's Second Law can be expressed as:[67-68]

$$m_i \cdot a(r_i) = F_i \quad , i = 1, 2, 3, \dots . N \tag{1.8}$$

$$m_i \frac{\partial^2 r_i}{\partial t^2} = F_i = -\frac{dV(r_i)}{dr_i} \tag{1.9}$$

F_i stands for the force acting on atom i, and the force F_i acting on this atom depends on the atomic position r_i. $V(r_i)$ represents the potential energy of atom i, arising from interactions of this atom with all other atoms in the system.

For any time point ($t_0 + \Delta t$) the velocity and position of every atom can be determined as:

$$v_i(t_0 + \Delta t) = v_i(t_0) + a(r_i) \cdot \Delta t \tag{1.10}$$

$$r_i(t_0 + \Delta t) = r_i(t_0) + [v_i(t_0) \cdot \Delta t] + \left[\frac{1}{2} a(r_i) \cdot (\Delta t)^2\right] \tag{1.11}$$

Δt is a time increment that can be on the order of 1 *fs* or 2 *fs*. The trajectory of the system is generated repeating these integration steps many thousands of times, and by plotting the coordinates r_i *(t)* as a function of time.

The most popular programs to perform MD simulations are AMBER and GROMACS. GROMACS is an open source software which includes most of the AMBER and CHARMM force fields.[65] However, all of these algorithms were designed for simulating the behavior of biological systems in solution (usually water). The application of these methods to gaseous protein ions requires additional considerations.

1.6.1. Limitations of Classical MD Simulations; Mobile Charge Carriers in the Gas Phase

MD simulations have become an essential tool for probing the behavior of large electrosprayed ions in the gas phase, considering that existing high resolution structure determination tools only work in the condensed phase.[26, 47] However, it is a concern that in traditional MD simulations the formation or dissociation of covalent bonds is not possible, implying that charges (arising from protonation or deprotonation of titratable sites) are static.[47] For solution phase simulations all Asp, Glu, C-termini (CT) are negatively charged, while Arg, Lys, and N-termini (NT) carry positive charges, as governed by their pK_a values, His can be treated as a neutral side chain, but it can also act as a protonated residue[3]). For gas phase simulations on $[M + zH]^{z+}$ ions it is often assumed that all carboxylates are neutral (protonated) and only Arg, Lys, His, and NT will carry the charges.[69-71]

Both of the aforementioned concerns have to be addressed when it comes to modelling the behavior of the electrosprayed protein ions. It is well known that in the gas phase protons are highly mobile.[72-75] Also, experiments have strong evidence that R-COO$^-$ are abundant in gaseous $[M + zH]^{z+}$ ions, giving rise to zwitterionic motifs and salt bridges.[6, 74, 76-80] A number of computational methods can deal with protons as mobile charge carriers, including quantum mechanics / molecular mechanics[74], *ab initio* MD[81] and density functional theory/MD.[82] Unfortunately, the staggering computational costs prevents the application of these methods to large proteins.[10] Another concern is there is no clear path to choose the most appropriate spatial protonation pattern.[47]

Thachuk et al.[75] were the first to propose a mobile proton model for gas-phase MD simulations which describes proton hopping during the dissociation of collisionally heated protein complexes, but they only considered protonation of basic sites, which does not allow salt bridges to form. They also employed a coarse-grained force field which uses simplified representations of molecular structure. Such force fields may be suitable for bulk solution simulations[83] but not for the gas phase.[10]

Konermann et al.[10, 84] developed an all-atom mobile proton algorithm. The total energy of a particular proton distribution in a gaseous protein is given as the sum of Coulombic energy V_{Coul} and a proton affinity term E_{PAint}:

$$E_{tot} = V_{Coul} + E_{PA_{int}} \tag{1.12}$$

The Coulombic potential energy V_{Coul} includes the contributions from all n atoms in the system. The proton affinity term is the sum of proton affinities over all the titratable sites (*PAint (x) > 0, x = 1,..., N*).

$$E_{PA_{int}} = -\sum_{x=1}^{N} PA_{int}(x) \cdot \delta_x \tag{1.13}$$

where $\delta_x = 1$ and $\delta_x = 0$ for protonated and deprotonated sites, respectively.

This algortithm allows the implementation of gas phase MD simulations that overcome the limitations of (i) static charges and (ii) the *a priori* exclusion of salt bridges. The simulations are conducted in a way that short (~1 ns or less) MD run segments with static charges alternate with proton redistribution steps, where protons are reshuffled within the protein such that the energy term 1.12 is minimized. As a result, this strategy describes the behavior of gaseous proteins with quasi-mobile protons.

1.7. Scope of This Research

My M.Sc. research focused on probing the behavior of ubiquitin and lysozyme (1UBQ and 1AKI in Protein Data Bank, respectively). I investigated the conformations of these proteins in the gas phase under "native" ESI conditions. Using MS and IM-MS in both positive and negative ion modes, experimental collision cross sections and the most abundant charges of native proteins in the gas phase were determined. The key components of this work are MD simulations using GROMACS and an in-house software for mobile protein simulations.[10] The ultimate goal of this work is to obtain a detailed understanding of protein structural changes that occur in the solvent-free environment within the mass spectrometer. To this end, we compared the structures of gaseous protein ions with data

acquired in the crystal state and in solution. Overall, we found the structures of gaseous proteins to be surprisingly resilient to unfolding, consistent with the proposal that native-like protein structures correspond to minima on the free energy landscape, and that proteins get trapped in these minima during ESI.[85]

1.8   References

1.   Wu, Y.; Engen, J. R., What mass spectrometry can reveal about protein function. *Analyst* **2004,** *129* (4), 290-296.

2.   Kaltashov, I. A.; Eyles, S. J.; Desiderio, D. M., *Mass Spectrometry in Structural Biology and Biophysics: Architecture, Dynamics, and Interaction of Biomolecules*. Wiley: 2012.

3.   Nelson, D. L.; Cox, M. M.; Lehninger, A. L., *Lehninger principles of biochemistry*. W.H. Freeman: New York, 2013.

4.   Ibba, M.; Söll, D., Genetic Code: Introducing Pyrrolysine. *Curr. Biol.* **2002,** *12* (13), R464-R466.

5.   Compiani, M.; Capriotti, E., Computational and Theoretical Methods for Protein Folding. *Biochemistry* **2013,** *52* (48), 8601-8624.

6.   Bonner, J.; Lyon, Y. A.; Nellessen, C.; Julian, R. R., Photoelectron Transfer Dissociation Reveals Surprising Favorability of Zwitterionic States in Large Gaseous Peptides and Proteins. *J. Am. Chem. Soc.* **2017,** *139* (30), 10286-10293.

7.   Kumar, S.; Nussinov, R., Salt bridge stability in monomeric proteins11Edited by J. M. Thornton. *J. Mol. Biol.* **1999,** *293* (5), 1241-1255.

8.   Kumar, S.; Nussinov, R., Relationship between ion pair geometries and electrostatic strengths in proteins. *Biophys. J.* **2002,** *83* (3), 1595-1612.

9.   Bosshard, H. R.; Marti, D. N.; Jelesarov, I., Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. *J. Mol. Recognit.* **2004,** *17* (1), 1-16.

10.   Konermann, L., Molecular Dynamics Simulations on Gas-Phase Proteins with Mobile Protons: Inclusion of All-Atom Charge Solvation. *J. Phys. Chem. B* **2017,** *121* (34), 8102-8112.

11.   Lee, C.-W.; Wang, H.-J.; Hwang, J.-K.; Tseng, C.-P., Protein Thermal Stability Enhancement by Designing Salt Bridges: A Combined Computational and Experimental Study. *PLoS One* **2014,** *9* (11), e112751.

12.   Smyth, M. S.; Martin, J. H., x ray crystallography. *Mol. Pathol. : MP* **2000,** *53* (1), 8-14.

13.   Thibault, P.; Elser, V., X-Ray Diffraction Microscopy. *Annu. Rev. Condens. Matter Phys.* **2010,** *1* (1), 237-255.

14.   Su, X.-D.; Zhang, H.; Terwilliger, T. C.; Liljas, A.; Xiao, J.; Dong, Y., Protein crystallography from the perspective of technology developments. *Crystallogr. Rev.* **2015,** *21* (1-2), 122-153.

15.   Roemer, S. C.; Donham, D. C.; Sherman, L.; Pon, V. H.; Edwards, D. P.; Churchill, M. E. A., Structure of the progesterone receptor-deoxyribonucleic acid complex: novel interactions required for binding to half-site response elements. *Mol. Endocrinol. (Baltimore, Md.)* **2006,** *20* (12), 3042-3052.

16.   David, W., NMR Spectroscopy and Protein Structure Determination: Applications to Drug Discovery and Development. *Curr. Pharm. Biotechnol.* **2005,** *6* (2), 105-120.

17.   Zuiderweg, E. R. P., Mapping Protein−Protein Interactions in Solution by NMR Spectroscopy. *Biochemistry* **2002,** *41* (1), 1-7.

18.   Bloch, F.; Hansen, W. W.; Packard, M., The Nuclear Induction Experiment. *Phys. Rev.* **1946,** *70* (7-8), 474-485.

19.   Purcell, E. M.; Torrey, H. C.; Pound, R. V., Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Phys. Rev.* **1946,** *69* (1-2), 37-38.

20.   Wüthrich, K., *NMR in Structural Biology*. World Scientific, 1995.

21.   Milne, J. L. S.; Borgnia, M. J.; Bartesaghi, A.; Tran, E. E. H.; Earl, L. A.; Schauder, D. M.; Lengyel, J.; Pierson, J.; Patwardhan, A.; Subramaniam, S., Cryo-electron microscopy – a primer for the non-microscopist. *FEBS J.* **2013,** *280* (1), 28-45.

22.   Fenn, J. B., Electrospray Wings for Molecular Elephants (Nobel Lecture). *Angew. Chem. Int. Ed.* **2003,** *42* (33), 3871-3894.

23.   Leney, A. C.; Heck, A. J. R., Native Mass Spectrometry: What is in the Name? *J. Am. Soc. Mass. Spectrom.* **2017,** *28* (1), 5-13.

24.   Jonsson, A. P., Mass spectrometry for protein and peptide characterisation. *Cell. Mol. Life Sci. CMLS* **2001,** *58* (7), 868-884.

25.   Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S., Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013,** *85* (1), 2-9.

26.   Wyttenbach, T.; Bowers, M. T., Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-Free Ion Mobility–Mass Spectrometry Environment. *J. Phys. Chem. B* **2011,** *115* (42), 12266-12275.

27.   El-Aneed, A.; Cohen, A.; Banoub, J., Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Appl. Spectrosc. Rev.* **2009,** *44* (3), 210-230.

28.   Warnke, S.; von Helden, G.; Pagel, K., Protein Structure in the Gas Phase: The Influence of Side-Chain Microsolvation. *J. Am. Chem. Soc.* **2013,** *135* (4), 1177-1180.

29.   Liu, Y.; Clemmer, D. E., Characterizing Oligosaccharides Using Injected-Ion Mobility/Mass Spectrometry. *Anal. Chem.* **1997,** *69* (13), 2504-2509.

30.   Jurneczko, E.; Barran, P. E., How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase. *Analyst* **2011,** *136* (1), 20-28.

31.   Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V., Ion mobility–mass spectrometry analysis of large protein complexes. *Nat. Protoc.* **2008,** *3*, 1139.

32.   Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T., Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Anal. Chem.* **2010,** *82* (22), 9557-9565.

33.   Dugourd, P.; Hudgins, R. R.; Clemmer, D. E.; Jarrold, M. F., High-resolution ion mobility measurements. *Rev. Sci. Instrum.* **1997,** *68* (2), 1122-1129.

34.   Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H., Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun. Mass Spectrom.* **2004,** *18* (20), 2401-2414.

35.   Shvartsburg, A. A.; Smith, R. D., Fundamentals of Traveling Wave Ion Mobility Spectrometry. *Anal. Chem.* **2008,** *80* (24), 9689-9699.

36.   May, J. C.; McLean, J. A., Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Anal. Chem.* **2015,** *87* (3), 1422-1436.

37.   Sundarapandian, S.; May, J. C.; McLean, J. A., Dual Source Ion Mobility-Mass Spectrometer for Direct Comparison of Electrospray Ionization and MALDI Collision Cross Section Measurements. *Anal. Chem.* **2010,** *82* (8), 3247-3254.

38.   May, J. C.; Morris, C. B.; McLean, J. A., Ion Mobility Collision Cross Section Compendium. *Anal. Chem.* **2017,** *89* (2), 1032-1044.

39.   Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F., Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem.* **1996,** *100* (40), 16082-16086.

40.   Shvartsburg, A. A.; Jarrold, M. F., An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **1996,** *261* (1), 86-91.

41.   Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F., Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem. A* **1997,** *101* (5), 968-968.

42.   Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F., Protein Structure in Vacuo: Gas-Phase Conformations of BPTI and Cytochrome c. *J. Am. Chem. Soc.* **1997,** *119* (9), 2240-2248.

43.   Shvartsburg, A. A.; Mashkevich, S. V.; Baker, E. S.; Smith, R. D., Optimization of Algorithms for Ion Mobility Calculations. *J. Phys. Chem. A* **2007,** *111* (10), 2002-2010.

44.   Shvartsburg, A. A.; Smith, R. D., Optimum waveforms for differential ion mobility spectrometry (FAIMS). *J. Am. Soc. Mass. Spectrom.* **2008,** *19* (9), 1286-1295.

45.   D'Atri, V.; Porrini, M.; Rosu, F.; Gabelica, V., Linking molecular models with ion mobility experiments. Illustration with a rigid nucleic acid structure. *J. Mass Spectrom. : JMS* **2015,** *50* (5), 711-726.

46.   Ewing, S. A.; Donor, M. T.; Wilson, J. W.; Prell, J. S., Collidoscope: An Improved Tool for Computing Collisional Cross-Sections with the Trajectory Method. *J. Am. Soc. Mass. Spectrom.* **2017,** *28* (4), 587-596.

47.   Konermann, L.; Metwally, H.; McAllister, R. G.; Popa, V., How to run molecular dynamics simulations on electrospray droplets and gas phase proteins: Basic guidelines and selected applications. *Methods* **2018,** *144*, 104-112.

48.   Heuvel, R. H. H. v. d.; Heck, A. J. R., Native protein mass spectrometry: from intact oligomers to functional machineries. *Curr. Opin. Chem. Biol.* **2004,** *8* (5), 519-526.

49.   Katta, V.; Chait, B. T., Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry. *J. Am. Chem. Soc.* **1991,** *113* (22), 8534-8535.

50.   Robinson, C. V.; Radford, S. E., Weighing the evidence for structure: electrospray ionization mass spectrometry of proteins. *Structure* **1995,** *3* (9), 861-865.

51.   Loo, J. A., Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **1997,** *16* (1), 1-23.

52.   Kaddis, C. S.; Loo, J. A., Native Protein MS and Ion Mobility: Large Flying Proteins with ESI. *Anal. Chem.* **2007,** *79* (5), 1778-1784.

53.   El-Hawiet, A.; Kitova, E. N.; Liu, L.; Klassen, J. S., Quantifying labile protein—Ligand interactions using electrospray ionization mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **2010,** *21* (11), 1893-1899.

54.   Marty, M. T.; Zhang, H.; Cui, W.; Blankenship, R. E.; Gross, M. L.; Sligar, S. G., Native Mass Spectrometry Characterization of Intact Nanodisc Lipoprotein Complexes. *Anal. Chem.* **2012,** *84* (21), 8957-8960.

55.   Dyachenko, A.; Gruber, R.; Shimon, L.; Horovitz, A.; Sharon, M., Allosteric mechanisms can be distinguished using structural mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110* (18), 7235.

56.   Marcoux, J.; Robinson, Carol V., Twenty Years of Gas Phase Structural Biology. *Structure* **2013,** *21* (9), 1541-1550.

57.   Heck, A. J. R., Native mass spectrometry: a bridge between interactomics and structural biology. *Nat. Methods* **2008,** *5*, 927.

58.   van Duijn, E., Current limitations in native mass spectrometry based structural biology. *J. Am. Soc. Mass. Spectrom.* **2010,** *21* (6), 971-978.

59.   Benesch, J. L. P.; Ruotolo, B. T., Mass spectrometry: come of age for structural and dynamical biology. *Curr. Opin. Struct. Biol.* **2011,** *21* (5), 641-649.

60.   Mehmood, S.; Allison, T. M.; Robinson, C. V., Mass Spectrometry of Protein Complexes: From Origins to Applications. *Annu. Rev. Phys. Chem.* **2015,** *66* (1), 453-474.

61.   Breuker, K.; Oh, H.; Horn, D. M.; Cerda, B. A.; McLafferty, F. W., Detailed Unfolding and Folding of Gaseous Ubiquitin Ions Characterized by Electron Capture Dissociation. *J. Am. Chem. Soc.* **2002,** *124* (22), 6407-6420.

62.   Breuker, K.; McLafferty, F. W., Stepwise evolution of protein native structure with electrospray into the gas phase, 10^{-12} to 10^2 s. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105* (47), 18145-18152.

63.   Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V., Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. *Science* **2005,** *310* (5754), 1658.

64.   Uetrecht, C.; Rose, R. J.; van Duijn, E.; Lorenzen, K.; Heck, A. J. R., Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* **2010,** *39* (5), 1633-1655.

65.   Makarewicz, T.; Kaźmierkiewicz, R., Molecular Dynamics Simulation by GROMACS Using GUI Plugin for PyMOL. *J. Chem. Inf. Model.* **2013,** *53* (5), 1229-1234.

66.   van Dijk, M.; Wassenaar, T. A.; Bonvin, A. M. J. J., A Flexible, Grid-Enabled Web Portal for GROMACS Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2012,** *8* (10), 3463-3472.

67.   Hinchliffe, A., *Molecular Modelling for Beginners*. John Wiley & Sons: 2003; p 428.

68.   Kukol, A., *Molecular Modeling of Proteins*. Springer: 2015; p 474.

69.   Schnier, P. D.; Gross, D. S.; Williams, E. R., Electrostatic Forces and Dielectric Polarizability of Multiply Protonated Gas-Phase Cytochrome c Ions Probed by Ion/Molecule Chemistry. *J. Am. Chem. Soc.* **1995,** *117* (25), 6747-6757.

70.   Mao, Y.; Ratner, M. A.; Jarrold, M. F., Molecular Dynamics Simulations of the Charge-Induced Unfolding and Refolding of Unsolvated Cytochrome c. *J. Phys. Chem. B* **1999,** *103* (45), 10017-10021.

71.   Hall, Z.; Politis, A.; Bush, M. F.; Smith, L. J.; Robinson, C. V., Charge-State Dependent Compaction and Dissociation of Protein Complexes: Insights from Ion Mobility and Molecular Dynamics. *J. Am. Chem. Soc.* **2012,** *134* (7), 3429-3438.

72. Dongré, A. R.; Jones, J. L.; Somogyi, Á.; Wysocki, V. H., Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *J. Am. Chem. Soc.* **1996,** *118* (35), 8365-8374.

73. Boyd, R.; Somogyi, Á., The mobile proton hypothesis in fragmentation of protonated peptides: A perspective. *J. Am. Soc. Mass. Spectrom.* **2010,** *21* (8), 1275-1278.

74. Li, J.; Lyu, W.; Rossetti, G.; Konijnenberg, A.; Natalello, A.; Ippoliti, E.; Orozco, M.; Sobott, F.; Grandori, R.; Carloni, P., Proton Dynamics in Protein Mass Spectrometry. *J. Phys. Chem. Lett.* **2017,** *8* (6), 1105-1112.

75. Fegan, S. K.; Thachuk, M., A Charge Moving Algorithm for Molecular Dynamics Simulations of Gas-Phase Proteins. *J. Chem. Theory Comput.* **2013,** *9* (6), 2531-2539.

76. Forbes, M. W.; Bush, M. F.; Polfer, N. C.; Oomens, J.; Dunbar, R. C.; Williams, E. R.; Jockusch, R. A., Infrared Spectroscopy of Arginine Cation Complexes: Direct Observation of Gas-Phase Zwitterions. *J. Phys. Chem. A* **2007,** *111* (46), 11759-11770.

77. Breuker, K.; Brüschweiler, S.; Tollinger, M., Electrostatic Stabilization of a Native Protein Structure in the Gas Phase. *Angew. Chem. Int. Ed.* **2011,** *50* (4), 873-877.

78. Yoo, H. J.; Wang, N.; Zhuang, S.; Song, H.; Håkansson, K., Negative-Ion Electron Capture Dissociation: Radical-Driven Fragmentation of Charge-Increased Gaseous Peptide Anions. *J. Am. Chem. Soc.* **2011,** *133* (42), 16790-16793.

79. Zhang, Z.; Browne, S. J.; Vachet, R. W., Exploring Salt Bridge Structures of Gas-Phase Protein Ions using Multiple Stages of Electron Transfer and Collision Induced Dissociation. *J. Am. Soc. Mass. Spectrom.* **2014,** *25* (4), 604-613.

80. Seo, J.; Warnke, S.; Pagel, K.; Bowers, M. T.; von Helden, G., Infrared spectrum and structure of the homochiral serine octamer–dichloride complex. *Nat. Chem.* **2017,** *9*, 1263.

81. Cautereels, J.; Blockhuys, F., Quantum Chemical Mass Spectrometry: Verification and Extension of the Mobile Proton Model for Histidine. *J. Am. Soc. Mass. Spectrom.* **2017,** *28* (6), 1227-1235.

82. Nakai, H.; Sakti, A. W.; Nishimura, Y., Divide-and-Conquer-Type Density-Functional Tight-Binding Molecular Dynamics Simulations of Proton Diffusion in a Bulk Water System. *J. Phys. Chem. B* **2016,** *120* (1), 217-221.

83. Marrink, S. J.; Tieleman, D. P., Perspective on the Martini model. *Chem. Soc. Rev.* **2013,** *42* (16), 6801-6822.

84. Popa, V.; Trecroce, D. A.; McAllister, R. G.; Konermann, L., Collision-Induced Dissociation of Electrosprayed Protein Complexes: An All-Atom Molecular Dynamics Model with Mobile Protons. *J. Phys. Chem. B* **2016,** *120* (23), 5114-5124.

85.   Clemmer, D. E.; Russell, D. H.; Williams, E. R., Characterizing the Conformationome: Toward a Structural Understanding of the Proteome. *Acc. Chem. Res.* **2017,** *50* (3), 556-560.

Chapter 2

2. Protein Ions Generated by Native Electrospray Ionization: Comparison of Gas Phase, Solution, and Crystal Structures

2.1. Introduction

Electrospray ionization (ESI) allows the transfer of intact proteins from solution into the gas phase. The solvent-free ions generated in this way can be interrogated by mass spectrometry (MS), ion mobility spectrometry (IMS),[1-5] and various other techniques.[6-12] Protein ESI studies are usually conducted on $[M + zH]^{z+}$ ions. However, ESI can also be performed in negative ionization mode, resulting in $[M - zH]^{z-}$ species.[1]

It remains controversial to what extent proteins retain solution-like structures after ESI.[4, 13-17] Some experiments suggested the occurrence of significant conformational changes (i.e., large-scale unfolding).[13, 15, 18-20] On the other hand, there is now substantial evidence that gaseous biomolecules can retain key aspects of their solution conformations and interactions, provided that the conditions are properly optimized.[2, 3, 17, 21-24] Such "native ESI" experiments employ nondenaturing aqueous solutions at near-neutral pH, with ion sampling and transmission conditions that avoid excess collisional heating.[7, 21, 25-28]

Native globular proteins in solution are tightly folded; most hydrophobic side chains are sequestered in the core, whereas titratable and polar side chains tend to be on the surface.[29] Proteins experience dramatic environmental changes during ESI. In water, the hydrophobic effect represents a key stabilizing factor.[30] The absence of water after ESI suggests that hydrophobicity is not a major contributor to protein stability in vacuo.[31] Also, the protonation states of titratable sites can be very different in solution and after ESI.[32] The fact that electrosprayed ions carry a net charge $z \neq 0$ implies that their internal electrostatics are dominated by repulsive forces,[33-35] an aspect that is compounded by the low dielectric constant in the gas phase ($\varepsilon_{vacuum} = 1$).[36] Considering all these factors, the retention of solution-like structures after ESI is a remarkable phenomenon.[2, 3, 17, 21-24] This retention has been attributed to kinetic trappings,[14, 37, 38] i.e., the presence of activation barriers that prevent large-scale transitions on the time scale of typical ESI-IMS/MS experiments.[16]

Even under properly optimized native ESI conditions, the transition from solution into the gas phase will be accompanied by some structural alteration.[16] Molecular dynamics (MD) simulations[33, 39] and microsolvation studies[40] suggest that extended surface side chains such as Lys^+ tilt towards the

protein surface where they bind to backbone carbonyls. Salt bridge formation among side chains on the protein surface may take place as well.[39] Such side chain collapse in the gas phase is consistent with the observation that IMS-derived collision cross sections (Ω) tend to be smaller than expected from crystal structures.[28, 33, 41-43] However, the exact nature of such gas phase compaction events remains incompletely understood, and predicted outcomes may depend on the modeling strategy used.[20, 34, 44-46]

Efforts to delineate protein structural differences before and after ESI are hampered by various factors. Chief among these is the lack of high-resolution structure determination methods in the gas phase. IMS-derived Ω values report on the overall compactness, but any Ω value may be consistent with a multitude of conformations.[47] A standard approach is to compare experimental Ω values with MD-derived candidate structures.[4, 34, 47, 48] One challenge associated with gas phase MD simulations is the proper placement of charges. Proteins contain numerous sites that can be protonated (N-terminus, Arg, Lys, His) or deprotonated (Asp, Glu, C-terminus), giving rise to an astronomical number of possible charge configurations.[49] The problem becomes even more daunting when allowing for zwitterionic motifs, i.e., the presence of both positively and negatively charged sites regardless of the ion polarity.[49-55] Identifying the most suitable protonation pattern has to consider interactions among protonated/deprotonated moieties, charge solvation effects, as well as the intrinsic proton affinities (PA_{int}) of individual sites.[36, 40, 56-58] Most MD studies treat protonation patterns as stationary,[34, 39, 59, 60] whereas in reality protons in gaseous proteins are mobile.[61-64] Mobile-proton MD approaches are a promising strategy for addressing these difficulties,[57, 58] but thus far such techniques are not widely used. Also, most MD simulations focus on ns time windows that are much shorter than the tens of milliseconds required for typical IMS experiments.[4, 34, 47, 48]

Understanding ESI-associated structural changes is further complicated by uncertainties regarding the exact properties of proteins in solution. Much of the ESI literature assumes that X-ray crystallography provides an accurate representation of solution-phase conformations. Consequently, gas phase vs. "solution" phase comparisons usually rely on crystallographic data.[14, 39, 41, 45] Although protein crystals contain some water,[65] X-ray and solution structures do not always match. Surface side chains and loops are highly dynamic in solution, while these elements tend to adopt well defined orientations after crystallization.[66-68] Such ordering arises from crystal packing, i.e., interactions with neighboring proteins in the lattice.[69, 70] Nuclear magnetic resonance (NMR) spectroscopy provides information on proteins in bulk solution. However, NMR structures are calculated largely from backbone distance and angle restraints.[71] The dynamic nature of exposed side chains implies that NMR coordinates of the corresponding atoms should be treated with caution.[68] In summary, the exact nature of protein structural differences before and after ESI remain to be uncovered. Existing

uncertainties are rooted in the absence of high-resolution information on gaseous ions, and from ambiguities in condensed phase reference data.

The current work addresses the issues outlined above by taking a critical look at the crystal, solution, and gas phase properties of ubiquitin and lysozyme. We compared ESI-MS/IMS data with the results of extended (1 µs) gas phase MD data for both $[M + zH]^{z+}$ and $[M - zH]^{z-}$ ions. Mimicking MD strategies used in previous studies[34, 39, 59, 60] we initially employed static charge models. The main focus of this work, however, is on mobile proton MD data generated with a recently developed mobile proton algorithm[57] that accounts for the known mobility of H^+ in gas phase proteins.[61-64] Proof-of-principle work using this algorithm has already been presented,[57] but a critical comparison of data obtained by this approach with ESI-MS/IMS experiments is still lacking. The current work closes this gap by combining experimental and computational data, thereby providing detailed insights into the properties of electrosprayed protein ions, and their relationship to solution and crystal structures.

2.2. Methods

2.2.1. Mass Spectrometry and Ion Mobility Spectrometry

Bovine ubiquitin (8565 Da) and chicken egg white lysozyme (14305 Da) were purchased from Sigma (St. Louis, MO, USA). Samples were prepared in 10 mM aqueous ammonium acetate (pH 7) at a protein concentration of 10 µM. Positive ion data were acquired on a Synapt HDMS G1 time-of-flight mass spectrometer (Waters, Milford, MA) with a Z-spray ESI regular source operated at +2.8 kV. Solutions were infused at 5 µL min^{-1} using a syringe pump. Source temperatures, as well as source voltages and ion transmission voltages were reduced as much as possible (source temperature 25 °C, desolvation temperature 40 °C, sampling cone 5 V, extraction cone 1 V, trap collision energy 2 V, trap DC bias 8-9 V, trap gas off). All other instrument parameters were as described.[72] These conditions were chosen to minimize collisionally or thermally induced structural changes of the protein,[28, 33] resulting in instrument operation just above the ion transmission threshold. A further drop in transmission was encountered for negative ion experiments, necessitating the use of a more sensitive instrument, a Synapt G2 (ESI voltage -2.8 kV) under otherwise identical conditions. Collision cross sections were measured using traveling wave ion mobility measurements with N_2 as buffer gas. The IMS data were converted to He Ω values. The calibration procedure used for this purpose was performed using a mix of collisionally activated reference proteins. This calibration procedure is well established for positive ion mode,[72] but its use for negative ion IMS is new (see Table 2.1 for details). The IMS reference mix was infused in 49:49:2 (v/v/v) methanol/water/acetic

acid with the cone set to 100 V, while keeping all other instrument parameters as described above. Ω errors represent the standard deviation of three independent measurements.

Table 2.1. Reference He Ω values for collisionally unfolded standard proteins (horse apo-myoglobin (Mb), horse heart cytochrome c (Cyt) and bovine ubiquitin (Ubq) at cone = 100 V, electrosprayed from a denatured solvent system (49:49:2 (v/v/v) methanol/water/acetic acid). Data are from David Clemmer's website:

http://www.indiana.edu/~clemmer/Research/Cross Section Database/Proteins/protein_cs.htm

Protein	Charge	Ω (Å^2)
Mb	8+	2352
Mb	9+	2704
Mb	10+	2796
Mb	11+	2942
Mb	12+	3044
Mb	13+	3136
Mb	14+	3143
Mb	15+	3230
Mb	16+	3313
Mb	17+	3384
Mb	18+	3489
Mb	19+	3570
Mb	20+	3682
Mb	21+	3792
Mb	22+	3815
Cyt	7+	2007
Cyt	8+	2061
Cyt	9+	2215
Cyt	10+	2226
Cyt	11+	2303
Cyt	12+	2335
Cyt	13+	2391
Cyt	14+	2473
Cyt	15+	2579
Cyt	16+	2679
Cyt	17+	2723
Cyt	18+	2766
Ubq	6+	1525
Ubq	7+	1580
Ubq	8+	1622
Ubq	9+	1649
Ubq	10+	1732
Ubq	11+	1802

Protein	Charge	Ω (Å^2)
Mb	6-	1795
Mb	7-	2025
Mb	8-	2203
Mb	9-	2535
Mb	10-	2656
Mb	11-	2879
Mb	12-	2968
Mb	13-	3057
Mb	14-	3136
Mb	15-	3226
Mb	16-	3281
Mb	17-	3395
Cyt	5-	1246
Cyt	6-	1535
Cyt	7-	1586
Cyt	9-	2024
Ubq	4-	1059

2.2.2. MD simulations
2.2.2.1. Solution MD Simulations

All simulations of this work were performed using Gromacs 2016.3[73] at 300 K with a 2 fs integration step. The crystal structures 1ubq (ubiquitin, 1.8 Å)[74] and 1aki (lysozyme, 1.5 Å)[75] served as starting conformations because they contained the complete set of protein atoms, and because of their high resolution (below 2 Å). For 1 µs solution simulations the Charmm36 force field[76] was used with TIP3P water because this combination has previously been shown to adequately model the behavior of small globular proteins.[77] The simulations employed periodic boundary conditions with a minimum 1 nm distance between protein atoms and the edge of the periodic box and particle mesh Ewald summation. NaCl was added at a concentration of 150 mM, and additional ions were added as needed to ensure charge neutrality. All other MD parameters and procedures were as described.[78]

2.2.2.2 Gas Phase MD Simulations

The behavior of gaseous ubiquitin and lysozyme ions was modeled in 1 µs runs at 300 K, using the crystal data 1ubq and 1aki as starting points. The simulations employed the OPLS/AA force field[79, 80] which has been widely used for earlier gas phase applications,[15, 24, 60, 81-83] and that has been validated against ab initio data on model compounds in the absence of solvent.[79, 80] Our simulations were performed in a vacuum environment without cut-offs for electrostatic or Lennard-Jones interactions as described.[57] For static charge runs the protonation states of all titratable sites were set using Gromacs pdb2gmx, and the initial charge patterns were retained throughout the 1 µs runs.

Mobile proton simulations were conducted by complementing Gromacs with in-house Fortran code and bash scripts,[57] and by breaking down 1 µs MD runs into 1 ns segments. After each of these segments the protons residing on the various sites (N-terminus$^+$ [referred to as NT$^+$], Arg$^+$, Lys$^+$, His$^+$, Asp^0, Glu^0, C-terminus^0 [denoted as CT^0]) were redistributed using a steepest descent energy minimization procedure that takes into account PA$_{int}$ of all possible acceptor sites,[84] as well as electrostatic interactions and intramolecular charge solvation involving all atoms in the protein. PA$_{int}$ values used were[84] 886.6 (NT), 918.0 (Lys), 1004.0 (Arg), 952.7 (His), 1452.7 (Asp$^-$), 1453.5 (Glu$^-$), 1423.8 (CT$^-$) kJ mol^{-1}. This mobile proton strategy allowed basic sites to switch between their protonated/neutral states, while acidic sites could switch between neutral/deprotonated. All proton hopping events were subject to preservation of total charge, as governed by the net z value of the protein ion.[57] The presence of two titratable sites for the N-terminal Lys in lysozyme (NT and side

chain) caused some difficulties with the code used here. This challenge was overcome by adding an N-terminal Gly to the sequence, such that the two charge sites were in two separate residues. X-ray data show that this N-terminus is part of a disordered tail,[75] and we conducted static charge simulations to ensure that the presence of this N-terminal Gly did not affect the structure or dynamics of lysozyme in any noticeable way.

2.3.   Results and Discussion
2.3.1. Experimental Characterization of Gaseous Protein Ions

Ubiquitin and lysozyme served as model systems for the native ESI experiments of this work. Both proteins possess a mixed α /β secondary structure. Ubiquitin does not have any disulfides, while lysozyme is stabilized by four disulfide bridges as seen in the pdb file.[74, 75] The positive ion ubiquitin mass spectrum was dominated by 6+ ions. The corresponding Ω distribution peaked at 960 Å^2. A less intense satellite band was present around 1140 Å^2, revealing a small sub-population of semi-unfolded protein (Figure 2.1. A, B). In negative ion mode the protein mainly formed 5- ions with a Ω maximum at ~990 Å^2 (Figure 2.1. C, D). Lysozyme in positive ion mode exhibited a dominant 8+ signal, and the collision cross section of the corresponding ions was around 1390 Å^2 (Figure 2.1. E, F). In negative ESI lysozyme generated 6- ions ($\Omega \approx$ 1470 Å^2, Figure 2.1. G, H). The standard deviation of the measured Ω maxima was around 1%. For both proteins the Ω distributions in positive ion mode were slightly wider, indicating a greater conformational heterogeneity. The gentle ion sampling conditions chosen here caused broadening of the lysozyme mass distribution in negative ion mode due to acetate adduction, but Ω values of adducted and non-adducted ions were almost indistinguishable (Figure 2.2.). The ESI-IMS/MS data reported here are consistent previous literature reports [14, 28, 42, 43] Similar to other proteins, the $|z|$ values in positive ion mode were slightly higher than in negative ion mode, while the corresponding Ω values were similar for both polarities.[1, 85]
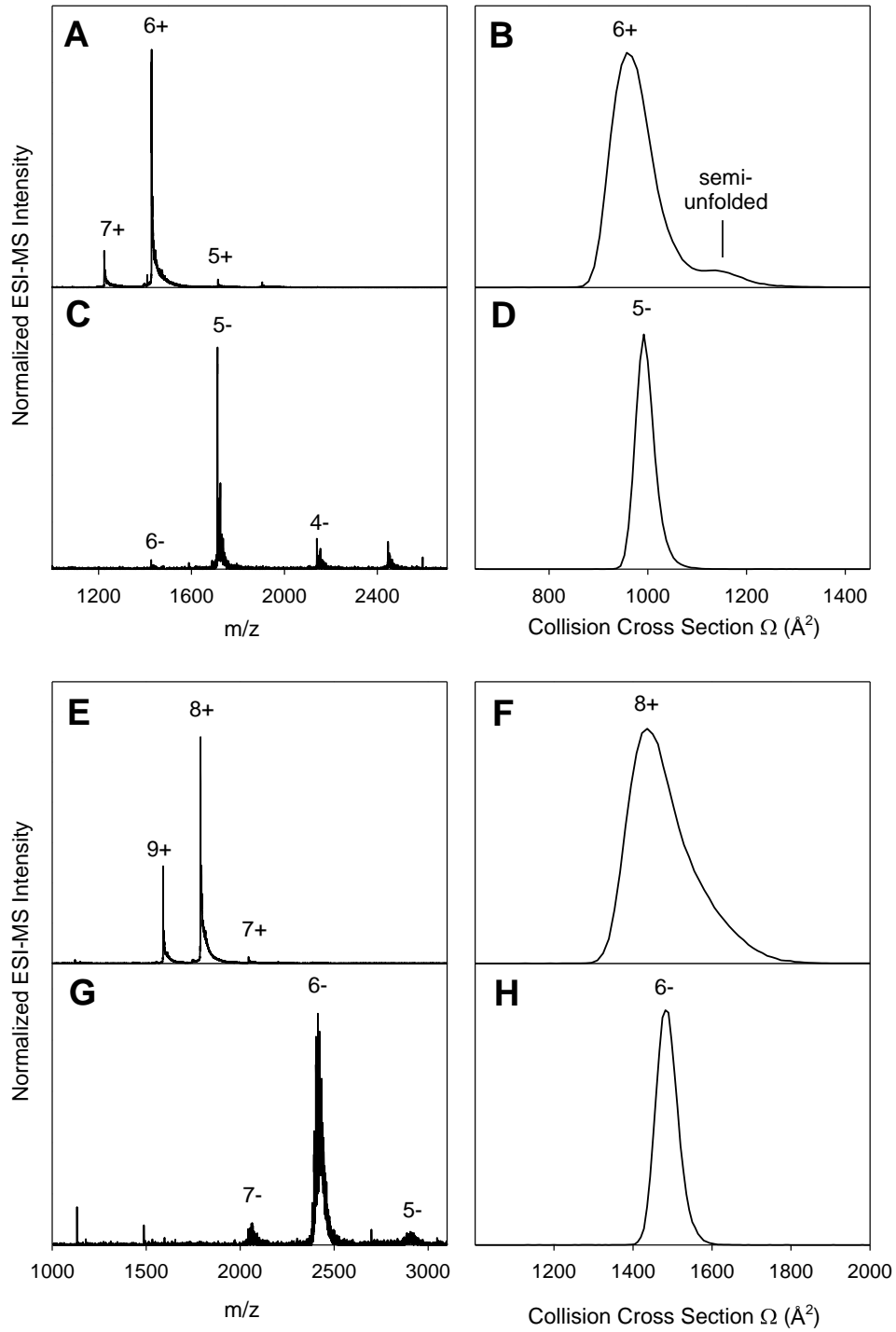
Figure 2.1. Native ESI-MS/IMS data. Ubiquitin in (A, B) positive ion mode and (C, D), negative ion mode; (E, F) lysozyme in positive ion mode and (G, H) negative ion mode.
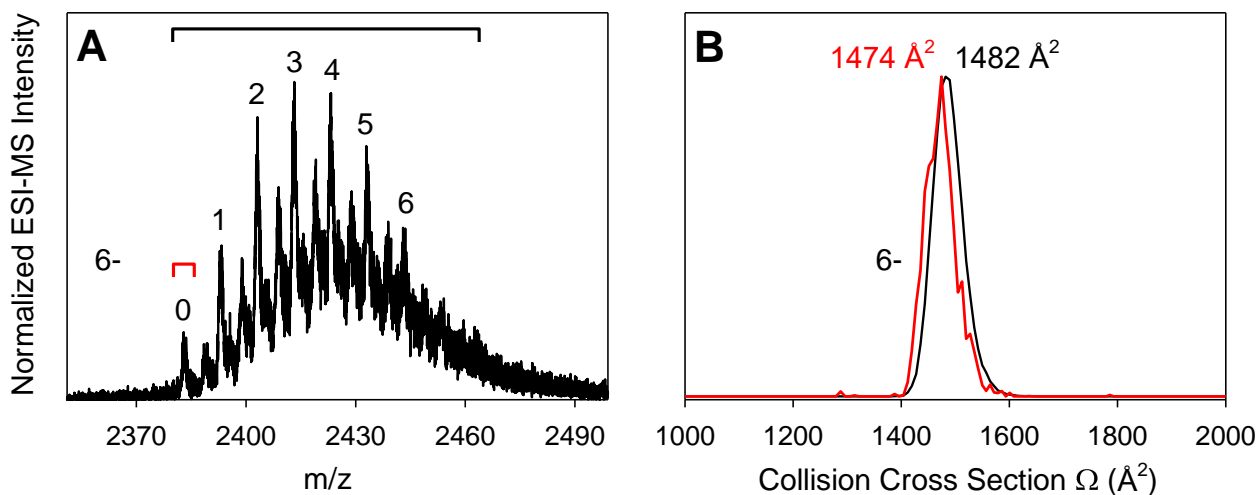
Figure 2.2. Lysozyme IMS/MS data acquired in negative ion mode under native ESI conditions. (A) Close-up of the 6- mass spectral region. The peaks are broadened due to acetate adducts (acetic acid, 60 Da), reflecting the presence of ammonium acetate in the sample solution. The number of acetate moieties is indicated as 0, 1, 2, ... (B) IMS profile extracted for adduct-free ions (red bracket in panel A), and for the entire ion population (black bracket in panel A). It is evident that the acetate adducts do not strongly affect the collision cross section of the protein. Panel B also includes the maxima of the two IMS profiles.

2.3.2. Crystal Contacts and Conformations

As noted in the Introduction, native ESI investigations commonly rely on crystal structures when comparing "solution" and gas phase proteins.[14, 39, 41, 45] The underlying assumption is that crystal structures are a faithful representation of solution conformations. One objective of the current study was to test the validity of this approach, a task that requires a careful analysis of crystal structures.

For both ubiquitin and lysozyme the pdb-archived asymmetric unit corresponds to a single polypeptide chain (green in Figure 2.3. A, B). Visualization of these pdb files in a viewer program results in single-protein images that omit all crystal neighbours. Figure 2.3. A, B provides an expanded view that comprises a central protein (green) with its neighbors (gray, generated using the symmetry mates feature of Pymol). Analysis of the crystal packing interactions revealed a multitude of salt bridges between adjacent proteins, with O-N distances on the order of 4 Å or less.[86] The participating moieties include Arg^+, Lys^+, Asp^-, Glu^- side chains, and C-terminal carboxylates (Figure 2.3. C-D). These contacts are formed because the corresponding side chains point away from the protein surface towards the crystal neighbors. This trend is particularly prevalent for ubiquitin, which exhibits tighter crystal packing than lysozyme. Both proteins have a protruding $R-COO^-$ terminus that is in contact with Lys^+ from a neighboring protein. In addition, lysozyme has a C-terminal Arg^+/Asp^-

contact (Figure 2.3. C-D). The C-terminal four residues of ubiquitin are disordered in the crystal, and the coordinates in this region represent a modeled average orientation.[74]

Crystal contacts such as those in Figure 2.3. C-D can alter protein conformations,[66, 67, 69, 70] but the severity of such distortion effects has to be evaluated on a case-by-case basis. In the context of native ESI-MS it is of particular interest to explore whether the extended Lys/Arg/Asp/Glu/CT orientations are caused by crystal packing. Alternatively, the corresponding moieties might already have a tendency to adopt extended orientations in bulk solution.[39] This paradox is discussed in the following section.

2.3.3. Crystal Structures vs. Solution MD Conformations

MD simulations in bulk water were conducted to explore the structural preferences of individual ubiquitin and lysozyme chains in the absence of crystal packing. Three independent 1 μs solution runs for each protein yielded root mean square deviation (RMSD) values of 0.1 - 0.2 nm relative to the X-ray data, indicating that the solution conformations remained very close to the crystal structures (Figure 2.4.). The resemblance of crystal and MD backbone coordinates is illustrated in Figure 2.3. E-F. Noticeable differences were confined to the C-terminal tail of ubiquitin which was quite mobile for the solution structures (Figure 2.3. E), in agreement with experimental data.[74] Of particular interest are the orientations of titratable side chains, many of which reach across to neighboring proteins in the crystal (see above, Figure 2.3. C-D). Remarkably, these extended side chain orientations were retained in solution, albeit with a greater degree of disorder compared to the crystals (Figure 2.3. G-H). This behavior is also evident from radius of gyration (R_g) analyses, which showed that the overall compactness of surface side chains in solution was close to the corresponding X-ray values (Figure 2.3. I-J).

The purpose of the preceding sections was to assess the appropriateness of crystal data as surrogates for solution conformations when discussing structural changes during ESI. For the two model proteins considered here the crystal conformations do indeed provide a faithful representation of the solution structure, validating the strategy used in earlier ESI studies.[14, 39, 41, 45] However, this equivalence of crystal and solution structures does not hold for all proteins, requiring careful case-by-case considerations.[66, 67, 69, 70] Figure 2.5. represents additional solution phase MD data for ubiquitin and lysozyme, complementing Figure 2.3. I-J of the main text.
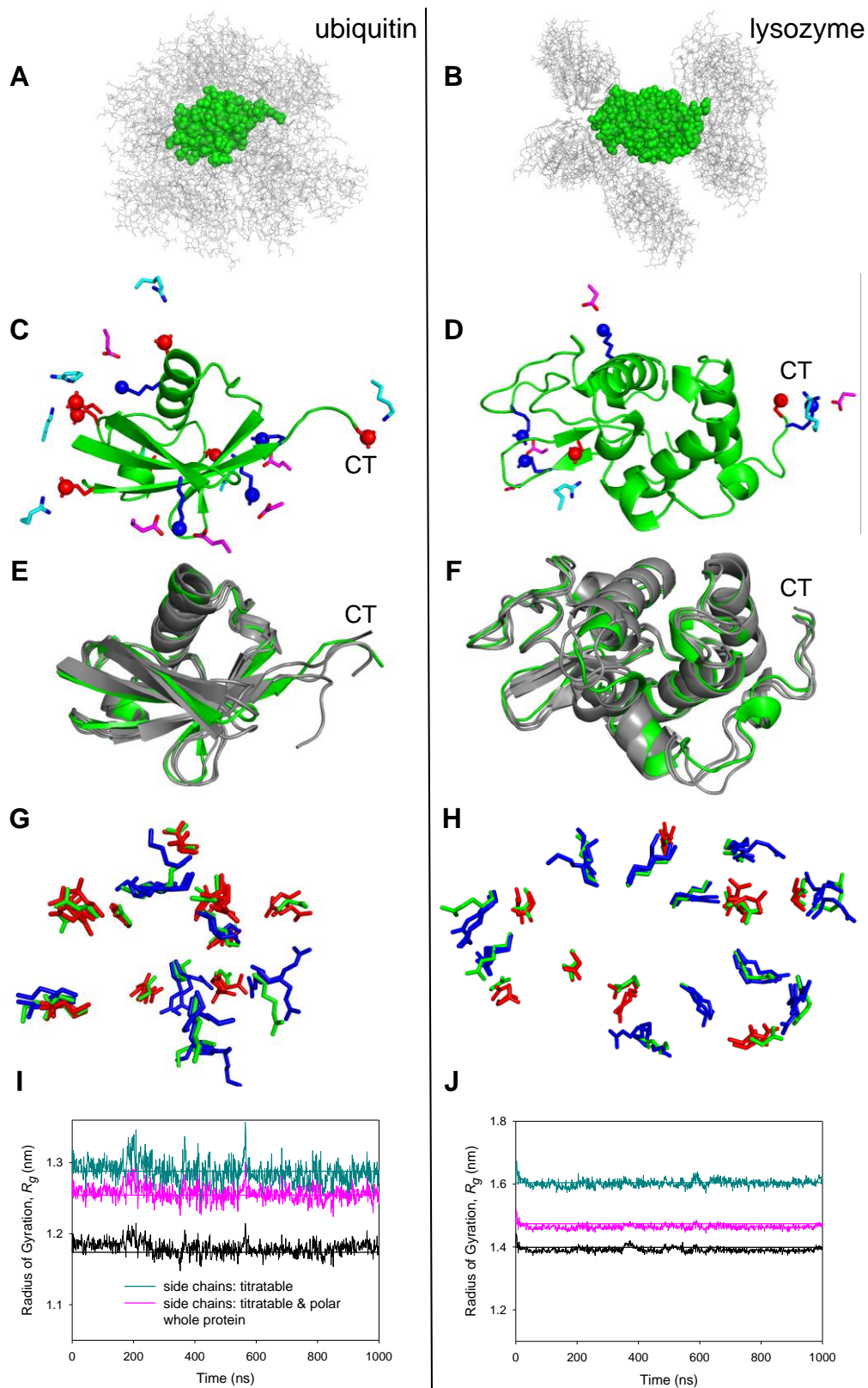
Figure 2.3. (A, B) Ubiquitin and lysozyme crystal structures, highlighting one chain (green) that is surrounded by others (gray). (C, D) Crystal salt bridges of Lys+/Arg+ (blue) and Asp-/Glu-/CT- (red) with side chains from adjacent proteins (Lys/Arg: cyan; Asp/Glu: magenta). (E, F) Crystal backbone structures (green) overlaid with three 1 μs solution MD runs (gray). (G, H) Titratable side chains in the crystal (green) and from three 1 μs solution MD runs (Lys/Arg: blue; Asp/Glu: red). Not all side chains are shown to reduce clutter. (I, J) Radius of gyration R_g for titratable side chains, titratable and

polar side chains, and for the whole protein during 1 µs solution MD runs. Horizontal lines represent crystal R_g values.
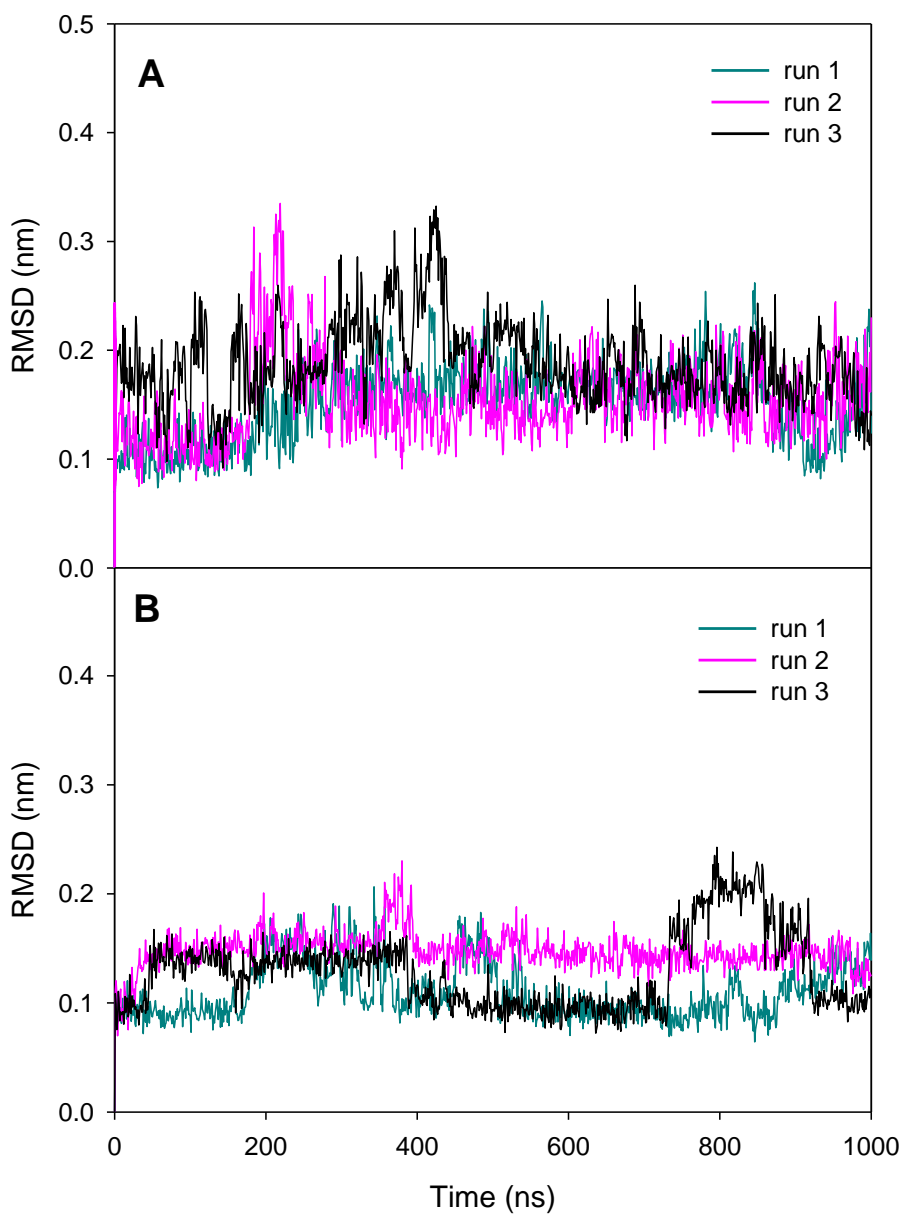
Figure 2.4. Solution phase MD data for ubiquitin (A) and lysozyme (B), displaying the root mean square deviation (RMSD) relative to the crystals structure as a function of simulation time. Both panels contain data from three independent runs. The low RMSD values indicate that the proteins retain conformations close to the crystal structure throughout the 1 µs simulation window.

ubiquitin   lysozyme
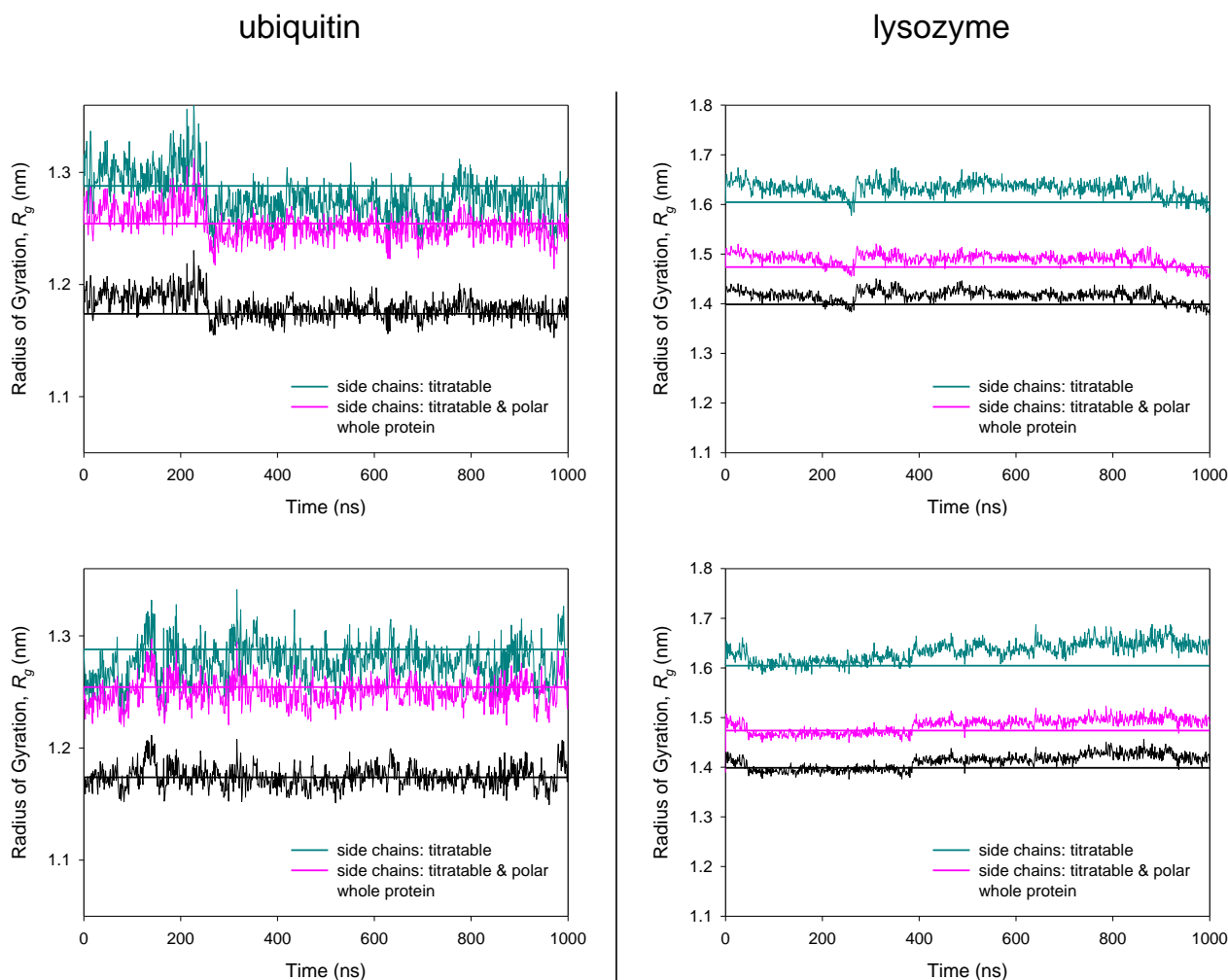
Figure 2.5. Additional solution phase MD data for ubiquitin and lysozyme, complementing Figure 2I-J of the main text. This figure displays the radius of gyration (R_g) as a function of simulation time for different groups of atoms. Horizontal lines refer to the crystal structures 1ubq and 1aki, respectively. The orientation of surface side chains and the overall compactness of the protein in solution stay close to the corresponding crystal parameters.

2.3.4. Modeling Protons in Gas Phase MD Simulations

For simulating the behavior of gaseous protein ions three different charge models were used.

(i) Static protons (positive-only or negative-only). This approach represents the simplest model tested here; it did not allow for proton transfer within the protein. Also, zwitterionic motifs were not considered, i.e., $[M + z\text{H}]^{z+}$ ions contained exactly z positive charges at basic sites (N-terminus, Arg, Lys, His), while all acidic sites remained in their neutral R-COOH state. Conversely,

[M - zH]$^{z-}$ ions were treated by having z negative charges on selected R-COO$^-$ moieties (Asp, Glu, C-terminus), without any positive charges.

(ii) Static protons (zwitterionic). This second model allowed the presence of zwitterionic motifs and salt bridges, reflecting the strong evidence that electrosprayed [M + zH]$^{z+}$ ions can contain R-COO$^-$ sites (and vice versa for [M - zH]$^{z-}$ ions).[49-55] Under these conditions the number of (de)protonated sites exceeds z due to the presence of opposite charges. However, model 2 also did not allow for proton transfer events. Models 1[34, 48] and 2[39, 60] mimic strategies that have been widely used for earlier gas phase MD studies. Charge patterns used for all static proton runs are summarized in Tables 2.2. and 2.3.

(iii) Mobile protons. The third model employed a recently developed technique[57] that accounts for the highly mobile nature of protons in gaseous proteins.[61-64] During MD runs all H$^+$ were allowed to move among basic and acidic sites, as governed by their mutual electrostatic interactions, charge solvation effects, and by the PA_{int} values of protonation sites. Model 3 allowed for salt bridges. The formation of such zwitterionic motifs was governed by the overall energetics of the system, without manual intervention by the user.[57]

Table 2.2. Ubiquitin Sequence and Protonation Patterns (1ubq.pdb)

```
1                   21                  41                  61
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
```

Acidic sites: 12/76 = 15.8%   Basic sites: 13/76 = 17.1%

Protonation patterns (bold/underlined residues are charged)

[ubiquitin + 6H]$^{6+}$ static protons / positive-only

```
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
```
(the last two patterns show reversible unfolding)

[ubiquitin + 6H]$^{6+}$ static protons / zwitterionic

```
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
```

[ubiquitin - 5H]$^{5-}$ static protons / negative-only

```
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
```

[ubiquitin - 5H]$^{5-}$ static protons / zwitterionic

```
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
```

Table 2.3. Lysozyme Sequence and Protonation Patterns (1aki.pdb)

(The sequence used here contains an N-terminal glycine that is not present in the X-ray structure – see Methods).

```
 1                      21                     41
G KVFGRCELAAAMKRHGLDNY RGYSLGNWVCAAKFESNFNT QATNRNTDGSTDYGILQINS
 61                     81                     101                    121
 RWWCNDGRTPGSRNLCNIPC SALLSSDITASVNCAKKIVS DGNGMNAWVAWRNRCKGTDV QAWIRGCRL
```

Acidic sites: 10/130 = 7.7%   Basic sites: 19/130 = 14.6%

<u>Protonation patterns (bold/underlined residues are charged)</u>

[lysozyme + 8H]$^{8+}$ static protons / positive-only

```
G KVFGRCELAAAMKRHGLDNY RGYSLGNWVCAAKFESNFNT QATNRNTDGSTDYGILQINS
  RWWCNDGRTPGSRNLCNIPC SALLSSDITASVNCAKKIVS DGNGMNAWVAWRNRCKGTDV QAWIRGCRL

G KVFGRCELAAAMKRHGLDNY RGYSLGNWVCAAKFESNFN TQATNRNTDGSTDYGILQINS
  RWWCNDGRTPGSRNLCNIPC SALLSSDITASVNCAKKIV SDGNGMNAWVAWRNRCKGTDV QAWIRGCRL

G KVFGRCELAAAMKRHGLDNY RGYSLGNWVCAAKFESNFN TQATNRNTDGSTDYGILQINS
  RWWCNDGRTPGSRNLCNIPC SALLSSDITASVNCAKKIV SDGNGMNAWVAWRNRCKGTDV QAWIRGCRL
```

[lysozyme + 8H]$^{8+}$ static protons / zwitterionic

```
G KVFGRCELAAAMKRHGLDNY RGYSLGNWVCAAKFESNFNT QATNRNTDGSTDYGILQINS
  RWWCNDGRTPGSRNLCNIPC SALLSSDITASVNCAKKIVS DGNGMNAWVAWRNRCKGTDV QAWIRGCRL

G KVFGRCELAAAMKRHGLDNY RGYSLGNWVCAAKFESNFNT QATNRNTDGSTDYGILQINS
  RWWCNDGRTPGSRNLCNIPC SALLSSDITASVNCAKKIVS DGNGMNAWVAWRNRCKGTDV QAWIRGCRL

G KVFGRCELAAAMKRHGLDNY RGYSLGNWVCAAKFESNFNT QATNRNTDGSTDYGILQINS
  RWWCNDGRTPGSRNLCNIPC SALLSSDITASVNCAKKIVS DGNGMNAWVAWRNRCKGTDV QAWIRGCRL
```

[lysozyme - 6H]$^{6-}$ static protons / negative-only

```
G KVFGRCELAAAMKRHGLDNY RGYSLGNWVCAAKFESNFNT QATNRNTDGSTDYGILQINS
  RWWCNDGRTPGSRNLCNIPC SALLSSDITASVNCAKKIVS DGNGMNAWVAWRNRCKGTDV QAWIRGCRL

G KVFGRCELAAAMKRHGLDNY RGYSLGNWVCAAKFESNFNT QATNRNTDGSTDYGILQINS
  RWWCNDGRTPGSRNLCNIPC SALLSSDITASVNCAKKIVS DGNGMNAWVAWRNRCKGTDV QAWIRGCRL

G KVFGRCELAAAMKRHGLDNY RGYSLGNWVCAAKFESNFNT QATNRNTDGSTDYGILQINS
  RWWCNDGRTPGSRNLCNIPC SALLSSDITASVNCAKKIVS DGNGMNAWVAWRNRCKGTDV QAWIRGCRL
```

[lysozyme - 6H]$^{6-}$ static protons / zwitterionic

```
G KVFGRCELAAAMKRHGLDNY RGYSLGNWVCAAKFESNFNT QATNRNTDGSTDYGILQINS
  RWWCNDGRTPGSRNLCNIPC SALLSSDITASVNCAKKIVS DGNGMNAWVAWRNRCKGTDV QAWIRGCRL

G KVFGRCELAAAMKRHGLDNY RGYSLGNWVCAAKFESNFNT QATNRNTDGSTDYGILQINS
  RWWCNDGRTPGSRNLCNIPC SALLSSDITASVNCAKKIVS DGNGMNAWVAWRNRCKGTDV QAWIRGCRL

G KVFGRCELAAAMKRHGLDNY RGYSLGNWVCAAKFESNFNT QATNRNTDGSTDYGILQINS
  RWWCNDGRTPGSRNLCNIPC SALLSSDITASVNCAKKIVS DGNGMNAWVAWRNRCKGTDV QAWIRGCRL
```

2.3.5. Overview of Gas Phase MD Results

Multiple 1 μs simulations were generated, focusing on the dominant charge states seen in the experimental spectra (Figure 2.6). The overall behavior seen in these MD runs is illustrated by the R_g values of Figure 2.6. for all three charge models.

Static proton/positive-only simulations on [ubiquitin + 6H]$^{6+}$ were conducted for various protonation patterns (Table 2.2.). Most of the trajectories generated using this model retained tightly folded conformations, while two of the runs showed fluctuations between semi-unfolded and compact structures (Figure 2.6. A). Subsequently we tested the behavior of [ubiquitin + 6H]$^{6+}$ in static proton/zwitterionic runs. All the trajectories produced in this way retained tightly folded structures (Figure 2.6. B). Mobile proton simulations on [ubiquitin + 6H]$^{6+}$ were dominated by compact conformers, with occasional unfolding/refolding for some of the trajectories (Figure 2.6. C). Most of these reversible structural events would have gone undetected in the much shorter (~10 ns or less) time windows used for earlier gas phase simulations,[34, 39, 48] highlighting the necessity of using an extended time scale.

An analogous set of simulations was performed for [ubiquitin - 5H]$^{5-}$, resulting in the R_g data of Figure 2.6. D-E. Across the board, the negative ions data displayed a high stability without unfolding, regardless of the charge model used. The greater stability of negative ubiquitin ions relative to their positive counterparts is attributed to their lower internal electrostatic repulsion, keeping in mind their z values (|5-| < |6+|).[33-35] To support this assertion, we repeated MD runs for positive ion patterns that consistently caused unfolding (red and blue, Figure 2.6. A), but lowered the charge state from 6+ to 5+. As expected, the 5+ ions retained a compact conformation (Figure 2.7.).

MD simulations were also conducted on lysozyme in the gas phase, using all three protonation models in both polarities. The resulting data showed lysozyme to be highly stable, without any large-scale unfolding for any of the conditions tested (Figure 2.8., Table 2.3.). The greater resilience of lysozyme$^+$ compared to ubiquitin is attributed to the presence of four disulfide bridges in the former.[75]

The key question explored here is to what extent ubiquitin and lysozyme retain their native solution structures in the gas phase. As discussed above, the crystal structures of both proteins provide an adequate representation of their solution conformations. Thus, Figure 2.9. compares mobile proton MD conformations (gray) with the corresponding crystal data (green). Figure 2.9. A exemplifies data from a [ubiquitin + 6H]$^{6+}$ trajectory that did not show unfolding, revealing that the backbone fold remained very close to that of the crystal structure throughout the 1 μs simulation window. The same is true for [ubiquitin - 5H]$^{5-}$, [lysozyme + 8H]$^{8+}$, and [lysozyme - 6H]$^{6-}$ (Figure 2.9. C-E). RMSD values relative to the corresponding crystal structures for these "folded" trajectories were on the order

of 0.3 to 0.4 nm (Figures 2.10., 2.11.), reinforcing the fact that the gas phase proteins remained close to their crystal structures (albeit not as close as in solution where RMSDs were 0.1 - 0.2 nm, Figure 2.4.).

Figure 2.9. B illustrates one of the reversible unfolding transitions experienced by [ubiquitin + 6H]$^{6+}$. For this particular trajectory the protein opened up at $t \approx 100$ ns, (magenta), but at $t = 200$ ns it had folded back to a compact state. The fact that gaseous ubiquitin returned close to its original structure after being semi-unfolded implies that the solution-like conformation represents a local free energy minimum, consistent with the kinetic trapping hypothesis outlined above.[14, 37, 38]
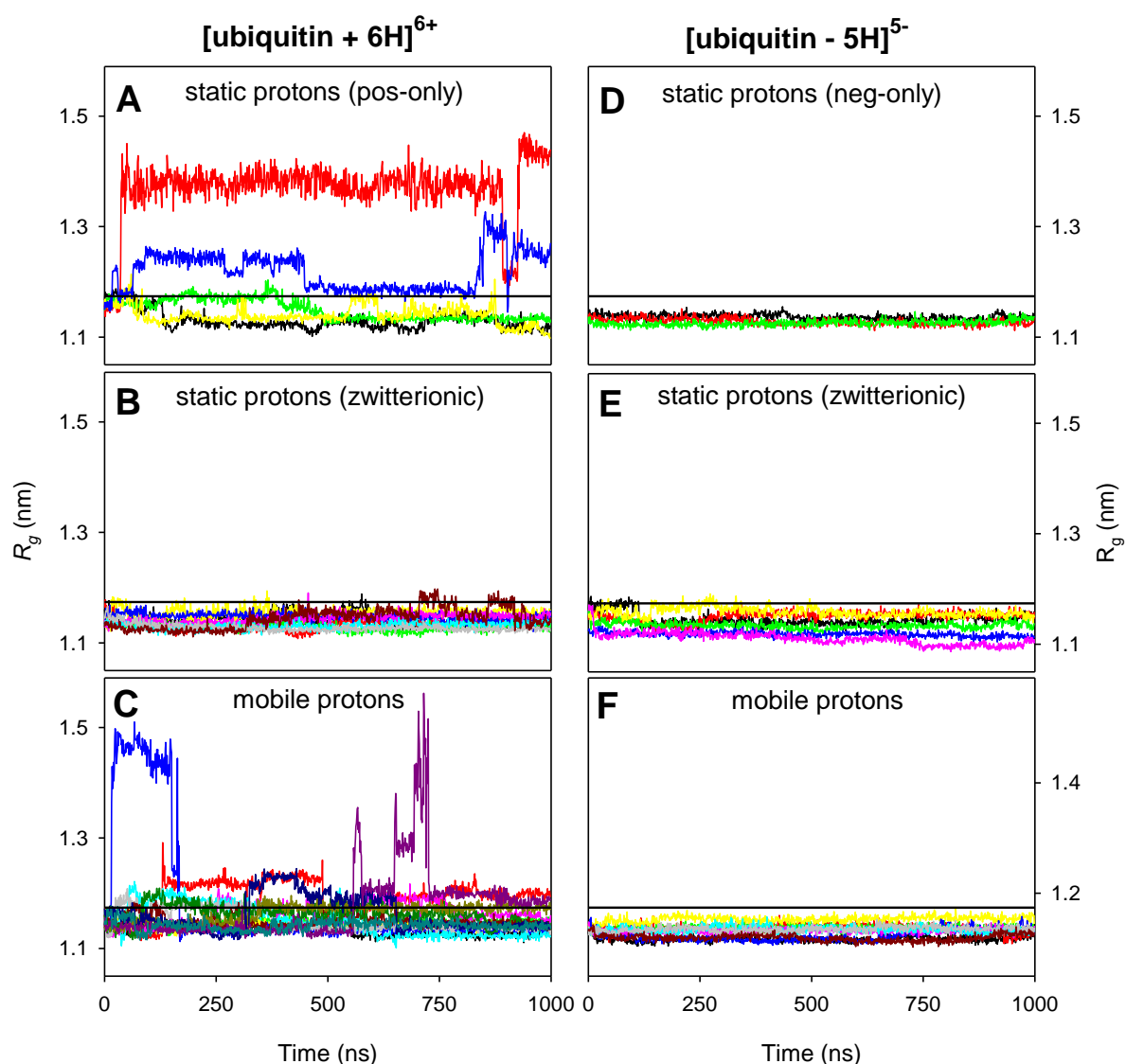

Figure 2.6. Whole protein R_g values obtained in 1 µs ubiquitin gas phase MD simulations (A-C, 6+ charge state; D-F, 5- charge state). The panels contain data from 4-14 independent runs. Data in the top row were obtained using stationary positive-only (A) and negative-only (D) charges. The center row data were generated using various stationary zwitterionic charge patterns (B, E). Panels C and F show the results of mobile proton MD runs.
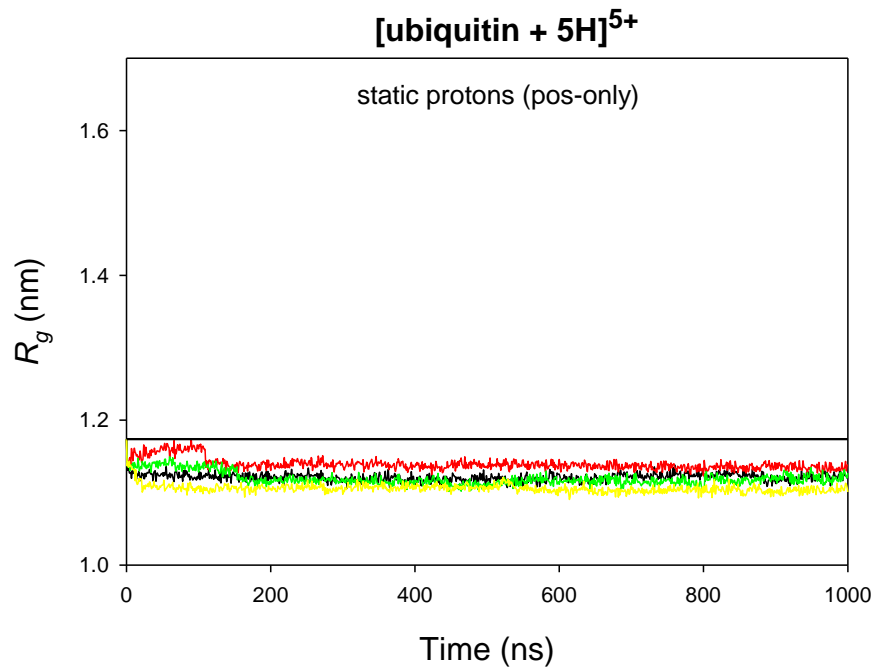
Figure 2.7. Gas phase MD data for [ubiquitin + 5H]$^{5+}$ with static protons, in positive-only mode. The proton configuration for these runs resembled those used for the runs in Figure 3. A. (main text) where unfolding was observed. After removal of one proton, a native-like compact conformation was retained for all the simulations of this Figure.

Protonation patterns used for these runs:

```
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
MQIFVKTLTGKTITLEVEPS DTIENVKAKIQDKEGIPPDQ QRLIFAGKQLEDGRTLSDYN IQKESTLHLVLRLRGG
```
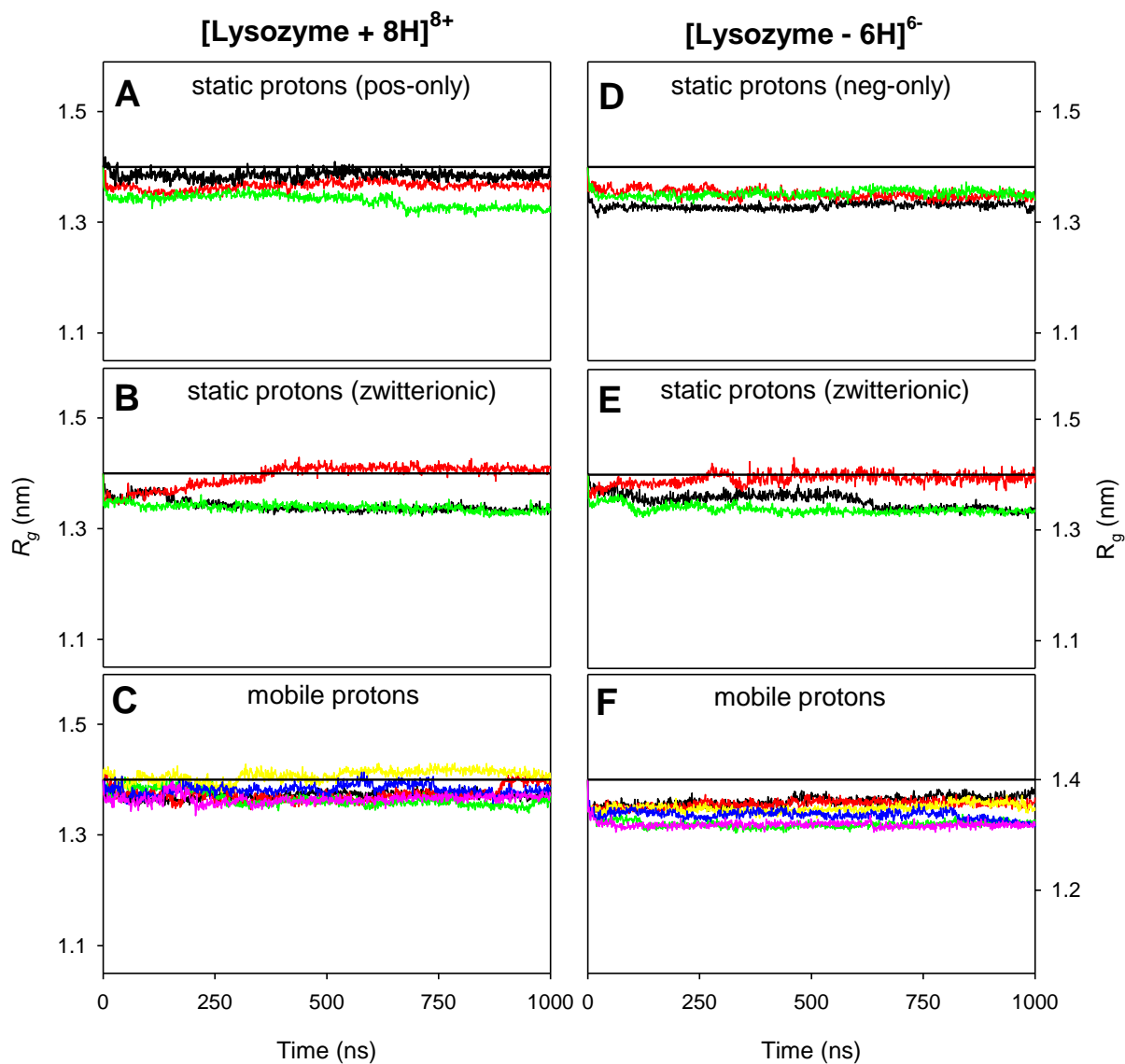
Figure 2.8. Radius of gyration (R_g) obtained in gas phase 1 µs lysozyme MD simulations in positive ion mode (A-C, 8+ charge state) and in negative (D-F, 6- charge state). Individual panels contain data from a several independent (3-6) runs. The data in the top row were obtained using stationary positive-only (A) and negative-only (D) charges. The center row data were generated using various stationary zwitterionic charge patterns (B, E). Panels C and F show the results of mobile proton MD runs.
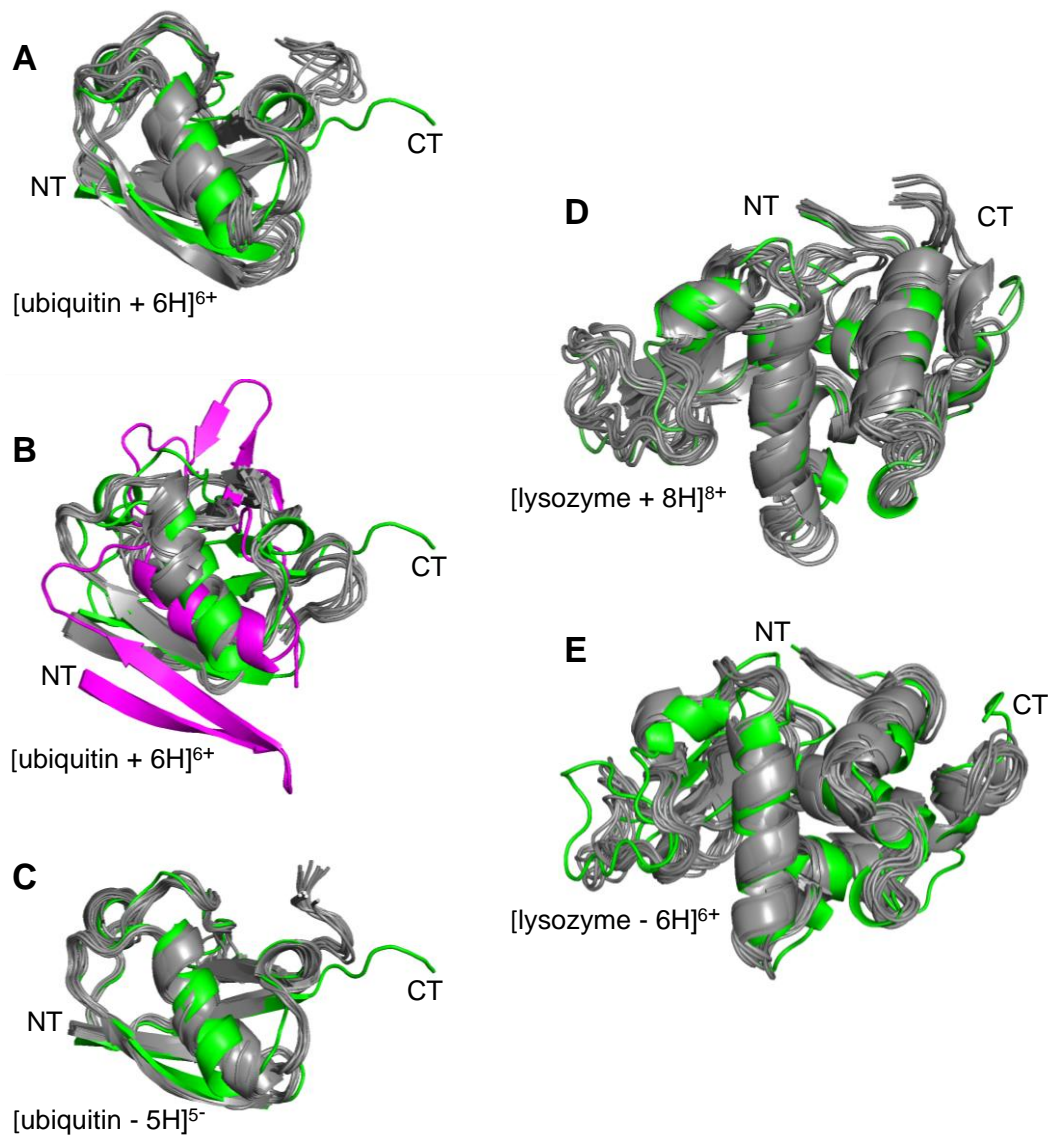
Figure 2.9. Gas phase mobile proton MD results on ubiquitin (A-C) and lysozyme (D, E). Each panel contains the X-ray structure (green), and MD snapshots representing t = 100 ns, 200 ns, … 1 μs. The magenta panel in (B) illustrates transient unfolding of [ubiquitin + 6H]$^{6+}$ at t = 100 ns (blue trajectory in Figure 2.6. C).
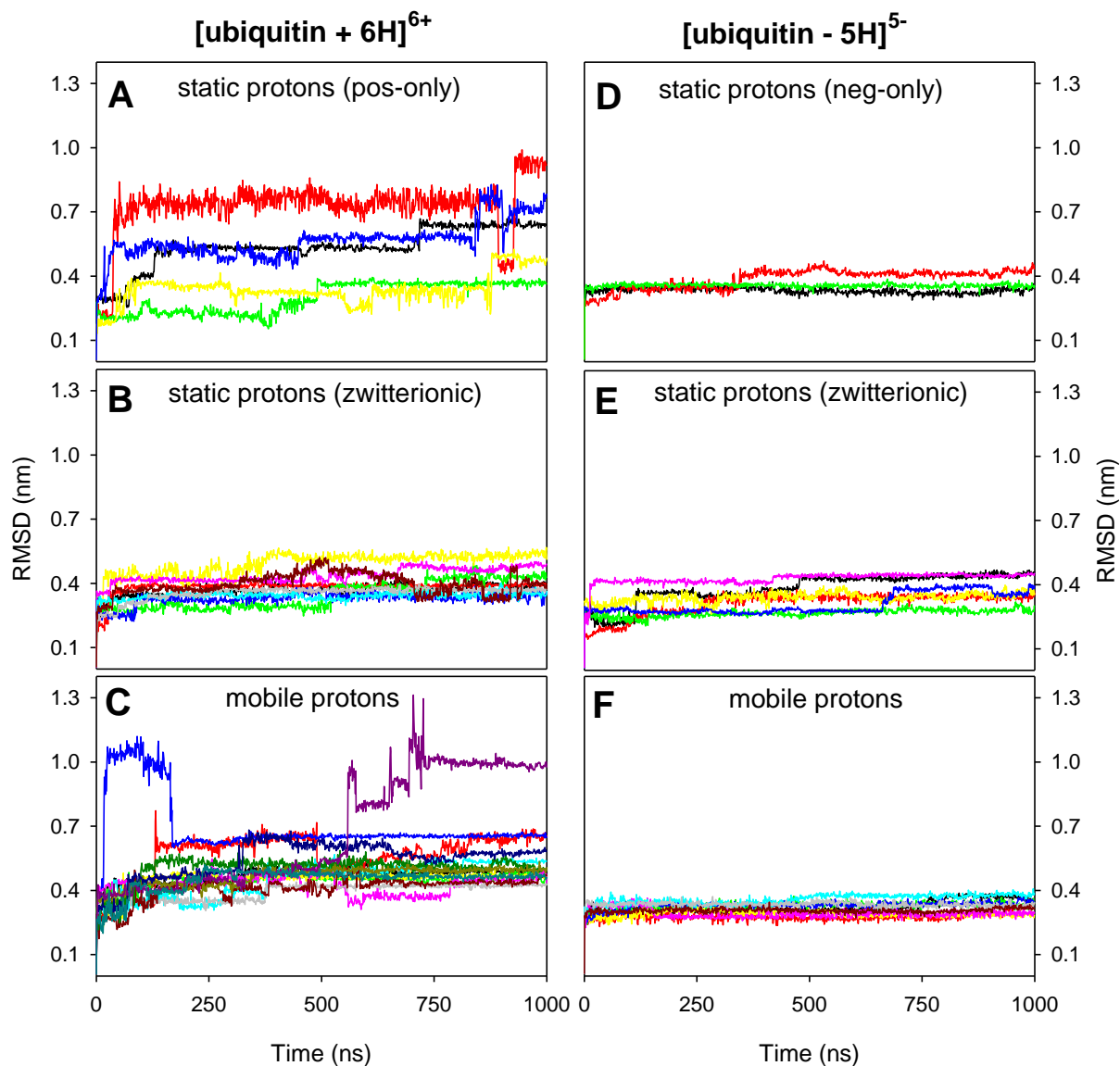
Figure 2.10. Gas phase RMSD data obtained for ubiquitin MD simulations in positive ion mode (A-C, 6+ charge state) and in negative (D-F, 5- charge state).
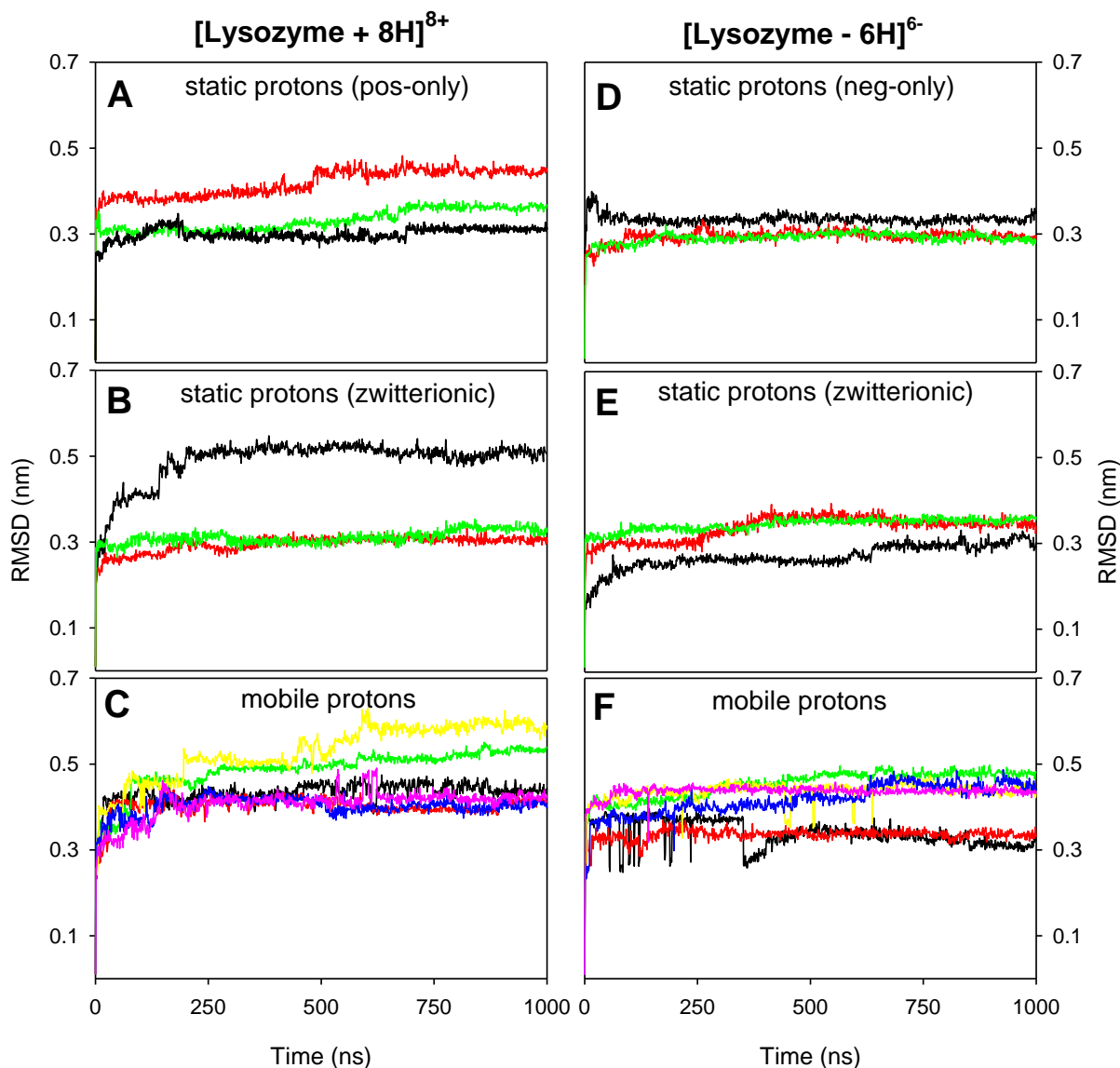
Figure 2.11. Gas phase RMSD data obtained for lysozyme MD simulations in positive ion mode (A-C, 8+ charge state) and in negative (D-F, 6- charge state).

2.3.6. Charge Patterns Produced in Mobile Proton Runs

Our mobile proton simulations used various initial charge patterns (coinciding with those of the static charge runs, Tables 2.2., 2.3.). Regardless of these starting configurations, the mobile proton runs showed a strong propensity to generate zwitterionic patterns (Figure 2.12.), i.e., R-COO$^-$ sites accumulated in $[M + zH]^{z+}$ ions, and Arg$^+$/Lys$^+$ sites accumulated in $[M - zH]^{z-}$ ions. H$^+$ migration had largely gone to completion after ~500 ns, producing fairly stable patterns (Figure 2.12.).

Shown along the right hand side of Figure 2.12. are representative 1 μs structures of ubiquitin and lysozyme. These snapshots reveal that oppositely charged sites were involved in networks of salt

bridges at the protein surface, i.e., all R-COO$^-$ moieties in [M + zH]$^{z+}$ ions were in contact with at least one positively charged sites (and vice versa for [M - zH]$^{z-}$ ions).

Arg was the preferred basic H$^+$ residence site, in accordance with its high PA_{int}. Negative charges were distributed over Asp, Glu and CT sites, reflecting their similar PA_{int} values.[84] The zwitterionic character after 1 μs was more pronounced for gaseous ubiquitin than for lysozyme (Figure 2.12.). This difference is attributed to the higher percentage of basic/acidic sites in ubiquitin which promotes the formation of Coulombically favorable salt bridge networks (Tables 2.2., 2.3.). A detailed summary of the average charge on each of the titratable sites in the mobile proton simulations is provided in Figure 2.13.

Overall, the current work marks the first time that the structural dynamics of gaseous ions was simulated on an extended (1 μs) time scale using a model that explicitly allowed for H$^+$ hopping within the gaseous protein ions. The strong preference for zwitterionic motifs seen in these simulations is consistent with experimental investigations.[49-55]
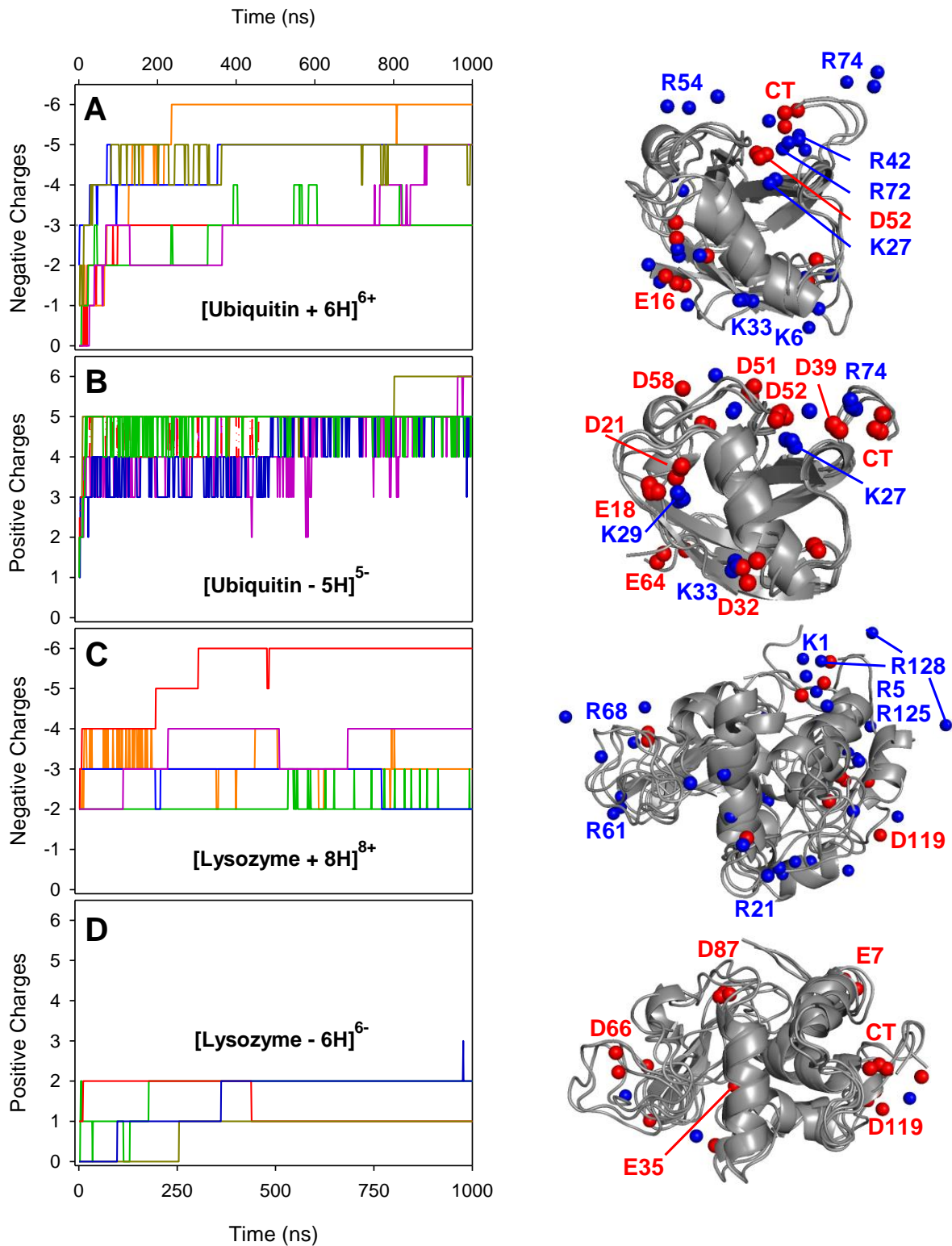
Figure 2.12. Gas phase mobile proton MD data for ubiquitin (A, B) and lysozyme (C, D) in positive and negative ion mode. Each panel shows the accumulation of "opposite" charges, i.e., the formation of salt bridges. The resulting zwitterionic structures are exemplified along the right hand side, using overlays of three representative 1 μs frames for each condition. Positive and negative charges are highlighted as blue and red spheres, respectively. Labels highlight sites that are consistently charged for all three 1 μs structures. For a detailed analysis of the zwitterionic patterns, see Figure 2.13.

Figure 2.13. Average charge on titratable sites as determined by averaging mobile proton simulation runs for (A) [ubiquitin + 6H]$^{6+}$ (not including runs that showed transient unfolding), (B) [ubiquitin - 5H]$^{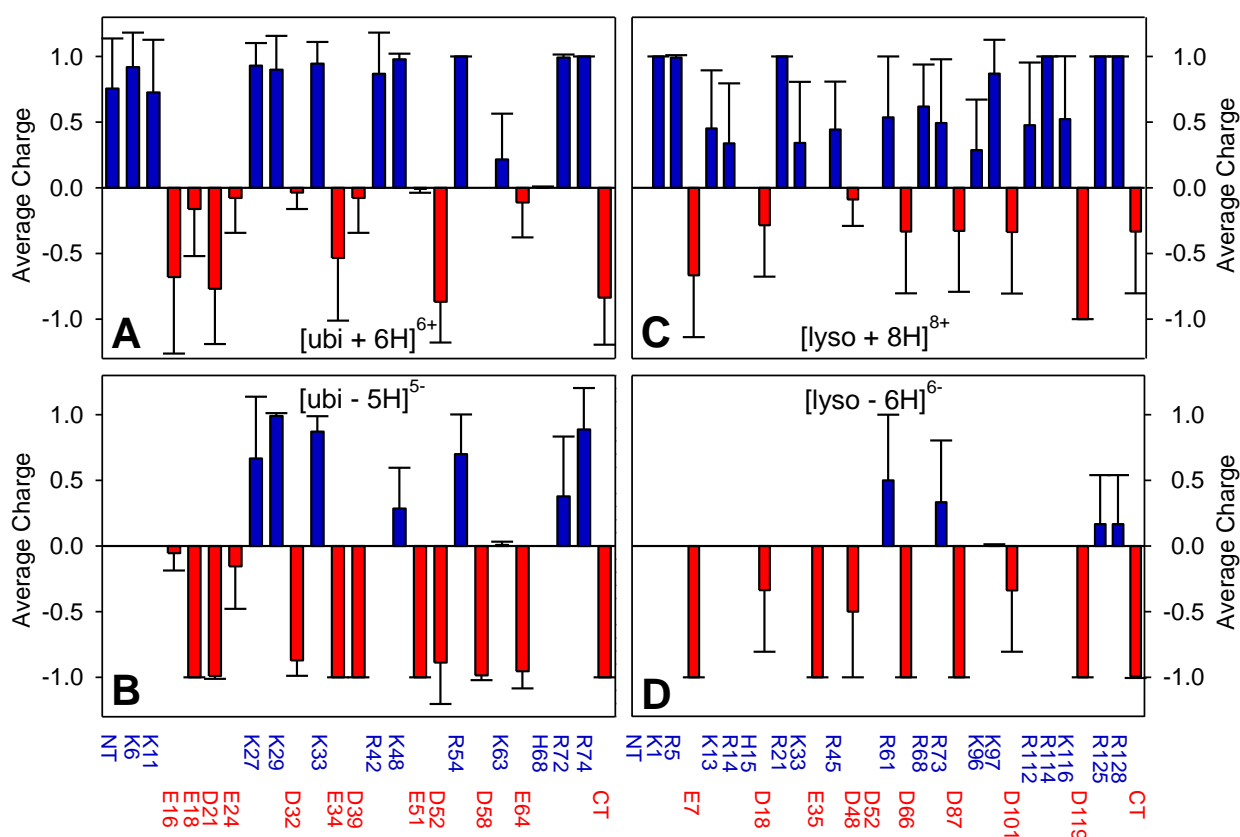5-}$, (C) [lysozyme + 8H]$^{8+}$, and (D) [lysozyme - 6H]$^{6-}$. Each colored bar represents the mean positive or negative charge with its standard deviation, averaged for the time window of 500 ns - 1 µs. The time interval between 0 and 500 ns was not considered to eliminate any possible dependence of the simulated charge patterns on the initial charge configuration.

2.3.7. Surface Side Chain Collapse

Gaseous ubiquitin and lysozyme retained an overall backbone fold close the corresponding solution/X-ray structures, as discussed above. However, inspection of Figure 2.6. reveals the gas phase R_g values for most of the simulations were 3-5% lower than for the X-ray coordinates, indicating compaction in the gas phase. To uncover the origin of this effect we calculated R_g values for various sub-groups of atoms. The most pronounced R_g drop was seen for surface side chains and, among these, the titratable sites Arg, Lys, His, Asp, Glu made the largest contribution. This side chain collapse occurred for both ubiquitin and lysozyme. This phenomenon was largely independent of the protonation model, and it took place for both [M + zH]$^{z+}$ and [M - zH]$^{z-}$ ions (Figures 2.14., 2.15.).

The origin of this side chain collapse is illustrated in Figure 2.16., which displays representative mobile proton MD structures of positive and negative ubiquitin ions. The Figure

reveals numerous intramolecular contacts that force the titratable side chains from their extended X-ray/solution structures (Figure 2.3.) into orientations close to the protein surface. In addition to salt bridges, these contacts include charge-dipole interactions where cationic sites are solvated by backbone CO groups, while R-COO$^-$ sites interact with backbone NH sites. Many contacts also involve H-bonds among uncharged titratable sites, reflecting capability of R-COOH, amino groups, and guanidine groups to act as both H-bond donors (OH, NH) and acceptors (C=O, N:). Side chain contacts similar to those discussed here were also observed in static proton/zwitterionic simulations (data not shown).

Considering that salt bridges are a major contributor to side chain collapse in zwitterionic runs, it may seem surprising that a similar protein contraction was also seen in simulations that did not allow for zwitterions (Figure 2.14., 2.15.). The reason for this effect is that even in the absence of oppositely charged moieties most of the titratable side chains formed a tightly connected network at the protein surface (Figure 2.18.). These non-zwitterionic contacts involved numerous H-bonds among neutral moieties, exemplified by the cluster of Lys27^0, Arg42^0, Glu51^0, Asp52^0, CT^0 in Figure 2.18. B. H-bonding and charge solvation involving non-titratable sites (Asn, Gln, Ser, and Thr, plus backbone NH and CO) took place under non-zwitterionic conditions as well.

The formation of extensive intramolecular contacts involving side chains, and the resulting protein compaction seen in our MD data is consistent with earlier simulations by Steinberg et al.[39] However, that earlier work[39] only tested a static proton/zwitterionic model in positive ion mode. The current study extends the work of Steinberg et al.[39] by demonstrating that similar side chain collapse takes place in both positive and negative ion mode, and under different simulation conditions (with or without zwitterions, and with or without mobile protons).

Figure 2.14. Representative gas phase MD results obtained for ubiquitin in positive (A-C) and negative ion mode (D-F) with different protonation modes. Each panel shows the radius of gyration (R_g) for titratable side chains, titratable and polar side chains, and for the whole protein. Horizontal lines represent the corresponding crystal R_g values.
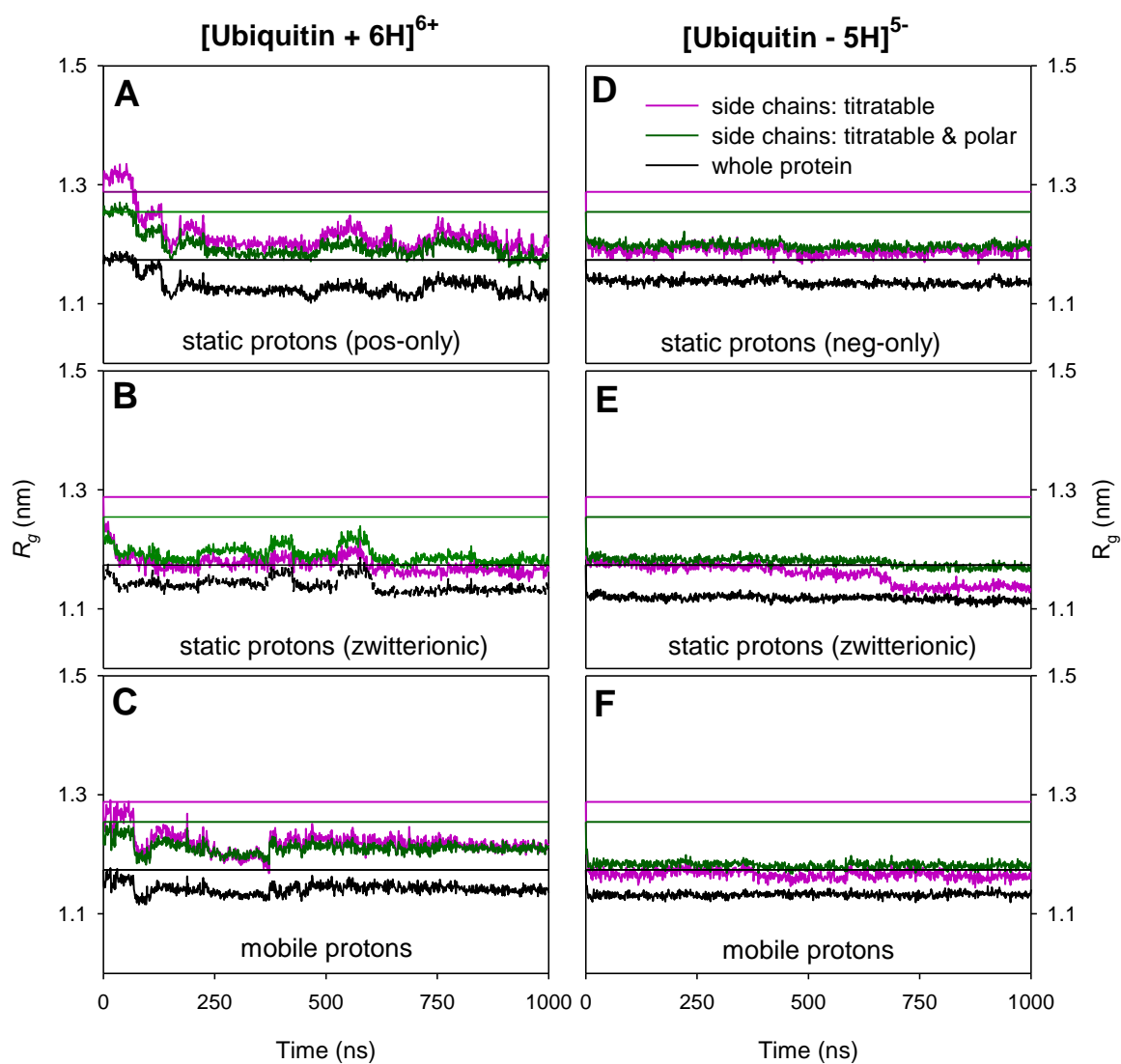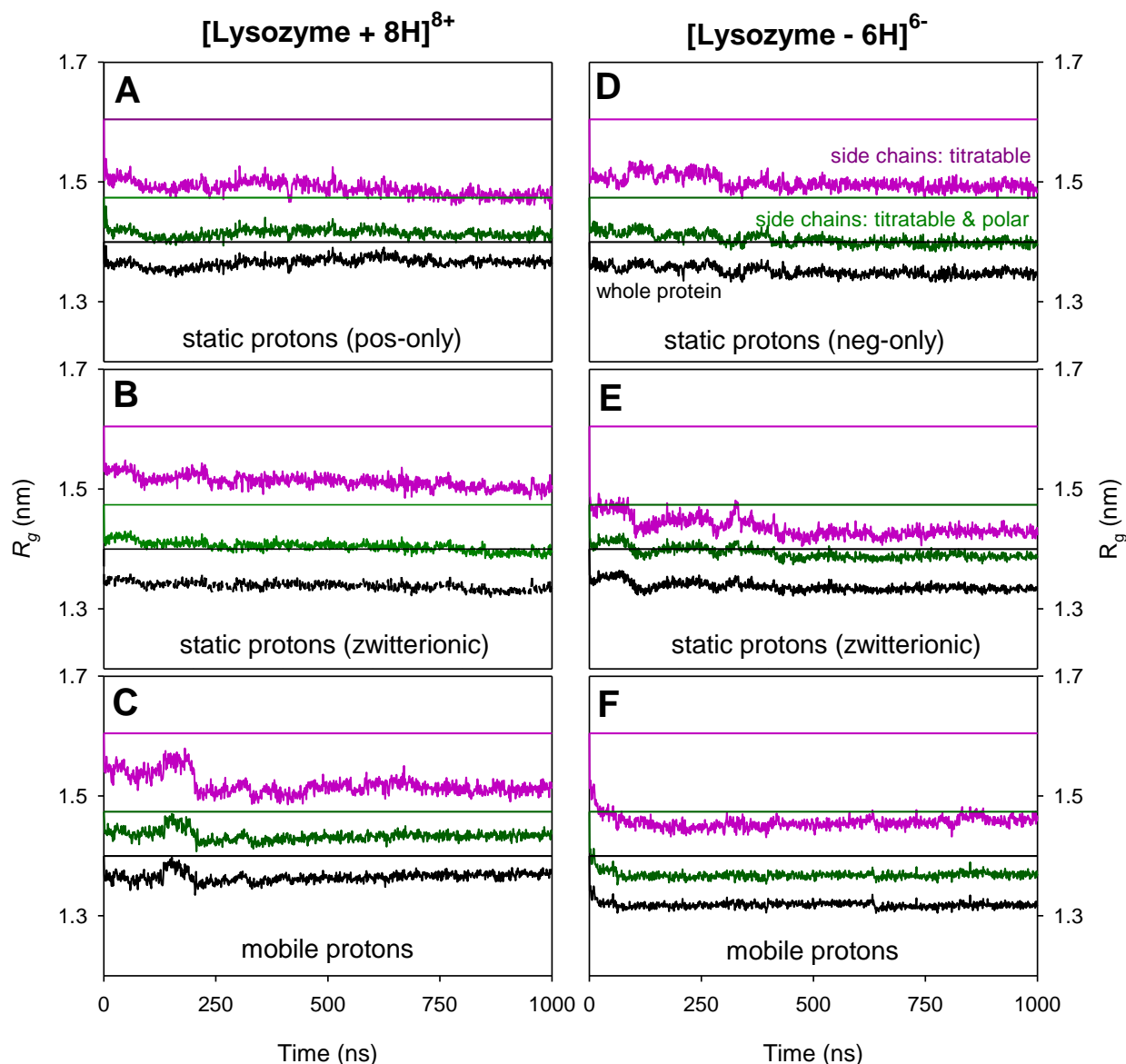
Figure 2.15. Representative gas phase MD results obtained for lysozyme in positive ion mode (A-C) and negative ion mode (D-F) with different protonation modes. Each panel shows the radius of gyration (R_g) for titratable side chains, titratable and polar side chains, and for the whole protein. Horizontal lines represent the corresponding crystal structure R_g values.
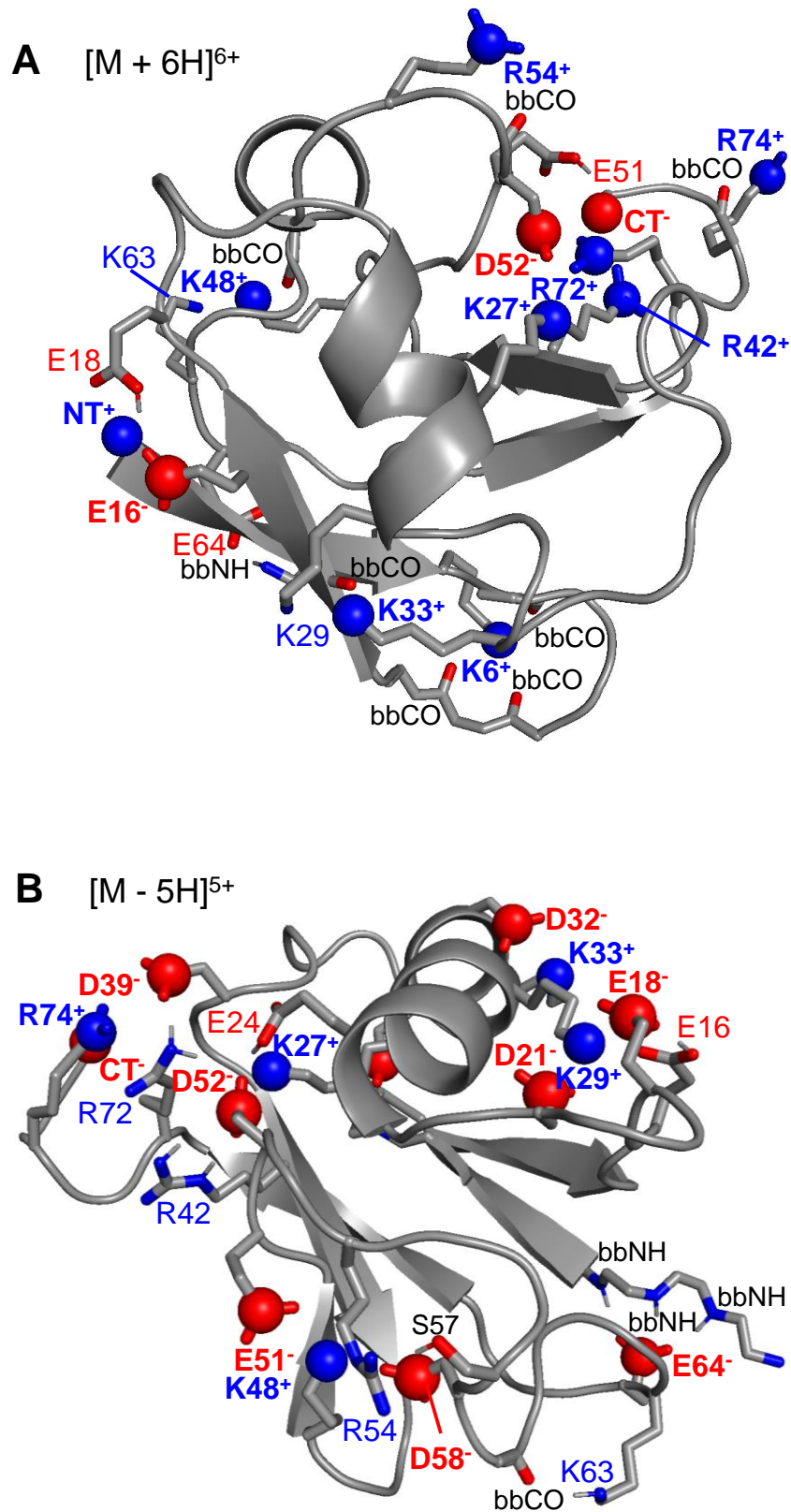
Figure 2.16. Packing of titratable/polar side chains in $t = 1$ μs mobile proton MD structures of ubiquitin in (A) positive and (B) negative ion mode. Cationic and anionic sites are highlighted as blue and red spheres. Selected N and O atoms are indicated as blue and red sticks, respectively. bbCO and bbNH denote backbone CO and NH groups. Not all titratable/polar moieties are shown to reduce cluttering.

2.3.8. MD and Experimental Ω Values

Figure 2.17. compares MD Ω values with the corresponding experiments. Ω values were also calculated for crystal and solution MD structures (after water removal). Like Figure 4, Figure 8 contains five panels because [ubiquitin + 6H]$^{6+}$ appeared in two forms, tightly folded and semi-unfolded. The following discussion initially focuses on folded species, excluding those ubiquitin trajectories that showed conformational changes (Figure 2.17. A, C-E). The following key points are apparent from these data:

(1) Crystal structures and solution MD conformers share virtually the same Ω values. This observation reinforces the view that the crystal coordinates of ubiquitin and lysozyme provide an adequate representation of their solution behavior, as highlighted in Figure 2.3.

(2) Experimental Ω values for electrosprayed ubiquitin and lysozyme are significantly smaller than expected from their crystal structures, revealing that the proteins become more compact upon transitioning from solution into the gas phase. This effect has previously been reported for a range of proteins in positive ion mode.[28, 33, 41-43] The current study reveals that the same effect also takes place in negative ion mode (Figure 2.17 C, E). For the proteins studied here this collapse is most prevalent for [ubiquitin + 6H]$^{6+}$ (~16%, Figure 2.17. A), and it is least pronounced for [lysozyme -6H]$^{6-}$ (~3.5%, Figure 2.17. E).

(3) As outlined in Figure 2.16., gas phase protein contraction is caused by titratable and polar side chains at the protein surface. After desolvation these side chains collapse from their initially extended crystal/solution structures onto the protein surface where they form tight networks involving salt bridges, charge-dipole contacts, and H-bonds. The occurrence of such collapse events has previously been proposed on the basis of positive ion zwitterionic MD runs.[39] The current work generalizes this finding by demonstrating that compaction takes place in both polarities, and for three different charge models.

(4) It is remarkable that the protein compaction in our simulations is virtually independent of the charge model used. All three models yielded Ω values that agreed closely with each other (Figures 2.17. A, C-E). We initially expected that the extent of side chain collapse would depend on the charge model used. However, our data show that different types of contacts (e.g. mostly salt bridged vs. primarily H-bonded, Figures 2.16., 2.18.) cause very similar side chain compaction. Gratifyingly, the collapsed MD-generated gas phase structures had Ω values close to the experimentally overserved data for both proteins and in both polarities.

(5) In our experiments, only [ubiquitin + 6H]$^{6+}$ showed the presence of semi-unfolded species – evident from the satellite signal at ~1140 Å^2 (Figure 2.1. B). Consistent with this experimental

observation, only MD runs for [ubiquitin + 6H]$^{6+}$ exhibited occasional transitions to semi-unfolded conformers (Figure 2.6. A, C). The Ω values of these non-native MD structures (with $R_g > 1.3$ nm) agreed closely with the experimental data (Figure 2.17. B). In other words, non-compact conformers such as the magenta species in Figure 4B represent suitable candidate structures for the ~1140 Å^2 signal seen in the IMS data of [ubiquitin + 6H]$^{6+}$ (Figure 2.1. B). The unfolding/refolding transitions of [ubiquitin + 6H]$^{6+}$ would likely have gone undetected in the much shorter (ns) simulation windows used in earlier studies.
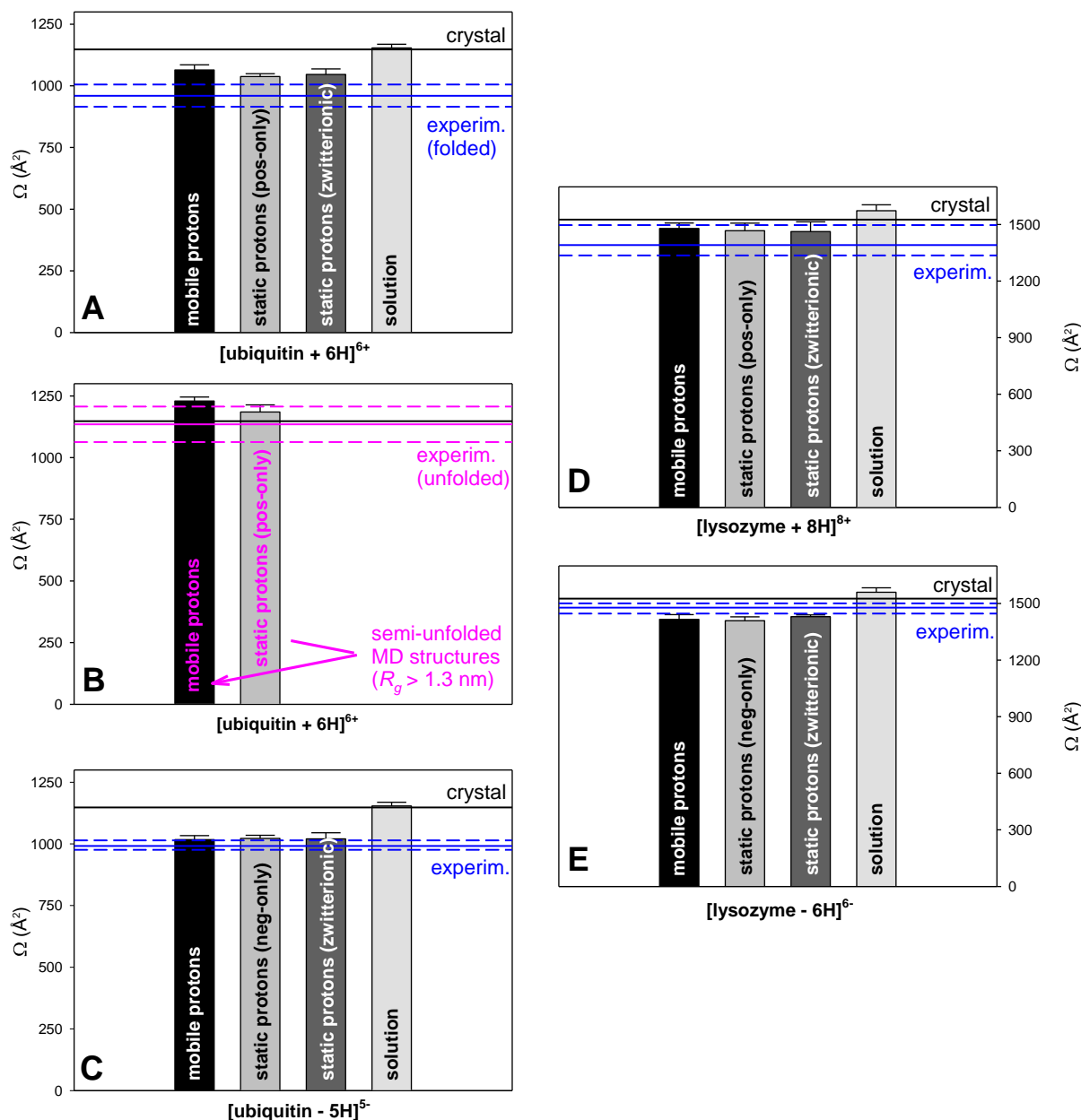
Figure 2.17. Collision cross sections (Ω) from experiments and MD simulations. Vertical bars represent gas phase and solution MD Ω values. The first two panels separately display data for (A) folded and (B) semi-unfolded [ubiquitin + 6H]$^{6+}$. Colored horizontal lines represent experimental average Ω maxima, dashed lines indicate the corresponding FWHMs. Horizontal black lines represent crystal Ω values.
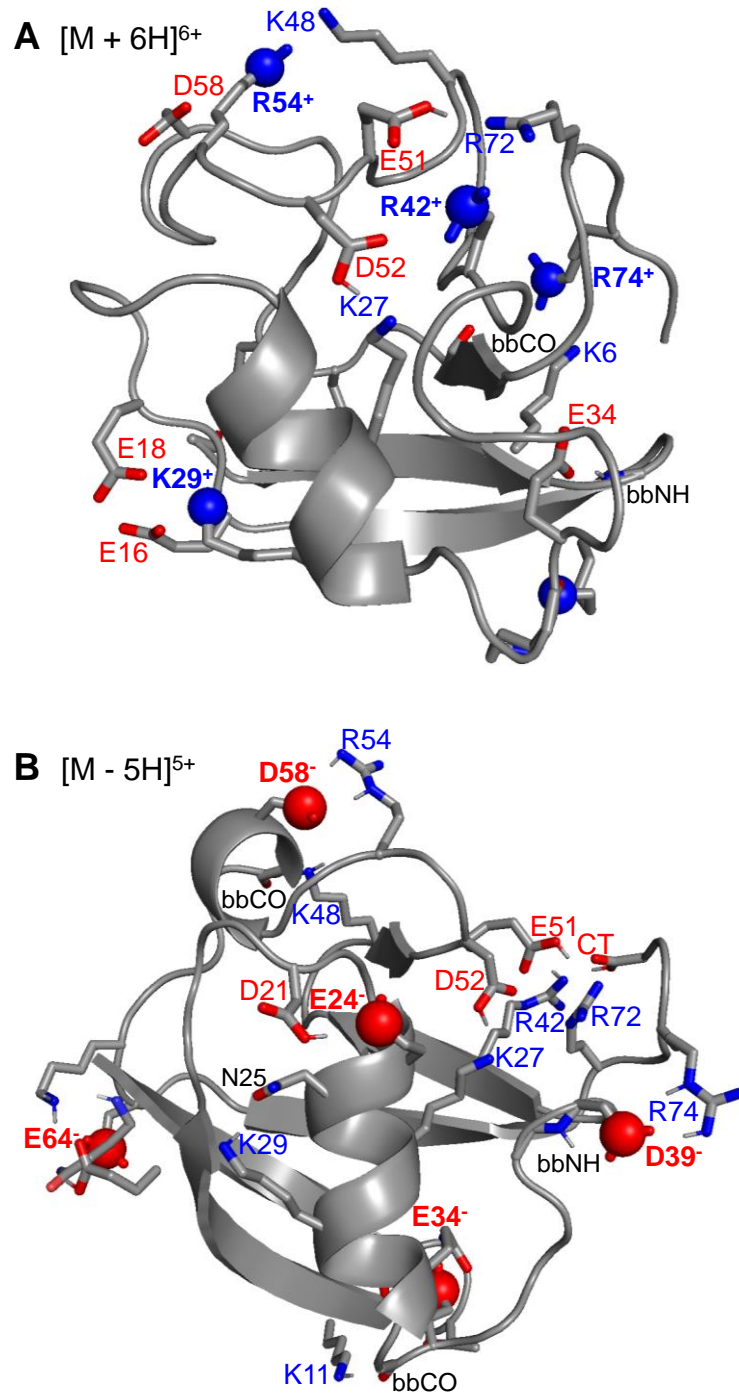
Figure 2.18. Packing of titratable/polar side chains, seen in $t = 1$ µs MD structures for ubiquitin obtained using static protons in (A) positive-only and (B) negative-only simulations. Positive and negative charges are highlighted as blue and red spheres. A number of neutral side chains are shown as well. Selected N and O atoms are indicated as blue and red sticks, respectively. bbCO and bbNH denote backbone CO and NH groups. Not all titratable/polar moieties are shown to reduce cluttering.

2.4.   Conclusions

This work conducted a comprehensive comparison of MD-generated gas phase protein structures with collision cross sections from native ESI experiments. We systematically studied the behavior of three different charge models for two proteins under both polarity conditions in an extended (1 μs) MD time interval. Most of the trajectories revealed the persistence of tightly folded gas phase conformers, with backbone conformers that closely resembled the initial solution conformations. It has been suggested that the thermodynamically stable state of gaseous proteins corresponds to "inside out" conformers, where nonpolar residues are exposed to the surface, while polar and titratable side chains are buried in the interior.[87] The persistence of proper "outside out" conformers (with charged/polar sites on the surface) in our simulations is consistent with the view that solution-like structures can survive after ESI due to kinetic trapping.[14, 37, 38]

Overall, the current work highlights the potential, but also the limitations of using gas phase investigations for deducing structural features of proteins in solution. We found the backbone fold to be largely retained (with the exception of flexible termini), while surface side chains underwent substantial changes in their orientations and interactions. It is hoped that advances in hardware and software will soon extend gas phase MD time scales even further, such that it will become possible to probe kinetic trapping and structural retention in simulations that match the time frame (tens of milliseconds) of typical IMS experiments.

2.5.   References

1.   Heck, A. J. R.; Van den Heuvel, R. H. H. Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom. Rev.* **2004,** *23*, 368-389.

2.   Mehmood, S.; Allison, T. M.; Robinson, C. V. Mass spectrometry of protein complexes: From origins to applications. *Annu. Rev. Phys. Chem.* **2015,** *66*, 453-474.

3.   Benesch, J. L. P.; Ruotolo, B. T. Mass spectrometry: Come of age for structural and dynamical biology. *Curr. Op. Struct. Biol.* **2011,** *21*, 641-649.

4.   Chen, S.-H.; Russell, D. H. How closely related are conformations of protein ions sampled by im-ms to native solution structures? *J. Am. Soc. Mass Spectrom.* **2015,** *26*, 1433-1443.

5.   Kaddis, C. S.; Loo, J. A. Native protein ms and ion mobility: Large flying proteins with esi. *Anal. Chem.* **2007,** *79*, 1779-1784.

6.   Quintyn, R. S.; Zhou, M.; Yan, J.; Wysocki, V. H. Surface-induced dissociation mass spectra as a tool for distinguishing different structural forms of gas-phase multimeric protein complexes. *Anal. Chem.* **2015,** *87*, 11879-11886.

7.   Zhang, H.; Wen, J.; Blankenship, R. E.; Gross, M. L. Native electrospray and electron-capture dissociation fticr mass spectrometry for top-down studies of protein assemblies. *Anal. Chem.* **2011,** *83*, 5598-5606.

8.   Morrison, L. J.; Brodbelt, J. S. Charge site assignment in native proteins by ultraviolet photodissociation (uvpd) mass spectrometry. *Analyst* **2016,** *141*, 166-176.

9.   Seo, J.; Hoffmann, W.; Warnke, S.; Huang, X.; Gewinner, S.; Schollkopf, W.; Bowers, M. T.; von Helden, G.; Pagel, K. An infrared spectroscopy approach to follow beta-sheet formation in peptide amyloid assemblies. *Nat. Chem.* **2017,** *9*, 39-44.

10.   Frankevich, V.; Barylyuk, K.; Chingin, K.; Nieckarz, R.; Zenobi, R. Native biomolecules in the gas phase? The case of green fluorescent protein. *ChemPhysChem.* **2013,** *14*, 929-935.

11.   Czar, M. F.; Zosel, F.; Koenig, I.; Nettels, D.; Wunderlich, B.; Schuler, B.; Zarrine-Afsar, A.; Jockusch, R. A. Gas-phase fret efficiency measurements to probe the conformation of mass-selected proteins. *Anal. Chem.* **2015,** *87*, 7559-7565.

12.   Mistarz, U. H.; Brown, J. M.; Haselmann, K. F.; Rand, K. D. Probing the binding interfaces of protein complexes using gas-phase h/d exchange mass spectrometry. *Structure* **2016,** *24*, 310-318.

13.   Skinner, O. S.; McLafferty, F. W.; Breuker, K. How ubiquitin unfolds after transfer into the gas phase. *J. Am. Soc. Mass Spectrom.* **2012,** *23*, 1011-1014.

14.   Wyttenbach, T.; Bowers, M. T. Structural stability from solution to the gas phase: Native solution structure of ubiquitin survives analysis in a solvent-free ion mobility–mass spectrometry environment. *J. Phys. Chem. B* **2011,** *115*, 12266-12275.

15.   Ly, T.; Julian, R. R. Elucidating the tertiary structure of protein ions in vacuo with site specific photoinitiated radical reactions. *J. Am. Chem. Soc.* **2010,** *132*, 8602-8609.

16.   Breuker, K.; McLafferty, F. W. Stepwise evolution of protein native structure with electrospray into the gas phase, 10^{-12} to 10^2 s. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105*, 18145-18152.

17.   Seo, J.; Hoffmann, W.; Warnke, S.; Bowers, M. T.; Pagel, K.; von Helden, G. Retention of native protein structures in the absence of solvent: A coupled ion mobility and spectroscopic study. *Angew. Chem.-Int. Edit.* **2016,** *55*, 14173-14176.

18.   Badman, E. R.; Hoaglund-Hyzer, C. S.; Clemmer, D. E. Monitoring structural changes of protein in and ion trap over ~ 10-200 ms: Unfolding transitions in cytochrome *c* ions. *Anal. Chem.* **2001,** *73*, 6000-6007.

19.   Bich, C.; Baer, S.; Jecklin, M. C.; Zenobi, R. Probing the hydrophobic effect of noncovalent complexes by mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2010,** *21*, 286-289.

20.   Devine, P. W. A.; Fisher, H. C.; Calabrese, A. N.; Whelan, F.; Higazi, D. R.; Potts, J. R.; Lowe, D. C.; Radford, S. E.; Ashcroft, A. E. Investigating the structural compaction of biomolecules upon transition to the gas-phase using esi-twims-ms. *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 1855-1862.

21.   Leney, A. C.; Heck, A. J. R. Native mass spectrometry: What is in the name? *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 5-13.

22.   Wang, W.; Kitova, E. N.; Klassen, J. S. Influence of solution and gas phase processes on protein-carbohydrate binding affinities determined by nanoelectrospray fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **2003,** *75*, 4945-4955.

23.   Gavriilidou, A. F. M.; Holding, F. P.; Mayer, D.; Coyle, J. E.; Veprintse, D. B.; Zenobi, R. Native mass spectrometry gives insight into the allosteric binding mechanism of m2 pyruvate kinase to fructose-1,6-bisphosphate. *Biochemistry* **2018,** *57*, 1685-1689.

24.   Konermann, L.; Metwally, H.; McAllister, R. G.; Popa, V. How to run molecular dynamics simulations on electrospray droplets and gas phase proteins: Basic guidelines and selected applications. *Methods* **2018,** *144*, 104-112.

25.   Porrini, M.; Rosu, F.; Rabin, C.; Darre, L.; Gomez, H.; Orozco, M.; Gabelica, V. Compaction of duplex nucleic acids upon native electrospray mass spectrometry. *ACS Central Sci.* **2017,** *3*, 454–461.

26.   Bonvin, G.; Bobst, C. E.; Kaltashov, I. A. Interaction of transferrin with non-cognate metals studied by native electrospray ionization mass spectrometry. *Int. J. Mass Spectrom.* **2017,** *420*, 74-82.

27.   Gabelica, V.; Livet, S.; Rosu, F. Optimizing native ion mobility q-tof in helium and nitrogen for very fragile noncovalent structures. *J. Am. Soc. Mass Spectrom.* **2018,** *29*, 2189-2198.

28.   Hopper, J. T. S.; Oldham, N. J. Collision induced unfolding of protein ions in the gas phase studied by ion mobility-mass spectrometry: The effect of ligand binding on conformational stability. *J. Am. Soc. Mass Spectrom.* **2009,** *20*, 1851-1858.

29.   Dill, K. A.; MacCallum, J. L. The protein-folding problem, 50 years on. *Science* **2012,** *338*, 1042-1046.

30.   Brini, E.; Fennell, C. J.; Fernandez-Serra, M.; Hribar-Lee, B.; Luksic, M.; Dill, K. A. How water's properties are encoded in its molecular structure and energies. *Chem. Rev.* **2017,** *117*, 12385-12414.

31.   Liu, L.; Bagal, D.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S. Hydrophobic protein-ligand interactions preserved in the gas phase. *J. Am. Chem. Soc.* **2009,** *131*, 15980-15981.

32.   Wang, G.; Cole, R. B. Disparity between solution-phase equilibria and charge state distributions in positive-ion electrospray mass spectrometry. *Org. Mass Spectrom.* **1994,** *29*, 419-427.

33.   Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. Protein structure in vacuo: The gas-phase conformation of bpti and cytochrome c. *J. Am. Chem. Soc.* **1997,** *119*, 2240-2248.

34.   Hall, Z.; Politis, A.; Bush, M. F.; Smith, L. J.; Robinson, C. V. Charge-state dependent compaction and dissociation of protein complexes: Insights from ion mobility and molecular dynamics. *J. Am. Chem. Soc.* **2012,** *134*, 3429-3438.

35.   Laszlo, K. J.; Munger, E. B.; Bush, M. F. Folding of protein ions in the gas phase after cation-to-anion proton-transfer reactions. *J. Am. Chem. Soc.* **2016,** *138*, 9581-9588.

36.   Schnier, P. D.; Gross, D. S.; Williams, E. R. Electrostatic forces and dielectric polarizability of multiply protonated gas-phase cytochrome c ions probed by ion/molecule chemistry. *J. Am. Chem. Soc.* **1995,** *117*, 6747-6757.

37.   Hendricks, N. G.; Julian, R. R. Leveraging ultraviolet photodissociation and spectroscopy to investigate peptide and protein three-dimensional structure with mass spectrometry. *Analyst* **2016,** *141*, 4534-4540.

38.   Clemmer, D. E.; Russell, D. H.; Williams, E. R. Characterizing the conformationome: Toward a structural understanding of the proteome. *Accounts Chem. Res.* **2017,** *50*, 556-560.

39.   Steinberg, M. Z.; Elber, R.; McLafferty, F. W.; Gerber, R. B.; Breuker, K. Early structural evolution of native cytochrome c after solvent removal. *ChemBioChem* **2008,** *9*, 2417-2423.

40.   Warnke, S.; von Helden, G.; Pagel, K. Protein structure in the gas phase: The influence of side-chain microsolvation. *J. Am. Chem. Soc.* **2013,** *135*, 1177-1180.

41.   Jurneczko, E.; Barran, P. E. How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase. *Analyst* **2011,** *136*, 20-28.

42.   Shi, H.; Atlasevich , N.; Merenbloom, S. I.; Clemmer, D. E. Solution dependence of the collisional activation of ubiquitin [m + 7h]7+ ions. *J. Am. Soc. Mass Spectrom.* **2014,** *25*, 2000-2008.

43.   Scarff, C. A.; Thalassinos, K.; Hilton, G. R.; Scrivens, J. H. Travelling wave ion mobility mass spectrometry studies of protein structure: Biological significance and comparison with x-ray crystallography and nuclear magnetic resonance spectroscopy measurements. *Rapid Commun. Mass Spectrom.* **2008,** *22*, 3297-3304.

44.   Pacholarz, K. J.; Barran, P. E. Distinguishing loss of structure from subunit dissociation for protein complexes with variable temperature ion mobility mass spectrometry. *Anal. Chem.* **2015,** *87*, 6271-6279.

45.   Hogan, C. J.; Ruotolo, B. T.; Robinson, C. V.; de la Mora, J. F. Tandem differential mobility analysis-mass spectrometry reveals partial gas-phase collapse of the groel complex. *J. Phys. Chem. B* **2011,** *115*, 3614-3621.

46.   Michaelevski, I.; Eisenstein, M.; Sharon, M. Gas-phase compaction and unfolding of protein structures. *Anal. Chem.* **2010,** *82*, 9484-9491.

47.   Gabelica, V.; Marklund, E. Fundamentals of ion mobility spectrometry. *Curr. Op. Chem. Biol.* **2017,** *42*, 51-59.

48.   Mao, Y.; Woenckhaus, J.; Kolafa, J.; Ratner, M. A.; Jarrold, M. F. Thermal unfolding of unsolvated cytochrome c: Experiment and molecular dynamics simulations. *J. Am. Chem. Soc.* **1999,** *121*, 2712-2721.

49.   Li, J.; Santambrogio, C.; Brocca, S.; Rossetti, G.; Carloni, P.; Grandori, R. Conformational effects in protein electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **2016,** *35*, 111-122.

50.   Yoo, H. J.; Wang, N.; Zhuang, S. Y.; Song, H. T.; Hakansson, K. Negative-ion electron capture dissociation: Radical-driven fragmentation of charge-increased gaseous peptide anions. *J. Am. Chem. Soc.* **2011,** *133*, 16790-16793.

51.   Forbes, M. W.; Bush, M. F.; Polfer, N. C.; Oomens, J.; Dunbar, R. C.; Williams, E. R.; Jockusch, R. A. Infrared spectroscopy of arginine cation complexes: Direct observation of gas-phase zwitterions. *J. Phys. Chem. A* **2007,** *111*, 11759-11770.

52.   Ogorzalek Loo, R. R.; Loo, J. A. Salt bridge rearrangement (sabre) explains the dissociation behavior of noncovalent complexes. *J. Am. Soc. Mass Spectrom.* **2016,** *27*, 975-990.

53.   Zhang, Z.; Browne, S. J.; Vachet, R. W. Exploring salt bridge structures of gas-phase protein ions using multiple stages of electron transfer and collision induced dissociation. *J. Am. Soc. Mass Spectrom.* **2014,** *25*, 604-613.

54.   Breuker, K.; Brüschweiler, S.; Tollinger, M. Electrostatic stabilization of a native protein structure in the gas phase. *Angew. Chem. Int. Ed.* **2011,** *50*, 873-877.

55.   Bonner, J. G.; Lyon, Y. A.; Nellessen, C.; Julian, R. R. Photoelectron transfer dissociation reveals surprising favorability of zwitterionic states in large gaseous peptides and proteins. *J. Am. Chem. Soc.* **2017,** *139*, 10286-10293.

56.   Marchese, R.; Grandori, R.; Carloni, P.; Raugei, S. On the zwitterionic nature of gas-phase peptides and protein ions. *PLoS Comput. Biol.* **2010,** *6*.

57.   Konermann, L. Molecular dynamics simulations on gas-phase proteins with mobile protons: Inclusion of all-atom charge solvation. *J. Phys. Chem. B* **2017,** *121*, 8102-8112.

58.   Fegan, S. K.; Thachuk, M. A charge moving algorithm for molecular dynamics simulations of gas-phase proteins. *J. Chem. Theory Comput.* **2013,** *9*, 2531-2539.

59.   Shelimov, K. B.; Jarrold, M. F. Conformations, unfolding, and refolding of apomyoglobin in vacuum: An activation barrier for gas-phase protein folding. *J. Am. Chem. Soc.* **1997,** *119*, 2987-2994.

60.   Patriksson, A.; Adams, C. M.; Kjeldsen, F.; Zubarev, R. A.; van der Spoel, D. A direct comparison of protein structure in the gas and solution phase: The trp-cage. *J. Phys. Chem. B* **2007,** *111*, 13147-13150.

61.   Dongré, A. R.; Jones, J. L.; Somogyi, Á.; Wysocki, V. H. Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model. *J. Am. Chem. Soc.* **1996,** *118*, 8365-8374.

62.   Boyd, R. K.; Somogyi, Á. The mobile proton hypothesis in fragmentation of protonated peptides: A perspective. *J. Am. Soc. Mass Spectrom.* **2010,** *21*, 1275-1278.

63.   Li, J. Y.; Lyu, W. P.; Rossetti, G.; Konijnenberg, A.; Natalello, A.; Ippoliti, E.; Orozco, M.; Sobott, F.; Grandori, R.; Carloni, P. Proton dynamics in protein mass spectrometry. *J. Phys. Chem. Lett.* **2017,** *8*, 1105-1112.

64.   Cautereels, J.; Blockhuys, F. Quantum chemical mass spectrometry: Verification and extension of the mobile proton model for histidine. *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 1227-1235.

65.   Carugo, O.; Blatova, O. A.; Medrish, E. O.; Blatov, V. A.; Proserpio, D. M. Packing topology in crystals of proteins and small molecules: A comparison. *Sci Rep* **2017,** *7*.

66.   Kowalski, J. A.; Liu, K.; Kelly, J. W. Nmr solution structure of the isolated apo pin1 ww domain: Comparison to the x-ray crystal structures of pin1. *Biopolymers* **2002,** *63*, 111-121.

67.   Vetter, S. W.; Leclerc, E. Novel aspects of calmodulin target recognition and activation. *Eur. J. Biochem.* **2003,** *270*, 404-414.

68.   Igumenova, T. I.; Frederick, K. K.; Wand, A. J. Characterization of the fast dynamics of protein amino acid side chains using nmr relaxation in solution. *Chem. Rev.* **2006,** *106*, 1672-1699.

69.   Carugo, O.; Argos, P. Protein-protein crystal-packing contacts. *Protein Science* **1997,** *6*, 2261-2263.

70.   Luo, J.; Liu, Z.; Guo, Y.; Li, M. A structural dissection of large protein-protein crystal packing contacts. *Sci Rep* **2015,** *5*.

71.   Billeter, M.; Wagner, G.; Wuthrich, K. Solution nmr structure determination of proteins revisited. *J. Biomol. NMR* **2008,** *42*, 155-158.

72.   Sun, Y.; Vahidi, S.; Sowole, M. A.; Konermann, L. Protein structural studies by traveling wave ion mobility spectrometry: A critical look at electrospray sources and calibration issues. *J. Am. Soc. Mass Spectrom.* **2016,** *27*, 31-40.

73.   Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1–2*, 19-25.

74.   Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of ubiquitin refined at 1.8 a resolution. *J. Mol. Biol.* **1987,** *194*, 531-544.

75.   Artymiuk, P. J.; Blake, C. C. F.; Rice, D. W.; Wilson, K. S. The structures of the monoclinic and orthorhombic forms of hen egg-white lysozyme at 6 angstroms resolution. *Acta Crystallogr. Sect. B-Struct. Commun.* **1982,** *38*, 778-783.

76.   Huang, J.; MacKerell, A. D. Charmm36 all-atom additive protein force field: Validation based on comparison to nmr data. *J. Comput. Chem.* **2013,** *34*, 2135-2145.

77.   Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110*, 5915-5920.

78.   McAllister, R. G.; Konermann, L. Challenges in the interpretation of protein h/d exchange data: A molecular dynamics simulation perspective. *Biochemistry* **2015,** *54*, 2683-2692.

79.   Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001,** *105*, 6474-6487.

80.   Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996,** *118*, 11225-11236.

81.   Fegan, S. K.; Thachuk, M. Suitability of the martini force field for use with gas-phase protein complexes. *J. Chem. Theory Comput.* **2012,** *8*, 1304-1313.

82.   Marchese, R.; Grandori, R.; Carloni, R.; Raugei, S. A computational model for protein ionization by electrospray based on gas-phase basicity. *J. Am. Soc. Mass Spectrom.* **2012,** *23*, 1903-1910.

83.   Cadenas, E.; Boveris, A.; Chance, B. Low-level chemiluminescence of hydroperoxide-supplemented cytochrome c. *Biochem. J.* **1980,** *187*, 131-140.

84.   Moser, A.; Range, K.; York, D. M. Accurate proton affinity and gas-phase basicity values for molecules important in biocatalysis. *J. Phys. Chem. B* **2010,** *114*, 13911-13921.

85.   Allen, S. J.; Schwartz, A. M.; Bush, M. F. Effects of polarity on the structures and charge states of native-like proteins and protein complexes in the gas phase. *Anal. Chem.* **2013,** *85*, 12055−12061.

86.   Kumar, S.; Nussinov, R. Salt bridge stability in monomeric proteins. *J. Mol. Biol.* **1999,** *293*, 1241-1255.

87.   Wolynes, P. G. Biomolecular folding in vacuo!!!(?). *Proc. Natl. Acad. Sci. U.S.A.* **1995,** *92*, 2426-2427.

Chapter 3

3.   Conclusions and Future Work

3.1.   Conclusions

Proteins are biological macromolecules that perform a variety of functions including catalysis, signaling, transport, and so forth. The fully folded, biologically active structure of a protein which is called the native structure, results from the interplay of various intra- and intermolecular interactions such as the hydrophobic effect, hydrogen bonding, van der Waals forces, and electrostatic interactions.[1-3] A variety of techniques can be used to study protein structures. Mass spectrometry (MS) has revolutionized life sciences by deciphering protein complexes and protein interaction networks in biological systems.[4-5] ESI allows the production of intact gaseous multiply charged ions.[4] The question to what extent these $[M + zH]^{z+}$ and $[M - zH]^{z-}$ ions retain solution-like conformations under "native" ESI conditions remains a matter of debate. One of the key questions in this context is the internal charge distribution of electrosprayed ions. All these issues are difficult to tackle experimentally, because traditional high resolution structure determination techniques only report on proteins in the condensed phase. For this reason, MD simulations play an important role for exploring the behavior of gas phase proteins.[6-7] A major limitation of presently available MD force fields is that charges are static, while excess protons are known to be highly mobile in the gas phase.[8-10]

In Chapter 2, we combined a newly developed mobile-proton algorithm with the GROMACS MD package to examine the properties of ubiquitin and lysozyme in the gas phase. This work takes advantage of IMS and ESI-MS under native (non-denaturing) conditions.

Focusing on ubiquitin and lysozyme, we examined several pertinent questions. Firstly, we examined the crystal phase structures to see how the main chain, the side chains, and the titratable residues are situated for proteins in the solid state.

As the second step, using MS and IM-MS in both positive and negative ion modes, experimental collision cross sections charge states of gaseous protein ions produced by native ESI were determined. We found 6+ and 5- to be the most abundant charge states of ubiquitin, while lysozyme spectra showed 8+ and 6- as the most abundant ion species.

In the third step, these data were compared with the solution phase structures which were obtained with MD simulations conducted at room temperature, paying particular attention to R_g and

RMSD values. We applied solvent MD runs to scrutinize whether the X-ray structures of both proteins are affected by crystal packing. Main and side chain orientations were fully retained in water, providing a justification for the hitherto unscrutinized standard approach of relying on crystal data for "solution" vs. gas phase comparisons.

Earlier modeling studies of gaseous proteins did not account for the highly mobile nature of H+ charge carriers. We compared for the first time the results from static and mobile H^+ simulations, focusing on both positively and negatively charged protein ions. MD simulations using GROMACS and an in-house software called "charge mover" which minimizes the total energy of the system and considers protons as mobile[8] were used to study the behavior of these two proteins in the gas phase to complete the fourth step. Traditional MD runs with non-mobile charges were conducted as well. Most earlier gas phase protein MD investigations employed very short (ns) simulation windows. By extending this time frame to 1 μs we were able to observe rare unfolding/folding transitions in ubiquitin. These predicted fluctuations were consistent with a semi-unfolded subpopulation seen in ion mobility spectrometry (IMS) experiments.

The fourth step was to analyze all the data amassed in the preceding four steps. Our goal was to develop a clear understanding of what happens to proteins in the gas phase under gentle conditions, both from an experimental and a computational perspective.

Our results showed that native-like conformations were largely retained in the gas phase. Very similar data were obtained in both mobile and non-mobile proton MD simulations in positive and negative ion mode. In mobile proton runs the multiply charged ions were stabilized by numerous salt bridges at the protein surface over extended time scales (1 microsecond). Our runs revealed a strong preference for retention of a solution-like backbone fold, while titratable/polar surface side chains collapsed towards the protein surface. This side chain collapse was caused by a multitude of intramolecular salt bridges, H-bonds, and charge-dipole interactions. Surface titratable side chains were found to adopt orientations that were less extended than in crystals and in solution (with R_g values 3-5% lower than for the X-ray coordinates), causing gaseous protein to be somewhat more compact than in the condensed phase. Calculated collision cross sections of these MD species were in a close agreement with experimental data. Our results generalize the findings of Steinberg et al. (ChemBioChem 9, 2417, 2008) who had first proposed the existence of such side chain contacts on the basis of short-term simulations using static H^+. Overall, this study supports the view that solution-like protein structures can be retained due to kinetic trapping on the time scale of typical ESI-IMS experiments.[11]

4.2. Future Work

The first aspect that needs to be developed in the scope of this research, is to extend the MD simulations time scale to milliseconds to simulate IMS experiments.

Unfolding and afterwards refolding of ubiquitin and lysozyme in the gas phase can be investigated. First the temperature needs to be increased in vacuum so the protein can unfold properly, then the refolding process will be explored while the temperature is returning to 300 K. The goal of this work will be to find out if the protein refolds properly (and re-acquires its native structure while refolding).

As a third project, higher charge states can be applied to the native protein in the gas phase to investigate if the protein still can preserve the native structure. As an example, 13+ and 12- charges (the maximum number of acidic and basic sites available on ubiquitin) can be investigated to see under what conditions the protein can still preserve the native structure; and what happens if the protein is examined by both mobile and non-mobile proton algorithms.

One of the interesting areas of study would be developing the mobile proton algorithm. It can be employed for studying the behavior of ligand-bound globular proteins and membrane proteins in the gas phase to see what the ligand reaction will be in the presence of mobile charge carriers.

4.3. References

1. Meuzelaar, H.; Tros, M.; Huerta-Viga, A.; van Dijk, C. N.; Vreede, J.; Woutersen, S., Solvent-Exposed Salt Bridges Influence the Kinetics of α-Helix Folding and Unfolding. *J. Phys. Chem. Lett.* **2014,** *5* (5), 900-904.

2. Nelson, D. L.; Cox, M. M.; Lehninger, A. L., *Lehninger principles of biochemistry*. W.H. Freeman: New York, 2013.

3. Kaltashov, I. A.; Eyles, S. J.; Desiderio, D. M., *Mass Spectrometry in Structural Biology and Biophysics: Architecture, Dynamics, and Interaction of Biomolecules*. Wiley: 2012.

4. Fenn, J. B., Electrospray Wings for Molecular Elephants (Nobel Lecture). *Angew. Chem. Int. Ed.* **2003,** *42* (33), 3871-3894.

5. Leney, A. C.; Heck, A. J. R., Native Mass Spectrometry: What is in the Name? *J. Am. Soc. Mass. Spectrom.* **2017,** *28* (1), 5-13.

6. Makarewicz, T.; Kaźmierkiewicz, R., Molecular Dynamics Simulation by GROMACS Using GUI Plugin for PyMOL. *J. Chem. Inf. Model.* **2013,** *53* (5), 1229-1234.

7. van Dijk, M.; Wassenaar, T. A.; Bonvin, A. M. J. J., A Flexible, Grid-Enabled Web Portal for GROMACS Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2012,** *8* (10), 3463-3472.

8. Konermann, L., Molecular Dynamics Simulations on Gas-Phase Proteins with Mobile Protons: Inclusion of All-Atom Charge Solvation. *J. Phys. Chem. B* **2017,** *121* (34), 8102-8112.

9. Konermann, L.; Metwally, H.; McAllister, R. G.; Popa, V., How to run molecular dynamics simulations on electrospray droplets and gas phase proteins: Basic guidelines and selected applications. *Methods* **2018,** *144*, 104-112.

10. Fegan, S. K.; Thachuk, M., A Charge Moving Algorithm for Molecular Dynamics Simulations of Gas-Phase Proteins. *J. Chem. Theory Comput.* **2013,** *9* (6), 2531-2539.

11. Clemmer, D. E.; Russell, D. H.; Williams, E. R., Characterizing the Conformationome: Toward a Structural Understanding of the Proteome. *Acc. Chem. Res.* **2017,** *50* (3), 556-560.

Curriculum Vitae

Name:	Maryam Bakhtiari
Post-secondary Education and Degrees:	Shahid Beheshti University Tehran, Tehran, Iran 2007-2011 B.Sc. University of Guilan Rasht, Guilan, Iran 2011-2014 M.Sc. The University of Western Ontario London, Ontario, Canada 2017-2019 M.Sc.
Honours and Awards:	Western Graduate Research Scholarship (WGRS) 2017-2018
Related Work Experience	Teaching Assistant The University of Western Ontario 2017-2018

Publications:

1. Bakhtiari, M., Ghalami-Choobar, B. *J. Electroanal. Chem.* **2015**, 754, 109-117.
2. Moradabadi, A., Bakhtiari, M., Kaghazchi, P. *Electrochim. Acta.* **2016**, 213, 8-13.