

---

Electronic Thesis and Dissertation Repository

---

7-8-2015 12:00 AM

## From Solution Into the Gas Phase: Studying Protein Hydrogen Exchange and Electrospray Ionization Using Molecular Dynamics Simulation

Robert G. McAllister  
*The University of Western Ontario*

Supervisor  
Lars Konermann  
*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  
© Robert G. McAllister 2015

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>

 Part of the [Physical Chemistry Commons](#)

---

### Recommended Citation

McAllister, Robert G., "From Solution Into the Gas Phase: Studying Protein Hydrogen Exchange and Electrospray Ionization Using Molecular Dynamics Simulation" (2015). *Electronic Thesis and Dissertation Repository*. 2938.  
<https://ir.lib.uwo.ca/etd/2938>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact [wlsadmin@uwo.ca](mailto:wlsadmin@uwo.ca).

From Solution Into the Gas Phase: Studying Protein Hydrogen Exchange and  
Electrospray Ionization Using Molecular Dynamics Simulation  
(Thesis format: Integrated Article)

by

Robert Gordon McAllister

Graduate Program in Chemistry

A thesis submitted in partial fulfillment  
of the requirements for the degree of  
Master of Science

The School of Graduate and Postdoctoral Studies  
The University of Western Ontario  
London, Ontario, Canada

© Robert Gordon McAllister 2015

## Abstract

Here, we apply Molecular Dynamics (MD) simulations to investigate fundamental aspects of structural mass spectrometry (MS). We first examine microscopic phenomena underlying Hydrogen/Deuterium exchange (HDX). HDX interrogates structural dynamics of proteins by measuring the rate of Deuterium uptake into backbone amides. We perform microsecond MD simulations on ubiquitin to investigate this process. We find that HDX protection often cannot be explained by H-bonding or solvent accessibility considerations. These findings caution against non-critical use of HDX data in structural contexts. We next use MD to examine the Electrospray ionization (ESI) mechanism of proteins. ESI is a soft ionization technique resulting in the production of gaseous protein ions. The mechanism of ion formation from nanometer sized droplets is unclear. We apply a trajectory stitching MD approach to simulate protein-containing nanodroplets, finding that natively-folded proteins remain solvated as droplets shrink. Residual charge carriers remain following desolvation, consistent with Dole's charged residue model.

## Keywords

Proteins, Molecular Dynamics, Mass Spectrometry, Electrospray Ionization, Hydrogen-Deuterium Exchange

## Co-Authorship Statement

The work in Chapter 2 was previously published and is reprinted with permission from:

McAllister, R. G. & Konermann, L. Challenges in the Interpretation of Protein H/D Exchange Data: A Molecular Dynamics Simulation Perspective. *Biochemistry* **54**, 2683–2692 (2015). Copyright 2015, American Chemical Society.

The work in Chapter 3 is being prepared for future publication:

McAllister, R. G., Metwally, H., Sun, Y. S. & Konermann, L. The Electrospray Mechanism of Natively Folded Proteins: A Molecular Dynamics Investigation.

The original drafts of these manuscripts were prepared by the author (R.G.M.) with subsequent revisions by the author and Dr. Lars Konermann. All simulation work presented was performed by the author. Mass Spectrometry and Ion Mobility Spectrometry data presented in Chapter 3 were gathered and analyzed by the author with assistance from Mrs. Haidy Metwally and Ms. Yu (Sherry) Sun.

*In loving memory of my grandparents.*

## Acknowledgments

First and foremost, I owe tremendous gratitude to my supervisor, Dr. Lars Konermann. I didn't approach Lars about joining his lab until only a couple of months before my proposed start date in September 2013. At that point, I had already been told by several potential supervisors that they could not accommodate me in their lab on such short notice. Despite only knowing me from a semester-long undergraduate course several years beforehand, Lars took a chance on me, and allowed me to join his group to begin work in molecular dynamics simulation – an area where I had little previous experience. Since then, he has consistently challenged me to exceed my own expectations, and to grow in my knowledge and skills. I am a better researcher, a better communicator, and a better person for his guidance. Working with Lars for these last two years has been a pleasure, and I will always be grateful for the opportunity he gave me.

I also thank my colleagues from the Konermann lab for many useful discussions and thoughtful questions over the years, as well as the help they have all lent to my research. Siavash, who has sat a meter to my left for the past year, is probably the hardest-working researcher I'll ever meet, and has provided guidance and advice on innumerable occasions; I have no doubts that he will be an excellent professor one day very soon. Haidy has been a wonderful collaborator on several of my projects, providing top notch experimental work to go with my simulations. I could not have completed the work in chapter 3 without her. I also thank Sherry for her help in gathering and analyzing the ion mobility data in chapter 3, particularly as she was writing her own thesis, at the time. Dupe, Ming, Courtney, Lauren, Danielle, Samuel, Antony, and those mentioned above have made my experience in the Konermann lab all the richer. I haven't the room, here, to give them half the thanks they deserve.

To Liz, my parents, and my late grandparents, I owe more gratitude than I could ever fully express. I thank them for always being there, and for believing in me even when I didn't. This work would not have been remotely possible without their love and support. It belongs to them as much as me.

# Table of Contents

Abstract.....	ii
Co-Authorship Statement.....	iii
Dedication.....	iv
Acknowledgments.....	v
Table of Contents.....	vi
List of Tables.....	x
List of Figures.....	xi
List of Symbols and Abbreviations.....	xiii
1 Introduction.....	1
1.1 Proteins.....	1
1.2 Mass Spectrometry.....	2
1.2.1 The Ion Source.....	2
1.2.2 The Mass Analyzer.....	3
1.2.3 The Detector.....	6
1.3 Electrospray Ionization.....	7
1.3.1 Charged Droplets.....	8
1.3.2 ESI Mechanisms.....	9
1.4 Structural Mass Spectrometry.....	11
1.4.1 Collision-Induced Dissociation and Tandem MS.....	11
1.4.2 Covalent Labelling and Cross-Linking.....	12
1.4.3 Ion Mobility Spectrometry.....	12
1.4.4 Hydrogen/Deuterium Exchange.....	13
1.5 Computer Simulations.....	15

1.5.1	<i>Ab Initio</i> Methods and Density Functional Theory.....	15
1.5.2	Molecular Mechanics.....	16
1.5.3	Monte Carlo Methods.....	16
1.5.4	Molecular Dynamics.....	16
1.6	Molecular Dynamics in Detail.....	16
1.6.1	Newton's Laws and Integration Algorithms.....	17
1.6.2	Force Fields.....	21
1.6.3	Thermostats.....	27
1.6.4	Additional Considerations.....	29
1.7	Scope of the Thesis.....	32
1.8	References.....	34
2	Challenges in the Interpretation of Protein H/D Exchange Data: A Molecular Dynamics Simulation Perspective.....	41
2.1	Introduction.....	41
2.2	Methods.....	44
2.2.1	MD Simulations.....	44
2.2.2	Data Analysis.....	44
2.3	Results and Discussion.....	46
2.3.1	Ubiquitin Structure and Dynamics.....	46
2.3.2	Experimental Results.....	48
2.3.3	Main Chain H-Bonds.....	50
2.3.4	Side Chain H-Bonds.....	50
2.3.5	Solvent Accessibility.....	51
2.3.6	Unprotected Sites.....	52
2.3.7	Problem Cases.....	52
2.3.8	Crystallographically Defined Water Molecules.....	53



2.3.9	Force Fields and Solvent Models.....	55
2.4	Conclusions.....	57
2.5	References.....	60
3	The Electrospray Mechanism of Natively Folded Proteins: a Molecular Dynamics Investigation.....	66
3.1	Introduction.....	66
3.2	Materials and Methods.....	68
3.2.1	Protein Solutions.....	68
3.2.2	Mass Spectrometry and Ion Mobility Spectrometry.....	69
3.2.3	MD Simulations: System Construction.....	69
3.2.4	MD Simulations: General Aspects.....	70
3.2.5	MD Trajectory Stitching.....	72
3.3	Results and Discussion.....	72
3.3.1	Folded Species in ESI-MS.....	72
3.3.2	Temporal Evolution of Protein Nanodroplets.....	76
3.3.3	Temperature Profile and Water Model Effects.....	78
3.3.4	Variations in Initial Protein Charge State.....	80
3.3.5	Additional Proteins.....	82
3.4	Conclusions.....	84
3.5	References.....	88
4	Conclusions and Future Work.....	93
4.1	Conclusions.....	93
4.2	Future Directions of Study.....	95
4.2.1	Extended Simulations and Additional Proteins in HDX/MS.....	95
4.2.2	Gas Phase HDX and Pulsed HDX.....	95
4.2.3	ESI-MS of Noncovalent Complexes.....	96

4.2.4 Additional Macromolecules in ESI-MS.....	96
4.3 References.....	97
Curriculum Vitae .....	98

## List of Tables

Table 2.1 Summary of backbone NH protection behavior. ....	58
Table 3.1 Summary of simulation conditions discussed in this chapter. ....	71

## List of Figures

Figure 1.1 Schematic of an ESI ion source.....	8
Figure 1.2 Proposed models of ESI ion formation. ....	10
Figure 1.3 Interactions in a typical force field.....	23
Figure 1.4 The Lennard Jones Potential. ....	24
Figure 1.5 The Coulombic potential for oppositely charged species.....	26
Figure 1.6 Schematic of Periodic Boundary Conditions. ....	31
Figure 2.1 Scematic of H-bond and bulk interaction rate algorithms.....	45
Figure 2.2 RMSD and RMSF plots for ubiquitin simulation.....	46
Figure 2.3 A structural fluctuation opens H-bonds near ubiquitin's C-terminus. ....	47
Figure 2.4 H-bonding properties of three NH sites.....	48
Figure 2.5 Overview of experimental HDX data, and various properties extracted from a 1 $\mu$ s CHARMM22*/TIP3P simulation. ....	49
Figure 2.6 Energy-minimized structure of ubiquitin, where blue/pink/red coloring denotes the level of HDX/NMR protection as in Figure 2.3.....	51
Figure 2.7 Problem cases and waters resolved in the crystal structure.....	53
Figure 2.8 Overview of experimental HDX data, and various properties extracted from a 1 $\mu$ s Amber99sb-ILDN/TIP4P simulation trajectory.....	56
Figure 3.1 ESI mass spectra of ubiquitin in 10 mM ammonium acetate under gentle conditions.....	73
Figure 3.2 ion mobility traces of adducted species in Figure 3.1c. ....	74

Figure 3.3 Snapshots from a 125 ns simulation of a ubiquitin-containing nanodroplet. ..	76
Figure 3.4 Root mean square deviation of ubiquitin vs the crystal structure in an evaporating nanodroplet.....	77
Figure 3.5 Mass spectra of ubiquitin sprayed from acidified and basified solution. ....	80
Figure 3.6 Ion mobility spectra of 6+ ubiquitin ions from acidified and basified solutions. ....	81
Figure 3.7 Mass spectra of cytochrome <i>c</i> and holo-myoglobin at pH 7. ....	83
Figure 3.8 Snapshots from MD simulations of holo-myoglobin and cytochrome <i>c</i> . ....	84
Figure 3.9 Comparison of experimental data, MD simulation results and a Rayleigh charge-based model for charge states of folded proteins. ....	86

## List of Symbols and Abbreviations

AMBER – Assisted Model Building with Energy Refinement

CEM – chain ejection model

CHARMM – Chemistry at Harvard Molecular Mechanics

CID – collision-induced dissociation

CRM – charged residue model

Cyt – equine heart cytochrome *c*

D – deuterium ( $^2\text{H}$ )

DFT – density functional theory

*e* – the elementary charge ( $1.602 \times 10^{-19}$  coulombs)

ESI – electrospray ionization

FTICR – Fourier transform ion cyclotron resonance

HDX – Hydrogen/Deuterium exchange

hMb – equine holo-myoglobin

IEM – ion evaporation model

IMS – ion mobility spectrometry

MALDI – matrix-assisted laser desorption/ionization

MD – molecular dynamics

MS – mass spectrometry

MS/MS – tandem mass spectrometry

$m/z$  – mass to charge ratio

NMR – nuclear magnetic resonance

OPLS/AA – Optimized Potential for Liquid Simulation – all atom

PBC – periodic boundary conditions

PME – Particle Mesh Ewald

RF – radio frequency

RMSD – root mean square deviation

RMSF – root mean square fluctuation

SASA – solvent accessible surface area

TOF – time of flight

Ubq – bovine ubiquitin

# 1 Introduction

## 1.1 Proteins

Proteins are a diverse class of biological macromolecules responsible for the vast majority of physiological functions *in vivo*, including structural, signaling, and catalytic processes. They are polymers of amino acids ranging in size from single chains of only a few kDa<sup>1</sup> to multi-chain complexes of several MDa.<sup>2</sup>

The dependence of correct protein function on structure is a long-established tenet of structural biology.<sup>3-5</sup> In the textbook example of serine protease action, the enzyme active site contains an “oxyanion hole” positioned ideally to coordinate a tetrahedral intermediate along the protein hydrolysis reaction coordinate. Similarly elegant geometries are present in the active sites of enzymes ranging from lysozyme<sup>4,6</sup> to F<sub>0</sub>F<sub>1</sub> ATPase.<sup>7</sup> Conversely, even small changes in the amino acid sequence of proteins – sometimes referred to as their “primary structure” – can result in deleterious consequences for the protein’s three-dimensional structure and function. For example, a single amino acid substitution to the oxygen carrier hemoglobin results in sickle cell disease.<sup>8</sup> As a result of this dependence, research on proteins often focuses on determining both structural and dynamic properties.

X-ray crystallography remains the gold-standard technique for high-resolution protein structural determination, accounting for more than 90% of structures deposited in the RCSB Protein Data Bank.<sup>9</sup> High resolution structures have led to revelations in understanding of biological systems from the first protein X-ray structure of sperm whale myoglobin<sup>10</sup> to solving the crystal structure of the bacterial ribosome.<sup>11</sup> Unfortunately, many proteins cannot be purified in sufficient quantity to form suitable crystals for X-ray diffraction, are intrinsically disordered, or are otherwise not amenable to x-ray crystallography studies. Nuclear magnetic resonance (NMR) spectroscopy and cryo-electron microscopy can obtain high-resolution structures of some of these proteins<sup>12,13</sup> but each still requires high concentrations of homogenous sample, and also suffers from size limitations and other disadvantages. When high resolution cannot be



easily obtained, one can still probe structure using lower-resolution techniques including ultraviolet-visible light spectroscopy<sup>14</sup>, circular dichroism spectroscopy<sup>15</sup>, small angle scattering experiments<sup>16</sup>, or structural mass spectrometry. Structural mass spectrometry is a widely-applicable and highly adaptable technique which can yield a wealth of structural data<sup>17</sup>, as will be discussed below.

## 1.2 Mass Spectrometry

Mass spectrometry (MS) is an analytical technique that measures the mass to charge ratio ( $m/z$ ) of analyte ions, where  $m$  is the ion mass in Daltons and  $z$  is the charge in terms of the elementary charge,  $e$ . The foundations for MS were laid during J. J. Thomson's search for the electron, and subsequent developments in the early 20<sup>th</sup> century extended its use to characterizing isotopic distributions and identifying organic compounds.<sup>18</sup> Further developments opened the door for MS on biological macromolecules and structural MS studies. Today, alongside their use in traditional research environments, MS is applied in fields as diverse as forensics<sup>19</sup> and space exploration.<sup>20</sup>

MS is a hugely useful technique for chemical and protein analysis largely because of its minimal analyte consumption, high sensitivity, low cost of operation, and tolerance for non-homogenous samples. Instruments are varied, and often specialized for particular applications, but they generally consist of three components: an ion source, a mass analyzer, and a detector. We will touch on each of these components briefly, highlighting aspects relevant to protein MS.

### 1.2.1 The Ion Source

The ion source generates charged gas-phase species. Many different ionization strategies have been developed. Electron impact ionization has been in use for both solid and gaseous analytes for nearly a century<sup>21,22</sup>, and has proven to be effective for detection and identification of small molecules. Other techniques, including chemical ionization approaches<sup>23</sup> have also been successful in this area. Unfortunately, these "harsh" ionization techniques, which rely on gas phase collisions between the analyte and charged particles, lead to extensive fragmentation, a problem that is exacerbated for large analytes.

In order to analyze proteins via MS, researchers typically rely on one of two “soft” ionization methods developed in the late 1980’s: matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI). MALDI, developed by Karas and Hillenkamp<sup>24</sup>, uses an ultraviolet laser to vaporize and ablate a protein co-crystallized with an organic “matrix.” The precise mechanism of ionization in MALDI is still not comprehensively understood<sup>25</sup>, but it is known to produce intact protein ions in the gas phase and can be tuned for resilience to the presence of contaminants such as salt and detergent.<sup>26,27</sup>

ESI, which will be discussed in more detail below, was developed by Fenn and co-workers.<sup>28</sup> It relies on the production of highly-charged protein ions from charged solvent droplets. ESI has many of its own advantages for protein analysis, including its ability to be directly coupled to upstream chromatography for efficient separation and analysis of complex mixtures, and has become the most-used ionization technique in protein MS.<sup>29</sup>

More recently, a number of “ambient” ionization techniques have been developed for *in situ* analyses that require efficient ionization outside the confines of the laboratory. These include desorption electrospray ionization<sup>30</sup> and laser ablation / electrospray ionization.<sup>31</sup>

### 1.2.2 The Mass Analyzer

The mass analyzer separates gas phase ions based on their  $m/z$  values. These analyzers typically consist of magnetic and electric fields forming ion “optics” which focus and guide the analyte ions to the detector, as governed by the classical equation of motion for ions:

$$\frac{m}{z} \frac{\mathbf{a}}{e} = \mathbf{E} + (\mathbf{v} \times \mathbf{B}) \quad (1.1)$$

where  $m$  is the mass of the particle,  $z$  is the ion charge in terms of the elementary charge ( $e$ ),  $\mathbf{a}$  is acceleration,  $\mathbf{E}$  is the electric field,  $\mathbf{v}$  is velocity, and  $\mathbf{B}$  is the magnetic field (bolded quantities represent vectors). Equation 1.1 is derived from Newton’s second law and the Lorentz force law.

In order to reduce collisions with background gas molecules, prevalent at ambient pressures, mass analyzers are housed inside a vacuum chamber held at a pressure that is typically less than  $10^{-6}$  Torr by vacuum pumps. Modern mass analyzers are able to achieve both high sensitivity and high resolution. Capacity for tandem MS experiments, low cost, and small size are additional desirable features in a mass analyzer.

Electric and/or magnetic sector mass analyzers were among the first to be developed<sup>32</sup> and are still in use today. In this type of instrument, travelling ions are deflected by static electric or magnetic fields in discrete regions of their flight path. For particles with a known initial velocity, their deflection can be calculated precisely using equation 1.1. These instruments are typically operated in a continuous scanning mode, which modulates the magnitude of the electric or magnetic field to transmit only a single  $m/z$  to the detector at a given time. This scanning is highly selective, but carries the risk that ions which are not “caught” by the specific parameters chosen will not be detected by the instrument. Sector instruments also tend to be expensive due to the strong static magnetic fields that they must generate and are less suitable for performing tandem MS experiments than other platforms, resulting in their being used primarily for small molecule analyses.

Quadrupole mass analyzers consist of four conducting cylindrical rods, which use a combination of radio-frequency (RF) voltage and DC voltage to filter out all but a small  $m/z$  window. The RF voltage is applied such that the potential on opposite rods is in-phase and that on adjacent rods is out-of phase. When only this RF voltage is applied, ions oscillate about the center of the quadrupole, but all are transferred through (except for ions with very low  $m/z$  such that they impact the rods). However, when a DC voltage is applied such that opposite rods have the same potential, trajectories of ions outside a small  $m/z$  window become destabilized, impacting the rods and not transmitting to the detector. Quadrupole mass analyzers are commonly used in protein MS because they are inexpensive, compact, and amenable for use in tandem MS, such as in a triple quadrupole arrangement.<sup>33</sup> Quadrupoles analyzers are typically operated in scanning mode, similar to sector instruments.

Pulsed mode analyzers accept packets of ions from an upstream ion gate or directly from a pulsed ion source such as MALDI. This type of analyzer includes time-of-flight (TOF) analyzers, Orbitrap analyzers, and Fourier transform ion cyclotron resonance (FTICR) analyzers. The latter two operate under a similar principle, whereby ions are trapped in a harmonic oscillating orbit passing close to a detector and inducing a current, which is recorded and later deconvoluted using a Fourier transformation to identify the individual  $m/z$  ratios of species in the ion packet. Orbitrap instruments use an electric field to accomplish this task<sup>34</sup>, while FTICR instruments use a strong magnetic field.<sup>35</sup> Both can achieve very high resolutions and mass accuracy, but require long acquisition times to do so for large analytes, making them somewhat less useful for coupling to continuous flow sources such as ESI. Superconducting magnets needed for the highest resolution ion cyclotron resonance instruments are also very expensive to cool and maintain.

TOF instruments have somewhat lower resolving power than Orbitrap or FTICR instruments, but still provide many of the same advantages, while being relatively inexpensive. Their high duty cycle allows efficient coupling to chromatography and use in tandem MS applications.<sup>36</sup> TOF analyzers accelerate ions through an electric field before allowing them to drift through a field-free region. The speed of the ions as they enter the field free region is given by:

$$E_{potential} = E_{kinetic} \quad (1.2)$$

$$ez\Delta U = \frac{1}{2}mv^2 \quad (1.3)$$

$$v = \sqrt{\frac{2ez\Delta U}{m}} \quad (1.4)$$

where  $v$  is the ion speed,  $z$  is the charge in terms of the elementary charge,  $e$ ,  $\Delta U$  is the potential difference across the electric field, and  $m$  is the ion mass. Since particles move at a constant speed in the vacuum environment of the TOF's field free region, flight time can be calculated as:

$$t_f = \frac{l}{v} \quad (1.5)$$

where  $t_f$  is the ion flight time,  $l$  is the length between the “pusher” field region and the detector. Combining equations 1.4 and 1.5 yields:

$$\frac{l}{t_f} = \sqrt{\frac{2ez\Delta U}{m}} \quad (1.6)$$

$$t_f = l \sqrt{\frac{m}{2ez\Delta U}} = \sqrt{\frac{m}{z}} \sqrt{\frac{l^2}{2e\Delta U}} \quad (1.7)$$

which can be simplified to

$$t_f = \sqrt{\frac{m}{z}} k \quad (1.8)$$

where  $k$  is a constant that is independent of the ion species. Ions with a greater  $m/z$  arrive later after the ion packet is released, and the flight time recorded can be used to determine the  $m/z$  precisely using equation 1.8. For improved resolution, modern TOF analyzers typically employ acceleration orthogonal to the ions' original direction of travel. A reflectron reduces peak broadening which is encountered when ions with the same  $m/z$  have different initial velocities. The reflectron also lengthens the ion path without substantially increasing instrument size. TOF analyzers are commonly used for MS analysis of both proteins and small molecules.

### 1.2.3 The Detector

The detector is the component in a MS instrument that is responsible for recognizing the presence of ions at a given  $m/z$  which have been separated in the mass analyzer. In instruments where ions make direct contact with the detector, such as TOF, quadrupole, or sector analyzers, some form of electron multiplier is typically used. In these devices, a single ion impacting the detector induces secondary emission of several electrons from the detector surface. This multiplication of signal occurs several times over as emitted electrons impact the multiplier again, resulting in a detectable current pulse. Modern instruments often make use of multi-channel plates with many small electron multiplier channels due to their high signal gain, short duty cycle, and ability to resolve ions in both space and time.<sup>36</sup> For Orbitrap and FTICR instruments, ions do not impact a detector, but

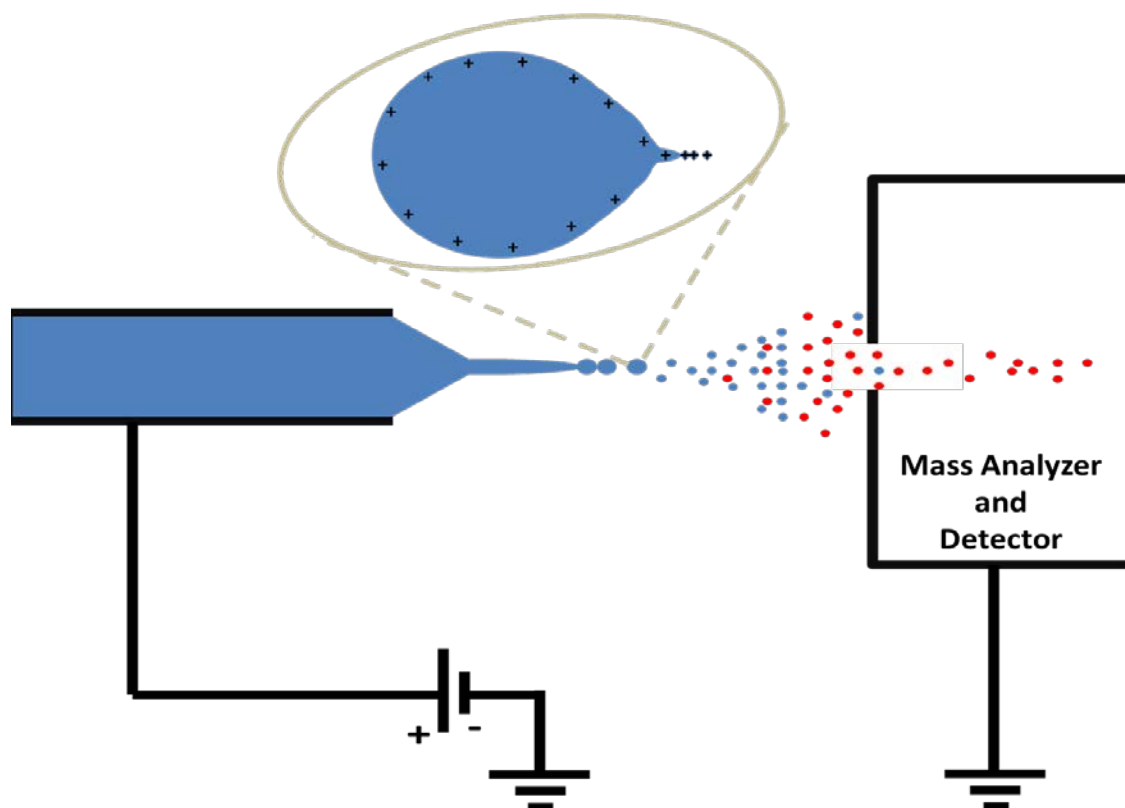
instead pass between pairs of metal-plate electrodes while being trapped in the analyzer, inducing a weak AC voltage, which can be transformed into discrete  $m/z$  signals.<sup>34,35</sup>

The detector signals are digitized and recorded on a computer, which performs the necessary mathematical transformations to calculate  $m/z$  values and intensities of ions. Many separate signals from the detector are combined to yield a full mass spectrum.

### 1.3 Electrospray Ionization

Electrospray ionization is a soft ionization technique, which is capable of ionizing large proteins to very high charge states. ESI is effective for a wide range of analyte sizes, from inorganic ions to small organic molecules to GDa proteins<sup>37</sup>, regardless of whether these species are charged or neutral in solution. ESI-MS was first demonstrated by Dole for analyzing masses of polystyrene<sup>38</sup>, with subsequent development by Fenn and co-workers extending the technique to other organic molecules, negatively charged analytes, and large proteins<sup>28,39,40</sup> for which he was awarded a part of the 2002 Nobel Prize in chemistry.<sup>29</sup>

In ESI, a low-concentration protein solution is introduced into a metal capillary. A potential of several kV is applied, with the counter electrode located at the orifice of the instrument's mass analyzer (Figure 1.1). Electrophoretic separation of charge in the analyte solution is driven by this potential, resulting in the formation of a Taylor cone at the capillary tip<sup>41</sup> as charge is accumulated and deforms the meniscus. When Coulombic repulsion exceeds surface tension, a highly-charged jet is emitted from the Taylor cone<sup>42</sup>, which quickly deforms and disintegrates into charged droplets. Charge balance is provided by electrolytic reactions at the capillary electrode, resulting in continuous emission of charged droplets as long as sample is infused. Charge in these droplets is predominantly carried by small ions including  $H^+$ ,  $Na^+$ ,  $Cl^-$ , or  $NH_4^+$  which were present in the initial analyte solution. For simplicity, the remainder of the discussion on ESI will assume that instrument is run in positive ion mode (*i.e.* a positive potential is applied to the capillary) such that ESI generates gaseous cations.



**Figure 1.1 Schematic of an ESI ion source.** Taylor cones are formed at the inlet capillary (left) and in evaporating droplets (inset), as charge accumulates at the surface. Late progeny droplets (blue) give rise to gas-phase ions (red), which are detected inside the instrument, at right.

### 1.3.1 Charged Droplets

Charged droplets are generated from the ESI source, carrying away a large portion of the charge accumulated in the Taylor cone, but a comparatively small volume of sample. The size of these droplets is strongly dependent on the size of the capillary tip, and varies based on voltage and sample flow rate, but they are typically on the order of  $10^{-5}$  m to  $10^{-7}$  m range.<sup>37,43</sup> Charge density at the surface of these droplets increases over time concomitant with solvent evaporation, until surface tension is overcome by Coulombic repulsion at the Rayleigh limit, which can be calculated by<sup>44</sup>:

$$z_R = \frac{8\pi}{e} \sqrt{\varepsilon_0 \gamma r^3} \quad (1.9)$$

where  $z_R$  is the number of elementary charges,  $e$ , at the Rayleigh limit,  $\varepsilon_0$  is the vacuum permittivity,  $\gamma$  is the surface tension of the solvent, and  $r$  is the droplet radius.

At the Rayleigh limit, micrometer-sized droplets form Taylor cones, and emit still smaller progeny droplets, which carry away a substantial portion of the charge, but comparatively little of the droplet volume.<sup>42,45</sup> After this fission, both parent and progeny droplets carry charges below the Rayleigh limit, and solvent evaporation occurs again. Several cycles of evaporation and fission may occur, until final-generation nanometer scale droplets are produced. These final progeny droplets emit gas phase ions, which ultimately enter the mass analyzer.

### 1.3.2 ESI Mechanisms

The mechanisms by which ion species are transferred from the final-generation droplets to the gas phase are still being debated.<sup>46-49</sup> Several mechanisms have been proposed which are relevant to the formation of gas-phase protein ions: the ion evaporation model (IEM), the charged residue model (CRM), and the chain ejection model (CEM) (Figure 1.2). All predict that ions are formed by nanometer scale droplets, but they differ in the mechanism of ion release.

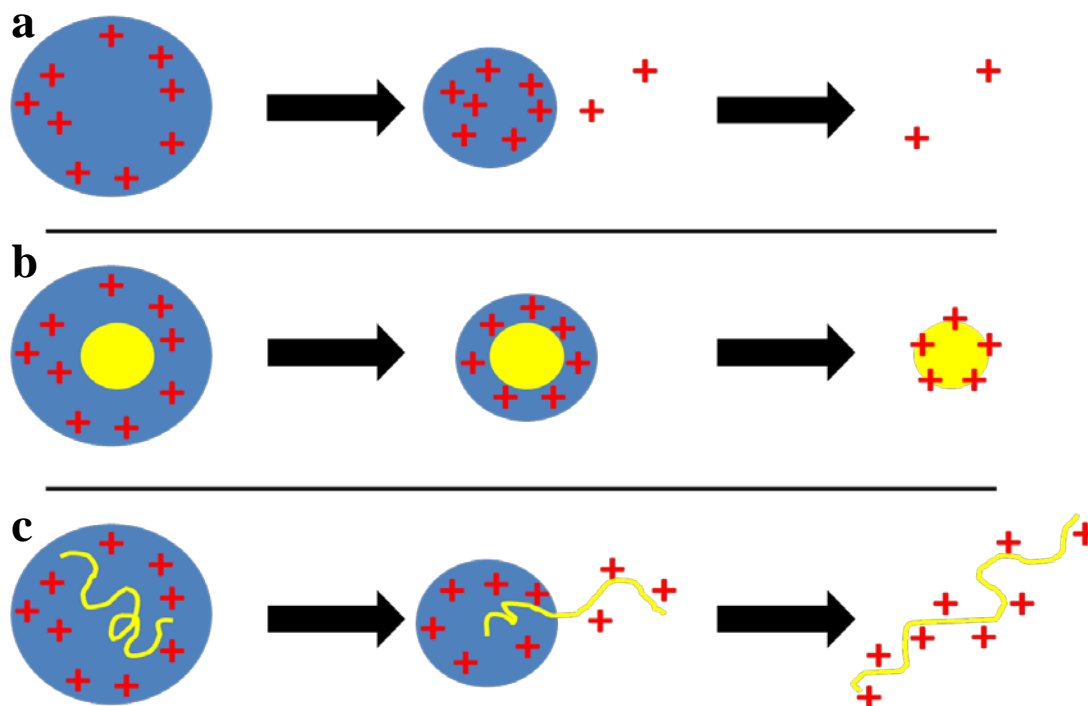
#### 1.3.2.1 The Ion Evaporation Model

The IEM suggests that once droplets are below a radius of several nanometers, charged ions are directly emitted from the droplet to shed charge as the surface charge density approaches the Rayleigh limit<sup>50</sup> (Figure 1.2a). The IEM is suspected to be the dominant ion formation mechanism for small pre-formed ions, such as  $\text{NH}_4^+$  and  $\text{CH}_3\text{COO}^-$  but has also been proposed as a mechanism by which protein ions may enter the gas phase.<sup>51</sup> Simulation studies of the ESI process have generally supported the view that IEM pertains predominantly to small molecules.<sup>47,48,52</sup>

#### 1.3.2.2 The Charged Residue Model

The CRM was proposed by Dole in his formative work on ESI.<sup>38</sup> It supposes that there exist many late-progeny droplets which contain only a single analyte ion. As solvent evaporates from these droplets, the remaining charge is left on the residual solute by charge transfer reactions or ionic attraction, resulting in a charge state that is approximately equal to that of an analyte-sized Rayleigh-charged droplet (Figure 1.2b).





**Figure 1.2 Proposed models of ESI ion formation.** The Ion Evaporation model is proposed for small ions (a), while the charged residue model (b) and chain ejection model (c) are proposed schemes for ionization of larger species such as proteins and polymers.

This model is well supported in both experimental and simulation studies of proteins and model polymers<sup>48,53,54</sup>, although it has not been observed directly. For small charged NaCl clusters and solvated metal ions, the CRM has been shown to be the dominant mechanism of ion formation by molecular dynamics simulation<sup>52,55</sup>, and it is suspected that this mechanism will also be dominant for large, hydrophilic species, such as folded proteins.

### 1.3.2.3 The Chain Ejection Model

A final model relevant to protein ion formation, the CEM, proposes that unfolded proteins are extruded from the droplet concomitant with charging via protonation on acidic and basic sites (Figure 1.2c). It suggests that this extension allows increased protein charging by spreading charges apart on the extended protein chain, which extends outside the droplet interior reducing Coulombic repulsion. Simulations using model polymers and polyethylene glycol polymers have supported this mechanism.<sup>47,48</sup>

## 1.4 Structural Mass Spectrometry

A number of techniques exist to elucidate protein structural information using MS. These techniques generally provide data at low-to-medium resolution, but can be informative when higher-resolution techniques are ineffective – when samples are at low concentration, are non-homogenous, or contain integral membrane and/or intrinsically disordered proteins. Many of these techniques use ESI as an ionization source due to its compatibility with chromatographic separation and soft ionization to very high charge states of large proteins, reducing instrument  $m/z$  range requirements.

Perhaps the simplest of these techniques is native ESI-MS, which attempt to reduce gas-phase collision and activation of proteins and protein-ligand complexes by tuning instrument parameters to lower pressures and voltages. In this way, non-covalent complexes can be preserved, and stoichiometries may be determined directly from the mass spectrum.<sup>56,57</sup> ESI charge state distributions can provide information about the conformations of proteins in the sprayed solution, as a consequence of the ionization mechanisms discussed above, with higher average charge states corresponding to unfolded species. Analysis of these distributions can be used to probe structural transitions, such as folding events, and even look at structural distributions in intrinsically disordered species.<sup>17,57,58</sup>

### 1.4.1 Collision-Induced Dissociation and Tandem MS

More detailed information about analytes can be acquired by analyzing not just full molecules, but also fragmented species by MS. Harsh ionization sources can be used to generate these fragments in-source, but this will result in undue loss of sensitivity for protein-containing samples, due to poor selectivity and extensive fragmentation. A better approach is the use of tandem mass spectrometry (MS/MS). In MS/MS, an ion of interest is isolated using a mass analyzer as a filter, then subjected to fragmentation before fragments are subsequently characterized in a second mass analyzer.<sup>59</sup>

In proteins, collision-induced dissociation (CID) is often used for fragmentation. Proteins or peptides are accelerated by an electric potential in a region of the mass spectrometer which has been filled with an inert background gas. High speed collisions with this

background gas result in activated species which generate ion fragments. This technique, in combination with MS/MS, can be used to sequence and subsequently identify proteins and peptides.<sup>60</sup> Although collision-induced dissociation and MS/MS do not provide protein structural data directly, they can be coupled to solution or gas phase chemical techniques such as covalent labelling and hydrogen-deuterium exchange which induce mass changes in the protein ions, yielding valuable conformational and dynamic data.

### 1.4.2 Covalent Labelling and Cross-Linking

Covalent labelling techniques, as the name implies, rely on modifying exposed region of a protein with a reactive species such as a hydroxyl radical.<sup>61</sup> These labels have a predictable mass shift when measured in MS/MS, and their location gives information about the solvent accessibility of amino acid side chains. Cross-linking studies are conceptually similar, but the reactive species is bifunctional, and therefore capable of reacting with two protein side chains. The functional groups are typically separated by a flexible linker, and identification of groups which have been chemically cross-linked by MS/MS can yield distance restraints which are useful in structural modelling.<sup>62</sup>

### 1.4.3 Ion Mobility Spectrometry

Ion mobility spectrometry (IMS) is being widely used in structural MS due to its ability to analyze gas phase structure, and its integration into commercially available MS instruments.<sup>63</sup> In IMS, ions are pushed through a region containing inert background gas by a weak electric field. Ions are accelerated at a rate proportional to charge, but experience a “drag” force opposite their direction of travel due to the background gas. As a result, they will move through the IMS cell at a rate that is proportional to their charge, but inversely proportional to their collisional cross section,  $\Omega$ . Compact, folded proteins have small  $\Omega$  values and move through the IMS cell quickly, while large, unfolded species move more slowly.<sup>64</sup> Recorded drift times can be compared to values calculated from possible structures using programs such as MOBCAL<sup>65</sup>, even allowing direct ion identification for some small species. IMS cells integrated into mass spectrometers provide gas-phase structural data, and can simultaneously act as an extra dimension of ion separation for discrimination of structurally-diverse compounds with the same  $m/z$ .

#### 1.4.4 Hydrogen/Deuterium Exchange

Hydrogen/Deuterium Exchange (HDX) probes protein structure and dynamics by taking advantage of the lability of protein O-H, S-H, and N-H bonds. When a protein is exposed to an isotopically-enriched solvent such as D<sub>2</sub>O, these labile hydrogens will be replaced with deuterons from the solvent. This process can be monitored by MS since D is 1 Da heavier than H, and so its incorporation into the protein results in a shift to higher *m/z*. NMR may also be used to monitor this process, as a result of the differing spins of the H and D nuclei. Regardless of the instrumentation used, HDX experimentalists often seek to obtain the rate of exchange,  $k_{HDX}$ , which is then related to protein structure and dynamics. Focus is usually on protein backbone amide groups which undergo HDX at a rate that is accessible in MS and NMR.<sup>66</sup> Exchange proceeds at each amide site according to the Linderstrøm-Lang scheme<sup>67,68</sup>:



where “open” conformations are exchange-competent, while “closed” configurations are exchange-incompetent,  $k_{open}$  and  $k_{close}$  are rates of structural transition between these states, and  $k_{ch}$  is an intrinsic rate constant which can be calculated based on the protein sequence, temperature and pD of the solution.<sup>66</sup> The experimentally-measured exchange rate,  $k_{HDX}$ , represents the rate of the overall conversion of NH→ND regardless of open or closed state. It can be related to the fundamental HDX constants above by:

$$\frac{d[NH_{open}+NH_{closed}]}{dt} = -k_{ch}[NH_{open}] \quad (1.11)$$

$$\frac{dNH_{open}}{dt} = k_{open}[NH_{closed}] - (k_{close} + k_{ch})[NH_{open}] \quad (1.12)$$

Under equilibrium conditions,  $NH_{open}$  is constant, so equation 1.12 gives:

$$(k_{close} + k_{ch})[NH_{open}] = k_{open}[NH_{closed}] \quad (1.13)$$

$$(k_{close} + k_{ch})[NH_{open}] = k_{open}[(NH_{open} + NH_{closed}) - NH_{open}] \quad (1.14)$$

$$NH_{open} = \frac{k_{open}}{k_{open}+k_{close}+k_{ch}} [NH_{open} + NH_{closed}] \quad (1.15)$$

Substituting equation 1.15 into equation 1.12 yields a first-order rate equation:

$$-\frac{d[NH_{open}+NH_{closed}]}{dt} = k_{HDX}[NH_{open} + NH_{closed}] \quad (1.16)$$

where:

$$k_{HDX} = \frac{k_{open}k_{ch}}{k_{open}+k_{close}+k_{ch}} \quad (1.17)$$

Based on the system under study, approximations can be made which allow  $k_{HDX}$  to be interpreted as a measure of opening free energy or kinetics. In general,  $k_{open} \ll k_{close}$ , and it can be left out of the denominator of equation 1.17. For most proteins at ambient temperature and near-neutral pH, amides spend a majority of time in the closed state, and only undergo rare, short-lived excursions to an exchange-competent structure giving the approximation  $k_{close} \gg k_{ch}$ , in the “EX2” limiting regime and:

$$k_{HDX} = \frac{k_{open}}{k_{close}} k_{ch} \quad (1.18)$$

allowing for estimation of the Gibbs’ free energy of the opening transition by the ratio of observed and intrinsic exchange rates:

$$\Delta G_{open} = -RT \ln \left( \frac{k_{open}}{k_{closed}} \right) = -RT \ln \left( \frac{k_{HDX}}{k_{ch}} \right) \quad (1.19)$$

Conversely, at high temperature or pH, the intrinsic exchange rate is elevated and proteins may be destabilized, occupying exchange-competent states for longer periods of time in the “EX1” regime where  $k_{close} \ll k_{ch}$ . In this case, the measured rate constant corresponds to the rate of structural opening events:

$$k_{HDX} = k_{open} \quad (1.20)$$

Experimentally, in HDX/MS, these regimes are observed at a peptide level as a gradual  $m/z$  increase in the case of EX2 kinetics, and a two-state distribution in the EX1 limit.

Mixed behavior is also possible.<sup>69</sup> This exchange occurs as a result of individual protein molecules exploring their full conformational space over time, allowing for the structural transitions that underlie these HDX schemes. The  $m/z$  shift observed over time can be used to estimate the rate of exchange by measuring at multiple time points. Exchange rates can be estimated at or near individual amide resolution using NMR, extensive proteolytic digestion or non-ergodic fragmentation techniques such as electron capture dissociation, in MS/MS.<sup>70,71</sup>

## 1.5 Computer Simulations

Simulation are a useful tool for studying phenomena that are difficult or even impossible to observe experimentally. Chemical simulations rely on computer algorithms to provide numerical solutions to equations derived from fundamental theories. The accuracy of these methods depends on the level of theory that is applied, but approximations are inherent to all chemical simulations. The level of accuracy required and the size of the system will determine which simulation strategies are employed, as greater size and accuracy require greater computational power.<sup>72</sup>

### 1.5.1 *Ab Initio* Methods and Density Functional Theory

*Ab initio* methods are based on the fundamental tenets of quantum mechanics, without input from empirical studies. These methods do not provide exact solutions to the Schrödinger equations, but they give the closest approximations available. Hartree-Fock simulations represent a commonly used *ab initio* method which is based on molecular orbital theory. It makes only a few approximations about the system under study, such as not explicitly considering electron-electron repulsion. These methods are highly computationally expensive since they include many-electron wavefunctions.<sup>72</sup> Density Functional Theory (DFT) is somewhat less expensive, since its models are based on single electron density rather than many-electron wavefunctions. Hybrid functional methods, which combine Hartree-Fock methods and DFT for greater accuracy with little additional computation time, have been developed<sup>73</sup>, but even the fastest of these methods are not sufficient for simulating proteins or other large biological molecules on experimentally-relevant timescales.

### 1.5.2 Molecular Mechanics

Molecular mechanics is a modelling technique which can be applied to simulate large systems such as proteins by treating the systems classically, avoiding costly quantum mechanical calculations. Atoms are treated as the smallest unit in the simulation, and are modelled as discrete point masses and/or charges. In these systems, potentials are calculated from empirically- or *ab initio* simulation-calibrated force fields that include bonded and non-bonded interaction terms. Bonded interactions such as stretching, bending, and torsion are generally modelled through harmonic potentials, while non-bonded interactions are modelled by various long-range potentials.<sup>72</sup> A more detailed look at these force fields is presented in 1.6.2 Force Fields.

### 1.5.3 Monte Carlo Methods

Monte Carlo methods rely on random permutations of a system to generate an ensemble of structures, whose properties and stabilities are usually determined based on molecular mechanics force fields. The Monte Carlo simulation scheme can provide excellent conformational sampling in protein systems, but it is limited to the study of systems at equilibrium and no correlation to time is possible.<sup>72</sup>

### 1.5.4 Molecular Dynamics

Molecular Dynamics (MD) simulations are another technique that typically relies on molecular mechanics force fields. In MD, Newton's equations of motion are integrated over discrete time steps to model the evolution of the simulated system over time. MD is a preferred simulation technique for proteins because of its low computational cost, scalability to large computer systems, and its ability to model kinetic processes.<sup>72</sup>

## 1.6 Molecular Dynamics in Detail

The growing use of MD simulations in recent decades is largely attributable to the wide availability of heavily-optimized simulation programs and steadily increasing computational power that have led to the many orders of magnitude increases in both system size and timescale accessible in MD simulation of proteins.<sup>74</sup> The earliest protein simulations were limited to systems of tens of particles evolving for tens of

picoseconds<sup>75</sup>, while the state-of-the-art today includes thousands of particles simulated for several milliseconds.<sup>76</sup>

Regardless of the particulars of the system, all MD simulations share some common features. Perhaps the most important of these features is the generation of a time-resolved trajectory of system configurations. This trajectory is a series of snapshots that correspond to individual microstates of the system under study. At equilibrium, these snapshots can be used to make predictions about the macroscopic system. This is due to the ergodic principle – that is, the ensemble average (replicate microstates in simulation) is equal to the time average (one system measured experimentally)<sup>77</sup>, provided enough microstates are sampled. The statistical ensemble that is modelled is generally selected to correspond to the experimental system under study. Examples include: the microcanonical ensemble (NVE) with constant number of particles, volume, and energy, which simulates an isolated system; the canonical ensemble (NVT) with constant number of particles, volume, and temperature, which simulates a system in thermal equilibrium with its surroundings; and the isothermal-isobaric ensemble (NPT), which simulates a system that is similar to NVT, but in a flexible container, such that pressure and temperature are in equilibrium with the surroundings.

### 1.6.1 Newton's Laws and Integration Algorithms

In MD, particles are treated classically, and are modelled as point masses. As a result, the evolution of the system can be modelled using Newtonian mechanics. In particular, the position of a particle,  $\mathbf{r}_i$ , can be fully described over the evolution of the system by Newton's second law:

$$\mathbf{F}_i = -\frac{\partial U(\mathbf{r}_i, \dots, \mathbf{r}_N)}{\partial \mathbf{r}_i} = m_i \frac{d^2 \mathbf{r}_i}{dt^2} \quad (1.21)$$

where  $\mathbf{F}_i$  is the force on particle  $i$ ,  $U(\mathbf{r}_i, \dots, \mathbf{r}_N)$  is the potential acting on the particle,  $m_i$  is the particle's mass, and  $t$  is time (bolded quantities are vectors). Since forces can be calculated from the system's current position, if it is assumed that acceleration is constant over some small timestep,  $\Delta t$ , and initial position and velocity are known, the system can be evolved over time by iterative calculation of new positions, velocities, and forces.



Several different integration schemes have been developed to perform these iterative calculations.

### 1.6.1.1 The Leapfrog Algorithm

With known initial conditions, an algorithm for updating positions and velocities might consist of simple first-order approximations:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \mathbf{v}_i(t) + O(\Delta t^2) \quad (1.22)$$

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \Delta t \frac{\mathbf{F}_i(t)}{m_i} + O(\Delta t^2) \quad (1.23)$$

where the  $O$  term represents a truncation error of the indicated order (in this case  $\Delta t^2$ ). This is known as the Euler integration scheme. Unfortunately, because velocity changes with time, equation 1.22 accumulates error quickly, since the value of the instantaneous velocity at  $t$  is not a good estimate of the average velocity over the full timestep. In order to improve accuracy, it is preferable to expand the Taylor series in equation 1.22 to include the second order term:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \mathbf{v}_i(t) + \frac{1}{2} \Delta t^2 \frac{\mathbf{F}_i(t)}{m_i} + O(\Delta t^3) \quad (1.24)$$

which we can rearrange to:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \left( \mathbf{v}_i(t) + \frac{1}{2} \Delta t \frac{\mathbf{F}_i(t)}{m_i} \right) + O(\Delta t^3) \quad (1.25)$$

In equation 1.25, we recognize the centre term as the right side of equation 1.23, but substituting  $\Delta t$  for a half time step. Thus, we may write:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \mathbf{v}_i \left( t + \frac{1}{2} \Delta t \right) + O(\Delta t^3) \quad (1.26)$$

Similarly, expanding equation 1.23 to include the second order term and rearranging will yield a center term that simplifies to the acceleration (force over mass) term a half-timestep ahead:

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \Delta t \left( \frac{1}{m_i} \mathbf{F}_i(t) + \frac{1}{2m_i} \Delta t \frac{d}{dt} \mathbf{F}_i(t) \right) + O(\Delta t^3) \quad (1.27)$$

$$= \mathbf{v}_i(t) + \Delta t \left( \frac{1}{m_i} \mathbf{F}_i\left(t + \frac{1}{2} \Delta t\right) \right) + O(\Delta t^3) \quad (1.28)$$

By substituting  $t=t+\frac{1}{2} \Delta t$ :

$$\mathbf{v}_i\left(t + \frac{3}{2} \Delta t\right) = \mathbf{v}_i\left(t + \frac{1}{2} \Delta t\right) + \Delta t \left( \frac{1}{m_i} \mathbf{F}_i\left(t + \Delta t\right) \right) + O(\Delta t^3) \quad (1.29)$$

The Leapfrog algorithm iterates between equations 1.26 and 1.29 to update the system over many time steps with considerably less error than the Euler algorithm.<sup>78</sup> The Leapfrog Algorithm gets its name from the positions/forces and velocities “hopping” one another in each iteration, with positions and forces being updated only at full time steps, and velocities being updated only at half-timesteps. It is an oft-used algorithm in MD because it requires only 3 calculations per iteration, making it efficient, with relatively low error.

### 1.6.1.2 Verlet and Velocity Verlet Integration

The Verlet algorithm is an alternative integration scheme that was popularized by Loup Verlet.<sup>79</sup> It is based on the Taylor expansion of  $\mathbf{r}_i(t)$  around  $t \pm \Delta t$ :

$$\mathbf{r}_i(t - \Delta t) = \mathbf{r}_i(t) - \Delta t \mathbf{v}_i(t) + \frac{1}{2} \Delta t^2 \frac{1}{m_i} \mathbf{F}_i(t) - \frac{1}{6} \Delta t^3 \frac{1}{m_i} \frac{d}{dt} \mathbf{F}_i(t) + O(\Delta t^4) \quad (1.30)$$

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \mathbf{v}_i(t) + \frac{1}{2} \Delta t^2 \frac{1}{m_i} \mathbf{F}_i(t) + \frac{1}{6} \Delta t^3 \frac{1}{m_i} \frac{d}{dt} \mathbf{F}_i(t) + O(\Delta t^4) \quad (1.31)$$

Adding equations 1.30 and 1.31 yields the Verlet algorithm, after rearranging:

$$\mathbf{r}_i(t + \Delta t) + \mathbf{r}_i(t - \Delta t) = 2\mathbf{r}_i(t) + \Delta t^2 \frac{1}{m_i} \mathbf{F}_i(t) + O(\Delta t^4) \quad (1.32)$$

$$\mathbf{r}_i(t + \Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \Delta t^2 \frac{1}{m_i} \mathbf{F}_i(t) + O(\Delta t^4) \quad (1.33)$$

The Verlet integration scheme is slightly more accurate than leapfrog integration, but it is seldom used due to lacking information on velocities, the need to know two successive

configurations of the system to begin integration, and the possibility of accumulating round-off errors in computer systems when adding the (very small) second order term to zero-order positions stored in memory.

A more commonly used variation of the Verlet algorithm is the velocity Verlet integration scheme:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \mathbf{v}_i(t) + \frac{1}{2} \Delta t^2 \frac{1}{m_i} \mathbf{F}_i(t) + O(\Delta t^3) \quad (1.34)$$

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \frac{1}{2m_i} \Delta t (\mathbf{F}_i(t) + \mathbf{F}_i(t + \Delta t)) + O(\Delta t^3) \quad (1.35)$$

This algorithm solves the major issues of the Verlet scheme, but it is not as accurate. It is very similar in performance and efficiency to the Leapfrog algorithm, but has the advantage of calculating velocities and positions at the same time points, which is useful in some cases. The second force calculation in equation 1.35 is typically stored in memory for the next iteration so that expensive force calculations are performed only once per timestep.

### 1.6.1.3 Energy Minimization Schemes

Energy minimization is an important step in an MD workflow, prior to the beginning of the actual simulation. The purpose of energy minimization is to move the modelled molecular configuration towards a local energy minimum in order to gently relax any highly unfavourable interactions, such as non-bonded atoms overlapping, resulting in overly-large repulsive forces at the first timestep (see 1.6.2.2 Van der Waals Forces). Energy minimization is a molecular mechanics method, since it is not correlated with time. A typical integration scheme is the method of steepest decent:

$$\mathbf{r}(n + 1) = \mathbf{r}(n) - \gamma \nabla U(\mathbf{r}(n)) \quad (1.36)$$

where  $\mathbf{r}(n)$  is a matrix containing all particle positions at step  $n$ ,  $\gamma$  is a small, scalar distance increment, and  $\nabla U$  is the gradient of the potential function. Equation 1.36 is iterated until the potential converges to a minimum, or the maximum repulsive force in the system is below some tolerance threshold.

## 1.6.2 Force Fields

The efficient integration of Newton's equations in MD depends on fast computation of potentials in a position-dependent manner. This would, ideally, be accomplished using the *ab initio* approaches previously discussed, but these calculations are prohibitively expensive for MD of large molecules. Instead, interactions between atoms are modelled by "effective" potentials which are calculated by a molecular mechanics force field. Force fields consist of a set of parameters describing the interactions between various atom types, and a set of equations to calculate the potential based on these parameters and the system's state.

Force fields differ in their level of detail in describing the system and how parameters are generated. Force fields commonly used in MD can be divided into several groups: "all-atom" potentials, which include explicit parameters for every atom in the system; "united atom" potentials, which save computation time by including contributions from non-polar hydrogens, such as those on methyl groups, in the parameters for the heavy atoms they are bonded to; and "coarse-grained" potentials, which reduce detail even further to improve efficiency, combining groups of heavy atoms and hydrogens into single, large pseudo-atoms. The parameters in united-atom and coarse-grained force fields are usually fitted to reproduce experimental results. All-atom force fields may be parameterized from *ab initio* calculations or an empirical fitting procedure. MD simulations of proteins are often carried out with all-atom force fields, including the Optimized Potential for Liquid Simulations – All Atom (OPLS/AA) force field, which is parameterized to fit experimental properties of liquids.<sup>80</sup> The Chemistry at Harvard Molecular Mechanics (CHARMM)<sup>81</sup> and Assisted Model Building with Energy Refinement (AMBER)<sup>82</sup> force fields, which derive their charge parameters from density functional theory calculations, are also common.

Each of these force fields contains similar sets of equations to determine the system potential, which can be summarized as:

$$U_{Total} = U_{Bonded} + U_{Non-Bonded} \quad (1.37)$$

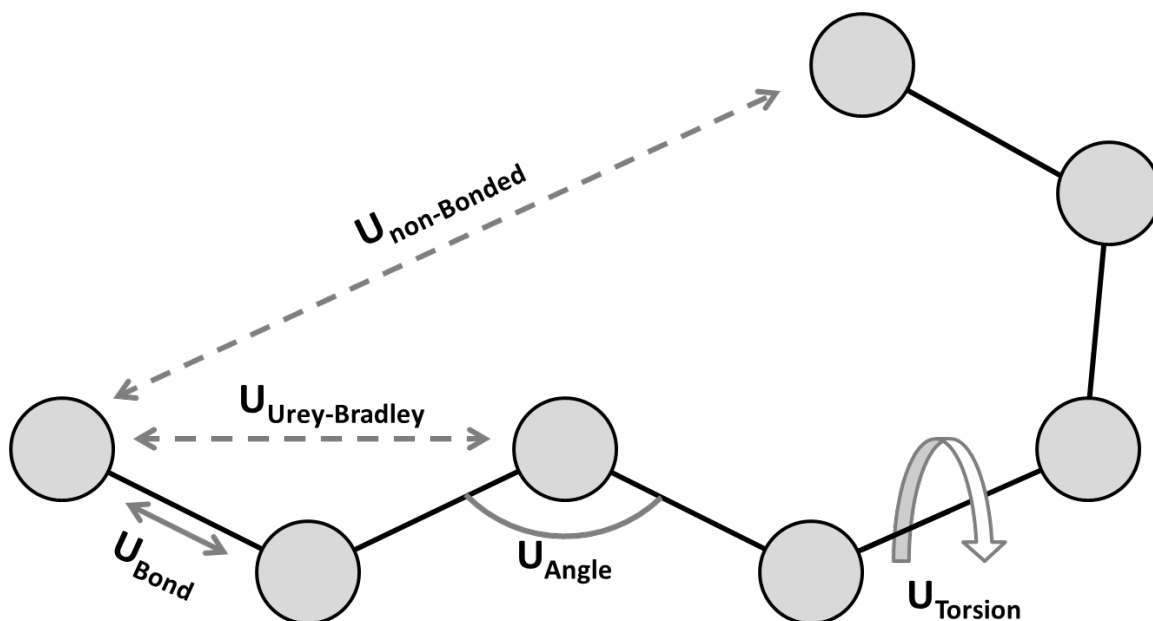
where the total calculated potential is a sum of bonded and non-bonded interaction terms (Figure 1.3). Non-bonded interactions are often ignored or reduced for atoms in the same molecule separated by fewer than 4 covalent bonds, as these interactions are captured by the bonded terms. The following sections will discuss how these terms are usually represented.

### 1.6.2.1 Bonded Interactions

Bonded interactions in protein MD are described in the force field by terms relating to covalent bond stretching (2-body term), bond bending (3-body term), and dihedral torsion (4 body term). Some force fields also include additional terms, such as an “improper” dihedral term which describes out-of-plane bending in planar systems (an additional 4-body term), or Urey-Bradley interactions which describe non-bonded force between atoms separated from one another by two covalent bonds (an additional 2-body term). The functional forms of these interactions are often based on harmonic potentials. For example, bond stretching potential in the AMBER force field is described by<sup>81</sup>:

$$U_{\text{Bond}} = \sum_{\text{bonds}} k_{\text{bond}}(d - d_0)^2 \quad (1.38)$$

where  $U_{\text{bond}}$  is the contribution of bond stretching to total potential,  $k_{\text{bond}}$  is a parameter determined by the identity of the bonded atoms,  $d$  is the current bond length,  $d_0$  is a parameter representing the minimum of the harmonic potential, and the sum is over all bonds in the system. Angle, improper dihedral, and Urey-Bradley interactions are also usually represented by harmonic potentials similar to equation 1.38.



**Figure 1.3 Interactions in a typical force field.** Bond, angle and Urey-Bradley interactions are often described with harmonic potentials, while torsion interactions are described by a Fourier series. Non-Bonded interactions include Coulombic and Lennard-Jones potentials. Not all interactions are labelled.

The dihedral angles term is somewhat more complicated, because the modelled systems may have more than one energy minimum at each bond. For example, in small molecules, several stable rotamers, such as anti and gauche configurations, may interconvert. In proteins, the backbone torsion angles are usually favoured to reside in certain ranges that are determined by local secondary structure. A harmonic potential with a single minimum would not effectively model this behavior, so the dihedral term is usually represented by a Fourier series, as in the AMBER force field<sup>81</sup>:

$$U_{dihedral} = \sum_{dihedrals} \sum_{j=1}^n k_j [1 + \cos(j\phi - \phi_{0,j})] \quad (1.39)$$

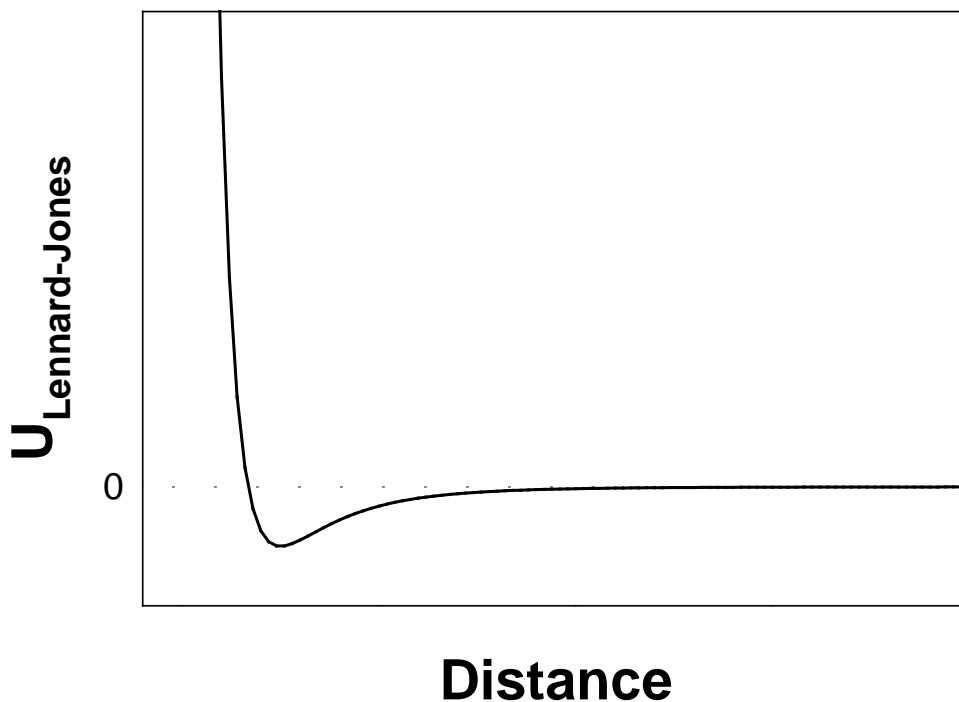
where  $k_j$  and  $\phi_{0,j}$  are parameters describing the amplitude and offset of the  $j$ th multiplicity term in the Fourier series,  $\phi$  is the current dihedral angle, and the sums are over all dihedral angles and all defined multiplicities, respectively.

### 1.6.2.2 Van der Waals Forces

Collectively, van der Waals forces describe the interactions between molecules that are not caused by covalent bonding or electrostatic effects. This includes permanent dipole/ permanent dipole interactions (Keesom forces), permanent dipole/ induced dipole interactions (Debye forces), induced dipole/ induced dipole interactions (London dispersion forces), and can be either attractive or repulsive. These interactions are usually modelled in MD force fields using a Lennard-Jones potential<sup>80-82</sup>:

$$U_{LJ} = \sum_{i,j} 4 \epsilon \left[ \left( \frac{\sigma}{d_{i,j}} \right)^{12} - \left( \frac{\sigma}{d_{i,j}} \right)^6 \right] \quad (1.40)$$

where  $\epsilon$  and  $\sigma$  are parameters which define the position and depth of the minimum in the Lennard-Jones potential, and  $d_{i,j}$  is the distance separating particles  $i$  and  $j$ . The potential is summed over all pairs of particles to which the non-bonded interaction terms apply. The  $10^{-6}$  term of the Lennard-Jones potential is attractive and there is net attraction between particles at distances past the minimum (Figure 1.4). This term approximates the



**Figure 1.4 The Lennard Jones Potential.** Units of distance and potential are arbitrary. The large negative slope as distance approaches zero prevents atoms from overlapping, modelling the Pauli repulsion.

averaged Keeson, Debye, and London forces. The  $10^{-12}$  term is repulsive, and accounts for the Pauli repulsion of overlapping electron orbitals at very short range (Figure 1.4). The repulsive term can also be represented by an exponential, as in a Buckingham potential, but the  $10^{-12}$  term is often chosen for this purpose because of its fast calculation as the square of  $10^{-6}$ .

The Lennard-Jones potential falls off quickly as distance between particles increases, meaning that it is often modelled with a cutoff,  $d_c$ , to improve computational efficiency. This only minimally impacts accuracy, if the cutoff is large, and a shifting adjustment is made such that the potential equals exactly 0 at the cutoff distance:

$$U_{LJ\text{-shifted}}(d) = \begin{cases} U_{LJ}(d) - U_{LJ}(d_c) & \text{for } d \leq d_c \\ 0 & \text{for } d > d_c \end{cases} \quad (1.41)$$

### 1.6.2.3 Electrostatic Interactions

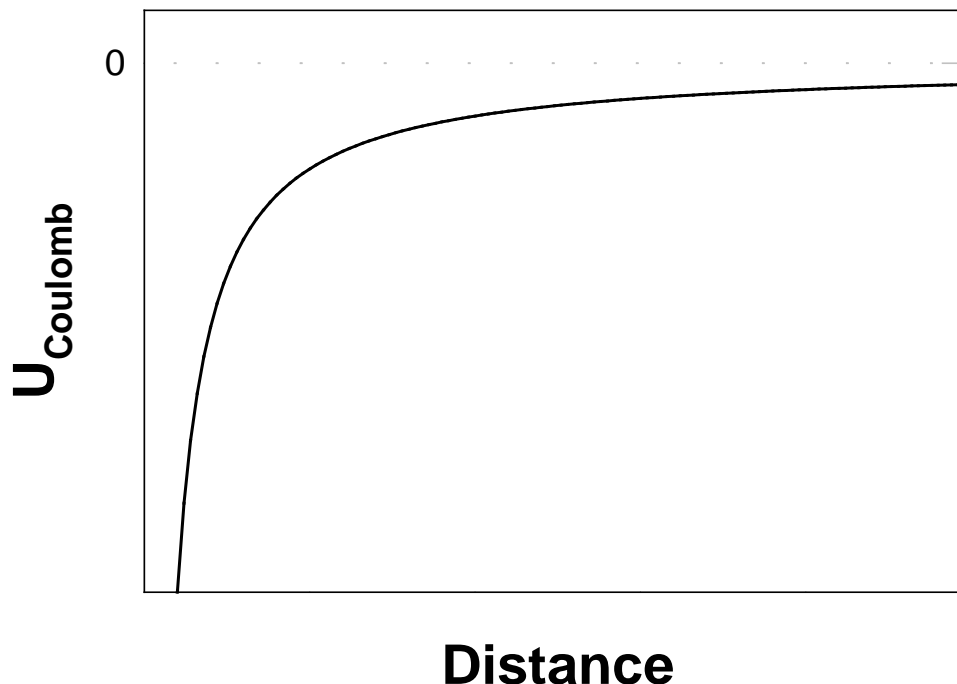
In MD, charged particle interactions are usually modelled with the Coulomb potential:

$$U_{Coulomb} = \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0 d_{i,j}} \quad (1.42)$$

where  $q_i$  and  $q_j$  are the charges of particles  $i$  and  $j$ ,  $\epsilon_0$  is the vacuum permittivity,  $d_{i,j}$  is the distance between the particles, and the sum is over all interacting pairs of particles (Figure 1.5). Equation 1.42 is often also shifted to include a cutoff, similar to the Lennard-Jones potential, but this may introduce problems, as the electrostatic interactions fall off much more slowly at long range. This can lead to undesirable artifacts in MD simulations, particularly in periodic systems which simulate bulk media (discussed in 1.6.4.2 Boundary Conditions and Simulation Cells).<sup>83</sup>

A solution to this problem is modelling the long range electrostatic interactions in periodic systems using the Ewald summation.<sup>84,85</sup> This summation technique replaces each point charge in the system with a point charge screened by a Gaussian charge distribution opposite in sign, whose contribution to the potential rapidly goes to zero at





**Figure 1.5 The Coulombic potential for oppositely charged species.** Units of distance and potential are arbitrary. The Coulomb potential converges to zero much more slowly than the Lennard-Jones potential (Figure 1.4).

large distances. To compensate for these screening charges, a second Gaussian charge distribution equal in magnitude, but opposite in sign to the screening charges is added. Short range electrostatic contributions are calculated directly using the screened point charges, while long range contributions can be calculated from the Fourier transform of the compensating Gaussians. The end result is summation that rapidly converges to the correct value for electrostatic potential, including the long range interactions.

To further improve efficiency of the Ewald summation, MD software packages often approximate it using a Particle Mesh Ewald (PME) algorithm. PME computes the long-range electrostatic interactions by distributing compensating charges on a discrete lattice, such that the overall charge density is maintained.<sup>86,87</sup> In this way, the compensating charge distribution can be computed quickly by a Fast Fourier Transform, allowing much-improved scaling in large systems, such as proteins.

### 1.6.2.4 Water Models

Although implicit solvation algorithms have been developed, water is usually modelled explicitly (*i.e.* as discrete particles in the system, rather than as a continuous dielectric medium). Despite the small size of H<sub>2</sub>O molecules, there are many different models to describe them in MD, which can be classified based on the number of simulated interaction sites, flexibility of covalent bonds, and polarizability. Polarizable and flexible models impose additional computation costs, meaning that most protein MD studies will benefit from use of a rigid, non-polarizable model, if it accurately reproduces properties of import.

Rigid three-site models such as TIP3P<sup>88</sup> and SPC/E<sup>89</sup> represent the H and O atoms explicitly with their own partial charges, and typically only consider Lennard-Jones interaction with the O atom. These models are highly efficient, and reasonably reproduce many of the physical properties of water, such as density and diffusion coefficient. Four-site models including TIP4P<sup>88</sup> and its modified relative TP4P/2005<sup>90</sup> place the O charge on a massless “virtual site” that sits on the bisector of the HOH angle. These models better reproduce experimental surface tension values relative to their three-site counterparts. Five site models that split the O charge between two virtual sites representing lone pairs of electrons have also been developed, with sites in a tetrahedral geometry, such as TIP5P.<sup>91</sup> However, no model perfectly reproduces all properties of water, and so selection of an appropriate model requires diligence on the part of the MD practitioner.

### 1.6.3 Thermostats

Thermostat algorithms are used in MD simulations in order to maintain an average temperature throughout the simulation run (for sampling in the NVT or NPT ensemble). This is particularly desirable for systems containing large molecules, such as proteins, since conformational changes in these molecules could otherwise lead to large temperature fluctuations in such a small system, which is not consistent with experimental conditions – for example, denaturing a protein and breaking H-bonds *in vitro* will not cause the system temperature to drop drastically, as it might *in silico*

without a thermostat. Initial thermalization of a system is usually accomplished by sampling velocities from a Maxwell-Boltzmann distribution, while the average temperature is maintained over time by one of several thermostat algorithms.

### 1.6.3.1 The Berendsen Weak Coupling Algorithm

The Berendsen thermostat is a velocity scaling scheme, which is implemented by multiplying all particle velocities in the system by a constant,  $\lambda$ , determined by<sup>92</sup>:

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau_T} \left( \frac{T}{T_0} - 1 \right)} \quad (1.43)$$

where  $\Delta t$  is the simulation timestep,  $\tau_T$  is a coupling constant,  $T$  is the instantaneous temperature of the system, and  $T_0$  is the thermostat's target temperature. The square root is taken since macroscopic temperature scales as the square of microscopic velocities. The coupling constant describes the strength of interaction with the simulated heat bath, with larger values producing a weaker coupling. Unfortunately, this thermostat does not reproduce a correct NVT ensemble, partly because it does not conserve angular momentum, however, this can be addressed by including a stochastic “noise” term in the algorithm, at relatively little computational cost.<sup>93</sup>

### 1.6.3.2 The Andersen Thermostat

The Andersen thermostat is a conceptually simple algorithm for coupling to a heat bath, and was the earliest thermostats to demonstrate NVT sampling in MD.<sup>94</sup> In this scheme, a subset of particles is randomly selected at each timestep, and velocities are re-sampled from a Maxwell-Boltzmann distribution for only these particles. The major shortcoming of this thermostat is that the repeated randomization of velocities results in poor modelling of time-dependent processes in the system, such as diffusion.

### 1.6.3.3 The Nose Hoover Extended Ensemble

The Nose Hoover thermostat is another commonly-used thermostat which couples the system to a heat bath by adding a non-Newtonian term to the equations of motion at each time step in order to maintain the average kinetic energy.<sup>95,96</sup> This additional term

simulating the heat reservoir is multiplied by the current velocities, and effectively adds a thermodynamic drag force. The thermostat modifies the equations of motion as follows (the Verlet integration scheme from equation 1.33 is chosen for simplicity, but the Nose Hoover thermostat may be used with any of the integration schemes):

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \Delta t^2 \left( \frac{\mathbf{F}(t)}{m} - \boldsymbol{\xi}(t + \Delta t) \cdot \mathbf{v}(t) \right) \quad (1.44)$$

$$\boldsymbol{\xi}(t + \Delta t) = \boldsymbol{\xi}(t) + [\sum_i m_i \mathbf{v}_i^2(t) - gT_0] \frac{\Delta t}{2Q_N} \quad (1.45)$$

where  $T_0$  is the thermostat target temperature,  $g$  is the number of degrees of freedom in the system, and  $Q_N$  is a coupling constant of the heat bath, which results in stronger coupling at greater values. This thermostat has the advantage of being time-reversible, and correctly reproduces the NVT ensemble, although it is somewhat more expensive to calculate than the Berendsen or Andersen schemes.

## 1.6.4 Additional Considerations

Several additional algorithms may be applied in MD simulations to better represent experimental conditions, or improve computational efficiency. Examples in the former category include applying barostat algorithms to control pressure and use of solid or periodic boundaries to achieve bulk or slab system geometries. The latter category includes optimizing simulation cell shape, generation of neighbor lists, and application of rigid constraints. Each of these will be discussed briefly.

### 1.6.4.1 Barostats

Barostat algorithms are analogous to thermostats, allowing simulation of a system under constant pressure, so long as boundaries are defined (*i.e.* the simulation is not performed in a vacuum environment). A number of different schemes have been developed for this purpose. The simplest of these is probably the Berendsen barostat<sup>92</sup>, which calculates a scaling factor  $\mu$  by:

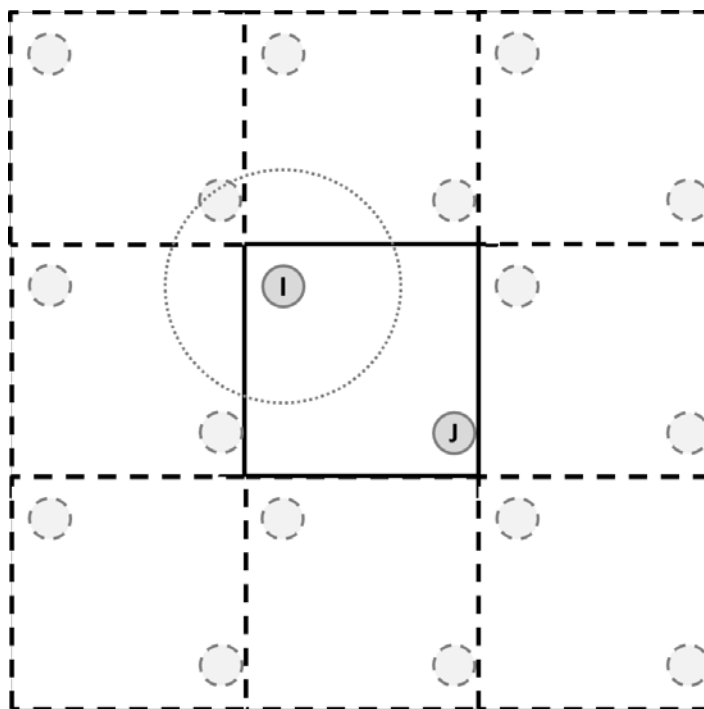
$$\mu = \sqrt[3]{1 - \frac{\Delta t}{\tau_P} (P - P_0)} \quad (1.46)$$

where  $\tau_P$  is a parameter controlling the strength of coupling to the barostat, with larger values representing tighter coupling,  $\Delta t$  is the simulation timestep,  $P$  is the instantaneous pressure, and  $P_0$  is target pressure of the barostat. At each timestep, the vectors defining the box, and the positions of all particles relative to the box center are scaled by  $\mu$ . The form of this barostat is analogous to the Berendsen thermostat, previously discussed.

#### 1.6.4.2 Boundary Conditions and Simulation Cells

In order to simulate realistic systems, various boundary conditions may be applied to MD. The finite region defined by the position of these boundaries is called the simulation or unit cell. The most commonly used boundary type is periodic boundary conditions (PBC) in all directions, in order to simulate a bulk system (Figure 1.6). PBC effectively surrounds the simulated system with “images” of itself in all directions, such that when a particle leaves the simulation cell along one face, it re-enters on the opposite face. Inter-atomic forces also extend into these images. Rigid boundaries can also be defined so that particles cannot exit the simulation cell along one or more axis. For example, a cell may have rigid boundaries along the  $z$  axis with PBC along the  $x$  and  $y$  axes, leading to a slab geometry that is infinite in the  $xy$  plane, which can be useful for MD simulations of membranes, interfaces, or surfaces.

In protein MD, when simulating in bulk solvent with PBC, the volume of the simulation cell determines the number of solvent molecules that must be included. The shape of the cell can be any infinitely tile-able 3d structure, such as a cube. The smallest size that can be used for this cell is defined by the minimum image convention, which states that images of a single particle should be farther apart than the short-range interaction cutoff of the Lennard-Jones and electrostatic potentials to prevent artificial self-interaction. With this in mind, other simulation cell shapes can be selected that maintain the same minimum particle-to-image distance as a cube, but with lower volume. The truncated octahedron and rhombic dodecahedron are often used for this purpose.



**Figure 1.6 Schematic of Periodic Boundary Conditions.** The simulation cell is located in the center, with images surrounding it on all sides. Only one “layer” of images is shown for clarity. Particle *I* interacts with the image of particle *J* in the Northwest image, since it is within the non-bonded cutoff (dotted circle). If *J* diffuses into the East image cell, its image from the West cell will enter the central simulation cell and replace it.

### 1.6.4.3 Neighbour Lists

Neighbour lists are used when non-bonded interactions have a cutoff distance to reduce computation time by efficiently identifying particles likely to interact with one another, and only calculating interactions between particles on those lists. The Verlet cutoff scheme<sup>97</sup> accomplishes this by storing a list of potential interaction partners for each particle that are within the cutoff distance plus some buffer distance, and reusing this list for several integration steps, since only those particles on the list will fall within the cutoff distance during a short period following the list generation. The Verlet cutoff scheme is much more efficient than a pairwise distance calculation at every timestep, and can be improved even further by applying domain decomposition when building the list. Domain decomposition splits the simulation cell into small, regular blocks whose size is based on some cutoff distance – such as the interaction cutoff plus the Verlet buffer distance – such that pairwise distance calculation to build the neighbor list for a particle

need only consider the same and adjacent blocks. When used together, these techniques allow non-bonded force calculation time to scale linearly with system size, instead of quadratically which is the case for an all-vs-all distance check at each timestep.

#### 1.6.4.4 Constraints

Constraint algorithms are used in MD simulation to allow a larger timestep by reducing the frequency of the fastest vibrational motions in the system. They model covalent bonds as rigid, instead of as a harmonic potential, since the stretching frequency of these bonds is much higher than the translational motions of unbounded atoms. Commonly used algorithms include the SHAKE<sup>98</sup>, SETTLE<sup>99</sup>, and LINCS<sup>100</sup> algorithms, which modify the forces on constrained particles using the method of Lagrange multipliers. The algorithms are iterative, and proceed until the constraint is satisfied within some small tolerance. When all covalent bonds in a protein MD simulation are constrained, the timestep may be increased several-fold, allowing for much longer simulations to be achieved at the same computational cost.

### 1.7 Scope of the Thesis

Here, we use molecular dynamics simulations to explore fundamental aspects of HDX, and ion desolvation during protein electrospray ionization. MD simulation provides an atomistic view of protein dynamics, giving insight on how the microscopic motions of proteins and unobservable processes within the ESI source give rise to the  $m/z$  signals that are observed experimentally. By taking advantage of the powerful computation capabilities of modern hardware and software, large protein systems can be investigated on timescales relevant to MS investigations. In this way, the marrying of simulations at an atomistic level with medium resolution structural mass spectrometry data can provide new insights into observed phenomena in both the solution and gas phase.

First, MD simulation is used to probe the structural and dynamic basis of amide protection in HDX (Chapter 1). We perform microsecond-scale simulations of ubiquitin. We compare amide-resolved protection factors from previously-published HDX/NMR datasets to the solution-phase structural environment of each NH group in the MD simulation. We evaluate the importance of microscopic properties including hydrogen

bonding to backbone carbonyls, hydrogen binding to side-chains, solvent accessible amide surface area, and solvent dynamics to experimentally observed HDX protection. We highlight anomalous protection in many amides in this well-studied model system, stressing the importance of critical use of HDX data in structural contexts.

Next, the final stages of the electrospray ionization process are investigated using MD (Chapter 2). In ESI, repeated cycles of droplet evaporation and fission give rise to nanometer-scale charged droplets containing few analyte ions. These droplets are too small to be directly observed, but are known to give rise to naked gas-phase ions which are detected in MS. Large, hydrophilic species such as folded proteins are usually thought to emerge from the droplet as charged species following complete solvent evaporation, the “Charged Residue Model,” although other mechanisms have been proposed. We use MD to investigate the evolution of charged protein-containing droplets in the ESI source, reporting the first simulation of such a droplet to dryness and supporting for the idea that proteins are CRM products. Moreover, we identify final-stage droplet radius as a primary factor determining experimentally observed charge states in protein ESI/MS.

We conclude with a brief discussion of MD as a complementary technique to structural MS of proteins, and highlight some future directions of research (Chapter 4).



## 1.8 References

- 1 Neidigh, J. W., Fesinmeyer, R. M. & Andersen, N. H. Designing a 20-residue protein. *Nature Structural & Molecular Biology* **9**, 425-430 (2002).
- 2 Walz, J. *et al.* 26S proteasome structure revealed by three-dimensional electron microscopy. *Journal of structural biology* **121**, 19-29 (1998).
- 3 Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223-230 (1973).
- 4 Blake, C. *et al.* Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature*, 757-761 (1965).
- 5 Cunningham, B. C. & Wells, J. A. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* **244**, 1081-1085 (1989).
- 6 Vocadlo, D. J., Davies, G. J., Laine, R. & Withers, S. G. Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* **412**, 835-838 (2001).
- 7 Abrahams, J. P., Leslie, A., Lutter, R. & Walker, J. E. Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature*, 621-628 (1994).
- 8 Frenette, P. S. & Atweh, G. F. Sickle cell disease: old discoveries, new concepts, and future promise. *Journal of Clinical Investigation* **117**, 850 (2007).
- 9 Berman, H. M. *et al.* The protein data bank. *Nucleic acids research* **28**, 235-242 (2000).
- 10 Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662-666 (1958).
- 11 Yusupov, M. M. *et al.* Crystal structure of the ribosome at 5.5 Å resolution. *science* **292**, 883-896 (2001).
- 12 Chang, S. *et al.* Cryo-EM Structure of Influenza Virus RNA Polymerase Complex at 4.3 Å Resolution. *Molecular Cell* (2015).
- 13 Cino, E., Fan, J., Yang, D. & Choy, W. Y. 1H, 15N and 13C backbone resonance assignments of the Kelch domain of mouse Keap1. *Biomol. NMR Assign.*, DOI 10.1007/s12104-12012-19398-12106 (2012).
- 14 Nienhaus, K. & Nienhaus, G. U. in *Protein-Ligand Interactions* 215-241 (Springer, 2005).
- 15 Sreerama, N., Venyaminov, S. Y. & Woody, R. W. Estimation of the number of  $\alpha$ -helical and  $\beta$ -strand segments in proteins using circular dichroism spectroscopy. *Protein Science* **8**, 370-380 (1999).
- 16 Svergun, D. I. & Koch, M. H. Advances in structure analysis using small-angle scattering in solution. *Current opinion in structural biology* **12**, 654-660 (2002).

- 17 Konermann, L., Vahidi, S. & Sowole, M. A. Mass Spectrometry Methods for Studying Structure and Dynamics of Biological Macromolecules. *Anal. Chem.* **86**, 213-232, doi:10.1021/ac4039306 (2014).
- 18 Griffiths, J. A Brief History of Mass Spectrometry. *Anal. Chem.* **80**, 5678-5683 (2008).
- 19 Ifa, D. R., Manicke, N. E., Dill, A. L. & Cooks, R. G. Latent fingerprint chemical imaging by mass spectrometry. *Science* **321**, 805-805 (2008).
- 20 Biele, J. & Ulamec, S. Capabilities of Philae, the Rosetta lander. *Space Science Reviews* **138**, 275-289 (2008).
- 21 Bleakney, W. A new method of positive ray analysis and its application to the measurement of ionization potentials in mercury vapor. *Physical Review* **34**, 157 (1929).
- 22 Dempster, A. A new method of positive ray analysis. *Physical Review* **11**, 316-325 (1918).
- 23 Munson, M. S. & Field, F.-H. Chemical ionization mass spectrometry. I. General introduction. *Journal of the American Chemical Society* **88**, 2621-2630 (1966).
- 24 Karas, M. & Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons. *Anal. Chem.* **60**, 2299-2301 (1988).
- 25 Chang, W. C. *et al.* Matrix-assisted laser desorption/ionization (MALDI) mechanism revisited. *Anal. Chim. Acta* **582**, 1-9 (2007).
- 26 Chen, R. *et al.* High-salt-tolerance matrix for facile detection of glucose in rat brain microdialysates by MALDI mass spectrometry. *Analytical chemistry* **84**, 465-469 (2011).
- 27 Amado, F. M., Santana-Marques, M. G., Ferrer-Correia, A. & Tomer, K. B. Analysis of peptide and protein samples containing surfactants by MALDI-MS. *Analytical Chemistry* **69**, 1102-1106 (1997).
- 28 Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **246**, 64-71 (1989).
- 29 Fenn, J. B. Electrospray Wings for Molecular Elephants (Nobel Lecture). *Angew. Chem. Int. Ed.* **42**, 3871-3894 (2003).
- 30 Takats, Z., Wiseman, J. M., Gologan, B. & Cooks, R. G. Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science* **306**, 471-473 (2004).
- 31 Nemes, P. & Vertes, A. Laser ablation electrospray ionization for atmospheric pressure, in vivo, and imaging mass spectrometry. *Analytical Chemistry* **79**, 8098-8106 (2007).
- 32 Thomson, J. J. Bakerian lecture: Rays of positive electricity. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 1-20 (1913).

- 33 Yost, R. & Enke, C. Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation. *Analytical chemistry* **51**, 1251-1264 (1979).
- 34 Hu, Q. *et al.* The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **40**, 430-443 (2005).
- 35 Comisarow, M. B. & Marshall, A. G. Fourier transform ion cyclotron resonance spectroscopy. *Chemical physics letters* **25**, 282-283 (1974).
- 36 Chernushevich, I. V., Loboda, A. V. & Thomson, B. A. An introduction to quadrupole–time-of-flight mass spectrometry. *J Mass Spectrom* **36**, 849-865 (2001).
- 37 Kebarle, P. & Verkerk, U. H. Electrospray: From Ions in Solutions to Ions in the Gas Phase, What We Know Now. *Mass Spectrom. Rev.* **28**, 898-917 (2009).
- 38 Dole, M. *et al.* Molecular beams of macroions. *J. Chem. Phys.* **49**, 2240-2249 (1968).
- 39 Yamashita, M. & Fenn, J. B. Negative Ion Production with the Electrospray Ion Source. *J. Phys. Chem.* **88**, 4671-4675 (1984).
- 40 Yamashita, M. & Fenn, J. B. Electrospray Ion Source. Another variation on the Free-Jet Theme. *J. Phys. Chem.* **88**, 4451-4459 (1984).
- 41 Taylor, G. DISINTEGRATION OF WATER DROPS IN ELECTRIC FIELD. *Proceedings of the Royal Society of London Series a-Mathematical and Physical Sciences* **280**, 383+, doi:10.1098/rspa.1964.0151 (1964).
- 42 Fernández de La Mora, J. The fluid dynamics of Taylor cones. *Annu. Rev. Fluid Mech.* **39**, 217-243 (2007).
- 43 Tang, K. & Gomez, A. Generation of monodisperse water droplets from electrosprays in a corona-assisted cone-jet mode. *Journal of Colloid and Interface Science* **175**, 326-332 (1995).
- 44 Rayleigh, L. On the Equilibrium of Liquid Conducting Masses charged with Electricity. *Philos. Mag.* **14**, 184-186 (1882).
- 45 Gomez, A. & Tang, K. Charge and fission of droplets in electrostatic sprays. *Phys. Fluids* **6**, 404-414 (1994).
- 46 Li, J. *et al.* Conformational effects in protein electrospray-ionization mass spectrometry. *Mass Spectrometry Reviews* (2015).
- 47 Consta, S., Oh, M. I. & Soltani, S. Advances in the theoretical and molecular simulation studies of the ion chemistry in droplets. *International Journal of Mass Spectrometry* (2014).
- 48 Konermann, L., Ahadi, E., Rodriguez, A. D. & Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **85**, 2-9 (2013).
- 49 Hogan, C. J., Carroll, J. A., Rohrs, H. W., Biswas, P. & Gross, M. L. Combined Charged Residue-Field Emission Model of Macromolecular Electrospray Ionization. *Anal. Chem.* **81**, 369-377 (2009).

- 50 Iribarne, J. V. & Thomson, B. A. On the evaporation of small ions from charged droplets. *J. Chem. Phys.* **64**, 2287-2294 (1976).
- 51 Nguyen, S. & Fenn, J. B. Gas-phase ions of solute species from charged droplets of solutions. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1111-1117 (2007).
- 52 Konermann, L., McAllister, R. G. & Metwally, H. Molecular Dynamics Simulations of the Electrospray Process: Formation of NaCl Clusters via the Charged Residue Mechanism. *J. Phys. Chem. B* **118**, 12025-12033, doi:10.1021/jp507635y (2014).
- 53 de la Mora, F. J. Electrospray Ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal. Chim. Acta* **406**, 93-104 (2000).
- 54 Yue, X., Vahidi, S. & Konermann, L. Insights into the Mechanism of Protein Electrospray Ionization From Salt Adduction Measurements. *J. Am. Soc. Mass Spectrom.* **25**, 1322-1331, doi:10.1007/s13361-014-0905-0 (2014).
- 55 Daub, C. D. & Cann, N. M. How Are Completely Desolvated Ions Produced in Electrospray Ionization: Insights from Molecular Dynamics Simulations. *Anal. Chem.* **83**, 8372-8376 (2011).
- 56 Kaltashov, I. A., Bobst, C. E. & Abzalimov, R. R. Mass spectrometry-based methods to study protein architecture and dynamics. *Protein Sci.* **22**, 530-544, doi:10.1002/pro.2238 (2013).
- 57 Kaltashov, I. A. & Abzalimov, R. R. Do Ionic Charges in ESI MS Provide Useful Information on Macromolecular Structure? *J. Am. Soc. Mass Spectrom.* **19**, 1239-1246 (2008).
- 58 Natalello, A., Benetti, F., Doglia, S. M., Legname, G. & Grandori, R. Compact conformations of alpha-synuclein induced by alcohols and copper. *Proteins-Structure Function and Bioinformatics* **79**, 611-621, doi:10.1002/prot.22909 (2011).
- 59 McLafferty, F. W. Tandem mass spectrometry. *Science* **214**, 280-287 (1981).
- 60 Hunt, D. F., Yates III, J. R., Shabanowitz, J., Winston, S. & Hauer, C. R. Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6233-6237 (1986).
- 61 Hambly, D. M. & Gross, M. L. in *Protein Mass Spectrometry* Vol. 52 (ed J. P. Whitelegge) 151-177 (Elsevier, 2009).
- 62 Leitner, A. *et al.* Probing Native Protein Structures by Chemical Cross-linking, Mass Spectrometry, and Bioinformatics. *Mol. Cell. Proteomics* **9**, 1634-1649, doi:10.1074/mcp.R000001-MCP201 (2010).
- 63 Giles, K., Williams, J. P. & Campuzano, I. Enhancements in travelling wave ion mobility resolution. *Rapid Commun. Mass Spectrom.* **25**, 1559-1566 (2011).

- 64 Bohrer, B. C., Merenbloom, S. I., Koeniger, S. L., Hilderbrand, A. E. & Clemmer, D. E. Biomolecule Analysis by Ion Mobility Spectrometry. *Annu. Rev. Anal. Chem.* **1**, 293-327 (2008).
- 65 Shvartsburg, A. A. & Jarrold, M. F. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **261**, 86-91, doi:10.1016/0009-2614(96)00941-4 (1996).
- 66 Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. Primary Structure Effects on Peptide Group Hydrogen Exchange. *Proteins: Struct., Funct., Genet.* **17**, 75-86 (1993).
- 67 Hvidt, A. & Nielsen, S. O. Hydrogen exchange in proteins. *Adv. Protein Chem.* **21**, 287-386 (1966).
- 68 Hvidt, A. & Linderstrom-Lang, K. Exchange of hydrogen atoms in insulin with deuterium atoms in aqueous solutions. *Biochem. Biophys. Acta* **14**, 574-575 (1954).
- 69 Xiao, Y. & Konermann, L. Protein structural dynamics at the gas/water interface examined by hydrogen exchange mass spectrometry. *Protein Science* (2015).
- 70 Cravello, L., Lascoux, D. & Forest, E. Use of different proteases working in acidic conditions to improve sequence coverage and resolution in hydrogen/deuterium exchange of large proteins. *Rapid Commun. Mass Spectrom.* **17**, 2387-2393, doi:10.1002/rcm.1207 (2003).
- 71 Pan, J., Han, J., Borchers, C. H. & Konermann, L. Electron Capture Dissociation of Electrosprayed Protein Ions for Spatially-Resolved Hydrogen Exchange Measurements. *J. Am. Chem. Soc.* **130**, 11574-11575 (2008).
- 72 Cramer, C. J. *Essentials of Computational Chemistry: Theory and Models*. 2nd edn, (John Wiley & Sons, Ltd., 2004).
- 73 Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **98**, 5648-5652 (1993).
- 74 Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature Structural & Molecular Biology* **9**, 646-652 (2002).
- 75 McCammon, J. A. Dynamics of folded proteins. *Nature* **267**, 16 (1977).
- 76 Piana, S., Lindorff-Larsen, K. & Shaw, D. E. Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5915-5920, doi:10.1073/pnas.1218321110 (2013).
- 77 Tolman, R. C. *The principles of statistical mechanics*. (Courier Corporation, 1938).
- 78 Birdsall, C. K. & Langdon, A. B. *Plasma physics via computer simulation*. (CRC Press, 2014).
- 79 Verlet, L. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical review* **159**, 98 (1967).

- 80 Jorgensen, W. L. & Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society* **110**, 1657-1666 (1988).
- 81 MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B* **102**, 3586-3616 (1998).
- 82 Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **117**, 5179-5197 (1995).
- 83 Patra, M. *et al.* Molecular dynamics simulations of lipid bilayers: major artifacts due to truncating electrostatic interactions. *Biophysical journal* **84**, 3636-3645 (2003).
- 84 Darden, T., Perera, L., Li, L. & Pedersen, L. New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure* **7**, R55-R60 (1999).
- 85 Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **369**, 253-287 (1921).
- 86 Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of chemical physics* **103**, 8577-8593 (1995).
- 87 Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *The Journal of chemical physics* **98**, 10089-10092 (1993).
- 88 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935 (1983).
- 89 Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269-6271 (1987).
- 90 Vega, C. & de Miguel, E. Surface tension of the most popular models of water by using the test-area simulation method. *J. Chem. Phys.* **126**, 154707, doi:10.1063/1.2715577 (2007).
- 91 Mahoney, M. W. & Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **112**, 8910-8922 (2000).
- 92 Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684-3690, doi:10.1063/1.448118 (1984).
- 93 Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 0141011-0141017, doi:10.1063/1.2408420 (2007).

- 94 Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of chemical physics* **72**, 2384-2393 (1980).
- 95 Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695-1697 (1985).
- 96 Nose, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255-268 (1984).
- 97 Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **159**, 98-103 (1967).
- 98 Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327-341 (1977).
- 99 Miyamoto, S. & Kollman, P. A. SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of computational chemistry* **13**, 952-962 (1992).
- 100 Hess, B., Bekker, H., Berendsen, H. J. & Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry* **18**, 1463-1472 (1997).

## 2 Challenges in the Interpretation of Protein H/D Exchange Data: A Molecular Dynamics Simulation Perspective

### 2.1 Introduction

Backbone H/D exchange (HDX) measurements are widely used for studying protein structure and dynamics. Both NMR spectroscopy<sup>1-4</sup> and mass spectrometry<sup>5-14</sup> can serve as detection methods. Ideally, these measurements yield the HDX rate constant  $k_{HDX}$  for each single NH site.

In short peptides the deuteration kinetics are governed by the two side chains adjacent to the amide of interest. HDX in near-neutral solution proceeds with  $OD^-$  catalysis through a  $R-C(O^-)=N-R$  imidate that subsequently interacts with  $D_2O$  to form  $R-CO-ND-R$ .<sup>17</sup> Positive charge density in the vicinity of the NH lowers the amide  $pK_a$  by stabilizing the imidate, thereby accelerating HDX. The opposite effect is encountered for negative charge.<sup>18-20</sup> Accordingly, the interactions between a peptide NH and its two adjacent side chains can be attributed to inductive effects, with some modulation by steric factors.<sup>21</sup> Empirical rules have been established to describe how these nearest neighbor interactions determine the second-order rate constant  $k_B$ , resulting in an overall peptide “chemical” rate constant<sup>21</sup>

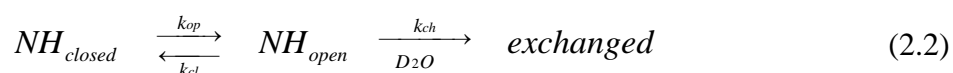
$$k_{ch} = k_B \times [OD^-] \quad (2.1)$$

Additional considerations are required to understand HDX in proteins. Disordered segments are said to be in an “open” state, and they exchange with rates close to those expected for short peptides ( $k_{HDX} \approx k_{ch}$ ).<sup>21</sup> In contrast, HDX in structured regions tends to proceed much more slowly. The factors contributing to the protection of these amides remain controversial. It is instructive to briefly highlight some of the pertinent issues.

Solvent accessibility is often quoted as an important determinant of HDX rates.<sup>5,24-28</sup> This notion is in sharp contrast to the view that protection is chiefly governed by H-bonding,<sup>29</sup> typically via backbone  $NH \cdots OC$  contacts.<sup>16</sup> Experiments confirm that H-bonded amides



usually exhibit strong protection, even if they are at the surface.<sup>29</sup> The H-bond-centric view emphasizes the role of conformational dynamics. Accordingly, NH groups in structured segments predominantly reside in a “closed” (exchange-incompetent) state, but they can undergo transient opening events that briefly disrupt H-bonds and provide NH contact with the solvent. HDX-relevant fluctuations can range from local events all the way to global unfolding/refolding.<sup>2</sup> During the brief time intervals that NH sites spend in the “open” state they may undergo deuteration with the rate constant  $k_{ch}$  (equation 2.1). The resulting HDX mechanism can be expressed as<sup>30</sup>



This framework yields an EX2 rate constant  $k_{HDX} = k_{op}/k_{cl} \times k_{ch}$ . The extent of protection relative to the peptide-calibrated state can be reported as<sup>8,16,29</sup>

$$\log P = \log(k_{ch} / k_{HDX}) \quad (2.3)$$

Although the H-bond-centric view is quite well established,<sup>3,8,16,29</sup> the relative importance of steric shielding (without H-bonding, e.g., in collapsed regions<sup>31</sup>) remains an open question.

The above considerations notwithstanding, some amides exchange very slowly, despite being apparently solvent accessible and free of H-bonds.<sup>29,32-34</sup> This phenomenon is difficult to reconcile with traditional views of the HDX process. It has been proposed that such anomalous protection arises from tertiary interactions that provide a unique electrostatic environment, thereby affecting the NH pK<sub>a</sub> in ways that go beyond the nearest neighbor effects seen in peptides.<sup>21</sup> This idea shows promise in some cases,<sup>32,35</sup> but fails in others.<sup>29</sup> A competing proposal<sup>29,36</sup> dismisses the importance of tertiary electrostatic contacts. Instead, it envisions that exposed NH sites can be protected by H-bonding to crystallographically defined waters at the protein surface. This alternative view re-emphasizes the role of H-bonding for HDX protection, by including tightly bound waters as suitable H-bond acceptors.<sup>37,38</sup> The idea of NH protection via water contacts is intriguing, but it has not been thoroughly tested yet.<sup>35</sup>

The inter-related issues outlined above complicate the interpretation of HDX experiments. Very basic concepts such as the exact nature of “closed” and “open” conformations remain nebulous. These difficulties may be partly rooted in the fact that HDX data are often interpreted in the context of static X-ray structures.<sup>11,16,25,27,35,36</sup> Crystal packing effects can significantly distort amide exposure and H-bonding.<sup>36,37</sup> Hence, X-ray structures may not properly represent the thermally activated ensemble encountered under ambient solution conditions. Molecular dynamics (MD) simulations<sup>15,39-44</sup> and related computational approaches<sup>26,38,45</sup> can help address these shortcomings by providing a dynamic view of the protein and its solvent environment. Unfortunately, the  $\mu\text{s}$  –  $\text{ms}$  range accessible in all-atom MD simulations<sup>46</sup> is much shorter than most HDX-relevant conformational fluctuations.<sup>29</sup> Coarse-grained simulations provide better conformational sampling<sup>15</sup> but lack structural details, thereby complicating comparisons with experimental data. Despite these limitations, MD data should provide a better comparison basis for the interpretation of HDX data than static X-ray structures.

Here we conduct all-atom MD simulations on ubiquitin to examine how well the solution phase behavior of this protein correlates with its HDX protection pattern. Ubiquitin is relatively small (76 residues, 72 backbone amides) with a well-defined native fold.<sup>47</sup> It was chosen for this work because its HDX behavior has been characterized in great detail, with  $\log P$  values that span more than six orders of magnitude.<sup>14,16,22,23,33</sup> The disordered C-terminus lacks protection and thus serves as internal standard. Our MD simulations provide a dynamic view of the protein ground state in solution. These data allow us to scrutinize on an amide-by-amide basis in how far HDX protection can be attributed to H-bonding, solvent accessibility, crystallographic waters, or other factors. We find that the interpretation of HDX data in a protein conformational context is surprisingly difficult; 21% of all backbone amides exhibit  $\log P$  values that are seemingly inconsistent with their structural environment. Our findings highlight severe shortcomings in the current understanding of structure-rate relationships, and they caution against a non-critical use of HDX data for deducing protein conformational features.

## 2.2 Methods

### 2.2.1 MD Simulations

All-atom ubiquitin simulations in explicit water were conducted using GROMACS 4.6.5.<sup>48,49</sup> Like other MD studies on ubiquitin<sup>46</sup> we used the 1.8 Å crystal structure 1UBQ<sup>47</sup> as starting point. Hydrogens were added using the PDB2GMX routine. All titratable moieties were set to their standard charge states expected for pH 7 (N-terminus<sup>+</sup>, R<sup>+</sup>, K<sup>+</sup>, D<sup>-</sup>, E<sup>-</sup>, C-terminus<sup>-</sup>) with H68 in its deprotonated state, resulting in a net protein charge of zero. The discussion below will focus on data generated using the CHARMM22\* force field<sup>50</sup> with TIP3P water.<sup>51</sup> In addition, simulations were conducted using Amber99sb-ILDN<sup>52</sup> with TIP4P water.<sup>51</sup> None of these (or any other commonly used) water models accounts for the self-dissociation of H<sub>2</sub>O into H<sup>+</sup> and OH<sup>-</sup>.<sup>53,54</sup> The protein was placed in a rhombic dodecahedral periodic box, with a minimum distance of 7 Å between protein atoms and the edge of the box. ~4400 water molecules were added from a pre-equilibrated configuration file. Randomly selected solvent molecules were replaced with Na<sup>+</sup> or Cl<sup>-</sup> for a total salt concentration of 150 mM. The systems were subjected to energy minimization for 800 iterations, prior to 100 ps of equilibration at 298 K and 1 bar using a velocity-rescaling thermostat<sup>55</sup> and Berendsen barostat.<sup>56</sup> Initial velocities were sampled from a Maxwell-Boltzmann distribution. 1 μs production runs were initialized from the equilibrated systems, using leap-frog integration with a time step of 2 fs. These simulations were carried out under NVT conditions with velocity rescaling at 298 K.<sup>55</sup> Bonds were constrained using the linear constraint solver algorithm for protein<sup>57</sup> and the SETTLE algorithm for water.<sup>58</sup> Short range interactions were modeled with a Lennard-Jones 10 Å potential-shifted cut-off, while electrostatics were modeled using the smooth particle-mesh Ewald method<sup>59</sup> with a grid spacing of 2.4 Å. The coordinates of all atoms were recorded every 2 ps for analysis.

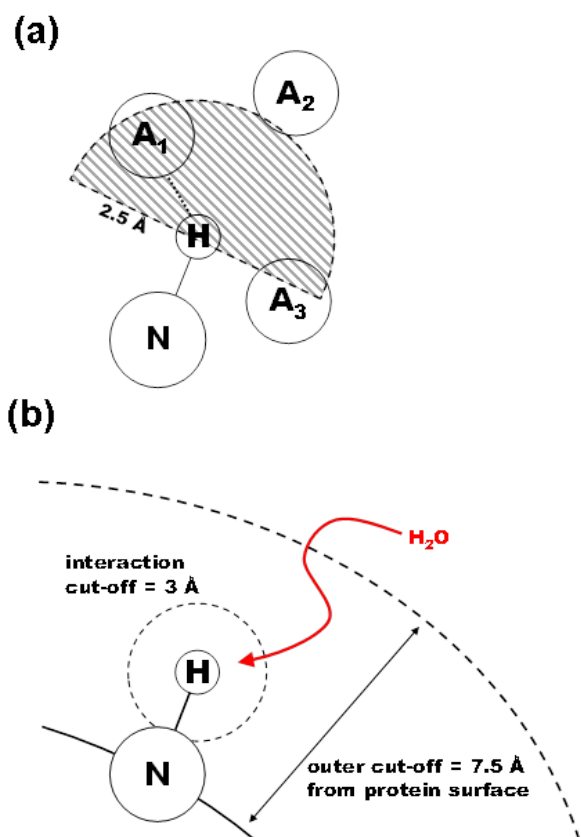
### 2.2.2 Data Analysis

An in-house program was written in C++ to detect backbone amide H-bonds. The program reports the distance between each amide hydrogen and the closest possible H-bond acceptor (A) that satisfies the N-H-A angle criterion of 90° - 180°.<sup>60</sup> A NH is

considered to be H-bonded if the angle criterion is satisfied while at the same time the H/A distance is less than 2.5 Å (Figure 2.1a). All oxygen and nitrogen atoms, including those in side chains, were considered as possible A moieties. The program was designed to record the fraction of time that each NH site interacts with each possible acceptor. When discussing H-bonds between two residues we will use the convention of naming the NH donor first, followed by the acceptor.

A second program was developed to measure the solvent accessible surface area (SASA) of specified atoms at each time point using the double cubic lattice method.<sup>61</sup> This technique models each atom as a sphere, with previously published effective van der Waals radii that represent protein interactions with water.<sup>62</sup> SASA values discussed below refer to the sum of the amide nitrogen and hydrogen for each residue.

To investigate the possible presence of tightly bound water networks at solvent-exposed NH sites<sup>29,36</sup> a “bulk interaction rate” was determined. This parameter represents a measure of how rapidly H<sub>2</sub>O (or D<sub>2</sub>O) molecules can diffuse from bulk solution into the direct vicinity of an NH site. For implementing this strategy, each water within 7.5 Å of



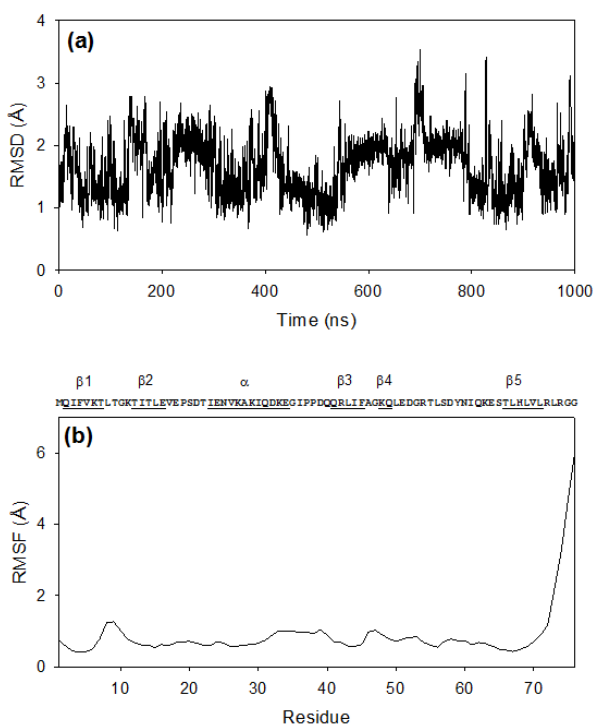
**Figure 2.1 Schematic of H-bond and bulk interaction rate algorithms.** (a) H-bond detection algorithm. An NH is considered to be H-bonded if a possible acceptor A is located within the shaded semi-sphere (exemplified by A1). A2 is too far away; A3 falls outside the permitted angular range. (b) Cartoon description of the algorithm used for assessing how rapidly bulk solvent can replace water molecules adjacent to a NH site (“bulk interaction rate”). One possible trajectory of a bulk H<sub>2</sub>O (or D<sub>2</sub>O) is indicated in red.

the protein surface was tagged. Next, we determined how much time elapsed before the first *untagged* water came within 3 Å of the amide hydrogen (Figure 2.1b). This algorithm is only meaningful for NH sites that are not permanently buried. The analysis was therefore only performed for residues with greater-than-mean SASA values. The calculations were repeated 1000 times with starting points that were spaced by 1 ns. The mean of these replicates was used to determine the interaction rate in units of  $\text{ps}^{-1}$ .

## 2.3 Results and Discussion

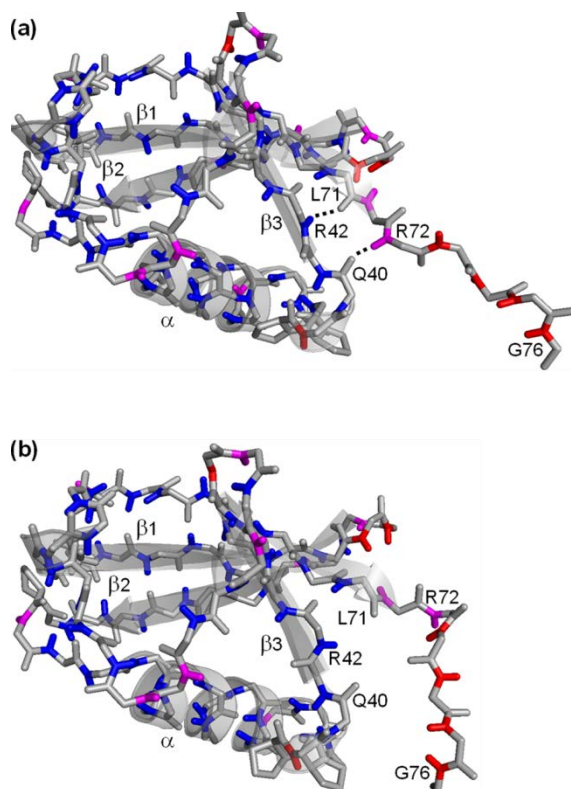
### 2.3.1 Ubiquitin Structure and Dynamics

All-atom MD simulations of native ubiquitin in explicit water were conducted at 298 K. During the 1  $\mu\text{s}$  simulation window the protein remained relatively close to its initial



**Figure 2.2 RMSD and RMSF plots for ubiquitin simulation.** (a) Root mean square deviation of ubiquitin backbone atoms relative to the energy-minimized crystal structure. (b) Root mean square fluctuation of backbone atoms averaged for each residue. Panel (b) also shows the protein sequence and key secondary structure elements.

structure, with root-mean-square deviation (RMSD) values between 1 and 3 Å (Figure 2.2a). Global unfolding/refolding transitions were not observed. This behavior is consistent with the known high stability of ubiquitin.<sup>47</sup> A root mean square fluctuation (RMSF) plot further illustrates the rigid nature of the protein, with RMSF values of no more than 1 Å for most residues (Figure 2.2b). Only the C-terminal tail (residues 73-76) is more flexible, in agreement with existing structural data.<sup>47</sup> Local regions occasionally undergo opening/closing transitions, as envisioned by eq. 2. For example, the H-bonds R42-L71 and R72-Q40 are intact at 944 ns (Figure 2.3a),

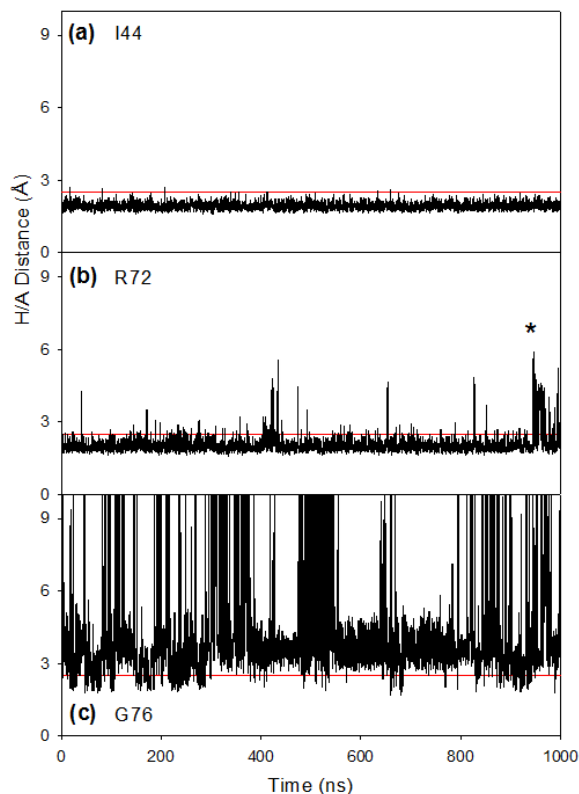


**Figure 2.3 A structural fluctuation opens H-bonds near ubiquitin's C-terminus.** MD snapshots for  $t = 944$  ns (a) and  $t = 946$  ns (b). Side chains are omitted. Backbone NH groups are colored according to their experimental protection factors (from Figure 2.5a): blue, strongly protected; pink, weakly protected; red, unprotected. Panel (a) highlights two backbone H-bonds, R42-L71 and R72-Q40. In (b) these two H-bonds are disrupted due to a local fluctuation.

whereas both contacts are broken at 946 ns (Figure 2.3b). A few ns later both contacts are regenerated (not shown in Figure 2.3).

The H-bonding status of all backbone NH sites was tracked as a function of time. For each residue we identified the closest possible acceptor A that satisfied the  $90^\circ - 180^\circ$  angle criterion, and we plotted the distance of this atom to the amide hydrogen. H/A distances below  $2.5 \text{ \AA}$  imply the presence of a H-bond (Figure 2.1a).<sup>60</sup> Figure 4 exemplifies three of the profiles obtained in this way. I44 is permanently H-bonded, R72 is H-bonded most of the time, whereas G76 is almost completely free. This analysis was conducted for all NH sites, yielding average values that are summarized in Figure 2.5 (discussed below).

Opening/closing events of the type illustrated in Figure 2.3 are quite rare in our simulations. Many sites (such as I44, Figure 2.4a) remain permanently H-bonded on the time scale considered here, despite undergoing deuteration with finite  $k_{HDX}$  values.<sup>14,16,22,23,33</sup> This reflects the well-known fact that all-atom MD studies cannot adequately sample all the conformational events probed by HDX, where labeling times extend to days.<sup>14,16,22,23,33</sup> Even state-of-the-art ( $\sim 1$  ms)<sup>46</sup> simulations are orders of magnitude too short for this purpose.<sup>29</sup> Fluctuations can be enhanced by running simulations at elevated temperature,<sup>46</sup> but such semi-denaturing



**Figure 2.4 H-bonding properties of three NH sites.** Shown is the amide hydrogen distance to the closest carbonyl acceptor (“A”) that falls within the permissible angular range (see Figure 2.1a). H/A distances below 2.5 Å (indicated by red lines) reflect the presence of a H-bond. (a) Data for I44, with H-bonding to H68 for 100% of the time; (b) R72, with 93% H-bonding to Q40. The asterisk marks the opening event of Figure 2.3. (c) G76, with 7% H-bonding to L73 and 1% to D39.

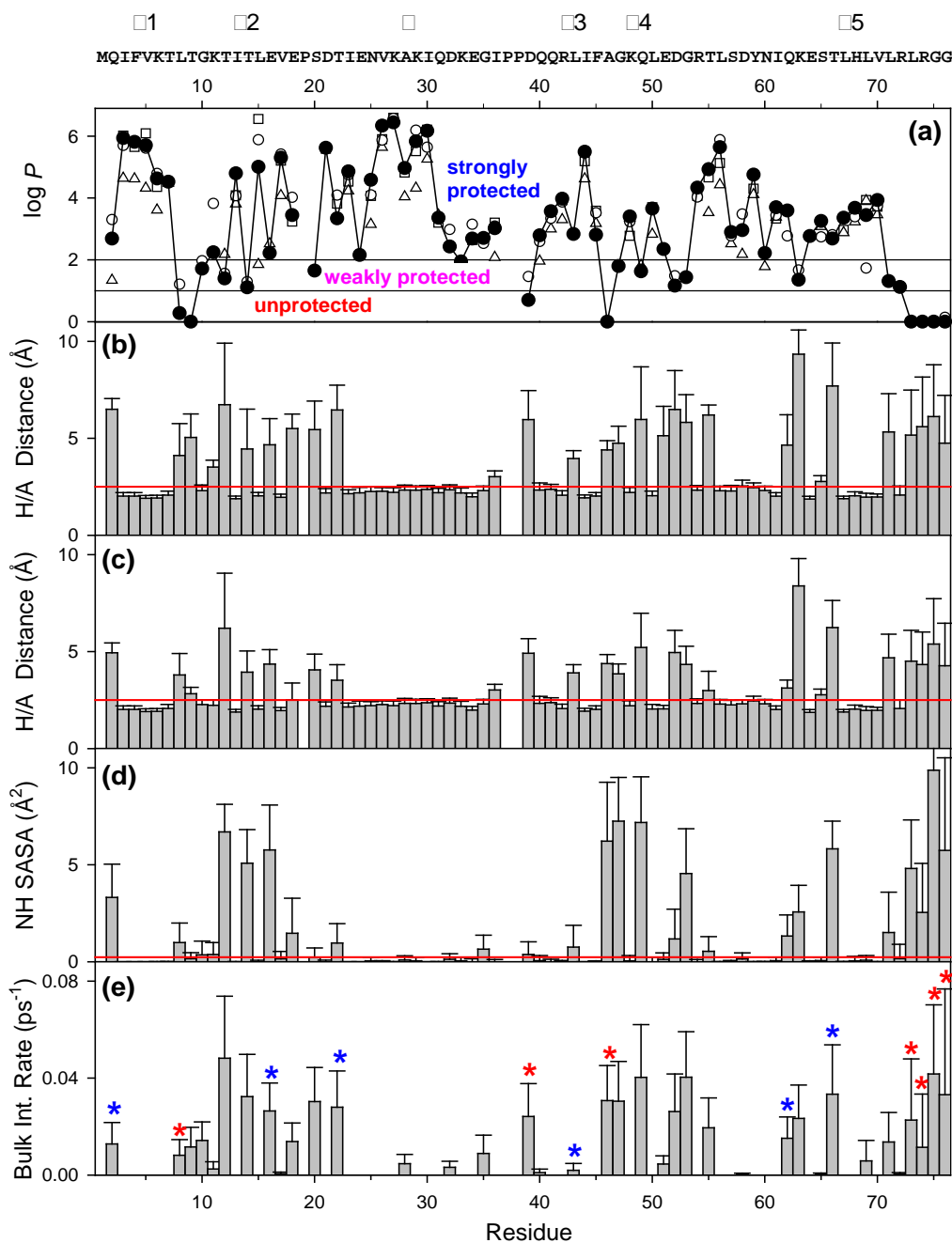
conditions favor large-scale dynamics that are not adequate for modeling a native HDX environment.<sup>2</sup>

In summary, although our data do not report on slow opening/closing transitions, they provide rich information on the ubiquitin ground state ensemble and its solvent environment. NH sites that are experimentally found to be protected must predominantly reside in a “closed” state under the simulation conditions used here. Conversely, sites with  $\log P \approx 0$  must be “open”.<sup>2,30</sup> Our trajectories, therefore, allow a detailed examination of the structural features that cause NH sites to be “open” or “closed”.

### 2.3.2 Experimental Results

Ubiquitin HDX/NMR data are available from several sources.<sup>16,22,23,33</sup>

The  $\log P$  values from different laboratories agree with each other quite well (Figure 2.5a), and they are also consistent with recent top-down mass spectrometry experiments.<sup>14</sup> Craig et al.<sup>15</sup> compiled an averaged  $\log P$  profile that will serve as foundation for the following considerations (filled symbols in Figure 5a). To simplify the discussion we categorize amides according to their experimental behavior. NH groups with  $\log P < 1$  are considered to be *unprotected*. The remaining sites are broken down into *weakly protected* ( $1 < \log P < 2$ ) and *strongly protected* ( $\log P > 2$ ). The locations of these NH groups are highlighted in Figure 3 using red, pink, and blue, respectively.



**Figure 2.5 Overview of experimental HDX data, and various properties extracted from a 1  $\mu$ s CHARMM22\*/TIP3P simulation.** (a) Experimental protection factors compiled by Craig et al.<sup>15</sup> are highlighted as solid circles. Also shown are individual data sets as open triangles,<sup>16</sup> open circles,<sup>22</sup> and open squares.<sup>23(4, 5)</sup> (b) Average distance between the backbone NH of each residue and the closest possible main chain carbonyl acceptor. Values below 2.5 Å (red line) reflect the presence of an H-bond. (c) Same as in panel b, but including side chains. (d) Average NH SASA values. The red line represents the SASA threshold of 0.23 Å<sup>2</sup>. (e) Average bulk interaction rate, describing how fast bulk water diffuses to an NH site. Red asterisks represent exposed amides that are unprotected. Blue asterisks represent “problem cases”, where exposed amides are strongly protected. Error bars shown in all panels represent standard deviations.



### 2.3.3 Main Chain H-Bonds

MD simulations reveal that 42 residues are H-bonded via main chain NH $\cdots$ OC contacts, evident from H/A distance values that fall below the 2.5 Å threshold<sup>60</sup> in Figure 2.5b. These NH sites are highlighted green in the Figure 2.6a. All of them fall into the strongly or weakly protected HDX/NMR categories. Many of these sites are located in the  $\alpha$  helix and the five-stranded  $\beta$  sheet, consistent with the view<sup>16</sup> that HDX protection is often correlated with the presence of secondary structure.

### 2.3.4 Side Chain H-Bonds

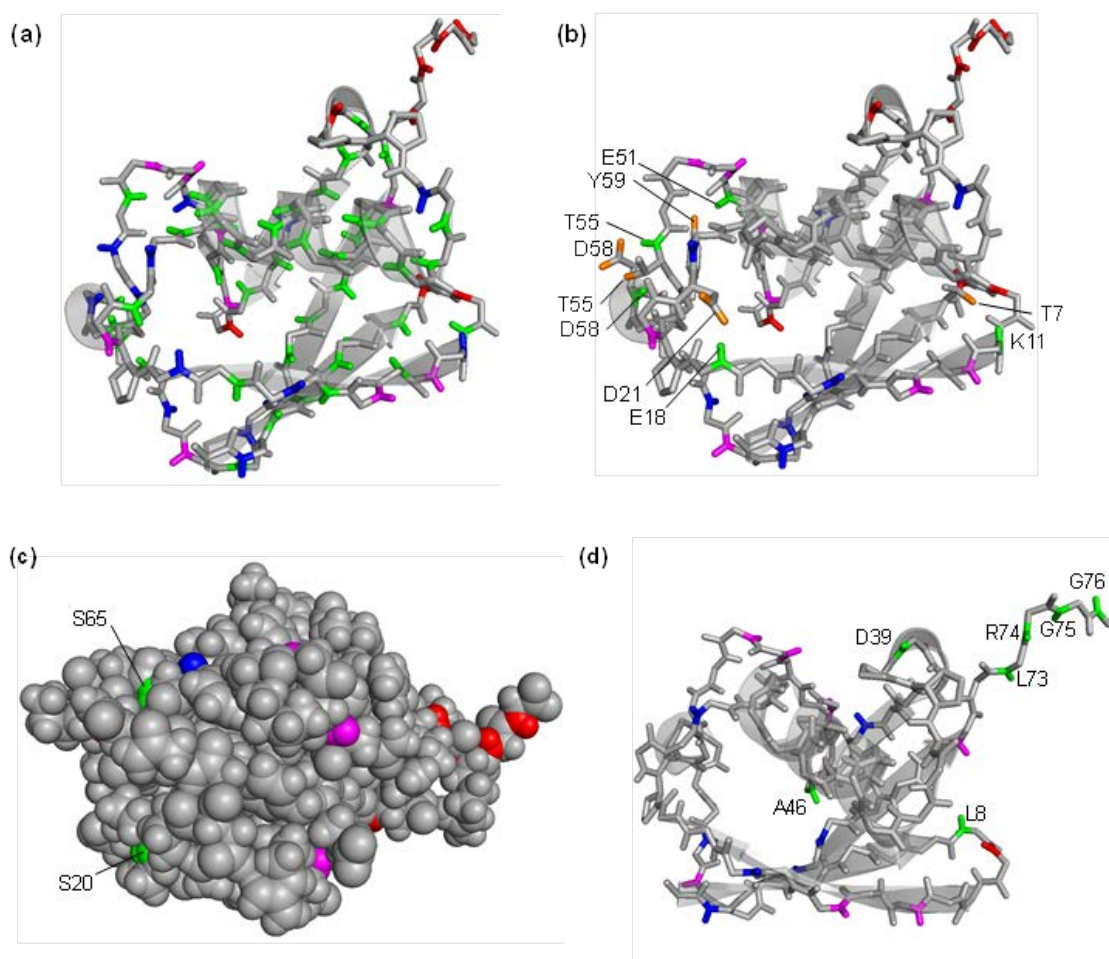
In addition to main chain contacts, NH interactions with side chains are encountered (Figure 2.6b). Experiments show that K11, E51, and D58 are protected (Figure 2.5a), despite lacking H-bonds to main chain carbonyls (Figure 2.5b). However, inclusion of side chain atoms in our analysis lowers their H/A distance below the 2.5 Å threshold (Figure 2.5c). Specifically, K11 is H-bonded to the T7 side chain 83% of the time, E51 is 98% bonded to the hydroxyl group of Y59, and D58 alternates between bonding to the T55 side chain (48%) and the T55 main chain carbonyl (38%).

E18 and T55 are strongly protected, and they interact with aspartate side chains. The corresponding contacts (E18-D21 and T55-D58) fall short of the 2.5 Å distance threshold (Figure 2.5c). These cases represent examples of “bifurcated” H-bonds, where a backbone NH interacts with both oxygens of a carboxylate. It has been noted previously that this type of H-bond goes undetected when applying standard geometric criteria.<sup>60</sup> Our data nonetheless underscore the importance of bifurcated H-bonds for HDX protection.

The MD results discussed so far confirm that H-bonding generally leads to protection. Interestingly, more than 10% of the HDX-relevant H-bonds do not constitute main chain contacts but side chain interactions, including bifurcated H-bonds. Yet, it is not possible to explain the protection pattern of ubiquitin solely on the basis of H-bonding. Eight strongly protected NH groups remain unaccounted for, as well as nine weakly protected sites (blue and pink, Figure 2.6b).

### 2.3.5 Solvent Accessibility

In an effort to rationalize the HDX properties of the remaining sites, we next examined whether a low solvent accessibility might contribute to protection of amide groups that are not H-bonded. NH SASA values range from zero for permanently buried sites all the way to  $\sim 10 \text{ \AA}^2$  in the disordered C-terminal tail (Figure 2.5d). It is not unreasonable to assume that SASA values below a certain threshold will slow down HDX by inhibiting



**Figure 2.6** Energy-minimized structure of ubiquitin, where blue/pink/red coloring denotes the level of HDX/NMR protection as in Figure 2.3. Green represents NH sites with experimental HDX properties that are consistent with features seen in the MD simulations. These “explainable” sites are then grayed out in the subsequent panels. (a) Green: Protected sites that are involved in main chain NH.....OC H-bonds. (b) Green: Protected sites that are H-bonded to side chains. Orange: Side chain H-bond acceptors. (c). Green: NH sites protected by low solvent accessibility. I36 is not visible in this all-atom spacefill representation. (d) Green: Unprotected sites that are solvent accessible and not H-bonded.

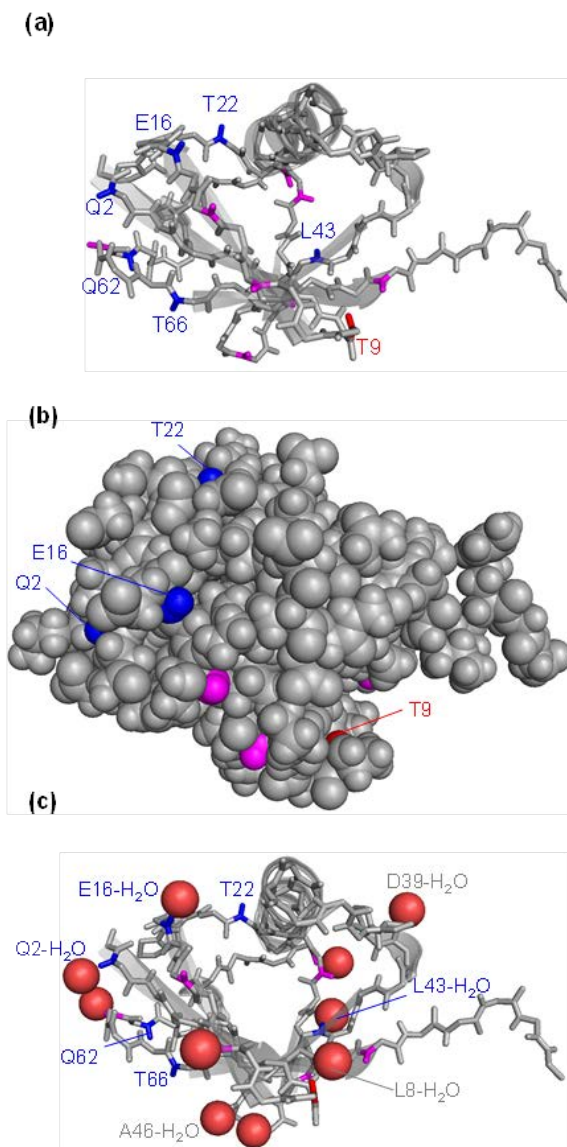
NH contact with catalyst and solvent.<sup>5,24,25,27-29</sup> Our data offer some clues as to what a suitable threshold might be. D39 is unprotected with SASA = 0.38 Å<sup>2</sup>, whereas S20 with SASA = 0.23 Å<sup>2</sup> is protected (Figure 2.5). This suggests that protection should be a significant factor for SASA values around 0.23 Å<sup>2</sup> and below. When adopting this threshold one can identify I36 and S65 as being occluded from the solvent (in addition to S20). All three residues are protected according to HDX/NMR, despite not being H-bonded (Figure 2.5). Readers might object to our somewhat heuristic choice of a SASA threshold. Indeed, it will be discussed below that the properties of T9 are not in line with the explanation attempts offered here. Our analysis is nonetheless compatible with the view that the low solvent accessibility of the non-H-bonded residues S20, I36, and S65 is a contributing factor to their experimentally observed protection (Figure 2.6c).

### 2.3.6 Unprotected Sites

The above considerations focused on sites that are protected according to HDX/NMR. We attempted to ascribe this protection to H-bonding and NH burial. It is also instructive to pursue the opposite strategy, i.e., examine unprotected sites and identify the reason(s) underlying their high HDX rates. From the data compiled in Figure 2.5 one can identify seven residues that have log *P* values close to zero, with high SASA values and no H-bonding. This group comprises L8, D39, A46, as well as the C-terminal tail (L73 - G76). In Figure 2.6d the corresponding amides are highlighted in green. These residues conform to the commonly held expectation that a lack of both H-bonding and burial will render NH sites prone to rapid exchange.

### 2.3.7 Problem Cases

The preceding discussion successfully addressed 79% of the ubiquitin backbone sites. Unfortunately, H-bonding and SASA considerations fail for the remaining 21%. In other words, 15 out of 72 amides do not behave in accordance with classical HDX expectations (Figure 2.7a). Most of these problem cases are NH groups that show experimental protection, despite having high solvent accessibilities and no H-bonding. This behavior is most dramatically illustrated by Q2, E16, T22, L43, Q62, and T66, which have log *P* values between 2.2 and 3.6. The opposite problem is encountered for T9, where log *P* =



**Figure 2.7 Problem cases and waters resolved in the crystal structure.** (a) Colored NH sites are problem cases, with HDX/NMR properties that are inconsistent with their H-bonding or SASA characteristics. Not all amides are annotated to prevent clutter. (b) Same as in (a), but in all-atom spacefill representation. (c) Ten crystallographically defined waters (red) are H-bonded to backbone amides in the pdb file 1UBQ. Only three out of six strongly protected surface amides interact with water molecules. Highlighted in gray are water-bound surface amides that are unprotected.

0. This residue is not H-bonded, but it is situated in a narrow pocket with a low SASA value of  $0.16 \text{ \AA}^2$  (Figure 2.7b). According to the threshold value identified above these conditions should provide significant protection. For comparison, S20 (SASA =  $0.23 \text{ \AA}^2$ ) exhibits  $\log P = 1.6$  without H-bonding.

### 2.3.8 Crystallographically Defined Water Molecules

It has been proposed that H-bonding of exposed amides to crystallographically defined waters can provide HDX protection.<sup>29,36</sup> This proposal envisions that specific solvent molecules are immobilized at the protein surface not only in the crystal, but also in bulk solution.<sup>35</sup> To examine this idea we inspected the X-ray structure of ubiquitin, which comprises 58 crystal waters.<sup>47</sup> Ten of these are H-bonded to exposed NH groups (Figure 2.7c). Is the presence of these H<sub>2</sub>O molecules correlated with HDX protection? Of the six exposed NH sites that exhibit anomalously high protection (Figure 2.7c, blue), three are H-bonded to crystal water (Q2, E16, and L43); the remaining three are not engaged in

defined solvent contacts (T22, Q62, and T66). Focusing on the weakly protected problem cases (Figure 2.7c, pink), it is seen that four of them (T12, G47, G53, and K63) interact with crystal waters, whereas no water contacts are evident for T14, Q49, D52, and L71. Most importantly, L8, D39, and A46 are H-bonded to crystal waters, despite exhibiting  $\log P$  values close to zero (Figure 2.7c, gray). This analysis demonstrates a complete lack of correlation between NH protection and H-bonding to crystallographically defined water molecules.

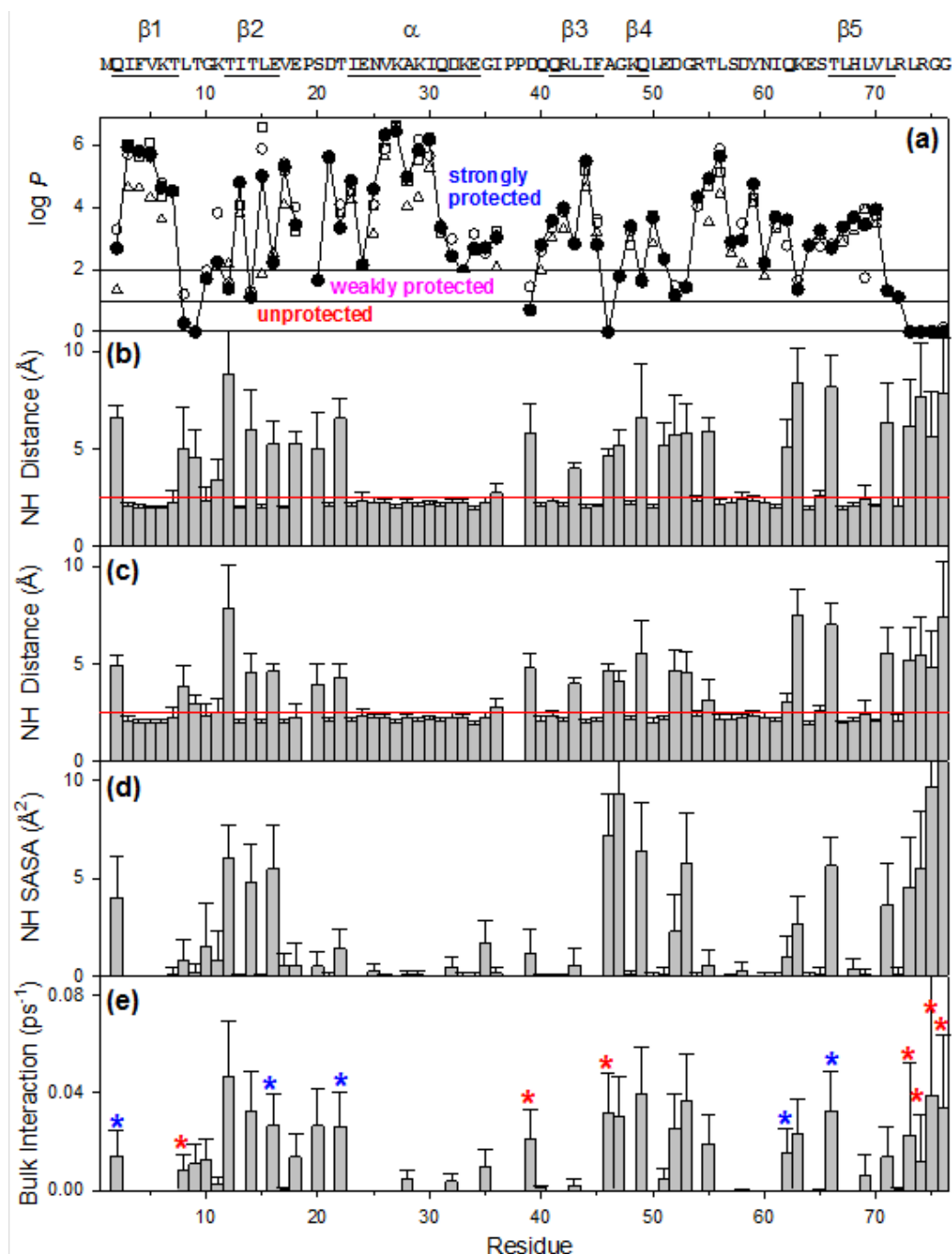
A discussion of NH hydration on the basis of isolated crystal waters may be too simplistic. Some studies suggest that protein surfaces in solution can give rise to H-bonded water cages, resulting in solvent regions that exhibit retarded exchange with the bulk.<sup>63-65</sup> It seems conceivable that such solvent cages might be able to provide HDX protection for exposed amides. To test for the presence of such protecting solvent cages we characterized the water dynamics at the protein surface around each NH site. A “bulk interaction rate” was determined from our MD data, which describes how fast solvent molecules can diffuse from the bulk into the direct vicinity of any given NH. A large interaction rate signifies that solvent molecules adjacent to an NH site are in rapid exchange with the bulk, implying the absence of protecting water networks. The unprotected sites in the C-terminal tail (residues 73-76) provide a reference context for this analysis. Bulk interaction rates for individual amides were generally found to be between zero to  $\sim 0.05 \text{ ps}^{-1}$  (Figure 2.5e). This range is consistent with earlier reports of surface solvent exchange on picosecond time scales.<sup>63</sup> The key question is the following: Do exposed NH sites with  $\log P \gg 0$  exhibit lower bulk interaction rates than NH sites with  $\log P \approx 0$ ?

It is most instructive to examine the bulk exchange characteristics of exposed NH sites by focusing on the two extreme cases of strongly protected vs. unprotected surface amides (marked with blue and red asterisks, Figure 2.5e). The average bulk exchange rate of the “blue” sites is  $0.020 \pm 0.01 \text{ ps}^{-1}$ , whereas the “red” sites have a bulk exchange rate of  $0.024 \pm 0.01 \text{ ps}^{-1}$ . The standard deviations of these average rates overlap with each other. While the “red” average is slightly higher, it would be far-fetched to propose that this small difference could be responsible for modulating HDX rates by more than two orders

of magnitude. We conclude that the solvent interactions for protected and unprotected exposed amides are not significantly different. In other words, our data do not support the idea that the HDX properties of exposed amides can be modulated by partially immobilized water molecules.

### 2.3.9 Force Fields and Solvent Models

It is unlikely that the lack of correlation between HDX protection and water immobilization in Figure 2.5e reflects inherent limitations of the MD strategy used here. The data of Figure 5b-e were obtained using the CHARMM22\* force field<sup>50</sup> with TIP3P water.<sup>51</sup> Earlier simulations using the same approach were found to be remarkably accurate in folding simulations on a range of proteins.<sup>66</sup> The TIP3P water model has recently been applied to address intricate solvation details at protein surfaces.<sup>65,67</sup> Nonetheless, it is clear that the results of MD simulations can be affected to some extent by the choice of force field.<sup>50,68</sup> To test the robustness of our findings we repeated the simulations using a different force field (Amber99sb-ILDN<sup>52</sup>) and a different water model (TIP4P<sup>51</sup>). The results obtained in this way are compiled in Figure 2.8. Comparison of these data with those of Figure 2.5 shows a very high degree of consistency. These additional data do not affect any of the aforementioned considerations.



**Figure 2.8 Overview of experimental HDX data, and various properties extracted from a 1  $\mu$ s Amber99sb-ILDN/TIP4P simulation trajectory.** (a) Experimental protection factors as in Figure 5. (b) Distance between backbone amide hydrogens and the closest possible main chain carbonyl acceptor. Values below 2.5 Å (red line) reflect the presence of an H-bond. (c) Same as in panel (b), but including side chains. (d) NH SASA values. (e) Bulk interaction rate, describing how fast bulk water diffuses to an NH site. Blue asterisks represent “problem cases”, where exposed amides are highly protected. Red asterisks represent exposed amides that are not protected.

## 2.4 Conclusions

It is sobering that after 50+ years of using HDX as a structural biology tool<sup>30</sup> practitioners have now concluded that it is necessary to revisit very simple proteins for conducting “*a detailed hydrogen by hydrogen analysis to examine the bases of structure-rate relationships*”.<sup>29</sup> The current work employs all-atom MD simulations in an effort to scrutinize some of the pertinent issues. Computational techniques of the type used here cannot model actual NH → ND conversion events; doing so would require *ab-initio* approaches that are not practical for large systems.<sup>69</sup> Nonetheless, even classical simulations can provide insights that go beyond those obtainable from static crystal structures.<sup>46,66</sup>

It remains undisputed that amides in rigid regions tend to be more protected than those in disordered segments. For ubiquitin, most of the amides in the structured sequence range (residues 1-72) are protected, while the disordered C-terminus (residues 73-76) is unprotected. However, the exact physicochemical features that provide amide NH protection are poorly defined. The results of the current analysis can be summarized as follows (Table 1): (i) H-bonding *always* leads to HDX protection. This includes H-bonds to backbone carbonyls, side chains, as well as bifurcated H-bonds. (ii) For NH sites that are not H-bonded, low SASA values are *often* (but not always) associated with HDX protection. (iii) A lack of H-bonding at solvent-accessible amides does *not* imply that the corresponding sites are unprotected. Instead, many of these amides are characterized by  $\log P \gg 0$ .<sup>29,32-34</sup> Aspect (iii) is the most troublesome finding of our work, as it goes against paradigms that are widely accepted in the HDX community. Most practitioners would concur that  $\log P \gg 0$  implies either H-bonding or NH burial (or both). Our work demonstrates that this view is incorrect for many surface amides. This issue is particularly worrisome for modeling initiatives that rely on HDX data or other “sparse” structural information for elucidating protein conformations.<sup>40,43,45,70</sup> The unexpected HDX behavior displayed by a considerable fraction of amides in a small model protein does not bode well for investigations on larger systems with unknown structures.



Efforts have been undertaken to account for unexpected surface amide protection on the basis of electrostatic factors.<sup>32,33,35</sup> As noted earlier, negative charges in the vicinity of an amide will slow down HDX by raising the NH pK<sub>a</sub>.<sup>18-20,71</sup> It remains unclear if such electrostatic effects can be responsible for the surface amide protection considered here. A recent study found no correlation between the HDX rates of surface amides and the calculated electrostatic field.<sup>29</sup> Also, our MD trajectories do not reveal enhanced charge density (from E<sup>-</sup> or D<sup>-</sup> side chains) in the vicinity of protected surface amides, with the exception of NH sites that engage in bifurcated H-bonds.<sup>60</sup> Electrostatic effects could nonetheless be a contributor to anomalous surface NH protection. Electrostatic modeling approaches require future refinement, as current results are strongly parameter-dependent.<sup>33</sup> Challenges include an adequate description of polarizabilities, dielectric properties, solvent contributions, salt-mediated screening,<sup>18,33</sup> as well as the choice of suitable reference structures.<sup>35</sup>

**Table 2.1 Summary of backbone NH protection behavior.**

NH Structural Context	HDX Protection Status
H-bonded to backbone	protected
H-bonded to side chain	protected
buried (not H-bonded)	often (not always) protected
located in <i>rigid</i> region, solvent-accessible, not H-bonded	often (not always) protected
located in <i>disordered</i> region, solvent-accessible, not H-bonded	unprotected
H-bonded to crystal water	no effect on protection

Our results do not support the proposal<sup>29,36</sup> that crystallographically defined water molecules can protect exposed amides via H-bonding contacts. The validity of this idea had previously been questioned by others.<sup>35</sup> Many of the surface amides in ubiquitin are indeed bound to crystal water, but these interactions are not correlated with the degree of NH protection. We further tested whether protection might arise from higher order water networks. Unfortunately, there is no evidence for differences in the solvation behavior of protected vs. unprotected exposed amides. It would be of interest to extend this analysis by including interaction rates of NH sites with OH<sup>-</sup> (or OD<sup>-</sup>) which acts as HDX catalyst. Regrettably, such calculations require *ab initio* strategies that are out of reach for systems of the size considered here.<sup>72,73</sup>

The interpretation of HDX kinetics in terms of log *P* values (eq. 3) relies on the adequacy of peptide-calibrated *k<sub>ch</sub>* data.<sup>21</sup> This approach has been criticized because peptides may not always properly mimic the environment experienced by protein NH groups (eq. 2).<sup>34</sup> Indeed, there are indications of discrepancies between protein and peptide-calibrated data in a few cases.<sup>74,75</sup> Some log *P* values can also be affected by measurement artifacts. These issues introduce uncertainties, especially for the “weakly protected” sites of Figure 5a. Data obtained for the “strongly protected” and “unprotected” amides are more robust, which is why most of our discussion focused on the latter two categories.

Overall, the results of this study emphasize that HDX data have to be interpreted with caution. Widely accepted tenets such as the putative correlation between HDX protection and amide H-bonding (and/or solvent exclusion) may have to be revised. At the current stage of development there is no consistent explanation for the fact that surface NH sites can be strongly protected, while structural data show them to reside in a seemingly “open” conformation. These inconsistencies show that the mechanism of protein HDX is far from being understood. Perhaps it is time to move beyond the simple Linderstrøm-Lang formalism (eq. 2),<sup>30</sup> which has governed the interpretation of HDX kinetics for decades. Quantum mechanical investigations that take into account the electronic properties of all interaction partners (protein, solvent, catalyst) as well as their conformational dynamics may be required to fully understand the intricacies associated with the seemingly trivial conversion of NH to ND.

## 2.5 References

- 1 Baldwin, R. L. Early days of protein hydrogen exchange: 1954-1972. *Proteins* **79**, 2021-2026 (2011).
- 2 Englander, S. W., Mayne, L. & Krishna, M. M. G. Protein folding and misfolding: mechanism and principles. *Quart. Rev. Biophys.* **40**, 287-326 (2007).
- 3 Fazelinia, H., Xu, M., Cheng, H. & Roder, H. Ultrafast Hydrogen Exchange Reveals Specific Structural Events during the Initial Stages of Folding of Cytochrome c. *J. Am. Chem. Soc.* **136**, 733-740, doi:10.1021/ja410437d (2014).
- 4 Ward, M. E. *et al.* Proton-detected solid-state NMR reveals intramembrane polar networks in a seven-helical transmembrane protein proteorhodopsin. *J. Am. Chem. Soc.* **133**, 17434-17443 (2011).
- 5 Balasubramaniam, D. & Komives, E. A. Hydrogen-exchange mass spectrometry for the study of intrinsic disorder in proteins. *Biochimica et Biophysica Acta* **1834**, 1202-1209 (2013).
- 6 Busenlehner, L. S., Salomonsson, L., Brzezinski, P. & Armstrong, R. N. Mapping protein dynamics in catalytic intermediates of the redox-driven proton pump cytochrome c oxidase. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15398-15403 (2006).
- 7 Keppel, T. R., Howard, B. A. & Weis, D. D. Mapping Unstructured Regions and Synergistic Folding in Intrinsically Disordered Proteins with Amide H/D Exchange Mass Spectrometry. *Biochemistry* **50**, 8722-8732, doi:10.1021/bi200875p (2011).
- 8 Konermann, L., Pan, J. & Liu, Y. Hydrogen Exchange Mass Spectrometry for Studying Protein Structure and Dynamics. *Chem. Soc. Rev.* **40**, 1224-1234 (2011).
- 9 Marciano, D. P., Dharmarajan, V. & Griffin, P. R. HDX-MS guided drug discovery: small molecules and biopharmaceuticals. *Curr. Op. Struct. Biol.*, 105-111 (2014).
- 10 Percy, A. J., Rey, M., Burns, K. M. & Schriemer, D. C. Probing protein interactions with hydrogen/deuterium exchange and mass spectrometry-A review. *Anal. Chim. Acta* **721**, 7-21 (2012).
- 11 Pirrone, G. F., Iacob, R. E. & Engen, J. R. Applications of Hydrogen/Deuterium Exchange MS from 2012 to 2014. *Anal. Chem.* **87**, 99-118, doi:10.1021/ac5040242 (2015).
- 12 Rand, K. D., Zehl, M., Jensen, O. N. & Jørgensen, T. J. D. Protein Hydrogen Exchange Measured at Single-Residue Resolution by Electron Transfer Dissociation Mass Spectrometry. *Anal. Chem.* **81**, 5577-5584 (2009).
- 13 Rob, T. *et al.* Measuring Dynamics in Weakly Structured Regions of Proteins Using Microfluidics-Enabled Subsecond H/D Exchange Mass Spectrometry. *Anal. Chem.* **84**, 3771-3779 (2012).

- 14 Wang, G., Abzalimov, R. R., Bobst, C. E. & Kaltashov, I. A. Conformer-specific characterization of nonnative protein states using hydrogen exchange and top-down mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20087-20092 (2013).
- 15 Craig, P. O. *et al.* Prediction of native-state hydrogen exchange from perfectly funneled energy landscapes. *J. Am. Chem. Soc.* **133**, 17463-17472 (2011).
- 16 Pan, Y. & Briggs, M. S. Hydrogen Exchange in Native and Alcohol Forms of Ubiquitin. *Biochemistry* **31**, 11405-11412 (1992).
- 17 Perrin, C. L. Proton Exchange in Amides: Surprises from Simple Systems. *Accounts Chem. Res.* **22**, 268-275, doi:10.1021/ar00164a002 (1989).
- 18 Abdolvahabi, A., Gober, J. L., Mowery, R. A., Shi, Y. H. & Shaw, B. F. Metal-Ion-Specific Screening of Charge Effects in Protein Amide H/D Exchange and the Hofmeister Series. *Anal. Chem.* **86**, 10303-10310, doi:10.1021/ac502714v (2014).
- 19 Fogolari, F., Esposito, G., Viglino, P., Briggs, J. M. & McCammon, J. A. pK(a) shift effects on backbone amide base-catalyzed hydrogen exchange rates in peptides. *J. Am. Chem. Soc.* **120**, 3735-3738, doi:10.1021/ja963133m (1998).
- 20 Molday, R. S., Englander, S. W. & Kallen, R. G. Primary Structure Effects on Peptide Hydrogen Exchange. *Biochemistry* **11**, 150-158 (1972).
- 21 Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. Primary Structure Effects on Peptide Group Hydrogen Exchange. *Proteins: Struct., Funct., Genet.* **17**, 75-86 (1993).
- 22 Bougault, C., Feng, L. M., Glushka, J., Kupce, E. & Prestegard, J. H. Quantitation of rapid proton-deuteron amide exchange using hadamard spectroscopy. *J. Biomol. NMR* **28**, 385-390, doi:10.1023/b:jnmr.0000015406.66725.30 (2004).
- 23 Johnson, E. C., Lazar, G. A., Desjarlais, J. R. & Handel, T. M. Solution structure and dynamics of a designed hydrophobic core variant of ubiquitin. *Struct. Fold. Des.* **7**, 967-976, doi:10.1016/s0969-2126(99)80123-3 (1999).
- 24 Fajer, P. G., Bou-Assaf, G. M. & Marshall, A. G. Improved Sequence Resolution by Global Analysis of Overlapped Peptides in Hydrogen/Deuterium Exchange Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **23**, 1202-1208 (2012).
- 25 Hughson, F. M., Wright, P. E. & Baldwin, R. L. Structural Characterisation of a Partly Folded Apomyoglobin Intermediate. *Science* **249**, 1544-1548 (1990).
- 26 Shan, Y., Arkhipov, A., Kim, E. T., Pan, A. C. & Shaw, D. E. Transitions to catalytically inactive conformations in EGFR kinase. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7270-7275, doi:10.1073/pnas.1220843110 (2013).
- 27 Sperry, J. B., Smith, C. L., Caparon, M. G., Ellenberger, T. & Gross, M. L. Mapping the Protein-Protein Interface between a Toxin and Its Cognate Antitoxin from the Bacterial Pathogen *Streptococcus pyogenes*. *Biochemistry* **50**, 4038-4045 (2011).

- 28 Yan, X. *et al.* Deuterium exchange and mass spectrometry reveal the interaction differences of two synthetic modulators of RXR $\alpha$ LBD. *Protein Sci.* **16**, 2491-2501 (2007).
- 29 Skinner, J. J., Lim, W. K., Bedard, S., Black, B. E. & Englander, S. W. Protein hydrogen exchange: Testing current models. *Protein Sci.* **21**, 987-995, doi:10.1002/pro.2082 (2012).
- 30 Hvidt, A. & Nielsen, S. O. Hydrogen exchange in proteins. *Adv. Protein Chem.* **21**, 287-386 (1966).
- 31 Sowole, M. A., Alexopoulos, J. A., Cheng, Y.-Q., Ortega, J. & Konermann, L. Activation of ClpP Protease by ADEP Antibiotics: Insights from Hydrogen Exchange Mass Spectrometry. *J. Mol. Biol.* **425**, 4508-4519, doi:http://dx.doi.org/10.1016/j.jmb.2013.08.005 (2013).
- 32 Anderson, J. S., Hernandez, G. & LeMaster, D. M. A billion-fold range in acidity for the solvent-exposed amides of Pyrococcus furiosus rubredoxin. *Biochemistry* **47**, 6178-6188, doi:10.1021/bi800284y (2008).
- 33 Hernandez, G., Anderson, J. S. & LeMaster, D. M. Polarization and Polarizability Assessed by Protein Amide Acidity. *Biochemistry* **48**, 6482-6494, doi:10.1021/bi900526z (2009).
- 34 Li, R. & Woodward, C. The hydrogen exchange core and protein folding. *Protein Sci.* **8**, 1571-1590 (1999).
- 35 Anderson, J. S., Hernandez, G. & LeMaster, D. M. Assessing the chemical accuracy of protein structures via peptide acidity. *Biophys. Chem.* **171**, 63-75, doi:10.1016/j.bpc.2012.10.005 (2013).
- 36 Skinner, J. J., Lim, W. K., Bédard, S., Black, B. E. & Englander, S. W. Protein dynamics viewed by hydrogen exchange. *Protein Sci.* **21**, 996-1005 (2012).
- 37 Alexopoulos, J. A., Guarnéa, A. & Ortega, J. ClpP: A structurally dynamic protease regulated by AAA+ proteins. *J. Struct. Biol.* **179**, 202-210 (2012).
- 38 Bahar, I., Wallqvist, A., Covell, D. G. & Jernigan, R. L. Correlation between Native-State Hydrogen Exchange and Cooperative Residue Fluctuations from a Simple Model. *Biochemistry* **37**, 1067-1075 (1998).
- 39 Garcia, A. E. & Hummer, G. Conformational dynamics of cytochrome c: Correlation to hydrogen exchange. *Proteins* **36**, 175-191, doi:10.1002/(sici)1097-0134(19990801)36:2<175::aid-prot4>3.0.co;2-r (1999).
- 40 Gsponer, J. *et al.* Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 99-104, doi:10.1073/pnas.0508667102 (2006).
- 41 Hsu, Y. H. *et al.* Fluoroketone Inhibition of Ca<sup>2+</sup>-Independent Phospholipase A(2) through Binding Pocket Association Defined by Hydrogen/Deuterium Exchange and Molecular Dynamics. *J. Am. Chem. Soc.* **135**, 1330-1337, doi:10.1021/ja306490g (2013).

- 42 Petruk, A. A. *et al.* Molecular Dynamics Simulations Provide Atomistic Insight into Hydrogen Exchange Mass Spectrometry Experiments. *J. Chem. Theory Comput.* **9**, 658-669, doi:10.1021/ct300519v (2013).
- 43 Sheinerman, F. B. & Brooks III, C. L. Molecular picture of folding of a small  $\alpha/\beta$  protein. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1562-1567 (1998).
- 44 Xu, J., Lee, Y., J., Beamer, L. J. & Van Doren, S. R. Phosphorylation in the Catalytic Cleft Stabilizes and Attracts Domains of a Phosphohexomutase. *Biophys. J.* **108**, 235-337 (2015).
- 45 Liu, T. *et al.* Quantitative assessment of protein structural models by comparison of H/D exchange MS data with exchange behavior accurately predicted by DXCOREX. *J. Am. Soc. Mass Spectrom.* **23**, 43-56 (2012).
- 46 Piana, S., Lindorff-Larsen, K. & Shaw, D. E. Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5915-5920, doi:10.1073/pnas.1218321110 (2013).
- 47 Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. Structure of Ubiquitin Refined at 1.8 Å Resolution. *J. Mol. Biol.* **194**, 531-544 (1987).
- 48 Berendsen, H. J., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **91**, 43-56 (1995).
- 49 Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **4**, 435-447 (2008).
- 50 Piana, S., Lindorff-Larsen, K. & Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **100**, L47-L49, doi:10.1016/j.bpj.2011.03.051 (2011).
- 51 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935 (1983).
- 52 Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-1958, doi:10.1002/prot.22711 (2010).
- 53 Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of chemical physics* **72**, 2384-2393 (1980).
- 54 Goyal, P. *et al.* Molecular Simulation of Water and Hydration Effects in Different Environments: Challenges and Developments for DFTB Based Models. *J. Phys. Chem. B* **118**, 11007-11027, doi:10.1021/jp503372v (2014).
- 55 Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 0141011-0141017, doi:10.1063/1.2408420 (2007).

- 56 Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684-3690, doi:10.1063/1.448118 (1984).
- 57 Hess, B., Henk, B., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463-1472 (1997).
- 58 Miyamoto, S. & Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **13**, 952-962, doi:10.1002/jcc.540130805 (1992).
- 59 Essmann, U. *et al.* A SMOOTH PARTICLE MESH EWALD METHOD. *J. Chem. Phys.* **103**, 8577-8593, doi:10.1063/1.470117 (1995).
- 60 Torshin, I. Y., Weber, I. T. & Harrison, R. W. Geometric criteria of hydrogen bonds in proteins and identification of bifurcated hydrogen bonds. *Protein engineering* **15**, 359-363 (2002).
- 61 Eisenhaber, F., Lijnzaad, P., Argos, P., Sander, C. & Scharf, M. The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *Journal of Computational Chemistry* **16**, 273-284 (1995).
- 62 Li, A. J. & Nussinov, R. A set of van der Waals and coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. *Proteins: Structure, Function, and Bioinformatics* **32**, 111-127 (1998).
- 63 Krishnamoorthy, A. N., Holm, C. & Smiatek, J. Local Water Dynamics around Antifreeze Protein Residues in the Presence of Osmolytes: The Importance of Hydroxyl and Disaccharide Groups. *J. Phys. Chem. B* **118**, 11613-11621, doi:10.1021/jp507062r (2014).
- 64 Mallajosyula, S. S., Vanommeslaeghe, K. & MacKerell, A. D. Perturbation of Long-Range Water Dynamics as the Mechanism for the Antifreeze Activity of Antifreeze Glycoprotein. *J. Phys. Chem. B* **118**, 11696-11706, doi:10.1021/jp508128d (2014).
- 65 Nibali, V. C. & Havenith, M. New Insights into the Role of Water in Biological Function: Studying Solvated Biomolecules Using Terahertz Absorption Spectroscopy in Conjunction with Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **136**, 12800-12807, doi:10.1021/ja504441h (2014).
- 66 Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **334**, 517-520 (2011).
- 67 Martin, D. R. & Matyushov, D. V. Dipolar Nanodomains in Protein Hydration Shells. *The Journal of Physical Chemistry Letters* **6**, 407-412 (2015).
- 68 Cino, E. A., Choy, W. Y. & Karttunen, M. Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations. *J. Chem. Theo. Comp.* **8**, 2725-2740, doi:10.1021/ct300323g (2012).

- 69 Ijaz, W., Gregg, Z. & Barnes, G. L. Complex Formation during SID and Its Effect on Proton Mobility. *The Journal of Physical Chemistry Letters* **4**, 3935-3939 (2013).
- 70 Moulton, J., Fidelis, K., Kryshtafovych, A., Schwede, T. & Tramontano, A. Critical assessment of methods of protein structure prediction (CASP) - round x. *Proteins-Structure Function and Bioinformatics* **82**, 1-6 (2014).
- 71 Shaw, B. F. *et al.* Neutralizing Positive Charges at the Surface of a Protein Lowers Its Rate of Amide Hydrogen Exchange without Altering Its Structure or Increasing Its Thermostability. *Journal of the American Chemical Society* **132**, 17411-17425, doi:10.1021/ja9067035 (2010).
- 72 Crespo, Y. & Hassanali, A. Unveiling the Janus-Like Properties of OH<sup>-</sup>. *The Journal of Physical Chemistry Letters* **6**, 272-278 (2015).
- 73 Tuckerman, M. E., Marx, D. & Parrinello, M. The nature and transport mechanism of hydrated hydroxide ions in aqueous solution. *Nature* **417**, 925-929 (2002).
- 74 Rand, K. D. & Jørgensen, T. J. D. Development of a Peptide Probe for the Occurrence of Hydrogen (<sup>1</sup>H/<sup>2</sup>H) Scrambling upon Gas-Phase Fragmentation. *Anal. Chem.* **79**, 8686-8693 (2007).
- 75 Del Mar, C., Greenbaum, E. A., Mayne, L., Englander, S. W. & Woods, V. L. Structure and properties of  $\alpha$ -synuclein and other amyloids determined at the amino acid level. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15477-15482 (2005).



## 3 The Electrospray Mechanism of Natively Folded Proteins: a Molecular Dynamics Investigation

### 3.1 Introduction

Electrospray ionization (ESI) is a soft ionization technique capable of producing multiply charged gas phase ions from macromolecular analytes in solution, including proteins.<sup>1-3</sup> ESI coupled to mass spectrometry (MS) can overcome  $m/z$  limitations of some types of mass analyzers by generating high charge states, and it supports facile coupling to upstream liquid chromatography, leading to its wide use as an analytical technique.<sup>3</sup> The ESI process is initiated by infusing analyte-containing solution through a capillary to which a high electric potential has been applied. Electrophoretic separation of charge and electrochemical reactions at the capillary electrode lead to buildup of excess charge at the tip and formation of a Taylor cone.<sup>5,6</sup> Charged droplets are emitted from this Taylor cone into a heated desolvation gas<sup>7,8</sup> where evaporation of solvent promotes charge buildup in droplets and subsequent jet fission.<sup>9</sup> Progeny droplets undergo repeated cycles of evaporation and fission, giving rise to a final generation of nanometer-sized droplets that produce the gaseous ions detected by MS.<sup>5,10</sup>

Coulombic fission in micrometer sized droplets has been directly observed<sup>9</sup>, and is expected to occur when charge repulsion on the droplet surface exceeds surface tension, as described by the Rayleigh limit<sup>11,12</sup>:

$$z_R = \frac{8\pi}{e} \sqrt{\varepsilon_0 \gamma r^3} \quad (3.1)$$

where  $z_R$  is the number of elementary charges,  $e$ ,  $\varepsilon_0$  is the vacuum permittivity,  $\gamma$  is the surface tension, and  $r$  is the droplet radius. Daughter droplets carry away only a small portion of the parent droplet's volume, but a disproportionate share of its excess charge.<sup>5,9,13</sup>

The mechanism of ion release from the final generation of nanodroplets remains a controversial topic<sup>5,14-16</sup>. Many groups have postulated that ions – particularly small ions such as  $\text{Na}^+$  and  $\text{NH}_4^+$  – are emitted from these droplets via an activated field emission

process, the ion evaporation model (IEM).<sup>16-19</sup> Conversely, it is envisioned that droplet evaporation to dryness leaves behind larger charged analytes such as ion clusters and macromolecules, the charged residue model (CRM).<sup>5,20-22</sup> Others have proposed intermediate mechanisms that include features of both the IEM and CRM.<sup>23,24</sup> Recent studies have also suggested that a third mechanism may apply to extended polymers such as polyethylene glycols, whereby the polymer is extruded from the droplet concomitant with charging, the chain ejection model (CEM).<sup>14,25</sup>

In protein ESI-MS, ion charge state is related to the conformation of ions in solution, with denaturing solutions giving rise to higher charge states.<sup>1,26</sup> Additionally, low charge species tend to be more heavily adducted when sprayed from salt-containing solution.<sup>27</sup> This suggests that unfolded protein ions are ejected from droplets via the CEM, while folded proteins emerge following solvent evaporation via the CRM.<sup>15</sup> Indeed, folded globular proteins with molecular weights ranging over several orders of magnitude produce ions with charges comparable to Rayleigh-charged droplets of the same mass, supporting the notion that these species are CRM products.<sup>20</sup> However, the droplets that produce these ions are small and short-lived, making them difficult to observe experimentally.

Molecular dynamics (MD) simulations can provide a detailed atomistic view of sub-microscopic processes, and are becoming an increasingly useful method to probe the mechanisms underlying ESI.<sup>14,15,28-33</sup> Studies in this direction have generally supported the idea that small carriers including  $\text{Na}^+$  and  $\text{NH}_4^+$  are ejected from the droplet via the IEM.<sup>15,28,30,33</sup> Simulations of disordered polymers from our group<sup>25</sup> and Consta<sup>14</sup>, have also supported the notion that the CEM is the dominant ionization mechanism for these species. However, simulations modelling the CRM have been limited. To our knowledge, only two previous studies have simulated a CRM process to completion – that is, the total desolvation of the analyte – and neither of these included macromolecules.<sup>22,29</sup> The issue in simulating this process is threefold. Firstly, the CRM occurs over a much longer timescale than the  $< 2$  ns simulations required to observe the IEM or CEM.<sup>14,15,22,25,28,30</sup> Second, and further compounding this issue, is that mechanistic MD studies of the ESI process on proteins necessitates very large droplets, which are computationally expensive

to simulate over long timescales. Finally, evaporative cooling of the simulated droplets results in drastic decreases in temperature which tends to halt solvent evaporation, even when a thermostat is applied.<sup>22,25,31,32</sup> As a result, MD simulations of proteins solvated in ESI droplets which result in solvent-free gaseous ions have not yet been reported.

In this work, we apply a trajectory stitching approach, which we have previously used in MD simulations of salt-containing nanodroplets<sup>22,34</sup>, to systems containing folded protein molecules and excess positive charge. We conduct 125 – 175 ns temperature-stabilized simulations on droplets of radius 3 nm which result in the formation of fully desolvated proteins in the gas phase. As solvent evaporates, droplets shed charge via emission of solvated single ions, consistent with the IEM. However, the protein remains solvated in the shrinking droplet until all solvent molecules have evaporated, consistent with the CRM. We find that these simulations yield clusters with charges comparable to the Rayleigh charge of a protein-sized water droplet across a range of analyte sizes, consistent with predictions of the CRM and folded species observed in ESI-MS. Further, through investigation of protein ions simulated with initial immobilized charges, we find that the remaining charge following solvent evaporation depends primarily on protein conformation, rather than initial charge in solution phase, in agreement with experimental data.

## 3.2 Materials and Methods

### 3.2.1 Protein Solutions

Bovine ubiquitin (Ubq), equine holo-myoglobin (hMb), and equine heart cytochrome *c* (Cyt) were purchased from Sigma (St. Louis, MO). pH 7 solutions were prepared with a protein concentration of 5  $\mu\text{M}$  in 10 mM aqueous ammonium acetate. Acidic and basic solutions of ubiquitin were prepared at 5  $\mu\text{M}$  protein concentration without buffer in 70 mM aqueous ammonium hydroxide (pH 11.2) and 2.65 mM aqueous formic acid (pH 3), respectively. Salt-containing samples additionally contained 200  $\mu\text{M}$  sodium acetate. The protein mixture used to calibrate the ion mobility spectrometry (IMS) cell consisted of 10  $\mu\text{M}$  each of Ubq, hMb, and Cyt in 49:49:2 (v/v/v) methanol:water:acetic acid (pH 2.2).

### 3.2.2 Mass Spectrometry and Ion Mobility Spectrometry

Mass spectra were gathered on a Synapt G1 HDMS instrument (Waters, Milford, MA) equipped with a Z-spray ESI source. Ion mobility data were gathered concurrently with MS data using the instrument's travelling-wave IMS cell. Samples were infused by a syringe pump at a flow rate of 5  $\mu\text{L}/\text{min}$ . ESI was carried out with capillary voltage of 2.8 kV and  $\text{N}_2$  desolvation gas at a rate of 500 L/h. Other instrument voltages were tuned to maximize the relative intensity of the most compact conformer observed in IMS for the 6+, 8+, and 9+ charge states of Ubq, Cyt, and hMb, respectively. Drift times observed in the IMS cell were converted to collision cross section ( $\Omega$ ) data using a calibration procedure similar to that published by Robinson and co-workers<sup>35</sup>, except that calibrant ions were collisionally activated using high sample cone voltage, and linear regression was against literature  $\Omega$  values of unfolded species.

### 3.2.3 MD Simulations: System Construction

All MD simulations were carried out using GROMACS version 4.6 or 5.0<sup>36</sup> on Linux workstations. Ubq, Cyt, and hMb were modelled based on the crystal structures with PDB accession codes 1UBQ<sup>37</sup>, 1HRC<sup>38</sup>, and 1WLA<sup>39</sup>, respectively. Solvent molecules were removed and hydrogens were added using the PDB2GMX routine of GROMACS. A slightly modified version of the CHARMM36 force field<sup>40</sup> which includes proper parameters for heme *c*<sup>41</sup> was used for all three proteins. An in-house script was used to build the Cyt topology which allowed for axial ligation of the heme iron by His18 and Met80 and correct geometry at the Cys14/heme and Cys17/heme thioether linkages. All titratable sites were set to their pH 7 protonation states (N-terminus<sup>+</sup>, Lys<sup>+</sup>, Arg<sup>+</sup>, His<sup>0</sup>, Cys<sup>0</sup>, Tyr<sup>0</sup>, C-terminus<sup>-</sup>, Asp<sup>-</sup>, Glu<sup>-</sup>), unless otherwise noted.

In order to build droplets, proteins were centered inside a rhombic dodecahedral box and subsequently solvated with water in a pre-equilibrated geometry using the GROMACS solvate utility. Approximately spherical droplets were formed from this dodecahedral system by applying an in-house Perl program called "*dropletCarver*" which removes all molecules in the system that are farther from the protein center of mass than a user-specified distance. For droplet construction, dropletCarver was set to create

spherical clusters with a radius of 3 nm. In each droplet, random solvent molecules were replaced with Na<sup>+</sup> ions to bring the total system charge to 16+. These sodium ions are the mobile charge carriers in our system since simulation of protons, which requires *ab initio* methods<sup>42</sup>, is not feasible for droplets of this size. The droplets were then centered in a 1 μm periodic cube for MD simulation, in order to take advantage of GROMACS GPU acceleration.

### 3.2.4 MD Simulations: General Aspects

Droplets were first subjected to up to 50000 iterations of steepest descent energy minimization such that the maximum force in the system was less than 250 kJ/mol/nm. Following energy minimization, the system was equilibrated for 2500 steps under constant temperature using the Nose Hoover thermostat<sup>43,44</sup>, with initial particle velocities sampled from a Maxwell-Boltzmann distribution. Production simulations were also carried out with initial thermalization from a Maxwell-Boltzmann distribution and maintained at constant temperature using a Nose Hoover thermostat. The leapfrog integrator was used for both equilibration and production phases of the simulations. Van der Waals forces were modelled using a Lennard-Jones potential, while electrostatic interactions were computed directly with a Coulomb potential. All non-bonded interactions used a potential-shifted cutoff at 333.3 nm. Neighbour lists were generated every 50 steps using the Verlet buffer method<sup>45</sup>. Linear momentum was removed from the system at every timestep. Constraints were applied to all bonds, using the LINCS algorithm<sup>46</sup> for proteins and the SETTLE algorithm<sup>47</sup> for water, to enable an integration timestep of 2 fs.

Different water models vary considerably with respect to the set of bulk properties which they can reproduce. We therefore performed simulations with both the 3-site TIP3P<sup>48</sup> and 4-site TIP4P/2005<sup>49</sup> water models, in order to examine the effects of water model choice on our simulations.

Droplets were coupled to a thermostat in order to simulate the collisional heating that occurs in the ESI source<sup>8</sup>, with activation depending on instrumental parameters. However, studies also show that ions may experience temperatures up to 450-800 K as

they are transported through the sampling region and ion guides of a mass spectrometer.<sup>50,51</sup> This is examined further in the simulations of Ubq by simulating droplets at either initial “lower activation” conditions of 330 K for 125 ns or “higher activation” conditions at 370 K for 75 ns. Droplets containing Cyt or hMb were only simulated under the higher activation conditions. All droplets were then subjected to a temperature of 450 K for 50 ns.

Finally, we investigated the effect of solution charge state of the protein on final system charge state by protonating or deprotonating Ubq side chains to achieve initial protein charge states of 0 (pH 7 default), 8-, 4-, 4+, and 8+. For positively charged Ubq, Glu residues were protonated to achieve the desired charge. Conversely, Lys residues were deprotonated to achieve negatively charged Ubq. Cyt and hMb were simulated only at their pH 7 protonation states, with initial protein charges of 6+ and 2-, respectively. Table 1 summarizes all simulation conditions discussed in this study, with each condition consisting of 5 replicates.

**Table 3.1 Summary of simulation conditions discussed in this chapter.**

Condition	Protein	Initial Protein Charge	Water Model	Na+	Activation Profile
1	Ubq	0	TIP4P/2005	16	Higher
2	Ubq	0	TIP3P	16	Higher
3	Ubq	0	TIP4P/2005	16	Lower
4	Ubq	8-	TIP4P/2005	24	Higher
5	Ubq	4-	TIP4P/2005	20	Higher
6	Ubq	4+	TIP4P/2005	12	Higher
7	Ubq	8+	TIP4P/2005	8	Higher
8	Cyt	6+	TIP4P/2005	10	Higher
9	hMb	2-	TIP4P/2005	18	Higher

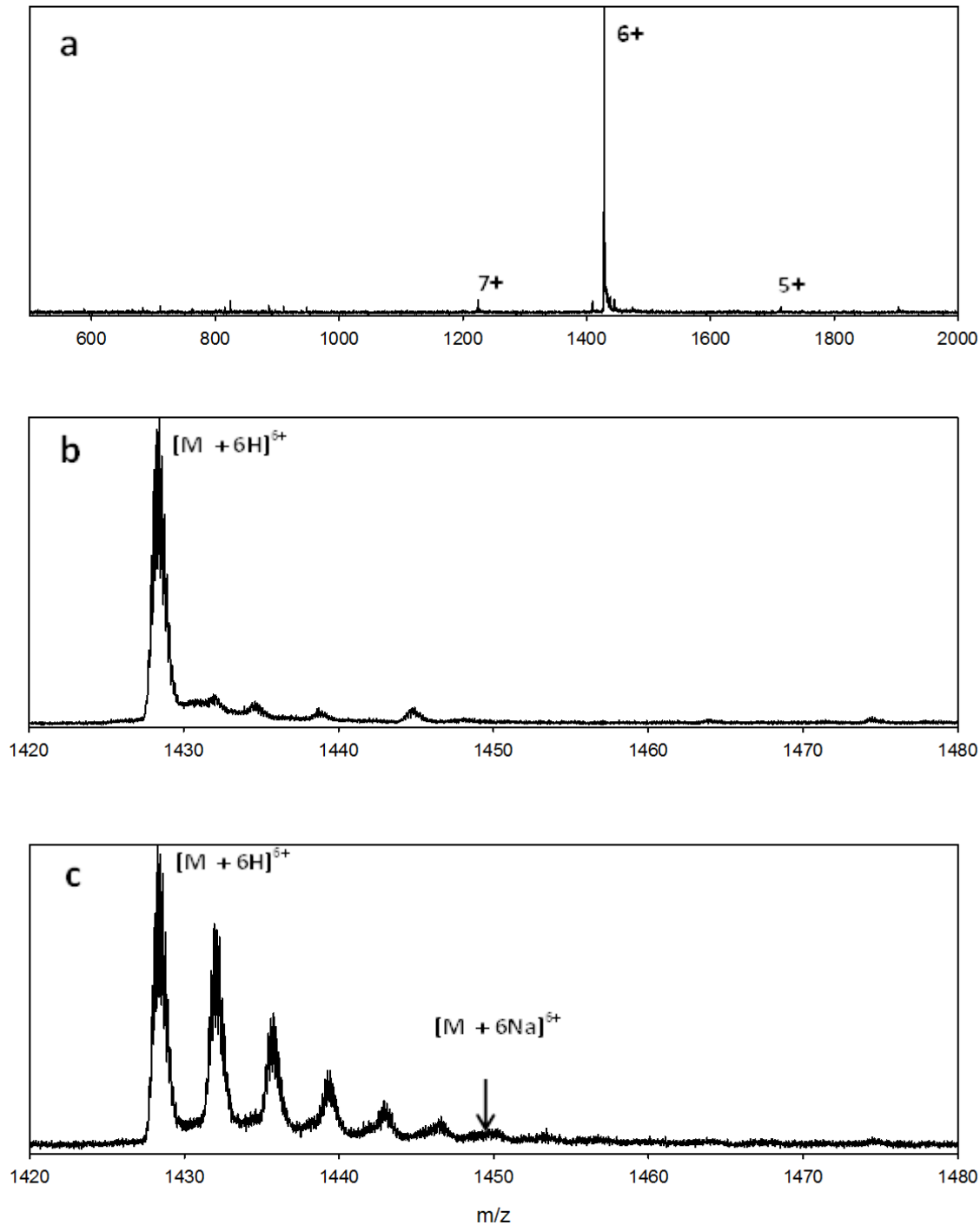
### 3.2.5 MD Trajectory Stitching

Simulating droplets in a vacuum environment under constant energy conditions has been previously shown to result in evaporative cooling as energetic solvent molecules escape the droplet, even when thermostats are used.<sup>22,28,32</sup> This is inconsistent with the collisional heating that the droplet undergoes in the source region, which eventually results in desolvation of ions.<sup>8</sup> In order to counteract this cooling, we applied a trajectory stitching approach, as discussed in previous studies from our lab.<sup>22,34</sup> Under this scheme, production simulations were not continuous, but broken up into discrete 500 ps simulation windows. At the end of each simulation window, the dropletCarver script was applied such that any molecule that had moved more than 15 nm away from the protein center of mass was removed from the system. The droplet was then re-centered in the simulation cell, and new velocities were sampled from a Maxwell-Boltzmann distribution prior to beginning the next simulation window. This process was iterated until the desired total run length was reached. We have previously shown that this approach maintains droplet temperatures over long trajectories.<sup>22</sup> The trajectory stitching approach used has the additional benefit of reducing the number of particles in the system over the course of the simulation, resulting in considerable speed-up versus simulating all particles for the full trajectory length.

## 3.3 Results and Discussion

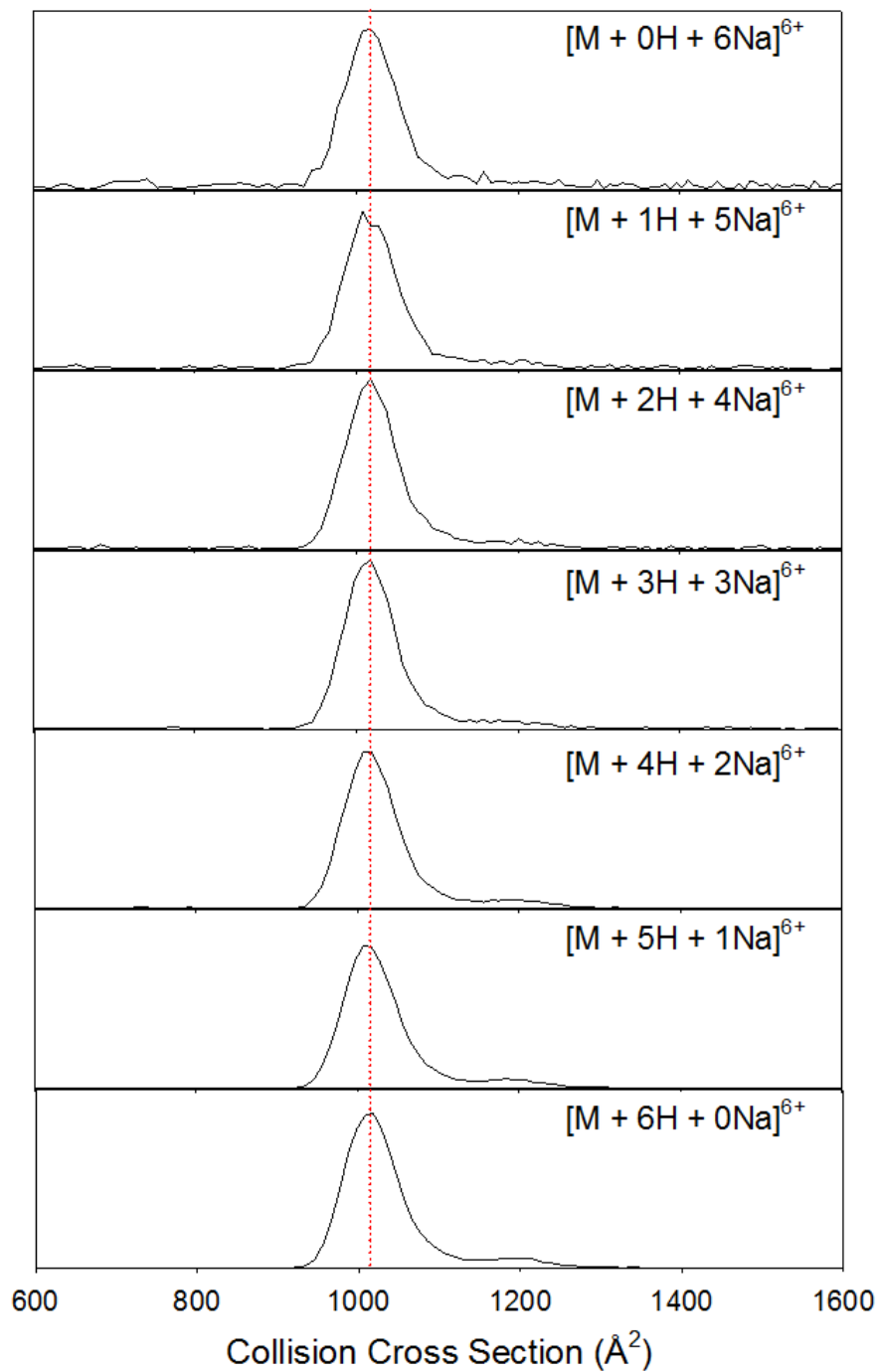
### 3.3.1 Folded Species in ESI-MS

We infused a 10  $\mu\text{M}$  solution of Ubq in 10 mM ammonium acetate (pH 7) into a positive mode ESI source, and applied gentle instrument parameters. Consistent with previous reports<sup>1,25,52</sup>, these conditions produce a fully desolvated protein charge state distribution centered at a low charge (Figure 3.1a). The most prominent peak in the spectrum is the  $[\text{M} + 6\text{H}]^{6+}$  species, while the 5+ and 7+ charge states are also observed at lower intensity (Figure 3.1a).



**Figure 3.1** ESI mass spectra of ubiquitin in 10 mM ammonium acetate under gentle conditions. (a) Full spectrum of Ubq in 10 mM ammonium acetate, showing a low charge state distribution. (b) Close up of the 6+ peak in (a). (c) close up of the 6+ peak when 2 mM sodium acetate is added to the sprayed solution, showing extensive sodium adduction.





**Figure 3.2 ion mobility traces of adducted species in Figure 3.1c.** Species display decreasing levels of sodium adduction from top to bottom, with the identity of each species identified at the right each panel. The dashed red vertical line indicates the collision cross section of the unadducted 6+ species from Figure 3.1b (1018 Å<sup>2</sup>).

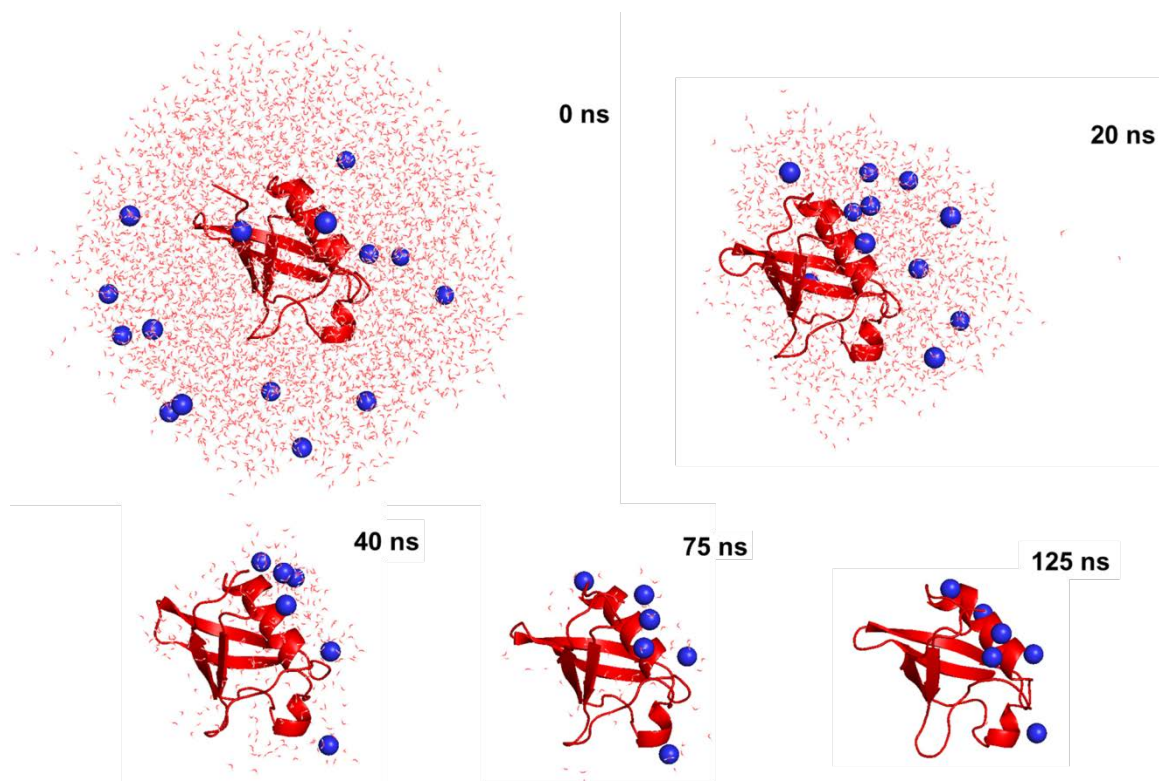
The full mass spectrum was also subjected to ion mobility spectrometry analysis using our instrument's in-line travelling wave ion mobility cell. The recorded drift times were converted to collision cross sections using a protocol similar to that of Robinson and co-workers.<sup>35</sup> The 6+ protonated ubiquitin species shows a collision cross section centered at 1018 Å<sup>2</sup> (Figure 3.2), which is in agreement with values previously measured by the Bowers group in a drift tube IMS device<sup>53</sup>, and theoretical values calculated from ubiquitin's X-ray crystal structure using the exact hard spheres scattering algorithm.<sup>54</sup> These data suggest that the 6+ species is a fully desolvated ubiquitin ion in a native-like conformation.

It is worth noting that these species are formed from a pH 7 solution in which Ubq molecules are expected to predominantly have a neutral charge (isoelectric point 6.56). Charging of the ion must therefore occur during the ESI process, as a result of protonation in the source or during desolvation. This is problematic for MD, since simulations including protons require use of much more computationally expensive *ab initio* methods.<sup>42,55</sup> We circumvent this issue by instead modelling mobile charges in our system using Na<sup>+</sup> ions, which can be treated classically.

We sought to examine the structural properties of multiply-sodiated Ubq ions by infusing a protein solution inoculated with 200 μM sodium acetate. This solution produces a charge state distribution comparable to that shown in Figure 3.1a for the sodium-free sample (data not shown), but the intensity of each protein peak is split between several sodium-adducted species (Figure 3.1c), an effect we have previously investigated with other salt additives.<sup>34</sup> Using these sodium acetate solutions, we are able to produce a ubiquitin ion whose excess charge is entirely carried by sodium adducts, labelled in Figure 3.1c. We also subjected these adducted species to IMS analysis. Multiply-sodiated species produced under these conditions display collision cross sections that are indistinguishable from the unadducted 6+ ion (Figure 3.2). They are, therefore, inferred to have a native-like structure comparable to the unadducted ion.<sup>53</sup> Based on these data, we suspect that formation of both protonated and sodiated Ubq ions proceeds via a similar mechanism. MD simulations using Na<sup>+</sup> ions as the excess charge carrier should therefore provide a reasonable model of the ESI process for folded proteins.

### 3.3.2 Temporal Evolution of Protein Nanodroplets

We first performed MD simulations on a Ubq-containing nanodroplet with excess charge carried by 16 Na<sup>+</sup> ions. Snapshots of a typical simulation are presented in Figure 3.3. At the beginning of the simulation, the droplet is roughly spherical, and Ubq is located at the center (Figure 3.3, 0 ns). As the simulation progresses, the protein migrates to the droplet's surface (Figure 3.3, 20 ns). We have previously seen this behavior with compact hydrophobic polymers<sup>25</sup>, but in this case, the effect seems to be Coulombically driven. Ubq's positively-charged basic side chains are excluded from the interior, remaining solvated close to the droplet surface. Gratifyingly, this agrees with Rayleigh's continuum model which envisions charge evenly distributed over the droplet surface<sup>12</sup>, as well as previous simulations in which the charge of small solvated cations was projected to the surface by orientation of solvent molecules.<sup>56</sup> The roughness of the droplet surface as solvent evaporates, particularly as charge density increases (Figure 3.3, 20 ns), has been previously observed in MD simulations, and may be analogous to Taylor cones that

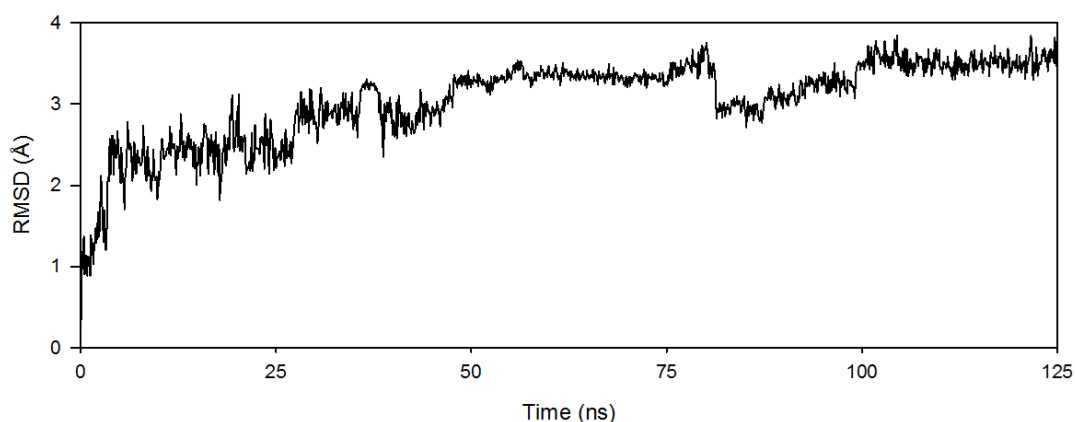


**Figure 3.3 Snapshots from a 125 ns simulation of a ubiquitin-containing nanodroplet.** Protein is shown in cartoon representation, in red. Sodium ions are in space-fill representation in blue. Water is displayed in line representation in red.

form in charged micrometer droplets, promoting field emission of solvated ions via the IEM.<sup>14</sup> Indeed, we observe many  $\text{Na}^+$  ions ejected in small solvent clusters to reduce system charge. These ejection events consist only of single ions, as we have previously seen in droplets containing salt species.<sup>22,34</sup>

Evaporation slows as the number of solvent molecules dwindles, and charged species “cling” to the last few waters in the system (Figure 3.3, 75 ns). These last solvent molecules are removed when the simulation temperature is increased to 450 K for 50 ns, modelling the higher collisional activation ions encounter in the sampling region and ion optics of a mass spectrometer.<sup>50,51</sup> These elevated temperatures do not result in loss of sodium adducts. The simulations culminate in desolvated protein ions clustered with 6 to 7 sodium ions (Figure 3.3, 125 ns), and an average charge of  $6.2 \pm 0.4$ . The adducted sodium ions tend to reside in the vicinity of the  $\alpha$ -helix, since Ubq has many acidic residues in this region which facilitates ionic cluster formation with  $\text{Na}^+$ . The charge states obtained are in remarkably good agreement with the mass spectrum displayed in Figure 3.1a, where the 6+ charge state of Ubq dominates and the 7+ species is the highest intensity minor constituent of the Ubq distribution.

An interesting result of these simulations is that the secondary and tertiary structure of ubiquitin appear to be unperturbed by desolvation and heating to at least 450 K in the gas phase, at least on the timescales studied here (Figure 3.3, 0 ns and 125 ns). Root mean square deviation (RMSD) analysis comparing the position of  $\alpha$  carbons in the protein



**Figure 3.4 Root mean square deviation of ubiquitin vs the crystal structure in an evaporating nanodroplet.** The RMSD values seen here are comparable to ubiquitin in bulk solvent, and an order of magnitude lower than denatured ubiquitin.<sup>4</sup>

backbone to the published crystal structure<sup>37</sup> confirm that solution phase structure is maintained throughout our simulations (Figure 3.4). This observation agrees with a previous ESI-MS study which suggests that Ubq may be kinetically trapped in a near-native configuration on at least microsecond timescales.<sup>53</sup> In IMS, observed collision cross section data generally agrees with theoretical values calculated from published high resolution structures when proteins are subjected to gentle ESI conditions, further supporting this idea.<sup>35,53</sup>

These simulations unambiguously characterize the folded Ubq ions discussed above as CRM products. As the droplet shrinks, excess charge is shed through IEM ejection of solvated Na<sup>+</sup> ions, but the protein remains solvated. Ionization depends on solvent evaporation, with residual charged species from the final stage droplet deposited on the protein surface. This is analogous to the formation of salt clusters discussed in a previous paper by our group.<sup>22</sup>

### 3.3.3 Temperature Profile and Water Model Effects

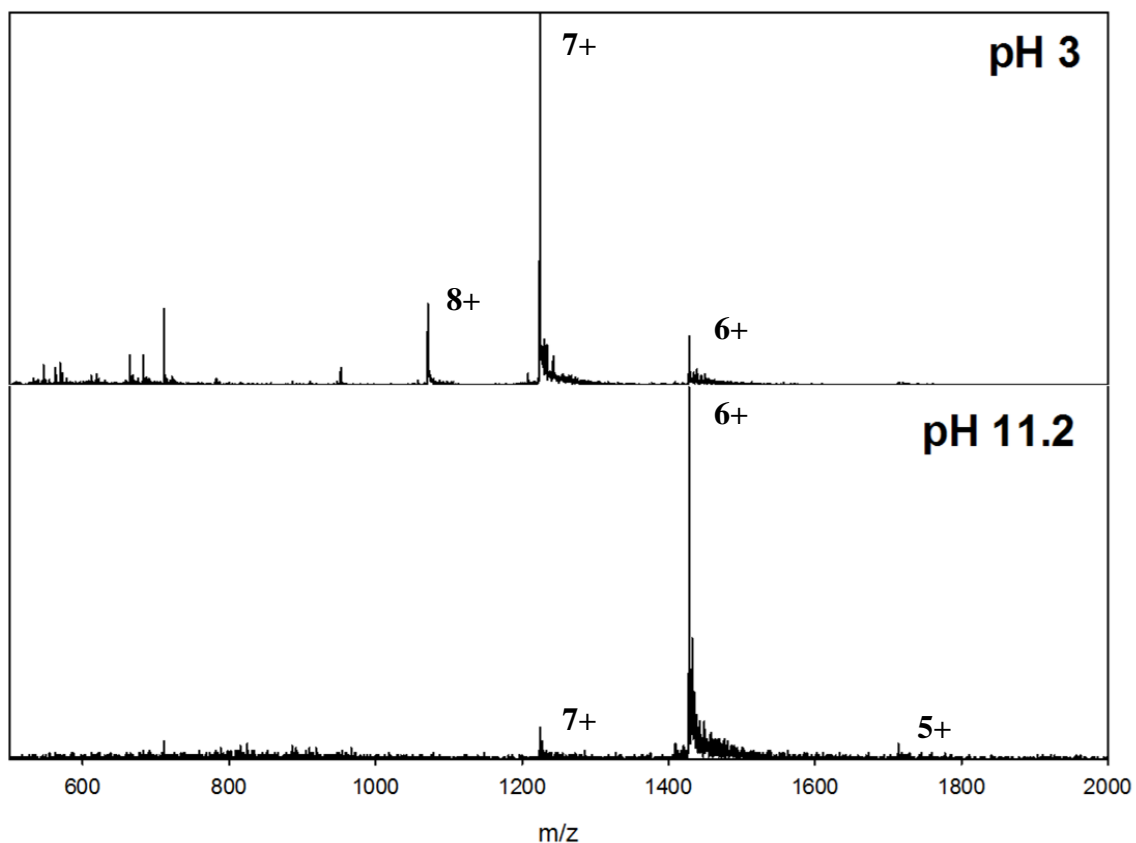
We next investigated whether the temperature profile applied in our MD simulations had an effect on the observed results. The simulations outlined above applied a thermostat a constant temperature of 370 K for 75 ns to simulate the heated gas environment in the source region followed by a temperature of 450 K to simulate collisional activation in the sampling region and ion optics. In-source activation can be made gentler by reducing instrument voltages and/or using a sub-atmospheric pressure electrospray source.<sup>5,57</sup> We therefore repeated our initial simulations in a “lower activation” regime in which droplets would be heated to a lesser extent over a longer period of time prior to reaching the sampling region. These simulations employed an initial temperature of 330 K for 125 ns, followed by a temperature of 450 K for 50 ns. Typical snapshots from these simulations are virtually indistinguishable from those in Figure 3.3, except that evaporation rate is slower during the 330 K simulation window (data not shown). Resultant protein ions from these runs are fully desolvated, near-natively folded Ubq/Na<sup>+</sup> clusters with 6 to 7 sodium ions and a mean charge of  $6.4 \pm 0.5$ . This suggests that the temperature profile in our simulations does not strongly affect Ubq charge states within the regime examined. The observation that changes in charge state caused by in-source activation appears to be

driven by protein conformational changes<sup>52</sup> suggests that a temperature regime which maintains Ubq structure ought to result in approximately the same charge state. We therefore conducted all additional simulations in the higher temperature regime, in order to reduce computation time.

We also examined the effect of the water model selected for our MD simulations. The runs discussed above employed the TIP4P/2005 water model due to its broad agreement with many different empirical properties of water across a range of temperatures<sup>49</sup>, including surface tension.<sup>58</sup> In previous MD studies, we have chosen to use a simple 3-site water model due to its low computational cost.<sup>22,25,56</sup> We therefore attempted simulations of the droplet system using the TIP3P water model.<sup>48</sup> In these simulations evaporation occurs more quickly (data not shown), possibly due to the decreased heat of vaporization in the TIP3P model relative to experiment, which may result from a less realistic description of hydrogen bonding in 3-site models.<sup>48</sup> TIP4P/2005, conversely, slightly overestimates the heat of vaporization.<sup>49</sup> Otherwise, the TIP3P droplets behaved similarly to the TIP4P/2005 droplets, but resulted in final products with charge states from 5+ to 6+, with a mean charge of  $5.6 \pm 0.5$ . This reduction in charge relative to the droplets previously discussed is possibly due to the lower surface tension of the TIP3P water model compared to TIP4P/2005.<sup>58</sup> The dependence of equation 3.1 on surface tension suggests that these droplets may eject  $\text{Na}^+$  via the IEM at a lower charge density. We find that  $\text{Na}^+$  ejection occurs at a significantly lower proportion of the Rayleigh limit in simulations using the TIP3P water model ( $78 \pm 4\%$ ) compared to TIP4P/2005 ( $89 \pm 3\%$ ). The TIP3P droplets eject sodium ions at approximately 1.14-fold lower mean charge density than the TIP4P/2005 droplets in agreement with the 1.15 to 1.30-fold reduction expected based on the square root dependence on  $\gamma$  in equation 3.1.<sup>58</sup> These data highlight both the importance of selecting an appropriate water model in MD simulation and the relevance of the Rayleigh model for our charged droplet system. We selected the TIP4P/2005 water model for all subsequent simulations, despite its higher computational cost, because of its excellent agreement with experimentally measured water surface tension values.<sup>58</sup>

### 3.3.4 Variations in Initial Protein Charge State

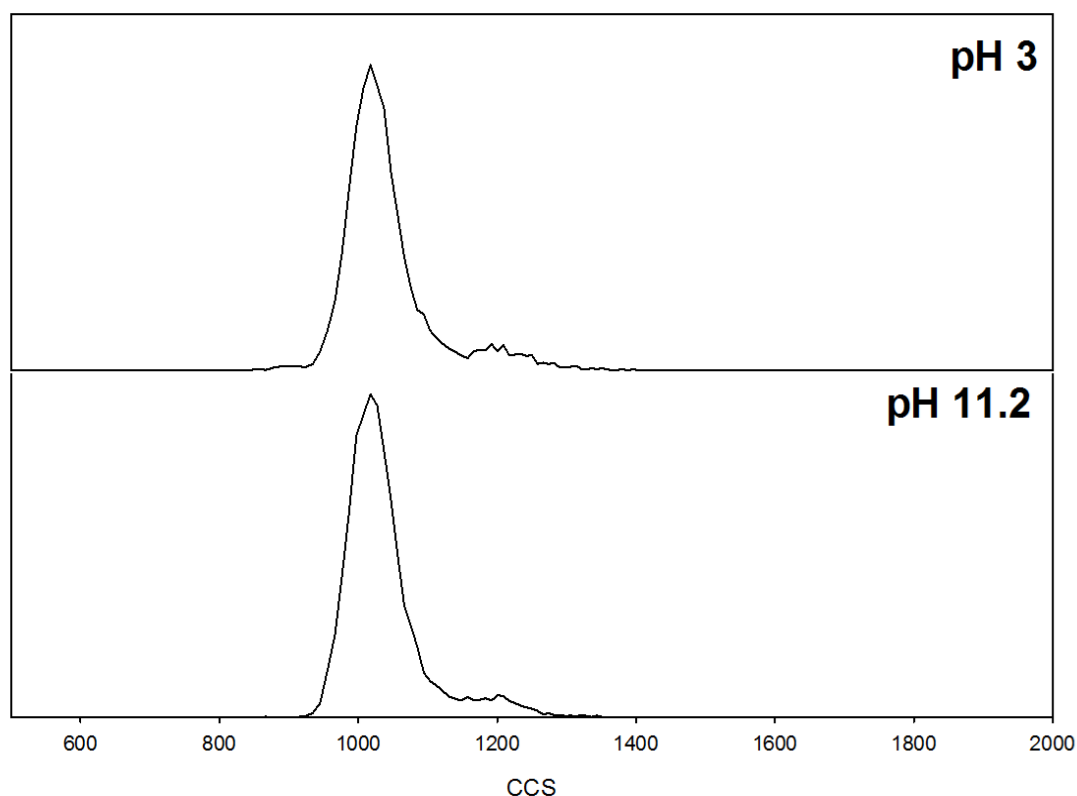
The charge states of proteins in solution can be modulated by adjusting the pH, which could potentially affect charge states of resultant ions when such a solution is used in ESI-MS. Acid is often added to protein solutions prior to ESI-MS and promotes denaturation of proteins and the formation of high charge states.<sup>25,26</sup> To investigate whether pH changes affected charge states produced for Ubq, we infused acidified (pH 3) and basified (pH 11.2) protein solutions into the ESI source. The acidic solution is sufficient to protonate most glutamic acid residues ( $pK_a = 4.25$ ) and so Ubq in solution should carry a net positive charge. Conversely the basic solution should deprotonate most lysine residues ( $pK_a = 10.5$ ), resulting in Ubq with a net negative charge. The spectra generated from these solutions are centered at low charge state (Figure 3.5), similar to pH 7 solutions (Figure 3.1a). The acidic solution is shifted to slightly higher charge (Figure



**Figure 3.5** Mass spectra of ubiquitin sprayed from acidified and basified solution. Upper panel was collected from ESI-MS of 2.65 mM aqueous formic acid, while bottom panel was from 70 mM  $\text{NH}_4\text{OH}$ .

3.5, upper panel), but lacks the distribution centered at high charge state characteristic of unfolded proteins.<sup>1</sup> We also examined these species by IMS (Figure 3.6). The 6+ charge state ions produced from both acidified and basified solutions have collision cross sections comparable to the pH 7 products (Figure 3.2), suggesting that these proteins are in native-like conformations.

We modelled these scenarios in MD by assigning non-default charges to side chains in the Ubq structure. We generated 8+ and 4+ Ubq ions by protonating glutamic acid side chains to simulate ions emerging from acidic solutions, as well as 8- and 4- ions by deprotonating lysine side chains to simulate ions emerging from basic solutions. These ions were then solvated in droplets and a sufficient number of Na<sup>+</sup> were added to bring the total system charge to 16+ (Table 3.1). Ubq with an initial charge of 8+ emerged from the droplet with 0 Na<sup>+</sup> adducts in all cases, suggesting that the protein carried sufficient charge to drive IEM ejection of Na<sup>+</sup> ions in the final stages of droplet evaporation. It's possible that one or more additional charge ejections would have



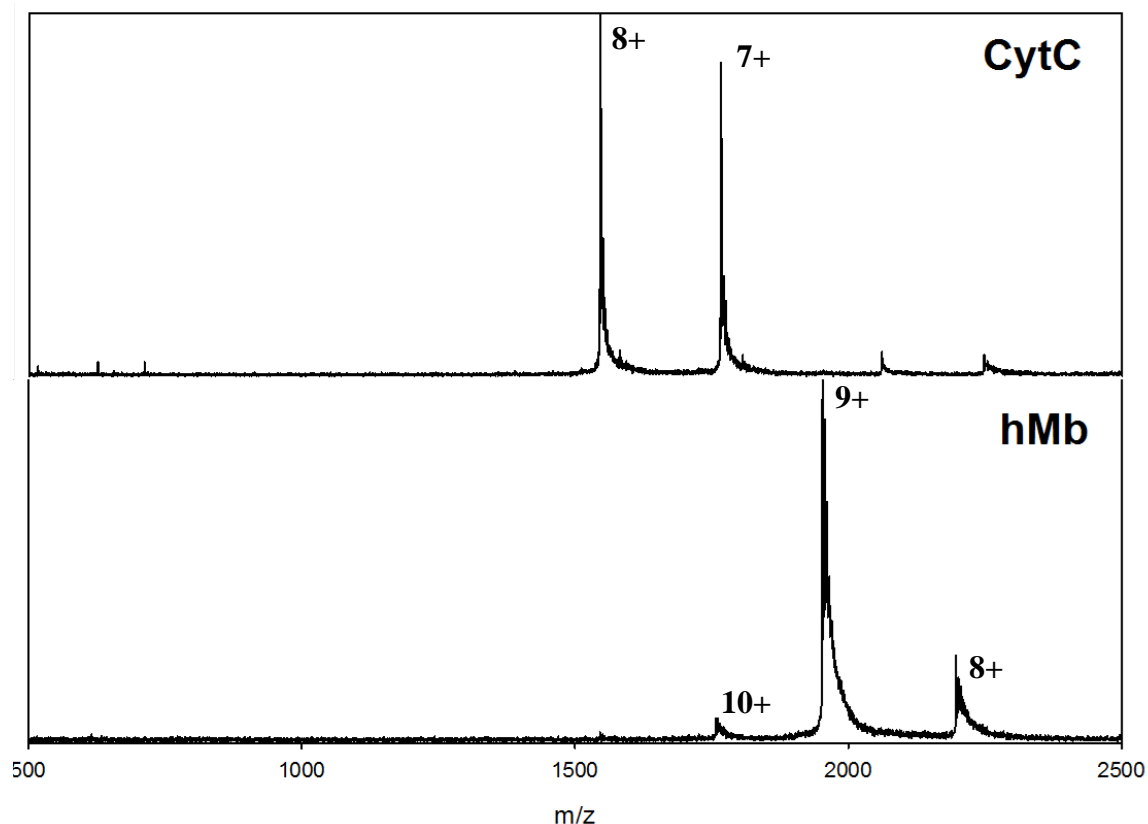
**Figure 3.6 Ion mobility spectra of 6+ ubiquitin ions from acidified and basified solutions.**



occurred in these simulations, but bonds cannot be broken in classical MD, so the ions cannot be deprotonated. However, the 4+, 4-, and 8- ions yielded final clusters with mean charges of  $6.8 \pm 0.4$  (2 to 3 adducts),  $5.8 \pm 0.4$  (9 to 10 adducts), and  $6.2 \pm 0.4$  (14 to 15 adducts), respectively. Final charge state is insensitive to the initial protein charge state for folded Ubq in our MD simulations, so long as the simulated species has a sufficient number of charge carriers to shed, in agreement with the experimental data, above (Figure 3.5 and Figure 3.6). These data suggest that protein conformation, rather than solution charge, is the primary factor determining final charge state of a protein in ESI-MS, as previously suggested.<sup>1,26</sup>

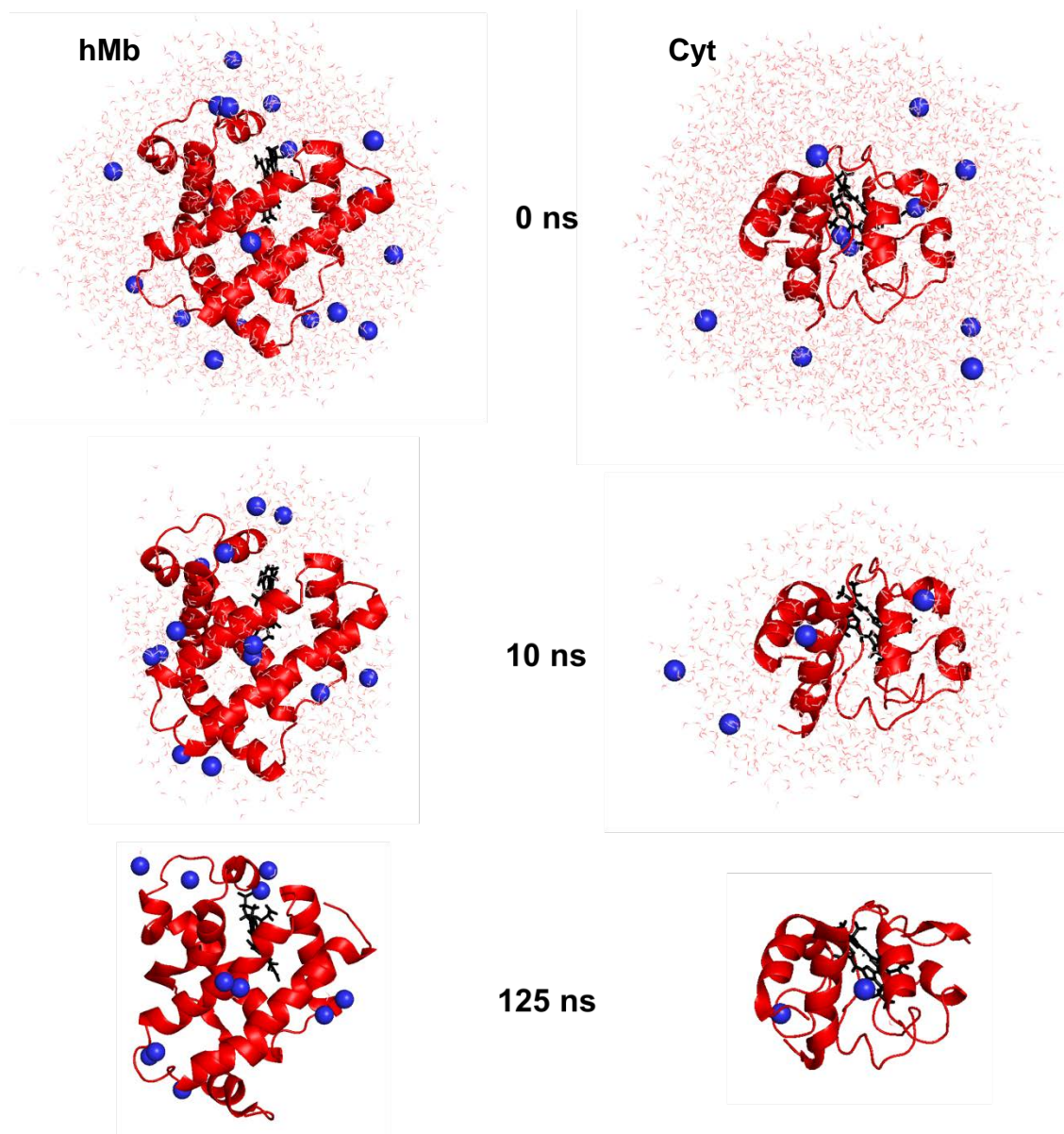
### 3.3.5 Additional Proteins

We performed additional experiments and simulations using Cyt and hMb. These proteins were selected as additional model systems so that the proteins studied would range across several different sizes (Ubq: 76 residues, 8.5 kDa; Cyt: 104 residues, 12.4 kDa; hMb: 153 residues, 17.5 kDa) and charges at pH 7 (Ubq: 0; Cyt: 6+; hMb: 2-). The mass spectra of hMb and Cyt are qualitatively similar to Ubq in that they consist only of low charge states (Figure 3.7). It is evident from the hMb spectrum that this protein has not unfolded, since there is no heme peak at  $m/z$  616, nor are there apomyoglobin peaks interleaved with the holo- species (Figure 3.7, lower panel). IMS traces of the most intense ion in each of these spectra also show that these proteins are compact and native-like (data not shown).



**Figure 3.7** Mass spectra of cytochrome *c* and holo-myoglobin at pH 7.

We performed MD simulations on droplets containing either Cyt or hMb. Typical snapshots of these simulations can be seen in (Figure 3.8). Similar to the simulations on Ubq (Figure 3.3), the proteins remain solvated until all solvent molecules have evaporated, and excess charge carriers remaining in the final stage droplet are deposited on the protein surface. The proteins remain essentially native with  $\alpha$ -carbon RMSD versus the crystal structure  $< 4 \text{ \AA}$ , even after heating in the gas phase at 450 K. The mean charge on products of these simulations is  $8.0 \pm 0.6$  for Cyt and  $8.6 \pm 0.8$  for hMb. The charge states generated in MD simulation again agree well with the distribution seen in experimental mass spectra. Our data consistently show that folded, soluble proteins emerge from nanodroplets as multiply-charged ions via a CRM process.



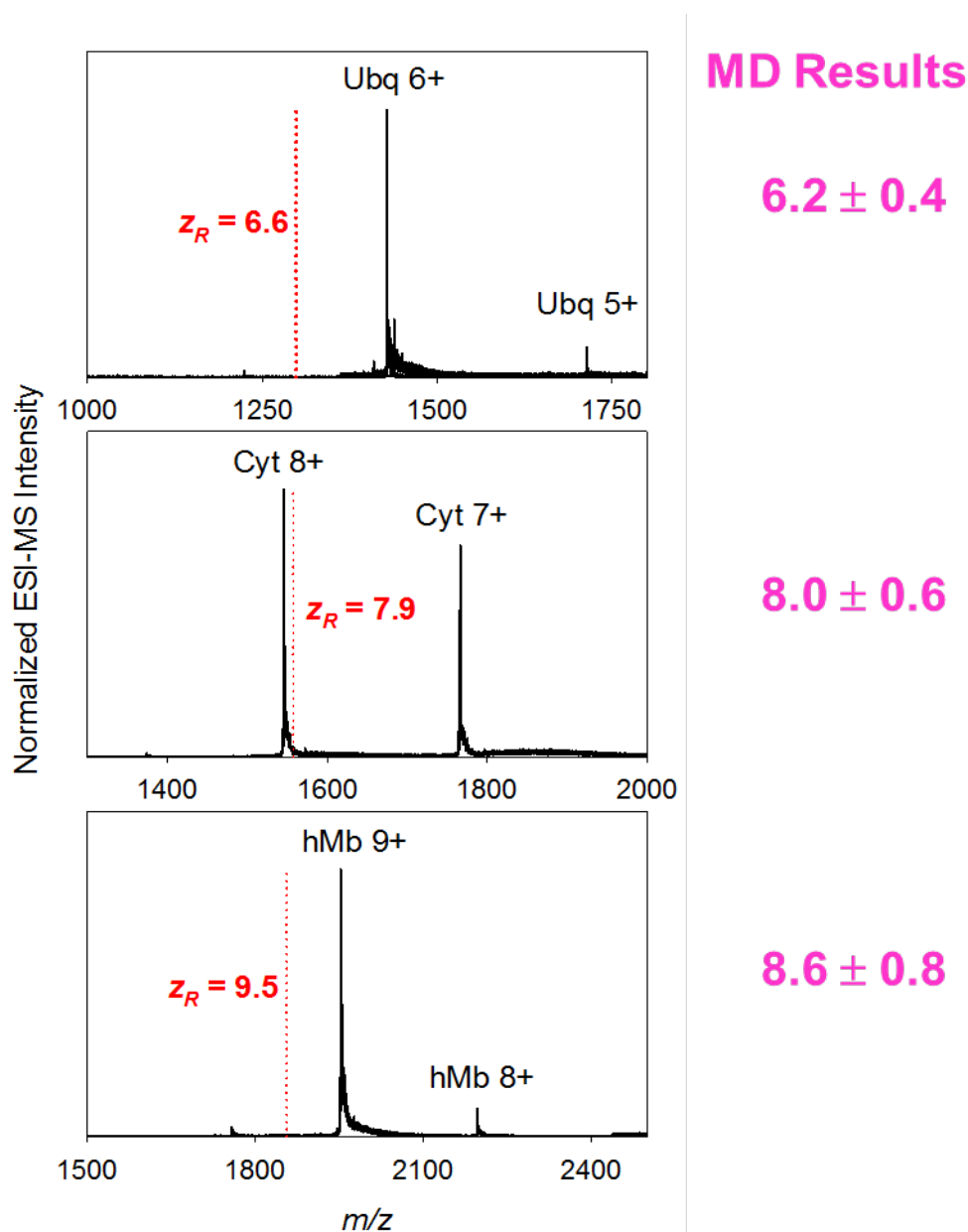
**Figure 3.8** Snapshots from MD simulations of holo-myoglobin and cytochrome *c*. Proteins are shown in cartoon representation, coloured red. Na<sup>+</sup> ions are shown in space fill representation, colored blue. Water molecules are shown as lines, coloured red. Heme groups are shown in stick representation, coloured black.

### 3.4 Conclusions

In this study, we conducted MD simulations on aqueous nanodroplets containing natively folded proteins and excess charge carried by sodium ions. As these droplets evolve, solvent molecules evaporate while proteins remain solvated, ultimately retaining the residual charge following complete desolvation. We unambiguously identify the protein

ions generated under gentle ESI conditions as CRM products, while droplet charge density is moderated by IEM ejection of small charge carrier ions, in line with previous models and simulation studies.<sup>22,23</sup> Our work adds to the emerging consensus within the ESI-MS literature that suggests the IEM applies for small, charged analytes; the CEM applies to elongated, hydrophobic analytes; and the CRM applies for compact, hydrophilic analytes.<sup>5,14,15</sup>

We note a general agreement between the charge states produced in our simulations and those observed in experimental ESI-MS spectra. Previous models have suggested that the charge remaining on these desolvated species is determined by the favourability of ion emission events which deplete the charge of the late stage droplet<sup>23,59</sup>. Our data agree with these notions insofar as the residual charged species that are not ejected from the shrinking droplet seem to become affixed to the protein surface upon desolvation. This and previous simulation studies from our lab<sup>15,22,34</sup> have shown that these ion emission events tend to maintain the droplet charge close to the Rayleigh limit. In this study, we also observed that the charge density in the evolving droplets depends on surface tension of the solvent model applied, in agreement with Rayleigh's continuum model. Further, through simulation of variously charged Ubq ion with excess Na<sup>+</sup> in these droplet systems, we showed evidence that the final charge of CRM products is determined primarily by macromolecular conformation. The ultimate charge states of our model folded systems, as well as recorded spectra, are well-predicted by the Rayleigh charge of a native-protein-sized water droplet regardless of initial solution conditions, as in the model of de la Mora<sup>20</sup> (Figure 3.8).



**Figure 3.9 Comparison of experimental data, MD simulation results and a Rayleigh charge-based model for charge states of folded proteins.** ESI-MS spectra for Ubq (top), Cyt (middle), and hMb (bottom). Vertical dashed red lines labelled  $z_R$  correspond to the charge predicted for a Rayleigh-charged droplet with the same volume as the protein. Values at right are charge states produced in our MD simulations for these systems.

This excellent agreement is promising, but it must be emphasized that the simulations presented in this work are not sufficient to reproduce an entire mass spectrum. Without additional insights into the breakup of droplets and partitioning of ions within the ESI plume, studies of this nature cannot be conducted, even if the computational demands of

such an endeavor were within reach. As well, the present MD approach is limited to study of charge carriers that can be modelled classically, such as  $\text{Na}^+$ . The use of *ab initio* methods<sup>42,55</sup> which allow for inclusion of protons and modelling of acid/base chemistry in these droplet systems would undoubtedly provide additional insights. Unfortunately, the size of the systems involved makes these approaches too computationally expensive to be feasible, at present.

It will also be informative to extend this method to larger proteins including non-covalent complexes. MD simulations to explore the ionization of unfolded proteins in ESI nanodroplets may also be within reach, but this approach will likely require some treatment of protons, since extrusion of elongated chains is predicted to be concomitant with analyte charging in these systems.<sup>14,25</sup> If *ab initio* strategies prove impractical, simulation schemes such as those developed by Fegan and Thachuk to deal with proton mobility in the gas phase<sup>60</sup> may be fruitful in such an investigation.

### 3.5 References

- 1 Kaltashov, I. A., Bobst, C. E. & Abzalimov, R. R. Mass spectrometry-based methods to study protein architecture and dynamics. *Protein Sci.* **22**, 530-544, doi:10.1002/pro.2238 (2013).
- 2 Morton, V. L., Stockley, P. G., Stonehouse, N. J. & Ashcroft, A. E. Insights into virus capsid assembly from non-covalent mass spectrometry. *Mass spectrometry reviews* **27**, 575-595 (2008).
- 3 Fenn, J. B. Electrospray Wings for Molecular Elephants (Nobel Lecture). *Angew. Chem. Int. Ed.* **42**, 3871-3894 (2003).
- 4 Piana, S., Lindorff-Larsen, K. & Shaw, D. E. Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5915-5920, doi:10.1073/pnas.1218321110 (2013).
- 5 Kebarle, P. & Verkerk, U. H. Electrospray: From Ions in Solutions to Ions in the Gas Phase, What We Know Now. *Mass Spectrom. Rev.* **28**, 898-917 (2009).
- 6 Van Berkel, G. J. *et al.* Electrochemical processes in electrospray ionization mass spectrometry. *J. Mass Spectrom.* **35**, 939-952 (2000).
- 7 Nemes, P., Marginean, I. & Vertes, A. Spraying Mode Effect on Droplet Formation and Ion Chemistry in Electrosprays. *Anal. Chem.* **79**, 3105-3116 (2007).
- 8 Covey, T. R., Thomson, B. A. & Schneider, B. B. Atmospheric Pressure Ion Sources. *Mass Spectrom. Rev.* **28**, 870-897 (2009).
- 9 Gomez, A. & Tang, K. Charge and fission of droplets in electrostatic sprays. *Phys. Fluids* **6**, 404-414 (1994).
- 10 Cech, N. B. & Enke, C. G. Practical Implication of Some Recent Studies in Electrospray Ionization Fundamentals. *Mass Spectrom. Rev.* **20**, 362-387 (2001).
- 11 Kebarle, P. & Peschke, M. On the Mechanisms by which the charged droplets produced by electrospray lead to gas phase ions. *Anal. Chim. Acta* **406**, 11-35 (2000).
- 12 Rayleigh, L. On the Equilibrium of Liquid Conducting Masses charged with Electricity. *Philos. Mag.* **14**, 184-186 (1882).
- 13 Konermann, L. A Simple Model for the Disintegration of Highly Charged Solvent Droplets during Electrospray Ionization. *J. Am. Soc. Mass Spectrom.* **20**, 496-506 (2009).
- 14 Consta, S., Oh, M. I. & Soltani, S. Advances in the theoretical and molecular simulation studies of the ion chemistry in droplets. *International Journal of Mass Spectrometry* (2014).
- 15 Konermann, L., Ahadi, E., Rodriguez, A. D. & Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **85**, 2-9 (2013).

- 16 Nguyen, S. & Fenn, J. B. Gas-phase ions of solute species from charged droplets of solutions. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1111-1117 (2007).
- 17 Iribarne, J. V. & Thomson, B. A. On the evaporation of small ions from charged droplets. *J. Chem. Phys.* **64**, 2287-2294 (1976).
- 18 Labowsky, M. A model for solvated ion emission from electrospray droplets. *Rapid Commun. Mass Spectrom.* **24**, 3079-3091 (2010).
- 19 Tang, L. & Kebarle, P. Dependence of ion intensity in electrospray mass spectrometry on the concentration of the analytes in the electrosprayed solution. *Anal. Chem.* **65**, 3654-3668 (1993).
- 20 de la Mora, F. J. Electrospray Ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal. Chim. Acta* **406**, 93-104 (2000).
- 21 Dole, M. *et al.* Molecular beams of macroions. *J. Chem. Phys.* **49**, 2240-2249 (1968).
- 22 Konermann, L., McAllister, R. G. & Metwally, H. Molecular Dynamics Simulations of the Electrospray Process: Formation of NaCl Clusters via the Charged Residue Mechanism. *J. Phys. Chem. B* **118**, 12025-12033, doi:10.1021/jp507635y (2014).
- 23 Hogan, C. J., Carroll, J. A., Rohrs, H. W., Biswas, P. & Gross, M. L. Combined Charged Residue-Field Emission Model of Macromolecular Electrospray Ionization. *Anal. Chem.* **81**, 369-377 (2009).
- 24 Wang, G. & Cole, R. B. Charged residue versus ion evaporation for formation of alkali metal halide clusters ions in ESI. *Anal. Chim. Acta* **406**, 53-65 (2000).
- 25 Ahadi, E. & Konermann, L. Modeling the behavior of coarse-grained polymer chains in charged water droplets: Implications for the mechanism of electrospray ionization. *J. Phys. Chem. B* **116**, 104-112 (2012).
- 26 Chowdhury, S. K., Katta, V. & Chait, B. T. Probing Conformational Changes in Proteins by Mass Spectrometry. *J. Am. Chem. Soc.* **112**, 9012-9013 (1990).
- 27 Yue, X., Vahidi, S. & Konermann, L. Insights into the Mechanism of Protein Electrospray Ionization From Salt Adduction Measurements. *J. Am. Soc. Mass Spectrom.* **25**, 1322-1331, doi:10.1007/s13361-014-0905-0 (2014).
- 28 Caleman, C. & van der Spoel, D. Evaporation from water clusters containing singly charged ions. *Phys. Chem. Chem. Phys.* **9**, 5105-5111 (2007).
- 29 Daub, C. D. & Cann, N. M. How Are Completely Desolvated Ions Produced in Electrospray Ionization: Insights from Molecular Dynamics Simulations. *Anal. Chem.* **83**, 8372-8376 (2011).
- 30 Ichiki, K. & Consta, S. Disintegration Mechanisms of Charged Aqueous Nanodroplets Studied by Simulations and Analytical Models. *J. Phys. Chem. B* **110**, 19168-19175 (2006).



- 31 Patriksson, A., Marklund, E. & van der Spoel, D. Protein Structures under Electrospray Conditions. *Biochemistry* **46**, 933-945 (2007).
- 32 Steinberg, M. Z., Breuker, K., Elber, R. & Gerber, R. B. The dynamics of water evaporation from partially solvated cytochrome c in the gas phase. *Phys. Chem. Chem. Phys.* **9**, 4690-4697 (2007).
- 33 Znamenskiy, V., Marginean, I. & Vertes, A. Solvated Ion Evaporation from Charged Water Droplets. *J. Phys. Chem. A* **107**, 7406-7412 (2003).
- 34 Metwally, H., McAllister, R. G. & Konermann, L. Exploring the Mechanism of Salt-Induced Signal Suppression in Protein Electrospray Mass Spectrometry Using Experiments and Molecular Dynamics Simulations. *Analytical chemistry* **87**, 2434-2442 (2015).
- 35 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. Protocols* **3**, 1139-1152 (2008).
- 36 Van der Spoel, D. *et al.* GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **26**, 1701-1718, doi:10.1002/jcc.20291 (2005).
- 37 Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. Structure of Ubiquitin Refined at 1.8 Å Resolution. *J. Mol. Biol.* **194**, 531-544 (1987).
- 38 Bushnell, G. W., Louie, G. V. & Brayer, G. D. High-resolution Three-dimensional Structure of Horse Heart Cytochrome c. *J. Mol. Biol.* **214**, 585-595 (1990).
- 39 Maurus, R. *et al.* A myoglobin variant with a polar substitution in a conserved hydrophobic cluster in the heme binding pocket. *Biochim. Biophys. Acta* **1341**, 1-13 (1997).
- 40 MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B* **102**, 3586-3616 (1998).
- 41 Autenrieth, F., Tajkhorshid, E., Baudry, J. & Luthey-Schulten, Z. Classical force field parameters for the heme prosthetic group of cytochrome c. *Journal of computational chemistry* **25**, 1613-1622 (2004).
- 42 Marx, D., Tuckerman, M. E., Hutter, J. & Parrinello, M. The nature of the hydrated excess proton in water. *Nature* **397**, 601-604 (1999).
- 43 Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695-1697 (1985).
- 44 Nose, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255-268 (1984).
- 45 Verlet, L. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical review* **159**, 98 (1967).

- 46 Hess, B., Bekker, H., Berendsen, H. J. & Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry* **18**, 1463-1472 (1997).
- 47 Miyamoto, S. & Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **13**, 952-962, doi:10.1002/jcc.540130805 (1992).
- 48 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935 (1983).
- 49 Abascal, J. L. & Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *The Journal of chemical physics* **123**, 234505 (2005).
- 50 Merenbloom, S. I., Flick, T. G. & Williams, E. R. How Hot are Your Ions in TWAVE Ion Mobility Spectrometry? *J. Am. Soc. Mass Spectrom.* **23**, 553-562 (2012).
- 51 Gabelica, V. & De Pauw, E. Internal Energy and Fragmentation of Ions Produced in Electrospray Sources. *Mass Spectrom. Rev.* **24**, 566-587 (2005).
- 52 Cassou, C. A. & Williams, E. R. Anions in Electrothermal Supercharging of Proteins with Electrospray Ionization Follow a Reverse Hofmeister Series. *Analytical Chemistry*, doi:10.1021/ac403398j (2014).
- 53 Wytenbach, T. & Bowers, M. T. Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-Free Ion Mobility–Mass Spectrometry Environment. *J. Phys. Chem. B* **115**, 12266-12275 (2011).
- 54 Shvartsburg, A. A. & Jarrold, M. F. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **261**, 86-91, doi:10.1016/0009-2614(96)00941-4 (1996).
- 55 Schmitt, U. W. & Voth, G. A. Multistate empirical valence bond model for proton transport in water. *J. Phys. Chem. B.* **102**, 5547-5551 (1998).
- 56 Ahadi, E. & Konermann, L. Surface charge of electrosprayed water nanodroplets: A molecular dynamics study. *J. Am. Chem. Soc.* **132**, 11270-11277 (2010).
- 57 Page, J. S., Tang, K., Kelly, R. T. & Smith, R. D. Subambient pressure ionization with nanoelectrospray source and interface for improved sensitivity in mass spectrometry. *Analytical chemistry* **80**, 1800-1805 (2008).
- 58 Vega, C. & de Miguel, E. Surface tension of the most popular models of water by using the test-area simulation method. *J. Chem. Phys.* **126**, 154707, doi:10.1063/1.2715577 (2007).
- 59 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.* **85**, 12055–12061 (2013).

- 60 Fegan, S. K. & Thachuk, M. A Charge Moving Algorithm for Molecular Dynamics Simulations of Gas-Phase Proteins. *J. Chem. Theory Comput.* **9**, 2531-2539, doi:10.1021/ct300906a (2013).

## 4 Conclusions and Future Work

### 4.1 Conclusions

In this work we have applied molecular dynamics simulations to shed light on fundamental aspects of structural mass spectrometry techniques. These simulations provide an atomistic view into the unobservable processes that underpin the data observed in experiment. We examined the behavior of proteins in both bulk solution and isolated nanometer sized droplets, resulting in new insights that will inform interpretation of Hydrogen/Deuterium exchange rates and electrospray ionization charge state distributions.

In Chapter 2, we examined the microscopic phenomena that underlie Hydrogen/Deuterium exchange (HDX), a common technique in protein structural investigations. The work presented focuses on ubiquitin for exploring the defining features that distinguish amides in “open” (exchange-competent) and “closed” (exchange-incompetent) environments. Instead of relying on static X-ray structures, we employed all-atom molecular dynamics (MD) simulations for obtaining a dynamic view of the protein ground state and its surrounding solvent. The HDX properties for 57 out of 72 NH sites can be readily explained on the basis of backbone and side chain H-bonding, as well as solvent accessibility considerations. Unexpectedly, the same criteria fail for predicting the HDX characteristics of the remaining 15 amides. Significant protection is seen for numerous exposed NH sites that are not engaged in intramolecular H-bonds, whereas other amides that seemingly share the same features are unprotected. We scrutinized the proposal that H-bonding to crystallographically defined water can cause the protection of surface amides. For ubiquitin, the positioning of crystal water is not compatible with this idea, nor do our MD data reveal a difference in the solvation properties of protected vs. unprotected surface amides, making it unlikely that restricted water dynamics can cause anomalous amide protection. The findings reported here suggest that efforts to deduce protein structural features on the basis of HDX protection factors may yield misleading results. This conclusion is relevant for initiatives that rely on sparse structural data as constraints for elucidating protein conformations. This study

demonstrates the utility of MD simulation for examining protein behavior observed in structural MS. We next sought to apply MD simulations to a mechanistic study of the electrospray ionization process, which produces the gas phase ions central to HDX and other structural MS experiments.

In Chapter 3, we studied the mechanism of Electrospray Ionization (ESI) for natively folded proteins. Electrospray ionization (ESI) coupled to mass spectrometry (MS) is a commonly used analytical technique for probing protein structure and dynamics. We employed all-atom molecular dynamics simulations to explore the process by which natively folded proteins are released from nanometer sized aqueous droplets into the gas phase. A trajectory stitching approach was applied to maintain system temperature and remove evaporated solvent as the droplet evolves. We find that shrinking droplets maintain charge densities close to the Rayleigh limit<sup>1,2</sup> through ejection of small charge carriers such as  $\text{Na}^+$ , while protein molecules remain solvated in the droplet interior. As the last solvent molecules evaporate, the proteins are ionized by residual charge carriers which remained at the surface of the vanishing droplet. Our data provide clear evidence that protein ions formed during native ESI are produced by Dole's charged residue model (CRM).<sup>3</sup> Moreover, we find that the ultimate charge states produced in these simulations agree with experimental MS charge state distributions, are predictable based on Rayleigh's continuum model<sup>1</sup>, and are insensitive to protein charge in the initial solution. To our knowledge, this study represents the first MD simulation to achieve complete desolvation of a protein ion in a CRM process. Once again, the utility of MD for interrogation of microscopic phenomena that lead to the  $m/z$  values observed in MS is highlighted.

Simulation and computation have been called the third pillar of modern science, alongside theory and experiment. Here, we have used MD simulations to explore fundamental aspects of structural mass spectrometry. The true value in our simulations comes in viewing them alongside experimental data and theory. In Chapter 2, we reaffirmed the H-bond-centric view of H/D exchange protection, but found that more than 20% of HDX protection observed in ubiquitin could not be rationalized by H-bonding or solvent accessibility. Researchers relying on HDX data for structural

initiatives will be well-served in applying similar simulation methods to better understand how their measured exchange rates relate to secondary and tertiary structure of proteins. Similarly, in Chapter 3, we found excellent agreement between our simulations and experiments, as well as Rayleigh's continuum model, for predicting the final charge state of a folded protein under gentle ESI-MS conditions. As computational power continues to expand, we foresee that studies marrying simulation and experiment, like those presented here, will be fruitful ground for critical evaluation and enhanced understanding of the fundamental aspects of structural mass spectrometry and other analytical techniques.

## 4.2 Future Directions of Study

### 4.2.1 Extended Simulations and Additional Proteins in HDX/MS

Extended, millisecond scale simulations on ubiquitin, such as those performed by Shaw and co-workers<sup>4</sup>, could provide additional insights into HDX at an amide-resolved level. In particular, these simulations could provide considerably greater conformational sampling, and the possibility for quantitative correlation of protection factors, based on amide properties in unfolded protein states. These simulations could also provide insights into how the extent of correlation in the opening and closing events of various amides, an important aspect of interpreting HDX at the peptide level.<sup>5</sup> Studies of additional proteins could also facilitate greater understanding of HDX protection. Detailed studies on Staphylococcal nuclease by Englander and co-workers<sup>6,7</sup> could make it a natural first choice for such endeavours.

### 4.2.2 Gas Phase HDX and Pulsed HDX

Our studies in Chapter 2 focused on solution-phase HDX data gathered in continuous labelling experiments, where a protein spends a long period of time solvated in D<sub>2</sub>O. However, HDX can also be carried out for comparatively short periods by pulsed labelling, where the protein is only exposed to D<sub>2</sub>O for as short a time as possible before exchange is "quenched" at low temperature and acidic pH.<sup>8</sup> Gas phase HDX can be carried out in a mass spectrometer through introduction of D<sub>2</sub>O vapor into the sampling or ion transport regions.<sup>9</sup> MD simulations could be especially useful in probing

the exchange phenomena in these regimes, since it can potentially achieve experimentally-relevant timescales. Challenges will be in generating appropriate starting structures, since these techniques are often applied to examine partially unfolded or disordered species, which are not well characterized by high-resolution structural techniques.<sup>8,10</sup>

### 4.2.3 ESI-MS of Noncovalent Complexes

In Chapter 3, it was found that proteins maintained their solution-phase structure upon transmission to the gas phase, but only monomeric proteins were discussed. The simulations and experiments on holo-myoglobin which show heme retained in the protein structure strongly suggest that this should also be the case for larger noncovalent complexes, including multimeric proteins. Since large proteins with many subunits are often the focus of structural mass spectrometry experiments<sup>11,12</sup>, a natural extension of the work presented would be exploration of the retention of quaternary structure in ESI-MS. Simulation of a tetrameric protein such as hemoglobin could be informative in such a study.

### 4.2.4 Additional Macromolecules in ESI-MS

We extensively interrogated ESI-MS of folded proteins, but examination of other analytes may reveal further insights. Peptides, such as those created from pepsin or trypsin digestion of proteins, are often analyzed by ESI-MS in protein studies in order to localize modifications on the protein structure. Short peptides may emerge from nanodroplets via a different mechanism than natively folded proteins due to their small size and disordered structure. Interrogation of the ESI mechanisms of additional macromolecules of biological or pharmaceutical importance including lipids, nucleic acids, oligosaccharides, or dendrimers could also be informative.

### 4.3 References

- 1 de la Mora, F. J. Electrospray Ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal. Chim. Acta* **406**, 93-104 (2000).
- 2 Rayleigh, L. On the Equilibrium of Liquid Conducting Masses charged with Electricity. *Philos. Mag.* **14**, 184-186 (1882).
- 3 Dole, M. *et al.* Molecular beams of macroions. *J. Chem. Phys.* **49**, 2240-2249 (1968).
- 4 Piana, S., Lindorff-Larsen, K. & Shaw, D. E. Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5915-5920, doi:10.1073/pnas.1218321110 (2013).
- 5 Weis, D. D., Wales, T. E., Engen, J. R., Hotchko, M. & Ten Eyck, L. F. Identification and characterization of EX1 kinetics in H/D exchange mass spectrometry by peak width analysis. *J. Am. Soc. Mass Spectrom.* **17**, 1498-1509, doi:10.1016/j.jasms.2006.05.014 (2006).
- 6 Skinner, J. J., Lim, W. K., Bédard, S., Black, B. E. & Englander, S. W. Protein dynamics viewed by hydrogen exchange. *Protein Sci.* **21**, 996-1005 (2012).
- 7 Skinner, J. J., Lim, W. K., Bedard, S., Black, B. E. & Englander, S. W. Protein hydrogen exchange: Testing current models. *Protein Sci.* **21**, 987-995, doi:10.1002/pro.2082 (2012).
- 8 Vahidi, S., Stocks, B. B., Liaghati-Mobarhan, Y. & Konermann, L. Mapping pH-Induced Protein Structural Changes Under Equilibrium Conditions by Pulsed Oxidative Labeling and Mass Spectrometry. *Anal. Chem.* **84**, 9124-9130 (2012).
- 9 Rand, K. D. *et al.* Gas-Phase Hydrogen/Deuterium Exchange in a Traveling Wave Ion Guide for the Examination of Protein Conformations. *Anal. Chem.* **81**, 10019-10028 (2009).
- 10 Zhang, Y. *et al.* Pulsed hydrogen-deuterium exchange mass spectrometry probes conformational changes in amyloid beta (A beta) peptide aggregation. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 14604-14609, doi:10.1073/pnas.1309175110 (2013).
- 11 Zhou, M., Jones, C. M. & Wysocki, V. H. Dissecting the Large Noncovalent Protein Complex GroEL with Surface-Induced Dissociation and Ion Mobility–Mass Spectrometry. *Anal. Chem.* **85**, 8262-8267, doi:10.1021/ac401497c (2013).
- 12 Marty, M. T. *et al.* Native Mass Spectrometry Characterization of Intact Nanodisc Lipoprotein Complexes. *Anal. Chem.* **84**, 8957-8960, doi:10.1021/ac302663f (2012).



## Curriculum Vitae

- Name:** Robert Gordon McAllister
- Post-secondary Education and Degrees:**
- The University of Western Ontario  
London, Ontario, Canada  
2009-2013 Honors B.M.Sc. (Biochemistry)  
Undergraduate Thesis: *“Identifying lentivector integration sites in a murine model of metachromatic leukodystrophy.”*  
Principle Investigators: Stephen D. Barr and C. Anthony Rugar
- The University of Western Ontario  
London, Ontario, Canada  
2013-2015 M.Sc. (Chemistry)  
Graduate Thesis: *“From Solution Into the Gas Phase: Studying Protein Hydrogen Exchange and Electrospray Ionization Using Molecular Dynamics Simulation”*  
Principle Investigator: Lars Konermann
- Honours and Awards:**
- Queen Elizabeth II Graduate Scholarship in Science & Technology  
2014-2015
- Dean’s Honor List – The University of Western Ontario  
2009-2013
- Continuing Education Scholarship – UWO  
2009-2013
- LAXESS P.A.I.R.S. Scholarship – LANXESS Canada (in support of studies at UWO)  
2009-2013
- Catherine Charlotte Wilson Scholarship – UWO  
2010
- Jean Ewing Martin Scholarship – Lambton-Kent District School Board (in support of studies at UWO)  
2009

**Related Work Experience**      Teaching Assistant  
The University of Western Ontario  
2013-2014

Summer Research Assistant  
The University of Western Ontario  
2013

### Peer-Reviewed Publications:

McAllister, R. G. & Konermann, L. Challenges in the Interpretation of Protein H/D Exchange Data: A Molecular Dynamics Simulation Perspective. *Biochemistry* **54**, 2683-2692 (2015).

Metwally, H., McAllister, R. G. & Konermann, L. Exploring the mechanism of salt-induced signal suppression in protein electrospray mass spectrometry using experiments and molecular dynamics simulations. *Anal. Chem.* **87**, 2434-2442 (2015).

Konermann, L., McAllister, R. G. & Metwally, H. Molecular dynamics simulation of the electrospray process: formation of NaCl clusters via the charged residue mechanism. *J. Phys. Chem. B* **118**, 12025-12033 (2014).

McAllister, R. G., Liu, J., Woods, M. W., Tom, S. K., Rupar, C. A. & Barr, S. D. Lentivector integration sites in ependymal cells from a model of metachromatic leukodystrophy: non-B DNA as a new factor influencing integration. *Mol. Ther. Nucleic Acids* **3**, e187 (2014).

### Conference Presentations:

McAllister, R. G. & Konermann, L. How well do we understand protein HDX protection patterns? A molecular dynamics simulation study. Poster. *63rd Annual ASMS Conference* (2015; international conference; ~5000 attendees)

McAllister, R. G. & Konermann, L. Exploring the mechanism of hydrogen exchange in solution using MD simulations. Talk. *31st Trent Conference on Mass Spectrometry* (2014; international conference; ~50 attendees)

McAllister, R. G., Barr, S. D. & Rupar, C. A. Identifying lentivector integration sites in a murine model of metachromatic leukodystrophy. Poster. *Harold B. Stewart Memorial Lecture and Research Showcase* (2013; institutional conference; ~100 attendees)