Charge Hopping in Coarse-Grained Simulations of Gas-Phase Protein Complexes

by

Sarah Katherine Fegan

B.Sc., The University of Waterloo, Waterloo, ON, Canada, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in
The Faculty of Graduate and Postdoctoral Studies
(Chemistry)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)
October 2014
© Sarah Katherine Fegan 2014
Abstract

With the invention of gentle ionization methods such as electrospray ionization, mass spectrometry has been used as a tool for studying large protein complexes. Simulations of these gas-phase protein complexes will allow for better understanding of the dissociation mechanism and lead to methods for controlling the dissociation. Controlling the dissociation will help obtain structural information from the mass spectrometry experiments.

In this work, the suitability of the MARTINI coarse-grained force field for gas-phase simulations is studied and a charge hopping algorithm is developed. Using a coarse-grained force field makes the simulations faster so that longer simulation times can be accessed. This is important because protein motions can take place on time scales of nanoseconds to milliseconds and these long times are not practical with all-atom simulations. Most molecular dynamics simulations use fixed charges, but including charge motion allows for better simulation of mobile protons.

Two protein complexes are studied here, one dimer and one tetramer. Hopping rates, energies, radii of gyration, and distances within the complexes are calculated. Simulations with the cytochrome c' dimer (no charge hopping) are compared to published all-atom results. The MARTINI force field is found to be good for qualitative results, but slightly more attractive than the OPLS all-atom (for the isolated protein complex). The transthyretin tetramer is used to study the hopping algorithm. Modifications of the protein (blocking N-termini from accepting charges and adding basic sites with a tether) are also explored.

The dissociation behavior of the protein complexes is controlled by the Coulomb repulsion model. Protein modifications near the N-termini show potential for controlling the dissociation.
Preface

The work presented in this thesis has been published by the author, Sarah K. Fegan, in co-authorship with research supervisor Prof. Mark Thachuk. Sarah K. Fegan is the main writer and first author of these works.

# Table of Contents

Abstract ................................................................. ii  
Preface ................................................................. iii  
Table of Contents ................................................... iv  
List of Tables ......................................................... vi  
List of Figures ......................................................... vii  
List of Abbreviations and Symbols ................................. ix  
Acknowledgements .................................................. xi  
Dedication ............................................................. xii  

1 Introduction ......................................................... 1  
  1.1 Mass Spectrometry Experiments ................................ 1  
  1.2 Computational Background ...................................... 6  
  1.3 Coarse Grained Models ........................................... 7  
  1.4 Goals ............................................................. 8  

2 Methods ............................................................ 9  
  2.1 Molecular Dynamics ............................................. 9  
  2.2 Input Parameters ................................................ 11  
  2.3 Properties for Analysis of Trajectories ....................... 13  
  2.4 Software and Force Field Particular to This Work .......... 14  
  2.5 Moving Charges ................................................ 17  

3 Suitability of the MARTINI Force Field For Use With Gas-Phase Protein Complexes ........................................... 22  
  3.1 Introduction ..................................................... 22  
  3.2 Method .......................................................... 22
3.3 Results and Discussion ........................................... 27
3.4 Conclusions ....................................................... 41

4 A Charge Moving Algorithm for Molecular Dynamics Simulations of Gas-
phase Proteins .......................................................... 45
  4.1 Introduction ....................................................... 45
  4.2 Method ............................................................ 46
  4.3 Results and Discussion ........................................... 47
     4.3.1 Hopping Rate and Charge Distributions ............... 47
     4.3.2 Structural Properties .................................... 53
     4.3.3 Energy Distributions and Pair Correlations .......... 53
     4.3.4 Computational Performance ............................. 61
  4.4 Conclusions ....................................................... 61

5 Controlling Dissociation Channels of Gas-Phase Protein Complexes Using
Charge Manipulation ..................................................... 65
  5.1 Introduction ....................................................... 65
  5.2 Method ............................................................ 65
  5.3 Results and Discussion ........................................... 66
  5.4 Conclusions ....................................................... 73

6 Concluding Remarks .................................................. 75
  6.1 Conclusions ....................................................... 75
  6.2 Future Work ....................................................... 76

Bibliography ............................................................ 78

Appendices

A Source Code .......................................................... 83
  A.1 mobile_charge.c ................................................. 83
  A.2 mobile_charge.h ................................................. 93
List of Tables

2.1 Proton binding energy values for basic amino acid sites[1]. . . . . . . . . . . 20

3.1 The residue numbers of the charged amino acids on each monomer for the
ten M8/M2 and ten M5/M5 charge distributions. . . . . . . . . . . . . . . . 24
## List of Figures

1.1 Experimental mass spectra of cytochrome c dimer ......................... 3  
1.2 Experimental mass spectra of shiga-like toxin pentamer .................. 5  
2.1 A skeletal diagram of the GROMACS molecular dynamics program. ...... 16  
2.2 A flowchart showing the hopping algorithm. .................................. 18  
3.1 Snapshot of cytochrome c monomer ........................................... 23  
3.2 Snapshot of ground state cytochrome c dimer ............................... 26  
3.3 Pull force for M5/M5 cytochrome c dimer ................................... 28  
3.4 Pull force for M8/M2 cytochrome c dimer ................................... 30  
3.5 Percentage dissociation of cytochrome c dimer ............................ 31  
3.6 Minimum separation of cytochrome c dimer .................................. 33  
3.7 Radius of gyration of cytochrome c dimer .................................. 35  
3.8 RMSD of cytochrome c dimer ................................................... 36  
3.9 Effect of secondary structure constraints on cytochrome c dimer ........ 39  
3.10 Intramolecular energies for cytochrome c dimer ......................... 40  
3.11 Intermolecular energies for cytochrome c dimer ......................... 42  
3.12 Snapshots of cytochrome c dimer ............................................ 43  
4.1 Hopping rate for transthyretin tetramer ..................................... 48  
4.2 Potential and kinetic energies of transthyretin tetramer ................ 50  
4.3 Charge configurations of transthyretin tetramer .......................... 51  
4.4 Histogram of charge distributions of transthyretin tetramer ............ 52  
4.5 Radius of gyration of transthyretin tetramer ............................... 54  
4.6 Snapshots of transthyretin tetramer ........................................ 55  
4.7 Energy distributions of attempted hops for transthyretin tetramer .... 56  
4.8 Energy distributions of attempted hops by donor type for transthyretin tetramer .................................................. 59  
4.9 Energy distributions of attempted hops by donor and acceptor type for transthyretin tetramer .................................................. 59  
4.10 Energy distributions of successful hops for transthyretin tetramer .... 60  
4.11 Hopping pairs of transthyretin tetramer ................................... 62
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Snapshots of transthyretin tetramer</td>
<td>67</td>
</tr>
<tr>
<td>5.2</td>
<td>Snapshots of transthyretin tetramer with tether</td>
<td>68</td>
</tr>
<tr>
<td>5.3</td>
<td>Charge configurations and radius of gyration for transthyretin tetramer</td>
<td>70</td>
</tr>
<tr>
<td>5.4</td>
<td>Minimum distances in transthyretin tetramer at 300K</td>
<td>71</td>
</tr>
<tr>
<td>5.5</td>
<td>Minimum distances in transthyretin tetramer at 600K</td>
<td>72</td>
</tr>
</tbody>
</table>
## List of Abbreviations and Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>hopping probability tuning parameter</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>dielectric constant</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>Lennard-Jones well depth</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Lennard-Jones cross section</td>
</tr>
<tr>
<td>BIRD</td>
<td>blackbody infrared radiative dissociation</td>
</tr>
<tr>
<td>CID</td>
<td>collisionally induced dissociation</td>
</tr>
<tr>
<td>E</td>
<td>energy</td>
</tr>
<tr>
<td>ESI</td>
<td>electrospray ionization</td>
</tr>
<tr>
<td>F</td>
<td>force</td>
</tr>
<tr>
<td>$f$</td>
<td>degrees of freedom</td>
</tr>
<tr>
<td>k</td>
<td>force constant</td>
</tr>
<tr>
<td>$k_B$</td>
<td>Boltzmann constant</td>
</tr>
<tr>
<td>LINCS</td>
<td>Linear Constraint Solver for Molecular Simulations</td>
</tr>
<tr>
<td>LJ</td>
<td>Lennard-Jones</td>
</tr>
<tr>
<td>m</td>
<td>mass</td>
</tr>
<tr>
<td>M2</td>
<td>monomer with +2 charges</td>
</tr>
<tr>
<td>M5</td>
<td>monomer with +5 charges</td>
</tr>
<tr>
<td>M8</td>
<td>monomer with +8 charges</td>
</tr>
<tr>
<td>MD</td>
<td>molecular dynamics</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>m/z</td>
<td>mass to charge ratio</td>
</tr>
<tr>
<td>N</td>
<td>number of beads in the system</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>q</td>
<td>charge</td>
</tr>
<tr>
<td>$R_g$</td>
<td>radius of gyration</td>
</tr>
<tr>
<td>RMSD</td>
<td>root mean square deviation</td>
</tr>
<tr>
<td>SID</td>
<td>surface induced dissociation</td>
</tr>
<tr>
<td>SORI</td>
<td>sustained off-resonance irradiation</td>
</tr>
<tr>
<td>T</td>
<td>temperature</td>
</tr>
<tr>
<td>t</td>
<td>time</td>
</tr>
</tbody>
</table>
TTR transthyretin tetramer
TOF time of flight (mass spectrometer)
v velocity
V(r) potential
Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor Dr. Mark Thachuk for all his guidance and encouragement. His support and advice in editing this thesis is greatly appreciated.

I would like to thank my committee members, especially Dr. Roman Krems, for reading the thesis.

I would also like to thank the members of the theory lab for making my time at UBC so enjoyable. In particular, Nalantha, Olga and Saheba have been a source of great help and friendship.

I would like to thank my family for their unconditional love and support.

I acknowledge financial support from the Natural Sciences and Engineering Research Council (NSERC) of Canada. All computations were performed using WestGrid computing resources, which are funded in part by the Canada Foundation for Innovation, Alberta Innovation and Science, BC Advanced Education, and the participating research institutions (www.westgrid.ca).
This thesis is dedicated to my parents.
Chapter 1

Introduction

Protein complexes are among the many systems studied by mass spectrometry. The biological function of proteins is dependant on the interactions between them[2]. In order to understand protein complexes it is necessary to determine the number, type, and spatial arrangement of the subunits within the complex[2]. Under the right conditions, the non-covalent interactions in these complexes can be dissociated in collision cells within mass spectrometers without disrupting the covalent bonds[3]. Identifying the interactions between monomers can be done by comparing the composition of sets of subcomplexes generated from dissociating the intact complex[4]. Understanding the dissociation pathways could lead to methods for controlling the dissociation. Ultimately this would aid in gaining information about the structures of protein complexes.

1.1 Mass Spectrometry Experiments

Numerous studies have used mass spectrometry (MS) to look at the dissociation of dimers[5–8], tetramers[9–13], and larger complexes[12, 14–17]. In mass spectrometers, gas-phase ions are separated based on their mass to charge ratio (m/z). When an ion moves through an electric field the force on the ion is proportional to the charge (z) on the ion, thus the acceleration is proportional to z/m. In time of flight (TOF) mass spectrometers, after the acceleration the ions move through a drift tube with no external field. The time it takes ions to reach the detector depends on the mass to charge ratio. Tandem mass spectrometry (MS/MS), where the first MS selects the parent ion of interest and the second MS is used to probe the dissociation products of the parent ion, is often used in these experiments[18]. Collision or dissociation cells can be placed between the two mass analyzers of the tandem mass spectrometer. The charge states used range from +11-15 for the smaller dimers and tetramers to over +30 for the largest complexes.

Electrospray ionization (ESI) is a method for creating protein ions in the gas phase. These experiments are possible because one of the advantages of ESI is that it is gentle enough to preserve non-covalent interactions allowing large protein complexes to be studied using mass spectrometry techniques[19]. In ESI, the solution flows through a narrow needle and charged droplets are formed. They become smaller until the protein is all that remains.
There are two mechanisms by which the droplets become smaller. One is the evaporation of neutral solvent molecules. The other is Raleigh fission, when the charge density on the surface of the droplet increases the repulsion between the charges causes the droplet to split into two smaller droplets.

Energy must be added to get non-covalently bound protein complexes to dissociate. The main types of non-covalent bonds in these complexes are Van der Waals, dipole-dipole interactions, and hydrogen bonding. Two of the methods used for this are collisionally induced dissociation (CID) and surface induced dissociation (SID). In CID, the protein complex collides with a bath gas, and through these collisions energy is gradually transferred to the complex. In SID, the complex collides with a much more massive surface. Because in SID the complex collides with a surface, it has a much faster rate of energy transfer than CID where there are many collisions with lighter bath gas particles. Some of these dissociation experiments are discussed in more detail below.

Jurchen and Williams[7] studied the dissociation of cytochrome c dimer (see Fig. 1.1). When the monomers could unfold asymmetric dissociation is observed. The higher charge state (+19 dimer) dissociates symmetrically into monomers with +10 and +9 charges. The +11 dimer dissociates asymmetrically into monomers with +7 and +4 or +8 and +3 charges. However, when cross-linkages were added to prevent the monomers from unfolding, only symmetric dissociation is seen. Although mass spectrometry does not give direct information on the size of the molecules, this indicates that unfolding is involved in the asymmetric dissociation pathway. They studied the dissociation mechanism and determined that proton transfer plays a role. The charge on a protein is correlated to the surface area (not the mass)[20]. This means that charge accumulation (coming from the charge transfer) accompanies unfolding.

In solution, the mechanism for proton transfer in proteins involves water molecules as intermediaries[21]. But mobile protons have also been found in mass spectrometry experiments. There is experimental and theoretical confirmation that charges are mobile in small gas-phase peptides[22–28], and a mobile proton model has been developed[23, 29]. Peptide fragmenting D/H exchange experiments performed by Jorgensen et al.[27] were able to selectively label the C- or N-terminal ends of peptides with deuterium and found 100% scrambling in CID. Charge motion will be discussed in more detail in Chapters 2 and 4.

Jones et al. dissociated a dimer cytochrome c using both SID and CID [8]. They selected a +11 charge state prior to the dissociation. When CID was used, the products were mostly monomers with +8 charges and the complementary monomers with +3 charges. When SID was used, the results were more symmetric with most monomers having +6 or +5 charges.

Blackbody infrared radiative dissociation (BIRD) is where thermal energy is exchanged with the surroundings by blackbody radiation. The homodimer ecotin was studied by
Figure 1.1: SORI-CAD spectra of odd charge states of cytochrome c dimer ions electro-sprayed from a 50:50 water/methanol solution with 2% acetic acid: 19+, 15+, and 11+ dimer charge states. The 11+ dimer ions are produced by gas-phase deprotonation of higher charge state dimers. Reprinted with permission from [7]. Copyright 2003 American Chemical Society.
Felitsyn et al.[6]. They used the BIRD method with temperatures of 399 to 448 K to dissociate dimers with charge states +14-17. Both symmetric and asymmetric monomer pairs result from the dissociation.

BIRD has also been used to study the dissociation of the B₅ pentamer of the shiga-like toxin [14, 17]. Figure 1.2 shows the results of the dissociation. In all four examples, one monomer is ejected with almost half of the total charge. They also calculated Arrhenius prefactors and from that values for the entropy of activation. The large values of Arrhenius factors and the entropy of activation suggest that the transition state has a high entropy (likely from unfolding)[14].

When tetramer streptavidin in the +14 charge state was dissociated using sustained off-resonance irradiation (SORI) the products were monomers with +7 or +6 charges and the complementary trimers with +7 and +8 charges[9]. SORI is a thermal dissociation method.

The transthyretin tetramer has been studied by several groups. When the +15 charge state is dissociated using CID, the main products are monomers with +8 or +9 (some +7) charges and the corresponding trimers with +7 and +6 (some +8) charges[10]. When SID is used to dissociate the same complex the products are mostly monomers with +4 charges, and dimers and trimers are also seen[12].

The dissociation of +11-14 charge states of pentamer shiga-like toxin I was studied using the BIRD method[14]. The monomers formed had 30-50% of the total charge and the tetramers had the remainder of the charge.

A dodecamer of small heat shock protein TaHSP16.9 was studied[15]. The dodecamer had a distribution of charges states centered at +33. After CID, monomers and 11-mers were formed with charge state distributions around +14 and +20 respectively.

The general trends were that collision induced dissociation and thermal dissociation produced asymmetric charge separation involving a monomer leaving with close to half the charge (or more for the dimers), and surface induced dissociation resulted in subunits with symmetric mass-to-charge ratios. This is true for many different complexes of varying sizes and composition, thus any general model will be widely applicable.

Using mass spectrometry, protein complex dissociation producing asymmetric charge distributions has been observed in protein dimers[7] and in larger multimeric complexes[10, 11, 14, 15]. Asymmetric charge dissociation occurs when the mass to charge ratios of fragments differ, that is one of the fragments has a larger fraction of the total charge than it has of the total mass. With slow dissociation methods, such as CID, unfolding and ejection of one monomer from the complex is almost universally observed. Typically, the ejected monomer carries away up to half the total charge on the complex. In some cases, the remaining complexes again eject another charge-enriched monomer. With fast dissociation
Figure 1.2: BIRD mass spectra of protonated Stx1 $B_5^{6+}$ ions: (a) $B_5^{11+}$ at a cell temperature of 166 C and a reaction time of 12 s; (b) $B_5^{12+}$ at 165 C and 2.5 s; (c) $B_5^{14+}$ at 165 C and 1.5 s; (d) Stx1 $B_5^{14+}$ at 143 C and 3.5 s. Reprinted with kind permission from [17]. Copyright 2007 Springer Science and Business Media.
methods like SID, more symmetric dissociation products are seen[8, 11, 12]. That is, single monomer ejection does not dominate, and the charge distribution of the fragments is close to what is expected for a uniformly charged protein complex.

In practice, usually the weakest bound monomer in a complex unfolds and dissociates upon CID activation. The predominance of this dissociation pathway excludes other pathways which may shed more light on the binding of monomers in the complex, and allow one to build a better picture of their assembly. These monomer interactions are often the key to understanding the relationship between structure and biological function in these complexes. Understanding the detailed mechanism for monomer ejection will help to provide methods by which this dissociation can be controlled and possibly other dissociation pathways selected. Hopefully, this will allow for more structural information about proteins to be gathered from experiments.

The relationship between the gas-phase structures and the solution structures is not immediately obvious. It is clear that harsh conditions can cause structural changes. However, structure can be preserved[30]. A recent study[31] compared the gas-phase collision cross sections and the solution-phase Stokes radii for several different proteins and protein complexes. They found good agreement between the values, indicating that if structural changes occurred during the ionization process they were not significant enough to change the size of the proteins.

1.2 Computational Background

In addition to experimental work, some theoretical work has also been done to explore the dissociation mechanism.

Klassen and coworkers [17] used the charged droplet model and a protein structure model to calculate energies for a single subunit leaving a pentamer. The charge droplet model[32] represents the dividing droplet as two spheres. Their protein structure model used more detailed structural information, but assumed that the displacement of one monomer was the only change in the structure on dissociation. Their results show that protein unfolding may be important to the asymmetric charge distribution and a simple Coulomb energy calculation is useful in understanding the dissociation.

Konermann and coworkers [33] used a simple model where folded proteins were represented by spheres and unfolded proteins by a string of smaller spheres. The number of beads used for the unfolded proteins was varied to model partial as well as fully unfolded proteins. They found that dimers dissociating without unfolding had a symmetric charge distribution, while when one of the subunits unfolded the charge distribution was asymmetric. An asymmetric charge partitioning was also found when one unit of a tetramer
unfolded while the other three remained folded. Their study also included collision induced
dissociation mass spectrometry experiments of dimer and tetramer proteins. By comparing
the experimental spectra to the charge distributions predicted by the model, they concluded
that dissociation was occurring with a range structures having varying degrees of unfolding.

The work by Csiszar and Thachuk [34] used charged ellipsoids to model surface charges.
They found that a constant surface charge density was the most favorable distribution. So
for a dimer with an asymmetric charge partitioning, the monomer with more charges would
want to unfold to have a surface area larger than that of the monomer with fewer charges.

All-atom studies were performed by Wanasundara and Thachuk [1, 35, 36] on the dissoci-
ation of cytochrome c’ dimers using the OPLS force field. Their work showed the symmetric
case with +5 charges on each monomer had a lower barrier to dissociation than the neutral
dimer, and the asymmetric cases (M8/M2 and M9/M1) had higher barriers. The higher
charged monomers in the asymmetric cases unfolded while the lower charged monomers
remained folded.

The essential conclusion of these studies[1, 17, 33–36] is that the monomer ejection
process is controlled by Coulomb repulsion. In particular, charges arrange themselves so
as to maintain a constant surface charge density and the lowest barrier for dissociation
occurs when two fragments have the same charge. Almost all experimental results can be
rationalized with this Coulomb repulsion model.

Lill and Helms created a method (Q-HOP MD) for including proton transport in molecu-
lar dynamics (MD) simulations by modeling hopping rates using transition state theory[37].
They used multicopy methods to increase the sampling in regions of interest. Donnini et
al.[38] performed simulations where the protonation states were changed in order to main-
tain constant pH.

1.3 Coarse Grained Models

All-atom force fields include parameters for every atom including hydrogen. In contrast,
coarse-grained force fields group atoms together and represent each group as a bead. Tozzini
reviewed coarse-grained models as used for molecular dynamics of proteins [39]. Although
coarse-grained models contain less detail than the all-atom ones, coarse-grained simulations
have the advantage of being faster. One reason that they are faster is that coarse-grained
systems have fewer particles. This means there are fewer calculations to perform for each
step, particularly for pairwise contributions to the energy. Also, the coarse-grained models
have removed the light atoms (like hydrogen), this removes some of the fast motions and
allows for a larger time step to be used. Thus fewer steps are needed to reach a particular
simulation time compared to when a smaller time step is used.
By reducing the number of particles in the simulation and removing the light atoms which have fast vibrations, coarse-grained models are computationally faster than all-atom models. One type of coarse-graining is the united atom model (for example the Amber united atom model [40] and the OPLS united atom model [41]). These have the hydrogen atoms included with the heavy atom they are bonded to, but each heavy atom is represented by a single interaction site. They are faster than all-atom models but not as fast as some coarser models. At the other end of the coarseness scale are models that use one or two beads for each amino acid. These models like the Go model [42, 43] are usually parameterized to reproduce the correct folding for a particular protein and are difficult to transfer to proteins with different structures.

1.4 Goals

We are hoping to find ways in which any one (or more) of the monomers could be selectively unfolded and/or dissociated from the rest of the complex. This might then allow for the relative binding strengths to be determined providing information about the structure of the complex. The goal of this is to provide some qualitative insights, not quantitative information.

The goals of this work are to use coarse-grained force fields and charge moving. The coarse-graining has two main advantages, it reduces the computational time and it makes the charge hopping easier (because there are no partial charges and moving the charge does not require changing the bonding connectivity). Using coarse-graining will increase the simulation time scales that can be reached for a given amount of computational time. This is important because while bond vibrations take place on femtosecond timescales, protein motions that involve higher order structure can take from picoseconds to minutes[30]. Also, protein complexes have many possible conformations, so good sampling is required.

Simulations of proton transfer in large protein systems are rare, because fixed charge simulations are much easier than charge moving. However, we want to include charge moving in order to better understand the role of mobile protons in the protein dynamics.
Chapter 2

Methods

2.1 Molecular Dynamics

Molecular dynamics (MD) studies the motion of molecules over time. Using a potential energy function, the forces on each atom are calculated. From these forces the velocities can be calculated using Newton’s law and the positions at a later time can be determined. This is repeated for many time steps. One benefit of molecular dynamics is that it can show fluctuations not just a static picture. Time dependent properties can be studied, or averages over long times can be used to calculate properties using statistical mechanics tools.

In classical MD, Newton’s equation of motion must be solved for each particle, that is

\[ F = m \frac{dv}{dt} = m \frac{d^2r}{dt^2}, \tag{2.1} \]

where \( F \) is the force, \( m \) is the mass, \( v \) is the velocity, \( r \) is the position, and \( t \) is time.

The Verlet algorithm is derived from Taylor expansions of the particle’s coordinates[44]. The new position of the particle is

\[ r(t + \Delta t) \approx 2r(t) - r(t - \Delta t) + \frac{f(t)}{m} \Delta t^2, \tag{2.2} \]

and the velocity is

\[ v(t) \approx \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t}. \tag{2.3} \]

From the Verlet algorithm the equivalent Leap Frog algorithm can be obtained by defining the velocities at half integer steps as

\[ v(t - \frac{\Delta t}{2}) \equiv \frac{r(t) - r(t - \Delta t)}{\Delta t}, \tag{2.4} \]

and

\[ v(t + \frac{\Delta t}{2}) \equiv \frac{r(t + \Delta t) - r(t)}{\Delta t}. \tag{2.5} \]

Rearranging Eq. (2.5) gives the new position as

\[ r(t + \Delta t) = r(t) + v(t + \frac{\Delta t}{2}) \Delta t. \tag{2.6} \]
Subtracting Eq. (2.4) from Eq. (2.5) and using Eq. (2.2) gives the velocity at time \( t + \Delta t \) as
\[
v(t + \Delta t) = v(t - \Delta t) + \frac{f(t)}{m} \Delta t.
\] (2.7)

The sets of functions and parameters used to calculate the forces acting on an atom (or coarse-grained bead) are called force fields. The equations for the force fields describe the potential energy for different types of interactions between atoms, and the parameters are used to tailor the calculations for specific atoms. The force on an atom is the derivative of the potential energy. Many force fields use harmonic oscillator forms for bonding interactions, that is
\[
V(r) = \frac{1}{2} k (r - r_0)^2,
\] (2.8)
where \( V(r) \) is the potential, \( k \) is the force constant for the bond stretching, \( r \) is the bond length, and \( r_0 \) is the equilibrium bond length. Lennard-Jones (LJ) potentials are often used for non-bonding (long range) interactions. They have both repulsive and attractive terms and produce a well shaped potential using the functional form
\[
V_{LJ}(r) = 4 \varepsilon_{ij} \left( \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right),
\] (2.9)
where \( \varepsilon_{ij} \) is the well depth, \( \sigma \) is the cross section diameter, and \( r \) is the distance between two atoms. The value of \( \varepsilon_{ij} \) depends on the interacting particle types and in the MARTINI force field it has values between 2.0 kJ/mol and 5.6 kJ/mol (higher values for more strongly interacting beads)[45]. The Coulomb potential is used for charges, that is
\[
V_{Coulomb}(r) = \frac{q_i q_j}{4 \pi \varepsilon_0 \varepsilon_r r},
\] (2.10)
where \( r \) is the distance separating two atoms, \( q_i \) is the charge on atom \( i \), \( \varepsilon_0 \) is the dielectric constant for a vacuum, and \( \varepsilon_r \) is the relative dielectric constant. The forces are calculated for every pair of atoms and added to determine the total force on each atom.

Energy minimization is often performed before molecular dynamics in order to remove any situations where atoms/beads are too close together. These bad contacts produce very large forces which can make the molecular dynamics trajectories unstable. Because the system is able to relax during molecular dynamics steps, reaching the absolute minimum is not necessary as long as there are no bad contacts in the initial structure for the MD run. Energy minimization is a process for finding a low energy structure by taking the potential energy function and performing a mathematical minimization with respect to the atom positions. These minimizations tend to find a local minimum not the global minimum, and rarely move far from the initial conformation[46]. There are three different
energy minimization algorithms which GROMACS can perform: steepest descent, conjugate gradient, and limited-memory BFGS[47]. Steepest descent is the simplest method. It moves atoms in the direction of the force with step-sizes which change at each iteration. The conjugate gradient method uses information from the previous step to adjust the change in the current iteration. L-BFGS is a quasi-Newton method[48].

For many large systems such as protein complexes, there is not a single correct configuration. Rather an ensemble of low energy configurations all contribute to the results measured. Different trajectories can be started from the same geometry and input parameters, but with different initial velocities. It is important to average the results from many trajectories in order to get proper sampling, because a single trajectory may become trapped in a local minimum and miss configurations that are important.

2.2 Input Parameters

Within the molecular dynamics software there are a number of parameters the user selects to control the simulation. The important ones discussed here are time step, electrostatics, thermostats, and constraints.

The distance an atom or bead moves in a single step is the product of the velocity and the time step. If the time step is too large, the particles move large distances which can produce large forces and make the dynamics unstable. However, the smaller the time step the more steps are needed to reach a particular simulation time. Thus the general preference is to use the largest time step for which molecular dynamics trajectories are stable. For all-atom simulations this is usually around 1 fs, but for coarse-grained simulations the time step can go much larger (20-40 fs in some cases).

Because all the systems studied here are isolated gas-phase molecules without solvent, no periodic boundary conditions were used. The absence of solvent also means that charges and intermolecular interactions are unattenuated, so for the Lennard-Jones and Coulomb electrostatics no cutoffs or shifts were employed. Thus, the electrostatic energy calculations are performed for all pairs of beads. The default value of the relative dielectric constant used with MARTINI accounts in part for the effect of solvent. In this study, charges residing on dissociating fragments act through empty space, for which a relative dielectric constant of unity is physically appropriate. Higher values are used in solution simulations to account for the screening effect of the solvent. In principle, charges within a monomer interact through the protein itself and would require a relative dielectric constant different from unity. However, as the protein conformation changes, the relative balance between through-space and through-protein interactions changes in a dynamic manner. Because the through-space contributions are most important for describing correctly the dissociation
process, the relative dielectric constant was set to unity for all simulations.

The simulation temperature is calculated from the kinetic energies of all the particles in the system using

\[ T = \frac{2E_{\text{kin}}}{k_B f}, \]  

where \( E_{\text{kin}} \) is the average kinetic energy, \( k_B \) is the Boltzmann constant, and \( f \) is the number of degrees of freedom. For the isolated molecule both center of mass and angular motion are removed (of the whole system), so the number of degrees of freedom is \( 3N - 6 \) (\( N \) is the number of beads). Thermostats modify the velocities of the particles to maintain a constant temperature. The simplest type of thermostat is velocity rescaling where at each time step each particle has its velocity multiplied by the factor \( (T_{\text{desired}}/T)^{1/2} \) to produce the desired temperature. However, this does not account for the fluctuations present in a physical system. The two most common thermostats are the Andersen and Nose-Hoover thermostats[44]. The Nosé-Hoover thermostat couples the system to a heat bath so that the temperature can fluctuate around the desired temperature[49–51]. This is done by extending the system with an additional coordinate \( s \), which provides a constraint on the system so that the average temperature remains constant. The coupling constant controls the time period of temperature fluctuations (usually on the order of 10 times the time step). The temperature in all of these simulations was controlled by a Nosé-Hoover thermostat using a coupling time constant of 0.3 ps.

Constraints are used to keep particular bond lengths fixed. After the positions of the atoms are updated in an unconstrained move, a constraint algorithm is used to calculate the additional force needed to keep the bond lengths at the values imposed by the constraints. The MARTINI force field adaptation for proteins uses constraints for the ring structures in some amino acid sidechains to prevent them from flipping/inverting in ways that the real (or all-atom) sidechains would not. Lagrange multipliers are often used to implement constraints. However, the resulting second order equations can be difficult to solve, thus approximations are used to make the problem linear.

The method used in this work is the LINCS (Linear Constraint Solver for Molecular Simulations)[52]. There is a set of equations \( g_i(r) = 0 \), one for each of the \( i \) constraints in the system. LINCS uses matrices for the equations. Newton’s equation with the Lagrange multipliers in the LINCS approach of adding the multipliers to the potential, \( V(r) \), is

\[ -M \frac{d^2 r}{dt^2} = \frac{\partial}{\partial r} (V - \lambda \cdot g), \]  

(2.12)

where \( M \) is the matrix of the masses of all the particles, and \( \lambda \) is the matrix of Lagrange multipliers (one for each constraint equation). When calculating matrix inverses power expansions of the matrix are used. Increasing the order of these expansions improves the
accuracy. The default lincs_order in GROMACS is 4, but for use with MARTINI in the
gas-phase I found that higher orders produced both better energy minimization and more
stable trajectories, and would recommend using lincs_order of 12.

SHAKE[53], another commonly used constraint algorithm, uses an iterative method.
The constraint equations are solved one at a time (without matrices), but solving the later
constraints may move atoms involved in earlier constraints. So, the procedure is repeated
until all the constraints are satisfied.

Elastic network bonds[54] can be used to connect selected backbone beads with harmonic
potentials with a force constant of 2500 kJ/(mol nm); a value much larger than those for
regular bonds. These are sometimes used for coarse-grained MD (not in atomistic MD)
because the bonds between backbone atoms which would keep the secondary structure in
place are not included. These network bonds are usually located in parts of the protein
where there are four or more amino acids which have an extended secondary structure (they
are not part of a helix or beta-sheet). Topologies can be created with or without elastic
network bonds. These bonds, when made between flexible parts of a protein, increase
the stability of the protein thereby allowing a larger time step to be used. Combining
elastic network bonds with the MARTINI force field allows for faster simulations than the
MARTINI force field alone. The effect of these bonds on the results will be considered in
Chapter 3 and all the simulations of later chapters include local elastic network bonds to
stabilize flexible parts of the structure.

2.3 Properties for Analysis of Trajectories

Trajectories are analyzed using energy and structural properties. The structural properties
calculated for this work are discussed below. Several different energy quantities can be
calculated from the positions and velocities including potential, Coulomb, Lennard-Jones,
and kinetic energies. The simulation temperature is derived from the average kinetic energy
of all the beads. The bond, angle, and dihedral angle contributions to the total energy are
very small compared to the Lennard-Jones and Coulomb energies, so they are not discussed.
All energies are reported in kJ/mol, and all distances in nm.

The radius of gyration ($R_g$),

$$R_g = \left( \frac{\sum_i r_i^2 m_i}{\sum_i m_i} \right)^{\frac{1}{2}},$$  \hspace{1cm} (2.13)

where $m_i$ and $r_i$ are the mass and position of atom $i$, is used to indicate the size of a protein.
A small compact protein has a small radius of gyration. If the protein becomes larger or
the shape changes to a more elongated shape the radius of gyration will increase. The more
a protein expands or unfolds the larger the radius of gyration becomes.

Another parameter that can be used to measure changes in the proteins is the root mean square deviation (RMSD),

\[
RMSD = \left( \frac{1}{M} \sum_i m_i (r_i - r_i^{(reference)})^2 \right)^{\frac{1}{2}}
\]

where \( M \) is the total mass, and \( m_i \) and \( r_i \) are the mass and position of atom \( i \). RMSD measures the difference between the positions of atoms in the protein and a reference structure.

Another property that is measured is the minimum distance. This is the smallest distance between the atoms on one monomer and atoms on the other monomer. This minimum separation was calculated using the GROMACS g_mindist utility. When the minimum separation is small the monomers have not dissociated, but after dissociation the minimum separation increases. Maximum distances between different beads or groups can be measured as a way of monitoring unraveling in the complex.

### 2.4 Software and Force Field Particular to This Work

GROMACS[55–57] is a readily available software package that can be used to perform molecular dynamics simulations. All of the molecular dynamics simulations in this work were performed using GROMACS. GROMACS has the capability for parallel computing, but here it was only used in serial (one processor per trajectory). Because the systems studied here are treated as single molecules, the distributed nature of the parallel version of GROMACS is not useful. Trying to divide a single molecule into different domains does not result in any computational savings. The hopping algorithm has been coded to work with the serial calculations, but there is no reason why it could not be adapted for parallel computing. GROMACS is available as open source software, so the code can be modified. It includes pull code that can be used to add a force to the system pulling groups of atoms apart.

To run the simulations, three kinds of input are needed. First is a starting geometry (positions and names of all the atoms). Second is the topology which describes the bonding and connectivity of the atoms. This topology includes all the parameters needed to calculate the potential energy such as bond force constants and Lennard-Jones parameters. The final input contains the molecular dynamics parameters (time step, number of steps, desired temperature, etc.). GROMACS has a preprocessing program (grompp) which turns all the input files into a single file that is formatted for the main program (mdrun) to use.

Figure 2.1 shows a skeletal outline of how the molecular dynamics part of the GROMACS program is organized (not every part of the code is shown, but it gives an overview of how
it works). The main program calls mdrunner, which initializes some of the data structures and then makes a call to the integrator. do_md is the integrator for molecular dynamics. There are others for stochastic and Brownian dynamics and for energy minimization (even though energy minimization is not an integration algorithm it takes the same place in the code). Within the do_md function are first some steps to set things up, and then a loop over the molecular dynamic time steps. During this loop, the forces are calculated and the velocities and coordinates of each bead are updated for each step.

To add charge hopping, I modified GROMACS by adding a new subroutine within the molecular dynamics loop. The position of this subroutine relative to the rest of the program is shown in Fig. 2.1 where the new function is highlighted in pink. Also, several new parameters were added to the molecular dynamics input (in the .mdp file) to control the hopping routine. These are mobile_charge (yes or no), mobile_cutoff, mobile_prob, mobile_freq, and mobile_seed (their use is discussed more in the next section).

The particular force field used in this work is MARTINI[45, 58]. It is a coarse-grained force field in which each bead represents four heavy atoms plus the hydrogens bonded to them. It was originally created to model lipids, but has been extended to proteins. Each amino acid is represented by between one and four beads. It has been parameterized by comparing the results of simulations to thermodynamic properties such as free energies of hydration, vaporization, and water/organic solvent partitioning.

Baron and coworkers[59–61] compared thermodynamic properties from all-atom and coarse-grained simulations. The coarse-grained model they used was an older version of MARTINI. They found the coarse-grained systems had a lower entropy than the corresponding all-atom systems, but the major structural features were similar. Monticelli et al. then extended the MARTINI force field to include proteins[45]. It was successfully used in several studies[62–65]. Simulations of a series of peptides at an interface were used to calculate partitioning free energies which were compared to experimental data[66]. Milani and coworkers used a modified MARTINI force field to study isolated nanofibers[67]. Their results were consistent with experimental results.

The MARTINI website (http://md.chem.rug.nl/cgmartini/) has files containing the parameters of the force field which are available for downloading. They also have scripts to convert a Protein Data Bank format structure to the coarse grained structure (atom2cg) and to create a MARTINI topology file (seq2itp). The seq2itp script places positive charges on all the basic amino acid residues and negative charges on all the acidic amino acid residues. For this work, I modified seq2itp so that the default is for all the amino acid residues to be neutral. Positive charges were then added to selected basic amino acid residues to create the desired charge distribution. No partial charges are used, and the beads charged by the seq2itp script are the MARTINI Qd type. In every positively charged monomer the charges
Figure 2.1: A skeletal diagram of the GROMACS molecular dynamics program.
are distributed among the basic amino acid side chains (arginine, lysine and histidine) and the N-terminus.

Two protein complexes have been studied here, cytochrome c' and transthyretin. Cytochrome c' (Protein Data Bank (PDB) ID: 1bbh)[68] is a homodimer with a total of 262 amino acid residues of which 30 are basic. The dimer was used in simulations without charge hopping. The results of this are discussed in Chapter 3 where they are compared to previous all-atom simulations[1, 35, 36]. Transthyretin (PDB ID: 3grg)[69] is a homotetramer with a total of 460 amino acid residues of which 56 are basic. The tetramer was used in the charge hopping study (results discussed in Chapters 4 and 5). The monomers within the tetramer are labeled as chains A, B, C, and D, and the same number of charges is placed on each monomer when the initial charge configurations are being generated.

For each given charge partitioning and configuration, charges were added to the ground state crystallographic PDB structure. The structures were first minimized using the steepest descent method, and then initialized with velocities randomly chosen consistent with the temperature. Short molecular dynamics simulations were performed to relax the structures to the equilibrium state. These relaxed structures were used as initial conditions for the simulations from which data were collected. For each case (different temperatures or parameters), 10 trajectories were run with each of 10 different charge configurations for a total of 100 trajectories. The results reported are the averages of the sets of 100 trajectories unless specified otherwise.

2.5 Moving Charges

The best description of charge hopping uses quantum mechanics. However, that is difficult and computationally expensive, so a simpler method is developed here. The scheme for the charge moving algorithm I created is shown in Fig. 2.2 and the source code is in Appendix A.

Overall, a charge hopping algorithm must have sites at which charges reside and criteria by which charges are moved. Since most mass spectrometry studies of protein complexes use electrospray ionization in positive ion mode, positive charges are sufficient for the current study. Thus, a positive charge can hop from one basic site to an uncharged basic site. All other amino acid residues remain neutral.

Two criteria for moving charges are employed. The first is a simple distance cutoff. If the distance between the charge donor and charge acceptor sites is less than \( r_{\text{cutoff}} \) then a hop is possible, and the energy will be considered. If the distance is larger than the cutoff then the move will not happen. The second criteria depends upon energy and uses a Monte Carlo like scheme to determine hopping probability. This change in energy for a hop
Find charged sites (donors) and uncharged basic sites (acceptors)

For each donor

- Find nearest acceptor.
- Is nearest acceptor closer than cutoff distance?
  - Yes
  - No
  - Add donor / acceptor pair to list of possible moves.

For each possible move

- Calculate change in Coulomb + binding energy.
- Is ΔE less than zero?
  - Yes
  - Hop
  - No
  - Calculate probability of hop (p) and a random number between 0 and 1.
  - Is random number less than p?
    - Yes
      - Hop
    - No
      - Don’t hop

Propagate trajectory

Figure 2.2: A flowchart showing the hopping algorithm.
is a combination of the difference in Coulomb energy between the final and initial charge
configurations and the difference in hydrogen binding energies for the acceptor and donor
residue types.

In MARTINI, no partial charges are used, so it is easy to change the charges on the
side chain beads of the amino acids when a hop occurs, since this involves simply increasing
or decreasing the total charge on a single bead. In particular, no structural changes are
required to change the charge state, so moving charges does not create any geometrical
problems, like bad contacts. More technical details of the algorithm are given below.

To organize sites, two lists are created. One is a list of all the sites containing charges.
These sites act as charge donors for the hops. The other is a list of the uncharged basic
sites which are available as charge acceptors.

The hopping algorithm examines the structure of the complex for possible hops before
the MD algorithm propagates it to the next time step. This is seen in the first black box
of Fig. 2.2. First, for each element in the list of charged sites, the empty basic site nearest
to it is found. If the distance to that site is less than $r_{\text{cutoff}}$, the pair is added to a list
of possible hops. The value of $r_{\text{cutoff}}$ was set to 0.57 nm. This value was chosen because
it is the diameter of a MARTINI coarse-grained bead plus 1 Å. In practice, the value of
$r_{\text{cutoff}}$ tunes the hopping rate. The larger $r_{\text{cutoff}}$, the greater the number of donor/acceptor
encounters and hence the greater probability a charge hop will occur. However, having too
large a value of $r_{\text{cutoff}}$ simulates proton transfer over larger distances, which at some point
is hard to physically justify. A value was chosen small enough to consider the transfer to
be localized but not so small as to decrease too severely the hopping rate. Results were
also calculated for $r_{\text{cutoff}} = 2.0$ nm and while the hopping rates did vary (as alluded to
above) the behavior of the protein structure did not. In other words, the results are not
particularly sensitive to the value of $r_{\text{cutoff}}$. In an atomistic model the angles between the
proton donor and acceptor would affect the probability of proton transfer. However, in the
coarse-grained model this structural detail is lost, so the angles between charge donor and
acceptor sites are not considered.

Second, for each pair of possible hops, the change in energy that would result from
the hop is calculated. This part of the hopping decision is seen in the second black box
of Fig. 2.2. The Coulomb and proton binding energies are considered for this calculation,
but not the Lennard-Jones energy because during the hop the coarse-grained beads do
not move. The proton binding energy is the difference in energy between the neutral and
protonated forms of the amino acid. Because this energy is not included in the MARTINI
force field it must be explicitly included in the energy calculation. Table 2.1 lists the values
of the proton binding energies for the basic sites used in this study. Ab initio calculations
of the energies of bare protonated and non-protonated amino acids were used to calculate
Table 2.1: Proton binding energy values for basic amino acid sites[1].

<table>
<thead>
<tr>
<th>Site</th>
<th>$\Delta E_{binding}$ (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arginine</td>
<td>-1028.8</td>
</tr>
<tr>
<td>Histidine</td>
<td>-957.7</td>
</tr>
<tr>
<td>Lysine</td>
<td>-937.2</td>
</tr>
<tr>
<td>N-terminus</td>
<td>-959.8</td>
</tr>
</tbody>
</table>

The binding energies[1]. These calculations do not consider the environment of the amino acid, but in the energy calculation the electrostatic environment is accounted for by the Coulomb energy contribution. The change in energy is calculated using

$$\Delta E = E_{final} - E_{initial}. \quad (2.15)$$

The energy of a particular configuration is

$$E = \sum_{i=1}^{N} E_{binding,i} + \sum_{j>i} \frac{e^2}{4\pi\epsilon r_{ij}}, \quad (2.16)$$

where $E_{binding,i}$ is the proton binding energy of the charged site $i$, $\epsilon$ is the dielectric constant in a vacuum, and $r_{ij}$ is the distance between the centers of charged sites $i$ and $j$. A relative dielectric constant of 1 is used in these calculations. The value of $E_{initial}$ is the energy $E$ of the current configuration while that of $E_{final}$ is the energy of the configuration that would result if the charge were moved from the donor to the acceptor site. All possible hopping pairs are sorted by their $\Delta E$ values from lowest to highest.

Starting with the one of lowest $\Delta E$, each possible hop is considered in turn. If $\Delta E \leq 0$, the charge is moved from the donor to the acceptor. If $\Delta E > 0$ then the probability of a hop occurring, $p$, is calculated using the Monte Carlo equation[44]

$$p = \exp \left[ -\frac{\alpha \Delta E}{kT} \right], \quad (2.17)$$

where $\alpha$ is an arbitrary scaling factor, $k$ is the Boltzmann constant, and $T$ is the temperature. The factor $\alpha$ is an adjustable parameter that allows the hopping rate to be increased or decreased without changing the temperature for the dynamics of the protein. As will be shown below, it will become necessary to increase the hopping rates in this way in order to make the timescales for charge hopping events smaller. At the same time, one does not wish to increase the hopping rates by raising the temperature greatly, since this will directly affect protein dynamics. By adjusting $\alpha$, the hopping rates can be manipulated indepen-
dently of the temperature. The value of $p$ is compared to a random number between 0 and 1. Only if the random number is less than $p$ is the charge moved from donor to acceptor.

When a hop occurs, the charge in the GROMACS mdatoms record for the donor site is decreased by one and the charge for the acceptor site is increased by one. The bead indexes in the lists of charged and uncharged basic sites are then swapped. The bead which had a charge of $+1$ before the hop is given a charge of 0 and is moved to the list of uncharged basic sites, and what had been the corresponding uncharged basic site is given a charge of $+1$ and is put in the list of charged sites. In the MARTINI force field, charged beads have different Lennard-Jones parameters than uncharged beads. However, in the implementation of the algorithm used in this study this difference was not taken into account, so the bead types were left unchanged. The total charge state of the system does not change. After each hop the values of $\Delta E$ are recalculated for all the remaining possible hopping pairs and the hopping pair list is resorted.

Five options have been added to the GROMACS mdp parameter file. To select moving charges the parameter mobile_charge is set to yes (the default is no). The largest distance between beads over which a charge can move is $r_{cutoff}$ (called mobile_cutoff in the mdp file and given a default value of 0.57 nm). The mobile_prob parameter is the scaling factor $\alpha$ (the default value is 1.0). The seed for the random number generator is set using mobile_seed (the default value of -1 uses time-of-day and process id to generate the seed).

If the mobile charge algorithm is executed at every time step, one would expect to see a lot of charges flip-flopping between two basic sites, but this is not physically correct. Therefore the protein, 10 MD steps are performed after each execution of the mobile charge algorithm to allow the protein to adapt to the charge configuration before more hops are considered. The mobile_freq parameter is used to control the number of MD steps performed between each hopping event.

The algorithm generates two output files (in addition to the standard GROMACS output). One lists the charged sites and the other gives information about all the hops considered each time the algorithm is executed. The time, temperature, number of hops considered, and the number of successful hops are given. As well, the charged donor site index, uncharged acceptor site index, $\Delta E$, $p$, and whether the hop was successful are listed for each move considered. The computational overhead of this algorithm is small (discussed in Chapter 4).
Chapter 3

Suitability of the MARTINI Force Field For Use With Gas-Phase Protein Complexes

3.1 Introduction

Since the MARTINI force field was parameterized with molecules in solution it is necessary to validate the force field for use in the gas-phase. The goal of this chapter is to use the MARTINI force field to replicate the all-atom studies performed with the cytochrome c′ dimer[1, 35, 36], compare the results, and thereby assess MARTINI’s utility for further gas-phase simulations.

Experimental evidence for protein unfolding has already been presented by Williams and coworkers [7]. The difficulty with all-atom simulations is that they are not practical for simulating long times or large systems. We would like to simulate protein complexes larger than dimers, and be able to include charge hopping.

3.2 Method

The system studied here is the cytochrome c′ dimer. Three partitionings of charges between the two monomers were considered. The first are neutral complexes with no charges. The second have a M5/M5 charge partitioning, being symmetric with five charges on each monomer for a total charge of +10, and the third have a M8/M2 charge partitioning, being asymmetric with eight charges on one monomer and two charges on the other for a total charge of +10. Figure 3.1 shows the positions of the basic amino acids available for the positive charges. All the charges are fixed. The same ten charge distributions are used for each of the charge partitionings as were used by Wanasundara and Thachuk [35]. They were chosen as the lowest energy configurations based upon short molecular dynamics relaxations. For details of the screening method see Ref. 35. The residue numbers of the charged amino acids are listed in Table 3.1 for each of the 20 charge distributions (residue 1 is the N-terminus).
Figure 3.1: Snapshot of a ground state monomer showing the available basic amino acid residues (and the N-terminus) in black, and all other amino acids in green.
Table 3.1: The residue numbers of the charged amino acids on each monomer for the ten M8/M2 and ten M5/M5 charge distributions.

<table>
<thead>
<tr>
<th></th>
<th>M8</th>
<th>M2</th>
<th>M5</th>
<th>M5</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1</td>
<td>1 25 27 70 72 74 89 111</td>
<td>64 74</td>
<td>25 64 72 89 111</td>
<td>25 64 72 89 111</td>
</tr>
<tr>
<td>c2</td>
<td>1 25 27 70 72 74 89 111</td>
<td>72 89</td>
<td>25 64 72 89 111</td>
<td>25 64 72 89 111</td>
</tr>
<tr>
<td>c3</td>
<td>1 25 27 70 72 74 89 111</td>
<td>74 89</td>
<td>25 64 72 89 111</td>
<td>25 64 72 89 111</td>
</tr>
<tr>
<td>c4</td>
<td>1 25 27 70 72 74 89 111</td>
<td>74 127</td>
<td>25 64 72 89 111</td>
<td>25 64 72 89 111</td>
</tr>
<tr>
<td>c5</td>
<td>1 25 27 70 72 74 89 111</td>
<td>89 127</td>
<td>25 64 72 89 111</td>
<td>25 64 72 89 111</td>
</tr>
<tr>
<td>c6</td>
<td>1 25 27 70 72 74 89 111</td>
<td>89 131</td>
<td>25 64 72 89 111</td>
<td>25 64 72 89 111</td>
</tr>
<tr>
<td>c7</td>
<td>1 25 27 64 74 89 111 129</td>
<td>89 131</td>
<td>25 64 72 89 111</td>
<td>25 64 72 89 111</td>
</tr>
<tr>
<td>c8</td>
<td>1 25 27 64 70 72 89 111</td>
<td>72 86</td>
<td>25 64 72 89 111</td>
<td>25 64 72 89 111</td>
</tr>
<tr>
<td>c9</td>
<td>1 25 27 64 70 74 89 111</td>
<td>89 129</td>
<td>25 64 72 89 111</td>
<td>25 64 72 89 111</td>
</tr>
<tr>
<td>c10</td>
<td>1 25 64 70 72 74 89 111</td>
<td>72 86</td>
<td>25 64 72 89 111</td>
<td>25 64 72 89 111</td>
</tr>
</tbody>
</table>
The reaction coordinate used for the dissociation of the dimer is the distance between the centers of mass of each of the monomers. The center of mass (COM) distance for the ground state dimer in the MARTINI force field is 1.4 nm and in the all-atom study, which employed the OPLS-AA/L force field, the ground state COM distance was 2 nm [35].

In addition to regularly bonded structures, trajectories were also propagated for these same structures supplemented with local elastic network bonds. The effect of these bonds on the results will be considered.

In the present case, 20 local elastic network bonds were created in the regions of the monomer with extended structure. The atoms involved in the elastic network bonds are shown in Fig. 3.2. Each monomer has two regions containing elastic network bonds but each such bond connects atoms only within its region. That is, no elastic network bonds connect atoms between the two regions. For complexes with elastic network bonds it was possible to use a time step of 30 fs while the same time step caused instabilities when the elastic bond network was removed. In the latter case, trajectories would propagate without incident for short times but for longer times would reach a point where their energies and temperatures would start to grow uncontrollably. This was traced to a fast motion occurring within the extended structure in which the elastic network was removed. At some point, an event occurs whereby two beads approach too closely (since the time step of 30 fs is too large to correctly account for the fast motion), a bad contact occurs and the energy increases too rapidly for the thermostat to compensate. This event is rare but for long simulation times it would occur in approximately 30% of the trajectories. Thus, elastic networks effectively remove fast motions arising in the extended structure regions. It is possible these fast motions are a difficulty only for gas-phase calculations where no solvent is present to damp them. To avoid the problem the time step was reduced to 2 fs, which decreased significantly the computational speed of the simulations when no elastic bond networks were included.

Three different types of trajectories were propagated - continuous pulling, stepwise pulling, and umbrella sampling. Then results of all three methods can be compared. The temperature was kept constant at 300K for all trajectories.

GROMACS parameters were set to produce simulations with continuous pulling. That is, the SHAKE algorithm [53] was used to constrain the center of mass distance to a 0.000001 nm tolerance, and this distance was increased by a constant amount at each time step. The resulting pull force needed to maintain each center of mass distance was also collected for later analysis. Continuous pulling simulations were run for a total time of 660 ns using a pull rate of 0.00001 nm/ps. The pull forces are more sensitive than structural parameters and it was necessary to investigate their convergence, as detailed below.

The convergence of the pull force results was gauged by comparing the predictions from a number of different methods. More specifically, for the M5/M5 charge distribution, three
Figure 3.2: Snapshots of the ground state dimer with one monomer shown in yellow and the other in green. (A) From the all-atom structure. (B) A stick model of the coarse-grained structure showing the atoms involved in elastic network bonds in black.
simulations were performed to calculate the pull force. The pull force is the amount of force needed to keep the monomer at a particular COM distance. The integral of the pull force gives the potential of mean force. The first case used the constraint forces from the continuous pulling simulations, as detailed above. Simulations with a faster pull rate of 0.001 nm/ps were also performed for comparison purposes.

The second case used a stepwise pull method. That is, starting from the ground state, the COM distance was increased in steps of 0.15 nm by pulling at a rate of 0.01 nm/ps for 15 ps. Each such pull was followed by a 600 ps simulation during which the center of mass was constrained at a fixed value using the SHAKE algorithm with a 0.000001 nm tolerance. The resulting constraint forces from the last 400 ps (the first 200 ps was allowed for relaxation to occur) of these simulations were time averaged to produce a single force value at that COM distance. This was repeated for 44 steps until the COM distance reached 7.88 nm. In addition to the pull force, the radius of gyration and the RMSD were calculated for the stepwise pull. This method was chosen because the all-atom studies [1, 35, 36] used a stepwise pull method, although the continuous pull is closer to the dissociation an experimental protein experiences.

The third case used simulations employing umbrella sampling [70]. An umbrella potential force constant of 3300 kJ/(mol nm$^2$) was used for simulations at a series of COM distances spanning from the ground state dimer distance to 8.4 nm. Each simulation was run for 1800 ps with the last 1600 ps being used for data collection. The resulting histograms were analyzed with the WHAM algorithm as implemented in the GROMACS g_wham utility program [71]. The pull force was then calculated by taking the derivative of the WHAM-generated potential of mean force curve as a function of COM distance.

### 3.3 Results and Discussion

The pull force for the M5/M5 charge partitioning, shown in Fig. 3.3, was calculated by three different methods in order to gauge its convergence. All three methods give the same values for COM distances larger than 7 nm where the pull forces attain a slightly negative value as a result of the Coulomb repulsion between the separated monomers. However, differences appear at smaller COM distances, especially for COM distances between 2 and 3 nm. Here the continuous pull method with the faster pull rate of 0.001 nm/ps produced the largest pull forces while the umbrella method produced the smallest ones. This indicates there is likely some irreversible work present in this continuous pull calculation possibly because the rate of pulling was too large in this COM distance range. This is supported by the fact that the stepwise pull calculation produced pull forces closer to the umbrella results. However, the continuous pull with the slower pull rate of 0.00001 nm/ps produced forces much closer
Figure 3.3: The pull force as a function of center of mass distance for the M5/M5 dimers. The black dotted line is the all-atom pull force. The dashed line is the dimer without elastic network bonds. The solid lines are the dimers with elastic network bonds; the fast continuous pull, slow continuous pull, stepwise pull and umbrella sampling results are red, brown, green, and magenta respectively.

to those of the stepwise pull. This slower continuous pull was then used for calculating the structural parameters discussed later. The pull forces from the all-atom study are similar in peak magnitude to the umbrella results, but at distances larger than 5 nm the all-atom pull force is more negative than the coarse-grained pull force. As discussed in Ref. 35, the all-atom pull force results contain some contributions from irreversible work which have not been removed from the values plotted in Fig. 3.3. Thus, the correct all-atom result is expected to be less than the curve shown. If one takes the umbrella result to be the best one for the coarse-grained calculation, then the pull forces in this case are generally somewhat larger than those corresponding to the all-atom study.

The pull force calculated with the same methods (except not with the umbrella sampling)
for the M8/M2 dimer can be seen in Fig. 3.4. As with the M5/M5 dimer, the pull force for the slower (0.00001 nm/ps) pull rate is lower than for the 0.001 nm/ps pull rate. The stepwise pull produces similar forces to the slower continuous pull. Unlike the pull force for the M5/M5 dimer, the force for the M8/M2 dimer remains positive for COM distances up to 15 nm. The all-atom results are similar to the coarse-grained slower pull until a COM distance of 4.5 nm. Trajectories which have dissociated have a very small pull force, while unfolding trajectories which are not dissociated require a significant pull force to move to larger COM distance. The percent of trajectories which have dissociated is discussed in the following paragraphs and the unfolding of the M8 monomers can be seen in the radius of gyration (Fig. 3.7). The pull force reported here is the average over both the dissociated and unfolding trajectories, so for COM distances larger than 6 nm the higher pull force for the all-atom method suggests that the all-atom simulations have a smaller percentage of dissociated dimers implying that there is more unfolding in the all-atom simulations.

Consider now the percentage of dissociated trajectories as a function of COM distance, as shown in Fig. 3.5. The ground state cytochrome $c'$ dimer has a minimum distance of approximately 0.43 nm. Using the minimum separation it is possible to determine the proportion of the trajectories that have dissociated. When the minimum distance for a particular trajectory was greater than 0.8 nm, that trajectory was considered to be dissociated. The neutral dimers all dissociate at a similar distance. Between 5.5 nm and 6.0 nm the percent dissociated changes rapidly from 0 to 100 percent. Neutral structures without elastic bond networks tend to dissociate at slightly smaller distances compared with those containing elastic bond networks but the difference is quite small. The M5/M5 complexes begin dissociating at approximately the same COM distance as the neutral ones, but the rate of dissociation with COM distance is slightly less than for the neutral complexes. However, in both cases, the steep rise in the curves indicates that the resulting complexes dissociate quickly and with little structural changes.

Contrasting this is the dissociation behavior for the M8/M2 complexes. These complexes begin dissociating at 6 nm and show a strong increase up to a COM distance of 8 nm at which approximately 50% of the complexes are dissociated. However, after this point, the rate of dissociation drops significantly so that at 14 nm only 80% of the M8/M2 complexes are dissociated. This implies there are two different behaviors in these complexes. The first, which occurs about 50% of the time, involves dissociation with small but significant structural changes. Complexes exhibiting this behavior have dissociated by the time their COM distances have reached 8 nm. The second, which occurs about 50% of the time, involves dissociation with large structural changes. Complexes exhibiting this behavior likely undergo significant amounts of unfolding during the dissociation process, and thus can attain large COM distances before dissociating. Figure 3.5 also shows that while there
Figure 3.4: The pull force as a function of center of mass distance for the M8/M2 dimers. The lines have the same meanings as in Fig. 3.3.
Figure 3.5: The percentage of the 100 trajectories that have dissociated as a function of center of mass distance. The solid, dotted, and dashed lines are the M5/M5, M8/M2, and neutral (M0/M0) charge partitionings, respectively. The black and red curves represent slow pulling of the complexes with and without elastic network bonds, respectively. The blue and green lines represent the results of the stepwise pulling.
is some variation, the dissociation behavior for the M8/M2 structures with and without elastic bond networks is similar.

The behavior seen in Fig. 3.5 can be contrasted with that seen in the corresponding all-atom simulation results, as indicated in Table 1 and Fig. 1 of Ref. 36. For the M5/M5 charge partitioning, the all-atom study showed complexes dissociated over a narrow range of center of mass distances centered at about 5.7 nm, while the neutral ones dissociated at 6.3 nm. Given that the ground state dimer COM distance was 2.0 nm in that study, the neutral and M5/M5 dimers dissociated after being stretched by approximately 4.3 nm and 3.7 nm, respectively. In the present case, the neutral and M5/M5 dimers do dissociate over a narrow range of COM distances but do so at a distance of approximately 6.0 nm which represents a stretch of approximately 4.6 nm. In other words, the neutral complexes in the coarse-grained model dissociate after about the same amount of stretching as in the all-atom model. However, the M5/M5 complexes in the coarse-grained model must be stretched to a larger distance before dissociating. The contrast is more noticeable for the M8/M2 dimer dissociation. The all-atom results, as seen in Fig. 1 of Ref. 36 show a gradual dissociation process and do not give evidence for different dissociation behavior at short versus long COM distances as seen in the present Fig. 3.5.

The minimum distance between monomers is expected to increase linearly with an increase in the COM distance. However, for small COM distances, the minimum distance, as shown in Fig. 3.6, remains small and constant. This indicates the monomers are bound together as a dimer even as the center of masses are pulled apart, that is the monomers stretch as the COM distance increases. At some point, dissociation occurs and the minimum distance begins to increase. For the neutral and M5/M5 complexes, this occurs at a COM distance just less than 6 nm, and for the M8/M2 complexes at about 6.5 nm. These distances correspond nicely with the points in Fig. 3.5 at which the dissociation curves begin to rise. For the neutral and M5/M5 complexes, the minimum distance line has a slope greater than 1 just after dissociation at COM distances between 6 and 7 nm. A slope greater than one implies the minimum distance is increasing faster than the rate at which the COM distances are increasing, in other words, the monomers recoil into a more compact form just after dissociation. Afterward, at about 7.2 nm, the minimum distance line becomes straight and the slope goes to 1 indicating that no further changes in size occur.

For the M8/M2 complexes with an elastic bond network, there is a dip in the minimum distance curve at larger COM distances. This is caused by bound complexes in which one of the monomers is unfolding. Such trajectories have a very small minimum distance value and this lowers the average for the ensemble. As seen in Fig. 3.5, approximately 50% of the complexes undergo this unfolding behavior.

The all-atom study [36] did not measure the minimum distance between monomers, but
Figure 3.6: The minimum distance separating the two monomers as a function of center of mass distance. The thick dashed black line has a slope of 1. The black and red curves represent slow pulling of the complexes with and without elastic network bonds, respectively. The blue and green lines represent the results of the stepwise pulling and the all-atom data, respectively.
instead the minimum intermolecular residual pair distance (the smallest distance between amino acids on the two monomers) was calculated. The minimum residual pair distance has slopes of less than 1 after dissociation indicating structural changes continuing at large COM distances. Most significantly, for no cases was a slope greater than 1 observed for any of the dissociating complexes.

Values of the radius of gyration are shown in Fig. 3.7. For the neutral and M5/M5 charge partitioning, the radius of gyration of the coarse-grained monomers remains approximately constant for all COM distances. In contrast, the all-atom monomers show an increase in radius of gyration for the M5/M5 case starting at a COM distance of 6 nm, that is after dissociation. For the M8/M2 charge partitioning in both the coarse-grained and all-atom models, the monomer with more charges has a larger increase in radius of gyration than the monomer with fewer charges, resulting from the unfolding of the higher charged monomer. The lower charged monomer has a radius of gyration value essentially the same as seen for the neutral and M5/M5 cases. Interestingly, the increase in radius of gyration is larger for coarse-grained M8/M2 complexes with an elastic bond network compared with those without. However, at COM distances larger than 12 nm, the M8/M2 complexes with elastic bond networks have a curve which begins to decrease. After some investigation it was found that this occurs because some trajectories which were unfolding in the predissociated state recoiled to a more compact structure after dissociation. This type of refolding after dissociation is not observed in the corresponding all-atom case, and is observed to a lesser extent for M8/M2 complexes without elastic network bonds. The stepwise dissociation had slightly higher values for the radius of gyration, but showed the same decrease at COM distances larger than 12 nm.

A similar conclusion is drawn by examining the values of the RMSD shown in Fig. 3.8. The M5/M5 charge configuration shows an increase in RMSD which then levels off for COM distances larger than 6 nm (when the structure is no longer changing). This is different than the all-atom data which shows the RMSD increasing more after 6 nm. However, at large COM distances the limiting values are the same for the M5/M5 all-atom and MARTINI models. In the M8/M2 charge configuration, the monomer with more positive charges has a larger increase in RMSD. The all-atom curves initially increase at a slower rate than the corresponding coarse-grained ones but eventually reach comparable values at the largest COM distances. As seen in Fig. 3.7, the coarse-grained RMSD values for the M8 monomers begin decrease sharply at larger COM distances, again indicating that structures are collapsing to a more compact structure. The neutral monomers have RMSD values slightly smaller than the M5/M5 charge configuration, however both these charge configurations show the same trend.

For all the structural parameters, standard deviations were calculated from the set of
Figure 3.7: The radius of gyration as a function of center of mass distance for the M5/M5 (top panel) and the M8/M2 (middle panel) charge partitionings, and the neutral complex (bottom panel). In all panels, the monomer with the higher and lower charges are shown in red and black, respectively. The solid, dashed, and dotted lines are the complexes with elastic network bonds, the complexes without elastic network bonds, and the all-atom data (M5/M5 and M8/M2), respectively. The green (higher charge) and blue (lower charge) curves are the stepwise pulling results.
Figure 3.8: The root mean square deviation (RMSD) as a function of center of mass distance for the M5/M5 (top panel) and the M8/M2 (middle panel) charge partitionings, and the neutral complex (bottom panel). The colors and line types have the same meanings as in Fig. 3.7.
100 trajectories at each COM distance. Because there is an ensemble of structures that contribute to the protein, these standard deviations give a measure of the width of the distribution of the properties. Since all the trajectories start from a relaxation of the ground state structure, the standard deviation is small at small COM distances (on the order of 5%). As the trajectories propagate, the monomers move away from the ground state and the distributions become wider (standard deviations on the order of 30%).

In the MARTINI force field, some secondary structural elements are imposed using the backbone-backbone bond, angle, and dihedral force constants [45]. For helices these values are much larger than they are for other secondary structures so the force field provides a large potential to enforce helical structures. The cytochrome c′ dimer has four helices arranged in a bundle in the ground state. Unfolding can occur in principle through the unbundling of these helices or through the unraveling of individual helices. The former is not hindered by the imposition of secondary structure while the latter is. In order to better understand the effect of secondary structure on our results, a set of M8/M2 (with local elastic network bonds) simulations were performed with the values of the helical backbone-backbone bond, angle, and dihedral force constants set to 625, 350, and 200 kJ/(mol nm) respectively. These values are half the default ones.

The results of these simulations are shown in Fig. 3.9. Generally, the results are qualitatively the same. However, with the weaker helices, the M8/M2 dimers have a larger percent dissociation and a smaller increase in radius of gyration and RMSD. The larger percent dissociation at small distances (rising from 0% to 60%) implies a smaller fraction of monomers unfold upon dissociation. This moves the results in worse agreement with the all-atom ones which show a greater fraction of unfolding, even greater than that seen with the default helical structure parameters. Note that the sharp decrease in RMSD and radius of gyration at larger COM distances is still present in both cases, and the decreases are comparable. Upon examining visually the structures at these large distances, it was found that just before dissociation, they consist of long extended structures, usually with one of the helices being stretched and partially unfolded. After dissociation, these structures compact to globular shapes. Comparable all-atom trajectories also show extended structures before dissociation, some with stretched helical structures. However, after dissociation, the structures remain extended. Caution is required when making comparisons here because the coarse-grained simulations speed up dynamic processes, and in this case, the force-field defined helical parameters additionally increase the rate of helix formation. It is quite likely given enough time, the all-atom extended structures would also collapse into globular structures, akin to the coarse-grained ones. The all-atom simulations were not propagated in time long enough for this to happen. However, the fact the all-atom simulations show no sign of compaction while the coarse-grained ones show a great deal of compaction likely
indicates the rate of compaction in the coarse-grained force field is too high relative to the all-atom case.

Taken together, the results in Fig. 3.9 indicate the monomers with weaker helices are more compact, which is expected if the rigid helical structure is being somewhat relaxed. This compactness appears to be creating more stable structures, less prone to unfolding. Overall, the relaxation of the helical constraints does not change the qualitative behavior of the complex, and in the case of the percentage dissociation, actually produces results which are further from the all-atom ones. At least in this particular case, the default helical parameters in MARTINI are producing satisfactory results, and do not seem to be producing behavior which is inconsistent with the all-atom results, provided detailed quantitative comparisons are not made.

The dissociation behavior of the protein complexes is affected by the balance between the attractive forces within and between the monomers and the repulsive forces caused by the net charges. In order to quantify somewhat this balance, particular components of the potential energy were examined. In the first instance, as seen in Fig. 3.10, the Lennard-Jones intramolecular energy, Coulomb intramolecular energy, and their sum were compared. For all cases, the Lennard-Jones (LJ) intramolecular energy decreases sharply up to a COM distance of 2 nm. Subsequently, the values for the neutral, M5, and M2 monomers become either slowly decreasing or constant up to a distance of about 6 nm after which point they become constant. In contrast, as a result of unfolding, the energies of the M8 monomers generally increase before becoming essentially constant at larger distances. The M5 and M2 monomers have higher LJ intramolecular energy than the neutral because the positive charges have caused the monomers to expand a bit thus increasing distances and lowering LJ energy. They have lower LJ intramolecular energy than the M8 monomer because the M5 and M2 monomers have not unfolded while some fraction of the M8 ones have, leading to less intramolecular contacts. The Coulomb intramolecular energy remains approximately constant in all cases except for the M8 monomers which show a slight decrease with increasing center of mass distance. This is consistent with the fact that the neutral, M2, and M5 monomers show little structural changes during the dissociation while some of the M8 ones unfold, allowing them to lower their intramolecular Coulomb energy. Because the Coulomb contribution is almost constant, the behavior of the total intramolecular energy, as seen in the lower panel of Fig. 3.10, is qualitatively the same as the LJ contribution. Interestingly, at large COM distances the energies are lower than the corresponding values at small distances, implying that upon dissociation the structures of the monomers change so as to increase their favorable LJ contacts. This is possible because they no longer have the configurational constraints present as part of the dimer.

The Lennard-Jones and Coulomb intermolecular energies are shown in Fig. 3.11. The LJ
Figure 3.9: The effect of the secondary structure constraints. The red, black, and green lines represent simulations with the default parameters, simulations with weaker helix structures, and the all-atom simulations, respectively.
Figure 3.10: The intramolecular contributions to the LJ (top panel) and Coulomb (middle panel) energies, as well as their sum (bottom panel). The sum is scaled so that the ground state energy is 0. In all panels, the solid, dash-dot, dotted, and dashed lines represent the monomers with +5, +8, +2, and 0 charges respectively. The black and red curves represent complexes with and without elastic network bonds, respectively.
intermolecular energy is steepest at COM distances less than 2 nm, and becomes flat at COM distances greater than 6 nm. This is the same behavior seen for the LJ energies in Fig. 4 of the all-atom study [36]. There is very little difference in the LJ intermolecular energy plots between the different charge partitionings. The Coulomb intermolecular energy is expected to have a $1/r$ decay for the M5/M5 and M8/M2 complexes. This smooth decrease in Coulomb energy can be seen in the middle panel of Fig. 3.11. As expected, the M5/M5 has a larger intermolecular Coulomb energy than the M8/M2 because the product of the charges on the monomers is higher in the former case. The relative total intermolecular energies are compared in the bottom panel of Fig. 3.11 where it can be seen that, as expected, the neutral complex has the highest barrier. This barrier is reduced as charges are added to the complex with the M5/M5 one having the lowest barrier. Note that the coarse-grained model used here does not use partial charges so its behavior upon dissociation is the same as would be expected for two monopoles separating. The all-atom study did use partial charges so the intermolecular Coulomb energy reported in Ref. 36 also contains energy from dipole and higher moments.

In addition to the quantitative comparisons given above it is possible to compare the coarse-grained and all-atom results on a qualitative level by comparing structures from typical dissociation events. The coarse-grained results are shown in Fig. 3.12 for the M5/M5 and M8/M2 complexes and these should be compared with the corresponding images in Fig. 5 of Ref. 35 and Figs. 5 and 6 of Ref. 36. Generally speaking the structural changes are consistent. At a finer level, the dissociated monomers for the coarse-grained M5/M5 complex have a more globular shape than the corresponding ones in the all-atom study.

3.4 Conclusions

Taken in their entirety, the results show that the MARTINI force field predicts structures and properties in qualitative agreement with the corresponding all-atom results [35, 36] and in many cases in semi-quantitative agreement. However, there are several differences worth noting. The radius of gyration values for the dissociated M5/M5 complexes are the same as those before dissociation for the coarse-grained case and don’t show the increase seen in the all-atom case. The minimum distance data shows that the M5/M5 and neutral coarse-grained complexes contract just after dissociation, a behavior not seen in the all-atom case. This is confirmed by the images in Fig. 3.12 showing a more globular structure for the dissociated M5 monomers. In addition, the pull forces for the M5/M5 dimer dissociation process are consistently higher for the coarse-grained model compared with the all-atom one.

All these differences indicate that the MARTINI force field appears to have a slightly
Figure 3.11: The intermolecular contributions to the LJ (top panel) and Coulomb (middle panel) energies, as well as their sum (bottom panel). The sum is scaled so that the ground state energy is 0. In all panels, the solid, dotted, and dashed lines are the M5/M5, M8/M2, and neutral (M0/M0) charge partitionings, respectively. The black and red curves represent complexes with and without elastic network bonds, respectively.
different balance of attractive and repulsive forces compared to the OPLS-AA/L force field. For the monomers with low charges (neutral, M2, and M5) it is too attractive. This is evidenced by the decrease in radius of gyration after the dissociation and in the minimum distance plot (Fig. 3.6) where immediately after dissociation the slope is greater than one. For the M8 monomers, the effect of the attractive forces is not seen at smaller distances, but is seen in the rapid decrease in radius of gyration (Fig. 3.7) and RMSD (Fig. 3.8) at COM distances greater than 12 nm. Also, while the M8 monomers do show some unfolding behavior, the all-atom pull forces at larger COM distances suggest that the coarse-grained model has less unfolding which indicates that the force field is too attractive. It is the repulsive charges that help monomers unfold, and lower the barrier for dissociation. The MARTINI results are consistent with a model in which the charges are not having the degree of influence as expected from the all-atom results, again indicating the attractive forces are too large.

The elastic network bonds have only a small effect on the results. However, they significantly increase the time step that can be used. A small number of elastic network bonds in the regions of the protein with the most extended structure have the effect of restricting some fast motions without changing the rest of the protein. As well, the default secondary structure parameters in MARTINI which act to constrain helices especially, do not seem to

Figure 3.12: Snapshots of the ground state, M5/M5 (dissociated), and M8/M2 (unfolded) dimers. These images were created using PyMOL [72]. In the M8/M2 dimer, the monomer with more charges is shown on the left.
be producing inconsistent results. In other words, the dissociation mechanism is not highly
dependent upon the details of the secondary structure changes in the proteins.

The goal of the present study was to determine the suitability of the MARTINI force field
for use in molecular dynamics calculations examining the effect of charge migration upon
protein complex dissociation. Such a study does not require a highly accurate potential since
the dissociation mechanism is expected to be fairly generic and not dependent upon fine
interactions. From this perspective, the MARTINI force field should be suitable for such
a qualitative study, although for a quantitative study improvements would be desirable.
Increasing the number of charges in the system or decreasing the value of the relative
dielectric constant would increase the repulsion and thus provide a correction for the balance
of attractive and repulsive forces.
Chapter 4

A Charge Moving Algorithm for Molecular Dynamics Simulations of Gas-phase Proteins

4.1 Introduction

The monomer ejection pathway is one where one monomer leaves the complex typically taking close to half the total charge with. Because the monomers in a multimeric complex start with the charges distributed evenly across the surface area, this pathway requires charges to adopt low energy configurations as the monomer unfolds, hence charges must be mobile. A rigorous treatment of proton transfer involves quantum mechanical calculation of the potential energy during proton motion. Such an approach has been used to study proton motion in water (for reviews see Refs. 73 and 74). In the present case, using a quantum mechanical method to describe proton transfer in protein complexes is computationally very intensive. Further, the time scale for the monomer ejection pathway is much larger than the time scale for proton transfer. Thus, the precise description of this transfer is not expected to play a major role. It is not possible to include the details of proton motion in a coarse-grained potential, because the degrees of freedom associated with protons are not present.

Donnini et al.[38] performed simulations where the protonation states were changed in order to maintain constant pH. However, in an isolated (gas-phase) protein complex, the total charge remains constant rather than the pH.

Lill and Helms created a method (Q-HOP MD) for including proton transport in molecular dynamics (MD) simulations by modeling hopping rates using transition state theory[37]. Their general procedure calculates energies and then uses those energies to determine the hopping probability. They used a multicopy method to enhance sampling at the proton transfer step. It would be technically challenging to implement this for charge hopping in protein complexes.

Instead, a simple method to incorporate moving charges into the simulation of a gas-phase protein complex that is compatible with a coarse-grained force field will be intro-
duced. This method can be used for gas-phase proteins (no solvent). The method involves adding a charge hopping routine to the loop over MD steps, using simple instantaneous hopping between sites and a Monte Carlo like scheme. This hopping algorithm is applied to transthyretin, a system for which experimental data is available[10, 75–77].

Results from some simulations of the homomeric tetramer are discussed to show the nature of the hops which occur and the effects of charge hopping. In addition to structural parameters, properties specific to hops are studied, including the hopping rate, distributions of the change in energy for a hop, and correlations between donor and acceptor sites. Conclusions follow at the end.

4.2 Method

To study the hopping, a transthyretin (TTR) tetramer was used as a model system with both high (+20) and low (+8) total charge states. TTR is a small homotetramer which has been studied experimentally[10, 75–77] using mass spectrometry employing collisionally activated dissociation with total charge states ranging from +8 to +15. In Chapter 3 the MARTINI force field was found to be a bit too attractive, therefore the maximum charge state in the simulations was chosen to be +20 to increase, as an offset, the Coulomb repulsion among the charges. Simulations were performed for temperatures of 300K, 400K, 500K, and 600K. Increasing the temperature from 300K to 600K has the effect of adding energy to the system giving it the energy needed to overcome the dissociation and unfolding barriers, mimicking the CID experiments. The temperature is being used to speed up the dynamics, so it does not give the true temperature dependence of the unfolding or an estimate of the experimental protein temperature. In the present work, simulations were run with $\alpha = 1$ and 0.023. The $\alpha = 0.023$ value was chosen because at 300K this gives a hopping probability of 0.5 or greater for $\Delta E$ values less than +75 kJ/mol (approximate midpoint of $\Delta E$ distribution for the $\alpha = 1$ simulations). A 30 fs time step was used for 75 ns simulations.

In all of these simulations the N-terminus of chain A was allowed to become positively charged while the other three N-termini remained neutral. This biases the trajectories towards the unfolding of one chain as is expected from the monomer ejection pathway. Although this is less realistic than having all four N-termini as basic sites, it serves as a test of the hopping algorithm.
4.3 Results and Discussion

4.3.1 Hopping Rate and Charge Distributions

In order to study the behavior of the hopping algorithm, a number of quantities were examined. Figure 4.1 shows the average hopping rate as a function of time. For each trajectory the simulation time was divided into bins of 150 ps length and the number of hops was counted in each bin and averaged over all trajectories. This number was then divided by 150 to calculate a hopping rate in units of hops per ps, and was normalized for the number of charges in the system. Figure 4.1 shows that the hopping rate is similar for all the temperatures considered, but slightly lower at 300K than at higher temperatures. It is higher at small time, and then after about 10 ns reaches a steady state value of about 0.002 hops/ps/charge for the high charge state and about 0.1 hops/ps/charge for the low charge state when $\alpha = 1$. This is much lower than transfer rates of 0.25 hops/ps/charge, 0.11 hops/ps/charge and 0.46 hops/ps/charge reported for an excess proton in water in Refs. 37, 78 and 79, respectively. The hopping rate is higher for the low charge state because that state has more available acceptor sites so each donor site has a higher probability of coming close to an acceptor site. When $\alpha$ is decreased to 0.023, the hopping rate increases to just above 0.1 hops/ps/charge for the high charge state and 0.4 to 0.6 for the low charge state. The decrease in hopping rate at small time is expected because the initial charge configurations were not optimized and the charges quickly rearrange to find lower energy configurations. However, at long times the hopping rate is not zero while the slope is very close to zero. This implies the system is moving among multiple charge distributions which are quite close to the lowest energy. This is consistent with predictions from simulations with static charges which imply that protein relaxation can stabilize many different charge configurations within a small band of lowest energy[1]. More hopping would be expected at higher temperatures, and the hopping rate does show a slight temperature dependence.

The hopping algorithm used here is a simple approximation to the real proton motion. Protein self-solvation and any other local environment effects are not included. The use of $r_{\text{cutoff}} = 0.57$ means that it is modeling direct proton transfer and any other possible mechanism is not considered. In effect, this $\alpha$ parameter is allowing us to tune the hopping rate without using a large increase in temperature which would change the dynamics of the whole system. Decreasing $\alpha$ from 1 to 0.023 increases the hopping rate for the high charge state from near zero to the same order of magnitude as the hopping rate reported for an excess proton in water[37, 78, 79].

The potential and kinetic energies averaged over all 100 trajectories are shown in Fig. 4.2 for the high charge state. For each temperature the kinetic energy is constant, as expected since a thermostat was used. The potential energy decreases at the beginning as the charges
Figure 4.1: A plot of the average hopping rate (hops/ps/charge) as a function of time for the +20 (bottom) and +8 (top) charge states. The temperatures 300K, 400K, 500K, and 600K are shown as black, red, green, and blue lines, respectively. The solid and dashed lines represent $\alpha = 1$ and $\alpha = 0.023$, respectively.
rearrange, and then becomes constant. This change in potential energy is expected as the charges move to more favorable arrangements. The plateau occurs when the charges are sampling micro-states with the lowest energy. The decrease in potential energy matches the higher hopping rate at small time, while the flattening out occurs at the same time as the hopping rate becomes constant. With the higher hopping rate, more states are accessed. This is because the increased hopping probability makes hops from arginine to lysine more frequently accepted. Thus each charge can undergo more movement. This is seen in Fig. 4.2 where the lower $\alpha$ simulations reach a lower potential energy.

The average number of charges on each monomer chain is shown as a function of time in Fig. 4.3. For the high charge state at 300K, all the chains have an average of +5 charges (a symmetric charge distribution). At 600K, chain A increases in average charge to +6 and the other chains decrease to +4.5 for the high charge state when $\alpha = 1$. Once the charge of chain A begins to increase, it does not return to lower charge values. This asymmetry in charge starts quickly (within the first 2 ns), and monotonically increases with the greatest changes occurring at small times. Intermediate temperatures show an increasing asymmetry in the charge configuration. The charge accumulation on chain A is larger (+9.5 compared to +6) when $\alpha$ is decreased to 0.023 from 1 (see Fig. 4.3) for the high charge state. This is accompanied by a complementary decrease in the number of charges on the other three chains. For the low charge state, the distributions are centered at +2 $\pm$ 1 at both 300K and 600K and this does not change when $\alpha$ is changed. This change is on the same time scale as changes in secondary protein structure. These changes easily fall within experimental time scales in typical time-of-flight mass spectrometry studies (on the order of ms). The fluctuations in charge for each chain are small compared to the average charge.

While Fig. 4.3 shows the average (over 100 trajectories) number of charges on each monomer, Fig. 4.4 shows the distribution of charges at 300K and 600K for $\alpha = 0.023$. The last 100 configurations of each trajectory were counted for the charge distribution histograms. At 300K, all the chains are equivalent and have distributions centered at +5 and +2 (for the high and low charge states respectively) with the majority of populations within $\pm$1 charge of the average. At 600K, the high charge state distributions are broader overall having widths of about $\pm$2 charges and there is asymmetry. One chain has a distribution of charges centered near +10, while the other three chains have lower charges (centered near +3).

While these simulations used a +20 total charge state, the highest charge state as reported in experiments was +15. For a +15 TTR tetramer, CID experiments show fragments with an average charge state +8 $\pm$ 1 and trimers with the remaining charge[10, 76, 77]. The experiments do not have a well defined temperature, but increasing collision energies are expected to correspond to higher internal temperatures. The 600K simulation results and
Figure 4.2: The potential and kinetic energies averaged over all 100 trajectories as a function of time for the +20 charge state at different temperatures. The temperatures 300K, 400K, 500K, and 600K are shown as black (brown), red (magenta), green (dark green), and blue (purple) lines for $\alpha = 1$ ($\alpha = 0.023$), respectively.
Figure 4.3: The average number of charges on each monomer is shown averaged over 100 trajectories as a function of time for different temperatures for the +20 charge state using $\alpha=1$ (middle) and $\alpha=0.023$ (top), and the +8 charge state with $\alpha=0.023$ (bottom). The colors have the same meanings as in Fig. 4.1 (there are four lines at each temperature, one for each chain).
Figure 4.4: A plot of the charge distributions in the last 100 steps of the trajectories for the +20 charge state (left) and the +8 charge state (right) both using $\alpha=0.023$ (the top panel is at 300K and the bottom panel is at 600K). Each chain is in a different color.
the experiments both show one monomer (out of the four) having half the total charge, but the 600K simulations have a broader distribution of charge states than the experiments. This broader distribution could be caused by the simulation time being shorter than the experimental time. If some of the trajectories have unfolded and have the asymmetric charge while others are caught in the middle of this process, it would make the distribution wider than it would be after a longer time.

4.3.2 Structural Properties

Plots of the radius of gyration (averaged over 100 trajectories) of each chain are shown in Fig. 4.5 as a function of time for different temperatures for the high charge state. Generally, the values stay relatively constant in time except for chain A which grows with time. This growth becomes more pronounced as temperature increases. Decreasing $\alpha$ increases the extent to which the radius of gyration increases. This large increase in radius of gyration indicates structural changes. Figure 4.6 shows snapshots of the protein complex at 300K and 600K demonstrating the unraveling of chain A. For the low charge state, the radius of gyration does not change for any of the chains at any temperature indicating that there is no unfolding.

The root mean square deviation (RMSD) was also calculated and shows the same qualitative pattern as the radius of gyration with chain A increasing in size with increasing temperature. Quantitatively, the increase is larger with the higher hopping rate (lower $\alpha$).

4.3.3 Energy Distributions and Pair Correlations

In order to examine the hopping algorithm in more detail, statistics related to the hopping process itself were examined for the +20 charge state. The first of these is shown in Fig. 4.7 in which the values of $\Delta E$ of Eq.(2.15) are plotted as distributions for all attempted hops. Each trajectory was divided into three time domains. The beginning, middle, and end correspond to $t \leq 25$ ns, $25$ ns $< t \leq 50$ ns, and $t > 50$ ns, respectively. All the instances with $\Delta E < 0$ represent successful hops independent of temperature. For $\Delta E > 0$, the temperature dependent hopping probability is determined from $\Delta E$ using Eq.(2.17) so only a fraction of hops will be successful in this case. The energy distributions have two peaks, one at approximately 30 kJ/mol and the other at about 100 kJ/mol for $\alpha = 1$. Most of the $\Delta E$ values are positive, which is consistent with the low hopping rate. The lower $\alpha$ simulations have an additional peak in the $\Delta E$ distributions (see Fig. 4.7) at approximately -100 kJ/mol which was not seen when $\alpha$ was 1.

The vertical lines in Figs. 4.7 and 4.10 represent the differences in $E_{\text{binding}}$ among all possible pairs of basic sites, as calculated from Table 2.1. Recall that $\Delta E$ has two contri-
Figure 4.5: The radius of gyration averaged over 100 trajectories as a function of time for trajectories run at different temperatures for the +20 charge state (left) and the +8 charge state (right). The first (bottom), second, third, and fourth (top) panels show results at 300K, 400K, 500K, and 600K, respectively. Chains A, B, C, and D are represented by black, red, green, and blue lines respectively. The solid and dashed lines correspond to $\alpha = 1$ and $\alpha = 0.023$, respectively.
Figure 4.6: Snapshots of the ground state tetramer and the final structures at the end of two trajectories for the +20 charge state with $\alpha = 0.023$ (one at 300K and the other at 600K). Each chain is in a different color, and charged beads are represented in black.
Figure 4.7: Distributions of $\Delta E$ for all attempted hops taken at times near the beginning, middle, and end of the +20 charge state trajectory propagation. The vertical lines show the differences in proton binding energies among all possible pairs of residue types, as calculated from the values in Table 2.1. The colors have the same meaning as in Fig. 4.1. The top panel shows data collected from the first 25 ns, the middle panel the next 25 ns, and the last panel the remaining 25 ns of the trajectories. The solid and dashed lines represent results for $\alpha = 1$ and $\alpha = 0.023$, respectively.
butions: i) from the differences in $E_{\text{binding}}$ and ii) from the differences in Coulomb energy resulting from moving a charge between two basic sites. If the difference in $E_{\text{binding}}$ dominated $\Delta E$ one would expect to see values of $\Delta E$ peaked near the vertical lines in Fig. 4.7. The peaks are not located at the same values as the differences in binding energy (but they are the same order of magnitude) and they have a peak width of about 40 kJ/mol, indicating the importance of the Coulomb energy.

Plotting the energy distributions separated by the donor site type produces the graph seen in Fig. 4.8. The energy distributions are further broken down by acceptor site type (data shown in Fig. 4.9). The arginine proton binding energy is approximately 90 kJ/mol lower than that of histidine or lysine. Thus, if the electrostatic environment was the same for all hops one would expect that it would take 90 kJ/mol more energy to move a proton from an arginine than to move a proton from a lysine or histidine. This difference in binding energy is the source of the peaks seen in Fig. 4.7. The hops with $\Delta E$ closer to 0 are from hopping pairs where the donor and acceptor have the same type (or lysine/histidine pairs), and the other peaks are when the donor and acceptor are different.

In the first few steps the charges tend to hop to arginine. In the low hopping rate situation, those charges remain on the arginine so most of the attempted hops seen in Fig. 4.7 are unsuccessful hops with an arginine donor. The high charge state has more charges than there are arginines, so there are still some charges on lysines and histidines which can hop to other lysines or histidines. In the high hopping rate situation, the charges do hop from arginine to other basic sites, thus there are more attempted hops with non-arginine donors (and more with arginine acceptors). The $\Delta E$ values of the successful hops can be seen in Fig. 4.10.

The discussion above relates to the high charge state $\Delta E$ distributions. The low charge state $\Delta E$ distributions have peaks at the same $\Delta E$ values but with different peak heights.

The hopping pairs graph shown in Fig. 4.11 plots the bead number of the donor site on the $x$-axis and of the corresponding acceptor site on the $y$-axis for all successful hops. Data is shown for 300K and 600K. Overall, Fig. 4.11 shows the correlations between donor and acceptor sites counted over all successful hops in the 100 trajectories. The number of hops in all 100 trajectories between any pair of sites is indicated by the color: black is 1 to 999, red 1000 to 9999, and green 10,000 to 49,999. Chains A, B, C, and D are represented by bead numbers 1-254, 255-508, 509-762, and 763-1014, respectively. Most of the hops are between residues in the same monomer. Hops between different monomers do occur, but they are less frequent. Many of the hops are between pairs that are close together in the sequence. These appear as points near the diagonal. The plot is approximately symmetric about the line $y=x$. This means that if a hop from residue A to residue B occurs, then the hop from residue B to residue A will often occur with a similar frequency (not necessarily
Figure 4.8: Energy distributions for all the hop attempts of the +20 charge state separated by the type of the donor site (arginine, lysine, and histidine donors are shown in black, red, and green respectively). The left column is at 300K and the right column is at 600K. Each trajectory was split into three time regions as in Fig. 4.7. The solid and dashed lines represent results for $\alpha = 1$ and $\alpha = 0.023$, respectively.
Figure 4.9: Energy distributions for the hop attempts of the +20 charge state in the last time division of Fig. 4.8 separated by the type of the donor site (arginine, lysine, and histidine donors are shown in the top, middle, and bottom rows respectively) and also by acceptor site (arginine, lysine, and histidine acceptors are shown as black, red, and green lines respectively). The left column is at 300K and the right column is at 600K. The solid and dashed lines represent results for $\alpha = 1$ and $\alpha = 0.023$, respectively.
Figure 4.10: Energy distributions for the accepted hops. The black, red, green, and blue lines represent results at 300, 400, 500, and 600K respectively. The solid and dashed lines represent results with $\alpha = 1$ and $\alpha = 0.023$, respectively. The vertical lines represent the differences in binding energies among all the possible residue types.
exactly the same). This seems counterintuitive since according to the values of $E_{binding}$ hops should be preferred when moving charge to more basic sites and not vice versa. However, hops between sites of the same type would be equally likely no matter which one was the donor.

The number of pairs for which hops occur is much larger at 600K than at 300K. In particular, there are many more hops between different chains. At both temperatures the pairs with the most frequent hops are those between sites that are close together in the sequence (near the diagonal).

All the possible basic sites have a positive charge at some point in the 100 trajectories. Not all sites are visited with the same frequency. The TTR tetramer has 16 arginine residues, 16 histidine residues, and 24 lysine residues. Arginine is the most basic of the four types of basic site considered here and the charges spend significantly more time on arginine than on any other residue. Lysine is the least basic, but it is the most common so lysine has a larger percentage of the charges than histidine. The low occupation of the N-terminus is also explained by the number of residues of each type. As the temperature increases, the percent occupation for arginine increases significantly (from 57% to 77%) and the occupation of the N-terminus increases slightly while the others decrease.

4.3.4 Computational Performance

All the computations performed in this work were done using the orcinus computer (part of the WestGrid/ Compute Canada computing consortium, www.westgrid.ca). For the transthyretin tetramer with +20 charges, a simulation of 2,500,000 steps (75 ns) took 5.75 hours with hopping and 5.5 hours with fixed charges. For the cytochrome c’ dimer with +10 charges, a simulation of 2,500,000 steps (75 ns) took 1.8 hours with hopping and 1.75 hours with fixed charges. There is only a small difference in computational time with and without the hopping. The time needed for the first stage of the hopping algorithm depends on the number of donor and acceptor sites (not the total number of beads) and the later parts of the hopping algorithm are only calculated for potential hops. However, the calculations of the forces in the molecular dynamics program are performed pairwise for all the atoms in the molecule and the updating of positions and velocities is also done for all the atoms, so these take most of the computational time.

4.4 Conclusions

Charge motion was studied in the +20 and +8 charge states of the TTR tetramer. Overall, the comparison of the high and low charge states shows that the Coulomb repulsion among the charges is a major contributor to unraveling a monomer in the complex, thereby
Figure 4.11: For each successful hop in the +20 charge state, the bead number of the donor site is plotted on the x-axis with the bead number of the corresponding acceptor site on the y-axis. Results are shown for 300K (top) and 600K (bottom) both using α=1. The number of hops between any pair of beads is indicated by the color: black is 1 to 999, red 1000 to 9999, green 10,000 to 49,999, and blue more than 50,000.
producing a state ready for dissociation. As seen in Fig. 4.5, monomer unfolding does not occur for the low charge (+8) states, regardless of the temperature or the charge hopping rate (that is, the value of α). For the high charge (+20) state, charge enrichment of a single monomer and significant monomer unfolding can occur. The rate of these processes is very slow when the temperature and hopping rate are low but they increase dramatically both with higher hopping rates and temperatures, as seen in Figs. 4.3 and 4.5. In this sense, the temperature and value of α act to decrease the timescale for these events but do not change their qualitative behavior.

This also speaks to the coupling between charge migration and monomer unfolding. With charge migration turned off, no monomer unfolding occurs regardless of the temperatures studied. Thus, it is the combination of charge migration, needed for charge enrichment, and energy activation, accomplished by increasing temperature, that leads to monomer unfolding. This is consistent with experimental studies[10, 75–77].

Because the Coulomb repulsion among the charges is the dominant factor for monomer unfolding, it is expected that particular microscopic details of the force field or protein structure are not critical for determining the qualitative behavior of the monomer unfolding process. Rather it is a result of larger scale variations in the charge distribution. This is a key assumption in these calculations and in the interpretation of the results. For example, the MARTINI force field provides strong bonds to maintain secondary structures, such as helices, so the particular way a monomer unfolds using the MARTINI force field will differ from that for an all-atom force field. In effect the MARTINI force field has a higher barrier for unfolding than typical all-atom force fields. However, the protein is still able to unravel. The more flexible coils between the strong secondary structures allow movement which contributes to the ability of the protein to reach an extended structure. This is shown graphically in Fig. 4.6 where it can be seen that a monomer can extend quite significantly, even though there are pockets of structure still locked in by the strong secondary structure bonds. These pockets do not change the fact that charge is associating with the unfolding monomer, and that this charge is being moved farther from the remaining trimer complex. The drop in Coulomb repulsion still occurs as a result. For this reason, the qualitative sequence of events leading to monomer unfolding should be reasonable using the MARTINI force field. In the simulations, the MARTINI bead types were not changed when the charge on a bead was changed. Again, the small change in potential parameters associated with the fitting for these different charge states is not expected to affect the qualitative behavior of the unfolding process.

Overall, the results of this study show that for the purposes of understanding charge motions in large, charged, gas-phase protein complexes, a detailed description of proton transfer is not necessary since processes are dominated by Coulomb repulsion. This is con-
sistent with experimental observations, and with previous static charge simulations from which the Coulomb repulsion model[1, 17, 33–36] was developed. For highly charged complexes, as studied here, total charge is the main physical parameter governing the process. The method used here is general enough to be applied to simulations of different protein complexes. Since the total charge and the associated Coulomb repulsion govern the process rather than the detailed protein structure, the general trends will be relevant for many protein complexes (although the quantitative details would differ).
Chapter 5

Controlling Dissociation Channels of Gas-Phase Protein Complexes Using Charge Manipulation

5.1 Introduction

For positively charged proteins, the N-terminus is expected to be important for monomer unfolding. This unfolding can occur as a sequence of steps moving from one end of the protein to the other similar to the zipper model for the folding of cytochrome c[80] discussed by Morozov et al.

In this chapter, results from modifying the protonation of N-termini sites and adding a tether to the protein complex are reported. This tether is a group of coarse-grained beads that is attached to the protein complex at a position of our choosing. The tether contains a basic bead that can accept a positive charge. By changing the number of basic sites on a monomer, one hopes to influence its unfolding behavior as this behavior determines the dissociation products seen in the mass spectra. Controlling the dissociation could provide more information about the structure of the complexes and lead to a better understanding of their biology[18, 81]. Our simulations do not directly model either CID or SID. Instead the simulations use increasing temperature and/or high hopping rates to mimic energy being transferred to the system.

5.2 Method

The charge hopping algorithm described in Chapter 2 was used. The $\alpha$ parameter for the charge hopping is set to 0.023 which leads to hopping probabilities centered around 0.5 for the typical range of $\Delta E$ values (at 300 K) and the default values were used for the other mobile charge parameters. Simulations were run at 300 and 600 K with a 30 fs timestep and total simulation times of 75 or 300 ps. The 300 K temperature is similar to normal room temperature, and the 600 K temperature corresponds to adding more energy more quickly (more like SID). All the charge configurations have a total charge of +20.
The transthyretin (TTR) homotetramer was used as a model system. Simulations were performed with and without an added tether. This tether consisted of two beads, one of MARTINI bead type N0 (non-polar) and the other of type Qd. The Qd bead is included in the list of basic sites. The hydrogen binding energy of the tether was set to -940 kJ/mol which is comparable to the value for lysine. A few simulations were also run with the hydrogen binding energy set to -1200 kJ/mol. No significant differences were seen between the two. The tether was attached to the cysteine sidechain at the first residue of chain A. For the case without the tether, simulations were also performed where the N-terminus from only one of the chains was allowed to accept a positive charge while the other three N-termini remained neutral. This was done by removing all the N-termini except one from the donor-acceptor list, so the chemical environment of the neutral N-termini (number of beads and their interactions) was not changed.

Several properties were calculated in order to examine the structural changes in the protein complex. The number of charges on each monomer was measured as a function of time, because the hopping algorithm allows charges to move from one monomer to another. Various distances were also calculated such as the minimum distance from the C- or N-terminus of one chain to the other three, as well as distances associated with positions within the length of the chain.

5.3 Results and Discussion

Some snapshots of sample trajectories are shown in Figs. 5.1 and 5.2. Figure 5.1 shows a set of snapshots for the complex with all four N-termini accepting charges and without the tether. The top snapshot shows the initial structure, the snapshot in the middle is at the end of a trajectory at 300K, and the bottom is a snapshot from trajectories at 600K. In this case, after a 300K trajectory, two chains are showing some unraveling while the other two are still compact showing the potential for asymmetric pathways. After a 600K trajectory, all four monomers are unraveling producing a symmetric result.

Figure 5.2 shows results from complexes with tethers (small, shown in red) attached to chain A (shown in grey). The top snapshot shows the initial structure, the snapshot in the middle is at the end of a trajectory at 300K, and the bottom two are snapshots from trajectories at 600K. In the initial structure all four monomers are folded. After the 300K trajectory, chain A is unfolded and the end with the tether is farthest from the rest of the complex. After the 600K trajectories, more than one of the monomers (including chain A) have unfolded. Again the tether is at the point farthest away from the center of the complex. The chain with the tether is unraveling more than the others producing asymmetry.

The number of charges per monomer and radius of gyration are shown in Figure 5.3.
Figure 5.1: Snapshots of the tetramer (with all four N-termini able to be protonated) are shown in the initial state (top), after a 300 ps trajectory at 300K (middle), and after a 300 ps trajectory at 600K (bottom). Each chain is shown in a different color (grey, orange, green, and blue).
Figure 5.2: Snapshots of the tetramer with an attached tether are shown in the initial state (top), after a 300 ps trajectory at 300K (middle), and after a 300 ps trajectory at 600K (bottom). Each chain is shown in a different color (grey, orange, green, and blue, and the tether is shown in red attached to the end of the grey chain.)
for simulations run at 300 K and 600 K. When looking at the data plots, please keep in mind that the y-axis scales are not the same for different rows. As seen in the middle column, when all four chains are identical, there is no accumulation of charge on one chain in preference to the others. The radius of gyration values are also the same for all four chains. When chain A has a basic site at the N-terminus and the other chains do not accept charges, as shown in the left column, there is charge accumulation on chain A (at a simulation temperature of 600K). There is also a corresponding increase in the radius of gyration. When chain A has an added basic site on the tether (placed near the N-terminus) as seen in the right column, charge accumulation can also be seen for chain A (at both 300K and 600K). In all cases, increasing the simulation temperature increases the radius of gyration.

Figures 5.4 and 5.5 show the minimum distances from beads on one monomer to the other three chains at simulation temperatures of 300K and 600K respectively. The minimum distance is measuring how close the bead is to the rest of the complex, so if a chain is unfolding those beads involved in the unraveling will move farther away from the other parts of the complex. The four beads used as reference points are the first bead for the N-terminus, the last bead for the C-terminus, a bead one third of the way along the chain, and a bead two thirds of the way along the chain. As seen in the middle column, when the four N-termini are the same, the minimum distance between the N-terminus and the other three chains increases, while the other three distances remain the same. This increase is faster and larger at 600K than at 300K. The left column shows that when one N-terminus is a basic site and the other three are not, at 300K the N-terminus minimum distance for chain A increases while the other distances do not. At 600K, the N-terminus minimum distance for chain A increases much more. The minimum distances for the beads one third and two thirds of the way down the chain increase, but with progressively smaller values and later times. The C-terminus minimum distance does not change. In the case with the tether, seen in the right column, the increase is similar to the case with all four chains the same, but chain A (the one the tether is attached to) has a slightly larger increase.

In all cases, the N-terminus has the largest movement away from the rest of the complex, and as one moves towards the C-terminus the changes in distance from the complex are smaller. It is clear that the N-terminus is leading the unraveling with the middle of the protein monomer following. The increases in N-terminus minimum distance match the increases in the radius of gyration.

Unsurprisingly, when all four chains are the same, they all show the same behavior. When only one of the four chains has an N-terminus which can accept a positive charge, that chain increases in radius of gyration more than any of the other chains. Adding a tether changes the behavior of the complex. The chain with the tether unravels more than
Figure 5.3: Averages over 100 trajectories run at 300K (top two rows) and 600K (bottom two rows). The top row of each pair shows the number of charges per monomer and the bottom row shows the radius of gyration (in nm). In the left column the N-terminus of chain A could be positively charged and the other three N-termini could not. In the middle column all four N-termini could accept a positive charge. In the right column a tether was added to chain A. The results for chains A, B, C, and D are shown in black, red, green, and blue, respectively.
Figure 5.4: Averages over 100 trajectories run at 300K. The top row shows the minimum distance from the N-terminus to the other three chains. The second row shows the minimum distance from a bead one third of the way along the monomer. The third row shows the minimum distance from a bead two thirds along the monomer. The bottom row shows the minimum distance from the C-terminus. In the left column the N-terminus of chain A could be positively charged and the other three N-termini could not. In the middle column all four N-termini could accept a positive charge. In the right column a tether was added to chain A. The results for chains A, B, C, and D are shown in black, red, green, and blue, respectively.
Figure 5.5: The same as Fig. 5.4, but with a simulation temperature of 600K.
the other three chains. However, it does not increase in size as much as chain A did when it was the only chain with an N-terminus which could accept charge.

The accumulation of charge and the unraveling behavior happen to a greater extent at 600K (compared to 300K) especially the unraveling. They also happen more rapidly.

One fundamental difference between CID and SID is that SID has a much faster energy transfer to the protein. These simulations do not directly model either CID or SID, but in the simulations a higher temperature corresponds to faster dynamics and faster changes in the charge configurations. It is possible that if the energetics for the initial step of unfolding is such that it is a rare event (as in CID where energy transfer is slow), once one monomer started unfolding it could completely unfold and dissociate before the other monomers had a chance to unravel. This would result in preferential charging of the unfolding monomer. Because SID is a faster and more energetic process, the monomers gain more energy in a short time which could lead to dissociation without unfolding or multiple monomers unfolding at the same time. This would result in the charge being spread out over all the monomers and thus there would be no preferential charging of one monomer and symmetric dissociation products would be seen.

If one chain is to have an increase in the number of charges, then the other chains must lose charges. This creates a competition between the chains for the charge accumulation and unfolding pathway. When all four chains are unfolding (as in Fig. 5.1 (c)), the charge will remain evenly distributed which is consistent with the symmetric dissociation seen in SID[8]. The mass spectra cannot distinguish between four folded monomers and four unfolded monomers when they all have the same charge.

5.4 Conclusions

The N-terminus is leading the unraveling. A positively charged terminus moves away from the rest of the positively charged residues. There is no evidence that the middle part of the protein chains breaks away first. The protein unraveling moves from one end of the protein to the other in a manner similar to the zipper mechanism[80]. However, one must be careful to look at the broad trends and not the quantitative values or residue level structure. The MARTINI force-field used here does not have an atomistic level of detail. Also, the secondary structure of the input imposes values for bonding and Lennard-Jones parameters which limit changes in secondary structure and thus could increase the unfolding barriers.

This suggests two possible ways for controlling the unraveling in experiments by making chemical modifications at or near the N-terminus. One is by blocking the N-termini for the chain(s) for which unfolding is not desired, so that they cannot become positively charged. In this set of simulations, some N-termini were blocked from accepting charge,
but the N-termini structure and interactions remained the same. This could be difficult to do experimentally. Adding bulkier groups to block protonation of N-termini would change the structure and this could lead to different results. In principle this could be simulated, but it has not yet been done. The second is by adding basic functional groups as a tether to increase the basic site density near the N-terminus of the chain for which unfolding is wanted. These two types of modifications could also be used in combination, that is by blocking unwanted sites and adding tethers to preferred ones.

Another possibility is to block protonation at all N-termini and then place tethers on the protein (not at the N-terminus) in order to direct the dissociation forces. Since the tethers will accumulate charge and move away from the complex, they act to focus forces on the complex at their points of connection. By directing these tethers to opposing positions on a complex, forces could be created that attempt to dissociate it by a means other than monomer ejection.

The simulation results shown here are preliminary ones in the sense that the vast parameter space involved with tether placement, length, and charge location has not been explored. However from the present results a definite degree of control is already seen, although it could be optimized. The hope is that these initial results will stimulate the idea of using charge manipulation through N-terminus blocking or tether addition to provide an experimental strategy to build a set of tools for rationally controlling protein complex dissociation.
Chapter 6

Concluding Remarks

6.1 Conclusions

Many protein complexes follow the monomer ejection mechanism where one monomer unfolds and leaves the complex with close to half the charge. This behavior is mostly driven by the Coulomb repulsion among the charges.

There are two ways for the charges to move farther apart - unfolding (where the protein expands and the charges within the monomer move away from each other as well as from the other chains) and dissociation (where the charges on one monomer are moved away from the rest of the complex). There is competition between these pathways. In this work, the protein monomers have unfolded more readily than they dissociated. In Chapter 3, the dimer dissociated when it was forced with the pull code. The other simulations run here, sometimes resulted in unfolding but we did not see dissociation. This is a topic we are pursuing further.

Chapter 4 contains results for tetramers with total charge states of +8 (low charge) and +20 (high charge). The high charge state has larger radius of gyration values than the low charge state and shows increases in radius of gyration not seen in the low charge state (see Fig. 4.5). This is expected because the higher charge state has more Coulomb repulsion which provides the energy needed to overcome the unfolding barrier. The +8 charge state has hopping rates up to 3 times larger than the +20 charge state. Unlike the collisionally induced dissociation experiments discussed in Chapter 1, we only saw asymmetric changes in radius of gyration when the simulations were biased (either by blocking charges from three of the N-termini or adding a basic tether).

Increasing the temperature increases the extent of unfolding (see Figs. 4.5 and 5.3). The radius of gyration increases for all cases, but this increase is most dramatic for the chains showing some unfolding. The hopping rate (Fig. 4.1) shows a very small increase in hopping rate with increasing temperature.

The ability to control dissociation is explored in Chapter 5. To control the unfolding it is necessary to manipulate the charges. Two types of protein modifications are considered, blocking charges from moving to the N-termini and adding a tether. The role of N-terminus in unfolding was shown in Fig. 5.4. The N-termini lead the unfolding and move away from
the rest of the complex, because of this the tether is added near the N-terminus. When a tether is added to the tetramer, the chain on which the tether is attached has more of the charges and a larger radius of gyration than when all four chains are the same. When only one of the N-termini is allowed to accept a charge (while the others are kept neutral), that chain accumulates more charges than the others and is the only one to unfold. These differences show the potential for controlling the unfolding, although it has not yet been optimized.

The reason that charge motion was included was that the charges are expected to move. The lowest energy charge distribution for a fully folded complex has the surface charge density as close to uniform as the discrete nature of the charges allows. For a complex with all the subunits the same this would mean that there are the same number of charges on each monomer. After the experimental CID dissociation, it is clear that not all the monomers have the same charges. In previous work with fixed charges, this was simulated by using different starting configurations, some with symmetrically distributed charges and some with asymmetric distributions. Starting all the trajectories with symmetric charge distributions, some charge accumulation is seen when the charges are allowed to move. The extent of unfolding of the monomer correlates to the amount of charge accumulated on that monomer.

### 6.2 Future Work

This work could be expanded by performing calculations on other protein complexes including complexes with two (or more) different types of monomers. Further exploration of the tether could be interesting, in particular placing the tether at different locations within the protein such as near the C-terminus.

The MARTINI force field is used for many systems[82]. The charge hopping algorithm could be applied to any simulation that uses MARTINI, as long as there are donor and acceptor sites for the charges. An area for future work is using this method to study other systems. The same code can easily be used for other protein systems and is not limited to protein complexes of the type studied here. One example of using this method for other systems is the study of an intrinsically disordered protein. It would also be possible to extend the method for use with other types of charged systems, such as detergent molecules. Adapting the hopping code for parallel computing would be a great advantage when studying larger systems with more molecules. The same algorithm could easily include negative charges. Only positive charges were considered here because the protein complex experiments almost exclusively use positive charge states, but negative charges also follow Coulomb’s Law so they would have the same physics.
The MARTINI force field has some limitations. Because of the coarseness, it is not able to capture secondary structure. There are no hydrogen bonds in the coarse-grained model, and it is these hydrogen bonds which define and maintain the secondary structure. This is partially offset by using high force constants for the backbone beads bonds and angles and by the elastic network bonds. However, that increases the energetic barrier for unfolding and reduces the flexibility for changes in secondary structure. It is slightly too attractive (Chapter 3), and is not reliable for quantitative measurements. Thus for future work, it would be interesting to use a more accurate and detailed force field.

Using an all-atom force field with this type of charge hopping would be extremely difficult, because in the all-atom case hydrogen atoms are considered explicitly and moving the charge would involve changing the bonding topology of the protein. Making and breaking covalent bonds is not common in molecular dynamics. Also, moving an atom (as opposed to just changing the charge) could easily form bad contacts resulting in numerical instabilities. However, it would be possible to adapt the charge hopping routine for use with a united atom force field. This will allow studies with more detail. You would have to select atoms carefully for calculating the donor to acceptor distance. It is a harder challenge than the coarse-grained model used here, because the net positive charges on basic side chains are distributed as partial charges (the MARTINI model used in this work only has integer charges). This makes moving the charge computationally more complicated because more than one atom is involved for both the donor and acceptor sites and more partial charges (both positive and negative) must be correctly included in the Coulomb energy calculation.

For each type of basic site, it would be necessary to determine the partial charges for each atom in the amino acid in both the charged and neutral forms. Then when a hop occurs, all the charges in the donor residue would have to change to the neutral form and all the charges in the acceptor residue would have to change to the charged form.
Bibliography


Appendix A

Source Code

A.1 mobile_charge.c

/*
 * This is a routine to allow positive charges to move from one bead
to another basic site.
 *
 * mdp options:
 * mobile_charge bool (default no)
 * mobile_cutoff real (default 0.57) MARTINI bead diameter plus 1 angstrom
 * mobile_prob real (default 1.0) scaling factor for probability
 * mobile_freq int (default 10)
 * mobile_seed int (default -1)
 *
 * Designed for use with coarse-grained force field (MARTINI).
 *
 */
#ifdef HAVE_CONFIG_H
#include <config.h>
#endif

#include <string.h>
#include "smalloc.h"
#include "typedefs.h"
#include "macros.h"
#include "random.h"
#include "gmxrandom.h"
#include "physics.h"
#include "xvgr.h"
#include "vec.h"
#include "pbc.h"
#include "txtdump.h"
#include "ionize.h"
#include "mobile_charge.h"
#include "names.h"
#include "futil.h"
#include "network.h"
#include "mtop_util.h"
#include "gmxfio.h"
#include "util.h"

/* END GLOBAL VARIABLES */

void mobile_charge(FILE *fplog,t_mdatoms *md,t_state *state,gmx_mtop_t *mtop,
        t_commrec *cr,t_inputrec *ir,gmx_enerdata_t *enerd,int step)
{
    /* define other variables here */
    t_atoms *atoms;
    static FILE *mobile_config;
    static FILE *mobile_out;
    static FILE *mobile_dist;
    static int start,end;
    static bool bFirst = TRUE;
    static gmx_rng_t x_rng;
    static int seed;

    /* donor_sites has the array of the charged beads, and acceptor_sites has the
       array of the basic beads which are available to accept a positive charge */
    static t_sites donor_sites[40];
    static t_sites acceptor_sites[300];
    t_sites temp_site;
    int num_res;

    /* number of charges, number of uncharged basic sites */
    static int num_charges,num_basic;

    /* number of pairs considered for hops, number of hops that occur */
    int pair_count,hop_counter;

    /* variables for loop indices */
    static real mobile_cutoff2,mobile_prob;
    rvec mobile_vec1,mobile_vec2,r_lm;
    real mobile_norm,mobile_test;
    int j_close,donor_num,acceptor_num;
    t_mobile mobile_data[40];
    real time,kt,random_number;
    real e_prot_lys = -937.216; /* kJ/mol */
    real e_prot_arg = -1028.8456; /* kJ/mol */
    real e_prot_his = -957.7176; /* kJ/mol */
    real e_prot_n = -959.8096; /* kJ/mol (N-terminus) */
    real e_prot_tet = -940; /* kJ/mol (tether) */

    if(bFirst){
        mobile_out = gmx_fio_fopen("mobile_out.out","w");
        mobile_config = gmx_fio_fopen("mobile_config.out","w");
        mobile_dist = gmx_fio_fopen("mobile_dist.out","w");

        start = md->start;
    }
end = md->start+md->homenr;
mobile_cutoff2 = ir->mobile_cutoff+ir->mobile_cutoff;
mobile_prob = ir->mobile_prob;

fprintf(fplog,"\nE\_binding\_tether=\%8.4f\n", e\_prot.n);

/* Find charged sites and available basic sites (and count them) */
num_charges = 0;
num_basic = 0;

/* N-terminus */
if (md->chargeA[0]==1) {
    donor\_sites[num\_charges].resname = "n";
    donor\_sites[num\_charges].resnum = 0;
    donor\_sites[num\_charges].bead\_num = 0;
    donor\_sites[num\_charges].e\_binding = e\_prot.n;
    num\_charges++;
}
else {
    acceptor\_sites[num\_basic].resname = "n";
    acceptor\_sites[num\_basic].resnum = 0;
    acceptor\_sites[num\_basic].bead\_num = 0;
    acceptor\_sites[num\_basic].e\_binding = e\_prot.n;
    num\_basic++;
}

/* specific to 3GRG chain B */
if (md->chargeA[255]==1) {
    donor\_sites[num\_charges].resname = "n";
    donor\_sites[num\_charges].resnum = 116;
    donor\_sites[num\_charges].bead\_num = 255;
    donor\_sites[num\_charges].e\_binding = e\_prot.n;
    num\_charges++;
}
else {
    acceptor\_sites[num\_basic].resname = "n";
    acceptor\_sites[num\_basic].resnum = 116;
    acceptor\_sites[num\_basic].bead\_num = 255;
    acceptor\_sites[num\_basic].e\_binding = e\_prot.n;
    num\_basic++;
}

/* specific to 3GRG chain C */
if (md->chargeA[508]==1) {
    donor\_sites[num\_charges].resname = "n";
    donor\_sites[num\_charges].resnum = 231;
    donor\_sites[num\_charges].bead\_num = 508;
    donor\_sites[num\_charges].e\_binding = e\_prot.n;
    num\_charges++;
}
else {
    acceptor_sites[num_basic].resname = "n";
    acceptor_sites[num_basic].resnum = 231;
    acceptor_sites[num_basic].bead_num = 508;
    acceptor_sites[num_basic].e_binding = e_prot_n;
    num_basic++;
}

/* specific to 3GRG chain D */
if (md->chargeA[761]==1) {
    donor_sites[num_charges].resname = "n";
    donor_sites[num_charges].resnum = 346;
    donor_sites[num_charges].bead_num = 761;
    donor_sites[num_charges].e_binding = e_prot_n;
    num_charges++;
}
else {
    acceptor_sites[num_basic].resname = "n";
    acceptor_sites[num_basic].resnum = 346;
    acceptor_sites[num_basic].bead_num = 761;
    acceptor_sites[num_basic].e_binding = e_prot_n;
    num_basic++;
}

/* basic amino acid residues. i is looping over all the beads */
for (i=start; i<end; i++) {
    num_res = mtop->molt->atoms.atom[i].resnr;
    if (!strcmp(*mtop->molt->atoms.resname[num_res],"ARG")){
        i=i+2;
        if (md->chargeA[i]==1) {
            donor_sites[num_charges].resname = "arg";
            donor_sites[num_charges].resnum = num_res;
            donor_sites[num_charges].bead_num = i;
            donor_sites[num_charges].e_binding = e_prot_arg;
            num_charges++;
        }
    } else {
        acceptor_sites[num_basic].resname = "arg";
        acceptor_sites[num_basic].resnum = num_res;
        acceptor_sites[num_basic].bead_num = i;
        acceptor_sites[num_basic].e_binding = e_prot_arg;
        num_basic++;
    }
}
if (!strcmp(*mtop->molt->atoms.resname[num_res],"LYS")){
    i=i+2;
    if (md->chargeA[i]==1) {
if (!strcmp(*mtop->moltype->atoms.resname[num_res], "lys")){
    donor_sites[num_charges].resname = "lys";
    donor_sites[num_charges].resnum = num_res;
    donor_sites[num_charges].bead_num = i;
    donor_sites[num_charges].e_binding = e_prot_lys;
    num_charges++;
}

} else{
    acceptor_sites[num_basic].resname = "lys";
    acceptor_sites[num_basic].resnum = num_res;
    acceptor_sites[num_basic].bead_num = i;
    acceptor_sites[num_basic].e_binding = e_prot_lys;
    num_basic++;
}

}

if (!strcmp(*mtop->moltype->atoms.resname[num_res], "HIS")){
    i = i + 3;
    if (md->chargeA[i] == 1) {
        donor_sites[num_charges].resname = "his";
        donor_sites[num_charges].resnum = num_res;
        donor_sites[num_charges].bead_num = i;
        donor_sites[num_charges].e_binding = e_prot_his;
        num_charges++;
    } else{
        acceptor_sites[num_basic].resname = "his";
        acceptor_sites[num_basic].resnum = num_res;
        acceptor_sites[num_basic].bead_num = i;
        acceptor_sites[num_basic].e_binding = e_prot_his;
        num_basic++;
    }
}

} else{

}

if (!strcmp(*mtop->moltype->atoms.resname[num_res], "TET")){ /*tether*/
    i = i + 1; /* small tether (two beads, one basic site) */
    if (md->chargeA[i] == 1) {
        donor_sites[num_charges].resname = "tet";
        donor_sites[num_charges].resnum = num_res;
        donor_sites[num_charges].bead_num = i;
        donor_sites[num_charges].e_binding = e_prot_tet;
        num_charges++;
    } else{
        acceptor_sites[num_basic].resname = "tet";
        acceptor_sites[num_basic].resnum = num_res;
        acceptor_sites[num_basic].bead_num = i;
        acceptor_sites[num_basic].e_binding = e_prot_tet;
        num_basic++;
    }
}
/* initialize random number generator */
if(ir->mobile_seed == -1) {
    seed = gmx_rng.make_seed();
} else {
    seed = ir->mobile_seed;
}
printf(mobile_out,"mobile_seed=%10d\n",seed);
x_rng = gmx_rng_init(seed);

bFirst = FALSE;
}

time = step*ir->delta_t;
kt = enerd->term[F_TEMP]*(1.3806504E-26)*(6.02214179E23) ; /*kJ/mol*/

/* Loop over charged sites and calculate distance to basic sites */
pair_count = 0;
for(charge=0; charge<num_charges; charge++) {
    mobile_test = 10000.0;
    copy_rvec(state->x[donor_sites[charge].bead_num],mobile_vec1);
    for(basic=0; basic<num_basic; basic++){
        copy_rvec(state->x[acceptor_sites[basic].bead_num],mobile_vec2);
        mobile_norm = distance2(mobile_vec1,mobile_vec2);
        /*For each charge find the available basic site which is closest */
        if(mobile_norm < mobile_test){
            j_close = basic;
            mobile_test = mobile_norm;
        }
    }
    /* If the charged site's nearest partner is close enough for a transfer to occur put the atom indices into mobile_data. pair_count will give a count of the number of pairs that are close enough. */
    if(mobile_test < mobile_cutoff2){
        mobile_data[pair_count].donor_index = charge;
        mobile_data[pair_count].acceptor_index = j_close;
        mobile_data[pair_count].mobile_distance = mobile_test;
        mobile_data[pair_count].hopped = 0;
        pair_count++;
    }
} /* pair_count should not be changed after this */
/* check that you aren't trying to move 2 charges to the same basic site */
if(pair_count>1){
  for(pair_index=0; pair_index<pair_count-1; pair_index++){
    for(m=pair_index+1;m<pair_count;m++){
      if(mobile_data[pair_index].acceptor_index==mobile_data[m].acceptor_index){
        fprintf(fplog,"\nWarning: 2 charges to same site\n")
      }
    }
  }
}

/* Now make a decision if charge transfer will occur (use coulomb energy and mobile probability). Change the charges, then update the charged and basic sites arrays */
hop_count =0;
/* for each charge that could move calculate delta E */
for(pair_index=0; pair_index<pair_count; pair_index++){
  mobile_energy(fplog, state, mtop, donor_sites, acceptor_sites, num_charges, mobile_data, pair_index);
}

/* sort possible moves by delta_energy highest to lowest */
qsort(mobile_data, pair_count, sizeof(t_mobile), compare_e);

/* starting with lowest delta_e – if delta_e is negative move, if delta_e is positive use probability*/
for(pair_index=pair_count-1; pair_index>=0; pair_index--){
  donor = mobile_data[pair_index].donor_index;
  acceptor = mobile_data[pair_index].acceptor_index;
  if(mobile_data[pair_index].delta_e <= 0){
    mobile_data[pair_index].probability =1;
    mobile_data[pair_index].hopped = 1;
  }
  else{
    mobile_data[pair_index].probability = exp(-mobile_data[pair_index].delta_e*mobile_prob/kt);
    random_number = gmx_rng_uniform_real(x_rng);
    if(random_number < mobile_data[pair_index].probability){
      mobile_data[pair_index].hopped = 1;
    }
  }
}

/* if hop occurs print positions of all charges to mobile_dist then move charge */
if(mobile_data[pair_index].hopped==1){
  hop_counter++;
}
```c
fprintf(mobile_dist,"%12.3f%3d%3d%3d%3d%s%3d%3d",time,hop_counter,
donor,donor_sites[donor].bead_num,donor_sites[donor].resname,
num_charges);
for(charge=0; charge < num_charges; charge++){
    fprintf(mobile_dist,"%12.3f%12.3f%12.3f",state->x[donor_sites[charge].bead_num][XX],state->x[donor_sites[charge].bead_num][YY],
    state->x[donor_sites[charge].bead_num][ZZ]);
}
fprintf(mobile_dist,"\n");
/* move charge */
md->chargeA[donor_sites[donor].bead_num] -= 1.0;
md->chargeA[acceptor_sites[acceptor].bead_num] += 1.0;
temp_site.resname = donor_sites[donor].resname;
temp_site.resnum = donor_sites[donor].resnum;
temp_site.bead_num = donor_sites[donor].bead_num;
temp_site.e_binding = donor_sites[donor].e_binding;
donor_sites[donor].resname = acceptor_sites[acceptor].resname;
donor_sites[donor].resnum = acceptor_sites[acceptor].resnum;
donor_sites[donor].bead_num = acceptor_sites[acceptor].bead_num;
donor_sites[donor].e_binding = acceptor_sites[acceptor].e_binding;
acceptor_sites[acceptor].resname = temp_site.resname;
acceptor_sites[acceptor].resnum = temp_site.resnum;
acceptor_sites[acceptor].bead_num = temp_site.bead_num;
acceptor_sites[acceptor].e_binding = temp_site.e_binding;
/* recalculate energy for remaining possible moves */
for(m=0; m < pair_index; m++){
    mobile_energy(fplog,state,mtop,donor_sites,acceptor_sites,
    num_charges,mobile_data,m);
}
/* resort possible moves */
qsort(mobile_data,pair_index,sizeof(t_mobile),compare_e);
}

/* if no hop occurs print positions of all charges to mobile_dist, the 0s are
place holders so the output has the same format as when there is
information about a charge hopping*/
if(hop_counter==0){
    fprintf(mobile_dist,"%12.3f%3d%3d%3d%3d%s%3d%3d",time,hop_counter,
    0,0,0,num_charges);
    for(charge=0; charge < num_charges; charge++){
        fprintf(mobile_dist,"%12.3f%12.3f%12.3f",state->x[donor_sites[charge].bead_num][XX],state->x[donor_sites[charge].bead_num][YY],
        state->x[donor_sites[charge].bead_num][ZZ]);
    }
```

Printf(mobile_dist,"\n");
}

/* Print hopping data */
Printf(mobile_out,"%12.3f%.4d%.4d%12.8f%.12.8f%.12.8f%.12.8f%.12.8f%.12.8f\n",time,pair_count,hop_counter);
for(pair_index=0;pair_index<pair_count;pai r_index++){
    donor = mobile_data[pair_index].donor index;
donor_num = donor_sites[donor].bead num;
acceptor = mobile_data[pair_index].acceptor_index;
acceptor_num = acceptor sites[acceptor].bead_num;
if(mobile_data[pair_index].hopped==0){
    Printf(mobile_out,"%.4d%.4d%.4d%.12.8f%.12.8f%.12.8f%.12.8f%.1d\n",donor,donor_num,acceptor_num,mobile_data[
pair_index].delta_bind,mobile_data[pair_index].delta coul,mobile_data[pair_index].delta_e,mobile_data[pair_index].probability,enerd->term[F_TEMP],mobile_data[pair_index].hopped);
}
else{
    Printf(mobile_out,"%.4d%.4d%.4d%.12.8f%.12.8f%.12.8f%.12.8f%.12.8f%.12.8f%.12.8f\n",donor, acceptor_num,donor_num,mobile_data[
pair_index].delta_bind,mobile_data[pair_index].delta coul,mobile_data[pair_index].delta_e,mobile_data[pair_index].probability,enerd->term[F_TEMP],mobile_data[pair_index].hopped);
}
}

/* Print the charged_sites array */
Printf(mobile_config,"%12.3f\n",time);
for(charge=0; charge < num_charges; charge++) {
    Printf(mobile_config,"%4d\n",donor_sites[charge].bead num);
}
Printf(mobile_config,"\n");

/* THE END */
}

/* the mobile_energy function calculates the change in coulomb (plus binding energy) a charge will experience when it moves from its charged site to the basic site for charges in the mobile_data structures*/
void mobile_energy(FILE *fplog,t_state *state,gmx_mtop,t *mtop,t_sites *
donor_sites,t_sites *acceptor_sites,int num_charges,t_mobile *mobile_data,
int pair_index)
{
    real e_coul_old,e_coul_new;
    int donor,acceptor,charge;
    rvec r_donor,r_acceptor,r_dm,r_am;
e_coul_old = 0.0;
e_coul_new = 0.0;
donor = mobile_data[pair_index].donor_index;
acceptor = mobile_data[pair_index].acceptor_index;
/* calculate Coulomb energy for the new and old configurations */
copy_rvec(state->x[donor_sites[donor].bead_num],r_donor);
copy_rvec(state->x[acceptor_sites[acceptor].bead_num],r_acceptor);
for(charge=0; charge < num_charges; charge++){
    if(donor != charge){
        rvec_sub(r_donor, state->x[donor_sites[charge].bead_num], r_dm);
        if(norm(r_dm)<=1e-10){
            e_coul_old += 1e20; /* this is an approximation (if distance between charges is zero the coulomb energy should go to infinity) */
        }
        else{
            e_coul_old += 138.9354558/norm(r_dm); /* units kJ/mol (r_lm is in nm) */
        }
    }
    else{
        rvec_sub(r_acceptor, state->x[acceptor_sites[charge].bead_num], r_am);
        if(norm(r_am)<=1e-10){
            e_coul_new += 1e20; /* this is an approximation (if distance between charges is zero the coulomb energy should go to infinity) */
        }
        else{
            e_coul_new += 138.9354558/norm(r_am); /* units kJ/mol (r_lm is in nm) */
        }
    }
}
/* calculate changes in energy */
mobile_data[pair_index].delta_coul = e_coul_new-e_coul_old;
mobile_data[pair_index].delta_bind = acceptor_sites[acceptor].e_binding-
    donor_sites[donor].e_binding;
mobile_data[pair_index].delta_e = mobile_data[pair_index].delta_coul+mobile_data[pair_index].delta_bind;
}
/* the qsort function (standard library function) sorts the elements of mobile_data by energy using compare_e */
int compare_e(const void *a, const void *b){
    t_mobile *ia = (t_mobile *)a;
    t_mobile *ib = (t_mobile *)b;
    if(ia->delta_e > ib->delta_e){
        return -1;
    }
}
if (ia->delta_e < ib->delta_e) {
    return 1;
} else return 0;

A.2 mobile_charge.h

/*
 * Header to go with mobile_charge.c
 */

#ifndef _mobile_charge_h
#define _mobile_charge_h

#include <stdio.h>
#include "typedefs.h"

typedef struct {
    int donor_index;
    int acceptor_index;
    real mobile_distance;
    real delta_coul;
    real delta_bind;
    real delta_e;
    real probability;
    int hopped;
} t_mobile;

typedef struct {
    char *resname;
    int resnum;
    int bead_num;
    real e_binding;
} t_sites;

extern void mobile_charge(FILE *fplog, t_mdatoms *md, t_state *state, gmx_mtop_t *mtop, t_commrec *cr, t_inputrec *ir, gmx_enerdata_t *enerd, int step);

extern void mobile_energy(FILE *fplog, t_state *state, gmx_mtop_t *mtop, t_sites *charged_sites, t_sites *basic_sites, int num_charges, t_mobile *mobile_data, int pair_index);

extern int compare_e(const void *, const void *);

#endif