Computer simulations of small molecules and proteins in solution

by

Nilusha Lakmali Kariyawasam Manachchige

B.S., University of Colombo, 2011

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Chemistry
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2019

Abstract

Computer simulations have been widely used in studying macromolecular systems due to a rapid increase in computer power. These simulations allow one to explore the structure, function and dynamics of biomolecules at atomistic level details and to predict unknown molecular properties. The accuracy of a computer simulation is mainly determined by the quality of the force field and the degree of sampling achieved during a simulation. Furthermore, the accuracy of calculated properties or results will depend on the methodology used to calculate these properties.

Most force fields are developed by fitting the bonded and non-bonded interaction parameters to the quantum mechanically or experimentally obtained data. In contrast, our effort to develop a simple, classical, non-polarizable, force field is based on fitting parameters, especially the partial atomic charges, to reproduce Kirkwood-Buff integrals (KBIs) for solution mixtures. Kirkwood-Buff (KB) theory is a theory of solution mixtures that can be applied to solutions with any number of molecules, regardless of their size and complexity. This theory allows us to obtain the correct balance between the solute-solute and solute-solvent interactions. A Kirkwood-Buff derived force field for polyols in solution will be discussed.

Fluctuation solution theory (FST) is an extension of KB theory which provides information regarding the local composition of solutions, or the deviation of local composition from bulk solution. The KBIs can be expressed in terms of particle number fluctuations and this allows us to calculate the KBIs without integrating the pair correlation function. A FST approach is used to calculate the partial molar volume and compressibility of proteins at infinite dilution without any subjective definitions of the protein volume and compressibility. These properties are solely determined using the solvent/water fluctuations in the presence and absence of the protein.

Furthermore, residue-based contributions to these properties are also available and are calculated. The results are compared among different proteins and force fields to establish trends.

Pressure perturbation is a powerful technique to study the hydration of macromolecules. Molecular dynamics techniques are used to identify the effect of pressure on the conformations of LacI and some variants of LacI. The *lac* repressor protein (LacI) is the regulatory unit of *lac* operon and it binds to the target site of the operon to repress the transition of the genes. The mutations studied here correspond to an experimentally known rheostat position, and we attempt to correlate the changes in activity for different mutants with the corresponding hydration changes.

Computer simulations of small molecules and proteins in solution

by

Nilusha Lakmali Kariyawasam Manachchige

B.S., University of Colombo, 2011

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Chemistry
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2019

Approved by:

Major Professor
Professor Paul E. Smith

Copyright

Abstract

Computer simulations have been widely used in studying macromolecular systems due to a rapid increase in computer power. These simulations allow one to explore the structure, function and dynamics of biomolecules at atomistic level details and to predict unknown molecular properties. The accuracy of a computer simulation is mainly determined by the quality of the force field and the degree of sampling achieved during a simulation. Furthermore, the accuracy of calculated properties or results will depend on the methodology used to calculate these properties.

Most force fields are developed by fitting the bonded and non-bonded interaction parameters to the quantum mechanically or experimentally obtained data. In contrast, our effort to develop a simple, classical, non-polarizable, force field is based on fitting parameters, especially the partial atomic charges, to reproduce Kirkwood-Buff integrals (KBIs) for solution mixtures. Kirkwood-Buff (KB) theory is a theory of solution mixtures that can be applied to solutions with any number of molecules, regardless of their size and complexity. This theory allows us to obtain the correct balance between the solute-solute and solute-solvent interactions. A Kirkwood-Buff derived force field for polyols in solution will be discussed.

Fluctuation solution theory (FST) is an extension of KB theory which provides information regarding the local composition of solutions, or the deviation of local composition from bulk solution. The KBIs can be expressed in terms of particle number fluctuations and this allows us to calculate the KBIs without integrating the pair correlation function. A FST approach is used to calculate the partial molar volume and compressibility of proteins at infinite dilution without any subjective definitions of the protein volume and compressibility. These properties are solely determined using the solvent/water fluctuations in the presence and absence of the protein.

Furthermore, residue-based contributions to these properties are also available and are calculated. The results are compared among different proteins and force fields to establish trends.

Pressure perturbation is a powerful technique to study the hydration of macromolecules. Molecular dynamics techniques are used to identify the effect of pressure on the conformations of LacI and some variants of LacI. The *lac* repressor protein (LacI) is the regulatory unit of *lac* operon and it binds to the target site of the operon to repress the transition of the genes. The mutations studied here correspond to an experimentally known rheostat position, and we attempt to correlate the changes in activity for different mutants with the corresponding hydration changes.

Table of Contents

List of Figures xi
List of Tables xvii
Acknowledgements xviii
Dedication xx
Chapter 1 - Introduction 1
1.1 Molecular Dynamics Simulations 1
1.1.1 Force Fields 4
1.2 Protein Denaturation 9
1.2.1 Pressure denaturation 11
1.2.1.1 Volume of a protein 12
1.2.1.2 Compressibility of a protein 14
1.3 The Kirkwood-Buff Theory of Solutions 16
1.4 Inversion of the Kirkwood-Buff theory 19
1.4.1 Fluctuation solution theory 20
1.5 Application of KB/FST to protein denaturation 20
1.5.1 Protein denaturation thermodynamics 21
1.6 Organization of dissertation 25
1.7 References 26
Chapter 2 - A Kirkwood-Buff Force Field for Simulations of Polyols 34
2.1 Introduction 34
2.2 Methodology 38
2.2.1 Kirkwood-Buff Analysis of the Experimental Data 38
2.2.2 Experimental Sources for Composition and Activity Data 40
2.2.3 Parameter Development 41
2.2.4 Molecular Dynamics Simulations 41
2.2.5 Analysis of simulated solution properties 43
2.3 Results and Discussion 45
2.4 Conclusions 59
2.5 References 59

Chapter 3 - A New View of Protein Compressibility 65
3.1 Introduction 65
3.1.1 Theory 69
3.1.2 Test systems 73
3.2 Methodology 74
3.2.1 Molecular dynamics simulations 74
3.2.2 Analysis 76
3.3 Results and discussion 77
3.4 Conclusions 88
3.5 References 88
Chapter 4 - Simulated Amino Acid Volumes and Compressibilities -A Force Field Comparison 92
4.1 Introduction 92
4.1.1 Theory 96
4.1.2 Ubiquitin 98
4.2 Methodology 99
4.2.1 Molecular dynamics simulations 99
4.2.2 Analysis 100
4.3 Results and Discussion 102
4.4 Conclusions 114
4.5 References 115
Chapter 5 - The Effects of Pressure on the Conformations of the *Lac* Repressor 119
5.1 Introduction 119
5.2 Methodology 124
5.2.1 System setup 124
5.2.2 Molecular dynamics simulations 125
5.2.3 Analysis 126
5.2.4 Pressure thermodynamics 127
5.3 Results and Discussion 129
5.4 Conclusions 138
5.5 Supporting information 139

5.6 References 142
Chapter 6 - Probing Hydration Changes in *Lac* Repressor Mutants 145
6.1 Introduction 145
6.2 Methodology 150
6.2.1 System setup 150
6.2.2 Molecular Dynamics simulations 150
6.2.3 Analysis 151
6.3 Results and discussion 152
6.4 Conclusions 160
6.5 Supporting Information 161
6.6 References 163
Chapter 7 - Conclusions and Future Work 166
Appendix A - Rationally Designed Peptide Nanosponges for Cell-Based Cancer Therapy 168
Appendix B - SLTCAP: A Simple Method for Calculating the Number of Ions Needed for MD Simulation 179
Appendix C - Peptide Nanosponges Designed for Rapid Uptake by Leucocytes and Neural Stem Cells 185
Appendix D - Copyright Clearance 195

List of Figures

Figure 2.1 Glycerol model with atom labels used for Bastiansen nomenclature..... 36
Figure 2.2 Experimental (lines) and simulated (points) data for the KBIs of EDL/HOH and EDL/MOH mixtures as a function of composition at 298 K. Error bars are the standard deviations calculated for 5 ns block averages of simulations. 46
Figure 2.3 Experimental (lines) and simulated (points) data for the KBIs of GLY/HOH and GLY/EOH mixtures as a function of composition. Two sets of experimental data obtained from two different set of activity data are shown for the GLY/HOH mixture at 298 K and 293 K. Error bars are the standard deviation calculated for 5 ns block averages of simulations. 47
Figure 2.4 Experimental (lines) and simulated (points) data for the KBIs of 1,2-PDL/HOH and 1,3-PDL/HOH mixtures, as a function of composition at 298 K. Error bars are the standard deviation calculated for 5 ns block averages of simulations..... 48
Figure 2.5 Experimental[32, 35, 51] (lines) and simulated (points) enthalpies of mixing for the six binary mixtures as a function of composition at 298 K. No experimental data were found for the GLY/EOH mixture. Error bars are the standard deviation calculated for 5 ns block averages of simulations. All the experimental data correspond to 298 K..... 49
Figure 2.6 Experimental (lines) and simulated (points) excess volumes of mixing for six binary mixtures as a function of composition at 298 K. 51
Figure 2.7 Experimental (lines) and simulated (points) translational self-diffusion coefficients (D_c and D_s) for GLY/HOH[52] and EDL/HOH[53] mixtures as a function of composition at 298 K..... 52
Figure 2.8 Experimental (lines) and simulated (points) dielectric constants for polyol/HOH mixtures as a function of composition at 298 K. Experimental dielectric constants in the full range of composition were only found for the EDL/HOH[70] mixture. For other systems experimental dielectric constants of the pure liquids are shown as triangles. 54
Figure 2.9 Simulated relative free energies of four different torsion angles of the pure GLY and two different torsion angles of the pure EDL at 298 K..... 55

Figure 2.10 Simulated conformational populations as a function of composition for four different torsion angles of GLY and two different torsion angles of EDL at 298 K. 56
Figure 3.1 Cartoon representation of native Lysozyme (PDB ID: 4LZT), Ribonuclease (PDB ID: 2AAS), and Ubiquitin (PDB ID: 1UBQ) colored by its secondary structure elements. Helix-red, sheet-yellow, hairpin-blue, coil-green. Images were generated with PyMOL molecular visualization software.[42] 73
Figure 3.2 Left: Water probability distribution around protein (g_{21}) as a function of distance from the surface of the free proteins at 1 bar. Right: pseudo protein volume as a function of integration distance at 1 bar for the free Lysozyme, RNaseA, and ubiquitin. 78
Figure 3.3 Pseudo volume of free, all atom position restrained (PR), and all atom position restrained with charged group neutralized (PR-Q) versions of the proteins as a function of pressure. Dots represent the raw volumes obtained by averaging the distant dependent volume, while lines represent the pressure fitted volumes. 79
Figure 3.4 Median, maximum, and minimum residue volumes after considering all the residues in free Lysozyme, RNaseA, and Ubiquitin at 1 bar. Midde horizontal line: median residue volume, upper horizontal line: maximum residue volume, lower horizontal line: minimum residue volume. HIS only shows a single value since there is only one HIS present in all three proteins. 1 nm= 602 cm^3 mol^{-1} 80
Figure 3.5 Pseudo compressibilities of free, all atom position restrained (PR), and all atom position restrained with charged group neutralized (PR-Q) versions of the proteins at 1 bar. Experimental partial molar compressibilities are also shown for pure water[8], Lysozyme[24], and RNaseA[24]. The pseudo compressibility of pure TIP3P water is also shown. 81
Figure 3.6 Pseudo compressibility contributions from positively charged (+ve), negatively charged (-ve), polar (P), nonpolar (NP) side chains and total backbone (BB) to the free proteins at 1 bar. 83
Figure 3.7 Water distribution around selected side chain atom (g_{i1}) of residues chosen to represent positively charged, negatively charged, polar, and nonpolar residues at 1 bar, 3 kbar, 6 kbar, and 10 kbar for ubiquitin. Residue name, the sequence number, and the atom type are denoted in labels. Enlarged second and third solvation shells are shown as insets for the top two figures. 84

Figure 3.8 Pseudo compressibility contributions from positively charged (+ve), negatively charged (-ve), polar (P), nonpolar (NP) side chains and total backbone (BB) to the free proteins as a function of volume fraction (ϕ) at 1 bar. The percent contributions from each group to the total compressibility are indicated. Total compressibility of each protein is given by the y-value of red shaded area and the contributions from each group is given by y-values of grey shaded bars. 86
Figure 3.9 The median, maximum, and minimum pseudo compressibilities of residues after considering all the residues in free Lysozyme, RNaseA, and Ubiquitin at 1 bar. Middle horizontal line: median residue compressibility, upper horizontal line: maximum residue compressibility, lower horizontal line: minimum residue compressibility. HIS only shows a single value since there is only one HIS present in all three proteins..... 87
Figure 4.1 Cartoon representation of the native Ubiquitin (PDB ID: 1UBQ) colored by its secondary structure elements. Helix-red, sheet-yellow, hairpin-blue, coil-green. The image was generated with PyMOL molecular visualization software.[44]..... 99
Figure 4.2 The C_α RMSDs calculated for first 70 residues of Ubiquitin with AMBER FF at 1 bar, 1 kbar, 2 kbar, and 3 kbar. All the simulations were performed at 300 K temperature..... 102
Figure 4.3 Water probability distribution around protein as a function of distance from the surface of the protein, Top: at 1 bar, 1 kbar, 2 kbar, and 3 kbar for AMBER FF. Bottom: at 1 bar for AMBER, CHARMM, GROMOS, and OPLS FFs. The line corresponds to AMBER FF is underneath the line correspond to OPLS FF and not visible. 103
Figure 4.4 Pseudo protein volume as a function of integrtion distance Top: at 1 bar, 1 kbar, 2 kbar, and 3 kbar. Bottom: at 1 bar for AMBER, CHARMM, GROMOS, and OPLS FFs. 104
Figure 4.5 Pseudo volume of Ubiquitin as a function of pressure for AMBER, CHARMM, GROMOS, and OPLS FFs. 105
Figure 4.6 Pseudo volume of polar, nonpolar, positively charged, and negatively charged groups as a function of pressure for AMBER, CHARMM, GROMOS, and OPLS FFs. Here, the volume of each group was obtained by summing over the residue volumes in each group and it should be noted that the different groups have a different number of residues as denoted within parenthesis. 106
Figure 4.7 Pseudo volume of the helix, sheet, hairpin, and coil groups as a function of pressure for AMBER, CHARMM, GROMOS, and OPLS FFs. Here, the volume of each group was

obtained by summing over the residue volumes in each group and it should be noted that the different groups have a different number of residues as denoted within parenthesis. 108
Figure 4.8 Simulated vs. experimental pseudo volumes of 76 residues in Ubiquitin at 1 bar. Experimental residue volumes were not obtained for Ubiquitin but determined from apparent molar volumes of amino acids at 298 K and 1 bar. 109
Figure 4.9 Average pseudo volumes of each amino acid residue present in Ubiquitin as a function of pressure. The total number of each residue in Ubiquitin is shown in parenthesis. The dashed line represents the experimental residue values at 1bar and used as a guideline to compare results. Experimental residue volumes were not obtained for Ubiquitin but determined from apparent molar volumes of amino acids at 298 K and 1 bar. 110
Figure 4.10 Pseudo compressibility of Ubiquitin as a function of pressure for AMBER, CHARMM, GROMOS, and OPLS FFs. 111
Figure 4.11 Pseudo compressibilities of polar, nonpolar, positively charged, and negatively charged groups as a function of pressure for AMBER, CHARMM, GROMOS, and OPLS FFs. The number of residues in each group is shown within the parenthesis. 112
Figure 4.12 Pseudo compressibilities of the helix, sheet, hairpin, and coil groups as a function of pressure for AMBER, CHARMM, GROMOS, and OPLS FFs. The number of residues in each group is shown within the parenthesis. 114
Figure 5.1 Cartoon representation of the LacI dimer (PDB id: 1EFA) bound to operator and anti-inducer, ONPF. Chain A of the dimer is shown in green and chain B in red. C-terminal residues are not shown here. The image was generated with PyMOL molecular visualization software.[13] 120
Figure 5.2 Cartoon representation of the core domains of the DNA bound (pink, PDB id: 1EFA) LacI after fitting to the C_α carbons in the C-subdomains of DNA unbound (green, PDB id:1LBH) structure. The image was generated with PyMOL molecular visualization software. 121
Figure 5.3 Schematic representation of the two vectors defined by connecting N and C subdomains of each chain. 126
Figure 5.4 The distance defined between two β-strands in chain A (green) and chain B (red) in the N subdomains of the LacI bound crystal structure is shown using a red double arrow

(only the core domains of the structure are shown). Image was generated using PyMOL molecular visualization software.[13] 127
Figure 5.5 Calculated RMSDs for the WT (a) RMSDs calculated for each individual subdomain after fitting to the same domain of the crystal structure at 1 bar (b) RMSDs for each individual subdomain at 3 kbar (c) RMSDs of N - subdomains after fitting to the C - subdomains of bound crystal structure (black) and unbound crystal structure (red) at 1bar (d) RMSDs of N - subdomains after fitting to the C- subdomains of bound (black) and unbound (red) crystal structures at 3 kbar. 130
Figure 5.6 Calculated vector angle between chain A and chain B of LacI dimer. (a) angle for WT at 1 bar (black) and 3 kbar (red) (b) angle for WT- headPR at 1 bar (black) and 3 kbar (red) (c) angle for WT/ONPF at 1 bar (black) and 3 kbar (red) (d) angle for WT-headPR/ONPF at 1 bar (black) and 3 kbar (red). Dashed lines represent the angle calculated for bound (lower dashed line) and unbound (upper dashed line) crystal structures. 131
Figure 5.7 Distance between the C_α atoms of residue 96 in chain A and chain B of WT at 1 bar (black) and 3 kbar (red). Dashed lines represent the distance obtained from the bound (upper dashed line) and unbound (lower dashed line) crystal structures. 133
Figure 5.8 RMSD from an ideal helix for HH in chain A (black) and chain B (red) of WT, and for HH alone (green) (a) at 1 bar (b) at 3 kbar. 134
Figure 5.9 Water probability distribution around selected residues (g_{iw}) as function of distance in the bound (black) and the unbound (red) state of the WT. Top two rows of the figure show the residues in chain A, while the bottom two rows show the corresponding residues in chain B. 137
Figure 5.10 Average number of water molecules within 5 Å of Val 52 and Ala 53 (a) at 1 bar (b) at 3 kbar as a function of time for the WT. 138
Figure 6.1 Cartoon representation of the LacI dimer (PDB id: 1EFA) bound to operator and anti-inducer, ONPF (blue). Chain A of the dimer is shown in green and chain B in red. C-terminal residues are not shown here. Residue 52 in both chains are colored in yellow. The image was generated with PyMOL molecular visualization software.[19] 147
Figure 6.2 Experimental β-galactosidase activity for 52nd position of LacI protein. The first black bar represents the repression in the absence of LacI protein. Bellow 13 Miller units (black solid line), any change in repression altered bacterial growth. The red dashed line

represents the activity for WT protein. The front series show the activity in the absence of effector molecule while the gray series shows the activity in the presence of effector. Error bars correspond to standard deviation of 2-4 bacterial colonies. The 'LacI-11' indicates that the LacI is just a dimer after deleting the C terminal tetramerization domain. This figure was reproduced from 'Meinhardt S, Manley MW Jr, Parente DJ, Swint-Kruse L (2013) Rheostats and Toggle Switches for Modulating Protein Function. PLoS ONE 8(12): e83502. https://doi.org/10.1371/journal.pone.0083502' This is an open-access article distributed under the terms of the Creative Commons Attribution License. 149

Figure 6.3 Calculated angle vector between chain A and chain B for WT, V52I, V52A, V52Q, and V52G LacI dimer. The angle at 1 bar is shown in black and at 3 kbar is shown in red. Dashed lines represent the angle calculated for the bound (lower dashed line) and unbound (upper dashed line) crystal structures..... 154

Figure 6.4 Distance between the C-alpha atoms of residue 96 in chain A and chain B of WT, V52I, V52A, V52Q, and V52G at 1 bar (black) and 3 kbar (red). Dashed lines represent the distance obtained from the bound (upper dashed line) and unbound (lower dashed line) crystal structures. 156

Figure 6.5 Schematic representation of distance vs angle measurements at 1 bar (left) and 3 kbar (right). Red squares represent the bound and unbound states determined according to the DNA bound and unbound crystal structures. 157

Figure 6.6 RMSD from an ideal helix for HH in chain A (black) and chain B (red) of WT, V52I, V52A, V52Q, and V52G at 1 bar (left column) and 3 kbar (right column). 158

Figure 6.7 Water probability distribution around selected residues (g_{iw}) as a function of distance in the bound (black) and the unbound (red) state of V52A at 1 bar. Top two rows of the figure show the g_{iw} for residues in chain A, while the bottom two rows show the g_{iw} for the corresponding residues in chain B. The bound and unbound states were determined only considering the angle data. 159

List of Tables

Table 2.1 Torsion angle parameters obtained for the KBFF models. 42
Table 2.2 Nonbonded parameters obtained for the KBFF models 43
Table 2.3 Experimental and simulated properties of pure liquids. All the simulations were performed at 298 K and experimental values were obtained from literature at 298 K unless stated otherwise. 53
Table 2.4 Comparison of QM[10] and simulated relative energies obtained for ten different conformations of pure EDL. 57
Table 2.5 Comparison of experimental[41] and simulated GLY backbone conformations. 57
Table 3.1 Number of residues in each physicochemical (polar, non-polar, positively charged, negatively charged) and secondary structure (helix, sheet, hairpin, coil) groups and their percentages for Lysozyme, RNaseA, and Ubiquitin. 74
Table 5.1 Details of the systems simulated. All the systems were simulated at 300 K temperature and 1 bar and 3 kbar pressures. Production runs were continued up to 1μs. 125
Table 5.2 Volume differences between the bound and unbound states ($\Delta V_{unbound\text{-}bound}$) for the WT residues at 1 bar. Residues in both chains are included. 135

Acknowledgements

First and foremost, I would like to extend my sincere gratitude towards my Ph.D. advisor Professor Paul E. Smith for his endless guidance and support throughout my graduate studies. This dissertation would not have been possible without your continuous support. It has been a great honor and pleasure to work with you as a graduate student. I really appreciate your valuable time, ideas and comments you provided me for all my research projects.

Secondly, I would like to express my sincere thanks to my advisory committee, Professor Daniel Higgins, Professor Christopher Culbertson, Professor Jeremy Schmit, and Professor Ruth Welti for their comments, suggestions and valuable time. Also, I would like to thank Professor Ryszard Jankowiak who was one of my former committee members for the last five years for his support and valuable time.

Also, I would like to thank my research collaborators, Professor Stefan H. Bossmann, Professor Jeremy Schmit, and Professor Liskin Swint-Kruse for their input and support.

Especially I would like to extend my sincere thanks to all the past and present members in Smith group, including Dr. Elizabeth A. Ploetz, Dr. Gayani Pallewela, Dr. Sadish Karunaweera, Dr. Nawavi Naleem, and Sandun Gajaweera for sharing knowledge with me and all the support provided me throughout my graduate studies.

I would also like to express my sincere gratitude towards my undergraduate advisors Professor Nalin De Silva, Professor Rohini De Silva, and Professor Samantha Weerasinghe in the University of Colombo, Sri Lanka for their guidance, support, and encourage me to peruse graduate studies.

I would like to thank Kansas State University, Department of Chemistry for accepting me as a graduate student and funded me. Also, I would like to thank all the academic and non-

academic staff including Ms. Mary Dooley, Ms. Kimberly Ross, Ms. Kathy Roeser, and Mr. Tobe Eggers for their support throughout my graduate studies in Department of Chemistry. I also would like to thank the staff in the international student center for their advice and support.

I sincerely thanks to all my friends who live in Manhattan for their support in many ways and the great time we shared together. Also, I would like to thank all my friends and relatives back in Sri Lanka.

Finally, I am immensely grateful to my loving mother, father, brother, and husband for their endless love, care, strength, and support and my loving daughter for filling our lives with love and happiness. This would not be possible without you.

Dedication

To my dearest mother, father, and brother

For their endless love, care, and support

To my loving husband, Irosha Wanithunga

For his love, strength and for being there for me always

To my loving daughter, Ranudi Wanithunga

For filling our lives with love and happiness

Chapter 1 - Introduction

1.1 Molecular Dynamics Simulations

Computer simulation studies can be used as an alternative method to performing experiments and has become a major area of research in Chemistry. It is difficult, risky and often expensive to perform experiments under extremes of temperature and pressure.[1-2] For example, to study the dangerous processes involved with nuclear reactors one can easily use computer simulations under the necessary conditions. Computer simulations can be performed at high or low temperature/pressure conditions as necessary. Furthermore, with the aid of computer simulations we can model and study the structures and the properties of systems that are yet to be realized.[3] This allows for the saving time and costs in chemical laboratories.

Computational approaches are widely used to study biomolecules as they allow us to discover many atomistic level details that are essential to revealing the activity and function of these macromolecules. The design and development of novel drugs is another major field in computational chemistry that has a huge impact on the pharmaceutical industry.[4] Furthermore, computer simulations allow us to study cause-effect relationships where one can change one parameter at a time in a model system, without affecting other parameters, to investigate the relationships between different properties.[5] In experiments, it is generally not possible to change individual parameters without disturbing the others.[5]

Molecular dynamics (MD) is one of the main simulation methods and this method has been used to study the behavior of a variety of different molecular systems such as liquids, solution mixtures, and proteins. Biomolecules are dynamical in nature and their dynamics range from the picosecond to second time scale.[6] MD simulations can provide the motion of atoms as a function

of time which is not possible with most experimental techniques.[5] This can then be used to calculate the structural and dynamical properties of these molecules. The major constraint of computer simulations is that it is generally approximate (see below).[7] Therefore, experimental data are still required to validate the results obtained from simulations, and to improve the simulation techniques.

Classical MD simulation provides the ability to model a system at the atomistic level. Here, the atoms are generally treated as the smallest unit of a molecule and electrons are not treated explicitly. Ignoring the electrons allows us to simulate larger biomolecular systems over longer time scales. MD uses Newton's equations of motion to calculate the forces acting on the atoms of the system and therefore their corresponding motion.[8]

According to Newton's second law, the forces and accelerations for a system consisting of N particles interacting with a potential energy $U(\vec{r_i})$, where i =1, 2… N, is given by,

$$\vec{F_i} = m_i \ddot{\vec{r}}_i \tag{1.1}$$

where forces are,

$$\vec{F}_i = -\frac{\partial U(\vec{r}_i)}{\partial \vec{r}_i} \tag{1.2}$$

Here, m_i, $\vec{r_i}, \ddot{\vec{r}}_i$ are the mass, position and second derivative of position with respect to time, respectively, for each particle.

The Verlet algorithm is widely used to calculate the atom position $\vec{r_i}$ at time (t+Δt) when the position and acceleration at time t and position at time (t-Δt) are known.[9]

$$\vec{r}_i(t+\Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t-\Delta t) + \ddot{\vec{r}}_t(\Delta t)^2 \tag{1.3}$$

The velocity can then be calculated using the following formula.

$$\vec{v}_i(t) = \frac{\vec{r}_i(t+\Delta t) - \vec{r}_i(t-\Delta t)}{2\Delta t} \tag{1.4}$$

Choosing an appropriate integration time step (Δt) is important as a too large a time step will lead to unstable system behavior due to large fluctuations and drifts in energy. Small time steps will increase the accuracy of the solution for the equations of motion, however one may need a large number of steps to achieve similar length simulations. This means that every simulation involves a trade-off between computational cost and integration accuracy. The main factor determines the size of the time step is the highest frequency motion associated with the system of interest. For most systems the highest frequencies arise from the stretching of bonds connecting hydrogen to heavy atoms (H-X stretching) and time step of ≈ 0.5 fs is required for these types of systems.[10] It is possible to remove motion associated with bond vibrations. The SHAKE algorithm is one of the most common constraint techniques used in biomolecular simulations to fix the H-X bond length, allowing for a time step of 2 fs to integrate the equations of motion.[10-11]

To perform a MD simulation an initial set of atomic coordinates for the system of interest are required. These can be obtained by x-ray crystallography, NMR experiments, or by model building.[6] Once the initial coordinates are available, the system is usually refined using energy minimization to remove any bad contacts. Then velocities are assigned to the atoms and the dynamics simulated for about 100 ps to equilibrate the system. Equilibration is performed in order to make sure that the system is stable and there are no irregular fluctuations.[10] Once the system is equilibrated, at the desired temperature and pressure, a production run is then performed. The length of the production run depends on the size of the system and the properties we are interested in studying. For biomolecules, long simulation times are usually required as they can adopt many different conformations and display slow dynamics. Lengthy simulations will allow

macromolecules to explore more of the possible configurations. Finally, the simulated trajectories corresponding to the production run are used to analyze the system. The simulated trajectory contains all the information – such as atomic positions, velocities and forces as a function of time – which are needed to calculate thermodynamic (heat capacity, enthalpy, density, compressibility etc), dynamical (self-diffusion coefficient) or structural properties (radial distribution function) of the system.[12]

The success of a computer simulation of a chemical system depends mainly on the accuracy of the force field and the degree of sampling achieved.[1] Whether we have achieved sufficient sampling is determined by how much of the important parts of configuration space have been sampled.[13] However, sampling all the possible configurations of a protein is impractical due to their size and complexity. To overcome this problem sampling techniques are often used. High temperature MD simulations are the most common way of overcoming limited sampling. Replica Exchange Molecular Dynamics (REMD) is widely used to enhance the sampling of biomolecule states.[14-16] In this method, replicas of the system are simulated independently and simultaneously at different temperatures.[11] This method allows non interacting replica pairs to exchange at different temperatures every few thousand steps, and this allows the system to relax faster than a single constant temperature MD simulation.[14, 16] Even though REMD simulations enhance the relaxation in biomolecules, it becomes computationally very expensive when the number of replicas and the size of the system increases.

1.1.1 Force Fields

The force field (FF) is the major aspect of a MD simulation. A FF can be defined as a set of equations and parameters used to determine the potential energy of a system of interest. These

mathematical functions consider atoms are the smallest unit in a molecule rather than the electrons and nuclei used in quantum mechanics.[10] Moreover, these mathematical functions include the terms that describe the structure and dynamic properties of biological molecules.[10] The quality of the FF is considered extremely important as it determines the accuracy of a computer simulation given sufficient sampling.

The total potential energy of a system can be written as a series of intramolecular terms, representing the covalently bonded atoms, and a series of intermolecular terms representing the nonbonded interactions between atoms.[4] The intramolecular interactions include bond stretching, bond rotation and angle bending terms while the nonbonded intermolecular interactions contain the van der Waals and electrostatic interactions. A typical FF contains the following terms,

$$V_{bonds} = \sum_{bonds} \frac{1}{2} K_b (b - b_0)^2 \tag{1.5}$$

$$V_{angles} = \sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_0)^2 \tag{1.6}$$

$$V_{properdihedrals} = \sum_{torsions} K_\phi \left[1 + \cos(\mathrm{n}\phi - \delta)\right] \tag{1.7}$$

$$V_{improperdihedrals} = \sum_{improper} \frac{1}{2} K_\varphi (\varphi - \varphi_0)^2 \tag{1.8}$$

$$V_{vanderWaals} = \sum_{i,j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{1.9}$$

$$V_{electrostatics} = \sum_{i,j} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{1.10}$$

Equations (1.5-1.8) include parameters, b_0, θ_0, and φ_0, corresponding to the equilibrium bond length, equilibrium angle and equilibrium improper dihedral angle, respectively.

Furthermore, K_b, K_θ, K_ϕ, and K_φ are the force constants associated with bond, angle, proper dihedral, and improper dihedral terms, respectively. The proper dihedral term includes the additional parameters n and δ, which are the multiplicity and the phase shift, respectively.

The nonbonded interactions are very important in the simulation of biological molecules as they include the environmental effects on the properties of a molecule. The nonbonded interactions are typically modeled by a Lennard Jones potential and Coulombic term. The Lennard Jones term, which represents the van der Waals interactions (equation 1.9), includes parameters for the magnitude of the favorable (dispersion) interaction between atoms i and j, $\varepsilon_{ij,}$ given the distance between atom i and j, r_{ij}; and the distance at which the interatomic interaction energy is zero, σ_{ij}. The $1/r^{12}$ term, the repulsive term represents the electron cloud repulsion due to the close overlap of atoms as explained by the Pauli exclusion principle. The power of 12^{th} indicates the strong distance dependence of the repulsion. The attractive $1/r^6$ term describes the long-range London's dispersion interactions, or instantaneous induced dipole-dipole interactions. In general, the ε_{ij} and σ_{ij} are not determined for every possible atom pair. These parameters are determined for the individual atom types (ε_{ii} and σ_{ii}) and then combination rules are used to calculate ε_{ij} and σ_{ij} between atom pairs. Typically, Lorentz–Berthelodt rules are used to calculate these parameters where the ε_{ij} is calculated via the geometric mean and the σ_{ij} is calculated via arithmetic mean as follows.[17]

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \tag{1.11}$$

$$\sigma_{ij} = \frac{(\sigma_{ii} + \sigma_{jj})}{2} \tag{1.12}$$

If someone wishes to transfer parameters between two different FFs, it is necessary to check whether both FFs use the same combination rules, as some FFs use the geometric mean for both

parameters.[17] It is not recommended to transfer parameters between FFs which use different combination rules.

The Coulombic term, which represents the electrostatic interaction (equation 1.10), depends on the partial atomic charges q_i, q_j, and the distance between atom pair, r_{ij}. In general, partial atomic charges are obtained from QM calculations in gas phase. However, these gas phase charges do not include any polarization effects in the condensed phase. Polarization effects are the changes in the electron density of molecules in the condensed phase due to the presence of intermolecular interactions. To account for this effect in biomolecular systems, optimized partial atomic charges specifically aimed to overestimate the dipole moments of small molecules are often used.[17-18] As an improvement, the explicit treatment of polarization has been included using induced dipole models or fluctuation charge models as most common approaches.[17] As an alternative to the explicit treatment of polarization effect, empirically optimized partial atomic charges to reproduce Kirkwood-Buff (KB) integrals have also been used.[19-28] This simple approach implicitly include the polarization effect, as the parameterization is performed to reproduce liquid mixture properties.[18]

The ability of the parameters to reproduce the experimental or quantum mechanically obtained target data determine the accuracy of the results of computational studies of biological systems.[3] The FFs available for biomolecular simulations can be all atom (AA), united atom (UA) or coarse grained (CG). In AA FFs all the atoms are treated explicitly, while in UA only the heavy atoms and the polar hydrogen atoms are represented explicitly. The remaining hydrogens are attached to carbon atoms in UA FF approach. One can sometimes choose the UA over the AA approach if one needs to have a large sampling of conformational space more efficiently. In the CG FF approach a few atoms are treated as a single particle and this model further enhances the

efficiency of simulations of macromolecules, thereby extending the time scales possible. Some of the most popular examples for AA FFs are AMBER[29], CHARMM[30], and OPLS[31] while GROMOS[32] and KBFF[21] are classic UA FFs. Martini[33] is a popular CG FF, where in this model four heavy atoms of a molecule or four water molecules are treated as a single bead which greatly reduces the number of degrees of freedom in the system. Most of these FFs are evolving with time as the developers optimize the parameters for particular systems and there are many versions of each FF are available. CHARMM22[34-35], CHARMM36[36-37]; AMBER94[38], AMBER96[39], AMBER99[40]; GROMOS96[41], GROMOS45A3[42], GROMOS53A5[32], GROMOS53A6[32] are some examples for different versions of common FFs.

Furthermore, FFs can be divided into two groups as polarizable and non-polarizable FFs, also known as additive and non-additive FFs. In non-polarizable FFs, polarization effects are included implicitly by using effective partial atomic charges, whereby in polarizable FFs the polarizability is explicitly treated by adding an extra term to the potential energy function that represents the energy associated with polarization of the charge distribution.[17] The major advantage reported for polarizable FFs is the improved accuracy achieved by the redistribution of the charges in response to variations in the local environment.[43] A major disadvantage of the polarizable FFs is the computational cost due to the use of a complex potential energy function. Therefore, nonpolarizable FFs are much popular for simulations of macromolecular systems. However, progress has been made for the application of fully polarizable parameters for biomolecular systems.[43-44] Some examples of polarizable FFs include the AMOEBA FF[45-46] including multipole electrostatics, the Drude FF[47-49] using point charges and the CHeq FF[50-51] based on the use of a fluctuating charge model.

The choice of a FF should depend on the systems that they have been parameterized for.[2] Different FFs and their versions are parameterized for different systems such as proteins, nucleic acids, and small organic molecules. Furthermore, these FFs can be parameterized in the presence or absence of the solvent. In the absence of a solvent (implicit solvent) the aqueous medium is often compensated for by the use of a distance dependent dielectric constant, while in the presence of an aqueous medium (explicit solvent) a dielectric constant of 1 is used.[2] Explicit solvent water models used in biomolecular simulations involve SPC[52], SPC/E[53], TIP3P,[54] and TIP4P[54] models. AMBER, CHARMM, OPLS FFs are developed for the TIP3P water model, while OPLS has also been tested with TIP4P. GROMOS has been developed for the SPC model, while the KBFF approach uses the SPC/E water model.[17] Therefore, it is important to use a FF with its specified water model unless special solvent requirements are needed.[17] Furthermore, when selecting a FF the properties of interest also matter, as different FFs are optimized to reproduce different experimental properties.

1.2 Protein Denaturation

Proteins are large biomolecules composed of combinations of 20 different amino acid residues bonded together via peptide bonds. The sequence of the amino acids is known as the primary structure of the protein, while the function and activity of a protein is determined by the way the protein is folded into a three-dimensional structure. This folded three-dimensional structure, which performs the biological function and activity, corresponds to the native state of the protein.[55] Hydrogen bonding and the hydrophobic interactions are known to be the major factors that contribute to the stability of the native structure of a protein.[56]

A protein in a solution is generally in an equilibrium between a folded and an unfolded state, with the folded state usually being more favorable under ambient conditions. One can perturb the equilibrium between folded and unfolded states by altering the thermodynamic state of the system by changing the temperature, pressure, and pH.[57] In addition, the equilibrium can be shifted by changing the composition of solution, i.e. by adding a cosolvent to the solution. Changing the above factors in a way that the equilibrium shifts towards the unfolded state is known as protein denaturation. Protein denaturation is an important process used to study the stability of a protein. Exploring the thermodynamic properties related to denaturation process, and the forces that govern the folding to unfolding transition is necessary to understand the folding behavior.[58-59] By perturbing a system from equilibrium to populate the unfolded state, we can study the thermodynamic properties, such as, enthalpy, entropy and free energy, to determine the stability of the folded state of protein. In general, the equilibrium between the native (N) and the denatured (D) state is quantified by an equilibrium constant,

$$K = \frac{[D]}{[N]} \tag{1.13}$$

and the standard free energy of the unfolded state (ΔG_U) is related to K as,

$$\Delta G_U = -RT \ln K \tag{1.14}$$

where K is the equilibrium constant, R is the gas constant and T is the absolute temperature.

Temperature has been widely used to perturb the equilibrium in an effort to study the thermodynamics properties. However, to gain a complete understanding of the thermodynamics of a system, we also need to study the response of a system to pressure perturbations.[60-61] One advantage of pressure perturbation studies are that one can perturb the system without adding heat

or changing chemical activity. Changing temperature can affect both the volume and thermal activity of a system, while pressure only changes the volume of the system.[60]

1.2.1 Pressure denaturation

More than half of the volume of the total biosphere is covered by the ocean where the pressure increases by ~1 MPa per 100 m up to ~100 MPa in the deepest oceans.[62-63] Therefore, it is important to study the structure-function relationship of biomolecules under pressure. Furthermore, the effect of pressure on proteins is different than temperature or cosolvent effects on proteins.[64] Upon applying pressure, water exchanges between the bulk and the surface of native proteins due to increased conformational fluctuations.[65-66] In response to the water penetration into the native protein, pressure can induce protein denaturation.[65] In general, proteins undergo pressure denaturation above 4 kbar pressure.[60] However, the pressure range over which proteins denature depend on the conformational properties of the protein. For some proteins, the primary and secondary structures are not affected even at 10 kbar pressure.[65]

Pressure denaturation is directly related to the volume and compressibility differences between the folded and the unfolded state of a protein.[67-68] Therefore, these properties provide a useful tool to characterize the conformational transitions of proteins. The volumetric properties provide information concerning the intra and intermolecular interactions of proteins, which is complementary to the data provided by the temperature related properties, namely heat capacity and enthalpy.[67]

1.2.1.1 Volume of a protein

The effect of pressure on chemical equilibrium is given by the following thermodynamic relation,

$$\left(\frac{\partial \ln K}{\partial P}\right)_T = \frac{\Delta \bar{V}}{RT} \tag{1.15}$$

where $\Delta \bar{V}$ is the partial molar volume difference between the folded and the unfolded states.[69]

The water surrounding a protein (hydration water) plays an important role in determining the structure, function, and dynamics of a protein.[70] The volume of a protein provides insight into solute-solvent interactions and is therefore directly related to protein hydration.[71] According to the Le Chatelier principle, upon applying pressure the equilibrium should favor the state with a lower volume. Experimentally, it has observed that the pressure induced volume change for native to unfolded protein is very small and negative.[72] Measured values for the volume change are around 50-100 ml/mol, and this amount is generally about 0.5% to 2% of the protein's molar volume.[73] The major factors affecting the volume change upon pressure denaturation are the elimination of packing defects/voids and the exposure of buried groups upon unfolding[65, 72]. In general, the partial molar volume of a solute can be split into two main groups, namely the intrinsic volume and the changes in hydration volume.[74] The intrinsic volume represents the geometric volume of the protein, and surrounding water cannot penetrate into this volume. For a globular protein, the intrinsic volume is typically given by the sum of van der Waals volume and the void volume.[67] Changes in the hydration volume occurs due to the solute-solvent interactions and depend on the solvent accessible surface area of polar, nonpolar, and charged groups.[75] The hydration of charged and polar groups leads to a decrease in the volume due to electrostriction or hydrogen bonding.[70, 76] Upon unfolding, the hydrophobic residues buried inside the protein

become exposed to the aqueous environment and this leads to a positive or negative volume change depend on the model compound selected.[70]

The volume of a protein is experimentally measured using an apparent molar volume approach, where the volume is calculated by following the change in system volume upon solvating the solute in the solvent.[77] In this approach, the apparent molar volume is calculated by measuring the density of solution.[77-78] Furthermore, to help understand the volume changes upon protein denaturation, the properties of small molecule models are often used. In this approach, the volume of a protein is decomposed in to several groups according to their contributions. Chalikian and Breslauer[74] have decomposed the partial molar volume ($\bar{V}$) of a globular protein into five groups.[69]

$$\bar{V} = V_{id} + V_W + V_V + V_T + V_I \tag{1.16}$$

Here, V_{id} is the volume related to the kinetic contribution to the pressure of a solute molecule due to the translational degrees of freedom. This term is given by κ_TRT where, κ_T is the isothermal compressibility of the solvent. This term can be ignored for large proteins as the ideal value of κ_TRT is about 1 cm^3mol^{-1}.[74] V_W and V_V are the van der Waals volume and void volume of the protein, respectively. The thermal volume V_T is due to the thermally induced molecular vibrations of solute and solvent. The interaction volume, V_I denotes the decrease in the solvent volume due to interactions of water molecules with charged (electrostriction) and polar groups (hydrogen bonding) of solute. However, these terms contain many unknown parameters and they are not strictly obtained from experimental measurements, especially V_T and V_I.

In computer simulation studies many subjective definitions of protein volume are used to calculate the volume of a protein. Post and coworkers have introduced a method which calculates the volume of a protein based on a grid point analysis.[79] In their approach, the total protein volume

is divided in to a van der Waals volume and an unoccupied interstitial volume, where the interstitial volume contains internal cavities and packing defects. The interstitial volume is calculated by extending the atomic van der Waals radii by a constant value. This approach is only valid for approximately spherical proteins. Hence this volume definition is only valid for native globular proteins and not accurate for denatured proteins. Alternatively, Levy and coworkers have studied the volume of a protein using three different definitions namely the van der Waals volume, the molecular volume and the excluded volume.[71] The van der Waals volume is the volume occupied by the atoms of the molecule and measured with radii assigned to each atom.[71] The molecular volume is the volume enclosed by the molecular surface, and the excluded volume is the volume enclosed by solvent accessible surface area.[71] The last two quantities were measured using a probe sphere, and these volumes are therefore dependent on the size of the probe. According to the results that they have observed, the volume is dependent on the definition as there are significant differences between three different measures.

Another way of calculating the volume of a protein is the use of Kirkwood-Buff (KB) theory explained in section 1.3. Using computer simulations, KB integrals (G_{ij}) can be calculated and then related to volume through $RT\kappa_T - G_{21}$, where G_{21} is the solute-solvent KB integral.[80-86] This is more rigorous way of calculating the volume as this definition is not subjective.[85]

1.2.1.2 Compressibility of a protein

Compressibility is the other volumetric property that can be used to characterize pressure denaturation. The isothermal compressibility (κ_T) of a protein is given by the negative pressure derivative of the protein volume (V) at constant temperature (T).

$$\kappa_T = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T \tag{1.17}$$

Researchers have shown that there are two opposite contributions to the compressibility of a protein, and these two terms can be summed up to obtain the total compressibility.[67, 87-89] The positive contribution comes from the imperfect packing of the amino acid residues that display no accessible surface area, and is known as the 'intrinsic compressibility'. A negative contribution arises due to the interaction of water molecules with the amino acid residues on the surface of the protein, and this is known as the 'hydration compressibility'.[87] Upon applying pressure, proteins tend to favor smaller volumes compared to the native state.[90] Therefore, the volume change upon denaturation is negative. Furthermore, the compressibility of a protein is usually very small, and the pressure denatured state has a higher compressibility compared to the native state.[67] According to the experimental data, the isothermal compressibility of a globular protein ranges from $5x10^{-6}$ - $15x10^{-6}$ bar^{-1}, while the intrinsic compressibility of a protein is $25x10^{-6}$ bar^{-1}.[91-93] For, the comparison compressibility of pure water is $45x10^{-6}$, while benzene and hexane have compressibilities of $96x10^{-6}$, $165x10^{-6}$, respectively.[67] This suggests that the protein has a well packed, solid like, rigid interior compared to pure liquids.[88, 93-94] Rashin and coworkers has shown that the empty space within a protein is only about 25% of the total volume of a protein.[95] Moreover, the density of most protein are around 0.75 g/cm^3, and this value is close to the density of closely packed spheres (0.74 g/cm^3).[67]

Experimentally, the compressibility of a protein can be determined using crystallography, fluorescence spectroscopy, nuclear magnetic resonance spectroscopy (NMR), ultrasound and hole burning techniques.[79]

The compressibility of proteins has also been calculated using computer simulations. Again, Post and co-workers have calculated the compressibility of a protein using protein

molecular volume averages and fluctuations as $1/k_BT(<\Delta V^2>_{NPT}/<V>_{NPT})$ where V is the protein molecular volume.[79, 96] Here the protein volume is calculated based on subjective definitions, as described in previous paragraph, and assuming that the system fluctuations (implied by the above formula) can be simply replaced by the protein volume fluctuations. Even though they have observed high correlation between experimental and calculated compressibilities, the definition is somewhat different than thermodynamically defined infinitely dilute partial molar compressibility.[85]

1.3 The Kirkwood-Buff Theory of Solutions

Kirkwood-Buff (KB) theory is one of the most important and powerful theories of liquid mixtures and was published in 1951.[83] The theory relates the molecular distribution of species of a multicomponent system in the μVT ensemble to the thermodynamic properties of the mixture such as partial molar volumes, derivatives of the chemical potentials, and the compressibility. The relationship between the thermodynamic properties and the molecular distribution is given by KB integrals (KBI) as follows,

$$G_{ij} = \int_0^\infty \left[g_{ij}(R) - 1 \right] 4\pi R^2 dR \tag{1.18}$$

where $g_{ij}(r)$ is the corresponding pair correlation function, or the radial distribution function (RDF), between species i and j, and R is the distance between the corresponding center of masses. The KB integral (G_{ij}) quantifies the deviation in the distribution of j molecules around a central i molecule when compared to that expected for a random distribution of j molecules. If the KB integral between species i and j is greater than zero, then there are favorable net interactions or an

affinity between those species. Alternatively, a negative KB integral indicates that there are net unfavorable interactions between the corresponding species.

Even though KB theory represents a powerful tool for the analysis of solution mixtures, when it was first published it was practically impossible to calculate the required pair correlation functions for mixtures. Therefore, there were very few publications citing KB theory in the first 20 years after its introduction. In 1977, Ben-Naim introduced the inversion procedure of KB theory which can compute the affinity (KBIs) between two species using the observed experimental thermodynamic quantities of the mixtures such as partial molar volumes, isothermal compressibilities and partial vapor pressures.[97] This is known as the inversion of KB theory and, ever since, the use of KB theory has rapidly increased. Since then, KB theory has been used in a variety of fields to investigate the molecular distributions and preferential solvation in solutions for many solution mixtures.[98] The main advantages of KB theory are that it is an exact theory without any approximations, and the fact it can used to analyze any stable solution mixture with any number of components. Furthermore, this theory can be applied to any solution regardless of size and complexity of molecules, and is ideally suitable for the analysis of computer simulations of solution mixtures.[99]

KB theory and the inversion of KB theory has been mainly applied to two component systems. For a binary mixture containing a cosolvent (c) and a solvent (s) we can obtain the thermodynamic quantities – partial molar volumes of each component ($\bar{V}_c$, $\bar{V}_s$), isothermal compressibility of solution mixture (κ_T), and chemical potential derivatives (μ_{cc}, μ_{ss}, μ_{cs}) using the KB integrals (G_{cc}, G_{ss} and G_{cs}) as follows,[100]

$$\bar{V}_c = \frac{1+\rho_s(G_{ss}-G_{cs})}{\eta} \tag{1.19}$$

$$\bar{V}_s = \frac{1+\rho_c(G_{cc}-G_{cs})}{\eta} \tag{1.20}$$

$$\kappa_T = \frac{\zeta}{kT\eta} \tag{1.21}$$

In above equations, the ρ_c, ρ_s are average number densities, k is the Boltzmann constant, and T is the temperature. Two auxiliary quantities, η and ζ are defined as the following expressions,

$$\eta = \rho_c + \rho_s + \rho_c\rho_s(G_{cc}+G_{ss}-2G_{cs}) \tag{1.22}$$

$$\zeta = 1+\rho_c G_{cc} + \rho_s G_{ss} + \rho_c\rho_s(G_{cc}G_{ss}-G_{cs}{}^2) \tag{1.23}$$

For an ideal solution where activity coefficients of all species are equal to unity for all compositions (Symmetric ideal solutions), the KB integrals are neither zero nor independent of composition.[101] The excess volumes and enthalpies for zero for this type of solutions for the whole composition range. Ben-Naim has shown that for a symmetric ideal (SI) solution in a binary system,

$$\Delta G_{ij} = G_{ii} + G_{jj} - 2G_{ij} = 0 \tag{1.24}$$

at constant temperature and pressure for any composition.[100] Smith has proposed a general expression for KB integrals of SI solutions as,[102]

$$G_{ij}^{SI} = RT\kappa_T - V_i - V_i + \sum_{k=1}^{n}\rho_k V_k^2 \tag{1.25}$$

where κ_T is the isothermal compressibility, and V_i, V_j are the molar volumes of the pure components. This expression is valid for any i and j combination in SI solutions with n components.

1.4 Inversion of the Kirkwood-Buff theory

KB theory was initially derived to obtain the thermodynamic properties of solution mixtures using molecular distributions, where the distributions functions were obtain using computer simulations or analytical calculations.[100] This can be symbolically written as follows,

$$\left\{G_{ij}\right\} \rightarrow \left\{\bar{V}_i, \kappa_T, \partial\mu_i/\partial\rho_j\right\} \tag{1.26}$$

Upon introducing the inversion of KB theory, the thermodynamic quantities are used to calculate the KB integrals. As it is easy to measure the thermodynamic properties experimentally, rather than computing the pair correlation function (g_{ij} (R)), this approach provides a powerful tool to study the local distribution of molecules in a solution mixture. The inversion of KB theory can be symbolically written as follows,

$$\left\{\bar{V}_i, \kappa_T, \partial\mu_i/\partial\rho_j\right\} \rightarrow \left\{G_{ij}\right\} \tag{1.27}$$

Hence, the KB integrals can be expressed in terms of experimental thermodynamic quantities. For a binary mixture we have,[100]

$$G_{cs} = kT\kappa_T - \rho\bar{V}_c\bar{V}_s / D \tag{1.28}$$

$$G_{cc} = kT\kappa_T - \frac{1}{\rho_c} + \frac{\rho_S\bar{V}_s^2\rho}{\rho_c D} \tag{1.29}$$

$$G_{ss} = kT\kappa_T - \frac{1}{\rho_s} + \frac{\rho_c\bar{V}_c^2\rho}{\rho_s D} \tag{1.30}$$

where $\rho = \rho_c + \rho_s$ and,

$$D = \frac{x_c}{kT}\left(\frac{\partial\mu_c}{\partial x_c}\right)_{T,p} \tag{1.31}$$

where x_c is the mole fraction of the cosolvent.

1.4.1 Fluctuation solution theory

Fluctuation solution theory (FST) is an extension of KB theory that also provides information regarding the local composition of solutions, or the deviation of local composition from bulk solution. Here, KB integrals are given by the particle-particle fluctuations corresponding to a local system, instead of integrating the radial distribution function between species, as shown below,[1]

$$G_{ij} = V\frac{\left\langle \delta N_i \delta N_j \right\rangle}{\left\langle N_i \right\rangle \left\langle N_j \right\rangle} - \frac{\delta ij}{\left\langle N_i \right\rangle} \tag{1.32}$$

$$\delta N_i = N_i - \left\langle N_i \right\rangle$$

where the δN_i is the deviation of number of i particles from the average number of i particles in a fixed volume (V) of a grand canonical ensemble (μVT), and δ_{ij} is the Kronecker delta function. Here, the angular brackets represent ensemble averages in the grand canonical ensemble.

1.5 Application of KB/FST to protein denaturation

Proteins undergo denaturation under the influence of temperature, pressure or cosolvents, and the denaturation process has been widely studied using experiments. However, using experimental studies it is not easy to assign thermodynamic properties to specific protein conformations, either the native or denatured form.[80, 103] Computer simulation studies provide atomic level details of interactions and structural changes of molecules. Nevertheless, using computer simulations it is not easy to follow the denaturation process due to the high computational cost. For instance, it is impossible to simulate the equilibrium constant for denaturation for all but a few small (20-30 residue) systems. Furthermore, it is not clear how to extract from simulation the properties of proteins that are related to thermal or pressure

denaturation.[80, 103] In the literature, properties related to protein denaturation have been studied using replica exchange molecular dynamics (REMD) simulations.[104] This enhanced sampling technique explores new conformational spaces at different temperatures and allows one to determine the equilibrium constant by calculating the fraction of folded and unfolded protein molecules.[16, 85] However, this provides relative thermodynamic properties and cannot assign these properties to specific individual conformations.

The KB and FST theories provide the most promise to be able to extract the relevant, conformation specific, thermodynamic properties such as partial molar volumes and compressibilities. As mentioned above, in KB theory the thermodynamic properties are related to an integral over the RDFs. In the FST approach the properties are related to particle number fluctuations in a local region of solution, and the integration of the distribution functions is unnecessary. More importantly, this latter method allows for a surface-based analysis of proteins. Therefore, this type of approach is more convenient than the integration of the KB equation (1.17) for nonspherical, or irregular shaped, proteins.[103, 105] Furthermore, even though the KB approach provides identical results to FST, the analysis over the surface of the protein, rather than for the centers of mass, provides more insight by enabling one to decompose the thermodynamic properties based on both distance and proximity to different residues.[103, 105]

1.5.1 Protein denaturation thermodynamics

FST theory can be used to study the thermodynamics of the chemical equilibrium corresponding to protein denaturation. From the pressure denaturation point of view, the volume and compressibility of a protein can be obtained using this theory. For a protein in an infinitely dilute solution, the equilibrium constant can be written as follows,

$$\ln K = -\beta(\mu_D^{*,\infty} - \mu_N^{*,\infty}) \tag{1.33}$$

where β is 1/RT, and $\mu_D^{*,\infty}$ and $\mu_N^{*,\infty}$ correspond to the pseudo chemical potential of denatured and the native states. The pseudo chemical potential represents a statistical mechanical expression for the chemical potential that allows one to calculate the chemical potential without reference to standard states.[80, 100] The pseudo chemical potential is defined as the change in Gibbs free energy for adding a single particle to a fixed position in space within the system.[100] Therefore, the added particle is free from translational degrees of freedom. The chemical potential of a protein solute (2) in a solvent (1) can be written as,

$$\begin{aligned} \mu_2 &= \mu_{2,m}^0 + RT\ln(\gamma_{2,m}\,\mathrm{m}_2/\mathrm{m}^0) \\ \mu_2 &= \mu_{2,c}^0 + RT\ln(\gamma_{2,c}c_2/c^0) \\ \mu_2 &= \mu_2^* + RT\ln(\Lambda_2^3\rho_2) \end{aligned} \tag{1.34}$$

where the first two expressions relate the solute chemical potential to the usual experimental measures, using the standard chemical potentials ($\mu_{2,m}^0$, $\mu_{2,c}^0$) and activity coefficients ($\gamma_{2,m}$, $\gamma_{2,c}$) in terms of the solute molal (m_2) or molar (c_2) concentration, respectively. The third expression in equation (1.34) provides the Ben Naim's statistical mechanical relationship for the pseudo chemical potential. Here, the pseudo chemical potential is related to the total chemical potential using the thermal de Broglie wavelength (Λ_2) and the number density of the solute (ρ_2). This approach allows to one to split the chemical potential into two parts; a contributions from the translational degrees of freedom (second term on the right hand side), and a contribution from the intermolecular interactions of the system (first term on the right hand side).[100, 106] The pseudo chemical potential is similar to the excess chemical potential when the internal partition function

is neglected.[107] The pressure derivatives of above equation (1.34) for biomolecular solute (2) at constant temperature are given by,

$$\bar{V}_2 = \left(\frac{\partial \mu_2}{\partial p}\right)_{m_2} = \left(\frac{\partial \mu_{2,m}^0}{\partial p}\right)_{m_2} + RT\left(\frac{\partial \ln \gamma_{2,m}}{\partial p}\right)_{m_2}$$

$$\bar{V}_2 = \left(\frac{\partial \mu_{2,c}^0}{\partial p}\right)_{m_2} + RT\left(\frac{\partial \ln \gamma_{2,c}}{\partial p}\right)_{m_2} + RT\kappa_T \tag{1.35}$$

$$\bar{V}_2 = \left(\frac{\partial \mu_2^*}{\partial p}\right)_{m_2} + RT\kappa_T$$

where m_2 is the solute molality, and κ_T is the isothermal compressibility of the solution. The effect of pressure on the chemical equilibrium is usually given by a Taylor expansion around a reference pressure p_0,

$$\ln\left(\frac{K}{K_0}\right) = \left(\frac{\partial \ln K}{\partial p}\right)_{m_2} \Delta p + \frac{1}{2}\left(\frac{\partial^2 \ln K}{\partial p^2}\right)_{m_2} (\Delta p)^2 \tag{1.36}$$

where K_0 is the equilibrium constant, and $\Delta p = (p\text{-}p_0)$ where p_0 is usually 1 bar. Consequently, the derivatives of the above expression can be written using equation (1.33) as follows,

$$-RT\left(\frac{\partial \ln K}{\partial p}\right)_{m_2} = \left[\left(\frac{\partial \mu_D^{*,\infty}}{\partial p}\right)_{m_2} - \left(\frac{\partial \mu_N^{*,\infty}}{\partial p}\right)_{m_2}\right] = \Delta V_2^{*,\infty} \tag{1.37}$$

$$-RT\left(\frac{\partial^2 \ln K}{\partial p^2}\right)_{m_2} = \left[\left(\frac{\partial^2 \mu_D^{*,\infty}}{\partial p^2}\right)_{m_2} - \left(\frac{\partial^2 \mu_N^{*,\infty}}{\partial p^2}\right)_{m_2}\right] = -\Delta K_2^{*,\infty}$$

Here, $\Delta V_2^{*,\infty}$ is the volume difference ($V_D^{*,\infty} - V_N^{*,\infty}$), and $\Delta K_2^{*,\infty}$ is the difference in compressibility factors. The effect of pressure on the equilibrium therefore given by the volume difference between folded and unfolded state, and this is modulated by the protein compressibilities.

In terms of the particle number fluctuations, the pseudo molar volume of a solute of any size in an infinitely dilute solution is given by the following expression,[80, 103, 105-106]

$$V_2^{*,\infty} = -[\langle N_1 \rangle_2 - \langle N_1 \rangle_0] V_1^0 \tag{1.38}$$

where, $<N_1>_2$ is the average number of water molecules within a fixed local volume centered on the protein, and $<N_1>_0$ is the average number of water molecules within the same volume in pure water. Here, N_1 denotes the number of water molecules, while the subscript 2 and 0 represents the solution and the pure solvent, respectively. The pseudo molar volume is then given by the difference in number of water molecules, multiplied by the volume of pure water (V_1^0) at the corresponding temperature and pressure. This simple expression for volume is not subjective, as it uses only the average number of water molecules and the volume of pure water. Using this type of approach, the volume of a residue in a protein can be calculated by assigning each water molecule to a heavy atom of the protein based on their proximity.[103, 105]

The expression for the pressure derivative of the pseudo volume using the particle number fluctuations is,[80, 103, 105-106]

$$K_2^{*,\infty} = \kappa_{T,1}^0 V_2^{*,\infty} + \beta (V_1^0)^2 [< \delta N_1 \delta N_1 >_2 - < \delta N_1 \delta N_1 >_0] \tag{1.39}$$

where, $< \delta N_1 \delta N_1 >_x = < N_1^2 >_x - < N_1 >_x^2$, and $\kappa_{T,1}^0$ is the partial molar compressibility of pure water. The compressibility is related to the solvent fluctuations in the presence of the protein, compared to the solvent fluctuations in the absence of the protein or in pure water. The compressibility of a protein calculated using this expression can be either positive or negative, depending on the solvent fluctuations in the protein solution and bulk water. If the average fluctuations in the presence of protein ($< \delta N_1 \delta N_1 >_2$) are smaller than in pure water ($< \delta N_1 \delta N_1 >_0$) , then the compressibility of the protein will be negative, and vice versa.

In addition to the above FST expression, the compressibility can be also calculated using the expression,

$$\kappa_{T,2}^{*} = -\frac{1}{V_2^{*}}(\partial V_2^{*} / \partial p)_{T,m_2} \tag{1.40}$$

This method is computationally more expensive as it requires the volume of a protein as a function of pressure. Therefore, simulations need to be performed at several pressures. However, this approach is statistically more favorable since the compressibility is calculated from a set of simulations at different pressure, rather than using a single simulation at one pressure.

1.6 Organization of dissertation

Molecular dynamics simulations were performed to explore the structure, dynamics and functions of a series of biomolecular related systems. In this thesis we will discuss the development of new FF parameters, the use of FST theory to study pressure denaturation thermodynamics of proteins, and the application of MD simulations to better explain experimental data concerning biomolecules.

In Chapter 2 we discuss the development of force field parameters for models of glycerol, ethylene glycol, 1,2- propanediol, and 1,3-propanediol in water, methanol and ethanol solvents based on Kirkwood-Buff theory. Here, the parameters are optimized to reproduce the experimental Kirkwood-Buff integrals calculated from the thermodynamic properties of solution mixtures. Furthermore, the enthalpy of mixing, the volume of mixing, the density, the diffusion coefficients, the dielectric constants, the viscosity and the torsion angle populations are then used to validate the FF parameters.

In Chapter 3 we investigate the thermodynamic properties that characterize the pressure denaturation of proteins. Infinitely dilute partial molar volumes and compressibilities of native hen

egg white Lysozyme, Ribonuclease A, and Ubiquitin are studied using the FST approach. This approach depends solely on the solvent distribution in the presence and absence of the protein, and also allows us to calculate the residue-based contributions to the volume of the proteins. This provides great detail concerning the role of specific residues to the overall protein volume. Subsequently, the volume is followed as a function of pressure to calculate the residue based compressibilities. Finally, we discuss about the factors that may contribute to the overall very low compressibility of proteins.

In Chapter 4 the partial molar volumes and compressibilities at infinite dilute solution are studied and the results are compared among four different FFs, CHARMM22*, AMBER 99SB-ILDN, OPLS-AA and GROMOS53a6. The properties of each FF are compared and contrasted in an effort to establish consistent trends for these properties from simulation.

In Chapter 5 we use MD simulations to help understand the motions of wild type *lac* repressor. The *lac* repressor is a DNA-binding protein which regulates the function of *lac* operon. The effect of pressure on the structure and dynamics of LacI is then investigated.

In Chapter 6 our focus is to examine some properties that allow one to identify rheostat positions in proteins. MD simulations are performed to identify the changes in behavior of the LacI protein upon mutating the V52 position of LacI with different amino acids, and how these mutants respond to increases in pressure. The differences are then related to changes in protein hydration.

Chapter 7 provides a summary of this dissertation and our future work.

1.7 References

1. Smith, P. E.; Matteoli, E.; O'Connell, J. P., Fluctuation theory of solutions: applications in chemistry, chemical engineering, and biophysics. CRC Press: 2016.
2. Van Gunsteren, W., *Molecular Simulation* **1989,** *3* (4), 187-200.

3. Meller, J., Encyclopedia of life sciences **2001**, 1-8.

4. Hospital, A.; Goñi, J. R.; Orozco, M.; Gelpí, J. L., *Advances and applications in bioinformatics and chemistry : AABC* **2015,** *8*, 37-47.

5. van Gunsteren, Wilfred F.; Dolenc, J., *Biochemical Society Transactions* **2008,** *36* (1), 11-15.

6. Karplus, M.; Petsko, G. A., *Nature* **1990,** *347* (6294), 631-639.

7. Karplus, M.; McCammon, J. A., *Nature Structural & Molecular Biology* **2002,** *9* (9), 646-652.

8. Allen, M. P.; Tildesley, D. J., *Computer simulation of liquids*. Oxford university press: 2017.

9. Verlet, L., *Physical review* **1967,** *159* (1), 98.

10. Becker, O. M.; MacKerell, A. D.; Roux, B.; Watanabe, M., *Computational Biochemistry and Biophysics*. Taylor & Francis: 2001.

11. Monticelli, L.; Salonen, E., *Biomolecular simulations: methods and protocols*. Springer: 2013; Vol. 924.

12. González, M., *École thématique de la Société Française de la Neutronique* **2011,** *12*, 169-200.

13. van Gunsteren, W. F.; Berendsen, H. J. C., *Angewandte Chemie International Edition in English* **1990,** *29* (9), 992-1023.

14. Nymeyer, H., *Journal of Chemical Theory and Computation* **2008,** *4* (4), 626-636.

15. Paschek, D.; García, A. E., *Physical review letters* **2004,** *93* (23), 238105.

16. Sugita, Y.; Okamoto, Y., *Chemical physics letters* **1999,** *314* (1-2), 141-151.

17. Mackerell, A. D., *Journal of computational chemistry* **2004,** *25* (13), 1584-1604.

18. Ploetz, E. A.; Rustenburg, A. S.; Geerke, D. P.; Smith, P. E., *Journal of Chemical Theory and Computation* **2016,** *12* (5), 2373-2387.

19. Weerasinghe, S.; Smith, P. E., *The Journal of chemical physics* **2003,** *119* (21), 11342-11349.

20. Weerasinghe, S.; Smith, P. E., *The Journal of chemical physics* **2003,** *118* (23), 10663-10670.

21. Weerasinghe, S.; Smith, P. E., *The Journal of Physical Chemistry B* **2003,** *107* (16), 3891-3898.

22. Weerasinghe, S.; Smith, P. E., *The Journal of Physical Chemistry B* **2005,** *109* (31), 15080-15086.

23. Kang, M.; Smith, P. E., *Journal of computational chemistry* **2006,** *27* (13), 1477-1485.

24. Bentenitis, N.; Cox, N. R.; Smith, P. E., *The Journal of Physical Chemistry B* **2009,** *113* (36), 12306-12315.

25. Gee, M. B.; Cox, N. R.; Jiao, Y.; Bentenitis, N.; Weerasinghe, S.; Smith, P. E., *Journal of chemical theory and computation* **2011,** *7* (5), 1369-1380.

26. Ploetz, E. A.; Smith, P. E., *Physical Chemistry Chemical Physics* **2011,** *13* (40), 18154-18167.

27. Jiao, Y. The development of accurate force fields for protein simulation. Kansas state university, 2012.

28. Ploetz, E. A.; Weerasinghe, S.; Kang, M.; Smith, P. E.; Smith, P.; Matteoli, E.; O'Connell, J., *Accurate Force Fields for Molecular Simulation*. CRC Press: Boca Raton, FL: 2013.

29. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., *Proteins: Structure, Function, and Bioinformatics* **2010,** *78* (8), 1950-1958.

30. Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E., *Journal of Chemical Theory and Computation* **2010,** *6* (2), 459-466.

31. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., *The Journal of Physical Chemistry B* **2001,** *105* (28), 6474-6487.

32. Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F., *Journal of Computational Chemistry* **2004,** *25* (13), 1656-1676.

33. Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H., *The Journal of Physical Chemistry B* **2007,** *111* (27), 7812-7824.

34. Mackerell Jr, A. D.; Feig, M.; Brooks III, C. L., *Journal of computational chemistry* **2004,** *25* (11), 1400-1415.

35. MacKerell Jr, A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S., *The journal of physical chemistry B* **1998,** *102* (18), 3586-3616.

36. Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell Jr, A. D., *Journal of chemical theory and computation* **2011,** *8* (1), 348-362.

37. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell Jr, A. D., *Journal of chemical theory and computation* **2012,** *8* (9), 3257-3273.

38. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., *Journal of the American Chemical Society* **1995,** *117* (19), 5179-5197.

39. Kollman, P.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A., The development/application of a 'minimalist'organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data. In *Computer simulation of biomolecular systems*, Springer: 1997; pp 83-96.

40. Wang, J.; Cieplak, P.; Kollman, P. A., *Journal of computational chemistry* **2000,** *21* (12), 1049-1074.

41. van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G., *Biomolecular Simulation: The {GROMOS96} manual and userguide*. Hochschuleverlag AG an der ETH Zürich: 1996.

42. Schuler, L. D.; Daura, X.; Van Gunsteren, W. F., *Journal of Computational Chemistry* **2001,** *22* (11), 1205-1218.

43. Huang, J.; Lopes, P. E. M.; Roux, B.; MacKerell, A. D., *The Journal of Physical Chemistry Letters* **2014,** *5* (18), 3144-3150.

44. Baker, C. M., *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2015,** *5* (2), 241-254.

45. Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio Jr, R. A., *The journal of physical chemistry B* **2010,** *114* (8), 2549-2564.

46. Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P., *Journal of chemical theory and computation* **2013,** *9* (9), 4046-4063.

47. Lopes, P. E.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell Jr, A. D., *Journal of chemical theory and computation* **2013,** *9* (12), 5430-5449.

48. Savelyev, A.; MacKerell Jr, A. D., *Journal of computational chemistry* **2014,** *35* (16), 1219-1239.

49. Chowdhary, J.; Harder, E.; Lopes, P. E.; Huang, L.; MacKerell Jr, A. D.; Roux, B., *The Journal of Physical Chemistry B* **2013,** *117* (31), 9142-9160.

50. Lucas, T. R.; Bauer, B. A.; Patel, S., *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2012,** *1818* (2), 318-329.

51. Bauer, B. A.; Patel, S., *Theoretical Chemistry Accounts* **2012,** *131* (3), 1153.

52. Berendsen, H. J.; Postma, J. P.; van Gunsteren, W. F.; Hermans, J., Interaction models for water in relation to protein hydration. In *Intermolecular forces*, Springer: 1981; pp 331-342.

53. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., *The Journal of Physical Chemistry* **1987,** *91* (24), 6269-6271.

54. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., *The Journal of chemical physics* **1983,** *79* (2), 926-935.

55. Lazaridis, T.; Archontis, G.; Karplus, M., Enthalpic Contribution to Protein Stability: Insights from Atom-Based Calculations and Statistical Mechanics. In *Advances in Protein Chemistry*, Anfinsen, C. B.; Richards, F. M.; Edsall, J. T.; Eisenberg, D. S., Eds. Academic Press: 1995; Vol. 47, pp 231-306.

56. Privalov, P. L.; Makhatadze, G. I., *Journal of Molecular Biology* **1993,** *232* (2), 660-679.

57. Canchi, D. R.; García, A. E., *Annual Review of Physical Chemistry* **2013,** *64* (1), 273-293.

58. Kauzmann, W., *Nature* **1987,** *325* (6107), 763.

59. Dill, K. A., *Biochemistry* **1990,** *29* (31), 7133-7155.

60. Royer, C. A., [16] Application of pressure to biochemical equilibria: The other thermodynamic variable. In *Methods in enzymology*, Elsevier: 1995; Vol. 259, pp 357-377.

61. Dubins, D. N.; Filfil, R.; Macgregor, R. B.; Chalikian, T. V., *Biochemistry* **2003,** *42* (29), 8671-8678.

62. Gross, M.; Jaenicke, R., *The FEBS Journal* **1994,** *221* (2), 617-630.

63. Yancey, P. H.; Gerringer, M. E.; Drazen, J. C.; Rowden, A. A.; Jamieson, A., *Proceedings of the National Academy of Sciences* **2014,** *111* (12), 4461-4465.

64. Harano, Y.; Kinoshita, M., *Journal of Physics: Condensed Matter* **2006,** *18* (7), L107.

65. Boonyaratanakornkit, B. B.; Park, C. B.; Clark, D. S., *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **2002,** *1595* (1-2), 235-249.

66. Tanaka, N.; Ikeda, C.; Kanaori, K.; Hiraga, K.; Konno, T.; Kunugi, S., *Biochemistry* **2000,** *39* (39), 12063-12068.

67. Taulier, N.; Chalikian, T. V., *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **2002,** *1595* (1-2), 48-70.

68. Royer, C. A., *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **2002,** *1595* (1-2), 201-209.

69. Imai, T.; Harano, Y.; Kovalenko, A.; Hirata, F., *Biopolymers* **2001,** *59* (7), 512-519.

70. Sirotkin, V. A.; Komissarov, I. A.; Khadiullina, A. V., *The Journal of Physical Chemistry B* **2012,** *116* (13), 4098-4105.

71. Murphy, L. R.; Matubayasi, N.; Payne, V. A.; Levy, R. M., *Folding and Design* **1998,** *3* (2), 105-118.

72. Chalikian, T. V., *Annual review of biophysics and biomolecular structure* **2003,** *32* (1), 207-235.

73. Royer, C. A., Why and how does pressure unfold proteins? In *High Pressure Bioscience*, Springer: 2015; pp 59-71.

74. Chalikian, T. V.; Breslauer, K. J., *Biopolymers* **1996,** *39* (5), 619-626.

75. Chalikian, T. V.; Totrov, M.; Abagyan, R.; Breslauer, K. J., *Journal of molecular biology* **1996,** *260* (4), 588-603.

76. Kauzmann, W., Some Factors in the Interpretation of Protein Denaturation1. In *Advances in protein chemistry*, Elsevier: 1959; Vol. 14, pp 1-63.

77. Zamyatnin, A., *Progress in biophysics and molecular biology* **1972,** *24*, 107-123.

78. Millero, F. J.; Lo Surdo, A.; Shin, C., *The Journal of Physical Chemistry* **1978,** *82* (7), 784-792.

79. Dadarlat, V. M.; Post, C. B., *The Journal of Physical Chemistry B* **2001,** *105* (3), 715-724.

80. Jiao, Y.; Smith, P. E., *The Journal of chemical physics* **2011,** *135* (1), 07B602.

81. Patel, N.; Dubins, D. N.; Pomes, R.; Chalikian, T. V., *The Journal of Physical Chemistry B* **2011,** *115* (16), 4856-4862.

82. Floris, F. M., *The Journal of Physical Chemistry B* **2004,** *108* (41), 16244-16249.

83. Kirkwood, J. G.; Buff, F. P., *The Journal of Chemical Physics* **1951,** *19* (6), 774-777.

84. Sangwai, A. V.; Ashbaugh, H. S., *Industrial & Engineering Chemistry Research* **2008,** *47* (15), 5169-5174.

85. Ploetz, E. A.; Smith, P. E., *The Journal of Physical Chemistry B* **2014,** *118* (45), 12844-12854.

86. Patel, N.; Dubins, D. N.; Pomès, R.; Chalikian, T. V., *Biophysical chemistry* **2012,** *161*, 46-49.

87. Chalikian, T. V.; Sarvazyan, A. P.; Breslauer, K. J., *Biophysical chemistry* **1994,** *51* (2-3), 89-109.

88. Chalikian, T. V.; Breslauer, K. J., *Current opinion in structural biology* **1998,** *8* (5), 657-664.

89. Sarvazyan, A. P., *Annual review of biophysics and biophysical chemistry* **1991,** *20* (1), 321-342.

90. Royer, C., *Brazilian journal of medical and biological research* **2005,** *38* (8), 1167-1173.

91. Chalikian, T. V.; Gindikin, V. S.; Breslauer, K. J., *Journal of molecular biology* **1995,** *250* (2), 291-306.

92. Kharakoz, D.; Sarvazyan, A., *Biopolymers* **1993,** *33* (1), 11-26.

93. Gekko, K.; Hasegawa, Y., *Biochemistry* **1986,** *25* (21), 6563-6571.

94. Gekko, K.; Noguchi, H., *Journal of Physical Chemistry* **1979,** *83* (21), 2706-2714.

95. Rashin, A. A.; Iofin, M.; Honig, B., *Biochemistry* **1986,** *25* (12), 3619-3625.

96. Dadarlat, V. M.; Post, C. B., *Biophysical journal* **2006,** *91* (12), 4544-4554.

97. Ben-Naim, A., *The Journal of Chemical Physics* **1977,** *67* (11), 4884-4890.

98. Matteoli, E.; Mansoori, G. A., *Fluctuation theory of mixtures*. Taylor & Francis New York:: 1990.

99. Pierce, V.; Kang, M.; Aburi, M.; Weerasinghe, S.; Smith, P. E., *Cell biochemistry and biophysics* **2008,** *50* (1), 1-22.

100. Ben-Naim, A., *Molecular theory of solutions*. Oxford University Press: 2006.

101. Ploetz, E. A.; Bentenitis, N.; Smith, P. E., *The Journal of chemical physics* **2010,** *132* (16), 164501.

102. Smith, P. E., *The Journal of chemical physics* **2008,** *129* (12), 124509.

103. Ploetz, E. A.; Smith, P. E., *The Journal of Physical Chemistry B* **2015,** *119* (25), 7761-7777.

104. Canchi, D. R.; Paschek, D.; García, A. E., *Journal of the American Chemical Society* **2010,** *132* (7), 2338-2344.

105. Ploetz, E. A.; Smith, P. E., *Biophysical chemistry* **2017,** *231*, 135-145.

106. Ploetz, E. A.; Smith, P. E., *Advances in chemical physics* **2013,** *153*, 311.

107. Ploetz, E. A.; Smith, P. E., *The journal of chemical physics* **2011,** *135* (4), 044506.

Chapter 2 - A Kirkwood-Buff Force Field for Simulations of Polyols

2.1 Introduction

Studying the structure and thermodynamic properties of lipid bilayers is very important due to their vital role in biological systems. Computer simulation can be employed to study these properties as it provides the atomistic level details of the system of interest. However, an accurate force field (FF), and proper sampling, are needed to investigate the above-mentioned properties of a system using molecular dynamics (MD) simulations.

Polyols are an important class of molecules which contain more than one hydroxyl group. These molecules have plenty of applications in different fields such as foods, polymers, and pharmaceuticals.[1-2] Their special ability to form hydrogen bonds with water makes them useful as antifreezing agents.[3] However, in contrast to aliphatic alkanes, the ability of polyols to form strong inter and intramolecular hydrogen bonds make it challenging to obtain reasonable FF parameters for these models.[4]

Numerous MD studies have been performed to model polyols in the liquid phase. Computer simulation studies using two different glycerol models,[5-6] based on CHARMM22 and AMBER FFs, and three other glycerol models, based on OPLS FF, have been performed at a wide range of temperatures (300-460 K) and 0.1 MPa pressure.[7] They have studied properties such as the density, the thermal expansion coefficient, the isobaric specific heat, the compressibility, and the diffusion coefficients. The AMBER and OPLS based models closely replicated experimental densities while the CHARMM based model underestimated this property. The CHARMM based model displays larger deviations for the thermal expansion coefficient compared to other models. All models overestimate the experimental specific heat while CHARMM, AMBER and one of three OPLS models well reproduce the experimental diffusion coefficients. Furthermore, the

populations of structural conformers have also been studied, using the Bastiansen nomenclature,[8] and the FFs have been shown to generate very different conformer populations. A study of glycerol (CHARMM22 and AMBER based models) and water (TIP3P and TIP4P models) mixtures have shown that all the mixtures overestimate the experimental density at low temperature while underestimating at high temperature.[9] They have summarized that the mixtures with TIP4P water model better reproduce the experimental thermodynamic and dynamic properties.[9]

Xibing and coworkers have studied pure glycerol and 1, 2-ethanediol liquids using the CHARMM polarizable FF.[10] Here, electronic polarizability is explicitly treated via a classical Drude oscillator. They have shown that, with the better treatment of electrostatic interactions in polarizable FF, the models provide good agreement with quantum mechanically obtained conformational energies, compared to the CHARMM additive FF for 1,2-ethanediol[11] and glycerol,[12] where the gas phase dipole moments are overestimated to mimic condensed phase properties. They have also observed improved agreement with experimental condensed phase properties, such as the density and enthalpy of vaporization, over the additive models. Correct conformational energies and enthalpy of vaporization indicate the correct balance between intra and intermolecular hydrogen bonding in the condensed phase.

Ferrando and coworkers have developed a new FF for polyalcohols based on an extension of the anisotropic united atom FF AUA4 developed for hydrocarbons.[13] In this FF they have introduced an anisotropic united atom corresponding to the hydroxyl of alcohols. The 1, 2-ethanediol and 1, 3-propanediol models have been developed using this approach. Both models underestimate the experimental density values while they provide reasonable agreement with experimental vapor pressures and vaporization enthalpies.

Another study by Kulschewski was performed to model the properties of liquid propanol, 1,2-ethanediol, glycerol, and some other aliphatic alcohols using OPLS all-atom (AA) FF.[14] With the modified charges in AAFF, they have obtained good agreement with experimental densities and self-diffusion constants except for glycerol. However, these models do not show reasonable agreement with experimental dielectric constants and thermal expansion coefficients.

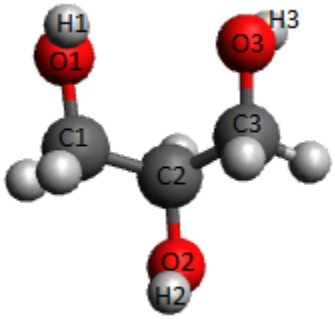


Figure 2.1 Glycerol model with atom labels used for Bastiansen nomenclature

The backbone rotamer population distribution of glycerol at infinite dilution has been studied using GLYCAM06/AMBER FF with the TIP3P water model.[15] The backbone conformations of glycerol are defined based on the Bastiansen nomenclature.[8] Six different conformations denoted as αα, αβ, αγ, ββ, βγ, γγ, have been defined using the terminal oxygen and three carbon atoms. According to the atom labels shown in Figure 2.1 the two dihedral angles are considered here are, O1C1C2O2 and O2C2C3O3. Rotamer αα is defined when the O1C1C2O2 is gauche$^-$ and the O2C2C3O3 is gauche$^+$. Rotamer ββ is defined when both angles are trans and, the γγ is defined when the O1C1C2O2 is gauche$^+$ and the O2C2C3O3 is gauche$^-$.[15] The heterogeneous atom pairs (βα, γβ) are not counted as different conformers since they are indistinguishable using this nomenclature.[16] Conventional MD simulations have been performed up to 1 μs, and replica exchange simulations up to 40 ns. They have observed similar limiting J

values (scalar $^{3}J_{HH}$ coupling constants) in both methods and simulation results were in very good agreement with the available experimental data. However, rotamer populations calculated from the MD data have shown some deviations from the populations calculated via NMR J-values. They suggest that this is due to the approximations involved in experimental calculations using NMR J-values.[15]

Our ultimate goal is to develop FF parameters for phospholipid molecules using consistent models for the ester, glycerol, phosphate, and hydrocarbon chain functional groups. In developing FF parameters for lipid bilayers, we combine FF parameters from models representing the functional groups of the phospholipid molecule. Here, we attempt to develop FF parameters for the glycerol part of the phospholipid head group. After determining the availability of experimental thermodynamic data for comparison, four polyol molecules were chosen, namely glycerol, 1, 2-ethanediol, 1, 2-propanediol and 1, 3-propanediol. Water, methanol, or ethanol were used as solvents to form binary mixtures.

Our approach to developing a simple, nonpolarizable, united atom FF is focused on reproducing the experimental Kirkwood Buff integrals (KBIs) as a function of composition for binary mixtures. The partial atomic charges of the models are parameterized to reproduce the KBIs of solution mixtures.[17-18] Kirkwood Buff (KB) theory is an exact theory which relates the microscopic structure of a solution to the thermodynamic properties of that solution.[19] In general, KBIs indicate how strong or weak the intermolecular interactions between solute-solute, solute-solvent and solvent-solvent are in a qualitative manner. This is beneficial in obtaining the correct balance between solute-solute and solute-solvent interactions in binary mixtures of our models.[18] Numerous solute models have been developed using the KBFF approach.[17, 20-27]

Nonpolarizable FFs are much simpler and easy to use for long time and large-scale system simulations. Polarizable FFs are computationally expensive. We have developed our non-polarizable FF using KB theory and the experimental solution activities to develop models for biomolecules in the condensed phase. As we attempt to reproduce the properties of solvent mixtures, not just the pure liquid, we obtain more accurate FFs with fixed effective partial atomic charges, which are usually sensitive to the changes in composition. Consequently, we would argue that polarizability is included in our FFs, but in an implicit manner.

2.2 Methodology

2.2.1 Kirkwood-Buff Analysis of the Experimental Data

KB integrals are calculated using the composition dependent chemical potential derivatives (μ_{ij}), partial molar volumes ($\bar{V}_i$), and isothermal compressibilities (κ_T) of binary mixtures obtained from the literature. Since the calculation of KBIs involves activity data, we can develop models that are sensitive to the properties of the solution mixtures.

The KBIs (G_{ij}) are calculated using the formula shown below,

$$G_{ij} = 4\pi \int_0^\infty \left[g_{ij}^{\mu VT}(r) - 1 \right] r^2 dr \tag{2.1}$$

where g_{ij} is the corresponding radial distribution function between species i and j, and r is the distance between the corresponding center of masses. The μVT indicates that the radial distribution function is calculated at constant number of chemical potentials, constant volume and constant temperature. (grand canonical ensemble). More details concerning KB theory can be found in our

previous studies.[18, 27-28] Large positive values for the G_{ij}s indicate net attractive interactions, while negative values indicate net repulsive interactions.

KB analyses of experimental data for six binary mixtures were performed, according to Ben Naim's inversion of KB theory,[29] and as discussed in previous studies.[17, 20-22, 24, 27] The experimental KBIs are calculated as follows,[30]

$$G_{ij} = RT\kappa_T - \frac{\bar{V}_i \bar{V}_j}{(1 + f_{cc})V_m} \tag{2.2}$$

$$G_{ii} = G_{ij} + \frac{1}{x_i}\left(\frac{\bar{V}_j}{1 + f_{cc}} - V_m\right) \tag{2.3}$$

where R is the gas constant, T is the temperature, V_m is the molar volume ($V_m = V/(N_c+N_s)$), and x_i is the mole fraction of component i. Here, c and s denote the cosolvent and solvent of a binary mixture, respectively, and

$$\frac{1}{RT}\left(\frac{\partial \mu_c}{\partial \ln x_c}\right)_{p,T} = 1 + \left(\frac{\partial \ln f_c}{\partial \ln x_c}\right)_{p,T} = 1 + f_{cc} \tag{2.4}$$

where μ_c is the chemical potential of the cosolvent, and fc is the cosolvent activity coefficient on the mole fraction scale with the pure cosolvent as the standard state.

The partial molar volumes at any given composition is given by,

$$Y_i = X_m^E - x_j\left(\frac{\partial X_m^E}{\partial x_j}\right)_{p,T} \tag{2.5}$$

with $Y = \bar{V}_m^E$ and $X=V$. The excess molar Gibbs energy values (G_m^E) are computed by assigning Y as excess chemical potential ($\mu_i^E/RT = \ln f_i$) and $X = G/RT$ in equation (2.5).[17]

Since it has shown that there is no significant effect of variations in κ_T on the KBIs,[31] the solution κ_T is calculated as follows,

$$\kappa_T = \emptyset_c \kappa_{T,c}^0 + \emptyset_s \kappa_{T,s}^0 \quad (2.6)$$

where $\kappa_{T,c}^0$ is the compressibility of the pure cosolvent, and $\emptyset_c$ is the volume fraction of the cosolvent in solution ($\emptyset_c = \rho_c \bar{V}_c$, where ρ_c is the number density and $\bar{V}_c$ is the partial molar volume).

2.2.2 Experimental Sources for Composition and Activity Data

Fitting constants for the excess molar volumes (V_m^E) and excess molar Gibbs energies (G_m^E) for glycerol/water (GLY/HOH) and 1, 2-ethanediol/water (EDL/HOH) systems were obtained directly from Marcus's book of solvent mixtures.[32] Another set of fitting constants for G_m^E of the GLY/HOH and glycerol/ethanol (GLY/EOH) systems were obtained at 293.15 K.[33] Excess molar volume data for GLY/EOH mixture were obtained at 298.15 K.[34] From the raw activity coefficients for the 1, 2-ethanediol-methanol (EDL/MOH) mixture,[35] G_m^E was calculated as a function of composition and fitted using Redlich-Kister polynomial.

$$X_m^E = x_c x_s \sum_{i=0}^{n} a_i (x_s - x_c)^i \quad (2.7)$$

where a_i's are fitting constants, x_c and x_s are the cosolvent and solvent mole fractions.

V_m^E data for this mixture were obtained from the same resource. Water activity data for 1,2-propanediol-water (1,2-PDL/HOH) and 1,3-propanediol-water (1,3-PDL/HOH) mixtures were used to calculate G_m^E for binary mixtures.[36] Fitting constants for V_m^E for propanediol systems were also obtained from literature.[37] Finally, the isothermal compressibility for all pure compounds were obtained from the literature.[38-39]

2.2.3 Parameter Development

In general, for the KBFF models, the bond, angle and torsion parameters are adopted from the Gromos FF[40]. However, there are some exceptions in this study. Torsion angle parameters were parameterized in order to reproduce experimentally,[41] and quantum mechanically,[10] obtained conformation energies for GLY and EDL respectively. The parameters obtained for H-O-C-C and O-C-C-O angles of EDL were then also used for 1, 2-propanediol. A summary of the torsion angle parameters used here is shown in Table 2.1. Nonbonded interactions are treated with a Lennard Jones (LJ) 6-12 and Coulomb potential. The σ and ε parameters for the LJ term of the carbon atoms were also adopted from the Gromos FF. In an effort to obtain a better agreement with experimental densities for GLY and 1, 2-PDL systems, the σ of the CH1 atom type was reduced by 20% from its original value. Effective partial atomic charges were iteratively varied to reproduce the experimental KBIs in the condensed phase as a function of composition. Here, for the GLY, EDL, and 1,2-PDL we could not use the charges already established for MOH,[17] even though we could simply model these molecules by just combining two or three MOHs. The partial atomic charges of MOH were scaled down to obtain the partial atomic charges for GLY, EDL, and 1,2-PDL as there are strong interactions between consecutive -C-O-H groups. A summary of the nonbonded parameters is shown in Table 2.2. Errors for the simulated KBIs were calculated from the averages of multiple 5 ns runs.

2.2.4 Molecular Dynamics Simulations

Molecular dynamics simulations were performed for six binary mixtures in the isothermal-isobaric ensemble (NpT) at 298.15 K and 1 atm pressure using the Gromacs software package (version 4.6 or 4.6.1)[42]. Water, methanol, and ethanol were used as solvents to make sure that our

models reproduce reasonable experimental results in different solvent environments. The previously developed KBFF methanol[17] and ethanol[43] models were used while SPC/E water model[44] was used for water. All the mixtures were simulated at mole fractions of 0.2, 0.4, 0.6, 0.8 and 1.0 in 10 nm length cubic boxes.

Table 2.1 Torsion angle parameters obtained for the KBFF models.

Model	Angle	$k_\emptyset$ (kJ/mol)	n	Model	Angle	$k_\emptyset$ (kJ/mol)	n
EDL	H-O-CH2-CH2	3.25	1	**GLY**	CH2-CH-CH2-O	-10.375	1
GLY	H-O-CH-CH2	0.5	2			0.25	2
1,2-PDL	H-O-CH-CH2	5.5	3			4.25	3
1,2-PDL	H-O-CH2-CH						
EDL	O-CH2-CH2-O	21.625	1	**1,2-PDL**	CH3-CH-CH2-O	-0.375	1
1,2-PDL	O-CH2-CH-O	0.25	2	**1,3-PDL**	CH2-CH2-CH2-O	0.25	2
		4.25	3			4.25	3
GLY	O-CH2-CH-O	20.0	1	**1,2-PDL**	H-O-CH-CH3	0.75	1
		0.25	2	**1,3-PDL**	H-O-CH2-CH2	0.5	2
		4.25	3			3	3
GLY	H-O-CH2-CH	0.75	1				
		0.5	2				
		5.5	3				

Dihedral angles were defined according to $V_\emptyset = k_\emptyset(1 + \cos\ (n\emptyset - \ \delta))$, for all the angles δ is taken to be zero.

Table 2.2 Nonbonded parameters obtained for the KBFF models

Atom type	σ (nm)	ε (kJ/mol)	q	
			EDL, GLY, 1,2-PDL	1,3-PDL
CH2	0.4070	0.4105	0.2910	0.3000
O3	0.3192	0.6506	-0.7954	-0.8200
H	0.1580	0.0880	0.5044	0.5200
CH1	0.4015	0.0949	0.2910	
CH3	0.3748	0.8672	0	

The temperature and pressure were maintained using weak coupling techniques.[45] In particular, the v-scale temperature coupling,[46] and Berendsen pressure coupling,[45] were used with 0.1 and 0.5 ps relaxation times. A compressibility of 4.5×10^{-5} bar^{-1} was used for the pressure coupling. All bonds were constrained with the LINCS algorithm.[47] The Particle Mesh Ewald (PME) was used to calculate the coulombic interactions[48]. Twin range cutoffs of 1.5 nm and 1.0 nm were used for van der Waals and electrostatic interactions. All the systems were equilibrated for 1 ns followed by production runs up to 20 ns, or further depending on the convergence of the systems. Glycerol mixtures were run up to 30 ns and while pure glycerol was run up to 50 ns. Configurations were saved at every 1.0 ps for analysis of the systems.

2.2.5 Analysis of simulated solution properties

The enthalpy of vaporization (ΔH_{vap}) of the pure liquid was calculated using the average total and intramolecular potential energies,

$$\Delta H_{vap} = \frac{-(E_{tot} - E_{intra})}{N} + RT \tag{2.8}$$

Here, E_{intra} term is the intramolecular potential energy in the liquid phase. The total potential energy of the liquid is denoted as E_{tot} and N is the total number of molecules in the liquid. The ΔH_{vap} was calculated by assuming that the intramolecular interactions in the gas phase are similar to the liquid phase, while intermolecular interactions in the gas phase are negligible. Furthermore, assuming ideal gas behavior the pV work done in the gas phase is RT, and since the volume of a liquid is negligible compared to the volume of a gas, the work done by the liquid phase is assumed to be zero.

The enthalpy of mixing values (H_m^E) were calculated using,

$$H_m^E = H_{m,sol} - x_c H_{m,c}^0 - x_s H_{m,s}^0 \tag{2.9}$$

where $H_{m,c}^0$, and $H_{m,s}^0$ are the molar enthalpies of the pure components, while $H_{m,sol}$ is the molar enthalpy of the solution mixture.

The V_m^E values were calculated using,

$$V_m^E = V_{m,sol} - x_c V_{m,c}^0 - x_s V_{m,s}^0 \tag{2.10}$$

$$V_{m,sol} = \frac{x_c M_c + x_s M_s}{d_{sol}}$$

where V_m is the molar volume of the solution, M_c and M_s are the molar masses of pure components, and d_{sol} is the mass density of the solution. The self-diffusion constants (D) were calculated using the code g_msd included in the Gromacs package and center of mass mean square deviation being employed according to the formula,

$$D = \lim_{t\to\infty} \frac{1}{6t} \langle [r_i(t) - r_i(0)]^2 \rangle \tag{2.11}$$

where t is the time and r_i is the position of the molecule.

The dielectric constants (ε) were computed using dipole moment fluctuations with the help of Gromacs code g_dipole. Here, the reaction field permittivity is assumed to be infinity.

Additional NVT simulations were performed up to 10 ns at the same temperature to calculate the shear viscosity (η) of the pure liquids. The Einstein relation,[49]

$$\eta = \frac{1}{2}\frac{V}{k_B T}\lim_{t\to\infty}\frac{d}{dt}\left\langle\left(\int_{t_0}^{t_0+t} P_{\alpha\beta}(t')dt'\right)^2\right\rangle_{t_0} \quad (2.12)$$

was then used where k_B is the Boltzmann constant, $P_{\alpha\beta}$ are the off-diagonal elements of the pressure tensor. P_{xy}, P_{xz}, and P_{yz}, were saved at every 2 fs. The Gromacs code g_energy was used to obtain the $P_{\alpha\beta}$ as a function of time. Errors were calculated using four runs of 2.5 ns length for all polyols, while only two runs of 2.5 ns length were used for SPC/E water. Linear regions of about 200-500 ps length were selected out of each 2.5 ns run to calculate η.

2.3 Results and Discussion

The KBIs obtained for the six binary mixtures are shown in Figure 2.2 to Figure 2.4. The experimental KBIs for EDL/HOH mixtures (Figure 2.2) are reasonably reproduced, although there were significant error bars for the simulated $G_{HOH/HOH}$ at higher molar fractions of EDL. The simulated KBIs for EDL/MOH system underestimate the experimental $G_{EDL/EDL}$ at lower mole fractions. However, we were not able to adjust these values and maintain the $G_{EDL/MOH}$ values in good agreement with the experimental data. The experimental KBIs for GLY/HOH mixtures well reproduced except for $G_{HOH/HOH}$ at higher mole fractions of GLY (Figure 2.3), as also seen for EDL/HOH mixture. This is due to sampling problems at extreme mole fractions where similar uncertainties allow in the experimental data as well (experimental data shown in Figure 2.3 for the

GLY/HOH mixture) The experimental data show that the KBIs of GLY/EOH mixtures are composition dependent. However, we do not observe such behavior in our simulation results. This mixture acts more like an ideal solution mixture as the KBIs are reasonably independent of composition.[50] This ideal behavior may be explained, to some extent, as both the GLY and EOH models have analogous functional groups. For the 1,2-PDL/HOH and 1,3-PDL/HOH mixtures the models reasonably well reproduced the KBIs (Figure 2.4) with exceptions for $G_{HOH/HOH}$ at some mole fractions. The partial atomic charge distribution for 1,3-PDL is different than for other three models as the two hydroxyl groups are far apart.

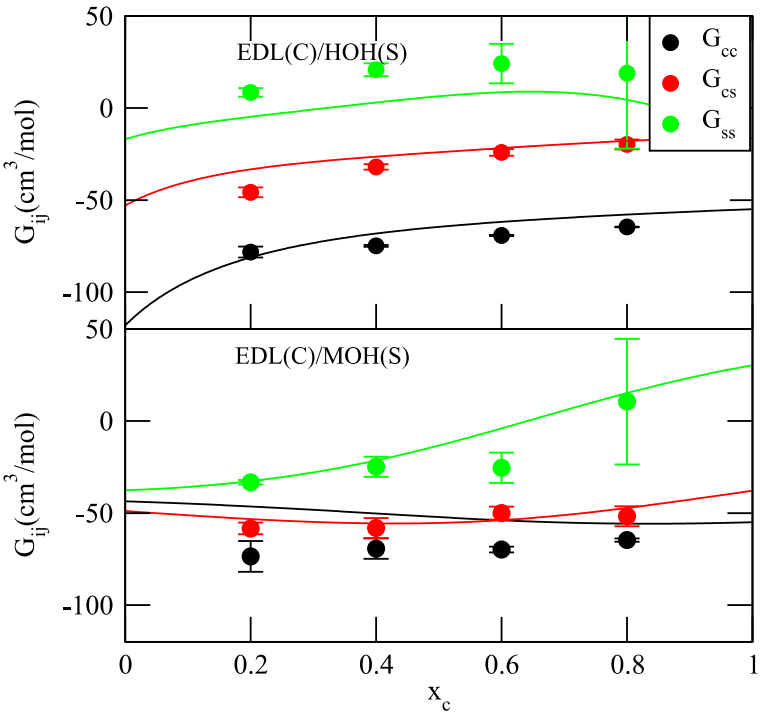


Figure 2.2 Experimental (lines) and simulated (points) data for the KBIs of EDL/HOH and EDL/MOH mixtures as a function of composition at 298 K. Error bars are the standard deviations calculated for 5 ns block averages of simulations.

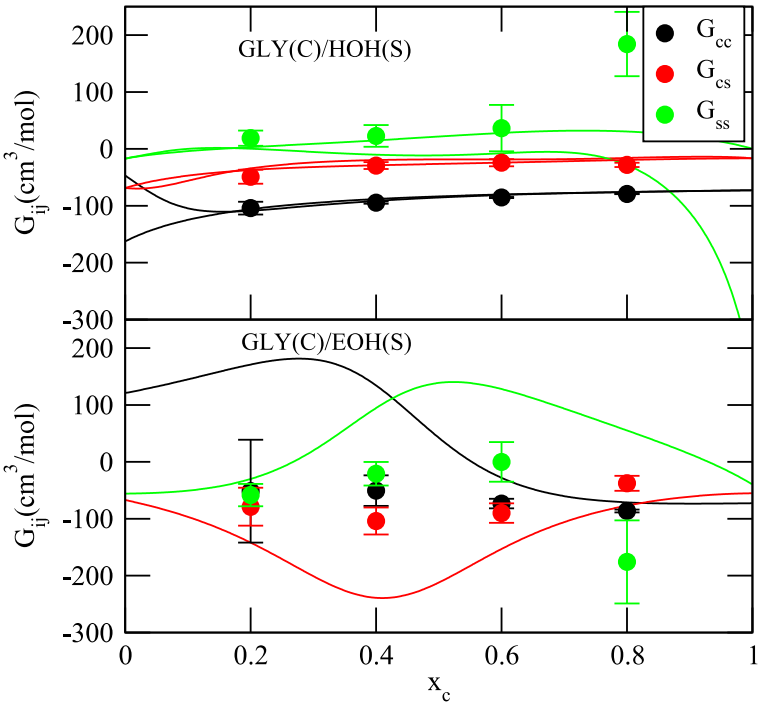


Figure 2.3 Experimental (lines) and simulated (points) data for the KBIs of GLY/HOH and GLY/EOH mixtures as a function of composition. Two sets of experimental data obtained from two different set of activity data are shown for the GLY/HOH mixture at 298 K and 293 K. Error bars are the standard deviation calculated for 5 ns block averages of simulations.

The H_m^E values for the six binary mixtures were calculated as a function of composition (Figure 2.5). According to the experimental results, all four polyol/HOH systems show a favorable H_m^E and our simulated data agree with that observation even though, in general, it is not easy to reproduce experimental H_m^E using simulations.[17] Unfortunately, we could not find experimental data for H_m^E of the GLY/EOH system in the literature. The EDL/MOH system shows an unfavorable enthalpy of mixing, and the simulated results agree with that, even though our simulations slightly overestimate the H_m^E values.

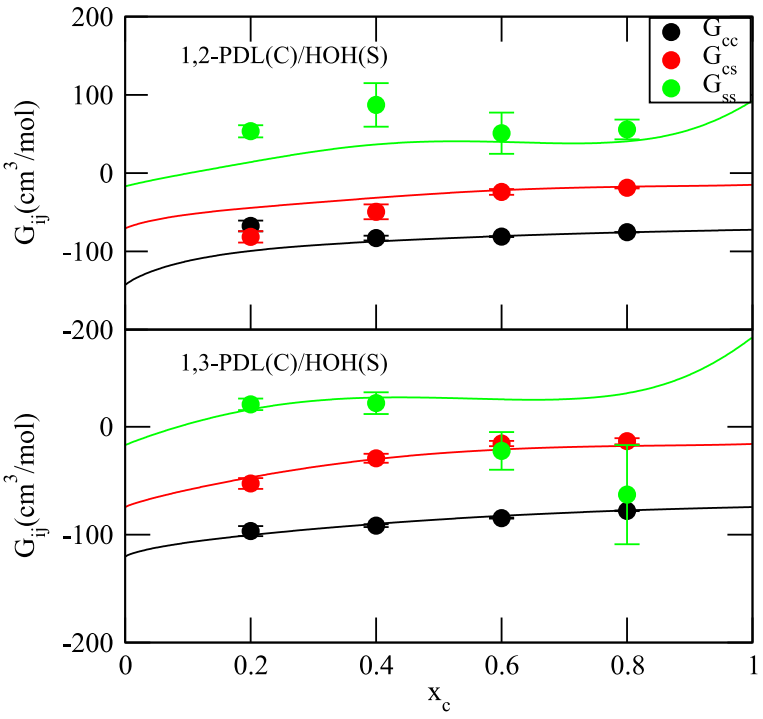


Figure 2.4 Experimental (lines) and simulated (points) data for the KBIs of 1,2-PDL/HOH and 1,3-PDL/HOH mixtures, as a function of composition at 298 K. Error bars are the standard deviation calculated for 5 ns block averages of simulations.

The experimental and simulated V_m^E values obtained for all the binary mixtures are displayed in Figure 2.6. Our four models display low simulated densities compared to the experimental value when mixed with water, as depicted in Table 2.3. However, while the polyol/HOH systems underestimate the experimental V_m^E, the polyol/alcohol systems overestimate the V_m^E. Again, it is not easy to correct both behaviors with the simple models used here. The united atom FF does not explicitly include hydrogen atoms in models and this might be a reason for the low density observed in our systems.

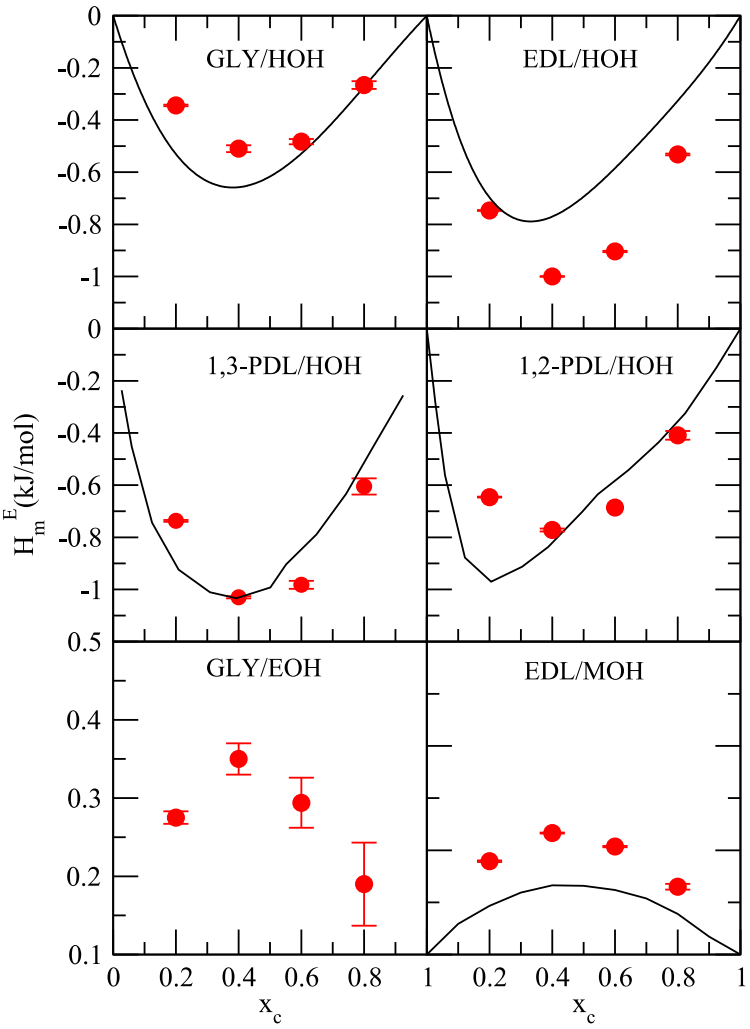


Figure 2.5 Experimental[32, 35, 51] (lines) and simulated (points) enthalpies of mixing for the six binary mixtures as a function of composition at 298 K. No experimental data were found for the GLY/EOH mixture. Error bars are the standard deviation calculated for 5 ns block averages of simulations. All the experimental data correspond to 298 K.

In addition to the thermodynamic properties of binary mixtures, the kinetic properties of mixtures, such as self-diffusion constants (D) and viscosity (η) were also investigated during the parameterization procedure. The translational self-diffusion constants for GLY/HOH and EDL/HOH were calculated as a function of composition (Figure 2.7). GLY has a very small self-diffusion constant being a heavy molecule and a viscous liquid. The GLY molecule's strong ability to make an extensive hydrogen bonding network due to their three hydroxyl groups reduces the

diffusion of both GLY and HOH with mole fraction. As can be seen in Figure 2.7, our GLY/HOH mixtures reproduce the experimental self-diffusion constant for almost all mole fractions. Pure EDL displays a higher D_c compared to the experiential value. This can be partially explained by the lower density of EDL provided by our model. Nevertheless, the appropriate experimental trend is reproduced by the EDL/HOH mixtures. We observe a significant error for D_c of pure 1, 2-PDL as it overestimates the experimental D_c even though there is not a much difference in the simulated and experimental density data. The reason for lower D_c of pure 1,3-PDL cannot be explained by the low-density value. In addition, the experiments indicate similar results for the D_c of pure 1,2-PDL and 1,3-PDL, while our models provide disparate results. We could not find a clear reasoning for the contrasting results observed for the diffusion constants of pure liquids.

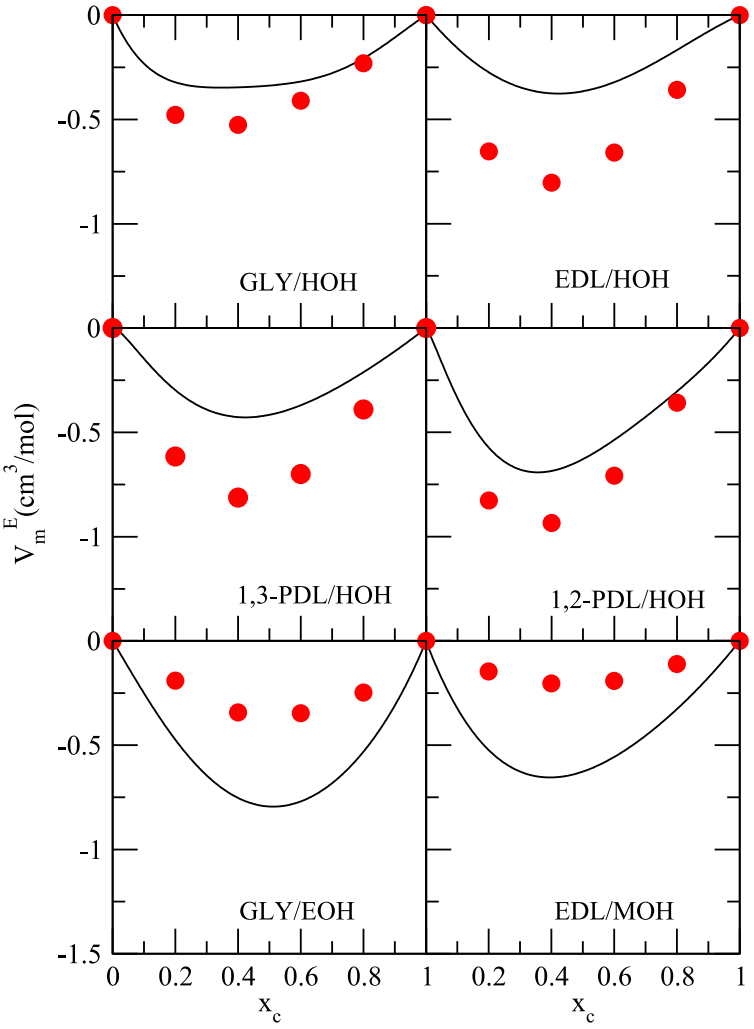


Figure 2.6 Experimental (lines) and simulated (points) excess volumes of mixing for six binary mixtures as a function of composition at 298 K.

Viscosity calculations for the pure liquids were performed using the Einstein equation as mentioned in the methods section. As can be seen in Table 2.3, our simulated η values for polyols are relatively low compared to the experimental results, except for 1,3-PDL where it gives a relatively high η. The viscosity of pure SPC/E water shows good agreement with experimental data. Furthermore, we found significantly high errors for the polyol calculations, especially for the highly viscous GLY systems. Viscosity calculations using the Einstein equation are not easy for

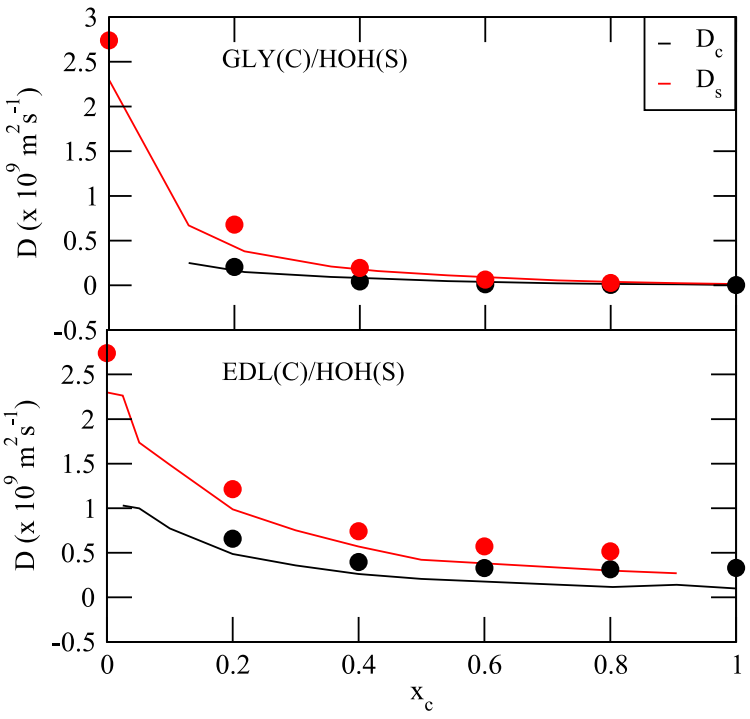


Figure 2.7 Experimental (lines) and simulated (points) translational self-diffusion coefficients (D_c and D_s) for GLY/HOH[52] and EDL/HOH[53] mixtures as a function of composition at 298 K.

polyols. First, viscosity is a system property, not a molecular property, and this makes it difficult to obtain good statistics.[54] Furthermore, in a simulation, the pressure fluctuates significantly and it is therefore difficult to converge with time. Interestingly, in the literature, we find different η values at somewhat similar temperatures, 396.5[55], 1670.1310[56] and 1412[57] cP at 298.15, 292.15 and 293.15 K respectively by different groups. This might indicate that it is difficult to get consistent result for GLY even with experimental techniques.

Table 2.3 Experimental and simulated properties of pure liquids. All the simulations were performed at 298 K and experimental values were obtained from literature at 298 K unless stated otherwise.

	Density/ρ (g/cm^3)		Diffusion/D ($10^{-9}m^2/s$)		Dielectric/ε		Viscocity/η (cP)		E_{pot} (kJ/mol)	ΔH_{vap} (kJ/mol)	
Model	**MD**	**exp**	**MD**	**exp**	**MD**	**exp**	**MD**	**exp**	**MD**	**MD**	**exp**
EDL	0.974	1.110[35]	0.33	0.0961[58]	42	[a]38.66[59]	6(2)	18.68[60]	-62.46	64.07	65.69[11]
GLY	1.146	1.258[61]	0.0022	0.0025[52]	36	[a]41.14[59]	317(135)	396.5[55]	-55.43	97.10	91.63[12]
1,2-PDL	0.974	1.032[62]	0.12	[a]0.041[63]	35	28.360[62]	9(2)	43.428[62]	-57.05	104.93	
1,3-PDL	0.975	1.049[62]	0.017	[a]0.041[63]	32	34.299[62]	56(8)	40.067[62]	-28.54	92.76	
HOH	0.996	0.997[64]	2.74	2.299[53]	65	78.54[59]	0.84(0.06)	0.890[65]	-46.70	49.18	43.98[66]
MOH	0.762	0.787[67]	2.10	2.42[68]	35	32.60[69]	0.59(0.08)	0.55[70]	-42.67	45.15	37.43[66]
EOH	0.754	0.785[67]	0.91	1.075[68]	22	24.35[69]	0.9(0.4)	1.1[71]	-41.50	43.97	42.26[66]

[a]Temperature at 293 K. Errors for the viscosity calculations are shown in parenthesis

The dielectric constants (ε) of the four polyol/HOH mixtures were computed and are displayed in Figure 2.8. Only the simulated ε of EDL/HOH as a function of composition is compared with experimental data. For other systems, only the pure components' experimental data are shown in Table 2.3. We note that the average dipole moment of pure and a few higher mole fractions of GLY, 1,2-PDL, and 1,3-PDL were not converged to zero with the simulation time. Pure GLY underestimates and 1,3-PDL overestimates the experimental ε values while EDL and 1,2-PDL show good agreement with experimental ε values. The SPC/E water model has a ε which is about 18% too low compared to the real water, and this has a consistent effect on the ε of polyol/HOH mixtures.

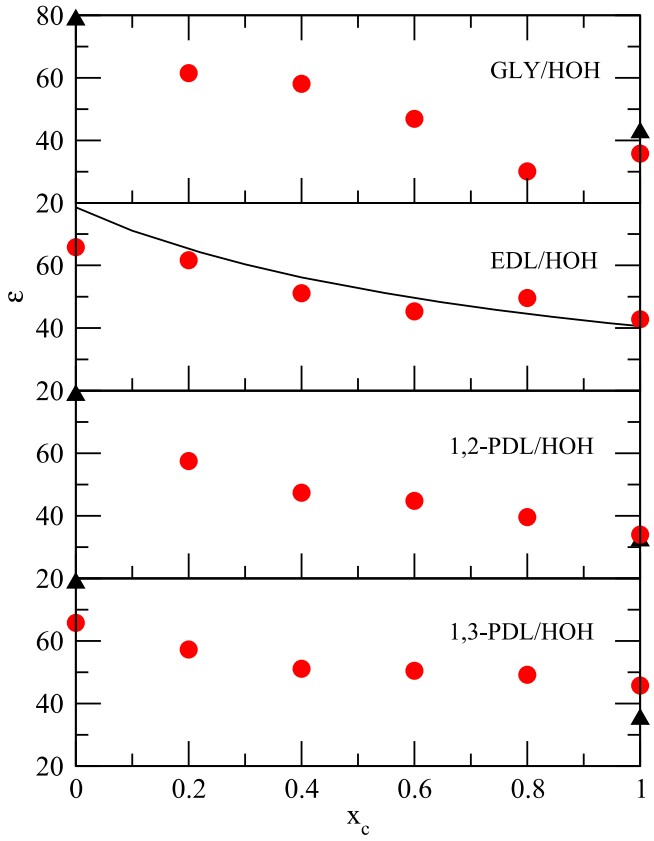


Figure 2.8 Experimental (lines) and simulated (points) dielectric constants for polyol/HOH mixtures as a function of composition at 298 K. Experimental dielectric constants in the full range of composition were only found for the EDL/HOH[72] mixture. For other systems experimental dielectric constants of the pure liquids are shown as triangles.

There are four different torsion angles for GLY and two for EDL. Figure 2.9 displays the relative free energies of these torsion angles as calculated for pure GLY and EDL. Four torsional angle potentials for GLY, and two torsional angle potentials for EDL can be observed. The percent gauche conformation (gauche$^-$ + gauche$^+$) for the EDL and GLY systems were then calculated as a function of composition.

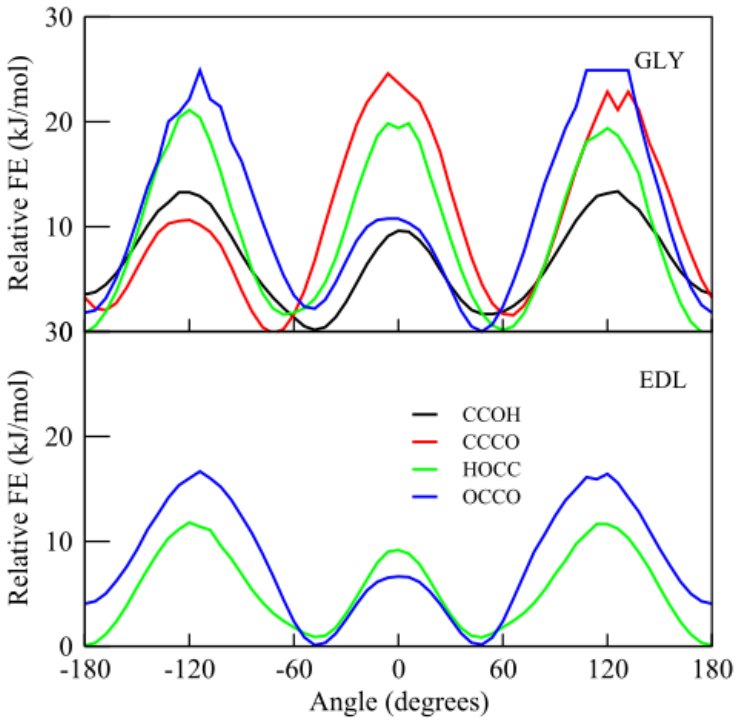


Figure 2.9 Simulated relative free energies of four different torsion angles of the pure GLY and two different torsion angles of the pure EDL at 298 K.

As can be seen in Figure 2.10, the percent gauche conformations do not depend on the composition for either of the systems except the OCCO angle of GLY. In addition, according to the simulated results of both systems, the gauche conformations are more prominent compared to the trans conformation. An experimental NMR study for EDL in nematic-lyprotic solution has shown that EDL exists in a 100% gauche conformation in aqueous solution.[73]Another NMR study by Pachler and Wessels has shown that 88% of EDL in aqueous solution, and 86% of EDL in the pure liquid, adopt the gauche conformation.[74] This agrees with our results. Experimental studies have also shown that the percent gauche conformation does not depend on the concentration of EDL.[75] Ten distinct conformations of EDL were defined using backbone torsion angles HOCC, OCCO and CCOH namely tGg', gGg', g'Gg', gTg', tTt, tTg, gTg, gGg, tGt, tGg where g, g' and t denotes

gauche clockwise, gauche counterclockwise and trans respectively.[10, 76] The comparison between the relative energies calculated by QM,[10] and the relative free energies observed in KBFF are shown in Table 2.4. We observe that our model for 1,2-EDL works reasonably well even without the explicit inclusion of polarizability in the FF.

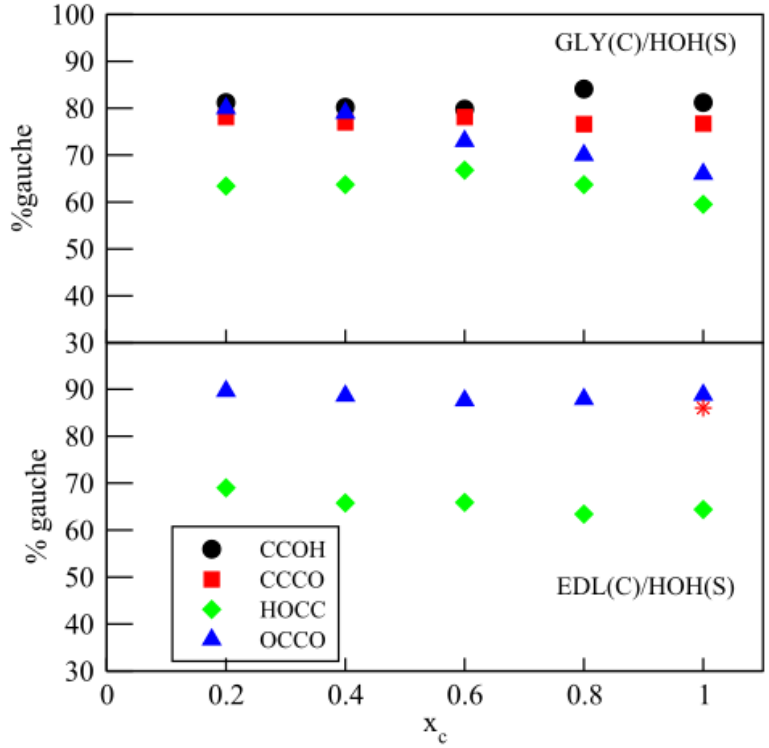


Figure 2.10 Simulated conformational populations as a function of composition for four different torsion angles of GLY and two different torsion angles of EDL at 298 K. Experimental gauche percentage of the pure EDL liquid is shown in red star.

Six different backbone conformations of GLY have been defined by Bastiansen in 1949 based on the terminal oxygen atoms and three carbon atoms.[8] Using the dihedral angles of O1-C1-C2-O2 (ϕ_1) and O2-C2-C3-O3 (ϕ_2) (Figure 2.1) six conformations, αα, αβ, αγ, ββ, βγ, γγ were defined as discussed in the introduction. A comparison between the experimental data and the simulated values for the conformer distribution is shown in Table 2.5. The experimental data were obtained from NMR spectroscopy experiments on 5% GLY in D_2O solution.[41] The simulated data

Table 2.4 Comparison of QM[10] and simulated relative energies obtained for ten different conformations of pure EDL.

Conformation	ΔE (kJ/mol)	
	QM	MD
tGg'	0	0.0
gGg'	1.7	3.7
g'Gg'	5.9	6.2
gTg'	12.6	7.9
tTt	12.7	10.0
tTg	12.8	6.4
gTg	12.9	8.5
gGg	14.2	7.8
tGt	15.5	19.2
tGg	16.8	11.3

Table 2.5 Comparison of experimental[41] and simulated GLY backbone conformations.

	Population (%)	
Conformation	MD	exp
αγ	43	28-30
αβ	27	20-21
αα	15	18-21
βγ	11	15-17
γγ	2	10-12
ββ	1	5

shown in Table 2.5. were calculated at the mole fraction of 0.2, which has the closest concentration to the experiment ($x_c \approx 0.01$). As can be seen, the GLY model shows reasonable agreement for most of the conformations. In particular, we have a well- distributed conformer population for GLY rather than a single conformation. Furthermore, an electron scattering analysis of gas phase glycerol has shown that αα and αγ configurations are most probable while the γγ arrangement is improbable.[8] As mentioned earlier, it is challenging to develop a FF for polar polyols due to their strong inter and intramolecular interactions.[4, 7] These results show similar features to the experimental aqueous phase distribution except for the αβ population which is also probable in aqueous phase.

Properties of the pure liquids calculated using the KBFF models are shown in Table 2.3. One of the main features is the densities of the pure liquids are underestimated by our models to some extent. One possible reason for the reduced density of polyols might be the united atom approach used here where the carbon and hydrogen atoms fused into a single bead. The united atom FF works well to reproduce the density of alkanes,[77] but it seems more problematic for aliphatic alcohols consisting of polar hydroxyl groups. As mentioned in the methods section we reduced the σ of CH1 atom type by 20% from its original value in order to improve the density of GLY and 1, 2-PDL models. However, we need to remain consistent with our previous models and therefore we only changed the σ of the CH1 atom type as the CH2 and CH3 atom types are already being used in our previously published models.[17, 21] Despite the lower density of the GLY model, it provides good agreement with the experimental self-diffusion constant. One other noticeable behavior is that the 1, 2-PDL model which provides the lowest error for the density, shows the largest error for the self-diffusion constant. Hence, we do not observe a correlation between density and self-diffusion. Therefore, it is hard to decide how to improve the properties.

The enthalpy of vaporization of pure liquids were calculated using the potential energy of systems. Our results show reasonable agreement with experimental values for pure GLY and EDL, even though we have not corrected our models for polarizability unlike the SPC/E water model.[44] However, we could not find experimental ΔH_{vap} for other pure liquids to compare with our simulation results.

2.4 Conclusions

Simple nonpolarizable FFs have been derived for EDL/HOH, EDL/MOH, GLY/HOH, GLY/EOH, 1,2-PDL/HOH, and 1,3-PDL/HOH binary mixtures using the KB theory of solutions to optimize he effective charges. All the models reasonably well reproduce the experimental KB integrals except for the GLY/EOH system. Good agreement with the experimental KB integrals ensures that there is a correct balance between solute-solvent interactions. In addition to the KB integrals, other thermodynamic, kinetic and physical properties for binary mixtures and the pure liquids were calculated. Most of the models reasonably well reproduce these properties, while there were some deviations observed for some properties such as densities, excess volumes of mixing, and viscosities. The population distributions obtained for EDL show reasonably good agreement with the experimental data, while for the GLY most of the conformations display reasonable agreement with experimental values. As these are simple nonpolarizable FF models we can not expect perfect correlation with all the simulated vs experimental properties.

2.5 References

1. Veneral, J. G.; Benazzi, T.; Mazutti, M. A.; Voll, F. A.; Cardozo-Filho, L.; Corazza, M. L.; Guirardello, R.; Oliveira, J. V., *The Journal of Chemical Thermodynamics* **2013,** *58,* 398-404.

2. Bär, A., *Food chemistry* **1985,** *16* (3-4), 231-241.

3. Ryabov, Y. E.; Hayashi, Y.; Gutina, A.; Feldman, Y., *Physical Review B* **2003,** *67* (13), 132202.

4. Vanderkooi, J. M.; Dashnau, J. L.; Zelent, B., *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2005,** *1749* (2), 214-233.

5. Reiling, S.; Schlenkrich, M.; Brickmann, J., *Journal of Computational Chemistry* **1996,** *17* (4), 450-468.

6. Blieck, J.; Affouard, F.; Bordat, P.; Lerbret, A.; Descamps, M., *Chemical physics* **2005,** *317* (2), 253-257.

7. Jahn, D. A.; Akinkunmi, F. O.; Giovambattista, N., *The Journal of Physical Chemistry B* **2014,** *118* (38), 11284-11294.

8. Bastiansen, O.; Borgiel, H.; Saluste, E., *Acta Chemica Scandinavica* **1949,** *3*, 415-421.

9. Akinkunmi, F. O.; Jahn, D. A.; Giovambattista, N., *The Journal of Physical Chemistry B* **2015,** *119* (20), 6250-6261.

10. He, X.; Lopes, P. E.; MacKerell, A. D., *Biopolymers* **2013,** *99* (10), 724-738.

11. Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D., *Journal of computational chemistry* **2008,** *29* (15), 2543-2564.

12. Hatcher, E. R.; Guvench, O.; MacKerell, A. D., *Journal of Chemical Theory and Computation* **2009,** *5* (5), 1315-1327.

13. Ferrando, N.; Lachet, V.; Teuler, J.-M.; Boutin, A., *The Journal of Physical Chemistry B* **2009,** *113* (17), 5985-5995.

14. Kulschewski, T.; Pleiss, J., *Molecular Simulation* **2013,** *39* (9), 754-767.

15. Yongye, A. B.; Foley, B. L.; Woods, R. J., *The Journal of Physical Chemistry A* **2008,** *112* (12), 2634-2639.

16. Towey, J.; Soper, A.; Dougan, L., *Physical Chemistry Chemical Physics* **2011,** *13* (20), 9397-9406.

17. Weerasinghe, S.; Smith, P. E., *The Journal of Physical Chemistry B* **2005,** *109* (31), 15080-15086.

18. Ploetz, E. A.; Weerasinghe, S.; Kang, M.; Smith, P. E.; Smith, P.; Matteoli, E.; O'Connell, J., *Accurate Force Fields for Molecular Simulation*. CRC Press: Boca Raton, FL: 2013.

19. Kirkwood, J. G.; Buff, F. P., *The Journal of Chemical Physics* **1951,** *19* (6), 774-777.

20. Weerasinghe, S.; Smith, P. E., *The Journal of chemical physics* **2003,** *119* (21), 11342-11349.

21. Weerasinghe, S.; Smith, P. E., *The Journal of chemical physics* **2003,** *118* (23), 10663-10670.

22. Weerasinghe, S.; Smith, P. E., *The Journal of Physical Chemistry B* **2003,** *107* (16), 3891-3898.

23. Weerasinghe, S.; Smith, P. E., *The Journal of chemical physics* **2004,** *121* (5), 2180-2186.

24. Kang, M.; Smith, P. E., *Journal of computational chemistry* **2006,** *27* (13), 1477-1485.

25. Bentenitis, N.; Cox, N. R.; Smith, P. E., *The Journal of Physical Chemistry B* **2009,** *113* (36), 12306-12315.

26. Gee, M. B.; Cox, N. R.; Jiao, Y.; Bentenitis, N.; Weerasinghe, S.; Smith, P. E., *Journal of chemical theory and computation* **2011,** *7* (5), 1369-1380.

27. Ploetz, E. A.; Smith, P. E., *Physical Chemistry Chemical Physics* **2011,** *13* (40), 18154-18167.

28. Ploetz, E. A.; Bentenitis, N.; Smith, P. E., *Fluid phase equilibria* **2010,** *290* (1), 43-47.

29. Ben-Naim, A., *The Journal of Chemical Physics* **1977,** *67* (11), 4884-4890.

30. Ben-Naim, A., *Molecular Theory of Solutions*. Oxford University Press: New York, 2006.

31. Matteoli, E.; Lepori, L., *The Journal of chemical physics* **1984,** *80* (6), 2856-2863.

32. Marcus, Y., *Solvent mixtures: properties and selective solvation*. CRC Press: 2002.

33. Zaoui-Djelloul-Daouadji, M.; Negadi, A.; Mokbel, I.; Negadi, L., *The Journal of Chemical Thermodynamics* **2014,** *69*, 165-171.

34. Savaroglu, G.; Caner, N., *Physics and Chemistry of Liquids* **2010,** *48* (1), 108-116.

35. Cratin, P. D.; Gladden, J., *The Journal of Physical Chemistry* **1963,** *67* (8), 1665-1669.

36. Marcolli, C.; Peter, T., *Atmospheric Chemistry and Physics* **2005,** *5* (6), 1545-1555.

37. Zemánková, K.; Troncoso, J.; Romaní, L., *Fluid Phase Equilibria* **2013,** *356*, 1-10.

38. Zorebski, E.; Dzida, M.; Piotrowska, M., *Journal of Chemical & Engineering Data* **2007,** *53* (1), 136-144.

39. Rodnikova, M. N.; Troitskii, V. M.; Solonina, I. A.; Val'kovskaya, T. M.; Sirotkin, D. A., *Russian Journal of Physical Chemistry A* **2013,** *87* (8), 1427-1429.

40. van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G., *Biomolecular Simulation: The {GROMOS96} manual and userguide*. Hochschuleverlag AG an der ETH Zürich: 1996.

41. Van Koningsveld, H., *Recueil des Travaux Chimiques des Pays-Bas* **1970,** *89* (8), 801-812.

42. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., *Bioinformatics* **2013,** *29* (7), 845-854.

43. Jiao, Y. The development of accurate force fields for protein simulation. Kansas state university, 2012.

44. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., *The Journal of Physical Chemistry* **1987,** *91* (24), 6269-6271.

45. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., *The Journal of Chemical Physics* **1984,** *81* (8), 3684-3690.

46. Bussi, G.; Donadio, D.; Parrinello, M., *The Journal of Chemical Physics* **2007,** *126* (1), 014101.

47. Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G., *Journal of computational chemistry* **1997,** *18* (12), 1463-1472.

48. Darden, T.; York, D.; Pedersen, L., *The Journal of chemical physics* **1993,** *98* (12), 10089-10092.

49. Hess, B., *The Journal of chemical physics* **2002,** *116* (1), 209-217.

50. Ben-Naim, A., *Molecular theory of solutions*. Oxford University Press: 2006.

51. Nagamachi, M. Y.; Francesconi, A. Z., *The Journal of Chemical Thermodynamics* **2006,** *38* (4), 461-466.

52. D'Errico, G.; Ortona, O.; Capuano, F.; Vitagliano, V., *Journal of Chemical & Engineering Data* **2004,** *49* (6), 1665-1670.

53. Ambrosone, L.; D'Errico, G.; Sartorio, R.; Costantino, L., *J. Chem. Soc., Faraday Trans.* **1997,** *93* (22), 3961-3966.

54. Smith, P. E.; van Gunsteren, W. F., *Chemical Physics Letters* **1993,** *215* (4), 315-318.

55. Lan, D.; Liu, L.; Dong, L.; Li, W.; Li, Q.; Yan, L., *Asian Journal of Chemistry* **2013,** *25* (5), 2709.

56. Fardad, K.; Farhoush, K.; Sasan, S.; Meysam, S.; Seyed Hamed, R., *Research Journal of Applied Sciences, Engineering and Technology* **2012,** *4* (17), 3095-3101.

57. Segur, J. B.; Oberstar, H. E., *Industrial & Engineering Chemistry* **1951,** *43* (9), 2117-2120.

58. Wellek, R. M.; Mitchell, R. D.; Moore, J. W., *Journal of Chemical & Engineering Data* **1971,** *16* (1), 57-60.

59. Akerlof, G., *Journal of the American Chemical Society* **1932,** *54* (11), 4125-4139.

60. Tsierkezos, N. G.; Molinou, I. E., *Journal of Chemical & Engineering Data* **1998,** *43* (6), 989-993.

61. Cristancho, D. M.; Delgado, D. R.; Martínez, F.; Abolghassemi Fakhree, M. A.; Jouyban, A., *Revista Colombiana de Ciencias Químico - Farmacéuticas* **2011,** *40*, 92-115.

62. George, J.; Sastry, N. V., *Journal of Chemical & Engineering Data* **2003,** *48* (6), 1529-1539.

63. Klein, M.; Fechete, R.; Demco, D. E.; Blümich, B., *Journal of Magnetic Resonance* **2003,** *164* (2), 310-320.

64. Vural, U. S.; Muradoglu, V.; Vural, S., *Bulletin of the Chemical Society of Ethiopia* **2011,** *25* (1).

65. Lide, D. R., *CRC Handbook of Chemistry and Physics, 84th Edition*. Taylor & Francis: 2003.

66. Polák, J.; Benson, G. C., *The Journal of Chemical Thermodynamics* **1971,** *3* (2), 235-242.

67. González, B.; Calvar, N.; Gómez, E.; Domínguez, Á., *The Journal of Chemical Thermodynamics* **2007,** *39* (12), 1578-1588.

68. Hurle, R. L.; Easteal, A. J.; Woolf, L. A., *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases* **1985,** *81* (3), 769-779.

69. Chmielewska, A.; Żurada, M.; Klimaszewski, K.; Bald, A., *Journal of Chemical & Engineering Data* **2009,** *54* (3), 801-806.

70. Thompson, J. W.; Kaiser, T. J.; Jorgenson, J. W., *Journal of Chromatography A* **2006,** *1134* (1), 201-209.

71. Khattab, I. S.; Bandarkar, F.; Fakhree, M. A. A.; Jouyban, A., *Korean Journal of Chemical Engineering* **2012,** *29* (6), 812-817.

72. Douheret, G.; Pal, A., *Journal of Chemical and Engineering Data* **1988,** *33* (1), 40-43.

73. Chidichimo, G.; Imbardelli, D.; Longeri, M.; Saupe, A., *Molecular Physics* **1988,** *65* (5), 1143-1151.

74. Pachler, K.; Wessels, P., *Journal of Molecular Structure* **1970,** *6* (6), 471-478.

75. Shimoaka, T.; Hasegawa, T., *Journal of Molecular Liquids* **2016,** *223*, 621-627.

76. Guvench, O.; MacKerell, A. D., *The Journal of Physical Chemistry A* **2006,** *110* (32), 9934-9939.

77. Schuler, L. D.; Daura, X.; Van Gunsteren, W. F., *Journal of Computational Chemistry* **2001,** *22* (11), 1205-1218.

Chapter 3 - A New View of Protein Compressibility

3.1 Introduction

A protein folds into a specific three-dimensional structure in aqueous solution under physiological conditions which is crucial for its function and activity.[1] The equilibrium state of a protein can be altered by changing the temperature, pressure, pH or by adding denaturants.[2-5] High pressure has been used as a powerful tool to study protein folding as it provides insight into the factors determine protein stability, most importantly, protein hydration.[6-7] The effect of pressure on the equilibrium is characterized by changes in the volume and compressibility associated with the conformational transition.[8-9]

The volume and compressibility of a protein are sensitive to solute-solvent interactions[8] and therefore, can be used to characterize the hydration of proteins. Water surrounding in a protein is very important as it determines the structure, function, and stability of a protein.[10-11] The partial molar volume of a solute is given by the first pressure derivative of the chemical potential. The isothermal compressibility of a protein is given by the first negative pressure derivative of the volume at constant temperature. In experiments, simple models are often used to define the volume and compressibility of proteins, and they mainly use small molecule data to evaluate these properties.[12] Amino acids, peptides, and amines are some of small molecule models that have been widely used to study the hydration properties of proteins.[13-15] However, the additive models based on small molecules have limitations for assessing protein volumes as they assume the volume to be insensitive to the environment of the small molecules, which is not true.[16-17]

---

*Simulations and analysis by Elizabeth A. Ploetz and Nilusha Kariyawasam. Figures by Elizabeth A. Ploetz.

In theoretical studies, the volume of a solute is defined as a summation of the intrinsic volume (V_M) and the hydration volume.[18] The hydration volume is the change in solvent volume due to solute-solvent interactions between charged, polar, or nonpolar groups on the surface of the protein.[2, 18] The hydration volume is often decomposed into three terms, the thermal volume (V_T), the interaction volume (V_I) and the volume contribution of the translational degrees of freedom of the solute ($\kappa_{T,1}RT$); where $\kappa_{T,1}$ is the isothermal compressibility of solvent, R is the gas constant and T is the absolute temperature. Therefore, the partial molar volume of infinitely dilute solute ($\bar{V}_2^{\infty}$) can be written as,

$$\bar{V}_2^{\infty} = V_M + V_T + V_I + \kappa_{T,1}RT \tag{3.1}$$

The thermal volume occurs due to the thermally induced vibrations of the solute and solvent molecules, while the interaction volume represents the reduction of the solvent volume due to the solvent interactions with polar and charged groups of the solute. However, these terms contain many unknown parameters, and these terms are not strictly obtained from experimental measurements, especially the V_T and V_I volumes.

The volume change upon pressure denaturation is usually negative, indicating that the denatured state has a small volume compared to the folded state.[4, 19] This difference is around 50-100 ml/mol or 0.5 - 2% of the total molar volume of a protein.[20] For a native protein, the volume fraction occupied by cavities ranges between 20-30%.[21-22] Royer and coworkers have suggested that the presence of cavities in the folded state, and the absence of cavities in the unfolded state, are main contributions to the volume change upon pressure denaturation.[19]

In theoretical studies, the compressibility of a globular protein is decomposed into two oppositely contributing factors involving a positive intrinsic compressibility and a negative hydration compressibility.[16] The intrinsic compressibility arises due to the imperfect packing of

the solvent inaccessible core of the protein while the hydration compressibility is due to the solute-solvent interactions over solvent accessible atomic groups of the protein. The hydration contribution to the compressibility is further decomposed into contributions from polar, nonpolar, and charged atomic groups on the surface of the protein. Chalikian and coworkers have shown that the contributions from polar, nonpolar, and charged groups on the surface to the compressibility is negative according to their model.[16]

The isothermal compressibility of a globular protein ranges from $5x10^{-6}$ - $15x10^{-6}$ bar^{-1}.[12, 23-24] The average intrinsic compressibility of a globular protein is determined to be $25x10^{-6}$ bar^{-1}.[16] However, the intrinsic compressibility is not uniform throughout the protein interior since there can be domains with different packing densities inside the protein structure.[25-26] . For, comparison the isothermal compressibility of pure water is $45x10^{-6}$ bar^{-1}, while benzene and hexane have the compressibility of $96x10^{-6}$ bar^{-1}, $165x10^{-6}$ bar^{-1}, respectively.[8] It is interesting that proteins have very low compressibility which are an order of magnitude smaller than the compressibility of pure water. This suggest that proteins have a tightly packed rigid like interior.[8] However, it is not clear why proteins have very low compressibility even in the presence of internal cavities. The difference in compressibility between the folded and molten globule states is found to be very small and positive, indicating that the unfolded state has a slightly higher compressibility.[8, 27] Still there is a lack of understanding of the compressibility of the folded and unfolded state upon pressure denaturation as it is difficult to explain the changes in compressibility in terms of cavities and hydration changes.[11]

Computer simulation studies have been performed to calculate the partial molar volume of proteins. As discussed in the first chapter, most studies involve subjective definitions of the protein volumes.[28-30] Furthermore, these subjective definitions are then used to calculate the

compressibility of proteins. Post and Dadarlat have calculated the isothermal compressibility of globular protein using protein volume fluctuations as shown below[28, 30],

$$\kappa_T = \frac{1}{k_B T} \frac{\langle \Delta V^2 \rangle}{\langle V \rangle} \tag{3.2}$$

where k_B is Boltzmann's constant, T is the temperature, $<V_2>$ is the average protein volume fluctuations and $<V>$ is the average volume. However, they use a subjective definition of the protein volume to calculate the protein compressibility. The definition they use is only possibly valid for globular proteins and cannot be used to calculate the volume of denatured proteins.[28] Moreover, they neglect the role of solvent by assuming that the volume fluctuation equation (used to calculate the compressibility of bulk system) can be applied to a component (protein) of a system.[31] The compressibility calculated from above expression can be possibly assign to intrinsic compressibility, but not the partial molar compressibility as the partial molar volume or compressibility includes the contributions from solute-solvent interactions.[32] Treating a protein as a bulk system has been both accepted,[33] and criticized,[32] in the literature. Furthermore, using the above expression one cannot obtain negative compressibilities which have been observed experimentally for small peptides and amino acids.[34] Except the subjective definitions of volume and compressibility, Hirata and coworkers have calculated the compressibility of a protein using the pressure derivative of the volume and they have calculated the volume of proteins using the reference interaction site model (RISM) theory coupled with Kirkwood–Buff (KB) solution theory.[32]

Rather than a subjective definitions of volumes KB theory provides a rigorous expression for the volume of a protein as follows[35],

$$\bar{V}_2^{\infty} = -4\pi \int_0^{\infty} (g_{21} - 1)\, \mathrm{r}^2\, dr \tag{3.3}$$

where g_{21} is the pair correlation function of the solvent (1) around the solute (2). However, this approach is not the most suitable way of calculating the volume of an irregular shaped objects such as proteins (see later discussion).

As the volume and compressibility changes upon denaturation are very small, it is crucial to have a method which can accurately capture these small changes. In this study, our focus is to explore the volume and compressibility of proteins under pressure and study the factors contribute to the low compressibility of proteins. Here, we use fluctuation solution theory (FST) to calculate the volume and compressibility of proteins without any subjective definitions or parameters. The results are compared for three different proteins. Furthermore, the residue contributions to these properties are discussed as the FST approach allows for the decomposition of the properties into residue-based contributions.

3.1.1 Theory

In this section, we discuss how to relate the experimental (macroscopic) thermodynamic quantities to computer simulations (microscopic) observations using statistical mechanical expressions. For this, the pseudo chemical potential approach of Ben-Naim is used.[35-37] The chemical potential (μ_2) of a solute (2) in a solvent (1) can be written as,

$$\begin{aligned} \mu_2 &= \mu_{2,m}^0 + RT \ln(\gamma_{2,m} m_2 / m^0) \\ \mu_2 &= \mu_{2,c}^0 + RT \ln(\gamma_{2,c} c_2 / c^0) \\ \mu_2 &= \mu_2^* + RT \ln(\Lambda_2^3 \rho_2) \end{aligned} \tag{3.4}$$

Here, the first two expressions relate the solute chemical potential to the experimental measures, involving the standard chemical potentials ($\mu^0_{2,m}$, $\mu^0_{2,c}$) and activity coefficients ($\gamma_{2,m}$, $\gamma_{2,c}$) in terms of the solute molal (m_2) or molar (c_2) concentration and the standard molality (m^0=1 mol/kg) or molarity (c^0=1 mol/L), respectively. The third expression of equation (3.4) provides the Ben Naim's statistical mechanical relationship for the pseudo chemical potential.[38] Here the pseudo chemical potential is related to the chemical potential using the thermal de Broglie wavelength (Λ_2) and number density of solute (ρ_2). As discussed in chapter 1, the pseudo chemical potential is defined as the change in Gibbs free energy when adding a single particle to a fixed position in space within the system.[38] The pseudo chemical potential approach eliminates the need for standard states and therefore, this is helpful to study single solute in solution (pseudo infinitely dilute) studied by computer simulations.[31, 39] After taking the pressure derivatives of equation (3.4) one find[40],

$$\bar{V}_2^{\infty} = V_{2,m}^0 = V_{2,c}^0 + RT\kappa_{T,1}^0 = V_2^{*,\infty} + RT\kappa_{T,1}^0 \tag{3.5}$$

where $\bar{V}_2^{\infty}$ is the partial molar volume of the solute, and $\kappa_{T,1}^0$ is the isothermal compressibility of the pure solvent. The $\bar{V}_2^{\infty}$is given by the pressure derivative of the chemical potential and the activity coefficient has disappeared as it is equal to unity for an infinitely dilute solute. This provides that the partial molar volume of solute at infinite dilution is equal to the pseudo volume ($V_2^{*,\infty}$) of solute within a constant. The difference between the standard and pseudo volumes is small under ambient conditions.[40] Pressure derivatives of equation (3.5) provide,[40]

$$\left(\frac{\partial \bar{V}_2^{\infty}}{\partial p}\right)_{m_2} = \left(\frac{\partial V_{2,m}^0}{\partial p}\right)_{m_2} = \left(\frac{\partial V_{2,c}^0}{\partial p}\right)_{m_2} + RT\left(\frac{\partial \kappa_{T,1}^0}{\partial p}\right)_{m_2} = \left(\frac{\partial V_2^{*,\infty}}{\partial p}\right)_{m_2} + RT\left(\frac{\partial \kappa_{T,1}^0}{\partial p}\right)_{m_2} \tag{3.6}$$

This shows the relationship between the pressure derivatives of the standard and pseudo volumes at infinite dilution. For a protein in equilibrium between native (N) and denatured state, the

equilibrium constant is given by, $K=N_D/N_N$ at infinite dilution. Using equation (3.4) and the equilibrium condition, $\mu_N = \mu_D$, we obtain,

$$-RT\ln K = \mu_D^{*,\infty} - \mu_N^{*,\infty} = \Delta G^{\circ} \tag{3.7}$$

The effect of pressure on the equilibrium can be expressed using a Taylor expansion as follows,

$$RT\ln\left(\frac{K}{K_0}\right) = RT\left(\frac{\partial \ln K}{\partial p}\right)_{m_2}\Delta p + \frac{RT}{2}\left(\frac{\partial^2 \ln K}{\partial p^2}\right)_{m_2}(\Delta p)^2 \tag{3.8}$$

where K_0 is the equilibrium constant at the reference pressure, and $\Delta p = p - p_0$ with p_0 being reference pressure, usually 1 bar. Then the effect of pressure on the equilibrium can be written in terms of properties of the native and denatured state.

$$-RT\left(\frac{\partial \ln K}{\partial p}\right)_{m_2} = \left[\left(\frac{\partial \mu_D^{*,\infty}}{\partial p}\right)_{m_2} - \left(\frac{\partial \mu_N^{*,\infty}}{\partial p}\right)_{m_2}\right] = \Delta V_2^{*,\infty} \tag{3.9}$$

$$-RT\left(\frac{\partial^2 \ln K}{\partial p^2}\right)_{m_2} = \left[\left(\frac{\partial^2 \mu_D^{*,\infty}}{\partial p^2}\right)_{m_2} - \left(\frac{\partial^2 \mu_N^{*,\infty}}{\partial p^2}\right)_{m_2}\right] = -\Delta K_2^{*,\infty}$$

Here, $\Delta K_2^{*,\infty}$ is the difference in compressibility factors.

The Fluctuation theory of solution is used to relate the local fluctuations in a solution to thermodynamic properties of bulk closed isothermal, isobaric (NpT) systems. Here, the thermodynamic properties are calculated in terms of particle number fluctuations in local regions of the solution. Now we use FST to calculate the pseudo volume and compressibility factors, the properties which characterize pressure denaturation. The pseudo molar volume of a solute at infinite dilution is given by,[35, 37, 40-41]

$$V_2^{*,\infty} = -[\langle N_1\rangle_2 - \langle N_1\rangle_0]V_1^0 \tag{3.10}$$

where $<N_1>_2$ is the average number of water molecules within a fixed local volume centered on the protein, and $<N_1>_0$ is the average number of water molecules within the same volume and shape of pure water. The size of the local volume should be large enough that the solvent distribution reaches the bulk distribution. Here, the N_1 denotes the number of water molecules while subscript 2 and 0 represents the solution and the pure solvent. The pseudo molar volume is provided when the difference in the number of water molecules is multiplied by the volume of pure water (V_1^0) at the corresponding temperature and pressure. This simple expression for volume is not subjective and does not involve any unknown parameters as it simply uses the average number of water molecules and the volume of pure water.

The pressure derivative of the pseudo volume is given by,[35, 37, 40-41]

$$K_2^{*,\infty} = \kappa_{T,1}^0 V_2^{*,\infty} + \beta(V_1^0)^2[<\delta N_1 \delta N_1>_2 - <\delta N_1 \delta N_1>_0] \tag{3.11}$$

where $<\delta N_1 \delta N_1>_x = <N_1^2>_x - <N_1>_x^2$, and $\kappa_{T,1}^0$ is the compressibility of pure water. The compressibility is based on the solvent fluctuations in the presence and the absence of the protein. The compressibility of a protein calculated using this expression can be either positive or negative, depending on the solvent fluctuations in the protein solution and bulk water. If the average fluctuations in the presence of protein ($<\delta N_1 \delta N_1>_2$) are smaller than in the pure water ($<\delta N_1 \delta N_1>_0$), then the compressibility of the protein will be negative. If they are bigger, the compressibility of the protein will be positive. This expression determines the compressibility of a protein purely based on the water distribution around the protein, i.e. the hydration waters. If there are strong solute-solvent interactions, the water fluctuations will be smaller. If the solute-solvent interactions are weak, the fluctuations will be larger in the presence of protein and the compressibility will therefore be higher. In the presence of charged residues on the surface of the protein, the fluctuations will be smaller due to strong solute-water interactions and this will reduce

the compressibility. Therefore, using our new approach we will possibly be able to explain the low compressibility of the native state of proteins.

3.1.2 Test systems

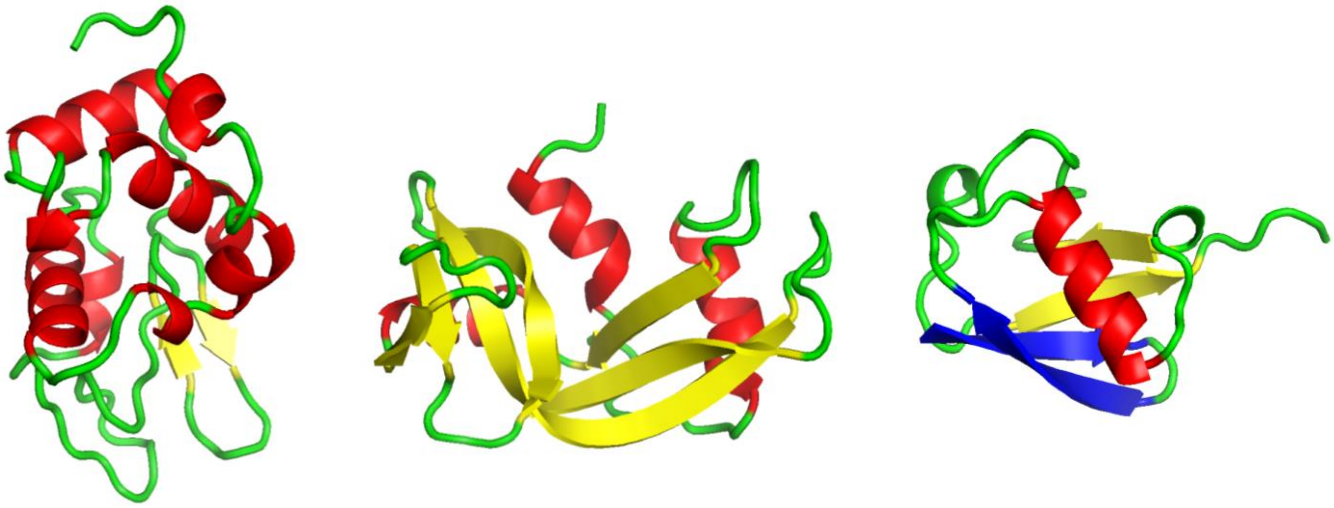


Figure 3.1 Cartoon representation of native Lysozyme (PDB ID: 4LZT), Ribonuclease (PDB ID: 2AAS), and Ubiquitin (PDB ID: 1UBQ) colored by its secondary structure elements. Helix-red, sheet-yellow, hairpin-blue, coil-green. Images were generated with PyMOL molecular visualization software.[42]

Hen egg white Lysozyme, Ribonuclease A (RNaseA), and Ubiquitin (Figure 3.1) were chosen as our test systems to calculate the volume and compressibility, and compared among proteins for consistency. These are relatively small proteins, where Lysozyme (PDB ID: 4LZT)[43] is a positively charged (+9) protein with 129 residues, RNaseA (PDB ID: 2AAS)[44], again a positively charged (+8) protein with 124 residues and ubiquitin (PDB ID: 1UBQ)[45], is a neutral protein with 76 residues. The number of residues in each physicochemical group and secondary structure group for the three proteins are shown in Table 3.1.

Table 3.1 Number of residues in each physicochemical (polar, non-polar, positively charged, negatively charged) and secondary structure (helix, sheet, hairpin, coil) groups and their percentages for Lysozyme, RNaseA, and Ubiquitin.

Category	Lysozyme		RNaseA		Ubiquitin	
	No: of residues	%	No: of residues	%	No: of residues	%
Total	129	100	124	100	76	100
Polar	45	35	56	45	20	26
Non-polar	56	43	39	31	32	42
Positive	18	14	18	15	12	16
Negative	10	8	11	9	12	16
Helix	52	40	26	21	12	16
Sheet	8	6	41	33	11	14
Hairpin	0	0	0	0	11	14
Coil	69	53	57	46	42	55

3.2 Methodology

3.2.1 Molecular dynamics simulations

Three types of simulations were performed in order to understand the factors contribute to the very low compressibilities of proteins. The first set of simulations were performed without any position restraints and will be known as 'free' simulations. The second set of simulations were performed with position restraints (1000 kJ mol^{-1} nm^{-2}) on all atoms (PR) to keep the structures fixed. This eliminates protein volume fluctuations. The third set of simulations were performed after neutralizing all the charged residues, except for the terminal residues, while keeping the restraints on all atoms (PR-Q). This set of simulations were performed to elucidate the effect of charged residues on the compressibility.

The crystal structures of the three proteins discussed above were used as the starting structures for the simulations. Proteins were solvated in a rhombic dodecahedrane box with TIP3P water. The distance between two parallel faces of the simulation boxes was 12 nm. Counterions (chlorides) were added to the Lysozyme and RNase systems to neutralize the charges. The CHARMM22* force field was used to simulate all the systems.[46]

All the simulations were performed with the GROMACS software package using versions 2016 or 2016.4.[47-48] Molecular dynamics simulations were performed in the isothermal-isobaric ensemble (NpT) at 300 K temperature and at 1 bar, 1 kbar, 2 kbar, and 3 kbar pressures. All protein bonds were constrained with the LINCS algorithm,[49] and water bonds with the Settle algorithm[50]. A time step of 2 fs was used to integrate the equation of motion with the Leap Frog algorithm.[51] The particle mesh Ewald (PME) technique[52] was used to calculate electrostatic interactions with a 0.12 nm Fourier grid spacing. The verlet cut-off scheme with a cut-off distance of 1.05 nm was used for van der Waals and electrostatic interactions. Periodic boundary conditions and the minimum image convention were applied to all the systems.

The systems were minimized for 1000 steps using the steepest descent algorithm followed by 100 ps equilibration with position restraints. For the equilibration, the Berendsen temperature coupling and Berendsen pressure coupling were used.[53] The Parrinello-Rahman pressure coupling,[54-55] and Nose-Hoover temperature coupling[56-57] were used for the production phase. For the PR and PR-Q simulations all atom position restraints (1000 kJ mol^{-1} nm^{-2}) were applied while for the free production simulations no position restraints were applied. The production runs were continued up to 100 ns. First 20 ns of the production runs were considered as an equilibration and 20-100ns runs were used for the analysis. Configurations were saved at every 10.0 ps for analysis of the systems. Pure TIP3P water was simulated for 20 ns using the same parameters at each

pressure for comparison. The 20-100ns of protein simulations were used for the analysis and the error bars were calculated for 20 ns block averages.

3.2.2 Analysis

For the protein simulations, the number of water molecules within a fixed local volume around the protein $(N_1)_2$ is counted at each time frame. Then, the snapshot of the protein at that time frame is superimposed on a pure water configuration and the number of water molecules within the same size and shape of the fixed local volume around the protein $(N_1)_0$ are counted. This process is repeated for all the frames in protein simulation, and at each frame, the number of water molecules is counted as a function of distance away from the surface of the protein. The volume of the protein is then calculated using equation (3.10) for each distance. The final protein volume is obtained by averaging the distance dependent volume over a region where the volume is not changing.

The residue volumes are obtained by assigning each water molecule to a heavy atom of a residue based on their proximity. The distance is calculated between the center of mass of each water molecule and the heavy atoms of each residue to determine proximity. For instance, if the distance between a heavy atom of the protein and water molecule A is shorter than the rest of the heavy atoms to water molecule A, then water molecule A is assigned to the residue containing the first heavy atom. The summation of residue volumes is equal to the total protein volume and this can be used to compare with the experimentally measured protein volume. Moreover, the residue volumes were decomposed into backbone and sidechain volume contributions by assigning residue atoms to backbone and sidechains.

The compressibility of a protein was calculated using the pressure derivative of the volume and after fitting the volumes to a quadratic fit according to,

$$\kappa^*_{T,2} = -\frac{1}{V_2^*}\left(\frac{\partial V_2^*}{\partial P}\right)_{T,m_2} \tag{3.12}$$

We are interested in this expression since it allows us to decompose the compressibility into residue-based contributions, in contrast to the fluctuation expression for compressibility (3.11). There is no clear way of decomposing the compressibility using equation (3.11) into residue-based contributions.[40] However, the compressibility results will be explained using the fluctuation expression. For the pseudo volume of the physicochemical or secondary structure group one uses (V_{Group}^*) in place of V_2^* to calculate the group contributions to the pseudo compressibility. The pseudo compressibility is related to the partial molar compressibility using the relationship,[40]

$$\bar{\kappa}_2^\infty = -\frac{1}{\bar{V}_2^\infty}\left(\frac{\partial \bar{V}_2^\infty}{\partial p}\right)_{T,m_2} = -\frac{RT}{\bar{V}_2^\infty}\left(\frac{\partial \kappa_1^0}{\partial p}\right)_{T,m_2} + \frac{\bar{V}_2^{*,\infty}}{\bar{V}_2^\infty}\kappa_2^{*,\infty} \tag{3.13}$$

for an infinitely dilute solute. The partial molar volume and pseudo volume, and partial molar compressibility and pseudo compressibility are differ by negligible amount for proteins in aqueous solutions under ambient conditions.[40]

3.3 Results and discussion

The root mean square deviations were calculated for all the proteins to see the stability of proteins over the simulation time and under pressure. Three proteins were remained stable and stayed closer to their native conformations.

Figure 3.2 depicts the relative probability of finding a water molecule around the proteins (g_{21}) compared to the bulk water distribution at 1 bar. All the distributions reach the bulk water distribution beyond 1 nm, after about three prominent solvation shells, and they all look similar. The pseudo volume as a function of local distance at 1 bar are shown in the right panel of Figure 3.2 and they also reach a plateau value beyond 1 nm. The volume of the protein is obtained by averaging over the plateau region. The three proteins have different volumes and the relative magnitude agrees with the number of residues in each protein.

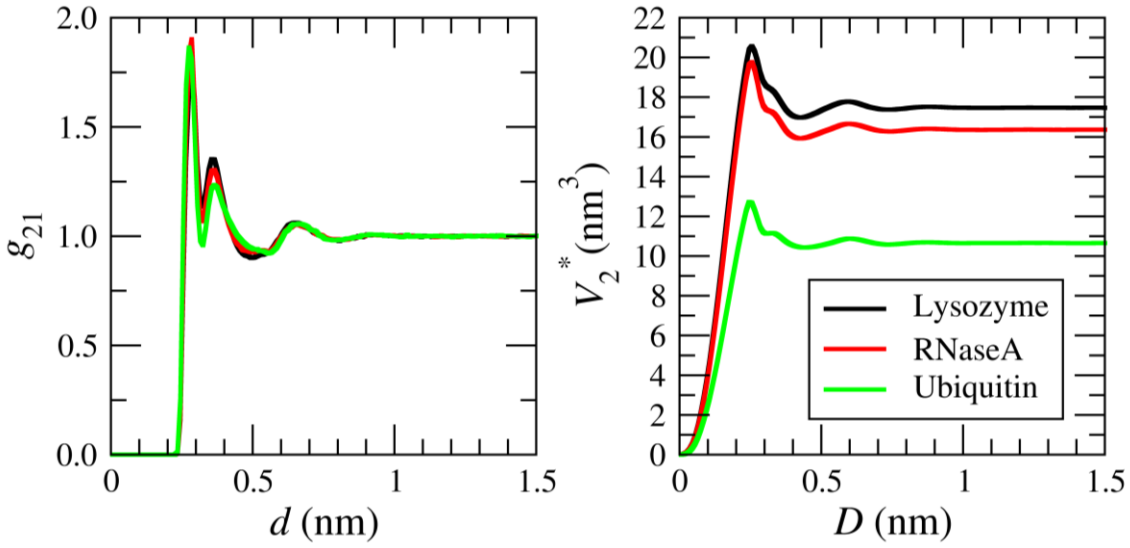


Figure 3.2 Left: Water probability distribution around protein (g_{21}) as a function of distance from the surface of the free proteins at 1 bar. Right: pseudo protein volume as a function of integration distance at 1 bar for the free Lysozyme, RNaseA, and ubiquitin.

Figure 3.3 shows the pseudo volume of the free, PR and PR-Q versions of the protein simulations as a function of pressure. The volume of the protein decreases with pressure for all the proteins and each version of the simulations. As expected, the Ubiquitin shows the lowest volume while the Lysozyme shows the highest volume. The position restrained versions, where the volume fluctuations are not allowed, show higher volumes compared to the free proteins except at 1 bar.

Among the two position restrained versions, PR-Q provides the highest volumes and all proteins follow the same trend. According to the volume expression (3.10), there should be less water around the frozen structure (PR) to have a higher volume than the free protein volume. This means that the flexibility of proteins plays a main role in its volume as it affects the surrounded water molecules. Since PR-Q version provides the highest volume it suggests that the charged residues give a negative contribution to the volume. Experimentally it has been shown that solvation of charged groups leads to a reduction in volume and this is known as 'electrostriction'.[18, 58]

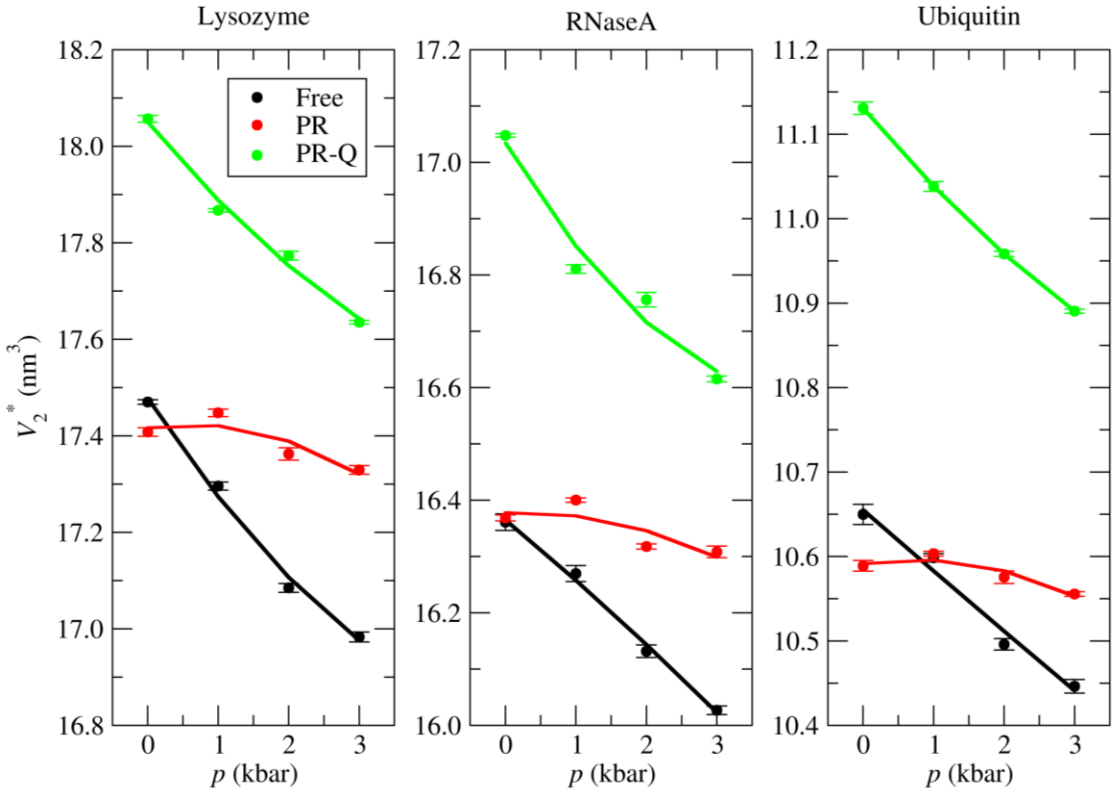


Figure 3.3 Pseudo volume of free, all atom position restrained (PR), and all atom position restrained with charged group neutralized (PR-Q) versions of the proteins as a function of pressure. Dots represent the raw volumes obtained by averaging the distant dependent volume, while lines represent the pressure fitted volumes.

Figure 3.4 displays the mean, maximum, and minimum residue volumes observed for Lysozyme, RNaseA, and Ubiquitin at 1 bar. This clearly shows that there can be more than a single

value for each amino acid residue volume depending on where it is in the sequence, or in space. This indicates that the residue volumes depend on their environment. As expected, we observe the highest residue volume for tryptophan and the lowest volume for glycine.

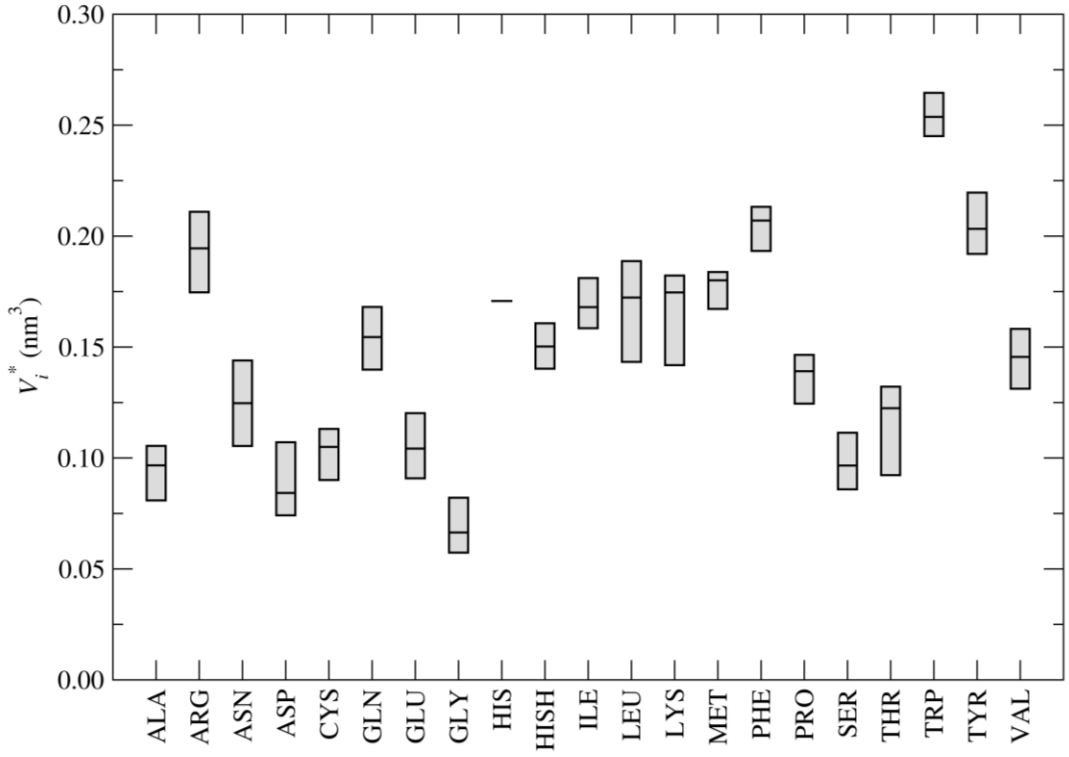


Figure 3.4 Median, maximum, and minimum residue volumes after considering all the residues in free Lysozyme, RNaseA, and Ubiquitin at 1 bar. Midde horizontal line: median residue volume, upper horizontal line: maximum residue volume, lower horizontal line: minimum residue volume. HIS only shows a single value since there is only one HIS present in all three proteins. 1 nm= 602 cm^3 mol^{-1}

Figure 3.5 displays the pseudo compressibility calculated for three simulations of each protein at 1 bar. Experimental compressibilities are also shown for the Lysozyme[24] and RNaseA[24] proteins, although we could not find the experimental compressibility for Ubiquitin. Free protein

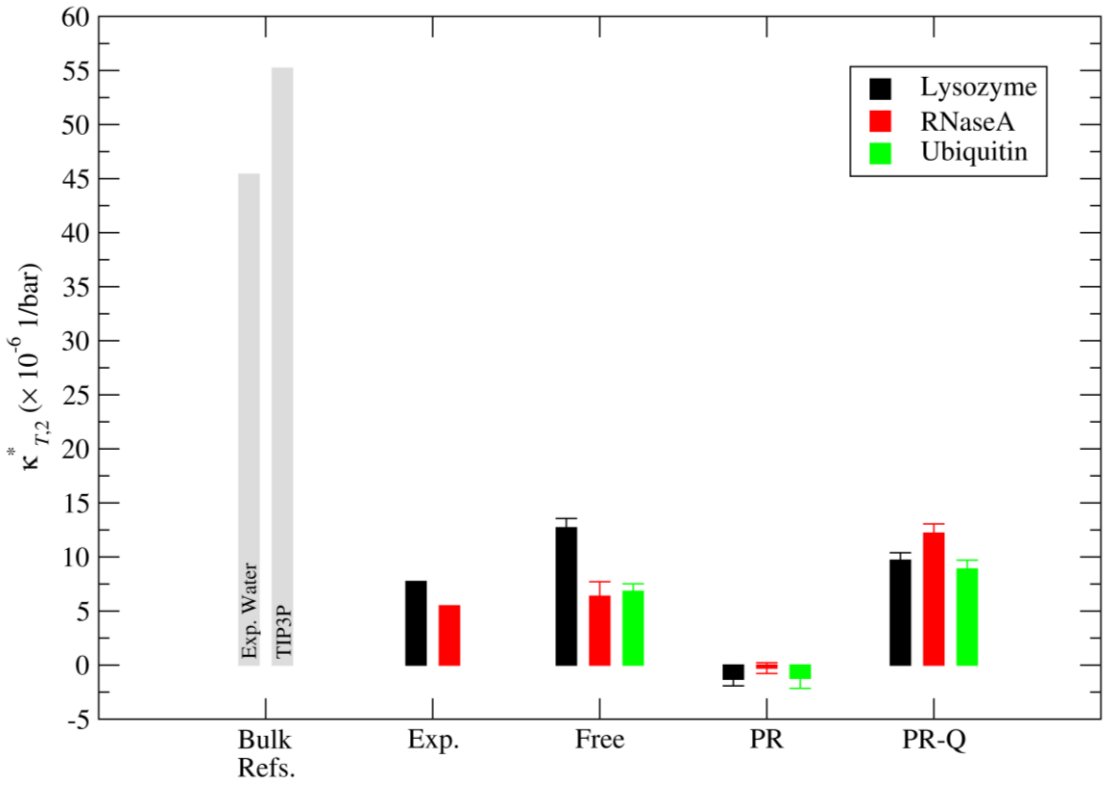


Figure 3.5 Pseudo compressibilities of free, all atom position restrained (PR), and all atom position restrained with charged group neutralized (PR-Q) versions of the proteins at 1 bar. Experimental partial molar compressibilities are also shown for pure water[8], Lysozyme[24], and RNaseA[24]. The pseudo compressibility of pure TIP3P water is also shown.

compressibilities calculated using the FST approach are in good agreement with the experimental data even though they were slightly overestimated. This clearly shows that all the three proteins have very low compressibilities compared to the pure water. Considering the PR and PR-Q versions we can observe that the PR versions display very small and negative compressibilities for all the proteins. This suggest that the volume fluctuations of proteins provide positive contribution to the compressibility of proteins. When the protein atoms are free to move then the water molecules will respond accordingly, and this will lead to higher water fluctuations around the protein. According to the FST expression for the compressibility (3.11), the compressibility will be higher with larger fluctuations in the presence of protein. In the PR-Q version, the amount of

compressibility reduced due to position constraints is regained by all proteins. This indicates that the charged amino acids contribute to the very low compressibilities of proteins.

The contributions from the physicochemical groups and the backbone (BB) to the pseudo compressibility of proteins are depicted in Figure 3.6. Each physicochemical group contains only the side chain (SC) atom contributions to the compressibility after decomposing each residue into BB and SC. Interestingly, both the positively charged and negatively charged groups show negative compressibilities for all the proteins except that Lysozyme shows a small positive compressibility for negatively charged groups. The negatively charged group of Ubiquitin shows the largest negative contribution to the compressibility, while all three proteins show consistent results for positively charged groups. Considering the number of positively and negatively charged residues as a percentage of the total number of residues in each protein (Table 3.1), we can see that all proteins have a similar percent composition for positively charged group (14%, 15%, 16% for Lysozyme, RNaseA, and Ubiquitin, respectively). Percent compositions for negatively charged group (8%, 9%, 16% for Lysozyme, RNaseA, and Ubiquitin, respectively) is also similar for Lysozyme and RNaseA, while it is about twice for the ubiquitin. This might be a reason that ubiquitin shows a large negative compressibility for negatively charged group. However, we do not observe similar compressibilities for Lysozyme and RNaseA even though they have similar percent composition for the negatively charged groups. Polar, nonpolar, and BB groups provide positive compressibilities and show similar contributions for all three proteins.

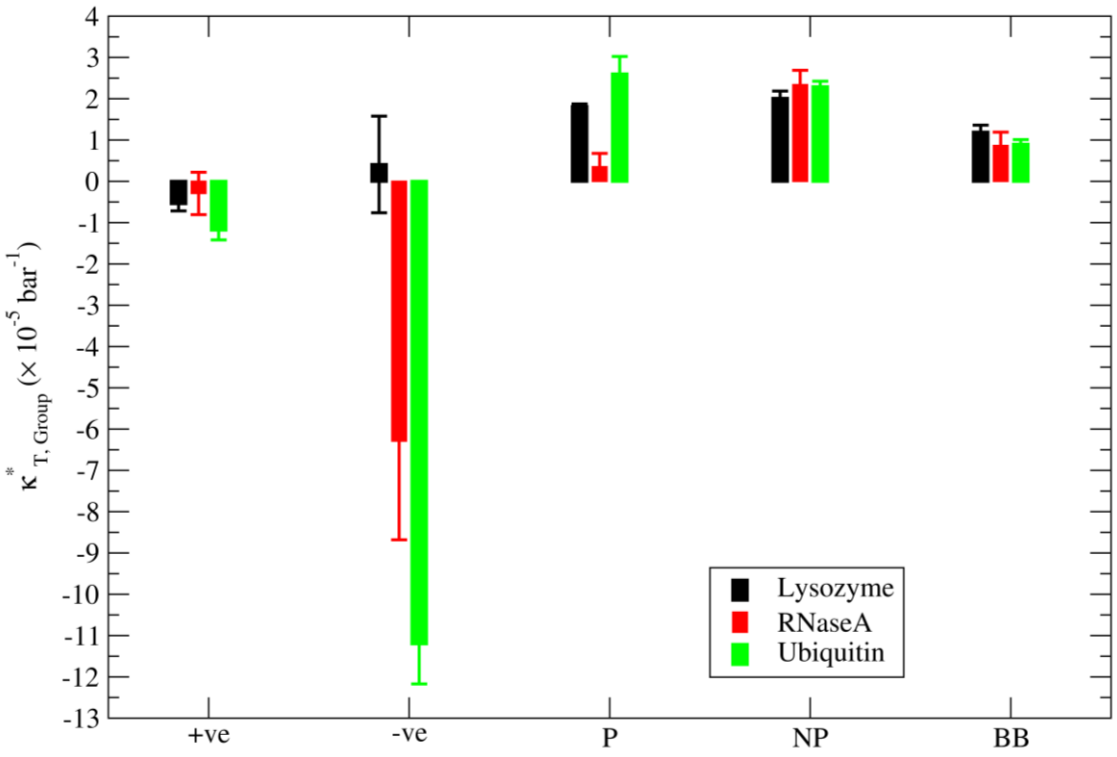


Figure 3.6 Pseudo compressibility contributions from positively charged (+ve), negatively charged (-ve), polar (P), nonpolar (NP) side chains and total backbone (BB) to the free proteins at 1 bar.

The negative compressibilities observed for charged groups can be explained using the fluctuation expression for the compressibility (equation (3.11)). According to this expression, negative compressibilities are given when the local fluctuations of waters in the vicinity of protein are smaller than the bulk water fluctuations and there will be smaller fluctuations if the protein-water interactions are stronger. This suggests that the strong interactions between charged side chains and water lead to low compressibilities for these groups.

To further explain the different behavior of negatively charged residues, the water distribution around a selected side chain atom for four different amino acid residues representing positively charged, negatively charged, polar, and nonpolar groups were determined at 1 bar, 3 kbar, 6 kbar, and 10 kbar (Figure 3.7). Here, the larger pressures were chosen since the differences

are much clearer at higher pressures. However, we observe a similar story at lower pressures but with a greater level of noise. For the positively charged (Lys), polar (Ser), and nonpolar (Leu) residues the water distribution around the selected atoms becomes more structured with pressure. In contrast, the structure gets weaker, or is destroyed, at higher pressures for negatively charged residues (Glu) and there are less water molecules at higher pressures This infer that, the volume of negatively charged group increases with pressure. This means that the solvation around negatively charged residues are particularly different than for the other groups.

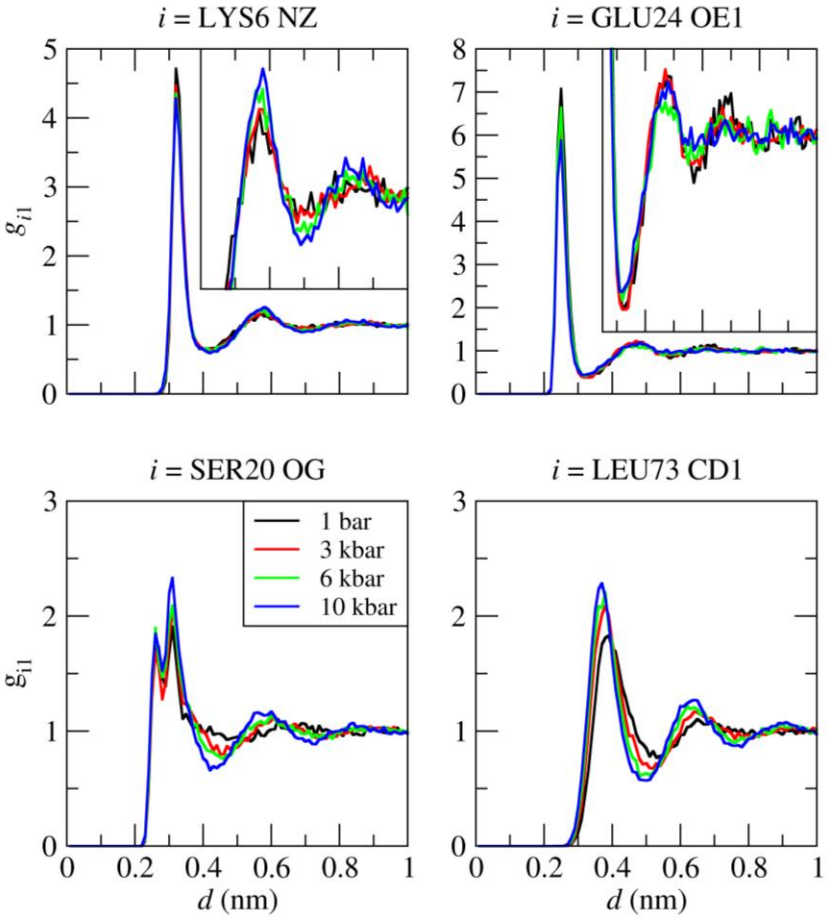


Figure 3.7 Water distribution around selected side chain atom (g_{i1}) of residues chosen to represent positively charged, negatively charged, polar, and nonpolar residues at 1 bar, 3 kbar, 6 kbar, and 10 kbar for ubiquitin. Residue name, the sequence number, and the atom type are denoted in labels. Enlarged second and third solvation shells are shown as insets for the top two figures.

To further illustrate the group contributions to the total compressibility of each protein, the percent contributions from each group are also included in Figure 3.8. The area under each bar gives the percent contribution from each group to the total compressibility where $\kappa_{T,2}^{*} = \sum_i \phi_i \kappa_{T,i}^{*}$. For all the proteins the highest positive contribution is given by the nonpolar groups. The highest negative contribution is given by the negatively charged groups for RNaseA and Ubiquitin, while it is the positively charged group for Lysozyme. The percent group contributions to Ubiquitin and RNaseA suggest that the overall small positive compressibilities are obtained by counteracting major positive contributions from nonpolar groups and negative contributions from charged groups. This is not consistent with the case for Lysozyme as it displays very small negative contributions from charged groups. Even though the backbones provide a larger volume fraction to the proteins, the contribution to the compressibility is smaller compared to the side chains.

In this study we have used relatively small proteins and therefore, they have very few buried residues within the core. (Lysozyme- 17 buried residues, RNaseA-18 buried residues, Ubiquitin-12 buried residues)) This might be a reason why we observe large negative contribution to the compressibility from charged residues.

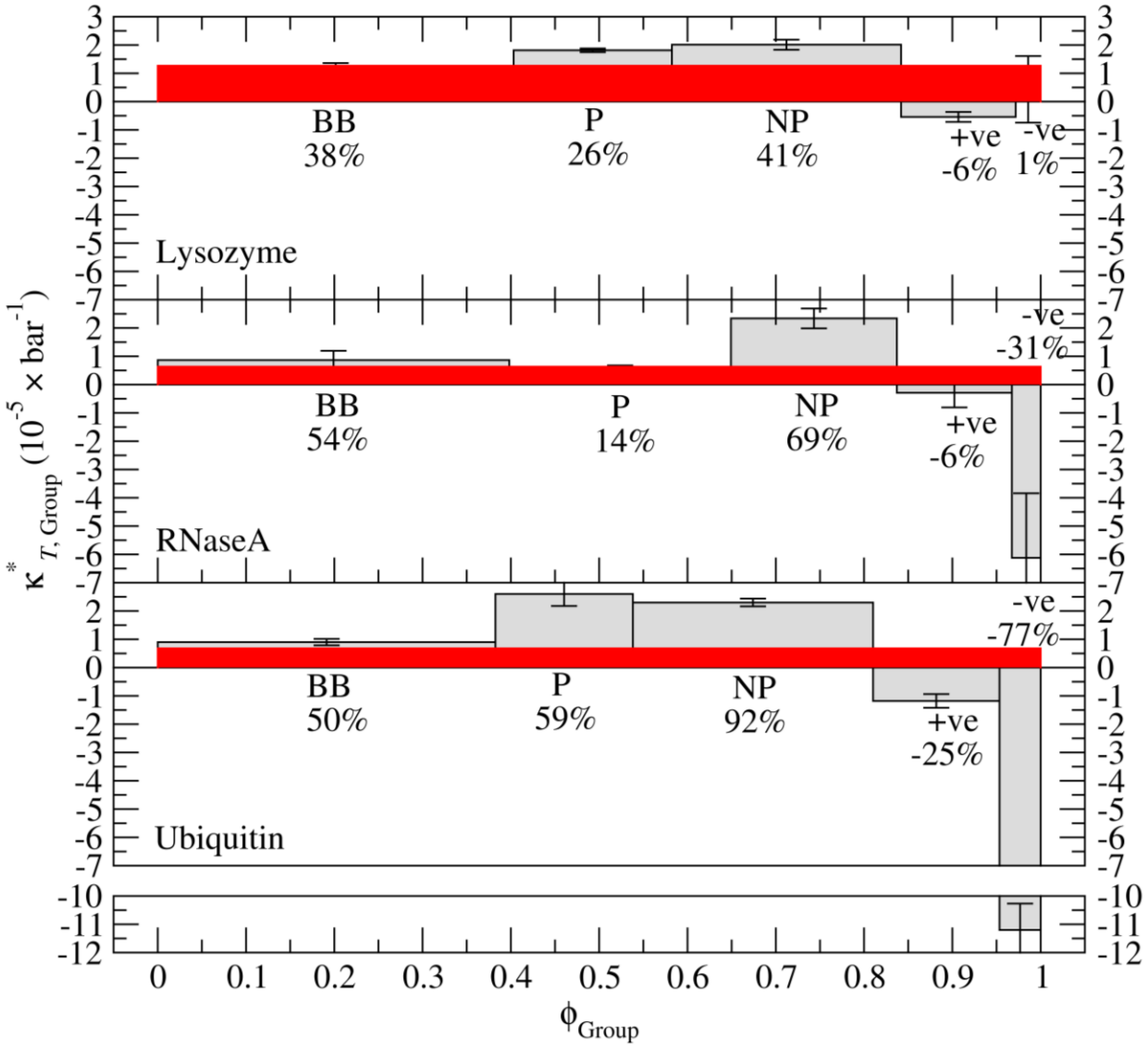


Figure 3.8 Pseudo compressibility contributions from positively charged (+ve), negatively charged (-ve), polar (P), nonpolar (NP) side chains and total backbone (BB) to the free proteins as a function of volume fraction (ϕ) at 1 bar. The percent contributions from each group to the total compressibility are indicated. Total compressibility of each protein is given by the y-value of red shaded area and the contributions from each group is given by y-values of grey shaded bars.

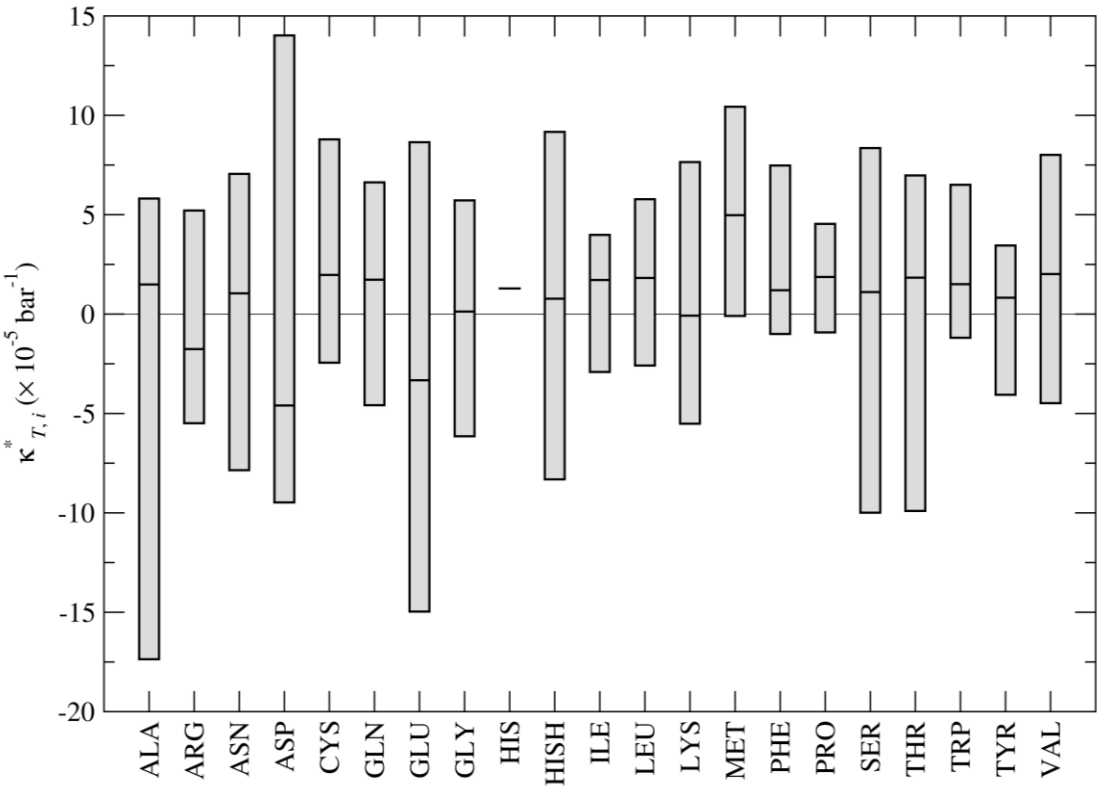


Figure 3.9 The median, maximum, and minimum pseudo compressibilities of residues after considering all the residues in free Lysozyme, RNaseA, and Ubiquitin at 1 bar. Middle horizontal line: median residue compressibility, upper horizontal line: maximum residue compressibility, lower horizontal line: minimum residue compressibility. HIS only shows a single value since there is only one HIS present in all three proteins.

Figure 3.9 displays the pseudo compressibilities for all the residues in the three proteins. Compared to the residue volume distribution, the range that the compressibilities are distributed over is much wider. This shows that any residues can have either positive or negative compressibilities. However, considering the median value, only the Arg, Asp, and Glu, which are charged amino acids show negative compressibilities.

3.4 Conclusions

The pseudo volume and the pseudo compressibility were calculated using FST for native Lysozyme, RNaseA, and Ubiquitin proteins. The FST approach discussed here allows us to calculate the residue-based contributions to the volume and compressibility of any shape, or size of solute without using subjective definitions or parameters. Local water distribution around protein is the major focus when using this approach.

Protein volume fluctuations provide a positive contribution to the compressibility of proteins. Furthermore, the nonpolar groups provide a larger positive contribution to the compressibility, while charged groups provide a large negative contribution to the compressibility. Among the charged groups, negatively charged residues show the largest negative compressibilities. However, the percent contributions from physicochemical groups to the compressibility seem somewhat dependent on the protein. Proteins studied here are relatively small, neutral and positively charged proteins. The size, and the charge of the protein may also affect the compressibility. Therefore, more proteins may need to be studied to better understand the overall low compressibilities shown by proteins.

3.5 References

1. Pace, C. N.; Shirley, B. A.; McNutt, M.; Gajiwala, K., *The FASEB journal* **1996,** *10* (1), 75-83.

2. Kauzmann, W., Some Factors in the Interpretation of Protein Denaturation1. In *Advances in protein chemistry*, Elsevier: 1959; Vol. 14, pp 1-63.

3. Dill, K. A., *Biochemistry* **1990,** *29* (31), 7133-7155.

4. Royer, C. A., *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **2002,** *1595* (1-2), 201-209.

5. Dobson, C. M., *Nature* **2003,** *426* (6968), 884.

6. Royer, C., *Brazilian journal of medical and biological research* **2005,** *38* (8), 1167-1173.

7. Royer, C. A., [16] Application of pressure to biochemical equilibria: The other thermodynamic variable. In *Methods in enzymology*, Elsevier: 1995; Vol. 259, pp 357-377.

8. Taulier, N.; Chalikian, T. V., *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **2002,** *1595* (1-2), 48-70.

9. Chen, C. R.; Makhatadze, G. I., *Nature communications* **2017,** *8*, 14561.

10. Sirotkin, V. A.; Komissarov, I. A.; Khadiullina, A. V., *The Journal of Physical Chemistry B* **2012,** *116* (13), 4098-4105.

11. Chalikian, T. V., *Annual review of biophysics and biomolecular structure* **2003,** *32* (1), 207-235.

12. Kharakoz, D.; Sarvazyan, A., *Biopolymers* **1993,** *33* (1), 11-26.

13. Häckel, M.; Hinz, H.-J.; Hedwig, G. R., *Biophysical chemistry* **1999,** *82* (1), 35-50.

14. Høiland, H., Partial molar volumes of biochemical model compounds in aqueous solution. In *Thermodynamic data for biochemistry and biotechnology*, Springer: 1986; pp 17-44.

15. Kharakoz, D., *Biophysical chemistry* **1989,** *34* (2), 115-125.

16. Chalikian, T. V.; Totrov, M.; Abagyan, R.; Breslauer, K. J., *Journal of molecular biology* **1996,** *260* (4), 588-603.

17. Chalikian, T. V., CHAPTER 21 Partial Molar Volumes of Proteins in Solution. In *Volume Properties: Liquids, Solutions and Vapours*, The Royal Society of Chemistry: 2015; pp 542-574.

18. Chalikian, T. V.; Breslauer, K. J., *Biopolymers* **1996,** *39* (5), 619-626.

19. Roche, J.; Caro, J. A.; Norberto, D. R.; Barthe, P.; Roumestand, C.; Schlessman, J. L.; Garcia, A. E.; García-Moreno E, B.; Royer, C. A., *Proceedings of the National Academy of Sciences of the United States of America* **2012,** *109* (18), 6945-6950.

20. Royer, C. A., Why and how does pressure unfold proteins? In *High Pressure Bioscience*, Springer: 2015; pp 59-71.

21. Richards, F. M., *Annual review of biophysics and bioengineering* **1977,** *6* (1), 151-176.

22. Liang, J.; Dill, K. A., *Biophysical journal* **2001,** *81* (2), 751-766.

23. Chalikian, T. V.; Gindikin, V. S.; Breslauer, K. J., *Journal of molecular biology* **1995,** *250* (2), 291-306.

24. Gekko, K.; Hasegawa, Y., *Biochemistry* **1986,** *25* (21), 6563-6571.

25. Akasaka, K.; Tezuka, T.; Yamada, H., Pressure-induced changes in the folded structure of lysozyme. Elsevier: 1997.

26. Kundrot, C. E.; Richards, F. M., *Journal of molecular biology* **1987,** *193* (1), 157-170.

27. Seemann, H.; Winter, R.; Royer, C. A., *Journal of molecular biology* **2001,** *307* (4), 1091-1102.

28. Dadarlat, V. M.; Post, C. B., *The Journal of Physical Chemistry B* **2001,** *105* (3), 715-724.

29. Murphy, L. R.; Matubayasi, N.; Payne, V. A.; Levy, R. M., *Folding and Design* **1998,** *3* (2), 105-118.

30. Dadarlat, V. M.; Post, C. B., *Biophysical journal* **2006,** *91* (12), 4544-4554.

31. Ploetz, E. A.; Smith, P. E., *The Journal of Physical Chemistry B* **2014,** *118* (45), 12844-12854.

32. Imai, T.; Hirata, F., *The Journal of chemical physics* **2003,** *119* (11), 5623-5631.

33. Rösgen, J.; Hinz, H.-J., *Biophysical chemistry* **2000,** *83* (1), 61-71.

34. Høiland, H.; Hedwig, G., 6 Compressibilities of amino acids, peptides and proteins in aqueous solution. In *Structural and Physical Data I*, Springer: 2003; pp 6001-6017.

35. Jiao, Y.; Smith, P. E., *The Journal of chemical physics* **2011,** *135* (1), 07B602.

36. Patel, N.; Dubins, D. N.; Pomes, R.; Chalikian, T. V., *The Journal of Physical Chemistry B* **2011,** *115* (16), 4856-4862.

37. Ploetz, E. A.; Smith, P. E., *Advances in chemical physics* **2013,** *153*, 311.

38. Ben-Naim, A., *Molecular Theory of Solutions*. Oxford University Press: New York, 2006.

39. Ben-Naim, A., *Molecular theory of solutions*. Oxford University Press: 2006.

40. Ploetz, E. A.; Smith, P. E., *Biophysical chemistry* **2017,** *231*, 135-145.

41. Ploetz, E. A.; Smith, P. E., *The Journal of Physical Chemistry B* **2015,** *119* (25), 7761-7777.

42. DeLano, W. L., *CCP4 Newsletter On Protein Crystallography* **2002,** *40* (1), 82-92.

43. Walsh, M. A.; Schneider, T. R.; Sieker, L. C.; Dauter, Z.; Lamzin, V. S.; Wilson, K. S., *Acta Crystallographica Section D* **1998,** *54* (4), 522-546.

44. Santoro, J.; González, C.; Bruix, M.; Neira, J. L.; Nieto, J. L.; Herranz, J.; Rico, M., *Journal of molecular biology* **1993,** *229* (3), 722-734.

45. Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J., *Journal of molecular biology* **1987,** *194* (3), 531-544.

46. Piana, S.; Lindorff-Larsen, K.; Shaw, D. E., *Biophysical journal* **2011,** *100* (9), L47-L49.

47. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., *Bioinformatics* **2013,** *29* (7), 845-854.

48. Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E., *Journal of Chemical Theory and Computation* **2010,** *6* (2), 459-466.

49. Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G., *Journal of computational chemistry* **1997,** *18* (12), 1463-1472.

50. Miyamoto, S.; Kollman, P. A., *Journal of computational chemistry* **1992,** *13* (8), 952-962.

51. Hockney, R.; Goel, S.; Eastwood, J., *Journal of Computational Physics* **1974,** *14* (2), 148-158.

52. Darden, T.; York, D.; Pedersen, L., *The Journal of chemical physics* **1993,** *98* (12), 10089-10092.

53. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., *The Journal of Chemical Physics* **1984,** *81* (8), 3684-3690.

54. Nosé, S.; Klein, M., *Molecular Physics* **1983,** *50* (5), 1055-1076.

55. Parrinello, M.; Rahman, A., *Journal of Applied physics* **1981,** *52* (12), 7182-7190.

56. Hoover, W. G., *Physical review A* **1985,** *31* (3), 1695.

57. Nosé, S., *Molecular physics* **1984,** *52* (2), 255-268.

58. Imai, T.; Harano, Y.; Kovalenko, A.; Hirata, F., *Biopolymers* **2001,** *59* (7), 512-519.

Chapter 4 - Simulated Amino Acid Volumes and Compressibilities - A Force Field Comparison

4.1 Introduction

It is important to understand the stability of proteins and, therefore, the thermodynamics of protein folding. An equilibrium between folded and unfolded states of a protein can be perturbed by changing the temperature, pressure, PH, or the addition of cosolvents. Temperature denaturation is characterized by the enthalpy, entropy, or heat capacity difference, while the pressure denaturation is characterized by the volume and compressibility difference. The properties related to pressure denaturation provide macroscopic details concerning the transition, contrary to the properties related to temperature denaturation.[1-2] Therefore, it is important to understand the pressure denaturation thermodynamics of proteins as it provides an insight into protein stability.

The partial molar volume and compressibility of a protein provide insights into the solute-solvent interactions (protein hydration) and the packing of amino acids.[3] The volume change upon protein denaturation is due to exposure of polar and non-polar groups, the electrostriction effect, and the elimination of internal cavities.[4] Experimental studies have shown that the volume change upon denaturation is very small and negative.[4] A computer simulation study by McCarthy and coworkers has suggested that high pressure affects the arrangement of water molecules, and the subsequent weakening of the hydrophobic effect is the main driving force for protein denaturation under high pressure.[5] Contrary to that, Royer and coworkers have concluded that the elimination of internal cavities was the major contribution to a negative volume change upon pressure denaturation.[4, 6] Therefore, the volume change upon denaturation is still subject to debate.

The isothermal compressibility of a protein is given by the negative pressure derivative of the protein volume at constant temperature, and this property is also used to explain the effect of pressure on protein stability. The isothermal compressibility of a globular protein is very small and ranges from $5x10^{-6}$ - $15x10^{-6}$ bar^{-1}.[7-9] For comparison, the compressibility of pure water is 45 $x10^{-6}$ bar^{-1}.[10] Having a compressibility value smaller than a liquid suggests that the protein has a well packed, solid like, interior.[10] However, it is not clear why native proteins have such a low compressibility, especially when internal cavities are present.

Considerable theoretical work has been performed to help understand the partial molar volume and compressibility of proteins. However, these studies have mainly been focused on experimental data for simple model peptides.[11] Thermodynamic properties of small molecules such as amino acids and short peptides are often studied and then used to interpret protein thermodynamic properties as a sum of group based contributions (additive approach).[12] It becomes more difficult to directly assess the thermodynamic properties for complex molecules such as proteins using simple models as these models do not include the environmental effects which are present in proteins. Furthermore, it is not easy to obtain the residue-based contributions to these properties using experimental studies.

There have been several efforts taken to calculate the partial molar volume of proteins using computer simulations. However, many of these calculations involve subjective definitions of the volume. For instance, Post and coworkers have calculated the volume using molecular dynamics simulations in the Gibbs ensemble (*NpT*).[13] This approach was based on an atomic van der Waals radius extension algorithm. Here, the total volume of a protein is divided in to a van der Waals volume and unoccupied interstitial volume. The interstitial volume is the unoccupied volume within the molecular boundary and includes internal cavities and packing defects. This

volume is calculated by extending the atomic van der Waals radii by a constant value. Nevertheless, this approach is only valid for approximately spherical shaped proteins and, therefore, not accurate for denatured proteins.

Kirkwood-Buff (KB) theory[14] has been successfully used to calculate the partial molar volume and other thermodynamic properties of solution mixtures.[15-21] This involves the integration of a pair correlation function or molecular distribution function. Since KB theory is an exact theory without any approximations, it provides a rigorous way of calculating volume without any subjective definitions or parameters.

Imai and coworkers have performed a theoretical study to calculate the PMVs of twenty amino acids in infinitely dilute solution based on Kirkwood-Buff theory and the reference interaction site model (RISM) equation of molecular liquids.[11] Their results show that ionization of the C and N termini give negative contributions to the volume. This volume reduction is then explained by electrostriction, which occurs due to the solute-solvent interactions around the charged atomic groups.[11] They also concluded that the contribution from a functional group (example - CH_2) to the volume depends on the location of that group in the protein and therefore, the partial molar volume determination using group contributions is not reliable. Later, they have developed this method by using the three-dimensional RISM equation, instead of traditional 1D-RISM, to help improve the results.[22-23]

The accuracy of a computer simulation is determined by the accuracy of the force field and the degree of sampling achieved.[24-25] Moreover, the quality of the methods used to interpret the results is another vital factor. In this study, we focus on calculating the volume and compressibility of Ubiquitin protein using our new approach based on fluctuation solution theory (FST).[15, 20, 26-27] The FST approach avoids the integration of molecular distribution function over irregularly shaped

objects such as proteins. Most importantly, this approach allows us to calculate the residue-based contributions to the thermodynamic properties of proteins.[20, 27] Here, molecular dynamics simulations were performed using several biomolecular force fields (FFs) and the result will be compared among different FFs to examine the consistency of results.

Specifically we perform a FF comparison between the AMBER99SB-ILDN[28], CHARMM22*[29], GROMOS 53A6[30], and the OPLS-AA[31-33] classical force fields for proteins. All these FFs are parameterized for biomolecular simulations and they differ from each other mainly due to the partial atomic charges, the parameters used in van der Waals interactions, and the dihedral potentials. As these FFs use different parameterization approaches such that they try to reproduce various properties obtained quantum mechanically, or experimentally, the final properties calculate with these FFs can be dissimilar. Thus, the solute-solute, solute-solvent, and solvent-solvent interactions will be different among the FFs for a identical system of interest.

Hess and co-workers have studied the hydration free energies, entropies, enthalpies, and heat capacities of amino acids with AMBER99, GROMOS 53A6, and OPLS-AA FFs combining with SPC, SPC/E, TIP3P, TIP4P, and TIP4P-Ew water models.[34] Their results suggest that the choice of water model is strongly affects the accuracy of the results while the differences in accuracy between FFs are small. Nillson and coworkers have performed a study on different water models (SPC, SPC/E, TIP3P, modified SPC and, modified TIP3P) and they have observed that different water models show different bulk water properties when simulated under the same conditions.[35]

Different FFs are developed in conjunction with different water models. The TIP3P water model is used for the AMBER99SB-ILDN and CHARMM22* FFs, while the SPC and TIP4P water models are used with the GROMOS-53A6 and the OPLS-AA force fields, respectively. The

AMBER99SB-ILDN FF is an improved version of the AMBER99SB[36] FF where the χ_1 torsion potentials for amino acid side chains are improved.[28] Thus, the improved AMBER99SB-ILDN is recommended over AMBER99SB for simulations of proteins. The CHARMM22* force field was developed with improving the backbone potentials of CHARMM22[37] FF. The GROMOS 53A6 is a united atom force field, while the other three FFs are all-atom FFs. In the GROMOS 53A6 FF, united atoms are used for aliphatic carbons. The GROMOS 53A6 is recommended for biomolecular simulations in explicit water as the parameters have improved hydration and solvation properties.[30] The OPLA-AA FF has refitted torsional coefficients to reproduce high level ab-initio data and refitted nonbonded interactions for sulfur containing dipeptides to reproduce gas phase dimerization energies, heat of vaporizations, and densities.[32]

4.1.1 Theory

For a protein in an equilibrium between the native (N) and the denatured state (D), the equilibrium constant (*K*) is given by,

$$K = \frac{[D]}{[N]} \tag{4.1}$$

The effect of pressure on the equilibrium is directly related to the partial molar volume difference between the native and the denatured states. If we consider a protein (2) in water (1) at infinite dilution, the first pressure derivative of the equilibrium constant at constant temperature (T) and solute molality (m_2) can be written as[27],

$$\left(\frac{\partial \ln K}{\partial P}\right)_{T,m_2} = -\beta(V_D^{*,\infty} - V_N^{*,\infty}) = -\beta\Delta V^* \tag{4.2}$$

where P is the pressure, the pseudo volume difference between the native and the denatured states is ($V_D^{*,\infty} - V_N^{*,\infty}$), and $\beta=1/RT$, R is the gas constant and T is the absolute temperature. The second derivative of the equilibrium constant under the influence of pressure is given by the difference in compressibility factors (ΔK^*) as shown below,[27]

$$\left(\frac{\partial^2 \ln K}{\partial P^2}\right)_{T,m_2} = -\beta\left(\frac{\partial \Delta V^*}{\partial p}\right)_{T,m_2} = -\beta \Delta K^* \tag{4.3}$$

As mentioned in the introduction, KB theory can be used to calculate the partial molar volume of proteins by integrating the pair correlation function as follows[15],

$$V_2^{*,\infty} = -4\pi \int_0^\infty (g_{21} - 1)\, \mathrm{r}^2\, dr \tag{4.4}$$

where $V_2^{*,\infty}$is the pseudo molar volume of an infinitely dilute solute, g_{21} is the pair correlation function of the solvent (1) around the solute (2). However, this approach is not the most useful/informative way of calculating the volume of an irregular shaped objects such as proteins.

Fluctuation Solution Theory (FST) can be used to calculate the partial molar quantities of solutes. This approach allows us to calculate the volume of a solute without integrating the molecular distribution function, which is ideal for an irregular shaped object like a protein. Moreover, this method allows us to decompose the properties into residue-based contributions. The pseudo molar volume of for an infinitely dilute solute ($V_2^{*,\infty}$) is given by[15, 20, 26-27],

$$V_2^{*,\infty} = -[\langle N_1 \rangle_2 - \langle N_1 \rangle_0] V_1^0 \tag{4.5}$$

where, $<N_1>_2$ is the average number of water molecules within a fixed local volume centered on the protein, and $<N_1>_0$ is the average number of water molecules within the same volume and shape of pure water. Here, the subscripts 2 and 0 represent the protein solute and the pure solvent. The size of the local volume should be large enough that the solvent distribution approaches bulk.

The volume can be calculated as a function of distance from the surface or center of mass of solute. The pseudo molar volume is calculated by multiplying the difference in the number of water molecules by the volume of a pure water molecule (V_1^0). This is a simple expression, without any approximations, that can be used to calculate the volume of any size or shaped protein.

The pressure derivative of the pseudo molar volume is given by,[15, 20, 27]

$$K_2^{*,\infty} = \kappa_{T,1}^0 V_2^{*,\infty} + \beta (V_1^0)^2 [< \delta N_1 \delta N_1 >_2 - < \delta N_1 \delta N_1 >_0] \tag{4.6}$$

where $< \delta N_1 \delta N_1 >_x = < N_1^2 >_x - < N_1 >_x^2$ and $\kappa_{T,1}^0$ is the partial molar compressibility of pure water. This expression determines the protein compressibility based on the fluctuations of solvent in the local volume centered around the protein ($< \delta N_1 \delta N_1 >_2$) and the fluctuations in the same volume of bulk solvent ($< \delta N_1 \delta N_1 >_0$). The compressibility will be smaller than the bulk solvent if the solvent fluctuations are smaller in the presence of protein or vice versa. In the presence of strong solute-solvent interactions there will be smaller solvent fluctuations. The compressibility determined using above expression can be either positive or negative. If the average number fluctuations in the presence of protein are smaller than the average solvent molecule number fluctuations in bulk pure water, then the compressibility will be negative. This is an important aspect of this expression as negative compressibilities have been observed experimentally.[38] It should be noted that the above expression can be applied to any solute regardless of its size.

The compressibility of a protein can also be studied by following the protein volume changes with pressure. However, an exact definition of protein volume should be used.

4.1.2 Ubiquitin

The Ubiquitin protein (PDB ID: 1UBQ, Figure 4.1)[39] was selected to perform our calculations due to its small size (76 residues) and the availability of some simulated denatured

structures which were used in our previous study.[27] Shaw and coworkers have performed 8 ms long simulation of Ubiquitin protein close to its melting temperature (390 K) at neutral pH to obtain these denatured structures.[40] Ubiquitin is a small globular protein with a molar mass of 8433 Da. Moreover, this is a thermally stable and highly soluble protein which does not have any disulfide bonds. Therefore, this has been used in many experimental protein folding studies.[41-43] The protein undergoes pressure denaturation at 5.4 kbar at ambient temperature.[42] Experimental measurements using FT-IR spectroscopy have shown that the volume change for Ubiquitin upon pressure denaturation is -50(±20) mL/mol, or -0.08 (±0.03) nm^3.[42] Moreover, experimental studies have shown that the C- terminal residues, 71-76 of Ubiquitin are highly flexible compared to the core.[43] Ubiquitin consists of 32 non-polar, 20 polar, 12 acidic, and 12 basic amino acid residues, while the native Ubiquitin has 11 hairpin, 12 helical, 11 sheet and 42 coil residues.[39]

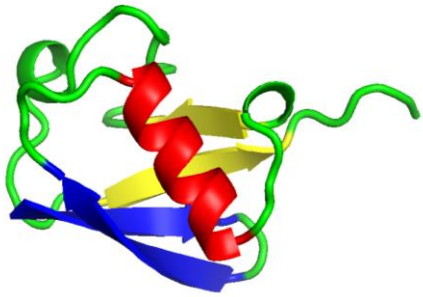


Figure 4.1 Cartoon representation of the native Ubiquitin (PDB ID: 1UBQ) colored by its secondary structure elements. Helix-red, sheet-yellow, hairpin-blue, coil-green. The image was generated with PyMOL molecular visualization software.[44]

4.2 Methodology

4.2.1 Molecular dynamics simulations

The crystal structure (PDB code 1UBQ) was used as the starting structure for the native Ubiquitin simulations. The structure was solvated in a rhombic dodecahedron simulation box with

the distance between two parallel faces being 12 nm. The TIP3P water model was used with AMBER99SB-ILDN, and CHARMM22* FFs, while TIP4P and SPC water models were used with OPLS-AA and GROMOS-53A6 FFs, respectively. Since Ubiquitin is a neutral protein no counterions were added. All simulations were performed with the GROMACS simulation package using version 2016 or 2016.4.[45-46] Molecular dynamics simulations were performed in the isothermal-isobaric ensemble (NpT) at 300 K temperature and 1 bar, 1 kbar, 2 kbar, and 3 kbar pressures. All protein bonds were constrained with the LINCS algorithm,[47] and water bonds with the Settle algorithm[48]. A time step of 2 fs was used to integrate the equation of motion with the Leap Frog algorithm.[49] The particle mesh Ewald (PME) technique[50] was used to calculate electrostatic interactions with a 0.12 nm Fourier grid spacing. The verlet cut-off scheme with a cut-off distance of 1.05 nm was used for van der Waals and electrostatic interactions, and the same cut-off distances were used for all the FFs even though the different FFs were developed with different cut-off values. Periodic boundary conditions and the minimum image convention were applied to all the systems.

The systems were minimized for 1000 steps using the steepest descent algorithm followed by 100 ps equilibration. For the equilibration, the Berendsen temperature coupling and Berendsen pressure coupling were used.[51] The Parrinello-Rahman pressure coupling,[52-53] and Nose-Hoover temperature coupling[54-55] were used for the production phase. The production runs were continued up to 100 ns.

4.2.2 Analysis

For the volume calculations, the number of water molecules around the protein for a given snapshot $(N_1)_2$ is calculated by counting the water molecules within a fixed local volume centered on protein. The corresponding number of water molecules in bulk water $(N_1)_0$ is calculated within

the same volume by superimposing the protein coordinates on pure bulk water configurations. An average number of water molecules in the presence and absence of the protein are then calculated by averaging over all the snapshots. The distance dependent volume of the protein (as shown in Figure 4.4) is calculated using equation (4.5), where the number of waters is calculated as a function of distance away from the surface of the protein. Then, the final protein volume is calculated by averaging the distance dependent volume over a region where the volume is not changing. In this study we have averaged the distance dependent protein volume over 1.5 - 2.0 nm region to obtain the final average protein volume.

The volume of a residue is calculated by assigning each water molecule to a heavy atom of residue based on their proximity. Each water is assigned to the closest heavy atom. The distance was calculated between the center of mass of each water molecule and the heavy atoms of each residue to determine proximity. The sum of the residue volumes provides the total protein volume which can be compared with experimentally measured thermodynamic protein volume as the residue-based thermodynamic properties themselves are not accessible.[20]

In this study the isothermal compressibility ($\kappa^*_{T,2}$) of protein and its residues were calculated using the pressure derivative of the volume after fitting the volumes to a quadratic in pressure.

$$\kappa^*_{T,2} = -\frac{1}{V_2^*}\left(\frac{\partial V_2^*}{\partial P}\right)_{T,m_2} \tag{4.7}$$

This expression was used to calculate the compressibility instead of equation (4.6), since there is no clear way of decomposing the compressibility in equation (4.6) into residue-based contributions.[20] In contrast, equation (4.7) enables the decomposition of compressibility into residue-based contributions.

4.3 Results and Discussion

The C_α root mean square deviations (RMSD) were calculated after a translational and rotational fit to the initial structure to determine the stability of the protein structure during the simulations at 1 bar, 1 kbar, 2 kbar, and 3 kbar. The RMSDs were calculated only for the first 70 residues of Ubiquitin since the C-terminal residues of Ubiquitin are very flexible at all pressures. Figure 4.2 shows the RMSDs calculated for Ubiquitin with the AMBER FF. As the RMSDs fluctuate around 0.1 nm for all the pressures, we can conclude that the simulated Ubiquitin is stable, and stays in native conformation over the simulation time (100 ns) at 1 bar and higher pressures. The RMSDs were also calculated for the other three FFs and a similar behavior was observed.

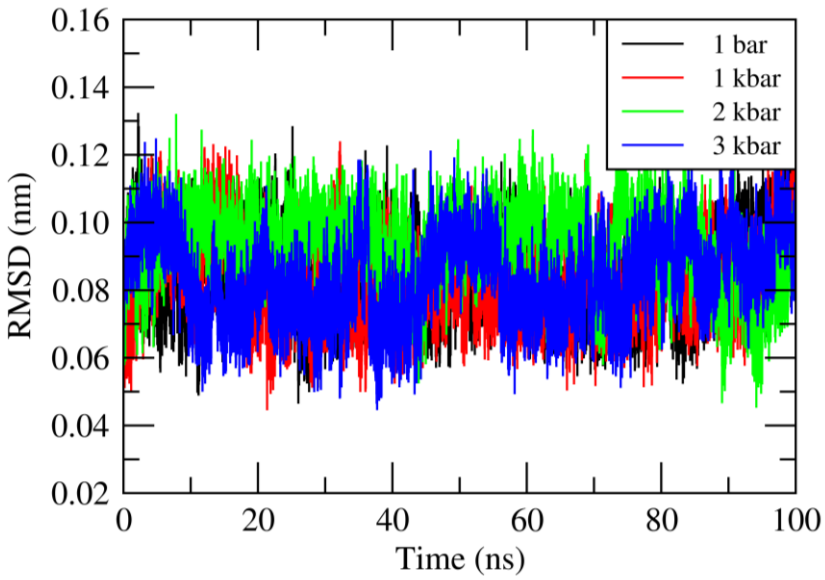


Figure 4.2 The C_α RMSDs calculated for first 70 residues of Ubiquitin with AMBER FF at 1 bar, 1 kbar, 2 kbar, and 3 kbar. All the simulations were performed at 300 K temperature.

Figure 4.3 depicts the water probability distribution around the protein (g_{21}) as a function of distance from the surface of the protein at different pressures for AMBER FF. After about 1 nm from the surface of the protein, the water distribution reaches the bulk pure water distribution (g_{21}=1) indicating that there are no significant protein-water interactions beyond that distance. Water becomes more structured around the protein as the pressure increases. The bottom figure shows the water probability distribution around the protein at 1 bar for the four different FFs.

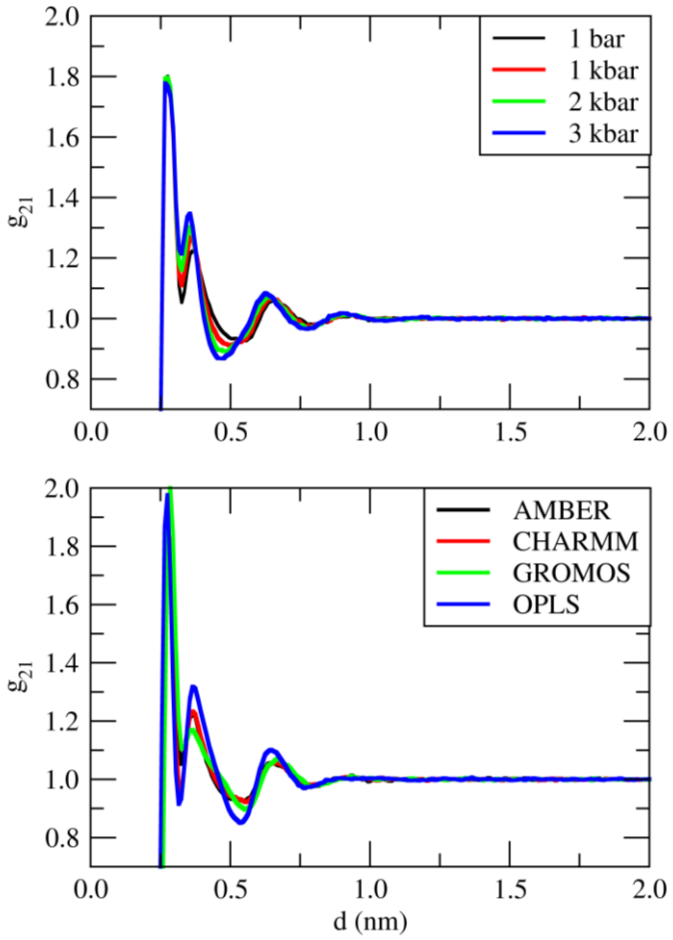


Figure 4.3 Water probability distribution around protein as a function of distance from the surface of the protein, Top: at 1 bar, 1 kbar, 2 kbar, and 3 kbar for AMBER FF. Bottom: at 1 bar for AMBER, CHARMM, GROMOS, and OPLS FFs. The line corresponds to AMBER FF is underneath the line correspond to OPLS FF and not visible.

We can observe that there are slight changes in the probability of finding a water molecule around the protein especially in the second and third solvation shells. However, the AMBER and OPLS FFs seem to have similar behavior and suggest more water around protein compared to the other two FFs.

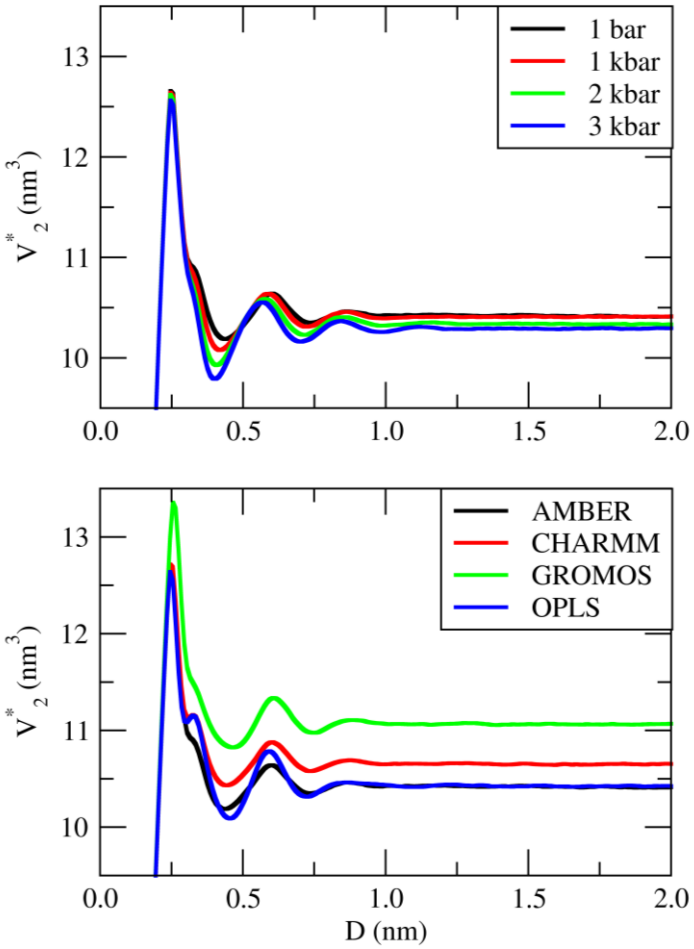


Figure 4.4 Pseudo protein volume as a function of integrtion distance Top: at 1 bar, 1 kbar, 2 kbar, and 3 kbar. Bottom: at 1 bar for AMBER, CHARMM, GROMOS, and OPLS FFs.

The pseudo volume of the protein (V_2^*) as a function of distance from the protein surface is shown in Figure 4.4. Here, the distance dependent volume of the protein reaches a constant value once the water distribution reaches the bulk random distribution. As expected, the volume of the protein decreases as the pressure goes up. A similar pattern is observed for the distance dependent

volume for different FFs at 1 bar. However, the pseudo volume is different for the different FFs. The highest Ubiquitin volume is given by the GROMOS FF, while the AMBER and OPLS FFs display the lowest volumes.

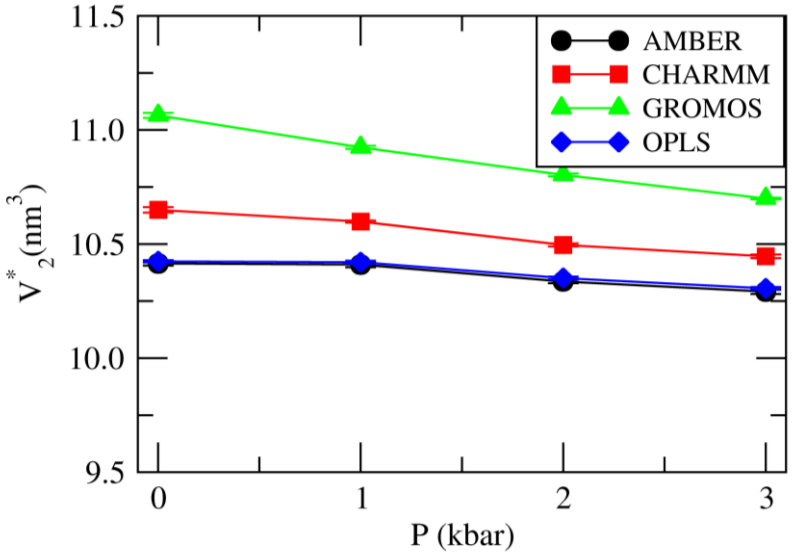


Figure 4.5 Pseudo volume of Ubiquitin as a function of pressure for AMBER, CHARMM, GROMOS, and OPLS FFs.

Figure 4.5 displays the pseudo volume of Ubiquitin as a function of pressure for different FFs. As observed in Figure 4.4, the volume of protein goes down with pressure for all the FFs.The AMBER and OPLS FFs seem to have a similar volume at all pressures while the CHARMM FF provides a slightly higher volume compared to these two FFs. The volumes obtained with the GROMOS FF shows significant deviation from the other three FFs and provides the highest volumes at all pressures. The volume difference between the AMBER and GROMOS FFs, which provides the maximum volume difference between the FFs at 1bar is 0.65 nm^3. Unfortunately, we could not find experimentally measured Ubiquitin partial molar volume in the literature. Hence, we cannot deduce which FF produces more accurate volumes. The changes in volume among different FFs can be due to the different partial charges and non-bonded parameters used by these

FFs. Moreover, as mentioned in the introduction, the choice of water model strongly affects the accuracy of the results.[34] Another possible explanation for the GROMOS FF acting differently would be that it is a united atom FF, while the others are all atom FFs. Considering the volume change from 1 bar to 3 kbar, it varies from -(0.13-0.36) nm^3 among the FFs with the AMBER and OPLS FFs showing the smallest change in volume (-0.13 nm^3), and the GROMOS FF showing the largest change in volume (-0.36 nm^3). For comparison, the volume of simulated pure water at 1 bar was 0.03 nm^3 for TIP3P, SPC, or TIP4P models. Therefore, the volume change from 1 bar to 3 kbar, in terms of number of water molecules, ranges from 4 -12.

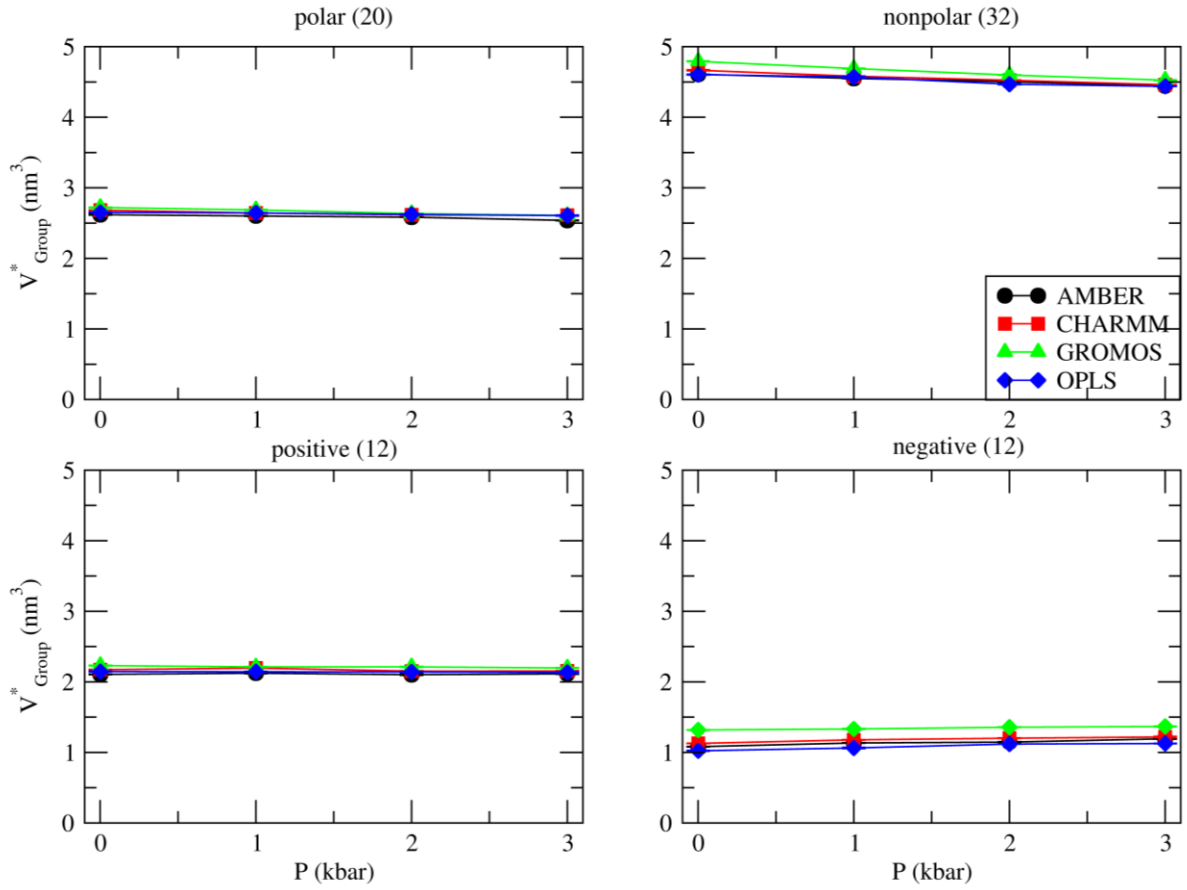


Figure 4.6 Pseudo volume of polar, nonpolar, positively charged, and negatively charged groups as a function of pressure for AMBER, CHARMM, GROMOS, and OPLS FFs. Here, the volume of each group was obtained by summing over the residue volumes in each group and it should be noted that the different groups have a different number of residues as denoted within parenthesis.

The volume of the protein was decomposed into contributions from the physicochemical groups as discussed in the analysis section and shown in Figure 4.6. Nonpolar group residues provide the major contribution to the volume of Ubiquitin and this is due to the 32 nonpolar residues out of 76 residues (42%) in Ubiquitin. The smallest contribution is given by the negatively charged group (12 residues). One interesting feature observed here is that the volume of acidic group increases while the volume of polar, and nonpolar groups decrease with pressure for all FFs. We do not see a significant change in volume with pressure for the basic group, and this is consistent among different FFs. Furthermore, the change in volume with pressure is significant for the nonpolar group and we could say that the nonpolar residues are more sensitive to the pressure and the overall negative protein volume change under pressure is due to the nonpolar residues. We observe approximately similar volumes for all the four FFs except for the GROMOS FF, where the GROMOS FF provides slightly larger volumes, especially for nonpolar and acidic groups compared to the other three FFs.

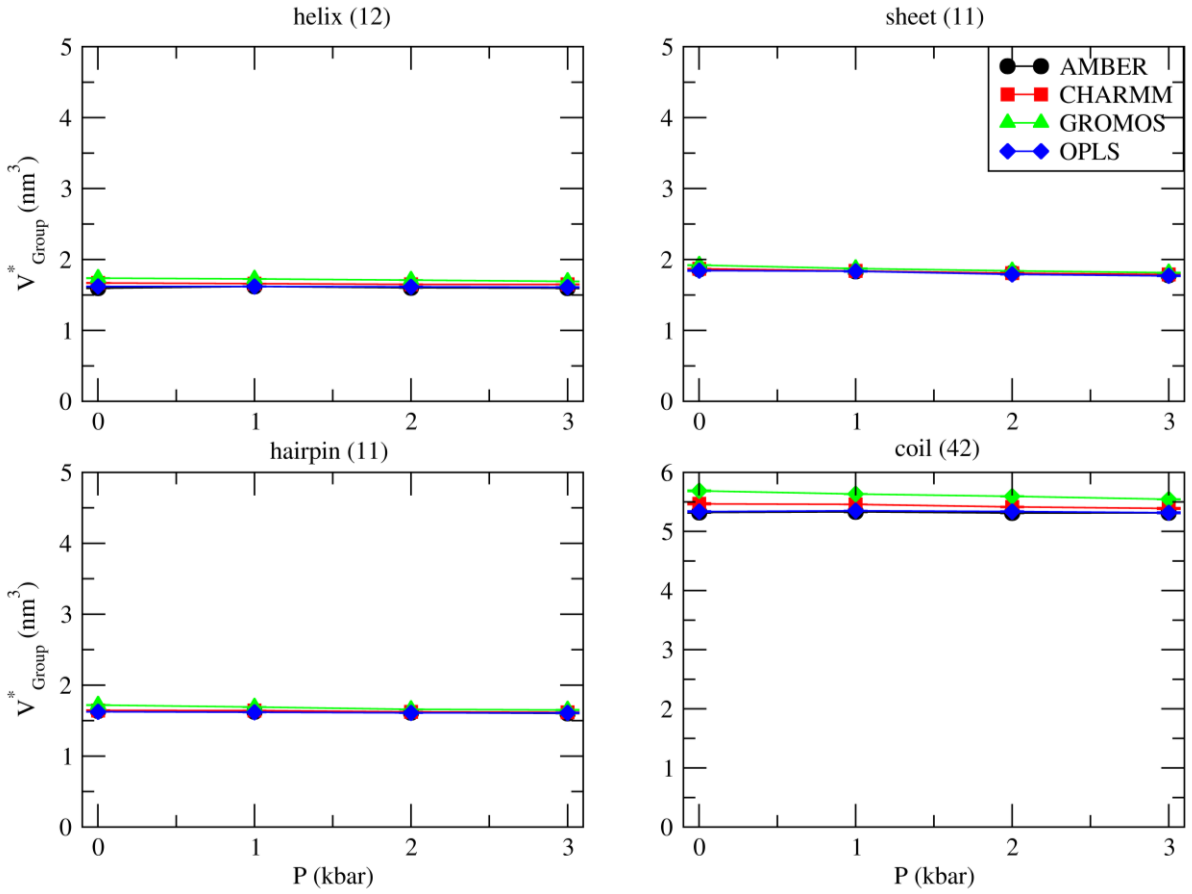


Figure 4.7 Pseudo volume of the helix, sheet, hairpin, and coil groups as a function of pressure for AMBER, CHARMM, GROMOS, and OPLS FFs. Here, the volume of each group was obtained by summing over the residue volumes in each group and it should be noted that the different groups have a different number of residues as denoted within parenthesis.

Figure 4.7 depicts the volume decomposition of Ubiquitin according to its secondary structure. Volumes of each group slightly decrease with pressure while the sheet and coil residues seem more affected by pressure than the helix and hairpin residues. Similar trends are observed for all the FFs and again the GROMOS FF seems to have slightly larger residue volumes, especially in the coil group, compared to other FFs.

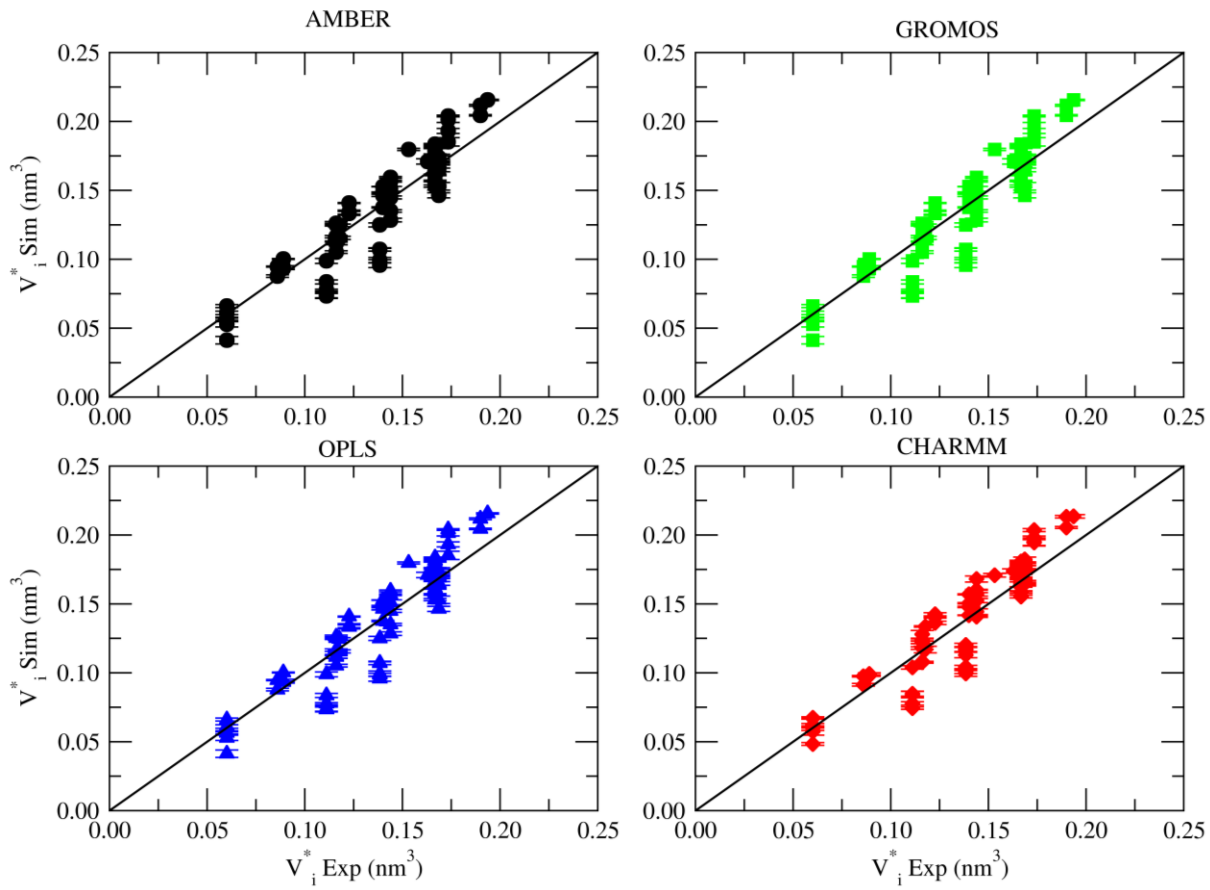


Figure 4.8 Simulated vs. experimental pseudo volumes of 76 residues in Ubiquitin at 1 bar. Experimental residue volumes were not obtained for Ubiquitin but determined from apparent molar volumes of amino acids at 298 K and 1 bar.

Calculated residue volumes were then compared with 'experimentally' obtained residue volumes. Figure 4.8 shows the experimental vs simulated volumes for all 76 residues in Ubiquitin at 1 bar. Zamyatnin has determined the experimental residue volumes based on the aparent molar volumes of amino acids at 1 bar.[56] There is a reasonable agreement between the experimental and simulated volumes for all the FFs. However, this is not a very accurate comparison as the experimental data shown here are not calculated for the residues in Ubiquitin. Therefore, the effects of the environment on the residue volumes are not included in these experimental values. There are several simulated volumes for the same residue (compared to a single experimental value)

indicating that the volume of each residue depends on where it is in the sequence, or in space, and this suggests that the residue volumes depend on their environment.

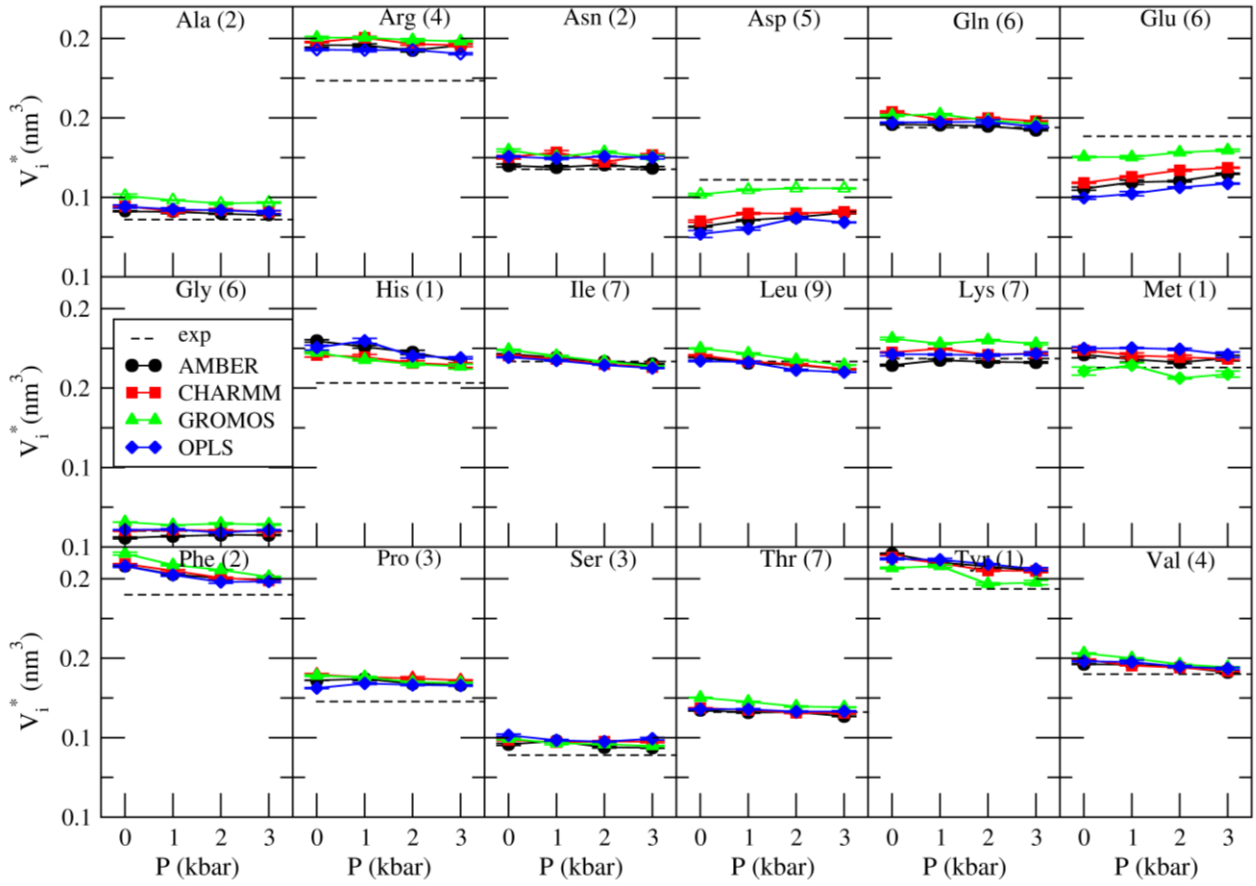


Figure 4.9 Average pseudo volumes of each amino acid residue present in Ubiquitin as a function of pressure. The total number of each residue in Ubiquitin is shown in parenthesis. The dashed line represents the experimental residue values at 1bar and used as a guideline to compare results. Experimental residue volumes were not obtained for Ubiquitin but determined from apparent molar volumes of amino acids at 298 K and 1 bar.

Figure 4.9 shows the average volume of each residue as a function of pressure for all the FFs. Again, the experimental data shown here were not calculated for residues in Ubiquitin and are just used as a guideline. We observe the smallest volume for glycine and the largest volume for tyrosine as observed experimentally (and expected). The volume of each residue decreases with pressure except for the aspartic and glutamic acids. This agrees with what we observed in Figure

4.6 where the volume of the negatively charged group decreases with pressure. Similar trends are observed for all the FFs while the GROMOS FF provides slightly higher volumes, especially for acidic residues, compared to other FFs. However, for both acidic residues, GROMOS produces volumes in more reasonable agreement with the experimental data compared to the other FFs.

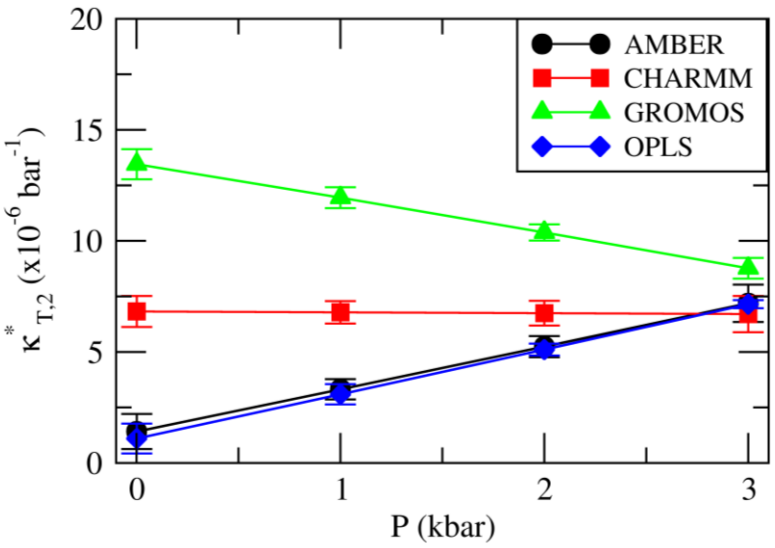


Figure 4.10 Pseudo compressibility of Ubiquitin as a function of pressure for AMBER, CHARMM, GROMOS, and OPLS FFs.

The pseudo compressibility calculated for Ubiquitin as a function of pressure is shown in Figure 4.10. Positive compressibility values are observed at all pressures, and for all the FFs. Simulated Ubiquitin compressibilities range from 1.4×10^{-6} – 13.4×10^{-6} bar^{-1} with different FFs. As discussed in the introduction, the compressibility of a protein is very small (for native globular proteins they range from 5×10^{-6}-15×10^{-6} bar^{-1}) and our compressibility values agree well with that observation. The trends in compressibility differ for the different FFs, except for the AMBER and OPLS FFs. This is expected as we noted a similar story for the volumes and the compressibility as given by the first pressure derivative of the volume. The compresibility of the protein goes up for

the AMBER and OPLS FFs, while it goes down for the GROMOS and CHARMM FFs with pressure increases.

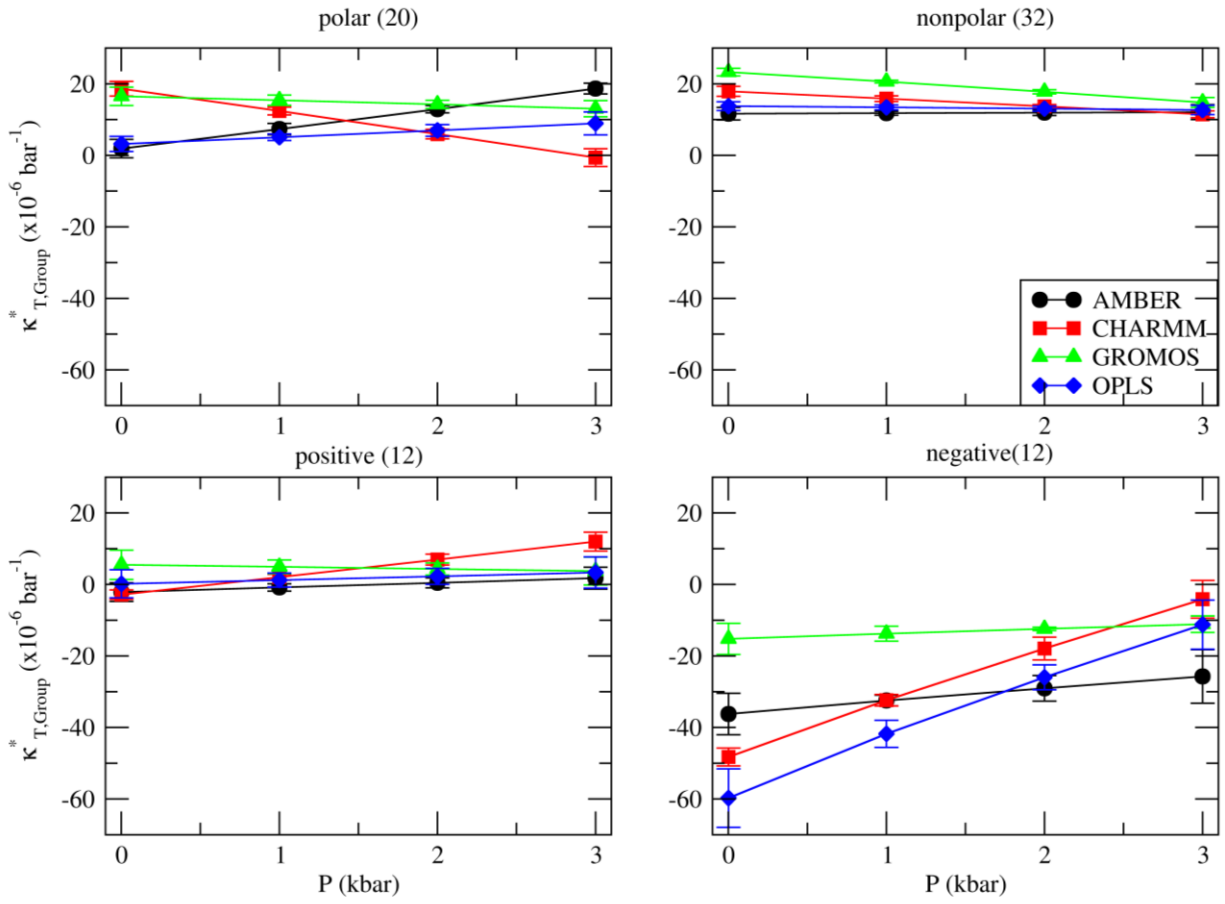


Figure 4.11 Pseudo compressibilities of polar, nonpolar, positively charged, and negatively charged groups as a function of pressure for AMBER, CHARMM, GROMOS, and OPLS FFs. The number of residues in each group is shown within the parenthesis.

Figure 4.11 shows the decomposition of the pseudo compressibility into physicochemical group contributions as a function of pressure. We observe positive compressibilities for polar, nonpolar, and positively charged groups, except there is a slightly negative compressibility for the positively charged group at 1 bar. As already noted, the negatively charged group acts differently by displaying a negative compressibility for all the FFs at all pressures. In agreement with this, an experimental study has shown that the zwitterionic groups of amino acids have a negative

compressibility.[57] The overall positive compressibility of Ubiquitin is due to the polar and nonpolar groups.

The negative compressibilty observed for negatively charged group can be explained using the compressibility expression involving water fluctuations (equation (3.11)). We obtain negative compressibilities when the water fluctuations in the presence of protein are lower than the water fluctuations in the absence of protein (pure water). The fluctuations will be lower when there are strong interactions between the protein and waters. Therefore, the negative compressibility observed for negatively charged group can be due to strong side chain-water interactions, which is known as electrostriction. Furthermore, the negatively charged group shows larger deviations for compressibilities among FFs than do other groups. This means that the negatively charged residues-water interactions are quite different among the different FFs. This can be due to the fact that the different FFs use slightly different approaches to obtain charges during the parameterization process. However, we did not observe significant differences for volumes since the size parameters are more similar among different FFs.

In addition, the pseudo compressibility was decomposed into the contributions from the secondary structure elements (Figure 4.12). Unlike the physicochemical groups, all the compressibilities are positive with the exceptions of two negative points for helix and coil residues at 1 bar using the AMBER and OPLS FFs, respectively. Here, again we observe that the compressibility goes up or down with pressure in each group depending on the FF and this is due to the slight changes in volume at each pressure for different FFs.

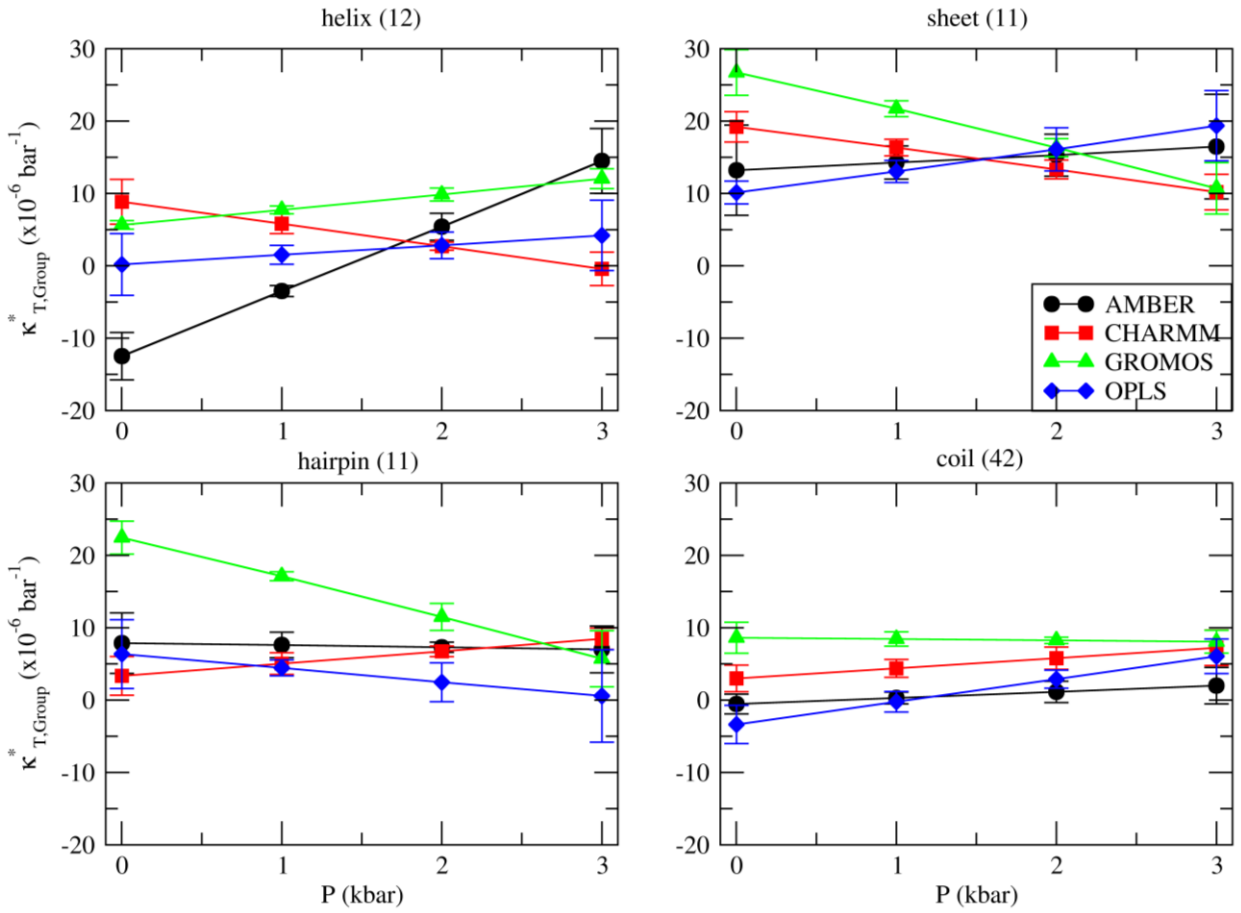


Figure 4.12 Pseudo compressibilities of the helix, sheet, hairpin, and coil groups as a function of pressure for AMBER, CHARMM, GROMOS, and OPLS FFs. The number of residues in each group is shown within the parenthesis.

4.4 Conclusions

The pseudo volume and pseudo compressibility of Ubiquitin were calculated using the FST approach at 1 bar, 1 kbar, 2 kbar, and 3 kbar, and the results were compared among the AMBER99SB-ILDN, CHARMM22*, GROMOS53A6, and OPLS-AA biomolecular FFs. Similar trends were observed for the pseudo volume of the protein and the residue-based volume contributions to the protein as a function of pressure for all the FFs. However, quantitatively there are slight variations in volumes among different FFs. This might be due to the fact that the non-bonded interactions are treated differently in different FFs, and that different water models are

used in each FF. Among the four FFs, GROMOS seems to produce higher volumes and this might be due to the fact that GROMOS is the only united atom FF used in this study. Still, we could not conclude which FF produces more accurate results without having experimental evidence.

Different trends were observed for the pseudo compressibility as a function of pressure as there were slight changes in volume among different FFs and the compressibility calculations involved pressure derivative of volume However, all the FFs produce the same sign for the total compressibility and group-based compressibilities. Moreover, the protein compressibilities given by the different FFs are within the acceptable range of compressibilities observed experimentally.

4.5 References

1. Dubins, D. N.; Filfil, R.; Macgregor, R. B.; Chalikian, T. V., *Biochemistry* **2003,** *42* (29), 8671-8678.

2. Royer, C. A., [16] Application of pressure to biochemical equilibria: The other thermodynamic variable. In *Methods in enzymology*, Elsevier: 1995; Vol. 259, pp 357-377.

3. Chalikian, T. V., CHAPTER 21 Partial Molar Volumes of Proteins in Solution. In *Volume Properties: Liquids, Solutions and Vapours*, The Royal Society of Chemistry: 2015; pp 542-574.

4. Royer, C. A., *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **2002,** *1595* (1-2), 201-209.

5. Grigera, J. R.; McCarthy, A. N., *Biophysical journal* **2010,** *98* (8), 1626-1631.

6. Roche, J.; Caro, J. A.; Norberto, D. R.; Barthe, P.; Roumestand, C.; Schlessman, J. L.; Garcia, A. E.; García-Moreno E, B.; Royer, C. A., *Proceedings of the National Academy of Sciences of the United States of America* **2012,** *109* (18), 6945-6950.

7. Chalikian, T. V.; Gindikin, V. S.; Breslauer, K. J., *Journal of molecular biology* **1995,** *250* (2), 291-306.

8. Kharakoz, D.; Sarvazyan, A., *Biopolymers* **1993,** *33* (1), 11-26.

9. Gekko, K.; Hasegawa, Y., *Biochemistry* **1986,** *25* (21), 6563-6571.

10. Taulier, N.; Chalikian, T. V., *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **2002,** *1595* (1-2), 48-70.

11. Imai, T.; Kinoshita, M.; Hirata, F., *The Journal of Chemical Physics* **2000,** *112* (21), 9469-9478.

12. Chalikian, T. V., *Annual review of biophysics and biomolecular structure* **2003,** *32* (1), 207-235.

13. Dadarlat, V. M.; Post, C. B., *The Journal of Physical Chemistry B* **2001,** *105* (3), 715-724.

14. Kirkwood, J. G.; Buff, F. P., *The Journal of Chemical Physics* **1951,** *19* (6), 774-777.

15. Jiao, Y.; Smith, P. E., *The Journal of chemical physics* **2011,** *135* (1), 07B602.

16. Patel, N.; Dubins, D. N.; Pomes, R.; Chalikian, T. V., *The Journal of Physical Chemistry B* **2011,** *115* (16), 4856-4862.

17. Floris, F. M., *The Journal of Physical Chemistry B* **2004,** *108* (41), 16244-16249.

18. Sangwai, A. V.; Ashbaugh, H. S., *Industrial & Engineering Chemistry Research* **2008,** *47* (15), 5169-5174.

19. Ploetz, E. A.; Smith, P. E., *The Journal of Physical Chemistry B* **2014,** *118* (45), 12844-12854.

20. Ploetz, E. A.; Smith, P. E., *Biophysical chemistry* **2017,** *231*, 135-145.

21. Patel, N.; Dubins, D. N.; Pomès, R.; Chalikian, T. V., *Biophysical chemistry* **2012,** *161*, 46-49.

22. Harano, Y.; Imai, T.; Kovalenko, A.; Kinoshita, M.; Hirata, F., *The Journal of Chemical Physics* **2001,** *114* (21), 9506-9511.

23. Yamazaki, T.; Imai, T.; Hirata, F.; Kovalenko, A., *The Journal of Physical Chemistry B* **2007,** *111* (5), 1206-1212.

24. van Gunsteren, W. F.; Berendsen, H. J. C., *Angewandte Chemie International Edition in English* **1990,** *29* (9), 992-1023.

25. van Gunsteren, Wilfred F.; Dolenc, J., *Biochemical Society Transactions* **2008,** *36* (1), 11-15.

26. Ploetz, E. A.; Smith, P. E., *Advances in chemical physics* **2013,** *153*, 311.

27. Ploetz, E. A.; Smith, P. E., *The Journal of Physical Chemistry B* **2015,** *119* (25), 7761-7777.

28. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., *Proteins: Structure, Function, and Bioinformatics* **2010,** *78* (8), 1950-1958.

29. Piana, S.; Lindorff-Larsen, K.; Shaw, D. E., *Biophysical journal* **2011,** *100* (9), L47-L49.

30. Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F., *Journal of Computational Chemistry* **2004,** *25* (13), 1656-1676.

31. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., *Journal of the American Chemical Society* **1996,** *118* (45), 11225-11236.

32. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., *The Journal of Physical Chemistry B* **2001,** *105* (28), 6474-6487.

33. Jorgensen, W. L.; McDonald, N. A., *Journal of Molecular Structure: THEOCHEM* **1998,** *424* (1-2), 145-155.

34. Hess, B.; van der Vegt, N. F. A., *The Journal of Physical Chemistry B* **2006,** *110* (35), 17616-17626.

35. Mark, P.; Nilsson, L., *The Journal of Physical Chemistry A* **2001,** *105* (43), 9954-9960.

36. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., *Proteins: Structure, Function, and Bioinformatics* **2006,** *65* (3), 712-725.

37. MacKerell Jr, A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S., *The journal of physical chemistry B* **1998,** *102* (18), 3586-3616.

38. Høiland, H.; Hedwig, G., 6 Compressibilities of amino acids, peptides and proteins in aqueous solution. In *Structural and Physical Data I*, Springer: 2003; pp 6001-6017.

39. Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J., *Journal of molecular biology* **1987,** *194* (3), 531-544.

40. Piana, S.; Lindorff-Larsen, K.; Shaw, D. E., *Proceedings of the National Academy of Sciences* **2013**, 201218321.

41. Wintrode, P. L.; Makhatadze, G. I.; Privalov, P. L., *Proteins: Structure, Function, and Bioinformatics* **1994,** *18* (3), 246-253.

42. Herberhold, H.; Winter, R., *Biochemistry* **2002,** *41* (7), 2396-2401.

43. Kitahara, R.; Yamada, H.; Akasaka, K., *Biochemistry* **2001,** *40* (45), 13556-13563.

44. DeLano, W. L., *CCP4 Newsletter On Protein Crystallography* **2002,** *40* (1), 82-92.

45. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., *Bioinformatics* **2013,** *29* (7), 845-854.

46. Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E., *Journal of Chemical Theory and Computation* **2010,** *6* (2), 459-466.

47. Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G., *Journal of computational chemistry* **1997,** *18* (12), 1463-1472.

48. Miyamoto, S.; Kollman, P. A., *Journal of computational chemistry* **1992,** *13* (8), 952-962.

49. Hockney, R.; Goel, S.; Eastwood, J., *Journal of Computational Physics* **1974,** *14* (2), 148-158.

50. Darden, T.; York, D.; Pedersen, L., *The Journal of chemical physics* **1993,** *98* (12), 10089-10092.

51. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., *The Journal of Chemical Physics* **1984,** *81* (8), 3684-3690.

52. Nosé, S.; Klein, M., *Molecular Physics* **1983,** *50* (5), 1055-1076.

53. Parrinello, M.; Rahman, A., *Journal of Applied physics* **1981,** *52* (12), 7182-7190.

54. Hoover, W. G., *Physical review A* **1985,** *31* (3), 1695.

55. Nosé, S., *Molecular physics* **1984,** *52* (2), 255-268.

56. Zamyatnin, A., *Progress in biophysics and molecular biology* **1972,** *24*, 107-123.

57. Kharakoz, D. P., *Biochemistry* **1997,** *36* (33), 10276-10285.

Chapter 5 - The Effects of Pressure on the Conformations of the *Lac* Repressor

5.1 Introduction

The lactose (*lac*) operon has been a popular model for studying and understanding the genetic and allosteric regulation of proteins.[1-2] Allostery is the thermodynamic process by which the binding of a first ligand affects the binding of the second ligand at a distance within the same protein.[3-4]The *lac* operon is a collection of genes with a single promoter which decomposes lactose into simple sugars in *Escherichia coli* bacteria. The lactose repressor protein (LacI) plays the main role in inhibiting the expression of *lac* operon by binding to the DNA sites known as operators within the operon.[5-6] LacI is a tetrameric protein with identical domains, with a molecular weight of 38,000, where each monomer contains 360 amino acids.[7-8] For LacI, sugar molecules such as allolactose,[9] or isopropyl-D-thiogalactoside (IPTG),[10] work as inducers which weaken the DNA binding affinity of the repressor through allosteric changes. In the presence of allolactose, LacI binds to allolactose and is released from the operator. This allows RNA polymerase to transcribe the genes. The anti-inducer orthonitrophenyl-β-D-fucoside (ONPF) strengthen the DNA binding affinity of the repressor. Nevertheless, there is no known function for ONPF in *E-coli*.[11]

There are two conformations of LacI, as the DNA bound and DNA unbound involved in the function of repression and induction.[5] X-ray crystal structures have been determined for both conformations. Figure 5.1 shows the DNA bound LacI structure (PDB id 1EFA).[11] Residues 1-49 compose the 'headpiece' domain which contains a helix-turn-helix motif and binds to the major groove of the DNA. Residues 50-58 are belong to the 'hinge helix' (HH) which interacts with the center of the operator in the minor grove.[11] In the absence of the DNA, the HH is unstable[12] and the residues 1-61 (head group) can move freely relative to the core domain (residues 62-333).[11]

The HH connects the DNA binding domain (DBD) to the core domain and the core domain consists of two subdomains namely, N-subdomain (residues 62-161 and 293-320) and C-subdomain (residues 162-289 and 321-329).[11] The anti-inducer molecule binds to a pocket in between the N and C subdomains. Lastly, the C-terminal residues 340-357 facilitate the tetramerization of LacI.[11]

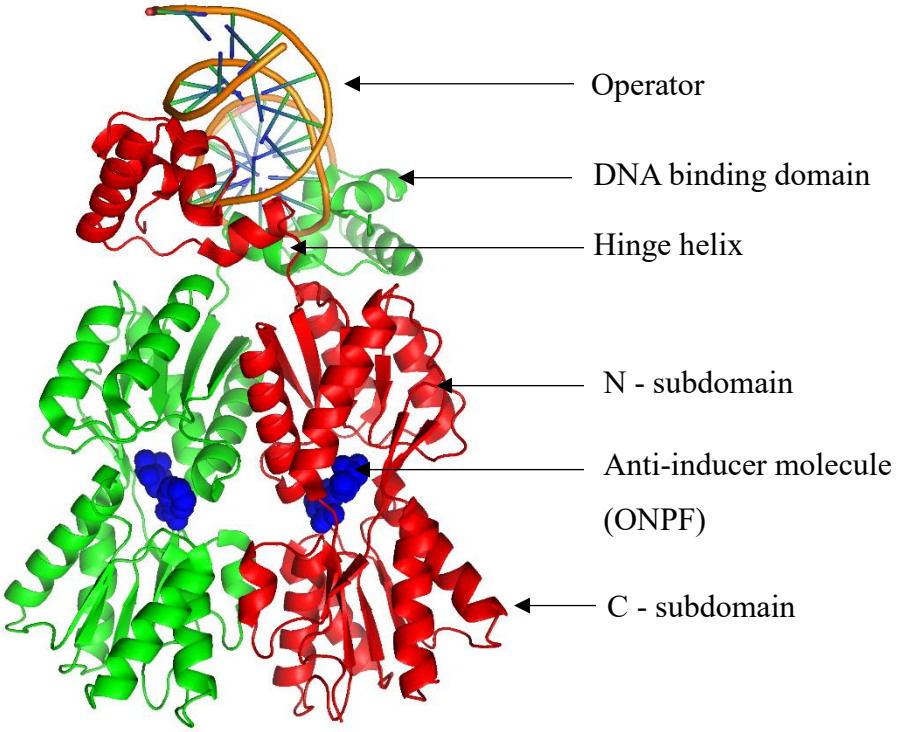


Figure 5.1 Cartoon representation of the LacI dimer (PDB id: 1EFA) bound to operator and anti-inducer, ONPF. Chain A of the dimer is shown in green and chain B in red. C-terminal residues are not shown here. The image was generated with PyMOL molecular visualization software.[13]

Analysis of the crystal structures have shown that an allosteric transition occurs upon inducer (IPTG) binding, and this involves the movement of the N-subdomains relative to each other and to the C-subdomains.[11-12] This alters the interactions between the core domain and the DNA binding domains which then leads to destabilization of the HHs and an increase in the mobility of the DNA binding domains.[12, 14-16] Ultimately, this leads to weaken the binding of the HH to the minor grove of the DNA.

Figure 5.2 shows the DNA bound and DNA unbound (PDB id – 1LBH)[12] structures after fitting to the C- subdomain of the DNA unbound structure. Here, the DNA, anti-inducer molecules (ONPF) and the head groups are not shown for the DNA bound structure and the inducer molecules (IPTG) are not shown for the DNA unbound structure. The root-mean-square deviation value of the N- subdomains is 0.2 nm after fitting the C- subdomains of both crystal structures.

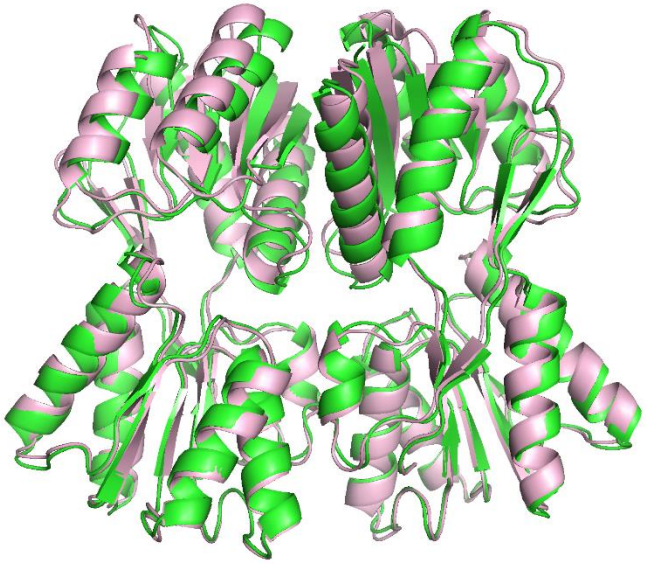


Figure 5.2 Cartoon representation of the core domains of the DNA bound (pink, PDB id: 1EFA) LacI after fitting to the C_α carbons in the C-subdomains of DNA unbound (green, PDB id:1LBH) structure. The image was generated with PyMOL molecular visualization software.

Experimental studies have shown that mutating some of the non-conserved positions (positions change during evolution) in proteins with different amino acids display progressive effects on function.[17] These types of positions are known as rheostat positions. Interestingly, the 52nd position (Valine 52/V52) of the LacI sequence is a rheostat position.[17] More details about the rheostat positions of the LacI will be discussed in next chapter.

Molecular Dynamics (MD) simulations have been performed to investigate the structural dynamics of LacI as the crystal structures do not provide a detailed mechanism for structural

changes. Pettitt and coworkers have studied the monomeric structure of the DBD of a variant of LacI (N*lac*-P) in the absence of DNA.[18] Their results suggest that the hinge helix destabilized in the absence of the DNA, which agrees with the experimental data.[14-16] Also, their simulation data indicates that the hinge helix moves independently from the DBD.[18] Schulten and coworkers have performed a multiscale simulation of DNA bound to the LacI tetramer.[6] They have observed that the domains in the protein structure are very stable and move relative to each other like rigid bodies. Moreover, the DBD movement with respect to the core of the protein was able to absorb strain from the DNA loop. They have concluded that the rotation of this head group is crucial for the functioning of LacI. Using targeted MD, Flynn and coworkers have shown that the monomers of homodimer follow asymmetric dynamics during the allosteric conformational pathway.[5] The conformational pathway from DNA-bound to DNA-unbound state is explained using the interactions of specific residues, especially at the interface of monomers of the dimer in N subdomains.[5] Sun and coworkers have investigated the effect of the hinge region on the non-specific binding of DNA using MD simulations.[19] Their findings show that the HH is disordered when the head group non-specifically binds to the DNA, and this contributes 50% towards the stability of the head group/DNA complex. Furthermore, the hinge region mainly stabilizes the head group/DNA via electrostatic interactions between protein-DNA and salt ions. The computer simulation of HH itself has shown that it is very disordered over a 1μs time scale in solution and can be trapped in a helical formation with high salt concentration.[20] The simulation of DBD including HH has shown that DBD is mainly contributing to the stability of HH. Their results suggest that the helix structure stability is mainly affected by environmental factors such as salt concentration and the presence of DNA.

Proteins are dynamical in nature and the function and activity of a protein is strongly dependent on its environment and hydration shell.[4, 21-22] Therefore, water plays an important role in macromolecular structure and stability.[23] Indeed, Levy and coworkers have defined water as 'twenty first amino acid' as the water is crucial in protein function and activity.[23] For instance, an experimental study by Salvay and coworkers on hemoglobin has shown the effect of hydration on the allosteric mechanism for regulating cooperative interactions.[24] Their findings suggest that excess water molecules on the surface of the hemoglobin are necessary to stabilize the high-affinity transition state in the hemoglobin cooperative reaction. Pressure perturbation is a powerful technique to study the changes in the hydration of macromolecules.[25-27] A high-pressure spectroscopy study on the allosteric mechanism of human cytochrome P450 3A has shown that there is a pressure sensitive equilibrium between two conformational states which differ by the degree of hydration and water accessibility of the heme pocket.[25] Royer and coworkers have investigated the *lac* repressor subunit interactions and protein-operator association using high pressure and fluorescence techniques.[26, 28] They conclude that, under high hydrostatic pressure, the tetramer of *lac* repressor undergoes dissociation and the volume change upon dissociation to a dimer is negative. They have observed that in the presence of inducer, IPTG, the tetramer is more stable and requires higher pressures for dissociation than without the inducer. Furthermore, the operator-tetramer dissociates more easily upon applying pressure compared to the tetramer itself indicating the destabilization of the tetramer in the presence of the operator.

Here, we perform MD simulations at higher pressure (3 kbar) to investigate the role of hydration on structural changes of *lac* repressor protein. A truncated dimeric structure of the LacI is employed in our simulations since the tetrameric protein is a 'dimer of dimers'[11-12, 29] and retains all functionality of the tetramer other than the DNA looping.[30-33]

5.2 Methodology

5.2.1 System setup

The crystal structure with PBD code 1efa was used as the starting structure after adding missing atoms (using AMBER leap) and removing the DNA and the c-terminal tetramerization domain. After modification, each chain of the dimer contains 328 residues with a total number of 9972 atoms. Since the crystal structure does not contain the 1st residue, only residues 2-329 in the crystal structure are used. The protein was solvated with TIP3P water in a rhombic dodecahedron simulation box where the distance between 2 parallel faces was 15 nm. Counterions were added (6 Na^+ ions) to neutralize the system. Simulations were performed in the absence and the presence of anti-inducer molecules. Charges for the anti-inducer, ONPF were obtained using the R.E.D. server.[34] The R.E.D server is an open web service designed to derive charges and non-polarizable force field parameters for new molecules. More details of the ONPF parameters are given in supporting information section.

Another set of simulations were performed with the position restraints (headPR) on the Cα atoms of head group residues (1-61 residues). Restraints on the head group were used to mimic the presence of the DNA as the DNA limits the motion of the head group.

A set of simulations were performed using only the HH peptide to understand the behavior of isolated HH itself (residues 50-58) and the role of its environment on stability of HHs. Simulations were performed by capping both the and N and C terminus of the helix using acetyl and N-methylamide groups respectively. The HH was solvated in a rhombic dodecahedron simulation box where the distance between 2 parallel faces was 6 nm. One chloride ion was added to neutralize the system. Details of the all simulated systems are shown in Table 5.1.

Table 5.1 Details of the systems simulated. All the systems were simulated at 300 K temperature and 1 bar and 3 kbar pressures. Production runs were continued up to 1µs.

System No:	System	No: of ions	No: of water molecules
1	WT LacI free, no ONPF (WT)	6 Na^+	74255
2	WT LacI -head PR, no ONPF (WT-headPR)	6 Na^+	74255
3	WT LacI free, with ONPF (WT/ONPF)	6 Na^+	74242
4	WT LacI -head PR, with ONPF (WT-headPR/ONPF)	6 Na^+	74242
5	Hinge helix (HH)	1 Cl^-	5074

5.2.2 Molecular dynamics simulations

Atomistic MD simulations were performed using the AMBER14SB[35] force field (FF) and the GROMACS 2016 or 2016.4 versions.[36] The AMBER14SB FF is an improved version of the AMBER99SB[37], where the adjustments have been made to the amino acid side chain and backbone parameters. Therefore, this version is recommended for protein and peptide simulations among the other AMBER FF versions.[35] All the simulations were performed in the isothermal isobaric ensemble (NpT) at 300 K temperature, and both 1 bar and 3 kbar pressure. All protein bonds were constrained with the LINCS algorithm,[38] and water with the Settle algorithm.[39] A time step of 2 fs was used to integrate the equations of motion with the Leap Frog algorithm.[40] The particle mesh Ewald (PME)[41] technique with 0.12 nm Fourier grid spacing was used to calculate the coulombic interactions. The Verlet cut-off scheme was used with a 0.9 nm cut-off distance for van der Waals and electrostatic interactions. Periodic boundary conditions and the minimum image convention were applied. All the systems were first minimized for 1000 steps using the steepest descent algorithm. Then the systems were equilibrated up to 100 ps by gradually increasing the temperature

(100, 200 and 300 K) with heavy atom position restraints (1000 kJ mol^{-1} nm^{-2}) and Berendsen T-coupling and Berendsen isotropic P-coupling. The Parrinello-Rahman pressure coupling,[42-43] and the Nose-Hoover temperature coupling[44-45] were used to maintain the constant pressure and temperature during the production simulations. The production simulations were continued up to 1 μs.

5.2.3 Analysis

The Cα Root Mean Square Deviation (RMSD) of each core domain were calculated after a translational and rotational fit to each domain in the initial crystal structure. This was performed to determine the stability of domains. Here, the N-subdomains of chain A and chain B are denoted as N_A and N_B respectively. The C-subdomains of chain A and chain B are denoted as C_A and C_B respectively. A vector angle (θ) was defined to crudely measure the movement of chain A relative to chain B as there is experimental evidence suggesting the asymmetric dynamics of monomers during the allosteric conformational pathway.[5] The schematic representation of the two vectors defined here are shown in, Figure 5.3, and the angle between two vectors is given by, $cos\theta = \overrightarrow{r_1}.\overrightarrow{r_2}/|\overrightarrow{r_1}||\overrightarrow{r_2}|$. The two vectors were defined using the center of mass of each subdomain. The measured angle for the 1efa (DNA bound) crystal structure was 14º, while for the 1lbh (DNA unbound) structure it was 23º.

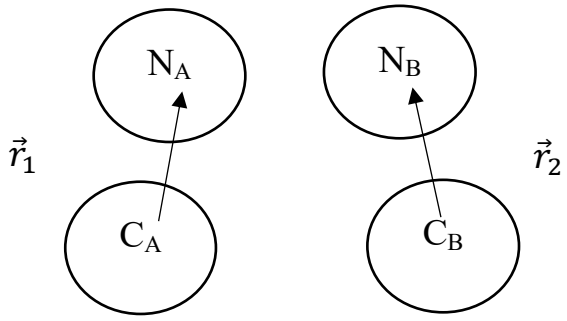


Figure 5.3 Schematic representation of the two vectors defined by connecting N and C subdomains of each chain.

Since the angle is an approximate measurement to define bound and the unbound states, we defined another measure to help differentiate bound and unbound states of LacI dimer. By examination of the movie of the simulated trajectory of WT at 1 bar, we observed that two β-strands (as shown in Figure 5.4) in N subdomains close to the monomer interface had moved towards each other over time. Therefore, the distance between the central Cα atoms (residues 96 and 424) was calculated to follow the movement of these two strands. The corresponding measured distance for the 1efa (DNA bound) crystal structure was 8 Å while for the 1lbh (DNA unbound) structure it was 5 Å.

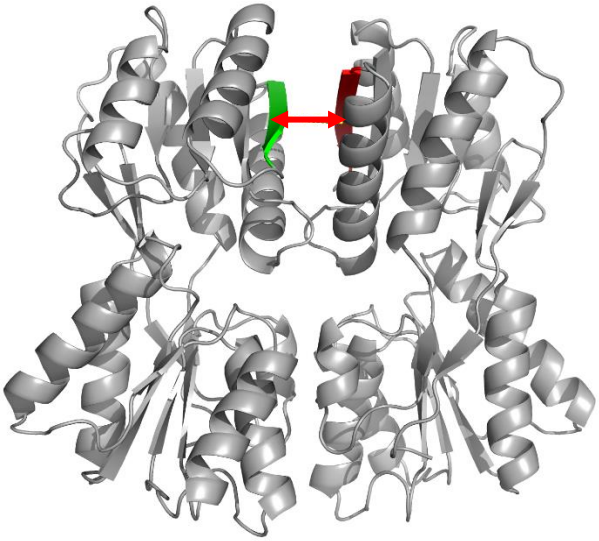


Figure 5.4 The distance defined between two β-strands in chain A (green) and chain B (red) in the N subdomains of the LacI bound crystal structure is shown using a red double arrow (only the core domains of the structure are shown). Image was generated using PyMOL molecular visualization software.[13]

5.2.4 Pressure thermodynamics

A system in equilibrium can be perturbed by changing temperature, pressure, pH or by adding a cosolvent.[46] Here, we assume that there is an equilibrium between DNA-bound and DNA-

unbound states of the *lac* repressor and it is affected by pressure. The effect of pressure on a chemical equilibrium is given by,[47]

$$\left(\frac{\partial \ln K}{\partial P}\right)_T = \frac{\Delta \bar{V}}{RT} \tag{5.1}$$

where K is the equilibrium constant, P is the pressure, $\Delta \bar{V}$ is the partial molar volume difference between the two states, T is the absolute temperature, and R is the gas constant. The volume of the protein was calculated using the particle number fluctuation approach as discussed in chapter 1. The pseudo molar volume of a solute of any size in an infinitely diluted solution is given by,

$$V_2^{*,\infty} = -[\langle N_1 \rangle_2 - \langle N_1 \rangle_0] V_1^0 \tag{5.2}$$

where, $<N_1>_2$ is the average number of water molecules within a fixed local volume centered on the protein, $<N_1>_0$ is the average number of water molecules within the same volume in pure water and V_1^0 is the volume of pure water at the corresponding temperature and pressure. The volume of each amino acid residue is calculated by assigning each water molecule to a heavy atom of an amino acid based on their proximity.

As mentioned in the introduction, the 52nd position of the LacI acts as a rheostat position. To examine how the residues around V52 are hydrated in the bound and unbound forms, the volume of residues within 5 Å and 10 Å from V52 were calculated. The first 60 ns of the production run was used to calculate the bound state properties since the calculated angle (θ) stays close to the bound angle during this time. The 100-160 ns segment was used to calculate the unbound state properties. Even though the angle stays close to the unbound value most of the time, only the 60 ns (100-160ns) was used for consistency since that is about how long the structure stays in the bound conformation.

The water probability distributions (g_{iw}) were calculated using the number of water molecules assigned to each residue based on their proximity. Here, the center of mass of a water molecule was used to determine the proximity. Normalization was performed by dividing the number of water molecules in each bin for a residue by the number of water molecules that can be found in bulk water within the same bin. For the bulk water analysis, the coordinates of the protein at a given time were superimposed on a bulk water configuration and the number of water molecules is counted in the same manner as for the protein solution.

5.3 Results and Discussion

The time history of the Cα RMSDs calculated for each subdomain of the WT at 1 bar and 3 kbar are shown in Figure 5.5a and Figure 5.5b. RMSD values of 0.1 nm or less were observed for each subdomain indicating that they are stable during the simulation. These results agree with the previous simulation data.[6] The C-subdomains have relatively small RMSD values compared to N-subdomains indicating that the C-subdomains are slightly more stable. Experimentally, it was shown also that the N subdomains can move relative to the C-subdomains.[11-12] Similar behavior is observed for all the domains at 3 kbar with low fluctuations. Figure 5.5c and Figure 5.5d show the Cα RMSDs calculated for N-subdomains of WT after a translational and rotational fit to the Cα atoms in the C-subdomains of the bound and unbound crystal structures. Since the C-subdomains are stable we could see how the N-subdomains move relative to the C-subdomains. There is a transition around 60 ns where the two RMSDs cross each other (Figure 5.5c). However, we could not conclude that the structure moves to the unbound state since the RMSD after fitting to the DNA unbound structure is around 0.2 nm, even though it is relatively small compared to the one

fitted to the DNA bound structure. At 3 kbar, the RMSD after fitting to the bound structure is slightly lower than the RMSD after fitting to the unbound structure.

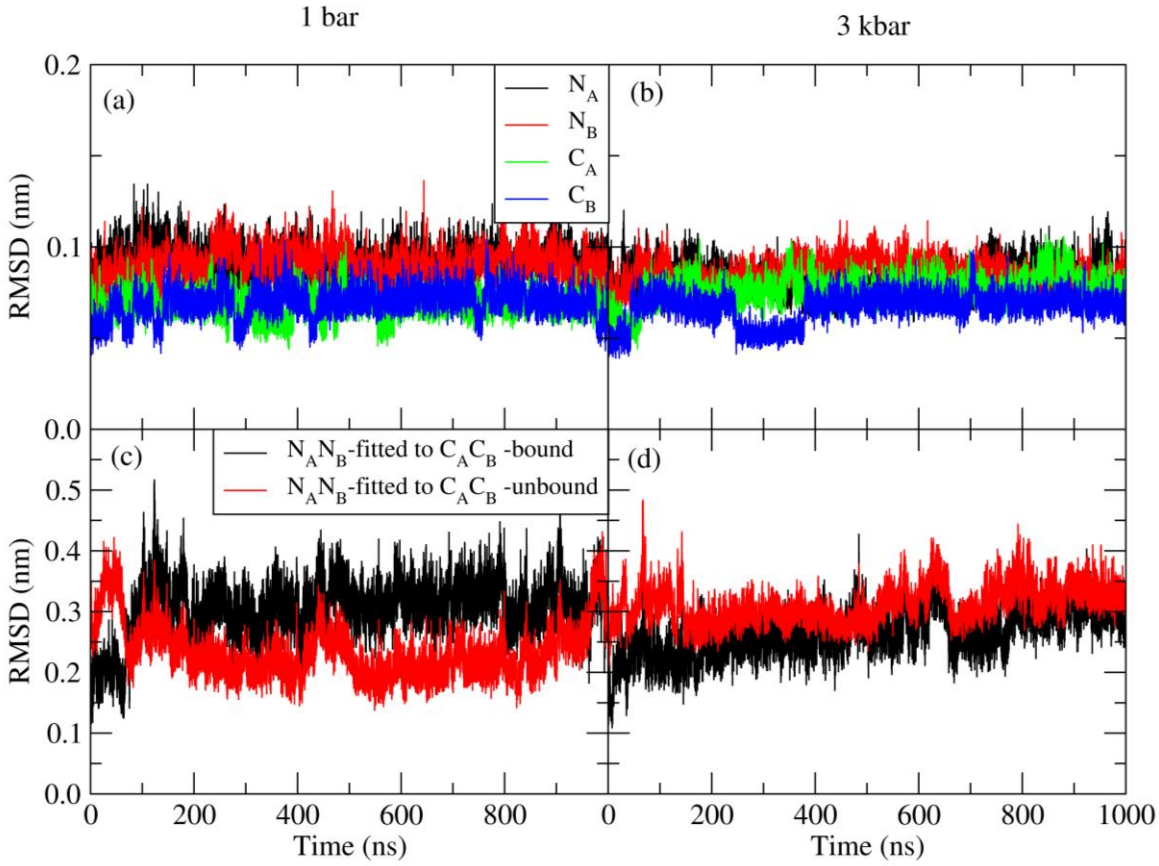


Figure 5.5 Calculated RMSDs for the WT (a) RMSDs calculated for each individual subdomain after fitting to the same domain of the crystal structure at 1 bar (b) RMSDs for each individual subdomain at 3 kbar (c) RMSDs of N - subdomains after fitting to the C - subdomains of bound crystal structure (black) and unbound crystal structure (red) at 1bar (d) RMSDs of N - subdomains after fitting to the C- subdomains of bound (black) and unbound (red) crystal structures at 3 kbar.

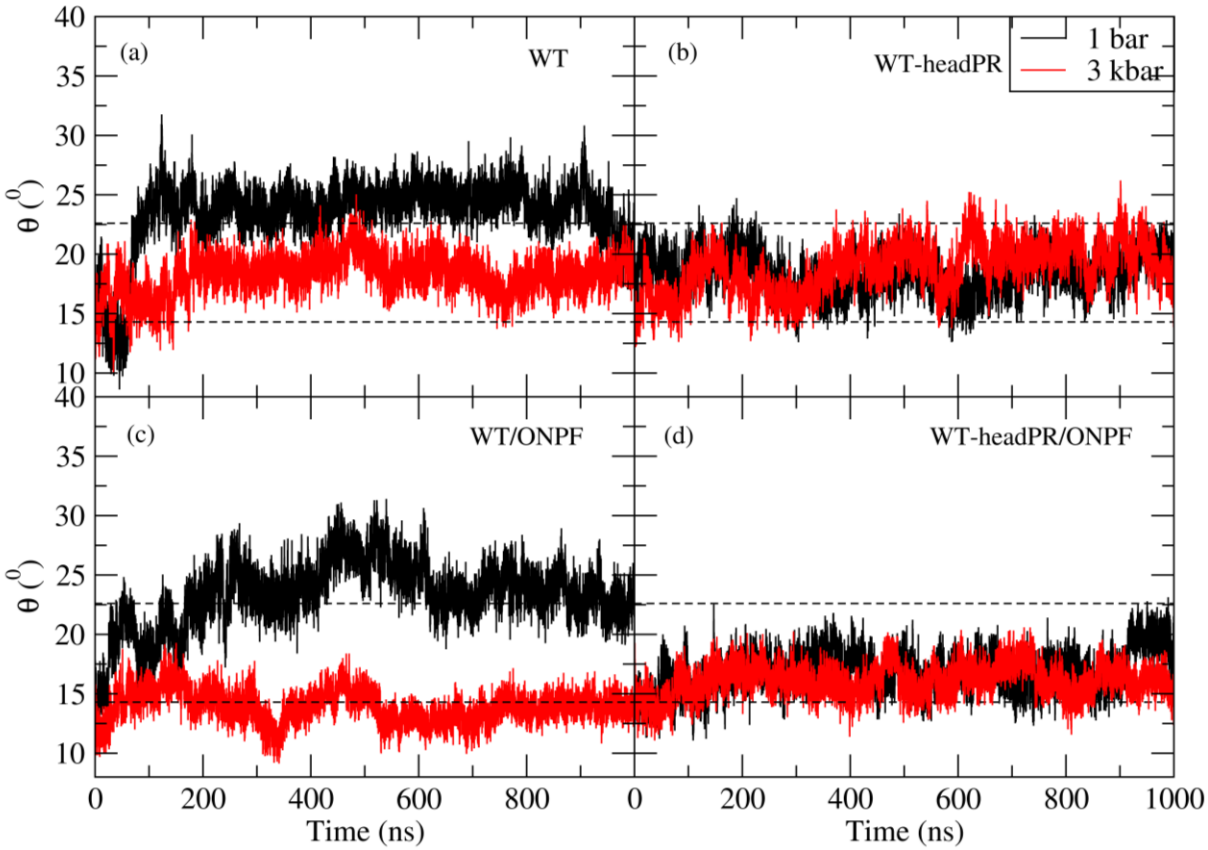


Figure 5.6 Calculated vector angle between chain A and chain B of LacI dimer. (a) angle for WT at 1 bar (black) and 3 kbar (red) (b) angle for WT- headPR at 1 bar (black) and 3 kbar (red) (c) angle for WT/ONPF at 1 bar (black) and 3 kbar (red) (d) angle for WT-headPR/ONPF at 1 bar (black) and 3 kbar (red). Dashed lines represent the angle calculated for bound (lower dashed line) and unbound (upper dashed line) crystal structures.

Figure 5.6 shows the angle vector calculated between chain A and chain B for systems 1-4 as a function of time. Under normal pressure, the WT angle moves from DNA-bound to DNA-unbound state around 60 ns. After that, the angle stays close to the unbound state angle (23^{o}) for almost all the time up to 1μs suggesting that the WT prefers the unbound state in the absence of DNA and ONPF at 1 bar. Nevertheless, at the end of the simulation we could see a drop in the angle. To make sure that the structure did not move to the bound state, the simulation at 1 bar was extended for another 100 ns. We observed that the angle shifted back to the unbound state during

that extra 100 ns simulation. The WT conformation does not move to the unbound state at 3 kbar. This suggests that pressure can control the transition of the WT from the DNA-bound to the DNA-unbound state. As pressure probes the hydration of proteins, we could argue that the bound state of the WT is more hydrated than the unbound state. Therefore, the volume of the bound state should be smaller than the volume of the unbound state.

The head group of the LacI is very flexible when the DNA is removed from the DBD. Therefore, we restrained the DBD to examine how it affects the motion of the core domains, as it then mimics the presence of DNA. The angle data for the WT-headPR simulation is shown in Figure 5.6b and demonstrates that the position restraints on the head group suppress the transition from the DNA-bound to the DNA-unbound state at 1 bar. Furthermore, this suppression is more effective at 3 kbar. Figure 5.6c shows the calculated angle for the WT/ONPF and it depicts the structural transition from the DNA-bound state to the unbound state is slower compared to the WT in the presence of ONPF at 1 bar. Furthermore, the structure remains in the bound state more effectively than the WT and WT-headPR at 3 kbar. In the presence of ONPF, and with PR, the WT-headPR/ONPF (Figure 5.6d) stays in the bound state even at 1 bar. Overall, this data indicates that there is an equilibrium between the DNA-bound and unbound states which is sensitive to pressure. Moreover, PR on the head group suppress the transition from the DNA-bound to unbound state, while with ONPF this is even more effective. These results agree with the experimental data since the PR head group mimics the DNA bound head group, and the role of the ONPF is to strengthen the DNA binding affinity.

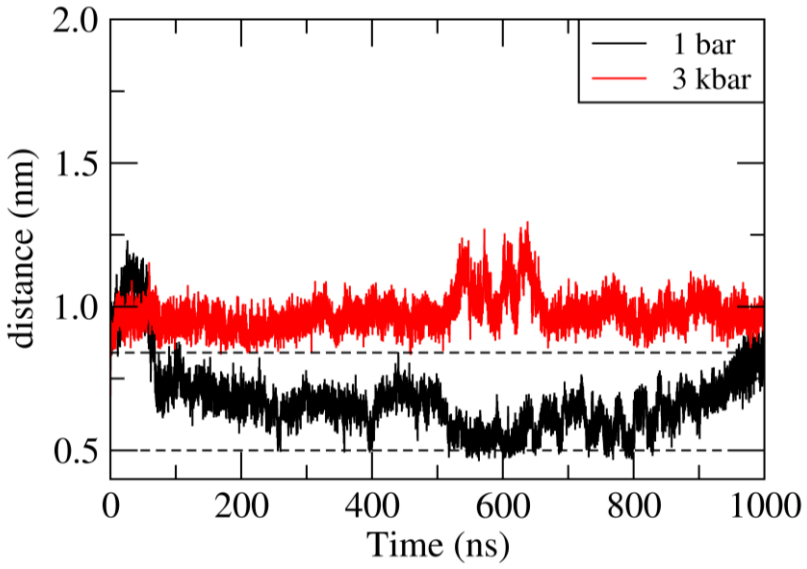


Figure 5.7 Distance between the C_α atoms of residue 96 in chain A and chain B of WT at 1 bar (black) and 3 kbar (red). Dashed lines represent the distance obtained from the bound (upper dashed line) and unbound (lower dashed line) crystal structures.

The distance between the two β-strands, indicated in the analysis section, is shown in Figure 5.7 for the WT simulations. These two β-strands are at the interface between the monomers in the N- subdomains. As observed for the WT angle motion, the distance between two β-strands moves from the DNA-bound to the DNA-unbound state. Furthermore, a higher pressure keeps the structure in the bound state for the whole simulation time. These results again suggest the presence of two main states and their sensitivity to pressure. Also, from these results, we could argue that the dimer does not fall apart at 3 kbar pressure as the conformation (angle and the distance) only moves between the DNA-bound and the DNA-unbound states.

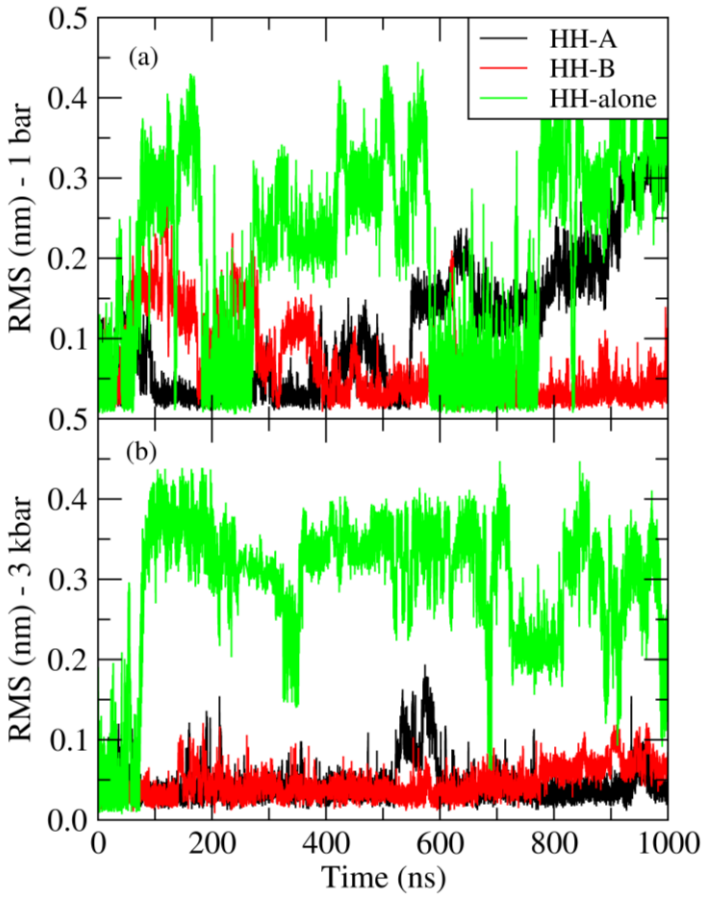


Figure 5.8 RMSD from an ideal helix for HH in chain A (black) and chain B (red) of WT, and for HH alone (green) (a) at 1 bar (b) at 3 kbar.

Figure 5.8 shows the RMSD from an ideal helix for the HHs in the WT and the isolated HH in solution. RMSDs were calculated for the HH in chain A and chain B of the WT. The HH in chain A is stable for about 500 ns and then it becomes unfolded at 1 bar. The HH in chain B unfolded more quickly around 100 ns, and then becomes folded again around 400 ns, and trapped in helical conformation for rest of the simulation. Both the HHs within the WT conformation seem stable at 3 kbar, indicating that pressure stabilizes the HHs. The HH alone is unstable at both pressures with large RMS values. Nevertheless, we could see that the HH alone does return to the helical formation and stays some time, after unfolding at 1 bar. The helix alone is fully disordered

at 3 kbar. These results suggest that the stability of HH depends on its environment and pressure only stabilizes the HH within the LacI environment.

Table 5.2 Volume differences between the bound and unbound states ($\Delta V_{unbound-bound}$) for the WT residues at 1 bar. Residues in both chains are included.

Criteria	No: of residues	Sum of $\Delta V_{unbound-bound}$ (nm^3)
Residues in the HHs	18	0.03
Residues within 5 Å from V52	22	0.06
Residues within 10 Å from V52	57	0.05
All residues	656	0.05

The residue volumes were calculated for WT at 1 bar and the sum of volume differences between bound and unbound states ($V_{unbound}$-V_{bound}) for selected residues are listed in Table 5.2. The summation of volume differences between the two states were positive for the residues in HHs, and within 5 Å and 10 Å from V52. Furthermore, the difference was also positive for the volumes of all 656 residues. Positive volume differences in all cases indicate that the volume of the unbound state is slightly higher than the volume of the bound state. According to our volume calculations using equation (5.2), there should be fewer water molecules around the unbound state compared to the bound state for a positive volume change. Therefore, we could argue that the bound state of the WT is more hydrated than the unbound state at 1 bar. However, these volume changes are very small and less than the volume of two water molecules - the volume of a TIP3P water molecule is 0.03 nm^3 at 300 K and 1 bar. Hence, we cannot come to a strong conclusion based on these volume calculations as this is a very large protein it is not easy to figure out which

regions of the protein are mainly affected. Furthermore, since the WT simulation mainly samples the unbound state at 1 bar, we do not have good statistics for the bound state.

Figure 5.9 shows the probability of finding a water molecule (local distribution functions, g_{iw}) around a residue for residues which are within 5 Å from V52. Here, V52 was chosen as our central residue since the 52^{nd} position of LacI has been identified as a rheostat position. Even though we did not observe a significant change in volume when considering selected sets of residues, the g_{iw} calculated for each residue indicate that there are differences in hydration between the bound and unbound states for some residues, especially for the residues 50-53. However, we do not see a significant difference in hydration for the bound and the unbound states for the residue 51 in chain B. The residues 50-53 have noisy g_{iw} curves for the unbound state indicating that there is less water around these residues in the unbound state. Other residues within 5 Å from the residue 52 do not show significant changes in hydration between two states.

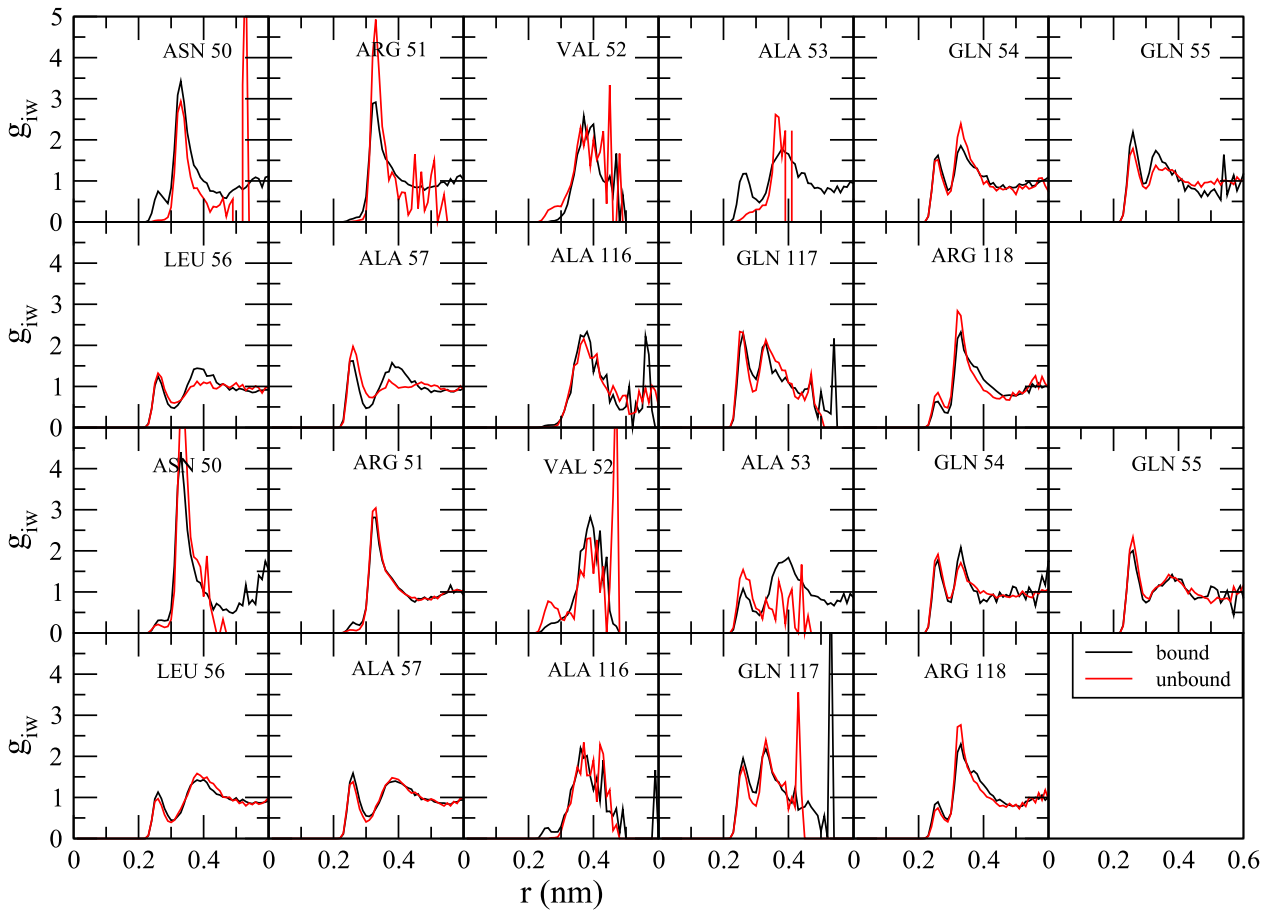


Figure 5.9 Water probability distribution around selected residues (g_{iw}) as function of distance in the bound (black) and the unbound (red) state of the WT. Top two rows of the figure show the residues in chain A, while the bottom two rows show the corresponding residues in chain B.

Figure 5.10 shows the average number of water molecules found within 5 Å from V52 and Ala 53 (A53) residues at 1 bar and 3 kbar. Since there is a significant difference in hydration for V52 and A53 between two states they were chosen to investigate how the number of water molecules changes as a function of time. First 100 ns segment was used for this analysis since the bound to unbound transition happens within the first 100 ns. As expected, the number of water molecules around A53 decreases with time indicating that the bound state is more hydrated at 1 bar. However, we do not see a significant change for V52. The number of water molecules around

both residues do not change at 3 kbar, since the structure stays in the bound form at this pressure as shown in the angle calculations (Figure 5.6).

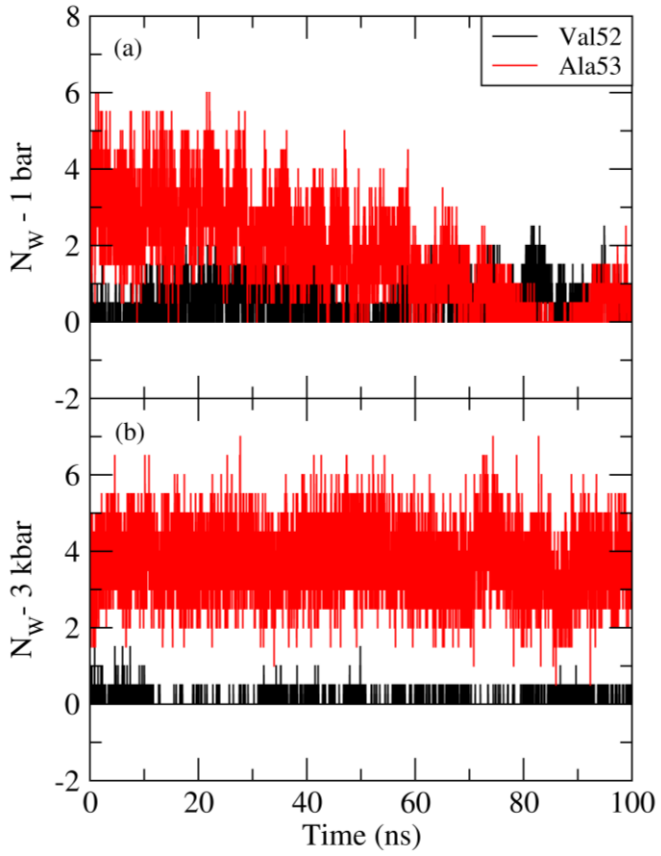


Figure 5.10 Average number of water molecules within 5 Å of Val 52 and Ala 53 (a) at 1 bar (b) at 3 kbar as a function of time for the WT.

5.4 Conclusions

Wild type *lac* repressor protein with and without anti-inducer molecules, and with and without head group position restraints, were simulated at 1 bar and 3 kbar to understand the structure and the dynamics of *lac* repressor. RMSD calculations for the N and C subdomains of each chain depict that each domain was stable, however, the N - subdomains can move relative to the C - subdomains. The wild type LacI dimer undergo as a conformational transition from a DNA-

bound to a DNA-unbound structure at 1bar. At higher pressure (3 kbar) this transition can be suppressed. Moreover, we observed that the PR head group also suppress the structural transition from the DNA-bound to the DNA-unbound state while the anti-inducer, ONPF, makes this even more effective. Interestingly, we observed that pressure could stabilize the HHs within the LacI. However, the simulations of isolated HH in solution indicated that the stability of HH depends on its environment, and pressure could only stabilize the HH within the LacI environment. Thus, the environment of the *lac* repressor plays a major role in the stability of the HHs. Considering the residues around V52, it appears that the bound and the unbound state of residues 50-53 have significantly different hydration properties, indicating that they are different in volumes.

Overall, we can suggest that there is an equilibrium between the DNA bound and unbound structures which can be governed by pressure. As the pressure probes changes in protein volume/hydration, this suggests that the structural transition observed here may have a strong hydration component.

5.5 Supporting information

Force field parameters for the anti-inducer molecule, ONPF were not available and therefore, the partial atomic charges for this molecule were obtained from the R.E.D. server website using the quantum mechanics software GAMESS. One bond and several angle parameters needed for ONPF were not included in AMBER14SB FF and are listed in Table S 5.1. The ONPF molecule with atom number labels are shown in Figure S 5.1.The partial atomic charges obtained from the R.E.D server are listed in Table S 5.2. The equilibrium bond length and equilibrium bond angles were measured from the ONPF crystal structure 1efa. The force constants were adopted from nitro compounds in OPLS all-atom FF[48]. Two improper angles were also not defined in

AMBER14SB FF, and the force constants for these two improper angles were chosen from available improper angle parameters in AMBER14SB based on atom types. For the two improper angles, multiplicity of 2 was used.

Table S 5.1 Bonded parameters used for ONPF. Parameters which were not available in AMBER14SB are shown here.

Bond/Angle	r_0 (Å) / θ_0 (deg)	Force constant
N-O	0.123	230120.0 $kJmol^{-1}nm^{-2}$
C-CA-N2	120.0	355.64 $kJmol^{-1}rad^{-2}$
CA-CA-N2	120.0	355.64 $kJmol^{-1}rad^{-2}$
CA-N2-O	116.5	334.72 $kJmol^{-1}rad^{-2}$
O-N2-O	121.0	334.72 $kJmol^{-1}rad^{-2}$
CA-O-N2-O	180.0	4.602 $kJmol^{-1}$
C-CA-CA-N2	180.0	43.932 $kJmol^{-1}$

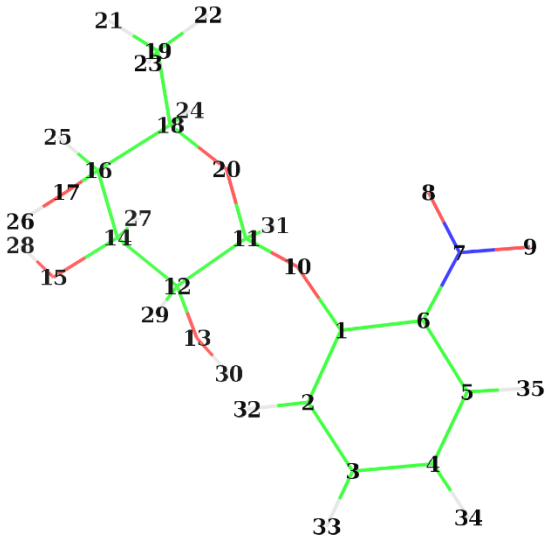


Figure S 5.1 ONPF molecule labeled with atom number.

Table S 5.2 Partial atomic charges of the ONPF obtained using R.E.D server.

No:	atom	atom type	charge
1	C1	C	0.1468
2	C2	CA	-0.0815
3	C3	CA	-0.1814
4	C4	CA	-0.1117
5	C5	CA	-0.2195
6	C6	CA	0.0277
7	N6	N2	0.7927
8	O6A	O	-0.4641
9	O6B	O	-0.4641
10	O1'	OS	-0.241
11	C1'	CT	0.0404
12	C2'	CT	0.119
13	O2'	OH	-0.5701
14	C3'	CT	0.127
15	O3'	OH	-0.6225
16	C4'	CT	0.0484
17	O4'	OH	-0.6292
18	C5'	CT	0.1631
19	C6'	CT	-0.2001
20	O5'	OS	-0.3059
21	H1	HC	0.0677
22	H2	HC	0.0677
23	H3	HC	0.0677
24	H4	H1	0.0376
25	H5	H1	0.0772
26	H6	HO	0.4418
27	H7	H1	0.0512
28	H8	HO	0.4385
29	H9	H1	0.1586
30	H10	HO	0.4121
31	H11	H2	0.181
32	H12	HA	0.1353
33	H13	HA	0.1602
34	H14	HA	0.1486
35	H15	HA	0.1808

5.6 References

1. Swint-Kruse, L.; Matthews, K. S., Current opinion in microbiology 2009, 12 (2), 129-137.

2. Swint-Kruse, L.; Zhan, H.; Fairbanks, B. M.; Maheshwari, A.; Matthews, K. S., Biochemistry 2003, 42 (47), 14004-14016.

3. Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J., Nature 2014, 508 (7496), 331-339.

4. Capdevila, D. A.; Braymer, J. J.; Edmonds, K. A.; Wu, H.; Giedroc, D. P., Proceedings of the National Academy of Sciences 2017, 201620665.

5. Flynn, T. C.; Swint-Kruse, L.; Kong, Y.; Booth, C.; Matthews, K. S.; Ma, J., Protein Science 2003, 12 (11), 2523-2541.

6. Villa, E.; Balaeff, A.; Schulten, K., Proceedings of the National Academy of Sciences 2005, 102 (19), 6783-6788.

7. Farabaugh, P. J., Nature 1978, 274 (5673), 765.

8. Platt, T.; Files, J. G.; Weber, K., Journal of Biological Chemistry 1973, 248 (1), 110-121.

9. Jobe, A.; Bourgeois, S., Journal of molecular biology 1972, 69 (3), 397-408.

10. Riggs, A. D.; Suzuki, H.; Bourgeois, S., Journal of molecular biology 1970, 48 (1), 67-83.

11. Bell, C. E.; Lewis, M., Nature Structural and Molecular Biology 2000, 7 (3), 209.

12. Lewis, M.; Chang, G.; Horton, N. C.; Kercher, M. A.; Pace, H. C.; Schumacher, M. A.; Brennan, R. G.; Lu, P., Science 1996, 271 (5253), 1247-1254.

13. DeLano, W. L., CCP4 Newsletter On Protein Crystallography 2002, 40 (1), 82-92.

14. Spronk, C. A.; Slijper, M.; van Boom, J. H.; Kaptein, R.; Boelens, R., Nature Structural and Molecular Biology 1996, 3 (11), 916.

15. Frank, D. E.; Saecker, R. M.; Bond, J. P.; Capp, M. W.; Tsodikov, O. V.; Melcher, S. E.; Levandoski, M. M.; Record Jr, M. T., Journal of molecular biology 1997, 267 (5), 1186-1206.

16. Nagadoi, A.; Morikawa, S.; Nakamura, H.; Enari, M.; Kobayashi, K.; Yamamoto, H.; Sampei, G.; Mizobuchi, K.; Schumacher, M.; Brennan, R., Structure 1995, 3 (11), 1217-1224.

17. Meinhardt, S.; Manley Jr, M. W.; Parente, D. J.; Swint-Kruse, L., PloS one 2013, 8 (12), e83502.

18. Swint-Kruse, L.; Matthews, K. S.; Smith, P. E.; Pettitt, B. M., Biophysical journal 1998, 74 (1), 413-421.

19. Sun, L.; Tabaka, M.; Hou, S.; Li, L.; Burdzy, K.; Aksimentiev, A.; Maffeo, C.; Zhang, X.; Holyst, R., PloS one 2016, 11 (3), e0152002.

20. Seckfort, D.; Montgomery Pettitt, B., Biopolymers, e23239.

21. Caro, J. A.; Harpole, K. W.; Kasinath, V.; Lim, J.; Granja, J.; Valentine, K. G.; Sharp, K. A.; Wand, A. J., Proceedings of the National Academy of Sciences 2017, 201621154.

22. Maciag, J. J.; Mackenzie, S. H.; Tucker, M. B.; Schipper, J. L.; Swartz, P.; Clark, A. C., Proceedings of the National Academy of Sciences 2016, 113 (41), E6080-E6088.

23. Levy, Y.; Onuchic, J. N., Annu. Rev. Biophys. Biomol. Struct. 2006, 35, 389-415.

24. Salvay, A. G.; Grigera, J. R.; Colombo, M. F., Biophysical Journal 2003, 84 (1), 564-570.

25. Davydov, D. R.; Baas, B. J.; Sligar, S. G.; Halpert, J. R., Biochemistry 2007, 46 (26), 7852-7864.

26. Royer, C. A.; Weber, G.; Daly, T. J.; Matthews, K. S., Biochemistry 1986, 25 (25), 8308-8315.

27. Royer, C., Brazilian journal of medical and biological research 2005, 38 (8), 1167-1173.

28. Royer, C. A.; Chakerian, A. E.; Matthews, K. S., Biochemistry 1990, 29 (20), 4959-4966.

29. Friedman, A. M.; Fischmann, T. O.; Steitz, T. A., Science 1995, 268 (5218), 1721-1727.

30. Daly, T. J.; Matthews, K. S., Biochemistry 1986, 25 (19), 5474-5478.

31. Chen, J.; Matthews, K., Journal of Biological Chemistry 1992, 267 (20), 13843-13850.

32. Matthews, K. S.; Nichols, J. C., Lactose repressor protein: functional properties and structure. In Progress in nucleic acid research and molecular biology, Elsevier: 1997; Vol. 58, pp 127-164.

33. Swint-Kruse, L.; Zhan, H.; Matthews, K. S., Biochemistry 2005, 44 (33), 11201-11213.

34. Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F.-Y., Nucleic Acids Research 2011, 39 (suppl_2), W511-W517.

35. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., Journal of chemical theory and computation 2015, 11 (8), 3696-3713.

36. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., Bioinformatics 2013, 29 (7), 845-854.

37. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Proteins: Structure, Function, and Bioinformatics 2006, 65 (3), 712-725.

38. Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G., Journal of computational chemistry 1997, 18 (12), 1463-1472.

39. Miyamoto, S.; Kollman, P. A., Journal of computational chemistry 1992, 13 (8), 952-962.

40. Hockney, R.; Goel, S.; Eastwood, J., Journal of Computational Physics 1974, 14 (2), 148-158.

41. Darden, T.; York, D.; Pedersen, L., The Journal of chemical physics 1993, 98 (12), 10089-10092.

42. Parrinello, M.; Rahman, A., Journal of Applied physics 1981, 52 (12), 7182-7190.

43. Nosé, S.; Klein, M., Molecular Physics 1983, 50 (5), 1055-1076.

44. Hoover, W. G., Physical review A 1985, 31 (3), 1695.

45. Nosé, S., Molecular physics 1984, 52 (2), 255-268.

46. Canchi, D. R.; García, A. E., Annual Review of Physical Chemistry 2013, 64 (1), 273-293.

47. Imai, T.; Harano, Y.; Kovalenko, A.; Hirata, F., Biopolymers 2001, 59 (7), 512-519.

48. Price, M. L.; Ostrovsky, D.; Jorgensen, W. L., Journal of Computational Chemistry 2001, 22 (13), 1340-1352.

Chapter 6 - Probing Hydration Changes in *Lac* Repressor Mutants

6.1 Introduction

Every person has unique variations in their genetic profile and these variations can make susceptible them to have certain diseases.[1] Therefore, it is important to identify individual genome profiles to treat these diseases. Personalized medicine is an effective way of treating patients with different genetic profiles as the treatments are tailored according to individual characteristics. However, even when the differences in genomes have been identified, the significance of these differences are complicated.[2] Many bioinformatics programs and algorithms have been developed to help identify the medically relevant or functionally important amino acid positions.[3-5] Many of these analyses involve the LacI/GalR (lactose repressor protein/galactose repressor protein) family of genes/proteins due to the availability of experimental data that can be used to validate any results.[5-6]

Many experimental studies have been performed involving mutations of highly conserved residues in proteins – those that do not change during evolution – as mutations at these positions generally lead to destabilization of the structure or changes in the function of a protein.[7-8] Mutations at highly conserved positions commonly act as a 'toggle' (on-off) switch, where conservative variants act similar to the parent protein, while other non-conserved mutants abolish function.[7] Non-conserved positions, that often change during the evolution, are not widely studied as they are assumed to cause small effects. However, some non-conserved positions are also known to have an important effect on protein activity.[7, 9] Interestingly, substitutions at these positions can display 'neutral' or 'rheostat' behavior. The positions where the amino acids

substitution show similar wild-type function are known as ‘neutral’ positions. The positions that show progressive functional changes upon substitution are known as ‘rheostat’ positions, and these positions are mainly observed to be in the non-conserved positions.[7]

By convention, substitution of similar amino acids typically allow normal protein function, while most other substitutions abolish or change function.[10] Mutations at conserved positions follow this substitution rule, and this rationale is commonly used to predict the effects of different variants. However, in contrast to the function of mutants at conserved positions, the functions of mutants at rheostat positions do not correlate with evolutionary frequency, or side chain similarities.[7]

As discussed in Chapter 5, the *lac* operon is responsible for the uptake and metabolism of lactose in *E.coli*. The function of the lactose repressor (LacI) is the allosteric regulation of DNA to modulate transcription.[11] Once the LacI bounds to the operator DNA in the *lac* operon, it prevents the transcription of downstream genes. If the LacI binds to the ligand allolactose,[12] or the gratuitous inducer isopropyl-D-thiogalactoside (IPTG)[13], then the LacI loses its high affinity to the operator and transcription is allowed.[12, 14] In addition, the anti-inducer molecule, orthonitrophenyl-β-D-fucoside (ONPF),[15] increases the affinity of the LacI for the operator. However, it has no known regulatory function in E. coli.[16]

There are two conformations of the LacI protein, representing the DNA bound and the DNA unbound forms involved in the function of repression and induction.[17] Figure 6.1 shows the DNA bound LacI structure (PDB id 1EFA).[16] The LacI protein has five structural units. The first 1-49 residues contain the helix-turn-helix that binds to the major groove of the DNA. Residues 50-58 belong to the ‘hinge helix’ (HH) which interacts with the center of the operator in the minor groove.[16] In the absence of DNA, the HH is unstable[18] and residues 1-61 (head group) can move

freely relative to the core domain (residues 62-333).[16] The HH connects the DNA binding domain (DBD) to the core domain (regulatory domains). The 18 amino acids (including HH) that link the DNA-binding and the core domain is known as the linker. The core domain consists of two subdomains, an N-subdomain (residues 62-161 and 293-320) and a C-subdomain (residues 162-289 and 321-329).[16] The anti-inducer molecule binds to the pocket in between the N and C subdomains. The C-terminal residues 340-357 facilitate the tetramerization of LacI.[16]

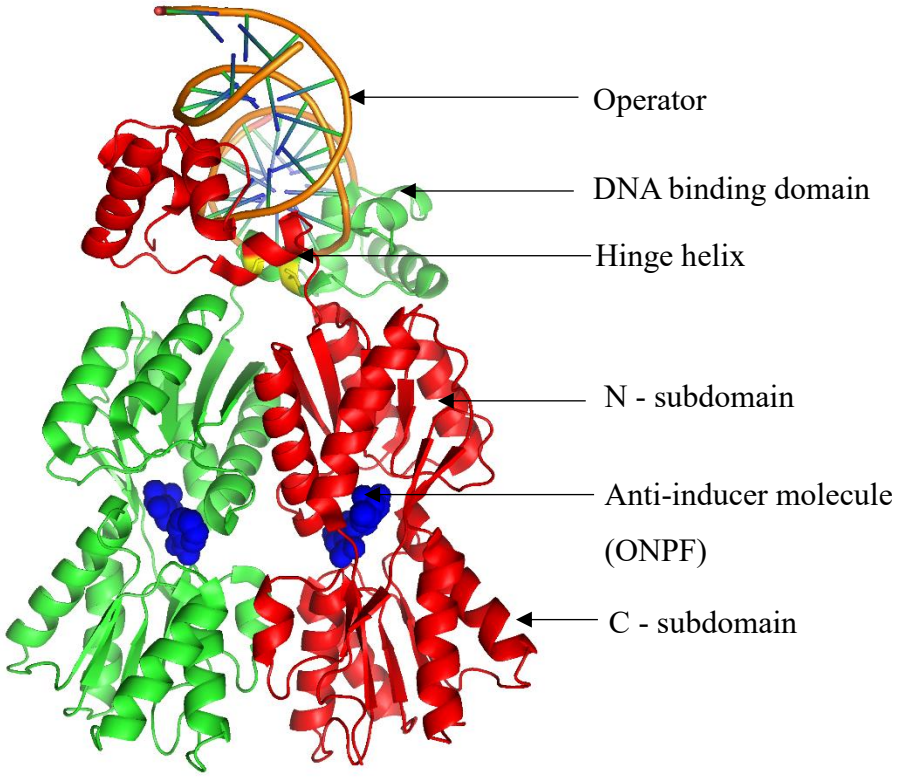


Figure 6.1 Cartoon representation of the LacI dimer (PDB id: 1EFA) bound to operator and anti-inducer, ONPF (blue). Chain A of the dimer is shown in green and chain B in red. C-terminal residues are not shown here. Residue 52 in both chains are colored in yellow. The image was generated with PyMOL molecular visualization software.[19]

A protein conformation change is required for the allosteric transition to occur. Experimental x-ray data have shown that the conformational transition involves the reorientation

of the N-subdomains while keeping the C-subdomains fixed during the transition.[18] Furthermore, the conformational change involves a helix to coil transition in the HH.[20]

Flynn and coworkers have simulated the DNA-bound LacI to the inducer-bound LacI transition pathway using target molecular dynamics (TMD).[17] TMD allows one to "switch" the initial structure to the target structure over a short period of simulation time (several hundred picoseconds). Only the core domains of the DNA-bound and inducer-bound dimers were used in their simulations with the ONPF and IPTG ligands. They have observed an asymmetric motion of the monomers during the simulations. Their simulation data suggest that the allosteric signal starts from the inducer binding site of one monomer and transmit to the other through many non-covalent interactions.

Experimental studies have also been performed on the non-conserved positions of LacI by mutating these positions with multiple amino acids.[2, 7, 21] For each variant, repression has been assayed using the activity of the β-galactosidase enzyme. Lower β-galactosidase activity indicates tighter transcription repression, while higher β-galactosidase activity indicates weaker transcription repression. Most of the non-conserved positions in the linker portion (residues 46-62) of the LacI are identified as rheostat positions following the mutation of these positions with multiple amino acids.[7, 22] The outcomes observed for the 52nd position of LacI are shown in Figure 6.2. It shows the β-galactosidase activity for 10 mutants in the presence and absence of the inducer molecule.

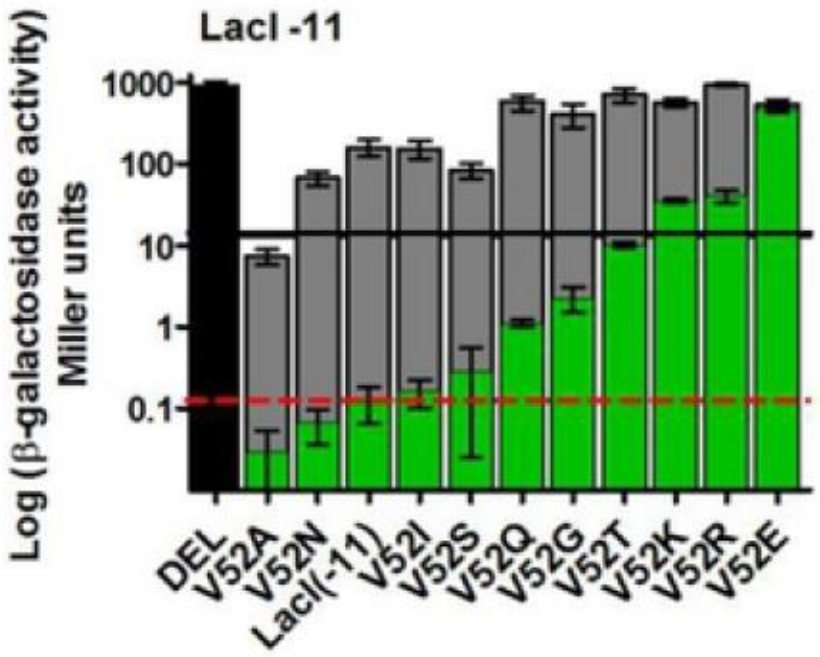


Figure 6.2 Experimental β-galactosidase activity for 52nd position of LacI protein. The first black bar represents the repression in the absence of LacI protein. Bellow 13 Miller units (black solid line), any change in repression altered bacterial growth. The red dashed line represents the activity for WT protein. The front series show the activity in the absence of effector molecule while the gray series shows the activity in the presence of effector. Error bars correspond to standard deviation of 2-4 bacterial colonies. The 'LacI-11' indicates that the LacI is just a dimer after deleting the C terminal tetramerization domain. This figure was reproduced from 'Meinhardt S, Manley MW Jr, Parente DJ, Swint-Kruse L (2013) Rheostats and Toggle Switches for Modulating Protein Function. PLoS ONE 8(12): e83502. https://doi.org/10.1371/journal.pone.0083502' This is an open-access article distributed under the terms of the Creative Commons Attribution License.

In this study, we are trying to investigate and understand some of the properties that may allow us to identify and rationalize the behavior of rheostat positions. The LacI protein is used as a model system since there exists plenty of experimental data characterizing the effects of mutations on this protein. In particular, molecular dynamics simulations at high pressure (3 kbar) are performed to investigate the role of hydration changes on the structural changes of multiple LacI mutants at an experimentally known rheostat position (position 52).

6.2 Methodology

6.2.1 System setup

The crystal structure with PBD code 1efa was used as the starting DNA bound structure after adding missing atoms, (using AMBER leap module), and removing the DNA and the c-terminal tetramerization domain. After modification, each chain of the dimer contains 328 residues. Since the crystal structure does not contain the 1st residue, only residues 2-329 in the crystal structure were used.

We selected four variants to perform the simulations based on the experimental data (Figure 6.2). The 52nd position of LacI (valine) was mutated with alanine (V52A), isoleucine (V52I), glutamine (V52Q), and glycine (V52G). According to the experimental data, V52A displays higher transcription repression, i.e. it binds more strongly to the operator DNA. V52I shows similar repression to the WT, while V52Q and V52G show a weaker affinity for DNA. Since the crystal structures are not available for the LacI variants, mutations were model with the PyMOL software.[19]

All the systems were solvated with TIP3P water in a rhombic dodecahedron simulation box where the distance between two parallel faces was 15 nm. Counterions were added (6 Na^+ ions in each system) to neutralize the systems. All the systems have the same charge since the selected mutants are neutral. All the simulations were performed in the absence of anti-inducer ligands.

6.2.2 Molecular Dynamics simulations

Atomistic MD simulations were performed using the AMBER14SB[23] force field and the GROMACS 2016 or 2016.4 versions.[24] All the simulations were performed in the isothermal

isobaric ensemble (NpT) at 300 K temperature, and both 1 bar and 3 kbar pressure. All the protein bonds were constrained using the LINCS algorithm,[25] and water with the Settle algorithm.[26] A time step of 2 fs was used to integrate the equations of motion with the Leap Frog algorithm.[27] Electrostatic interactions were determined using the particle mesh Ewald (PME)[28] technique with 0.12 nm Fourier grid spacing being used. The Verlet cut-off scheme was used with a 0.9 nm cut-off distance for van der Waals and electrostatic interactions. Periodic boundary conditions and the minimum image convention were applied to all the systems. The steepest descent algorithm was used for energy minimizations. The systems were then equilibrated up to 100 ps by gradually increasing the temperature (100, 200 and 300 K) with heavy atom position restraints (1000 kJ mol^{-1} nm^{-2}) and the Berendsen T-coupling and Berendsen isotropic P-coupling approaches. The constant temperature and pressure were achieved using Parrinello-Rahman pressure coupling,[29-30] and the Nose-Hoover temperature coupling, for the production simulations.[31-32] All the production simulations were continued up to 500 ns.

6.2.3 Analysis

The angle between the two vectors which connect the N-subdomain and C-subdomain of chain A and chain B is calculated for each system. More details on how the angle was calculated can be found in section 5.2.3. This calculation was performed as a crude measure to determine the dynamic motions of the domains in each variant. The measured angle for the 1efa (DNA bound) crystal structure was 23^{o}, while for the 1lbh (DNA unbound) structure it was 14^{o}.

As a second measurement to capture the domain motions, the distance between the C_{α} atoms of residue 96 in chain A and chain B was calculated. The measured distance for the 1efa

(DNA bound) crystal structure was 8 Å while that for the 1lbh (DNA unbound) structure was 5 Å. This measure roughly describes the separation of the two domains.

The probability distribution of water molecules (g_{iw}) around selected residues were determined for each variant at 1 bar. The first 10 ns of the production runs were used to calculate the bound state g_{iw} of all variants, since the calculated angle (θ) is close to the bound angle of crystal structure during this time. The 90-100 ns segments were used to calculate the unbound state g_{iw}.

6.3 Results and discussion

Figure 6.3 shows the angle calculated between chain A and chain B for the WT, and four variants, at 1 bar and 3 kbar. Considering the angle calculated at 1 bar, we can infer that the WT and all four variants move towards the unbound state easily as the DNA is removed in all five systems. By ramping up the pressure, the transition from bound to unbound state is slowed down, or stopped, depending on the system variant. It is clearly evident from this transition that the V52A mutant stays in the bound state during the whole simulation time at 3 kbar. This indicates that increased pressure has stabilized the bound state for V52A. This agrees with the experimental repression data, as the V52A shows tighter transcription repression, or strong DNA binding affinity compared to the other variants and the WT. For the WT, the calculated angle stays in between the reference bound and unbound angles. The V52I, V52Q, and V52G mutants display less response to pressure compared to the WT and V52A. According to the experimental results, the V52I variant is expected to behave like the WT. However, we do not observe similar behavior for these two systems. The V52Q and V52G show weaker transcription repression, or weaker DNA binding affinity, and our results agree with this observation since both these systems move to the DNA-

unbound state even at high pressure. It should be noted that the bound and unbound reference lines shown in figures are based on the WT crystal structure as the crystal structures for mutants are not available. Therefore, we cannot expect the mutants to have exactly the same bound and unbound angles as for the WT.

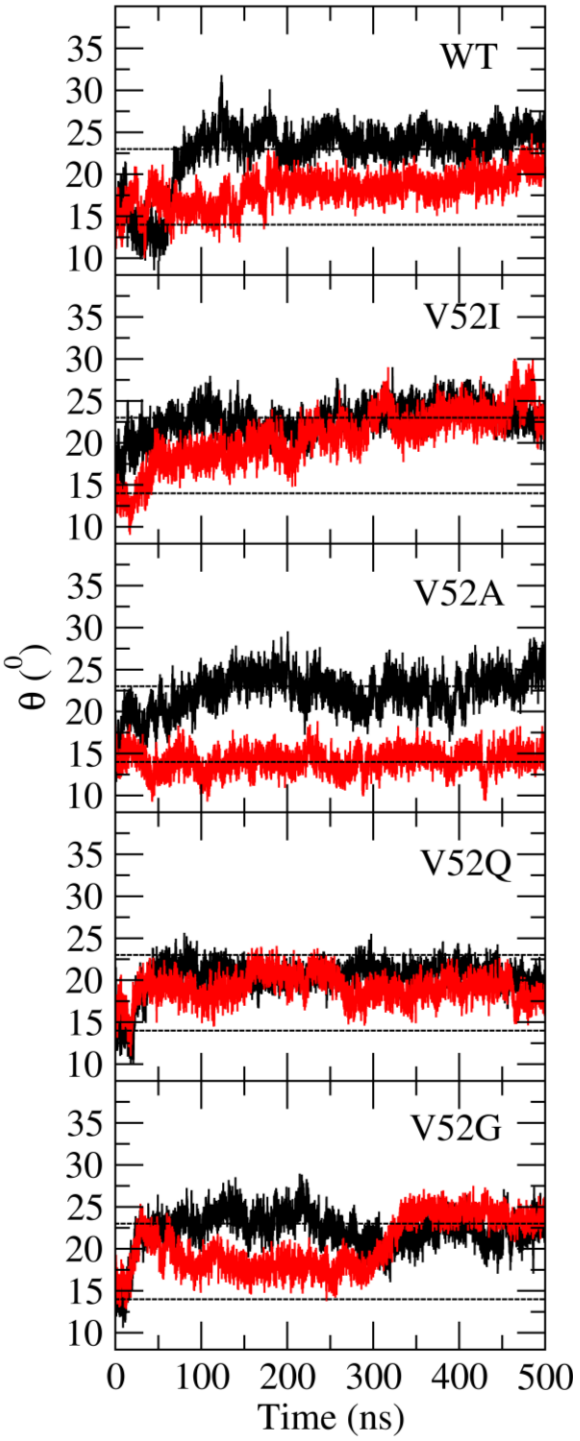


Figure 6.3 Calculated angle vector between chain A and chain B for WT, V52I, V52A, V52Q, and V52G LacI dimer. The angle at 1 bar is shown in black and at 3 kbar is shown in red. Dashed lines represent the angle calculated for the bound (lower dashed line) and unbound (upper dashed line) crystal structures.

As another measurement to distinguish between the bound and unbound states, the distance between the C-alpha atoms of residue 96 in chain A and chain B of N-subdomains were calculated and are shown in Figure 6.4. For the WT, the distance is close to the unbound reference distance, while for the V52I and V52Q mutants the distance lies in between the bound and unbound references at 1 bar. The distance is close to the bound reference distance for the V52G, contrary to what we observed for the simple angle calculations. Since the angle calculations suggested that the V52G mutant prefers the unbound state at both normal and higher pressures. It may be inferred that there are more than two possible states, and a single angle or distance measurement may not fully represent the complexity of these states.

A schematic representation of the distance vs angle for the WT and the four variants at 1 bar is shown in Figure 6.5. The bound and unbound states were determined according to the angle and distance measurements of the corresponding WT crystal structures. All the systems displayed calculated angles that are close to the unbound crystal structure angle values, although the distances do vary between the bound and unbound crystal structure distances. Considering both the measurements we can cluster the WT, V52I, and V52Q into the unbound state at 1 bar. However, it is difficult to relate the V52A and V52G to either the bound or unbound states at 1 bar.

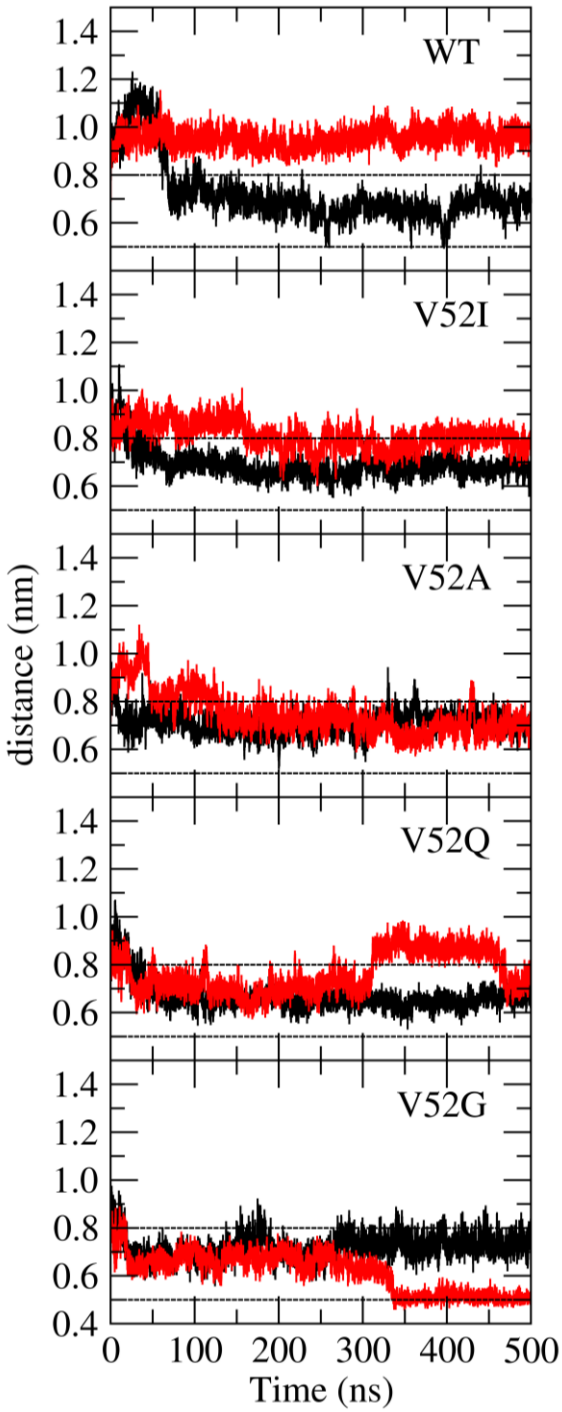


Figure 6.4 Distance between the C-alpha atoms of residue 96 in chain A and chain B of WT, V52I, V52A, V52Q, and V52G at 1 bar (black) and 3 kbar (red). Dashed lines represent the distance obtained from the bound (upper dashed line) and unbound (lower dashed line) crystal structures.

The corresponding data at 3 kbar indicates that the V52G mutant prefers the unbound state, while the V52A mutant prefers the bound state at higher pressure. Therefore, the hydration level of the bound state appears to be higher, and therefore the volume appears to be lower for the V52A mutant. We could not relate the WT, V52I, and V52Q mutants to the bound or unbound state at higher pressure. However, it must be kept in mind that the bound and unbound reference states correspond to the WT crystal structure, and this may change for the structure of the variants. Moreover, it is difficult to assign the conformations to the bound or unbound since the differences are very small for these two states. As mentioned in the previous chapter, the root-mean-square deviation (RMSD) between the N-subdomains after fitting to the lower C-subdomains is 0.2 nm for the bound and the unbound crystal structures.

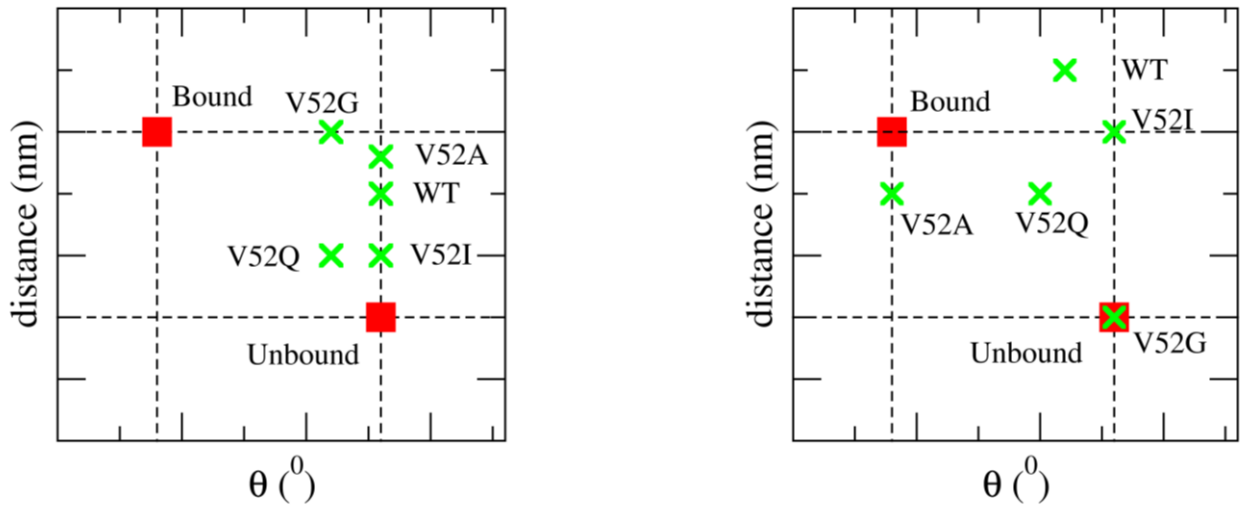


Figure 6.5 Schematic representation of distance vs angle measurements at 1 bar (left) and 3 kbar (right). Red squares represent the bound and unbound states determined according to the DNA bound and unbound crystal structures.

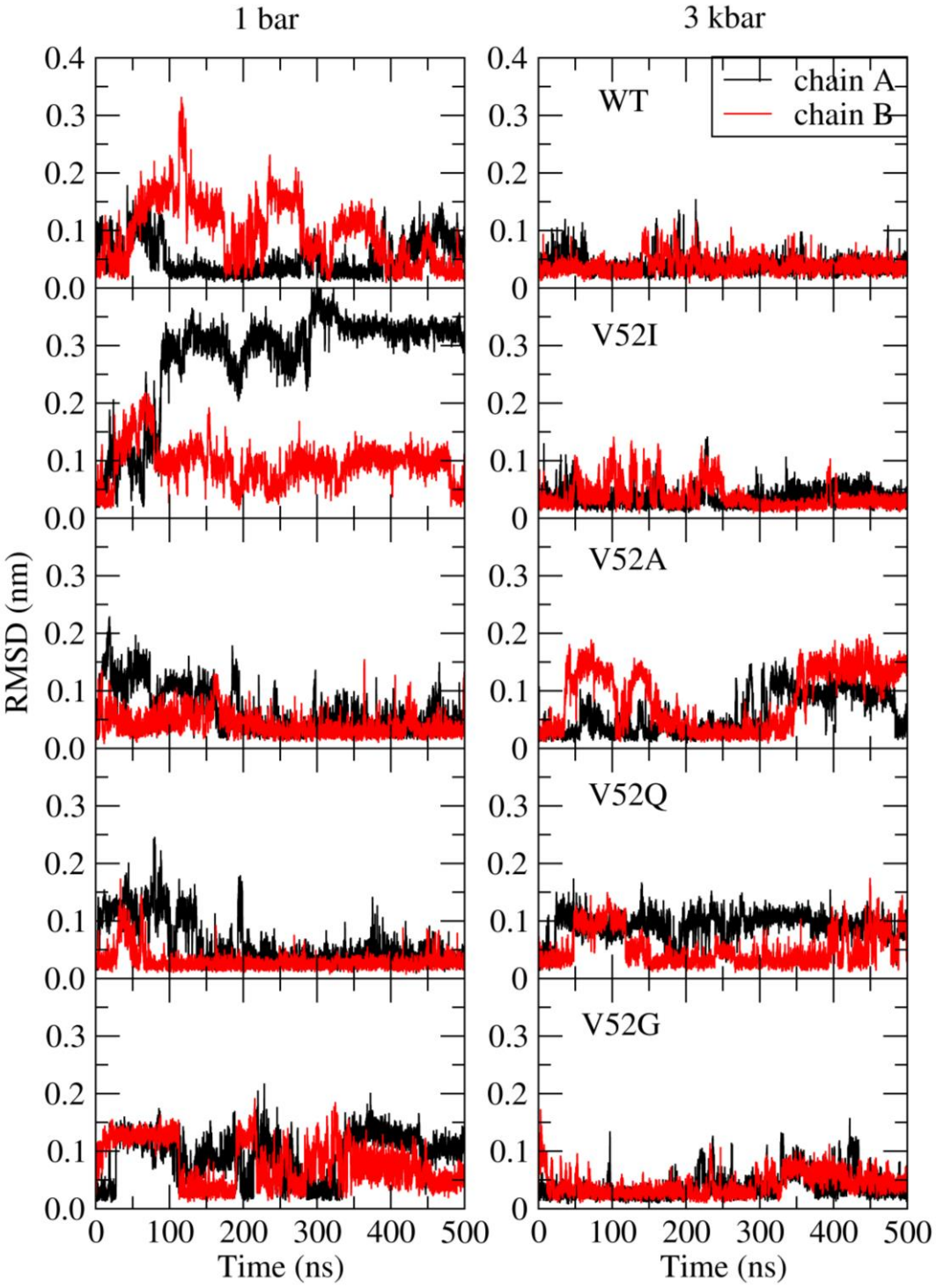


Figure 6.6 RMSD from an ideal helix for HH in chain A (black) and chain B (red) of WT, V52I, V52A, V52Q, and V52G at 1 bar (left column) and 3 kbar (right column).

Figure 6.6 shows the time histories for the RMSDs from an ideal helix for all the HHs. We can observe that the HHs display different stabilities among the variants under ambient conditions and in response to pressure. The RMSDs for the HHs of the V52A and V52Q mutants are very small, indicating that the HHs are stabilized upon mutating to alanine or glutamine at 1 bar. We observe that the HHs are stabilized under pressure for most of the systems except for V52A and V52Q. We expected the V52A mutant to display the most stable HHs at higher pressure since the

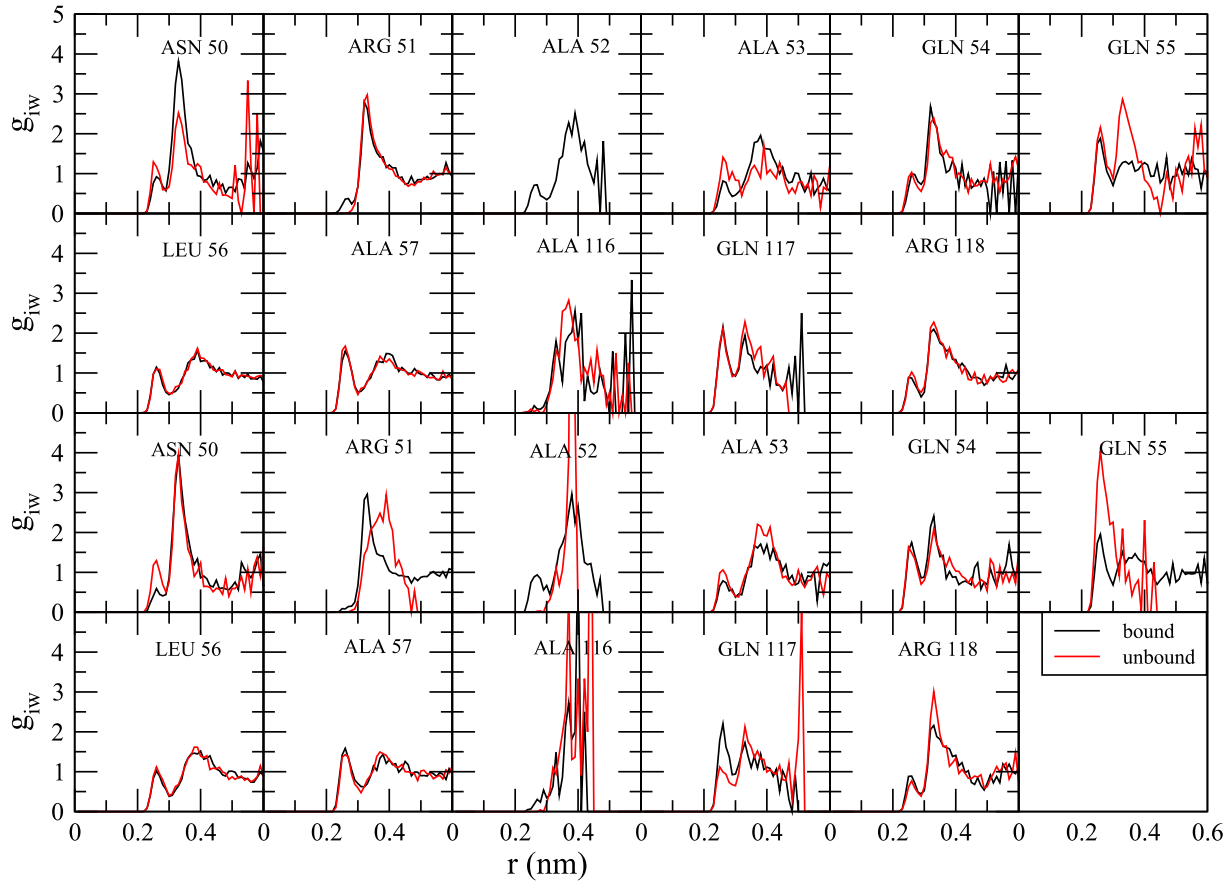


Figure 6.7 Water probability distribution around selected residues (g_{iw}) as a function of distance in the bound (black) and the unbound (red) state of V52A at 1 bar. Top two rows of the figure show the g_{iw} for residues in chain A, while the bottom two rows show the g_{iw} for the corresponding residues in chain B. The bound and unbound states were determined only considering the angle data.

experimental data indicate that it strongly binds to DNA, and the HHs are stable in the presence of DNA. However, the HHs of V52A are less stable at 3 kbar compared to that of 1 bar. Longer simulations might be needed to draw further conclusions from the HHs stabilities.

Figure 6.7 shows the probability of finding a water molecule around the residues that are within 5 Å of residue 52 of the V52A mutant. We notice that there was no water in the local vicinity of A52 in the unbound state. Additionally, the A53 and Q55 residues in both chains show a difference in hydration between the bound and unbound states. This suggests that the hydration of residues 52, 53 and 55 are most affected during the conformational transition. We observed that the residues 50-53 are most affected during the conformational transition for the WT (chapter 5-Figure 5.9) However, there was no significant difference in hydration for the Q55. The water probability distributions for the V52I, V52Q, and V52G mutants are shown in the appendix. However, for these calculations the bound and unbound states were defined by considering only the angle calculations as described in the analysis section. The angle and distance calculations give contradicting observations and, therefore, the angle calculations only may not give an accurate definition for the bound and unbound states.

6.4 Conclusions

Four variants of the LacI dimer, which were obtained by mutating the 52nd position of the LacI, were studied at 1 bar and 3 kbar using molecular dynamics simulations. The motion of the N-subdomains of each variant were captured by calculating the angle vector between the domains and the distance between the residue 96 in Chain A and chain B. Our results suggest that there can be intermediate states during the conformational transition from DNA bound to the DNA unbound state. The variations in binding affinity for the different variants can be somewhat explained using

the simulated hydration changes as we have observed different levels of hydration for these variants. However, we need to perform longer simulations and further analysis to correlate and rank the observed simulation data with the experimental results, and to identify the particular residues in each variant that are responsible for different binding affinities towards the DNA.

6.5 Supporting Information

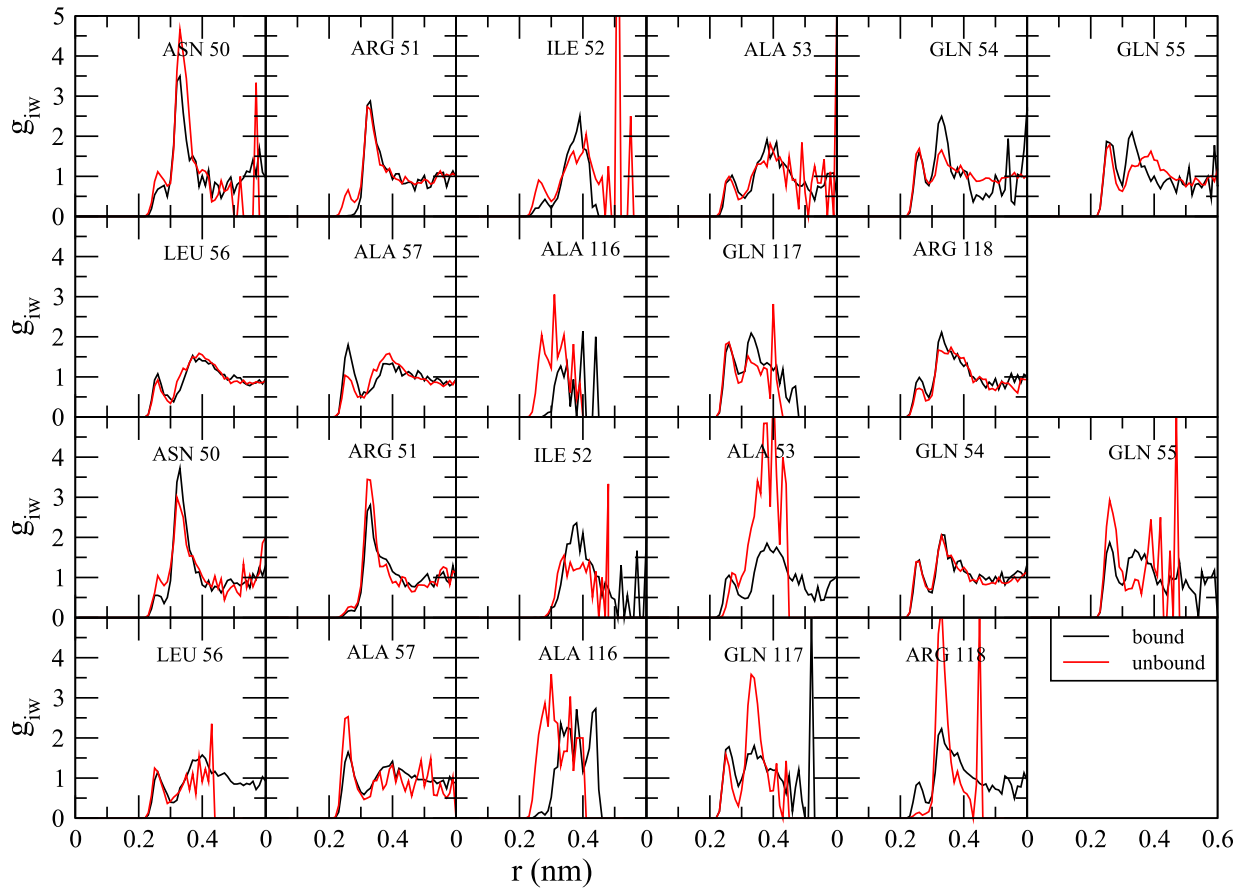


Figure S 6.1 Water probability distribution around selected residues (g_{iw}) as a function of distance in the bound (black) and the unbound (red) state of V52I at 1 bar. Top two rows of the figure show the g_{iw} for the residues in chain A, while the bottom two rows show the g_{iw} for the corresponding residues in chain B. The bound and unbound states were determined only considering the angle data.

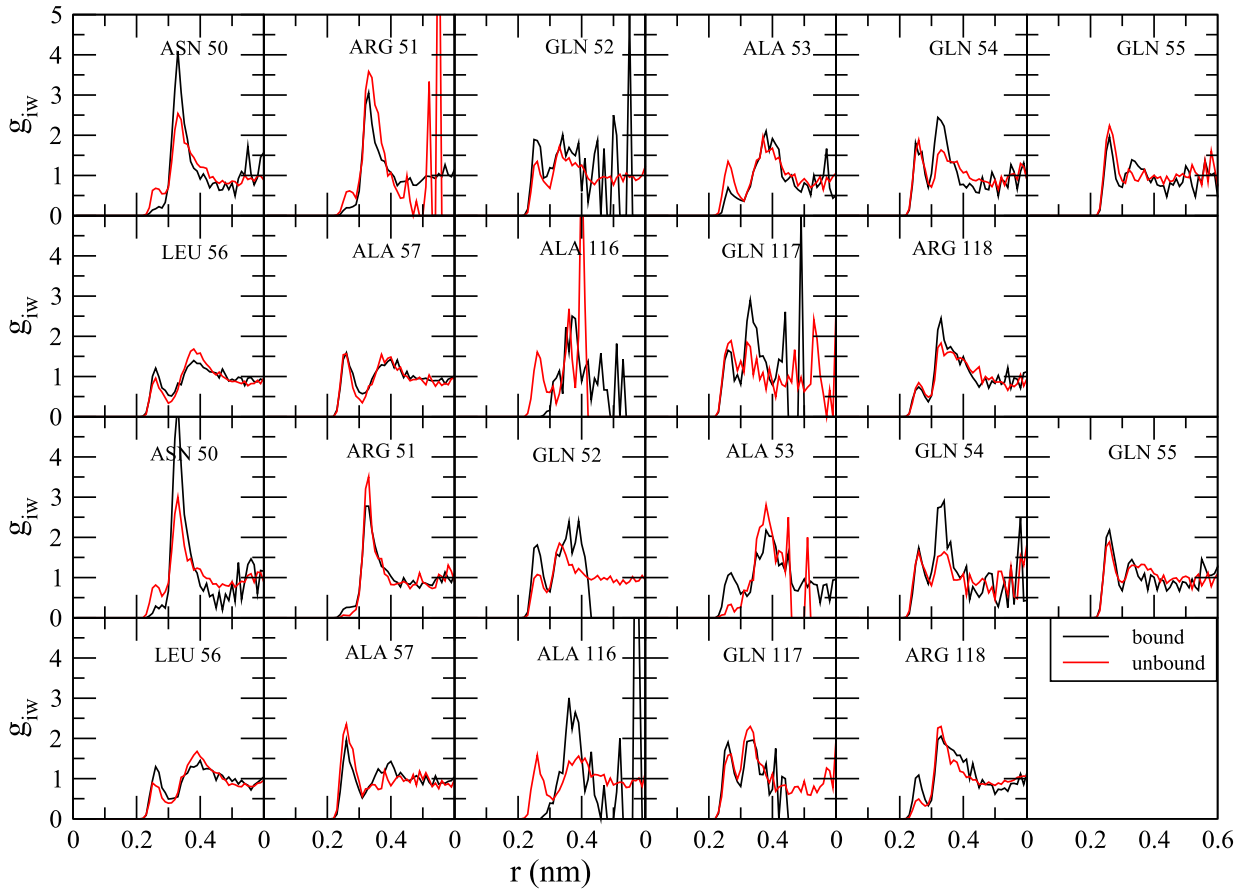


Figure S 6.2 Water probability distribution around selected residues (g_{iw}) as a function of distance in the bound (black) and the unbound (red) state of V52Q at 1 bar. Top two rows of the figure show the g_{iw} for the residues in chain A, while the bottom two rows show the g_{iw} for the corresponding residues in chain B. The bound and unbound states were determined only considering the angle data.

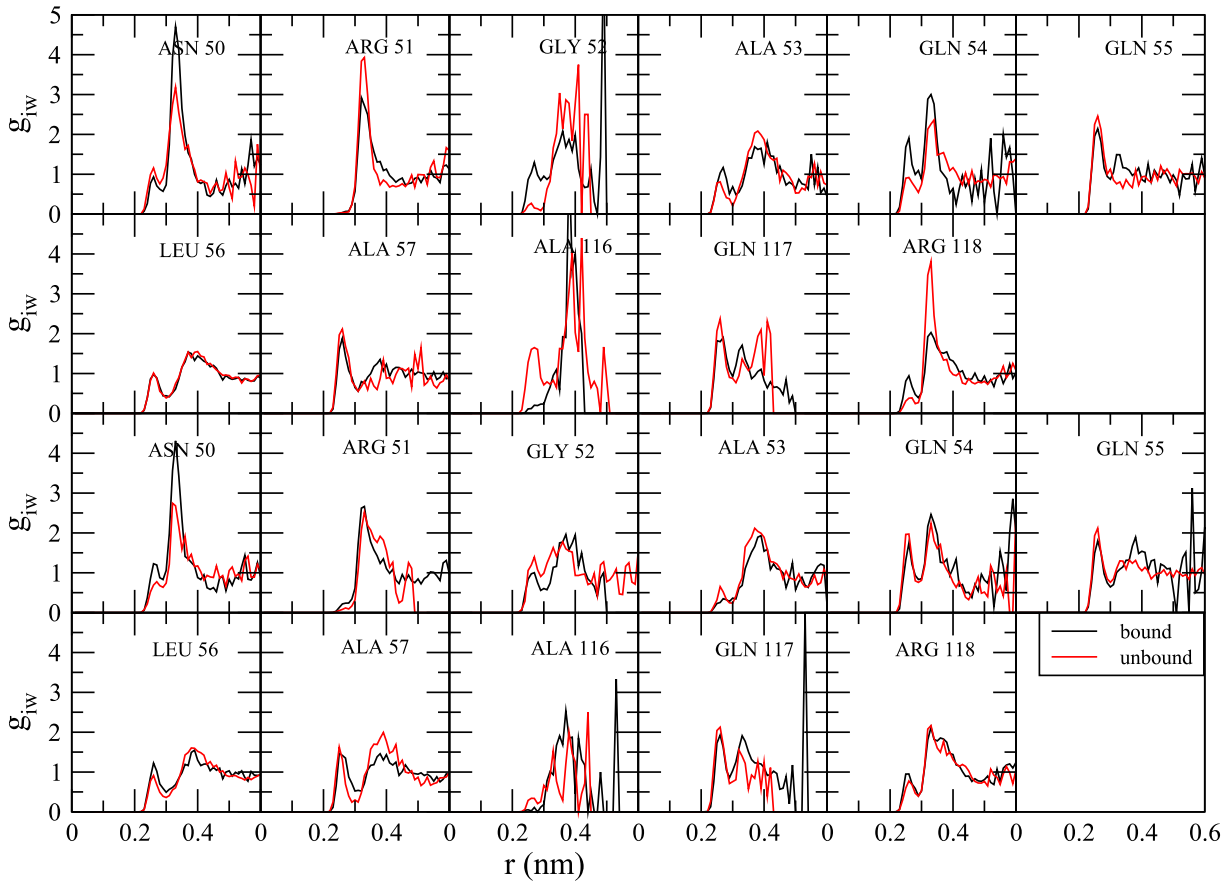


Figure S 6.3 Water probability distribution around selected residues (g_{iw}) as a function of distance in the bound (black) and the unbound (red) state of V52I at 1 bar. Top two rows of the figure show the g_{iw} for the residues in chain A, while the bottom two rows show the g_{iw} for the corresponding residues in chain B. The bound and unbound states were determined only considering the angle data.

6.6 References

1. Ng, P. C.; Levy, S.; Huang, J.; Stockwell, T. B.; Walenz, B. P.; Li, K.; Axelrod, N.; Busam, D. A.; Strausberg, R. L.; Venter, J. C., PLOS Genetics 2008, 4 (8), e1000160.

2. Tungtur, S.; Parente, D. J.; Swint-Kruse, L., Proteins: Structure, Function, and Bioinformatics 2011, 79 (5), 1589-1608.

3. Lee, B. C.; Park, K.; Kim, D., Proteins: Structure, Function, and Bioinformatics 2008, 72 (3), 863-872.

4. Dukka Bahadur, K.; Livesay, D. R., Bioinformatics 2008, 24 (20), 2308-2316.

5. Pei, J.; Cai, W.; Kinch, L. N.; Grishin, N. V., Bioinformatics 2005, 22 (2), 164-171.

6. Ye, K.; Anton Feenstra, K.; Heringa, J.; IJzerman, A. P.; Marchiori, E., Bioinformatics 2007, 24 (1), 18-25.

7. Meinhardt, S.; Manley Jr, M. W.; Parente, D. J.; Swint-Kruse, L., PloS one 2013, 8 (12), e83502.

8. Gray, V. E.; Kukurba, K. R.; Kumar, S., Bioinformatics 2012, 28 (16), 2093-2096.

9. de Beer, T. A.; Laskowski, R. A.; Parks, S. L.; Sipos, B.; Goldman, N.; Thornton, J. M., PLoS computational biology 2013, 9 (12), e1003382.

10. Miller, M.; Bromberg, Y.; Swint-Kruse, L., Scientific reports 2017, 7, 41329.

11. Swint-Kruse, L.; Matthews, K. S., Current opinion in microbiology 2009, 12 (2), 129-137.

12. Jobe, A.; Bourgeois, S., Journal of molecular biology 1972, 69 (3), 397-408.

13. Riggs, A. D.; Suzuki, H.; Bourgeois, S., Journal of molecular biology 1970, 48 (1), 67-83.

14. Shis, D. L.; Hussain, F.; Meinhardt, S.; Swint-Kruse, L.; Bennett, M. R., ACS synthetic biology 2014, 3 (9), 645-651.

15. Müller-Hill, B.; Rickenberg, H.; Wallenfels, K., Journal of molecular biology 1964, 10 (2), 303-318.

16. Bell, C. E.; Lewis, M., Nature Structural and Molecular Biology 2000, 7 (3), 209.

17. Flynn, T. C.; Swint-Kruse, L.; Kong, Y.; Booth, C.; Matthews, K. S.; Ma, J., Protein Science 2003, 12 (11), 2523-2541.

18. Lewis, M.; Chang, G.; Horton, N. C.; Kercher, M. A.; Pace, H. C.; Schumacher, M. A.; Brennan, R. G.; Lu, P., Science 1996, 271 (5253), 1247-1254.

19. DeLano, W. L., CCP4 Newsletter On Protein Crystallography 2002, 40 (1), 82-92.

20. Spronk, C. A.; Slijper, M.; van Boom, J. H.; Kaptein, R.; Boelens, R., Nature Structural and Molecular Biology 1996, 3 (11), 916.

21. Tungtur, S.; Meinhardt, S.; Swint-Kruse, L., Journal of molecular biology 2010, 395 (4), 785-802.

22. Meinhardt, S.; Manley Jr, M. W.; Becker, N. A.; Hessman, J. A.; Maher III, L. J.; Swint-Kruse, L., Nucleic acids research 2012, 40 (21), 11139-11154.

23. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., Journal of chemical theory and computation 2015, 11 (8), 3696-3713.

24. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., Bioinformatics 2013, 29 (7), 845-854.

25. Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G., Journal of computational chemistry 1997, 18 (12), 1463-1472.

26. Miyamoto, S.; Kollman, P. A., Journal of computational chemistry 1992, 13 (8), 952-962.

27. Hockney, R.; Goel, S.; Eastwood, J., Journal of Computational Physics 1974, 14 (2), 148-158.

28. Darden, T.; York, D.; Pedersen, L., The Journal of chemical physics 1993, 98 (12), 10089-10092.

29. Parrinello, M.; Rahman, A., Journal of Applied physics 1981, 52 (12), 7182-7190.

30. Nosé, S.; Klein, M., Molecular Physics 1983, 50 (5), 1055-1076.

31. Hoover, W. G., Physical review A 1985, 31 (3), 1695.

32. Nosé, S., Molecular physics 1984, 52 (2), 255-268.

Chapter 7 - Conclusions and Future Work

Simple nonpolarizable force field parameters were developed for the study of glycerol, 1,2-ethanediol, 1,2-propanediol, and 1,3-propanediol in solution using the Kirkwood-Buff theory of solution as a guide. All the models reasonably well reproduced the experimental Kirkwood-Buff integrals. Some other thermodynamic, kinetic and physical properties for the pure liquids and binary mixtures were also calculated and most of them were in good agreement with the experimental data. The parameters developed for the glycerol can be used to model the head group of the phospholipid molecules in progress.

Chapter 3 and 4 illustrated our new approach to calculate the partial molar volume and compressibility of proteins at infinite dilution using fluctuation solution theory. This approach does not use any subjective definitions and, for parameters for the protein volume and compressibility. In chapter 3, the factors that contribute to the very low compressibilities of proteins were discussed by decomposing the compressibility into residue-based and physicochemical group-based contributions. Furthermore, the contributions from volume fluctuations to the compressibility were considered. Our results suggest that the negatively charged residues provide a large negative contribution to the compressibility. However, we need to study more proteins to explain the overall small compressibilities of proteins as we have studied only the relatively small, neutral and positively charged proteins. In chapter 4, the partial molar volume and compressibility of the Ubiquitin protein, and the residue-based contributions to these properties, were calculated and compared among the major force fields; AMBER99SB-ILDN, CHARMM22*, GROMOS-53A6, and OPLS-AA. Similar trends were observed for the volume as a function of pressure. However, different trends were observed for the compressibility as a

function of pressure for the different FFs. This might be due to the fact that the different FFs use slightly different approaches to obtain partial atomic charges. As future work, we are planning to study the same properties using the Kirkwood-Buff derived force field (KBFF) for proteins developed by Smith group for a complete comparison. Also, the results need to be compared by only varying the water model to examine the effect of water models on the results.

The effect of pressure on the conformation of wild-type (WT) *lac* repressor protein and some of its mutants were studied (chapter 5 and 6). The mutations were made at the 52^{nd} position of LacI, which is an experimentally known rheostat position. For the WT LacI, we observed that there is an equilibrium between the DNA-bound and unbound states which is sensitive to pressure. Our on-going research to understand the effect of pressure on mutants showed that there can be intermediate states during the conformational transition from DNA bound to the DNA unbound state. We observed different levels of hydration for different mutants, and yet we need longer simulations further analysis to fully explain the experimental results. All the simulations of the mutants will be extended up to 1 μs. Furthermore, the mutants need to be studied in the presence of anti-inducer ligand (ONPF) to examine the differences compared to the WT in the presence of ONPF.

Appendix A - Rationally Designed Peptide Nanosponges for Cell-Based Cancer Therapy

BASIC SCIENCE

Nanomedicine: Nanotechnology, Biology, and Medicine
13 (2017) 2555–2564

nanomedicine
Nanotechnology, Biology, and Medicine

ELSEVIER

Original Article

nanomedjournal.com

Rationally designed peptide nanosponges for cell-based cancer therapy

Hongwang Wang, PhD[a,1], Asanka S. Yapa, PhD[a,1], Nilusha L. Kariyawasam, BSc[a,1], Tej B. Shrestha, PhD[b,1], Madumali Kalubowilage, PhD[a], Sebastian O. Wendel, PhD[a,b], Jing Yu, MS[a], Marla Pyle, MSc[a], Matthew T. Basel, PhD[b], Aruni P. Malalasekera, PhD[a], Yubisela Toledo, BSc[a], Raquel Ortega, BSc[a], Prem S. Thapa, PhD[c], Hongzhou Huang, PhD[d], Susan X. Sun, PhD[d], Paul E. Smith, PhD[a], Deryl L. Troyer, PhD, DVM[b], Stefan H. Bossmann, PhD[a,*]

[a]*Department of Chemistry, Kansas State University, Manhattan, KS, USA*
[b]*Department of Anatomy & Physiology, Kansas State University, Manhattan, KS, USA*
[c]*Microscopy and Analytical Imaging Laboratory, University of Kansas, Lawrence, KS, USA*
[d]*Grain Science and Industry, Kansas State University, Manhattan, KS, USA*

Received 15 March 2017; accepted 12 July 2017

Abstract

A novel type of supramolecular aggregate, named a "nanosponge" was synthesized through the interaction of novel supramolecular building blocks with trigonal geometry. The cholesterol-$(K/D)_n$DEVDGC$)_3$-trimaleimide unit consists of a trigonal maleimide linker to which homopeptides (either K or D) of variable lengths (n = 5, 10, 15, 20) and a consensus sequence for executioner caspases (DEVDGC) are added via Michael addition. Upon mixing in aqueous buffer cholesterol-$(K)_n$DEVDGC$)_3$-trimaleimides and a 1:1 mixture of cholesterol-$(K/D)_n$DEVDGC$)_3$-trimaleimides form stable nanosponges, whereas cholesterol-$(D)_n$DEVDGC$)_3$-trimaleimide is unable to form supramolecular aggregates with itself. The structure of the novel nanosponges was investigated through explicit solvent and then coarse-grained molecular dynamics (MD) simulations. The nanosponges are between 80 nm and several micrometers in diameters and virtually non-toxic to monocyte/macrophage-like cells.

Key words: Nanosponges; Drug delivery; Nanomaterials; Supramolecular; Aggregation; Explicit solvent molecular dynamics simulations; Coarse-grained molecular dynamics simulations

Classic liposomes and peptide vesicles

Targeted delivery of therapeutics to the tumor site is of vital importance in cancer treatment. This approach is able not only to maximize the treatment efficacy of therapeutics at the cancer site(s), but also to minimize the side effects caused by the therapeutics in conventional cancer treatment.[1,2] One method to achieve targeting delivery is to use a delivery modality designed to carry the therapeutics to the desired site, and then release them in the tumor.[3–6] Liposomes are a well-recognized example of drug delivery devices. Composed of a simple lipid bilayer, liposomes are non-cytotoxic, biocompatible, biodegradable, and capable of integrating or encapsulating large payloads of both hydrophilic and hydrophobic drugs.[7] Drugs incorporated into these nanocarriers can be accumulated in tumor tissue through the enhanced permeability and retention (EPR) effect.[8–10] Numerous liposome based drugs have been approved by the FDA, and many more are at different stages of clinical trials.[11] Despite their successes, liposomes have limitations. It is noteworthy that the preparation of liposome based drug requires

Abbreviations: PBS, phosphate-buffered saline buffer; AFM, Atomic Force Microscopy; TEM, Transmission Electron Microscopy; HBTU, (2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate); DIEA, N,N-Diisopropylethylamine; CDI, (carbonyl-di-imidazole); TIPS, triisopropylsilane; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid.

This work was funded by NSF (DMR 1242765 and CBET EAGER 1656989), and the Johnson Cancer Center at Kansas State University.

The authors declare no competing financial interest.

* Corresponding author at: Kansas State University, Department of Chemistry, Manhattan, KS 66506-0401.

E-mail address: sbossman@ksu.edu (S.H. Bossmann).

[1] These authors have contributed equally.

http://dx.doi.org/10.1016/j.nano.2017.07.004
1549-9634/

Please cite this article as: Wang H, et al, Rationally designed peptide nanosponges for cell-based cancer therapy. *Nanomedicine: NBM* 2017;13:2555-2564, http://dx.doi.org/10.1016/j.nano.2017.07.004

multistep procedures (i.e., hydration,[12,13] sonication,[14,15] extrusion,[16,17] using a size selective column,[18] etc.) in order to obtain narrow particle size distribution and separation from unloaded drugs. These tedious processes are associated with a high risk of damaging the entrapped drugs. Furthermore, the EPR effect is only slightly selective, thus achieving rarely more than 5 percent delivery of a nanotherapeutic drug to the tumor site(s).[8–10] Finally, liposomes are prone to systemic leaking of drugs, especially at longer circulation times.[19]

Self-assembling peptides are an attractive alternative to liposomes. For example, short amphiphilic sequences, acetyl-AAVVLLLW-$(E)_{n=2/7}$-COOH, form nanosized vesicles spontaneously in aqueous media at neutral pH. Hydrophilic molecules can be incorporated inside the vesicles.[20] Longer block copolypeptides poly(L-lysine)-*b*-poly(L-leucine), K_xL_y (x = 20 to 80, y = 10 to 30), form stable vesicles and micelles in aqueous solution with size ranging from 1 to 10 μm in diameter. These assemblies showed high degrees of membrane fluidity; as a result, they can be resized with precise control from ten to hundreds nanometers in diameter using liposome-based extrusion techniques.[21] The Tomich group reported two branched peptides with different lengths, mimicking diacyl glycerols, form water-filled vesicles, which can entrap water soluble dyes.[22–25]

The EPR effect works well in mice, but not in humans

During the last decade, it has become more and more evident that both, classic liposomes and peptide vesicles are facing the problem of ineffective drug delivery in humans. It is an emerging paradigm that the Enhanced Permeation and Retention Effect (EPR) works well in mouse models of cancer, but not in the clinic.[10,26,27] Drug transport as a payload of either stem cells[28] or defensive cells,[27] which migrate to tumors following their cytokine/chemokine secretion, is a new concept that has been proven effective in animal models. Currently, clinical translation of cell-based treatment methods for cancer and other diseases is rapidly progressing.[29] Therefore, we have developed "peptide nanosponges" for efficient targeting of defensive cells in peripheral blood, as well as cultured stem cells.

Peptide nanosponges

Peptide nanosponges that are reported here, are capable of effectively delivering their payload to defensive cells and stem cells. Especially autologous cells have the potential of truly personalized medicine when treating solid tumors and metastases.[30]

Here, we report the synthesis of a series of $(K)_n$DEVDGC, and $(D)_n$DEVDGC peptide sequences, where *n* equals to 5, 10, 15, 20 respectively. We have capped the N-terminal of the peptides with cholesterol, and further linked the peptides to a trimaleimide scaffold via Michael-addition.[31] We have obtained one positively and one negatively charged adduct [(cholesterol-$(K)_n$DEVDGC)$_3$-trimaleimide and (cholesterol$(D)_n$DEVDGC)$_3$-trimaleimide]. Upon mixing of the adduct pairs (*n* equals to 15 or 20) equimolarly under physiological conditions, nanosponges of very low polydispersity form instantaneously, which were characterized with dynamic light scattering (DLS), transmission electron microscopy (TEM), and atomic force microscopy (AFM).

Our computer modeling has indicated that the structure of the nanosponges is indeed "sponge-like": numerous hydrophilic and hydrophobic nanodomains exist in direct proximity. We attribute the novel (bio)physical properties of the nanosponges to their formerly unknown structure (Figure 1).

Our studies demonstrated that hydrophobic molecules, for example the cyanine 3.0 dye PKH26, can be incorporated inside these nanosponges. In the presence of cancer related proteases (e.g. caspase-3, 6 or 7[32]), these nanovesicles can be – principally – digested, leading to the possibility of triggered release of the payload. We found that the nanosponges are essentially non-toxic, and that these cells internalize them with high efficiency.

Based on these proof-of-concept experiments, our novel peptide-based nanosponges are very well suited for applications in specific drug delivery to solid tumors and metastases by means of cell-based therapy.

Methods

Trimaleinimide scaffold synthesis

A flexible trimaleinimide scaffold was synthesized by means of a two-step reaction. In the first step, reacting tris(2-aminoethyl) amine with 3 equivalents of maleic anhydride in acetic acid at room temperature produces the trimaleimic acid adduct.[33] In the second step, the trimaleimic acid adduct and sodium acetate were heated in acetic anhydride for 30 min at 100 °C to give the desired product.[34] The crude product was recrystallized from saturated ethyl acetate, and fully characterized by ^{1}H, ^{13}C NMR, and single crystal x-ray analysis (see SI section).

Peptide synthesis

Oligopeptides were synthesized by means of solid phase peptide synthesis on 2-chlorotrityl resin.[35,36] Three equivalents of F_{moc} (N-(9-fluorenyl)methoxycarbonyl) protected amino acid and HBTU were dissolved in a DIEA/DMF solution, and added to the 2-chlorotrityl resin preloaded with 0.20 mmol of amino acid per g. The solution was drained from the resin after 30 min of reaction. This process was repeated one more time. Then, the F_{moc} group of the newly introduced amino acid was removed by using 20% (v/v) piperidine in DMF. Following this procedure, stepwise addition of F_{moc}-protected amino acids resulted in the desired peptides. The N-terminal of the peptides was capped with cholesterol while still being on the resin by reacting with CDI activated cholesterol in DMF solution.[35,36] The final product was cleaved off the resin in TFA/water/TIPS (95:2.5:2.5, v/v/v) cocktail for 3 h at room temperature.[36] White solid product formed when adding the cocktail into cold anhydrous diethyl ether. The product was collected by centrifugation (3000 rpm, 10 min), washed with cold diethyl ether for three times, and dissolved in water prior to lyophilization. The products were purified by using a GE peptide column (mobile phase: aqueous 0.05 M TEA/acetic acid buffer, pH = 7.0), and dried in high vacuum.

Cholesterol-peptide-trimaleimide adduct formation

3.5 equivalents of cholesterol-peptide and 1 equivalent of trimaleimide were dissolved in deoxygenated PBS buffer (pH =

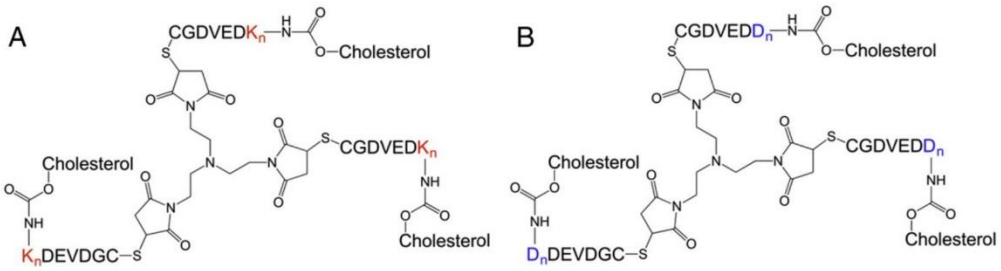


Figure 1. Tri-maleimide based peptide structures: components for the spontaneous formation of nanosponges. (**A**) Lysine-based materials, n = 5, 10, 15, 20. (**B**) Aspartic acid-based materials, n = 5, 10, 15, 20.

7.4), and stirred under argon atmosphere for 24 h.[37] After removing the solvent by lyophilization, the crude product was purified by dialysis (molecular weight cutoff 3500) against distilled water. The final product inside the membrane bag was lyophilized and further dried under high vacuum.

Nanosponge formation and DLS characterization

Separate solutions of (cholesterol-(K)$_n$DEVDGC)$_3$-trimaleimide and (cholesterol-(D)$_n$DEVDGC)$_3$-trimaleimide in deoxygenated PBS buffer were prepared and filtered through 200 μm filters. The prepared stock solutions were 0.050 mM and 0.50 mM. All other stock solutions were prepared by diluting the original solutions with deoxygenated PBS buffer. The two solutions were quickly mixed and vortexed for 30 s. The hydrodynamic diameters and polydispersity indexes (PDI) of the formed nanosponges were measured by dynamic light scattering (DLS, ZetaPALS, Brookhaven Instruments Corp., Holtsville, NY).[38] All measurements were carried out at 298 K, using 658 nm laser wavelength, and 90° detection angle. Data were collected from an average of three measurements over 60 s. DLS was also used to estimate the critical micellar concentration (cmc) of the nanosponges.

AFM characterization

Samples for atomic force microscopy (AFM) were prepared by adding one drop the nanosponge stock solution (0.050 M of each component in PBS) onto a freshly peeled MICA sheet, followed by removing of the solvent by using a gentle nitrogen stream (2 min). AFM images were taken by the Bruker Innova AFM image system (Bruker, Camarillo, CA) utilizing TESPA-HAR probes in tapping mode. The spring constant of the tip was 50 N/m and the frequency was 350 kHz. The set point, P gain and I gain were set at 1.2, 0.6 and 0.5, respectively. The images were gathered with 256 × 256 pixel resolution at a scan rate of 1 Hz. The images were then analyzed by the Nanoscope software (Bruker).

TEM characterization

Samples for transmission electron microscopy (TEM) were prepared by dropping 10 μL of 0.005% type DK20 solution in PBS directly on a glow discharged TEM grid. Uranyl acetate was used as a positive staining agent. In all cases electron microscopy was performed at an accelerating voltage of 200 kV. Nanosponge morphology on HOPG was examined by bright-field and dark-field transmission electron microscopy (TEM) using a FEI Technai G_2 transmission electron microscope at an electron acceleration voltage of 200 kV. Dark-field TEM did not reveal a characteristic diffraction pattern. High resolution images were captured using a standardized, normative electron dose and a constant defocus value from the carbon-coated surfaces. All TEM measurements were performed at the Microscopy and Analytical Imaging Laboratory of the University of Kansas.[39]

Molecular dynamics simulations

Classical molecular dynamics were performed using the gromacs software.[40] Simulations using the all atom (AA) Gromos force field (FF) were performed in the NpT ensemble at 300 K and 1 bar using the v-rescale and Berendsen temperature and pressure algorithms,[41,42] respectively. Electrostatic interactions were evaluated using the particle mesh Ewald approach,[43] while van der Waals interactions were truncated at 1.5 nm. The timestep was 2 fs and all solute bonds were constrained using Lincs,[44] while all solvent bonds (and angles) were constrained using Settle.[45] The coarse-grained (CG) simulations were also performed in the NpT ensemble at 310 K and 1 bar using the v-rescale and Parrinello-Rahman temperature and pressure algorithms,[41,46] respectively, as suggested by the MARTINI FF developers (http://www.cgmartini.nl/). Electrostatic and van der Waals interactions were evaluated using shifted potentials with a relative permittivity of 15.[47] The timestep was 25 fs and all solute bonds were constrained using Lincs. The CG simulations were checked to ensure that there was no freezing of water beads by calculating the diffusion constants of water periodically during the simulations. A variety of system sizes were simulated with the largest involving 108 (cholesterol-(K)$_{20}$DEVDGC)$_3$-trimaleimide and 108 (cholesterol-(D)$_{20}$DEVDGC)$_3$-trimaleimide molecules, 2376 sodium ions, and 776,763 water beads in a 45 nm cube box for 4 μs. Details of the FFs are provided in the Supporting Information.

Entrapment of PKH26 within peptide nanosponges

PKH26 solution was prepared by dissolving 10 μL of the PKH26 ethanolic dye solution (Sigma-Aldrich) to 1.0 mL of Diluent C in a polypropylene centrifuge tube, followed by addition of double-distilled water (pH = 6.90) to bring the total volume to 2.0 mL (final PKH26 concentration 5.0×10^{-6} M). Equal molar amounts of cholesterol-(K)$_{20}$DEVDGC)$_3$-trimaleimide and cholesterol-(D)$_{20}$DEVDGC)$_3$-trimaleimide (1.0 mM of each component) were added to the above dye solution. After brief sonication, the homogeneous solution was incubated at 37 °C for 6

h without light exposure. The free dye was removed by passing the entire sample through a Sephadex G-50 gel filtration column using double-distilled water as eluent. The collected fractions containing the nanosponges were lyophilized to dryness, and re-hydrated with PBS buffer (pH = 7.4). The DLS measurement showed that the hydrodynamic diameters of the nanovesicles were between 110 to 130 nm.

Cell experiments and MTT assays

The cytotoxicity of the PKH26 containing nanosponges was assessed by utilizing the MTT assay[48] on RAW264.7 monocyte/macrophage-like cells.[35] Cell experiments were carried out in RPMI 1640 medium with 10% FBS. The percentage of viable cells was determined after 24 and 48 h of incubation. Cells were seeded in a T-25 flask. After 24 h of incubation at 37 °C and 5% CO_2, cells were re-plated in a 96 well plate at 20000/cm^2 density and further incubated for 24 h at 37 °C, 5% CO_2, to obtain 80% confluency before the nanosponges were added.

A concentration series of the nanosponge composed of (cholesterol-$(K)_{20}$DEVDGC)$_3$-trimaleimide and (cholesterol-$(D)_{20}$-DEVDGC)$_3$-trimaleimide (0.0, 0.1, 0.2, 0.5, 1.0, 2.0, 5.0, 10, 20, 40, 60, 80, 100 μmol L^{-1} in total, molar ratio 1:1) was prepared by dissolving the nanosponge components in the same media that were used for culturing the cells. Cells were incubated for 24/48 h at 37 °C. Eight replicates were prepared for each concentration. A portion of 10 μL of MTT reagent (5 mg/ml in PBS) was added to each well, and the plates were incubated for another 4 h at 37 °C. Finally, 100 μL of 10% sodium dodecyl sulfate in 0.010 M HCl was added into each well and incubated for 24 h at 37 °C. Their absorbance was recorded by using a plate reader at 550 nm and 690 nm. PBS solution was used as control for all the experiments. The solution with of μmol L^{-1} of nanosponge served as control.

RAW264.7 cells were imaged by using a Zeiss, Axiovert 40 CFL microscope with darkfield, brightfield, phase contrast and epifluorescence illumination, a camera system and Jenoptik, a ProgRes C3 Cool camera and ProgRes Capture Pro 2.10.0.0 software.

Results

DLS characterization of the nanosponges

The effective diameters and the polydispersity index (PDI) values of the nanosponges obtained by dynamic light scattering measurements (DLS) are summarized in Table 1. These results indicate that nanosponge formation depends on the number of lysine/aspartic acid units. Larger aggregates with higher polydispersity are observed when *n* equals 5. A significant decrease in both size and PDI (polydispersity index) was observed when increasing n to 10 (440 nm and 0.26). A further increase of K and D to 15 and 20 led to virtually mono-dispersed nanosponges with effective diameters of approx. 200 nm. It is noteworthy that the formation of these nanosponges is spontaneous upon mixing of the adduct solutions. Continuous monitoring by DLS for 12 h at 298 K revealed that the nanosponges are very stable in aqueous solution (PBS). For drug delivery purposes, we are particularly interested in nanosponges of 100 to 200 nm in diameter. Further characterization was carried out for n = 15 and 20. The corresponding correlation curves and number-averaged size distributions are shown in the SI section.

Table 1
Effective hydrodynamic diameters, polydispersity indexes (PDI), and standard deviations (SD) for (cholesterol-$(K)_n$DEVDGC)$_3$-trimaleimide + (cholesterol-$(D)_n$DEVDGC)$_3$-trimaleimide nanosponges (0.050 mM of each component in PBS).

n (D and K) =	Effective diameter ± SD (nm)	PDI ± SD
5	1200 ± 240	0.642 ± 0.07
10	440 ± 50	0.26 ± 0.04
15	200 ± 5.0	0.077 ± 0.01
20	180 ± 25	0.201 ± 0.03

Hydrodynamic diameters remained virtually constant for 12 h.

CMC of the nanosponges

In analogy to the formation of micelles, a critical concentration at which spontaneous aggregation to nanosponges occurs, was determined. This molar concentration was named cmc in analogy to "critical micellar concentration". In a monodisperse nanomaterial solution, the correlation curve (C(t)) of the measured data in a dynamic light scattering (DLS) experiment is a smooth, single exponential decay function. The diffusion coefficient (D) is proportional to the lifetime of the exponential decay and can be calculated by fitting the correlation curve to an exponential function. The hydrodynamic diameter of particles can be obtained by using a variation of the Stokes-Einstein equation with known D value.[49] Based on basic DLS theory, we measured the critical concentrations for the nanosponge formation of (cholesterol-$(K)_{20}$DEVDGC)$_3$-trimaleimide (K) and a 1:1 mixture of (cholesterol-$(D)_{20}$DEVDGC)$_3$-trimaleimide (D) and (cholesterol-$(K)_{20}$DEVDGC)$_3$-trimaleimide (K). The concentrations of D, K, and DK were stepwise increased by adding microliter aliquots from stock solutions of 0.50 M each to 1.0 mL of PBS buffer in a cuvette. The endpoint indication of this titration curve is the appearance of a smooth, single exponential decay correlation curve. (Cholesterol-$(D)_{20}$DEVDGC)$_3$-trimaleimide (D) did not show significant aggregation even after increasing its concentration to 0.30 mM. For (cholesterol-$(K)_{20}$DEVDGC)$_3$-trimaleimide a smooth, single exponential decay curve was observed when its concentration reached 0.080 mM. The equimolar mixture of (cholesterol)-$(D)_{20}$DEVDGC)$_3$-trimaleimide and (cholesterol-$(K)_{20}$DEVDGC)$_3$-trimaleimide (DK) showed most facile nanosponge formation at a concentration as low as 0.0050 mM (total concentration, 0.0025 mM (D) and 0.0025 mM (K)). In comparison, sodium dodecyl sulfate micelles possess a cmc of approx. 8 mM at 298 K, which corresponds to 2.31 g L^{-1}. In comparison, only about 0.055 g L^{-1} of type DK nanosponges and 0.90 g L^{-1} of type K nanosponges are required to achieve spontaneous aggregation (Figure 2).

AFM characterization

Figure 3 shows the AFM images of (cholesterol-$(D)_{15}$-DEVDGC)$_3$-trimaleimide/cholesterol-$(K)_{15}$DEVDGC)$_3$-trimaleimide and (cholesterol-$(D)_{20}$DEVDGC)$_3$-trimaleimide/ (cholesterol-$(K)_{20}$-DEVDGC)$_3$-trimaleimide nanosponges. Type DK15 nanosponges formed 0.5-0.9 μm aggregated bundles. The height of the bundles is

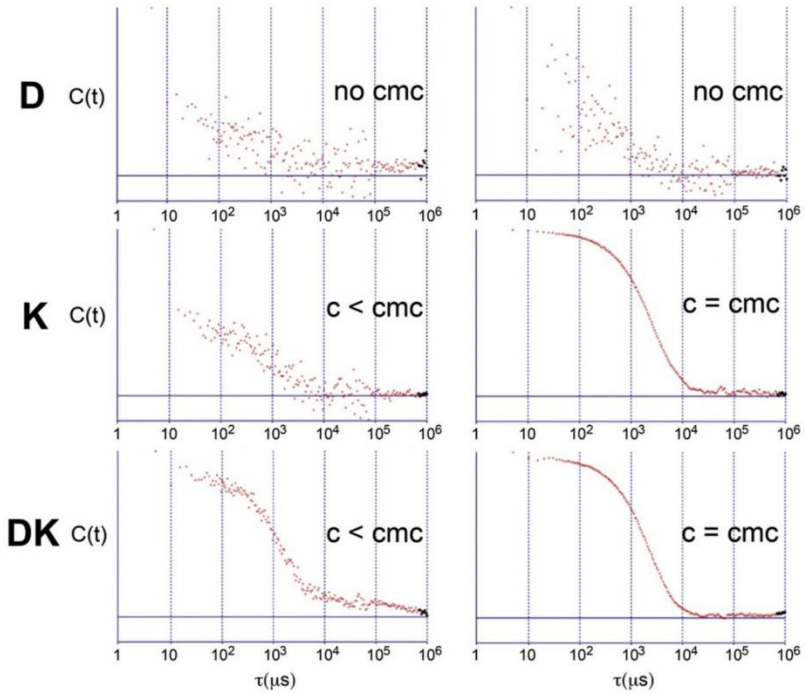


Figure 2. Correlation curves (C(t)) of dynamic light scattering measurements of (cholesterol-$(D)_{20}$DEVDGC)$_3$-trimaleimide (D), (cholesterol-$(K)_{20}$DEVDGC)$_3$-trimaleimide (K), and mixture (1/1 molar ratio) of both nanosponge components (DK) in 1× PBS buffer. In the left column, the concentrations of K, and DK are below the cmc (critical micellar concentration, here: concentration at which aggregation occurs). In the right column the concentrations are at their respective cmc (0.080 mM for type K and 0.0050 mM for type DK). No aggregation was observed for D in the concentration interval from 0.01 (left) to 0.30 mM (right).

between 150 to 250 nm. At higher magnification, it can be discerned that each bundle was formed by 3 to 5 smaller subunits. The diameter of the subunits ranges from 150 nm to 200 nm. Type DK20 nanosponges formed well defined individual nanosponges of 85 to 110 nm in size. Their height falls into the same range, indicating the formation of spherical nanosponges.

TEM characterization

TEM images for type DK20 nanosponges are shown in Figure 4. 2D projections of spherical sponges with diameters between 85 and 100 nm are clearly discernible (Figure 4, *A*). Their size distribution is displayed in Figure 4, *B*. However, smaller structures that are 35 to 45 nm in size can also be found in the TEM images. It is noteworthy that the exterior of the nanosponges acquired strong uranyl stains. This is an indication that cholesterol-$(D)_{20}$DEVDGC)$_3$-trimaleimide is enriched at the exterior of the nanostructures.

Molecular dynamics simulations

In an effort to elucidate the structure of the peptide aggregates we have performed all atom (AA) explicit solvent and coarse-grained (CG) molecular dynamics (MD) simulations. Technical aspects of the simulations, together with a detailed description of the models used, are provided in the Methods and the Supporting Information, respectively. However, before simulating the aggregation process itself it is important to check that the CG models are sufficiently accurate that reasonable results can be obtained. CG simulations are required as the systems under investigation involve large molecular aggregates. However, CG models generally provide rather crude representations of electrostatic interactions and conformational flexibility. As both these aspects are clearly present in the systems to be studied here, we have also investigated the ability of our CG models to mimic the more accurate AA explicit solvent analogues.

The simulation described here involves molecules for which no force fields (FFs) are currently available. Here, we describe our approach to provide reasonable descriptions of these systems using all atom (AA) and coarse-grained (CG) approaches. Highly accurate FFs for these systems would require significant development and may also necessitate experimental data that are not available. Hence, we have taken a more approximate, but practical, approach. We feel that this is appropriate as we are probing the overall behavior of the systems, and the requirement of CG models to study such large systems already introduces significant approximation.

The results from 100 ns AA and 1 μs CG MD simulations of the (cholesterol-$(K)_{20}$DEVDGC)$_3$-trimaleimide and (cholesterol-$(D)_{20}$-DEVDGC)$_3$-trimaleimide peptides have been compared. Electrostatic interactions followed the usual approach for the MARTINI models,[47] while partial conformational flexibility was introduced as

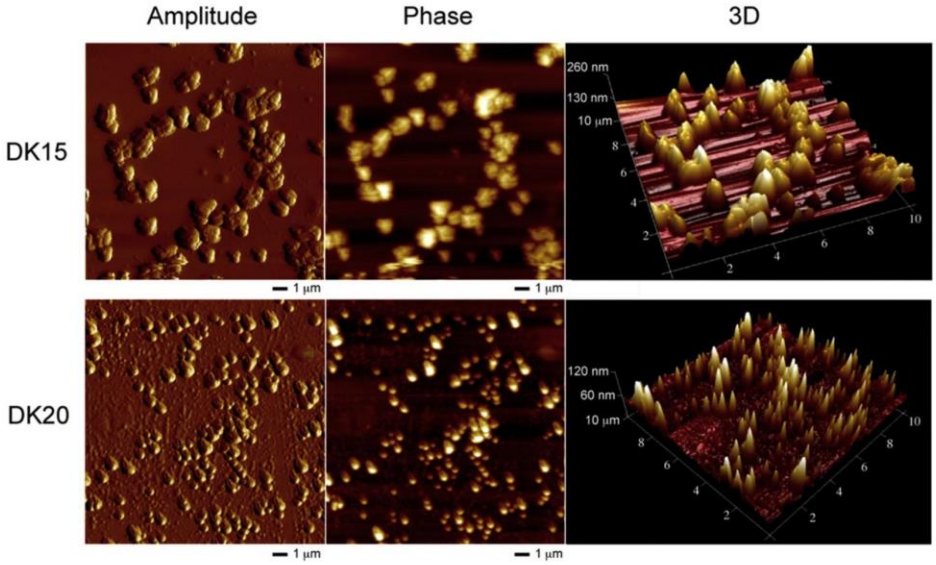


Figure 3. AFM (amplitude, phase, and 3D) images of type DK15 and type DK20 nanosponges. "15" and "20" refer to the number of D and K units in the oligopeptides that are attached to trimaleimide linkers.

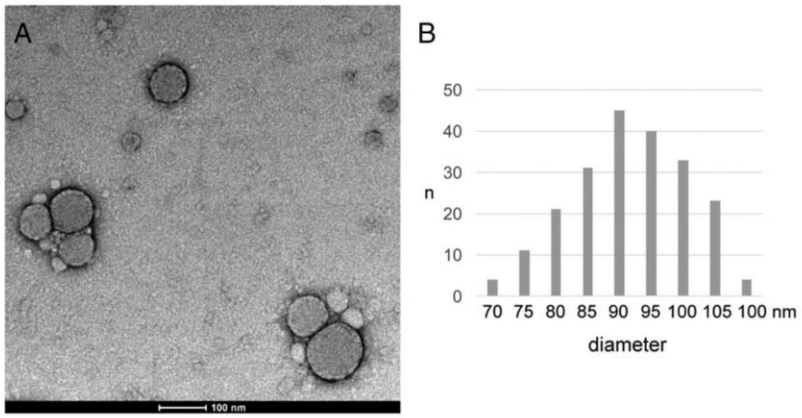


Figure 4. (**A**) TEM image of type DK20 nanosponges on graphite with uranyl acetate as positive staining. (**B**): Histogram of the size distribution of the larger nanosponges, obtained by using Image J (n: number of nanosponges counted in each group).[50]

described in the Supporting Information. The results are illustrated in Figure 5, *A*. The most notable behavior of the two peptide strands was the extended structures observed for (cholesterol-$(D)_{20}$-DEVDGC)$_3$-trimaleimide, and the collapsed structures observed for (cholesterol-$(K)_{20}$DEVDGC)$_3$-trimaleimide strands, as indicated by the AA simulations. The collapse of the (cholesterol-$(K)_{20}$-DEVDGC)$_3$-trimaleimide chain appears to require cholesterol as removal of this group eliminated any chain collapse (data not shown). This later observation is then in agreement with experimental data on poly-lys and poly-asp strands,[51–53] where the chains adopt extended or random coil structures. Clearly, the presence of cholesterol modifies this behavior. However, the same is not true for (cholesterol-$(D)_{20}$DEVDGC)$_3$-trimaleimide, which remains extended even in the presence of the cholesterol linkage. Most importantly, this difference in behavior is well reproduced in the CG simulations which also give rise to an extended (cholesterol-$(D)_{20}$DEVDGC)$_3$-trimaleimide and collapsed (cholesterol-$(K)_{20}$DEVDGC)$_3$-trimaleimide structures. Further examination of the (cholesterol-$(K)_{20}$DEVDGC)$_3$-trimaleimide simulation did not reveal any secondary structure formation upon collapse. Nevertheless, the identical behavior observed for the AA and CG models suggests that conformational

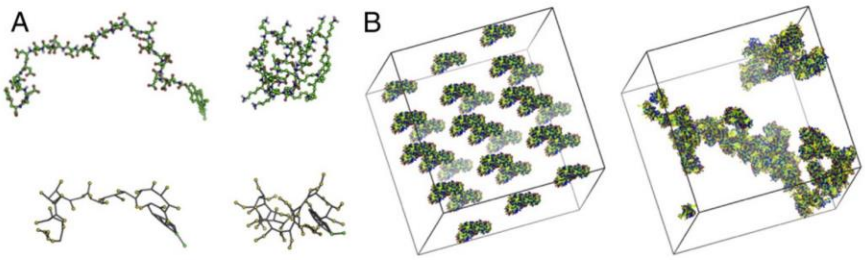


Figure 5. (**A**) Final structures obtained from the AA (top) and CG (bottom) simulations of (cholesterol-$(D)_{20}$DEVDGC)$_3$-trimaleimide (left) and (cholesterol-$(K)_{20}$DEVDGC)$_3$-trimaleimide (right). (**B**) Initial and final (4 μs) structures obtained from the CG simulation of (cholesterol-$(K)_{20}$DEVDGC)$_3$-trimaleimide and (cholesterol-$(D)_{20}$DEVDGC)$_3$-trimaleimide.

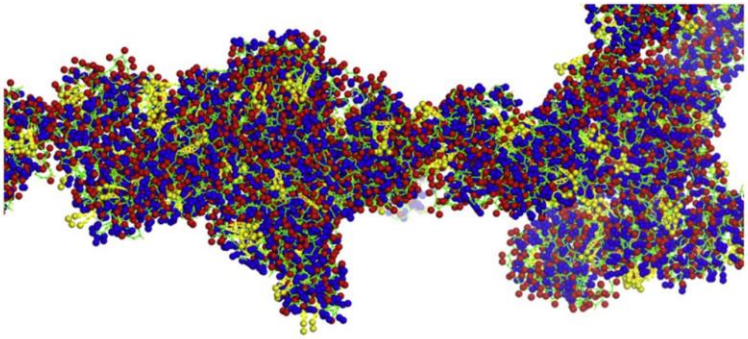


Figure 6. Expanded view of the final structure obtained from the CG MD simulations. The peptide backbone is displayed as green sticks, the Asp side chains are displayed as red balls, the Lys side chains are displayed as blue balls, while the cholesterol molecules are colored yellow. The structure resembles a "nanosponge" with hydrophobic and hydrophilic areas, as well as solvent-filled cavities.

flexibility and electrostatic interactions in these systems are sufficiently well represented that one can have confidence in the CG simulations.

The aggregation of equimolar mixtures of (cholesterol-$(K)_{20}$-DEVDGC)$_3$-trimaleimide and (cholesterol-$(D)_{20}$DEVDGC)$_3$-trimaleimide was performed in two steps using just CG MD simulations. In the first step, we randomly placed four (cholesterol-$(K)_{20}$-DEVDGC)$_3$-trimaleimides and four (cholesterol-$(D)_{20}$DEVDGC)$_3$-trimaleimides in a relatively small simulation box (15 nm in length), then transferred this peptide arrangement to a larger solvated box (40 nm in length) and simulated for 1 μs. This places the molecules in close proximity and they quickly formed a single relatively compact aggregate. In the second step the peptide aggregate was resolvated in a 15 nm length box, and then replicated in all three directions to form the final simulation box (45 nm in length) that was then simulated for 4 μs. During this period the smaller aggregates formed larger aggregates. This behavior is illustrated in Figure 6. Here, the smaller aggregates formed larger worm-like structures. Indeed, after 4 μs there were no isolated aggregates as all peptides chains contacted at least one other peptide chain. Clearly, the final structure obtained here does not represent that of a typical spherical vesicle, but more of a nanosponge. However, this is not too surprising as the peptides used here do not possess significant amphiphilic character compared to lipids, for example. Nevertheless, aggregation is observed in agreement with the experimental results described above, and other studies of poly-lys and poly-asp mixtures.[54]

While appearing largely amorphous the final structure obtained in Figure 5, *B* does display some interesting features. There was no strong evidence for secondary structure formation by either the (cholesterol-$(K)_{20}$DEVDGC)$_3$-trimaleimide or (cholesterol-$(D)_{20}$DEVDGC)$_3$-trimaleimide chains. While water did appear to be largely excluded from the chain contacts, there were visible cavities that appeared large enough to contain small molecules. An enlarged view of a section of the aggregate is display in Figure 6. Here one can see a preference of Asp side chains, over Lys side chains, for the surface. There was significant aggregation of cholesterol molecules to form stacked structures. However, these do not appear to be large enough to hold the aggregate together. Rather, electrostatic interactions appeared to be the main stabilizing force. The Asp-Lys side chain coordination numbers were determined to be 2.8 for the intermolecular contacts out to a distance of 0.7 nm.

Uptake of PKH26-containing nanosponges by RAW264.7 cells

Cell loading of the PKH26 entrapped nanosponges was tested on RAW264.7 monocyte/macrophage-like cells. This cell type

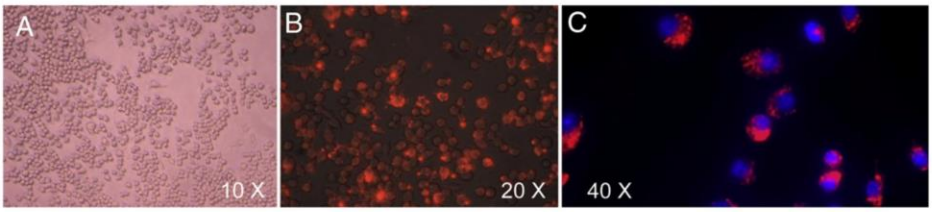


Figure 7. (**A**) RAW264.7 cells (control). (**B**) Fluorescent microscope image (taken with TRITC filter) of RAW264.7 cells after 2 h of incubation with 50 μM of PKH26-containing type DK nanosponges. (**C**) Fluorescence microscopy overlay of RAW264.7 cells featuring PKH26-containing nanosponges 72 h after uptake (image taken with TRITC filter) followed by a DAPI counterstain (image taken with UV filter).

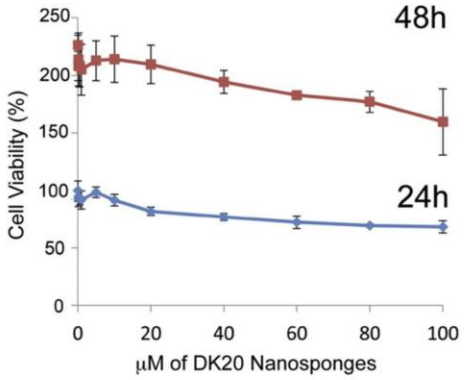


Figure 8. Cell viability of RAW264.7 cells as a function of type DK20 nanosponge concentration and incubation time (24 h and 48 h), as measured by the MTT assay.[35,55–57] Nanosponges were added to the cell culture medium in their respective concentrations (see experimental section).

was selected, because it can be used as carrier in cell-mediated cancer therapy.[55–57] Results indicated that type DK20 nanosponges can be loaded into RAW264.7 within 2 h. Under the fluorescence microscope, intensive red fluorescence spots can be discerned inside the cells, which are very different from labeling cells with free PKH26, which leads to uniform labeling (see SI). This indicates that after being taken up by these cells, the PKH26 is still entrapped inside the nanosponges. The PKH26-containing type DK20 nanosponges in Raw264.7 cells were studied over 24 h, 48 h, and 72 h. Virtually no leaching of the dye was observed by fluorescence microscopy within 72 h (Figure 7).

Cell toxicity of the peptide nanosponges

We have performed classic MTT cell proliferation assays[35,55–57] to measure the cell viability of RAW264.7 cells after incubation with PKH26-containing type DK20 nanosponges. As Figure 8 indicates, the type DK20 nanosponges are essentially not toxic to monocyte/macrophage-like cells, even at 100 μM concentration.

Discussion

A good targeted drug delivery system should have the following characteristics: a) composed by biocompatible and biodegradable materials, b) fast assembly and cargo loading, c) minimal systemic leaking during delivery, and d) fast release upon arrival at interested site.[58] We have designed (cholesterol-$(K)_n$ $DEVDGC)_3$-trimaleimide and (cholesterol-$(D)_n$$DEVDGC)_3$-trimaleimide units that both feature a trigonal linker, a cleavable sequence designed for executioner caspases-3,6, and 7 (DEVDGC[32]) and either an oligo-lysine or oligo-aspartic acid sequence of variable length (n = 5, 10, 15, 20). Whereas the cysteine at the C-terminus of each oligopeptide is used to attach it to the trigonal linker via Michael addition to maleimide,[31] the N-terminus is tethered to cholesterol, which has the function of a hydrophobic anchor. As Coarse-Grained Molecular Dynamics simulations indicate, a sponge-like dynamic structure is spontaneously assembled, due to the formation of ion pairs, intense hydrogen bonding, and the occurrence of hydrophobic regions and water-filled nanocavities. For nanosponges from (cholesterol-$(K)_{20}$$DEVDGC)_3$-trimaleimide/(cholesterol-$(D)_{20}$-$DEVDGC)_3$-trimaleimide units, AFM, TEM and DLS are in principal agreement about the diameter of the nanosponges. Furthermore, all three methods indicate the highly dynamic nature of the sponge-like aggregates, which is in very good agreement with the results obtained from Coarse-Grained Molecular Dynamics. The calculation discussed here predicts the relative enrichment of aspartate units at the nanosponges' surfaces. We can observe distinctly stronger staining at the exterior by uranyl acetate of the nanostructures observed in TEM. This is in perfect agreement with the predictions by modeling. The main difference between the TEM results and the principal outcome of the Molecular Dynamics Simulations is that the structures observed by TEM are spherical, whereas the simulated structures are not. The most probable cause for this discrepancy is that the TEM images were recorded on carbon-coated surfaces, which are very hydrophobic. Consequently, the cholesterol units are oriented towards the surface, causing a collapse of the 3D structure into a 2D coating. This effect is even more pronounced in high vacuum, which leads to a (partial) desiccation of the structure. Contrary to TEM, the structures observed by AFM resemble the results from Molecular Dynamics Simulations much closer.

The size of the nanosponges can be adjusted from several micrometers down to approx. 80 nm in diameter, depending on the concentration and the chemical composition (especially chain-length of the monopeptides (D or K)) of the supramolecular building blocks. The resulting nanosponges can be generated by simply mixing their components in aqueous buffer.

They are stable in size for up to 72 h. Therefore, in distinct contrast to classic liposomes, numerous applications can be envisioned in which the nanosponge will be long-term stored in desiccated form and mixed with aqueous buffer immediately before clinical use.

In recent years, cytotherapy has attracted massive attention as a targeting cancer therapy. Cytotherapy uses cancer targeting cells as delivery vehicles to carry therapeutics into the tumor site.[18,30,35,55–57,59,60] Studies have shown that macrophages are capable of delivering therapeutics to tumors sites.[55–57] We envision entrapping therapeutics into these nanovesicles, and then loading them into transport cells to achieve targeted delivery. However, the nanosponges themselves have to be non-toxic to ensure high survival rates during transport. Therefore, it is encouraging that virtually no toxic effects have been found during our initial cell proliferation tests with monocyte-macrophage-like cells.

Summary

Nanosponges form spontaneously by mixing two trimeric peptide building blocks, (cholesterol-$(K)_n$DEVDGC)$_3$-trimaleimide and (cholesterol-$(D)_n$DEVDGC)$_3$-trimaleimide (n = 5, 10, 15, 20) in aqueous buffers. The resulting sponge-like supramolecular aggregates are long-term stable and do not significantly change their diameter within 72 h. Their structure was elucidated with the help of Coarse-Grained Molecular Dynamics. Since the nanosponges were virtually non-toxic in cell experiments with monocyte/macrophage-like cells (RAW264.7 cells), they are promising candidates for drug-delivery to transporting cells in cytotherapy of solid tumors (leucocytes or stem cells). The fundamental features of this novel and structurally unique supramolecular system have been elucidated in this initial study. In further studies, we will investigate the suitability and adaptability of this system for tailored applications in targeted cancer therapy.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.nano.2017.07.004.

References

1. Cassidy J, Schatzlein AG. Tumour-targeted drug and gene delivery: principles and concepts. *Expert Rev Mol Med* 2004;**6**(19):1-17.
2. Hamad I, Moghimi SM. Critical issues in site-specific targeting of solid tumours: the carrier, the tumour barriers and the bioavailable drug. *Expert Opin Drug Delivery* 2008;**5**(2):205-19.
3. Moses MA, Brem H, Langer R. Advancing the field of drug delivery: taking aim at cancer. *Cancer Cell* 2003;**4**(5):337-41.
4. Allen TM, Cullis PR. Drug delivery systems: entering the mainstream. *Science* 2004;**303**(5665):1818-22.
5. Tiwari G, Tiwari R, Sriwastawa B, Bhati L, Pandey S, Pandey P, et al. Drug delivery systems: an updated review. *Pharm Invest* 2012;**2**(1):1-11.
6. Devadasu VR, Bhardwaj V, Kumar MNVR. Can controversial nanotechnology promise drug delivery? *Chem Rev* 2013;**113**(3):1686-735.
7. Sawant RR, Torchilin VP. Liposomes as smart' pharmaceutical nanocarriers. *Soft Matter* 2010;**6**(17):4026-44.
8. Yapa AS, Bossmann SH. Development of magnetic theranostic agents. In: Bossmann SH, Wang H, editors. *Magnetic Nanomaterials: Applications in Catalysis and Life Sciences*. London: The Royal Society of Chemistry; 2017.
9. Yu M, Zheng J. Clearance pathways and tumor targeting of imaging nanoparticles. *ACS Nano* 2015;**9**(7):6655-74.
10. Nakamura Y, Mochida A, Choyke PL, Kobayashi H. Nanodrug delivery: is the enhanced permeability and retention effect sufficient for curing cancer? *Bioconjug Chem* 2016;**27**(10):2225-38.
11. Lytton-Jean AKR, Kauffman KJ, Kaczmarek JC, Langer R. Cancer nanotherapeutics in clinical trials. *Cancer Treat Res* 2015;**166**:293-322 [Nanotechnology-Based Precision Tools for the Detection and Treatment of Cancer].
12. Colletier J-P, Chaize B, Winterhalter M, Fournier D. Protein encapsulation in liposomes: efficiency depends on interactions between protein and phospholipid bilayer. *BMC Biotechnol* 2002;**2**:9.
13. Glavas-dodov M, Fredro-kumbaradzi E, Goracinova K, Calis S, Simonoska M, Hincal AA. 5-Fluorouracil in topical liposome gels for anticancer treatment - formulation and evaluation. *Acta Pharm* 2003;**53**(4):241-50.
14. Templeton NS, Lasic DD, Frederik PM, Strey HH, Roberts DD, Pavlakis GN. Improved DNA: liposome complexes for increased systemic delivery and gene expression. *Nat Biotechnol* 1997;**15**(7):647-52.
15. Yatvin MB, Weinstein JN, Dennis WH, Blumenthal R. Design of liposomes for enhanced local release of drugs by hyperthermia. *Science* 1978;**202**(4374):1290-3.
16. Mayer LD, Hope MJ, Cullis PR. Vesicles of variable sizes produced by a rapid extrusion procedure. *Biochim Biophys Acta* 1986;**858**(1):161-8.
17. Olson F, Hunt CA, Szoka FC, Vail WJ, Papahadjopoulos D. Preparation of liposomes of defined size distribution by extrusion through polycarbonate membranes. *Biochim Biophys Acta* 1979;**557**(1):9-23.
18. Basel MT, Shrestha TB, Troyer DL, Bossmann SH. Protease-sensitive, polymer-caged liposomes: a method for making highly targeted liposomes using triggered release. *ACS Nano* 2011;**5**(3):2162-75.
19. Gabizon AA, Shmeeda H, Zalipsky S. Pros and cons of the liposome platform in cancer drug targeting. *J Liposome Res* 2006;**16**(3):175-83.
20. van Hell AJ, Costa CICA, Flesch FM, Sutter M, Jiskoot W, Crommelin DJA, et al. Self-assembly of recombinant amphiphilic oligopeptides into vesicles. *Biomacromolecules* 2007;**8**(9):2753-61.
21. Holowka EP, Pochan DJ, Deming TJ. Charged polypeptide vesicles with controllable diameter. *J Am Chem Soc* 2005;**127**(35):12423-8.
22. Sukthankar P, Gudlur S, Avila LA, Whitaker SK, Katz BB, Hiromasa Y, et al. Branched oligopeptides form nanocapsules with lipid vesicle characteristics. *Langmuir* 2013;**29**(47):14648-54.
23. Sukthankar P, Whitaker SK, Garcia M, Herrera A, Boatwright M, Prakash O, et al. Thermally induced conformational transitions in nascent branched amphiphilic peptide capsules. *Langmuir* 2015;**31**(10):2946-55.
24. Avila LA, Aps LRMM, Sukthankar P, Ploscariu N, Gudlur S, Simo L, et al. Branched amphiphilic cationic oligopeptides form peptiplexes with DNA: a study of their biophysical properties and transfection efficiency. *Mol Pharm* 2015;**12**(3):706-15.
25. Jia Z, Whitaker SK, Tomich JM, Chen J. Organization and structure of branched amphipathic oligopeptide bilayers. *Langmuir* 2016;**32**(38):9883-91.
26. Danhier F. To exploit the tumor microenvironment: since the EPR effect fails in the clinic, what is the future of nanomedicine? *J Controll Release* 2016;**244**(Part_A):108-21.
27. Si J, Shao S, Shen Y, Wang K. Macrophages as active nanocarriers for targeted early and adjuvant cancer chemotherapy. *Small* 2016;**12**(37):5108-19.
28. Rachakatla RS, Balivada S, Seo G-M, Myers CB, Wang H, Samarakoon TN, et al. Attenuation of mouse melanoma by a/C magnetic field after delivery of bi-magnetic nanoparticles by neural progenitor cells. *ACS Nano* 2010;**4**(12):7093-104.
29. Heathman TRJ, Nienow AW, McCall MJ, Coopman K, Kara B, Hewitt CJ. The translation of cell-based therapies: clinical landscape and manufacturing challenges. *Regen Med* 2015;**10**(1):49-64.

30. Basel MT, Shrestha TB, Bossmann SH, Troyer DL. Cells as delivery vehicles for cancer therapeutics. *Ther Deliv* 2014;**5**(5):555-67.
31. Fontaine SD, Reid R, Robinson L, Ashley GW, Santi DV. Long-term stabilization of maleimide-thiol conjugates. *Bioconjug Chem* 2015;**26**(1):145-52.
32. Linder M, Tschernig T. Vasculogenic mimicry: possible role of effector caspase-3, caspase-6 and caspase-7. *Ann Anat* 2016;**204**:114-7.
33. Rich DH, Gesellchen PD, Tong A, Cheung A, Buckner CK. Alkylating derivatives of amino acids and peptides. Synthesis of N-maleoylamino acids, [1-(N-maleoylglycyl)cysteinyl]oxytocin, and [1-(N-maleoyl-11-aminoundecanoyl)cysteinyl]oxytocin. Effects on vasopressin-stimulated water loss from isolated toad bladder. *J Med Chem* 1975;**18**(10):1004-10.
34. Pieken, W.; Hill, K.; Eaton, B.; McGee, D.; Vagle, K.; Gold, L.; Stephens, A. Conjugating macromolecules using cycloaddition reactions. US6737236B1, 2004.
35. Wang H, Shrestha TB, Basel MT, Dani RK, Seo G-M, Balivada S, et al. Magnetic-Fe/Fe3O4-nanoparticle-bound SN38 as carboxylesterase-cleavable prodrug for the delivery to tumors within monocytes/macrophages. *Nanotechnol* 2012;**3**:444-55 [12 pp].
36. Coin I, Beyermann M, Bienert M. Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. *Nat Protoc* 2007;**2**(12):3247-56.
37. Cheronis JC, Whalley ET, Nguyen KT, Eubanks SR, Allen LG, Duggan MJ, et al. A new class of bradykinin antagonists: synthesis and in vitro activity of bissuccinimidoalkane peptide dimers. *J Med Chem* 1992;**35**(9):1563-72.
38. Perera AS, Wang H, Basel MT, Pokhrel MR, Gamage PS, Kalita M, et al. Channel blocking of MspA revisited. *Langmuir* 2013;**29**(1):308-15.
39. https://mai.ku.edu/about-mai-lab.
40. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 2013;**29**(7):845-54.
41. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys* 2007;**126**(1):014101.
42. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys* 1984;**81**(8):3684-90.
43. Darden T, York D, Pedersen L. Particle mesh Ewald: an n.Log(N) method for Ewald sums in large systems. *J Chem Phys* 1993;**98**(12):10089-92.
44. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 1997;**18**(12):1463-72.
45. Miyamoto S, Kollman PA. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem* 1992;**13**(8):952-62.
46. Parrinello M, Rahman A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys* 1981;**52**(12):7182-90.
47. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink S-J. The MARTINI coarse-grained force field: extension to proteins. *J Chem Theory Comput* 2008;**4**(5):819-34.
48. Stockert JC, Blazquez-Castro A, Canete M, Horobin RW, Villanueva A. MTT assay for cell viability: intracellular localization of the formazan product is in lipid droplets. *Acta Histochem* 2012;**114**(8):785-96.
49. Bhattacharjee S. DLS and zeta potential — what they are and what they are not? *J Control Release* 2016;**235**:337-51.
50. https://imagej.nih.gov/nih-image/.
51. Chou PY, Scheraga HA. Calorimetric measurement of enthalpy change in the isothermal helix–coil transition of poly-L-lysine in aqueous solution. *Biopolymers* 1971;**10**(4):657-80.
52. Pivcova H, Saudek V. 13 C NMR relaxation study of poly (aspartic acid). *Polymer* 1985;**26**(5):667-72.
53. Saudek V, Stokrova S, Schmidt P. Conformational study of poly(alpha-L-aspartic acid). *Biopolymers* 1982;**21**(6):1011-20.
54. Ismail AA, Mantsch HH. Salt bridge induced changes in the secondary structure of ionic polypeptides. *Biopolymers* 1992;**32**(9):1181-6.
55. Basel MT, Balivada S, Shrestha TB, Seo G-M, Pyle MM, Tamura M, et al. A cell-delivered and cell-activated SN38-dextran prodrug increases survival in a murine disseminated pancreatic cancer model. *Small* 2012;**8**(6):913-20.
56. Seo G-M, Rachakatla RS, Balivada S, Pyle M, Shrestha TB, Basel MT, et al. A self-contained enzyme activating prodrug cytotherapy for preclinical melanoma. *Mol Biol Rep* 2012;**39**(1):157-65.
57. Basel MT, Balivada S, Wang H, Shrestha TB, Seo GM, Pyle M, et al. Cell-delivered magnetic nanoparticles caused hyperthermia-mediated increased survival in a murine pancreatic cancer model. *Nanomedicine* 2012;**7**:297-306.
58. Lehner R, Wang X, Marsch S, Hunziker P. Intelligent nanomaterials for medicine: carrier platforms and targeting strategies in the context of clinical application. *Nanomedicine* 2013;**9**(6):742-57.
59. Shrestha TB, Seo GM, Basel MT, Kalita M, Wang H, Villanueva D, et al. Stem cell-based photodynamic therapy. *Photochem Photobiol Sci* 2012;**11**(7):1251-8.
60. Alshetaiwi HS, Balivada S, Shrestha TB, Pyle M, Basel MT, Bossmann SH, et al. Luminol-based bioluminescence imaging of mouse mammary tumors. *J Photochem Photobiol B* 2013;**127**:223-8.

Appendix B - SLTCAP: A Simple Method for Calculating the Number of Ions Needed for MD Simulation

JCTC Journal of Chemical Theory and Computation

Letter

Cite This: J. Chem. Theory Comput. 2018, 14, 1823−1827

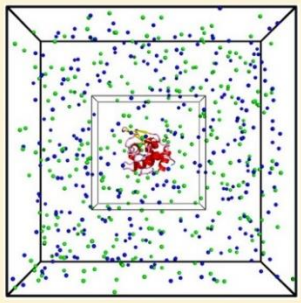
pubs.acs.org/JCTC

SLTCAP: A Simple Method for Calculating the Number of Ions Needed for MD Simulation

Jeremy D. Schmit,*[,†] Nilusha L. Kariyawasam,[‡] Vince Needham,[†] and Paul E. Smith[‡]

[†]Department of Physics and [‡]Department of Chemistry, Kansas State University, Manhattan, Kansas 66506, United States

ABSTRACT: An accurate depiction of electrostatic interactions in molecular dynamics requires the correct number of ions in the simulation box to capture screening effects. However, the number of ions that should be added to the box is seldom given by the bulk salt concentration because a charged biomolecule solute will perturb the local solvent environment. We present a simple method for calculating the number of ions that requires only the total solute charge, solvent volume, and bulk salt concentration as inputs. We show that the most commonly used method for adding salt to a simulation results in an effective salt concentration that is too high. These findings are confirmed using simulations of lysozyme. We have established a web server where these calculations can be readily performed to aid simulation setup.



While electrostatic interactions are indisputably an essential contribution to biomolecule interactions,[1] it still remains challenging to model them theoretically. Continuum models, like Poisson−Boltzmann (PB), provide key intuition about salt screening effects, however, such mean field theories cannot handle temporal or spatial fluctuations. Furthermore, the relevant length scales for biomolecule interactions are often too small to justify the averaging that underlies mean field approaches. Explicit solvent models address precisely these limitations by explicitly representing all charges and allowing mean-field phenomena, like screening, to emerge organically from the microscopic dynamics.

In setting up a simulation, care must be taken in considering the contents of the box. The solvent environment near a solute is perturbed by the presence of the solute and cannot be assumed to be identical to bulk solution. This perturbation comes with significant entropic cost that contributes strongly to the free energy of biomolecule interactions. For example, the reduction in H-bonding partners for water molecules near nonpolar surfaces gives rise to the hydrophobic effect.[2] Similarly, and more to the point of this paper, the presence of charges on the solute distorts the ionic environment surrounding the solute, which results in non-intuitive many-body corrections to the association free energy.[3,4] Therefore, it is important that the solvent composition in the simulation box is representative of the perturbed environment near the solute.

There are two main approaches to computing the number of salt ions that should be added to a simulation (not including grand canonical methods). The first is to simply add enough counterions to achieve a neutral box. Since this method omits co-ions entirely, it is representative of a salt-free solution where the only counterions present are those that dissociated from the solute. The second is to add enough co-ion/counterion pairs to achieve the desired solution salt concentration and then to add enough additional counterions to achieve a neutral box. We refer to this as the add-then-neutralize (AN) method. The problem with this approach is that charged solutes are expected to deplete the surrounding environment of co-ions. Therefore, this method is actually representative of a higher bulk salt concentration than intended.

Here we present a method, screening layer tally by container average potential (SLTCAP), to compute the number of salt ions that should be added to a simulation given an external salt concentration. This method requires no additional information beyond what is needed for AN. To do this, we take advantage of the fact that the ion fluctuations will be less significant when averaging over the entire simulation box. This allows us to employ a mean field formalism to compute the number of ions. The fluctuations will be restored, at least locally, by the ion dynamics in the final simulation.

Consider a solute of charge Q and volume v_p immersed in a simulation box with volume $v_t = v_p + v_w$, where v_w is the volume occupied by solvent. We assume that the simulation box is in equilibrium with a large solvent reservoir with a concentration c_0 of symmetric, monovalent salt. Inside the box, the ion concentrations are perturbed by interactions with the solute charges. The perturbed concentrations are related to the electrostatic potential, ϕ, by

$$c_+(\mathbf{x}) = c_0 e^{-e\phi(\mathbf{x})/k_B T} \tag{1}$$

$$c_-(\mathbf{x}) = c_0 e^{e\phi(\mathbf{x})/k_B T} \tag{2}$$

The total number of ions in the box can be obtained by integrating over the solvent volume:

$$N_\pm = \int_{v_w} c_0 e^{\mp e\phi(\mathbf{x})/k_B T} d^3\mathbf{x} \tag{3}$$

Received: December 14, 2017
Published: March 5, 2018

DOI: 10.1021/acs.jctc.7b01254
J. Chem. Theory Comput. 2018, 14, 1823−1827

where e is the electron charge, k_B is the Boltzmann constant, T is the absolute temperature, and the integral is evaluated over the ion accessible volume. While eq 3 is the optimal method for computing the ion numbers, it requires the cumbersome step of

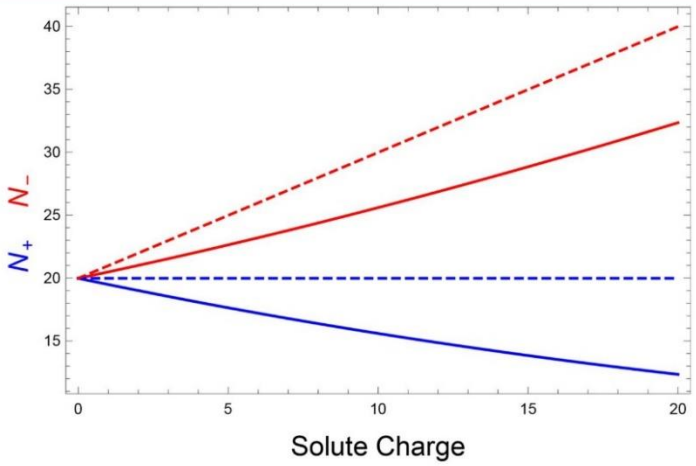


Figure 1. Calculated number of salt ions needed to achieve a bulk concentration of 150 mM in a solvent volume of 221 nm^3. The number of co-ions (cations in this case) required by the AN method (dashed lines) is independent of the solute charge, which is unrealistic because electrostatic repulsion between the solute and co-ions will deplete the local environment of co-ions. The SLTCAP method (solid lines) accounts for co-ion depletion, which also reduces the required number of counterions.

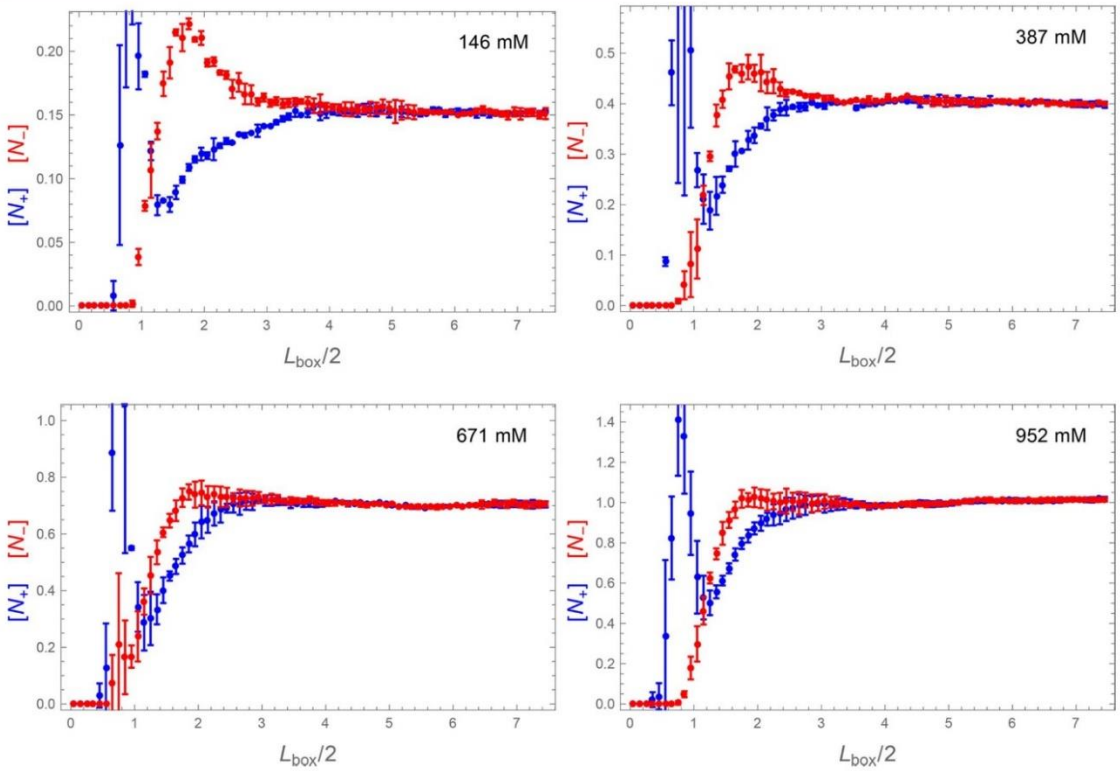


Figure 2. Local ion molality within cubic shells with an edge length L_{box} and a shell thickness 0.1 nm. For L_{box} values somewhat larger than the lysozyme radius ~1.6 nm, there is an enrichment of counterions (red) and depletion of co-ions (blue). At distances much greater than the screening length, the ion concentrations are both equal to the bulk concentration. The high concentration of co-ions at small distances indicates a short-ranged affinity between Na^+ and the protein.

DOI: 10.1021/acs.jctc.7b01254
J. Chem. Theory Comput. 2018, 14, 1823−1827

finding the potential by solving the PB equation. Although excellent tools exist for this,[5,6] this level of effort is unnecessary given that the important features of ion screening will be captured by the ion dynamics in the simulation. Instead, we employ an approximation in which we replace the spatially dependent potential $\phi(\mathbf{x})$ with an average potential $\bar{\phi}$ that is uniform across the simulation box. With this approximation, eq 3 becomes $N_{\pm} = v_w c_0 e^{\mp e\bar{\phi}/k_B T}$. The value of the average potential can be determined from the charge neutrality condition $-Q/e = N_+ - N_-$ (the conditions underlying this assumption are discussed below) which yields

$$\frac{e\bar{\phi}}{k_B T} = \mathrm{ArcSinh}\left(\frac{Q}{2ev_w c_0}\right) \tag{4}$$

The required ion numbers are then given by

$$N_{\pm} = v_w c_0 \exp\left[\mp \mathrm{ArcSinh}\left(\frac{Q}{2ev_w c_0}\right)\right] \tag{5}$$

Thus, it is possible to estimate the required ion numbers knowing only the solute charge, solvent volume, and bulk salt concentration.

Figure 1 plots eq 5 as a function of the protein charge (solid lines). A key feature is that the box is depleted of co-ions in addition to being enriched in counterions. This feature is not present in the AN method (dashed lines) which simply adds counterions on top of ion pairs present at bulk concentration. This means that AN results in the simulation of a system that has an effective salt concentration higher than intended. We can estimate the effective concentration generated by the AN method as follows. Assuming that we have a solute with a positive charge $Q > 0$, the number of ions added to the box under the AN method is

$$\begin{aligned} N_+ &= v_w c_0 \\ N_- &= v_w c_0 + Q \end{aligned} \tag{6}$$

The same number of ions can be obtained if the box is in equilibrium with a solution at an effective salt concentration c_{eff}

$$\begin{aligned} N_+ &= v_w c_{\mathrm{eff}} e^{-e\phi_{\mathrm{eff}}/k_B T} \\ N_- &= v_w c_{\mathrm{eff}} e^{e\phi_{\mathrm{eff}}/k_B T} \end{aligned} \tag{7}$$

where ϕ_{eff} is the average potential inside the box. Eliminating N_+, N_-, and ϕ_{eff} from eqs 6 and 7, we find that the effective salt concentration is

$$c_{\mathrm{eff}} = c_0\sqrt{1 + Q/(ev_w c_0)} \tag{8}$$

We see that the key parameter is the ratio of the solute charge Q/e to the number of ions that would be found in the box with an uncharged solute $v_w c_0$. When this ratio is small, such as when the box is large or the solute charge is small, the difference between AN and SLTCAP is minimal. However, the ratio can be quite large when the system of interest contains highly charged molecules like nucleic acids.

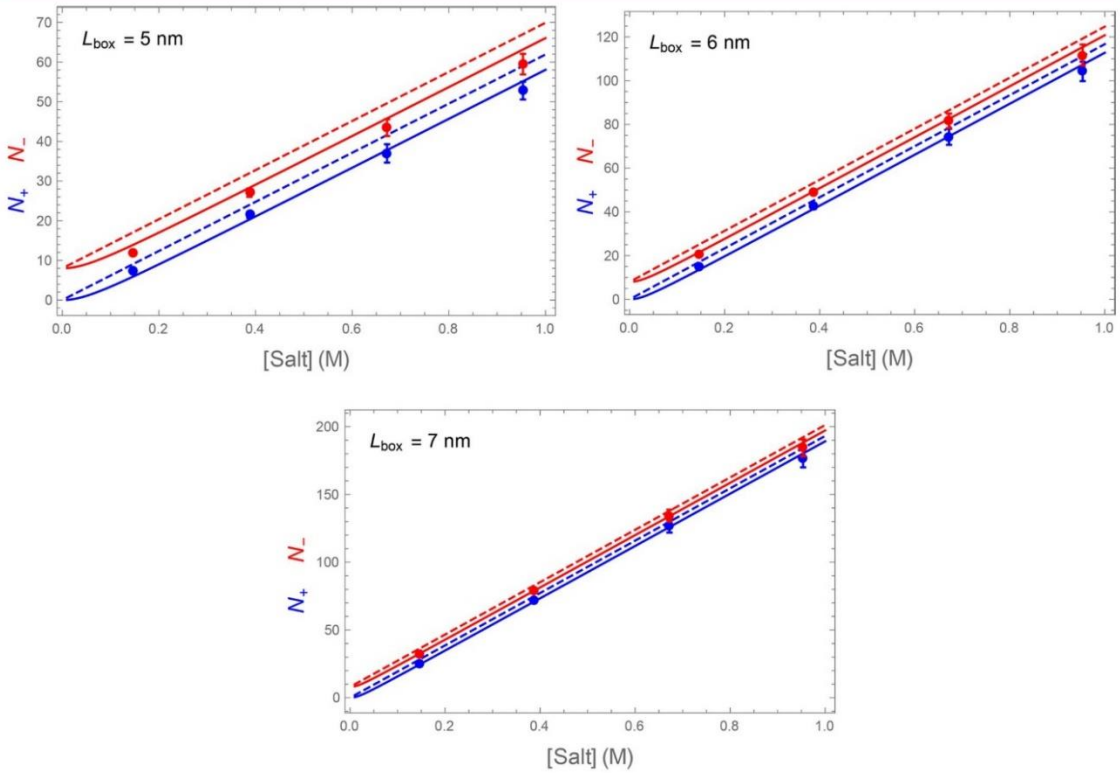


Figure 3. Number of Na^+ (blue dots) and Cl^- (red dots) found within a fictitious box of width L_{box} centered on a lysozyme molecule of charge $Q = +8$. The observed number of ions is consistent with the SLTCAP calculation (solid lines) but is systematically lower than the AN method (dashed lines).

DOI: 10.1021/acs.jctc.7b01254
J. Chem. Theory Comput. 2018, 14, 1823−1827

We compared the ion distributions predicted by AN and SLTCAP using simulations of a model protein. Lysozyme (PDB id: 1AKI) was simulated at 300 K and 1 bar in the presence of four different NaCl concentrations, 0.15, 0.39, 0.67, and 0.95 M. The MD simulations were performed for 40−60 ns at each salt concentration with the Amber 99SB-ILDN force field,[7] as implemented in the Gromacs simulation package,[8] using the particle mesh Ewald (PME) approach to determine the electrostatic interactions.[9] The charge on lysozyme (Q = +8 at pH 7.0) corresponds to the assumption that all the amino acids display standard pK_a values. These simulations were conducted in a large ≃15 nm simulation box where there are almost 6 Debye lengths between the surface of the protein and the edge of the box at the lowest salt concentration investigated. This setup approximates dilute solution with a large reservoir of ions that can relax around the protein. Ions were added to the box according to the AN method due to the fact that the large reservoir makes the system insensitive to this choice. We then count the number of anions and cations contained within a fictitious box of variable dimension L_{box} that is centered around the protein. The protein volume (22.2 nm^3) was determined by plotting the difference between the system volume and the water volume (N_w times molar volume of TIP3P water) as a function of bulk salt molality, followed by an extrapolation to zero salt.

Figure 2 shows the local ion concentrations as a function of distance from the center of the box. Interestingly, we observed large, but highly variable, concentrations of co-ions at distances less than the lysozyme radius of ∼1.6 nm, indicating a short-ranged affinity between the protein and Na^+ ions. At distances somewhat greater than the molecular radius, the system shows a significant depletion of co-ions and enrichment of counterions. This imbalance signifies the presence of the screening layer. The perturbed concentrations decay back to the bulk concentration with a characteristic distance of the screening length.

Figure 3 plots the number of ions contained within the fictitious box as a function of the salt concentration. The solid and dashed lines show the expected number of ions as calculated by SLTCAP and AN, respectively. The simulation data are in good agreement with the SLTCAP simulation and are systematically lower than the AN method. This agreement is despite the fact that the simulation box was prepared using the AN method and the presence of specific cation−protein interactions that are not accounted for by our theory.

Figure 4 plots the number of entrained ions as a function of the fictitious box size. The agreement with SLTCAP is quite good over the plotted range of $L_{box}/2$ = 2.5−4.0 nm. For box sizes above ∼5.0 nm, the number of entrained ions approaches the AN result, which is a consequence of the fact that the simulations were prepared by AN. For smaller box sizes, the number of entrained ions dips below the SLTCAP calculation, and even further below the AN method. The deviation is more pronounced at low salt concentration where the fictitious box passes inside the screening layer. In these cases, charge neutrality is not enforced, which can be seen by the fact that the simulation values for

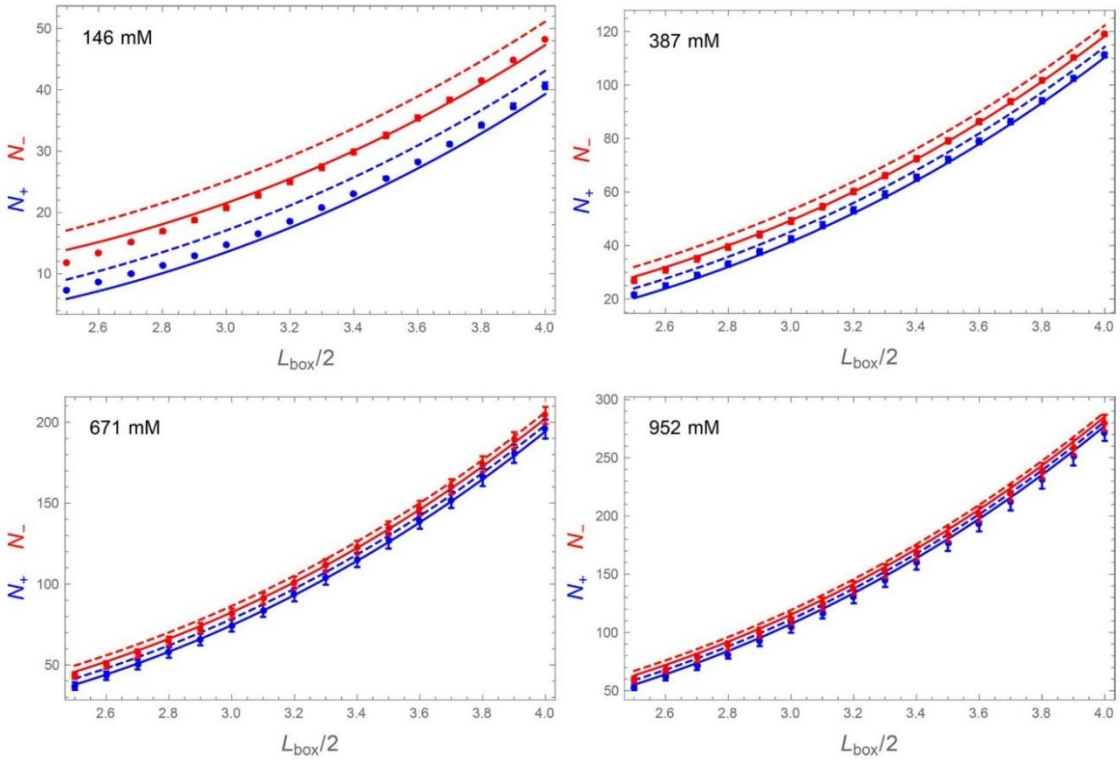


Figure 4. Number of Na^+ (blue dots) and Cl^- (red dots) found within a fictitious box of width L_{box} centered on a lysozyme molecule of charge Q = +8. The SLTCAP and AN methods are shown in solid and dashed lines, respectively. The simulation data agree well with the SLTCAP method except for a small amount of cation adsorption to the protein surface and the small and large box effects discussed in the text.

DOI: 10.1021/acs.jctc.7b01254
J. Chem. Theory Comput. 2018, 14, 1823−1827

N_- and N_+ are closer together than the theoretical values from either SLTCAP or AN (which both enforce neutrality). For these small box sizes, our simulations are not a good test of the theory because the large box size of 15 nm mimics dilute solution, whereas the overlap of screening layers indicates the onset of the concentrated regime.

The charge neutrality requirement should be considered in light of the conditions that are intended to be replicated by the simulations. The periodic boundary conditions employed by most simulations effectively produce a system with a protein concentration v_t^{-1}. Usually this is much higher than the concentration of the system of interest. However, assuming the enhanced periodicity effects due to the use of PME are small, the absence of direct protein−protein interactions closely corresponds to the infinitely dilute protein solute case provided the solution is allowed to relax back to the bulk state between replicas. This relaxation is not possible for highly charged proteins because there is no place for the co-ions to escape from the protein environment. The SLTCAP method allows the simulation to mimic dilute conditions by effectively allowing the ions to exchange with a large buffer reservoir.

When the system of interest is at high concentration, the correct handling of the ionic atmosphere is more complicated. If the solution is formed by dissolving a protein salt, then the dissociated counterions from the protein will add to the buffer salt. This is the situation described by the AN method. However, if the system is allowed to exchange with an external buffer, either in a centrifugal concentrator or by dialysis, then SLTCAP becomes the correct description. Of course, the condition of greatest interest is *in vivo*. While this environment is highly crowded, it is better described by SLTCAP because the heterogeneous environment allows for the depletion of co-ions near a charged solute.

In order to simulate a protein under conditions resembling dilute solution, it is necessary that the potential is small near the edges of the box $\phi(\mathbf{x}) \ll k_B T/e \simeq 25$ mV. Since the potential decays on a length scale set by the Debye screening parameter $\kappa^2 \equiv 2e^2 c_0/(\varepsilon k_B T)$, this condition can be approximated as $\bar{\phi} e^{-\kappa L_{buff}} \ll 25$ mV, where L_{buff} is the distance between the solute and the edge of the box, and ε is the permittivity of water. In practice, a few nanometers is usually sufficient for L_{buff} since $\kappa^{-1} \simeq 1$ nm for 100 mM salt.

We have established a web server that will rapidly perform SLTCAP calculations to facilitate the setup of molecular simulations.[10] The server provides several methods for computing the solvent volume, either directly from the number of water molecules or by estimating the solute volume using an average protein specific volume of 0.72 cm^3/g and subtracting this from the box volume. This estimate of the protein volume will be adequate provided the volume error δv_p is smaller than c_0^{-1}. The server then outputs N_+, N_-, and $\bar{\phi}$. Our hope is that this tool will lead to an improved representation of screening effects in biomolecule simulations.

AUTHOR INFORMATION

Corresponding Author

*E-mail: schmit@phys.ksu.edu.

ORCID

Jeremy D. Schmit: 0000-0002-0104-5468

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by NIH Grants GM107487 (J.D.S.), GM118719 (P.E.S.), and GM118589 (P.E.S.). We would like to thank A. Roitberg for suggesting this project, J. Comer for feedback on the web server, and E. Ploetz for comments on the manuscript.

REFERENCES

(1) Schreiber, G.; Fersht, A. R. Rapid, Electrostatically Assisted Association of Proteins. *Nat. Struct. Biol.* **1996**, *3*, 427−431.

(2) Dill, K. A.; Truskett, T. M.; Vlachy, V.; Hribar-Lee, B. Modeling Water, the Hydrophobic Effect, and Ion Solvation. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 173−199.

(3) Schmit, J. D.; Whitelam, S.; Dill, K. A. Electrostatics and Aggregation: How Charge Can Turn a Crystal Into a Gel. *J. Chem. Phys.* **2011**, *135*, 085103.

(4) Dahal, Y. R.; Schmit, J. D. Ion Specificity and Nonmonotonic Protein Solubility from Salt Entropy. *Biophys. J.* **2018**, *114*, 76−87.

(5) Li, L.; Li, C.; Sarkar, S.; Zhang, J.; Witham, S.; Zhang, Z.; Wang, L.; Smith, N.; Petukh, M.; Alexov, E. DelPhi: A Comprehensive Suite for DelPhi Software and Associated Resources. *BMC Biophys.* **2012**, *5*, 9.

(6) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037−10041.

(7) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1950−1958.

(8) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845−854.

(9) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(10) *Screening Layer Tally by Container Average Potential (SLTCAP)*, https://www.phys.ksu.edu/personal/schmit/SLTCAP/SLTCAP.html.

DOI: 10.1021/acs.jctc.7b01254
J. Chem. Theory Comput. 2018, 14, 1823−1827

Appendix C - Peptide Nanosponges Designed for Rapid Uptake by Leucocytes and Neural Stem Cells

RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2018, **8**, 16052

Peptide nanosponges designed for rapid uptake by leukocytes and neural stem cells†

Asanka S. Yapa,‡[a] Hongwang Wang,‡[a] Sebastian O. Wendel, ‡§[bc] Tej. B. Shrestha,‡[b] Nilusha L. Kariyawasam,[a] Madumali Kalubowilage,[a] Ayomi S. Perera, ¶[a] Marla Pyle,[b] Matthew T. Basel,[b] Aruni P. Malalasekera,‖[a] Harshi Manawadu, **[a] Jing Yu,[a] Yubisela Toledo,[a] Raquel Ortega,[a] Prem S. Thapa,[c] Paul E. Smith,[a] Deryl L. Troyer*[b] and Stefan H. Bossmann *[a]

The structure of novel binary nanosponges consisting of (cholesterol-(K/D)$_n$DEVDGC)$_3$-trimaleimide units possessing a trigonal maleimide linker, to which either lysine (K)$_{20}$ or aspartic acid (D)$_{20}$ are tethered, has been elucidated by means of TEM. A high degree of agreement between these findings and structure predictions through explicit solvent and then coarse-grained molecular dynamics (MD) simulations has been found. Based on the nanosponges' structure and dynamics, caspase-6 mediated release of the model drug 5(6)-carboxyfluorescein has been demonstrated. Furthermore, the binary (DK20) nanosponges have been found to be virtually non-toxic in cultures of neural progenitor cells. It is of a special importance for the future development of cell-based therapies that DK20 nanosponges were taken up efficiently by leucocytes (WBC) in peripheral blood within 3 h of exposure. The percentage of live cells among the WBC was not significantly decreased by the DK20 nanosponges. In contrast to stem cell or leucocyte cell cultures, which have to be matched to the patient, autologous cells are optimal for cell-mediated therapy. Therefore, the nanosponges hold great promise for effective cell-based tumor targeting.

Received 24th January 2018
Accepted 23rd April 2018

DOI: 10.1039/c8ra00717a

rsc.li/rsc-advances

Introduction

One of the grand challenges in nanomedicine is the effective targeting of tumors and metastases.[1] For almost a generation, Enhanced Permeation and Retention (EPR),[2,3] the passive diffusion of nanosize delivery vehicles (*e.g.* vesicles,[4] liposomes,[5] exosomes,[6] nanoparticles,[7] polymer-based nanostructures[8–10]) through gaps in the vasculature that have been built rapidly around tumor tissue, has been hailed as an important breakthrough in the fight against cancer. Unfortunately, emerging evidence clearly suggests that the EPR effect works well in rodent models (especially in nude mice), but not in humans, who feature distinctly different vasculature and, compared to rodents, significantly slower tumor growth.[2,3,11] Therefore, alternative targeting approaches are urgently needed. Active targeting strategies use either antibodies,[12–14] antibody-fragments,[15–17] peptide sequences[15,16,18] or aptamers,[15,16] which are capable of targeting receptors that are overexpressed in solid tumors, as for instance members of the integrin family.[19,20] However, active targeting processes can be impaired by physiological barriers, such as high interstitial fluid pressures and the formidable physical barrier imposed by tumor stroma.[21] Therefore, cell-mediated transport of anticancer drugs into the tumor tissue is, in the opinion of the authors, the most viable strategy to develop intelligent alternatives to chemotherapy.[10,22–27] Transport cells have the ability to migrate to tumors and metastases following cytokine/chemokine gradients.[28] Among them are stem cells,[29] monocytes/macrophages[30,31] and neutrophils.[32,33] Neural stem cells, which can be, principally, cultured and matched to patient-types, have been successfully utilized for cell-mediated therapies in rodent models,[22,27,34] as well as neutrophils[35,36] and monocytes.[10,23,24,27]

[a]*Department of Chemistry, Kansas State University, Manhattan, KS 66506, USA. E-mail: sbossman@ksu.edu; Fax: +1-785-532-6666; Tel: +1-785-532-6817*

[b]*Department of Anatomy & Physiology, Kansas State University, Manhattan, KS, USA. E-mail: troyer@vet.k-state.edu; Tel: +1-785-532-6405*

[c]*Microscopy and Analytical Imaging Laboratory, University of Kansas, Lawrence, KS, USA*

† Electronic supplementary information (ESI) available: Supporting Information is available free of charge on the RSC Advances website. It contains the chemical structure of 5(6) carboxy-fluorescein (CF), as well as its excitation and emission spectrum, the calibration curve for measuring CF fluorescence *vs.* concentration, and the DLS characterization of the DK20 nanosponges. An estimation of the number of CF molecules per nanosponge is provided as well. See DOI: 10.1039/c8ra00717a

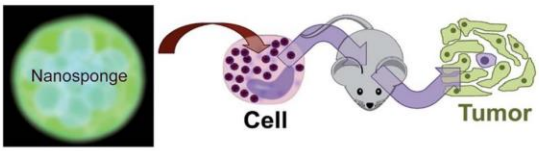
‡ These authors contributed equally.

§ Sebastian O. Wendel, Department of Biology, Kansas State University, KS, USA.

¶ Ayomi S. Perera, Centre for Nature Inspired Engineering, Chemical Engineering, University College London, Torrington Place, London WC1E 7JE, UK.

‖ Aruni P. Malalasekera, Department of Chemistry, Southwestern College, Winfield, KS, USA.

** Harshi Manawadu, University of Ruhuna, Department of Chemistry, Matara, Sri Lanka.



Scheme 1 Principles of cell therapy utilizing nanosponges. Neutrophils in peripheral blood will be loaded by targeting them with peptide nanosponges. After the blood has been given back intravenously to the patient, it is anticipated that the neutrophils will home to tumors within 6–12 h. Alternatively, neural stem cells can be cultured, loaded with peptide nanosponges and injected intravenously into the patient.

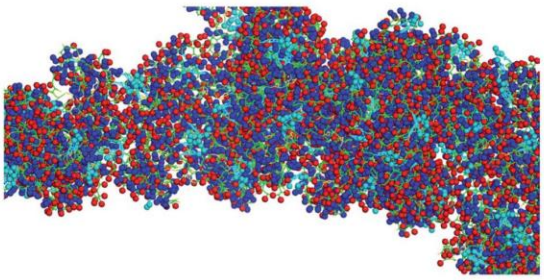


Fig. 1 Typical structure of a nanosponge according to coarse grained molecular dynamics simulations, which are described in detail in ref. 38. Red: aspartate groups, blue: lysine groups, cyan: cholesterol aggregates, green: peptide backbone.

The use of autologous cells has the potential of developing truly patient specific therapies, and also of significantly lowering the regulatory barriers for cell-based human cancer therapies.[37] Targeting neutrophils and monocytes in peripheral blood will avoid the necessity for their time-consuming isolation and culturing, and further reduce the regulatory hurdles since cell isolation is not necessary. In Scheme 1, the principles of cell-based cancer therapy are shown. In step 1, the selected transport cell type is targeted. In order to maintain high transport cell viabilities, carrier particle (vector) recognition and uptake by the transport cell(s) has to be very efficient. Furthermore, the vector that is used to facilitate uptake has to be virtually non-toxic. After the cells have been returned to the host, they actively migrate to tumors and metastases following cytokine/chemokine gradients. The last step consists in the triggered release of the payload and uptake of the latter by the tumor and stromal cells.[27]

This team has recently reported the design, synthesis, and characterization of designer peptide-nanosponges for efficient uptake by delivery cells in drug delivery.[38] Their supramolecular building blocks consist of (cholesterol-$(K/D)_n$DEVDGC)$_3$-trimaleimide units featuring a trigonal maleimide linker to which either lysine $(K)_{20}$ or aspartic acid $(D)_{20}$ are attached. Furthermore, a consensus sequence for caspase-6 (DEVD-GC)[39] is integrated into the structures. This consensus sequence can also be activated by other executioner caspases, such as caspase-3, and -7. There is emerging evidence that caspases can actively contribute to the development and progression of tumors.[40] This is in agreement with clinical evidence for the presence of active caspases in tumors.[40] Caspases −3 and −6 are taken up by cells and are, therefore, suitable to cleave the consensus sequences of nanosponges that have been taken up by transport cells, thus triggering their release by means of apoptotic processes, which enhance the porosity of the transport cells and then dissect them into apoptotic bodies.[41] Both, (cholesterol-$(K)_{20}$DEVDGC)$_3$-trimaleimide and mixtures of (cholesterol-$(K)_{20}$DEVDGC)$_3$-trimaleimide and (cholesterol-$(D)_{20}$DEVDGC)$_3$-trimaleimides form stable nanosponges (short notation: DK20). The structure of the novel nanosponges was investigated through explicit solvent and then coarse-grained molecular dynamics (MD) simulations. As Fig. 1 indicates, the nanosponge structure is featuring aspartate- and lysine-rich regions, together with cholesterol domains and (aqueous) solvent filled nanoholes. The resulting structure is fluctuating, depending on the temperature. Upon mixing with aqueous buffers DK20, nanosponges are immediately formed. They possess very low polydispersities and are long-term stable (up to 72 h as experimentally determined). They are capable of incorporating fluorescent dyes (*e.g.* carboxyfluorescein (discussed here) or PKH26 (ref. 38)), which can be used for fluorescence imaging and payload release studies.

It is noteworthy that the Coarse Grained MD Simulations did not result in spherical nanosponges due to the limited number

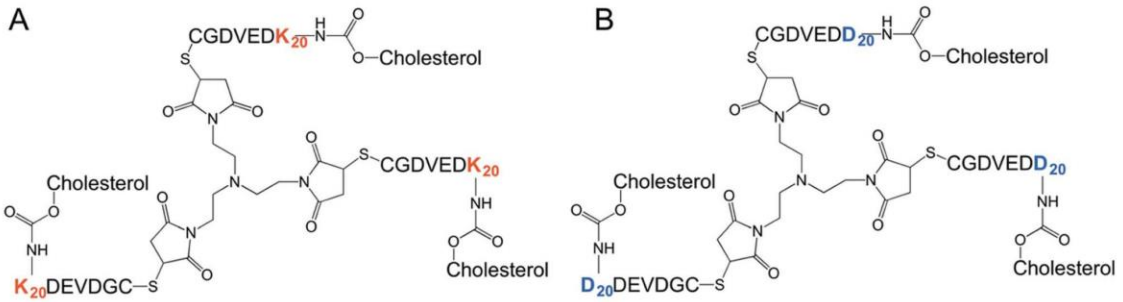


Fig. 2 Tri-maleimide based peptide structures: components for the spontaneous formation of type DK20 nanosponges. (A) Lysine-based component K20 (MW = 11 334.74 g mol^{-1}); (B) aspartic acid-based component D (MW = 10 439.51 mol^{-1}).

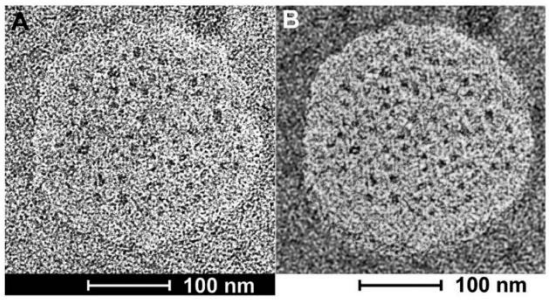


Fig. 3 (A) TEM image of type DK20 nanosponges on HOPG, as deposited from PBS solution. Bright field transmission TEM (200 kV) of a type DK20 nanosponge. Water-filled vesicles are discernible as dark spots within the bright nanosponge. (B) Same image as in 3A after black/white correction filter function in Photoshop.

of small trigonal units in these simulations. However, as shown below in Fig. 3, the principal findings from these simulations were corroborated by means of TEM.

In order to function properly, the nanosponges should be taken up quickly by the transport cells, transported to the tumor sites, and then released. The latter will be achieved by means of programmed cell death (apoptosis), which will occur naturally in neutrophils 12 to 24 h after reaching the tumor environment,[42] or by means of triggered apoptosis (other leucocytes and neural stem cells[27]). Caspase activation is the hallmark of apoptosis.[43] We will utilize caspases, which are proteolytic enzymes, to activate the nanosponges for drug delivery purposes.

In this report, we will describe refined structural investigations by TEM, *in vitro* release studies of the model drug carboxyfluorescein by caspase-6 activation, as well as cell targeting experiments of cultured neural stem cells and leucocytes in peripheral (bovine) blood. The data obtained from these experiments will demonstrate the unique properties of type DK20 nanosponges.

Experimental

Synthesis and characterization of the nanosponges

The synthesis of all building blocks required for the assembly of DK20 and K20 nanosponges, as well as their characterization by NMR and MALDI-TOF has been described in an earlier report (Fig. 2).[38]

TEM characterization

Samples for transmission electron microscopy (TEM) were prepared by dropping 10 μL of 0.050 mM type DK20 solution in PBS directly on a glow discharged TEM grid. Uranyl acetate was used as a positive staining agent in all TEM experiments. In all cases electron microscopy was performed at an accelerating voltage of 200 kV. Nanosponge morphology on HOPG was examined by bright-field and dark-field transmission electron microscopy (TEM) using a FEI Technai G_2 transmission electron microscope at an electron acceleration voltage of 200 kV. Dark-field TEM did not reveal a characteristic diffraction pattern. High resolution images were captured using a standardized, normative electron dose and a constant defocus value from the carbon-coated surfaces. All TEM measurements were performed at the Microscopy and Analytical Imaging Laboratory of the University of Kansas.[44]

Nanosponge formation and DLS characterization

The hydrodynamic diameter and polydispersity index (PDI) of the formed nanosponges were measured by dynamic light scattering (DLS, ZetaPALS, Brookhaven Instruments Corp., Holtsville, NY). All measurements were carried out at 25 °C, with 658 nm laser wavelength, and 90 degree detection angle. Data were collected from an average of three measurements over 60 seconds.

Carboxyfluorescein encapsulation

Equal molar ratios of (cholesterol-$(K)_{20}$DEVDGC)$_3$-trimaleimide and (cholesterol-$(D)_{20}$DEVDGC)$_3$-trimaleimide (5.0×10^{-4} M of each component) were dissolved in 10 μM carboxyfluorescein PBS (pH = 7.4) solution. After incubating at room temperature for 2 hours, the solution was transferred to a 3500 Da molecular weight cutoff dialysis bag. Free carboxy-fluorescein was removed by means of continuous dialysis against 1 × PBS buffer until virtually no fluorescence could be detected in the solution using a Fluoromax-2 spectrometer. Using a fluorescence calibration curve, it was estimated that the concentration of free carboxyfluorescein was < 1 nM. At this point, a dark red color was still retained inside the dialysis bag. This finding provided a good indication that carboxyfluorescein had been trapped inside the peptide nanosponges. From the integrated UV/Vis-absorption of the dialysis solution we have estimated that 65 ± 4 mol% of carboxyfluorescein was encapsulated in the procedure. After lyophilizing to dryness, a yellow/brown powder was obtained, which could be easily re-dispersed in PBS by vortexing for 5 min. In a subsequent dialysis experiment, it was found that virtually no carboxyfluorescein was leached after 24, 48, and 72 h. The UV/Vis and fluorescence spectra of carboxyfluorescein, as well as the fluorescence calibration curve as a function of carboxyfluorescein concentration can be found in the ESI section (Fig. S1, S2, and S3†). The average number of encapsulated carboxyfluorescein molecules per DK20 nanosponge was estimated to 8.5 ± 2.

Caspase-6 triggered dye release

The dye release experiment was performed using a fluorescence plate reader (BioTek Synergy H1). 200 μL of carboxyfluorescein loaded nanosponges in PBS solution (0.20 mg mL^{-1}) were added to each well of a 96-well black clear-bottom plate. To each control well, 10 μL of PBS buffer was added, and to each experimental well, 10 μL of caspase-6 PBS solution (0.1 μg mL^{-1}, 2.2×10^{-7} M, Enzo LifeSciences) was added.

The plate was incubated at 37 °C, the fluorescence intensity at 520 nm was recorded every 5 min.

Cell experiments and MTT assays

The cytotoxicity of the PKH26 containing nanosponges was assessed by utilizing the MTT assay[45] on C17.2 neural progenitor cells (NPCs),[34] which were a gift from Dr. V. Ourednik (Iowa State University) to Dr. D. L. Troyer, DVM (Kansas State University, Anatomy & Physiology). NPCs were originally developed by Dr. Evan Snyder.[46] These cells were maintained in DMEM supplemented with 10% FBS (Sigma-Aldrich), 5% horse serum (Invitrogen), 1% glutamine (Invitrogen), and 1% penicillin/streptomycin (Invitrogen). PKH26 is a hydrophobically modified cyanine 3.0 dye. The preparation of PHK26-loaded type DK20 nanosponges was described earlier.[38] Cell experiments were carried out in the culturing medium described above. The percentage of viable cells was determined after 24 and 48 hours of incubation. Cells were seeded in T-25 flask. After 24 h of incubation at 37 °C, cells were re-plated in a 96 well plate at 20 000 cm^{-2} density and further incubated for 24 h at 37 °C to obtain 80% confluency before the nanosponges were added.

Concentration series of type DK20 nanosponges (0.0, 0.1, 0.2, 0.5, 1.0, 2.0, 5.0, 10, 20, 40, 60, 80, 100 μmol L^{-1} in total, molar ratio 1 : 1) were prepared by dissolving the nanosponge components in the same media that were used for culturing the cells. Cells were incubated for 24/48 h at 37 °C. Eight replicates were prepared for each concentration. A portion of 10 μL of MTT reagent (5 mg mL^{-1} in PBS) was added to each well, and the plates were incubated for another 4 h at 37 °C. Finally, 100 μL of 10% sodium dodecyl sulfate in 0.010 M HCl was added into each well and incubated for 24 h at 37 °C. Their absorbance was recorded by using a plate reader at 550 nm and 690 nm. PBS solution was used as control for all the experiments.

Murine stem cells were imaged by using a Zeiss, Axiovert 40 CFL microscope with darkfield, brightfield, phase contrast and epifluorescence illumination, a camera system and Jenoptik, ProgRes C3 Cool camera and a ProgRes Capture Pro 2.10.0.0 software.

Cell uptake from peripheral blood

Bovine blood was obtained at the Kansas state feed lot. Blood was collected in citrated (0.105 M) 4.5 ml tubes (BD Vacutainer, Franklin Lakes, NJ, USA). The collected blood was pooled and split into 3.0 ml samples. Samples were supplemented with 1.0 mL of serum free RPMI medium to ensure supply of nutrients. The samples were incubated with 1.0 mL of 1.0 mg mL^{-1} type DK20 nanosponges in PBS at 37 °C. Leukocytes (WBC) were extracted *via* removal of the buffy coat after centrifugation.[47,48] Red Blood Cell Lysis Buffer (Sigma-Aldrich, St. Louis, MO, USA) was used to remove any remaining red blood cells and the samples were washed with PBS (10 min, 500 g). Cells were counted *via* hemocytometer and diluted to achieve a concentration of 5×10^5 cells per mL, suitable for analysis by flow cytometry (Guava EasyCyte, EMD Millipore). The survival of WBC incubated type DK20 nanosponges was detected with a annexin V/propidium iodine apoptosis kit (Novus Biologicals). The protocol provided with the kit was exactly followed.

Results

TEM-analysis of the nanosponge structure

Bright field transmission electron microscopy was able to reveal structures that are formed after depositing the nanosponges directly onto HOPG grids and exposing them to the high vacuum inside the TEM. The flattened nanosponges contain dark spots, which are indicatives of water/buffer-filled pockets inside the structure. Furthermore, after applying a black/white correction filter function available in Adobe Photoshop, brighter than average spots can be discerned within the nanosponge structure, which are indicative of cholesterol-rich regions. The average grey within the structure shown on Fig. 3B suggests the presence of both, lysine and aspartate-rich regions, which retain some of their water-content in high vacuum. These findings are in excellent agreement with the principal results of the molecular dynamics (MD) simulations of nanosponge structure.

It is noteworthy that the nanosponges, obtained under the experimental conditions described here, appear to be larger (240 ± 30 nm in diameter), whereas their diameter reported

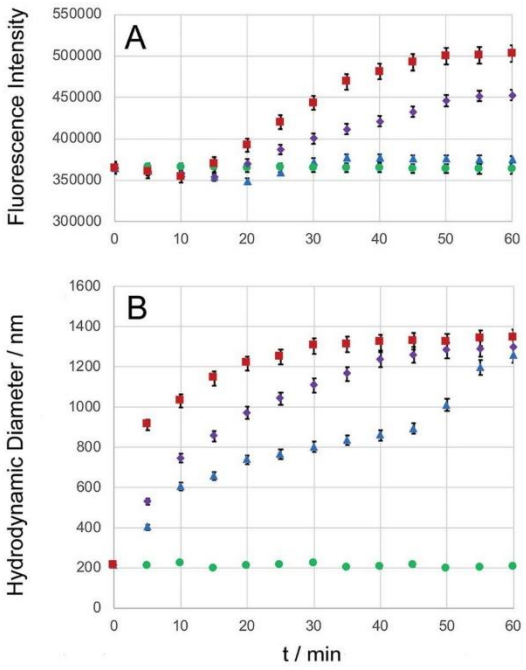


Fig. 4 Caspase-6 triggered carboxyfluorescein (CF) release. Type DK20 nanosponges containing carboxyfluorescein (0.0005 M of D20 and K20, see ESI†) were incubated with PBS (control, green dots), as well as 1.0×10^{-8} M caspase-6 in PBS (blue triangles), 1.0×10^{-7} M caspase-6 in PBS (purple diamonds), and 1.0×10^{-6} M caspase-6 in PBS (red squares) at 37 °C (pH = 7.4). (A) The observed fluorescence emission intensity (relative units) occurring from carboxyfluorescein was recorded at 513 nm with a 5 nm bandpass filter, $\lambda_{exc} = 493$ nm, as a function of time. (B) Hydrodynamic diameters (measured by means of DLS) *vs.* time. Experimental errors are shown for all cases where they extend beyond the size of the symbols.

earlier was 90 ± 15 nm.[38] However, a DK20 concentration approx. Five times lower (than reported in ref. 38).

Caspase-6 triggered carboxyfluorescein release

Carboxyfluorescein is a fluorescent dye, which has been used for microscopy and cell-tracking purposes. It is established that at increased concentrations, carboxyfluorescein undergoes intense self-quenching.[9,49] A detailed investigation of the concentration-dependent quenching of carboxyfluorescein in liposomes revealed both, monomer–monomer and monomer–dimer energy transfer processes.[50] Carboxyfluorescein dimers are non-fluorescent. Because carboxyfluorescein fluorescence can increase as a function of decreasing dye concentration, it has become a popular probe detecting drug release from a delivery system.[9,51,52] After entrapping carboxyfluorescein into type DK20 peptide nanosponges and subsequent lyophilizing to a powder, the obtained solid was dissolved in 3.0 mL PBS buffer (pH = 7.4). DLS measurements showed that the hydrodynamic diameter was 213 ± 25 nm before adding caspase-6 (see Fig. S4†). The nanosponges remained stable in PBS over a 24 h period. Dye release experiments using three concentrations of capase-6 are shown in Fig. 4A, together with the hydrodynamic diameters of the nanosponges, observed by means of DLS (Fig. 4B). The hydrodynamic diameters of the nanosponges steadily increased in the presence of caspase 6, indicating cleavage of the caspase-6 cleavage site DEVDGC,[39] and subsequent formation of new micro-sized structures. At all three investigated concentrations of caspase-6, a maximum hydrodynamic diameter of 1300 ± 50 nm was detected my means of DLS, albeit at different reaction times. The observed fluorescence intensities of carboxyfluorescein were clearly a function of the added caspase-6 concentration. When 1.0×10^{-8} M of caspase-6 was added, the fluorescence intensity decreased during the first 25 min., which was followed by an increase. After 35 min. A plateau was reached, which corresponded to an overall fluorescence increase of 3.7 ± 0.1%. When 1.0×10^{-7} M of caspase-6 was added, the observed decrease in fluorescence was smaller, and a plateau corresponding to an overall fluorescence increase of 24.4 ± 0.5% was reached after 55 min. When 1.0×10^{-6} M of caspase-6 was added, the observed decrease in fluorescence was smallest, followed by the largest observed increase in fluorescence intensity (38.2 ± 0.9% after 55 min.). The observed concentration-dependent decrease/increase pattern suggest a rearrangement of nanosponge structure during caspase-6 digestion. This hypothesis is corroborated by TEM results, which are shown in Fig. 5. As shown in Fig. 4A, the amount of released carboxyfluorescein is clearly dependent on the concentration of caspase-6 that was added. The use of PBS medium ensured that the pH of all systems was 7.4. However, due to the observed rearrangement of the nanosponge structure after enzymatic cleavage, the amount of released fluorescent dye is not a linear function of the observed fluorescent increase. Nevertheless, the experimental data summarized in Fig. 4 clearly demonstrates the potential of DK20 nanosponges for cell-based drug delivery. It should be noted that DK20 nanosponges in PBS containing 5 percent (by volume) of human serum were stable for 24 h, indicating virtually no proteolytic cleavage. This result is important, because it demonstrated that DK20 nanosponges are at least short-term stable in the presence of serum, due to the absence of active caspases. This is of importance for the uptake experiments of type DK20 nanosponges by leucocytes in peripheral blood discussed below. At higher concentrations of serum in PBS, DLS could not be utilized as a detection method due to the presence of thousands of human proteins.

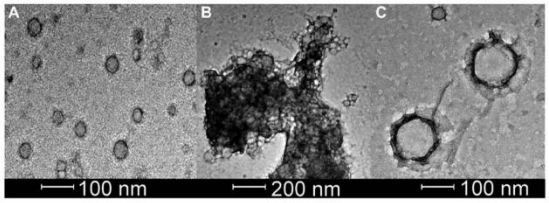


Fig. 5 Bright field transmission TEM (200 kV) of 0.20 mg mL^{-1} of carboxy-fluorescein-loaded CF-DK20 nanosponges: (A) Nanosponges deposited from PBS before adding caspase-6. (B) Reactive mixture deposited from PBS containing caspase-6 (2.60×10^{-10} M) after 15 min of reaction at 37 °C. (C) Novel nanostructures, which were formed in the reaction, deposited from PBS containing caspase-6 (2.60×10^{-10} M) after 60 min of reaction at 37 °C.

TEM-analysis of caspase-6 activation

Bright field TEM was also successfully used to visualize the effect of caspase-6 activation of carboxyfluorescein-loaded DK20 nanosponges. In Fig. 5, a sequence of three TEM images is shown: (A) 0.20 mg mL^{-1} of CF-DK20 nanosponges, deposited from PBS dispersion onto HOPG. (B) 0.20 mg mL^{-1} of CF-DK20 nanosponges after 15 min. Of incubation at 37 °C with commercially available caspase-6 (2.60×10^{-10} M), deposited from PBS dispersion onto HOPG. (C) 0.20 mg mL^{-1} of CF-DK20 nanosponges after 60 min. Of incubation at 37 °C with commercially available caspase-6 (2.60×10^{-10} M), deposited from PBS dispersion onto HOPG. TEM images were recorded immediately after the deposition of the (reactive) nanosponges on the carbon surfaces. Uranyl staining was added shortly before depositing the dispersions onto HOPG.

Fig. 5 shows that caspase-digestion of CF-DK20 nanosponges leads to the formation of a novel supramolecular structure. We have observed that the presence of a charged molecule (here: 5(6)-carboxyfluorescein) within the DK20 framework will influence the size of the formed aggregates. This effect is responsible for the differences in diameter that are observed compared to the TEM shown in Fig. 3 and ref. 38. Furthermore, the spherical nanosponges are deposited onto a carbon surface for the purpose of TEM. This will flatten their structures to a 2D coating and, at least partially, lead to the orientation of the hydrophobic cholesterol labels towards the carbon surface. Therefore, in contrast to dynamic light scattering, the nanosponges' structure will be somewhat distorted by the procedures necessary to record TEM. The resulting nanosponge diameters were 45 ± 10 nm (Fig. 6).

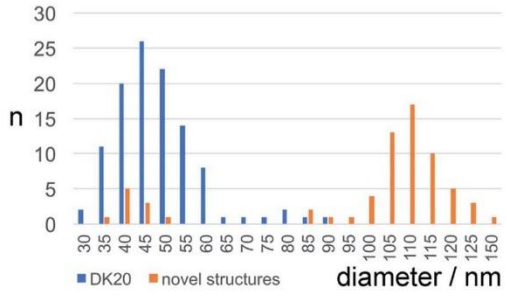


Fig. 6 Size distribution of carboxyfluorescein-loaded DK20 nanosponges before and after "digestion" with caspase-6 (as measured by means of TEM, *n*: number of particles counted).

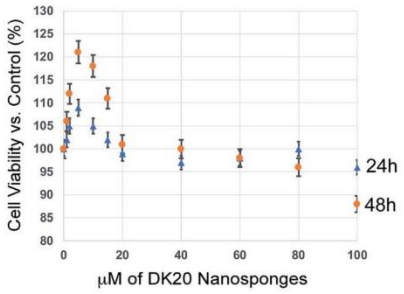


Fig. 7 Cell viability of C17.2 neural progenitor cells (NPCs) as a function of DK20 nanosponge concentration and incubation time (24 h and 48 h), as measured by the MTT assay.[45] Nanosponges were added to the cell culture medium in their respective concentrations (see ESI†). The cell viabilities after 24 h and 48 h in the absence of DK20 nanosponge, shown at 0 µM of DK20 nanosponges, were used as references to calculate viabilities.

In Fig. 5B the originally observed organic structures have completely vanished and a mesh of organic structures has formed. It is our interpretation of this observation that caspase-6 was able to cleave at least a fraction of the DEVDGC, thus disrupting the structure of the nanosponges. Enzymatic cleavage releases cholesterol-K_{20}-DE and cholesterol-D_{20}-DE units, and according to the results shown in Fig. 5c, these units (or at least cholesterol-K_{20}-DE) are able to form novel supramolecular structures. The re-formation of well-ordered structures may be responsible for the observed release of "only" about 3.7 ± 0.1% of carboxyfluorescein, which was observed by means of quantitative fluorescence recording. Interestingly, the supramolecular structures formed after 1 h of "digestion" with caspase-6 (2.60×10^{-10} M) are larger than the original nanosponges (110 ± 20 nm, Fig. 6).

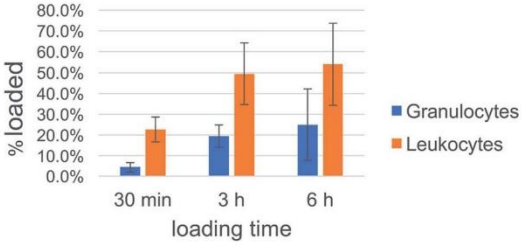


Fig. 8 Uptake efficacy of type DK20 nanosponges by non-granulocyte leukocytes (mainly lymphocytes and monocytes) and granulocytes (neutrophils) in peripheral blood as function of incubation time.

Cell toxicity of the peptide nanosponges

We have performed classic MTT cell proliferation assays[45] to determine the cell viability of murine C17.2 neural progenitor cells (NPCs)[34] after incubation with DK20 nanosponges. DK20 nanosponges are not toxic to NPCs, even at concentrations as large as 100 µM (Fig. 7). Only a slight increase in cell proliferation was observed at lower concentrations.

Nanosponge-uptake by leucocytes in peripheral blood

Cell uptake kinetics were recorded to determine the uptake efficiencies of the peptide nanosponges by neutrophils and leukocytes in peripheral blood. The results are summarized in Table 1 and Fig. 8. They indicate that the targeting of defensive cells within peripheral blood, followed by cell-based transport to the tumor site, is a feasible treatment strategy.

White blood cells from cattle blood show a time dependent uptake of DK20 nanosponges. The entire WBC (white blood cells) population was subclassified into granulocytes (neutrophils) and other leukocytes such as lymphocytes and monocytes. Here, the other leukocyte group loaded twice as well (>50%) compared to the granulocyte group (~25%). The loading was observed over a timeframe of 6 hours with the maximum loading completed after 3 hours.

Table 1 Uptake of DK20 nanosponges by leukocytes in peripheral blood

	DK20 Nanosponges					
	30 min		3 h		6 h	
	Neutrophils loaded	Other leukocytes loaded	Neutrophils loaded	Other leukocytes loaded	Neutrophils loaded	Other leukocytes loaded
Ave-rage	4.3%	22.6%	19.5%	49.4%	25.0%	54.0%
StDev	2.4%	6.0%	5.4%	14.8%	17.3%	19.7%

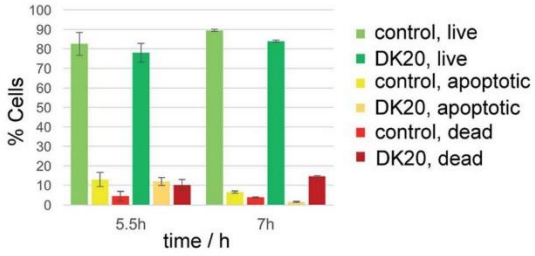


Fig. 9 Survival of WBC when exposed to DK20 peptide nanosponges, compared to the survival of an unexposed WBC control group. No significant difference ($p > 0.05$) between the treated and untreated live cell populations was detected.

Survival of the WBC population was analyzed by detecting apoptotic cells with the annexin V fluorescent marker and dead cells with propidium iodine. The relative survival was measured after 5.5 and 7 hours of incubation with DK20 nanosponges and compared to a control group. The live cell population remained between 78% and 89% relative to the total cell count and with no significant difference between the control and the DK20 group for the duration of the experiment. Apoptotic cells were at approximately 12% after 5.5 hours, again with no significant difference between the two groups. The apoptotic cell count in the DK20 nanosponge group drops to 2% after seven hours while the dead cell count is significantly increased and measured at 15%. Our hypothesis is that this is observed due to the stress exerted on the WBS from endocytosis and processing of the DK20 nanosponges (Fig. 9).

Discussion

Three main obstacles to efficient cell-mediated therapy of cancer and infectious diseases remain today: (1) fast uptake of drug formulations by the transport cells. (2) Effective migration of the transport cells to their intended target. (3) Efficient drug release by the transport cells once the target is reached. The nanosponges that are discussed here will be able to efficiently target neural progenitor cells[38,53] and leucocytes, either *ex vivo* or, preferentially, in peripheral blood to utilize the advantages of autologous cells for patient-specific cell therapies. Because of their fast uptake kinetics, and virtually non-existent toxicity, the nanosponges are well-suited for loading numerous drug formulations into various types of transport cells. Furthermore, because of their low toxicity to the transport cells, they make a very important contribution to facilitating effective cell migration to targets *in vivo*, because the viability of the transport cells will remain high during the migration phase of several days. In our previous work, we have observed by means of fluorescence microscopy that type DK20 nanosponges retained their structures within the cytoplasm of monocyte/macrophage-like cells (RAW 264.7) for up to 72 h (ref. 38). Since RAW 264.7 cells belong to the group of leukocytes, this finding has importance for the study reported here. Finally, as we have demonstrated here by utilizing 5(6)-carboxyfluorescein as model drug, caspase-mediated drug release can be achieved. The observed increase in carboxyfluorescein emission was 38.2 $\pm$ 0.9% after 55 min in the presence of 1.0×10^{-6} M of caspase-6 in PBS, compared to PBS alone. Furthermore, the "caspase storm" during apoptosis has the potential of further degrading the nanosponges and to create numerous apoptotic bodies, to which the nanosponge-derived components will be adsorbed. Therefore, we anticipate a very high degree of drug release and consequent re-uptake *in vivo*.

We have utilized caspase-6 (MEROPS, ID: C14.005) in our studies, because it is one of the "effector caspases of apoptosis".[54] Caspases 3 and 6 are responsible for significant morphological changes in the nucleus at the onset of apoptosis. Caspase-6 cleaves nuclear lamina and the protein NuMa of the nuclear mitotic apparatus. Caspase-6 is also a suitable protease to cleave the DEVDGC motif of the nanosponges. According to this mechanism, drug release from the nanosponges within the transport cells can be triggered by apoptosis.[55] It is noteworthy that the consensus sequence DEVGDC is also capable of reacting with the other effector caspases-2, -3, and -7.[39]

Neutrophils make up a significant percentage of leucocytes.[56] As already discussed, neutrophils undergo apoptosis within hours after reaching tumors and metastases.[56] This makes them very attractive autologous cells for tumor targeting. Neural progenitor cells constitute a second class of delivery cells, which can migrate to solid tumors and metastases in large numbers.[29,34,46] However, the release of the payload has to be triggered by introducing apoptosis, as described above of by designing a TetOn gene regulation system, which silences a specific gene unless a tetracycline, such as doxycycline, is present.[10,57]

Conclusion

The structure predictions for the supramolecular binary nanosponges (type DK20) through explicit solvent and then coarse-grained molecular dynamics (MD) simulations, have been confirmed by transmission electron microscopy and dynamic light scattering studies. The structural and dynamic understanding of the nanosponges has enabled several applications of these novel materials, which, principally, prove them as advanced biomaterials in cell-mediated drug transport to solid tumors/metastases and infectious diseases: caspase-activated (model) drug release was demonstrated with 5(6)-carboxyfluorescein-loaded nanosponges. PKH26-loaded nanosponges were essentially non-toxic to cultured neural progenitor cells (NPC).[38] Targeting of leucocytes (WBC) in peripheral blood was successful. After 3 h of incubation, maximal uptake into non-granulocytic leukocytes and granulocytes in peripheral blood was observed. No significant difference between the untreated and DK20-nanosponge-treated live cell populations was detected. These results indicate that leucocytes in peripheral blood can be targeted by DK20 nanosponges without the requirement of previous isolation. We regard this as an important step towards cell-mediated therapy of tumors utilizing autologous cells as delivery vectors.

Author contributions

The manuscript was written through contributions of all authors./All authors have given approval to the final version of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Abbreviations

PBS:	Phosphate-buffered saline buffer
AFM:	Atomic force microscopy
TEM:	Transmission electron microscopy
HBTU:	(2-(1*H*-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate)
DIEA:	*N*,*N*-Diisopropylethylamine
CDI:	(Carbonyl-di-imidazole)
TIPS:	Triisopropylsilane
HEPES:	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid

Acknowledgements

The authors acknowledge support from NSF (DMR 1242765 and CBET 1656989) and the Johnson Cancer Center at Kansas State University.

References

1 A. S. Yapa and S. H. Bossmann, in *Magnetic Nanomaterials: Applications in Catalysis and Life Sciences*, ed. S. H. Bossmann and H. Wang, RCS, London, 2017.
2 J. Si, S. Shao, Y. Shen and K. Wang, *Small*, 2016, **12**, 5108–5119.
3 F. Danhier, *J. Controlled Release*, 2016, **244**, 108–121.
4 L. Tavano and R. Muzzalupo, *Colloids Surf., B*, 2016, **147**, 161–171.
5 S. Fathi and A. K. Oyelere, *Future Med. Chem.*, 2016, **8**, 2091–2112.
6 A. Srivastava, A. Babu, J. Filant, K. M. Moxley, R. Ruskin, D. Dhanasekaran, A. K. Sood, S. McMeeki and R. Ramesh, *J. Biomed. Nanotechnol.*, 2016, **12**, 1159–1173.
7 S. Swain, P. K. Sahu, S. Beg and S. M. Babu, *Curr. Drug Delivery*, 2016, **13**, 1290–1302.
8 W. S. Neto, L. I. Pena, G. R. Ferreira, F. G. Souza Junior and F. Machado, *Curr. Org. Chem.*, 2017, **21**, 4–20.
9 M. T. Basel, T. B. Shrestha, D. L. Troyer and S. H. Bossmann, *ACS Nano*, 2011, **5**, 2162–2175.
10 M. T. Basel, S. Balivada, T. B. Shrestha, G.-M. Seo, M. M. Pyle, M. Tamura, S. H. Bossmann and D. L. Troyer, *Small*, 2012, **8**, 913–920.
11 Y. Nakamura, A. Mochida, P. L. Choyke and H. Kobayashi, *Bioconjugate Chem.*, 2016, **27**, 2225–2238.
12 G. Di Lorenzo, S. De Placido, M. Pagliuca, M. Ferro, G. Lucarelli, S. Rossetti, D. Bosso, L. Puglia, P. Pignataro, I. Ascione, O. De Cobelli, M. Caraglia, M. Aieta, D. Terracciano, G. Facchini, C. Buonerba and G. Sonpavde, *Expert Opin. Biol. Ther.*, 2016, **16**, 1387–1401.
13 S. Parakh, A. C. Parslow, H. K. Gan and A. M. Scott, *Expert Opin. Drug Delivery*, 2016, **13**, 401–419.
14 P. E. Hughes, S. Caenepeel and L. C. Wu, *Trends Immunol.*, 2016, **37**, 462–476.
15 R. Bazak, M. Houri, S. El Achy, S. Kamel and T. Refaat, *J. Cancer Res. Clin. Oncol.*, 2015, **141**, 769–784.
16 B. Mukherjee, B. S. Satapathy, L. Mondal, N. S. Dey and R. Maji, *Curr. Pharm. Biotechnol.*, 2013, **14**, 1250–1263.
17 H. Sun and Y. Zu, *Small*, 2015, **11**, 2352–2364.
18 X. Wu, J. Chen, M. Wu and J. X. Zhao, *Theranostics*, 2015, **5**, 322–344.
19 H. Hamidi, M. Pietila and J. Ivaska, *Br. J. Cancer*, 2016, **115**, 1017–1023.
20 C.-C. Sun, X.-J. Qu and Z.-H. Gao, *Anti-Cancer Drugs*, 2014, **25**, 1107–1121.
21 R. K. Jain and T. Stylianopoulos, *Nat. Rev. Clin. Oncol.*, 2010, **7**, 653–664.
22 T. B. Shrestha, G. M. Seo, M. T. Basel, M. Kalita, H. Wang, D. Villanueva, M. Pyle, S. Balivada, R. S. Rachakatla, H. Shinogle, P. S. Thapa, D. Moore, D. L. Troyer and S. H. Bossmann, *Photochem. Photobiol. Sci.*, 2012, **11**, 1251–1258.
23 G.-M. Seo, R. S. Rachakatla, S. Balivada, M. Pyle, T. B. Shrestha, M. T. Basel, C. Myers, H. Wang, M. Tamura, S. H. Bossmann and D. L. Troyer, *Mol. Biol. Rep.*, 2012, **39**, 157–165.
24 M. T. Basel, S. Balivada, H. Wang, T. B. Shrestha, G. M. Seo, M. Pyle, G. Abayaweera, R. Dani, O. B. Koper, M. Tamura, V. Chikan, S. H. Bossmann and D. L. Troyer, *Int. J. Nanomed.*, 2012, **7**, 297–306.
25 D. Troyer, *J. Cancer Sci. Ther.*, 2013, **5**, 142–143.
26 H. S. Alshetaiwi, S. Balivada, T. B. Shrestha, M. Pyle, M. T. Basel, S. H. Bossmann and D. L. Troyer, *J. Photochem. Photobiol., B*, 2013, **127**, 223–228.
27 M. T. Basel, T. B. Shrestha, S. H. Bossmann and D. L. Troyer, *Ther. Delivery*, 2014, **5**, 555–567.
28 H. Kulbe, N. R. Levinson, F. Balkwill and J. L. Wilson, *Int. J. Dev. Biol.*, 2004, **48**, 489–496.
29 N. Ullah, S. Liaqat, S. Fatima, F. Zehra, M. Anwer and M. Sadiq, *Asian Pac. J. Trop. Dis.*, 2016, **6**, 406–419.
30 R. Z. Panni, D. C. Linehan and D. G. DeNardo, *Immunotherapy*, 2013, **5**, 1075–1087.
31 D. M. Richards, J. Hettinger and M. Feuerer, *Cancer Microenviron.*, 2013, **6**, 179–191.
32 S. B. Coffelt, M. D. Wellenstein and K. E. de Visser, *Nat. Rev. Cancer*, 2016, **16**, 431–446.
33 A. Mantovani, *New J. Sci.*, 2014, 1–15, DOI: 10.1155/2014/271940.
34 R. S. Rachakatla, S. Balivada, G.-M. Seo, C. B. Myers, H. Wang, T. N. Samarakoon, R. Dani, M. Pyle, F. O. Kroh, B. Walker, X. Leaym, O. B. Koper, V. Chikan, S. H. Bossmann, M. Tamura and D. L. Troyer, *ACS Nano*, 2010, **4**, 7093–7104.
35 C. Tecchio and M. A. Cassatella, *Semin. Immunol.*, 2016, **28**, 119–128.

36 S. O. Wendel, S. Menon, H. Alshetaiwi, T. B. Shrestha, L. Chlebanowski, W.-W. Hsu, S. H. Bossmann, S. Narayanan and D. L. Troyer, *PLoS One*, 2015, **10**, e0128144.
37 J. Copier, M. Bodman-Smith and A. Dalgleish, *Immunotherapy*, 2011, **3**, 507–516.
38 H. Wang, A. S. Yapa, N. L. Kariyawasam, T. B. Shrestha, M. Kalubowilage, S. O. Wendel, J. Yu, M. Pyle, M. T. Basel, A. P. Malalasekera, Y. Toledo, R. Ortega, P. S. Thapa, H. Huang, S. X. Sun, P. E. Smith, D. L. Troyer and S. H. Bossmann, *Nanomed. Nanotechnol. Biol. Med.*, 2017, **13**, 2555–2564.
39 G. P. McStay, G. S. Salvesen and D. R. Green, *Cell Death Differ.*, 2008, **15**, 322–331.
40 R. Jager and R. M. Zwacka, *Cancers*, 2010, **2**, 1952–1979.
41 M. Linder and T. Tschernig, *Ann. Anat.*, 2016, **204**, 114–117.
42 D. R. Powell and A. Huttenlocher, *Trends Immunol.*, 2016, **37**, 41–52.
43 S. E. Logue and S. J. Martin, *Biochem. Soc. Trans.*, 2008, **36**, 1–9.
44 https://mai.ku.edu/about-mai-lab, (accessed 01/31/2017).
45 J. C. Stockert, A. Blazquez-Castro, M. Canete, R. W. Horobin and A. Villanueva, *Acta Histochem.*, 2012, **114**, 785–796.
46 J. Ourednik, V. Ourednik, W. P. Lynch, M. Schachner and E. Y. Snyder, *Nat. Biotechnol.*, 2002, **20**, 1103–1110.
47 G. Rees and R. Gough, *J. Med. Lab. Technol.*, 1968, **25**, 117–118.
48 J. Cid, M. Claparols, A. Pinacho, J. M. Hernandez, P. Ortiz, L. S. Puig and R. P. Pla, *Transfus. Apher. Sci.*, 2007, **36**, 243–247.
49 G. Podaru, S. Ogden, A. Baxter, T. Shrestha, S. Ren, P. Thapa, R. K. Dani, H. Wang, M. T. Basel, P. Prakash, S. H. Bossmann and V. Chikan, *J. Phys. Chem. B*, 2014, **118**, 11715–11722.
50 R. F. Chen and J. R. Knutson, *Anal. Biochem.*, 1988, **172**, 61–77.
51 J. E. Lee, N. Lee, T. Kim, J. Kim and T. Hyeon, *Acc. Chem. Res.*, 2011, **44**, 893–902.
52 G. Podaru, S. Ogden, A. Baxter, T. Shrestha, S. Ren, P. Thapa, R. K. Dani, H. Wang, M. T. Basel, P. Prakash, S. H. Bossmann and V. Chikan, *J. Phys. Chem. B*, 2014, **118**, 11715–11722.
53 H. Wang, T. B. Shrestha, M. T. Basel, R. K. Dani, G.-M. Seo, S. Balivada, M. M. Pyle, H. Prock, O. B. Koper, P. S. Thapa, D. Moore, P. Li, V. Chikan, D. L. Troyer and S. H. Bossmann, *Beilstein J. Nanotechnol.*, 2012, **3**, 444–455.
54 E. A. Martinova, *Russ. J. Bioorg. Chem.*, 2003, **29**, 471–495.
55 P. Gimenez-Bonafe, A. Tortosa and R. Perez-Tomas, *Curr. Cancer Drug Targets*, 2009, **9**, 320–340.
56 A. S. Cowburn, A. M. Condliffe, N. Farahi, C. Summers and E. R. Chilvers, *Chest*, 2008, **134**, 606–612.
57 C. Le Guiner, K. Stieger, A. Toromanoff, F. Rolling, P. Moullier and O. Adjali, in *A Guide To Human Gene Therapy*, World Scientific, 2010, pp. 163–180.

Appendix D - Copyright Clearance

Title: Rationally designed peptide nanosponges for cell-based cancer therapy

Author: Hongwang Wang,Asanka S. Yapa,Nilusha L. Kariyawasam,Tej B. Shrestha,Madumali Kalubowilage,Sebastian O. Wendel,Jing Yu,Marla Pyle,Matthew T. Basel,Aruni P. Malalasekera,Yubisela Toledo,Raquel Ortega,Prem S. Thapa,Hongzhou Huang,Susan X. Sun et al.

Publication: Nanomedicine: Nanotechnology, Biology and Medicine

Publisher: Elsevier

Date: November 2017

Logged in as:
Nilusha Kariyawasam
Kansas State University

LOGOUT

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights

BACK CLOSE WINDOW

Peptide nanosponges designed for rapid uptake by leukocytes and neural stem cells

Reproduced material from:

A. S. Yapa, H. Wang, S. O. Wendel, Tej. B. Shrestha, Nilusha L. Kariyawasam, M. Kalubowilage, A. S. Perera, M. Pyle, M. T. Basel, A. P. Malalasekera, H. Manawadu, J. Yu, Y. Toledo, R. Ortega, P. S. Thapa, P. E. Smith, D. L. Troyer and S. H. Bossmann, RSC Adv., 2018, 8, 16052

DOI: 10.1039/C8RA00717A - Published by The Royal Society of Chemistry.