

UNIVERSITY OF CALGARY

Novel Electrostatic Mechanisms Controlling the Conformational Switching of L-plastin

by

John Keenan Fanning

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FUFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE MASTER OF SCIENCE

GRADUATE PROGRAM IN BIOLOGICAL SCIENCES

CALGARY, ALBERTA

JUNE, 2018

© John Keenan Fanning 2018

0.1 - Abstract

L-plastin is an actin-bundling protein that promotes the motility of both hematopoietic and metastatic cancer cells. The high definition structure of the calcium-binding regulatory domain of human L-plastin was recently determined, allowing computational research on this portion of the protein. The Drude polarizable force field was used to provide accurate computational simulations of the calcium-binding domain in conjunction with experimental validation to shown that L-plastin can regulate calcium-binding, and thus actin-bundling, through internal electrostatic interactions. Through this work the Drude force field was also benchmarked, to show that it provides comparable results to classical force fields with the added ability to simulate polarizability. Overall, a novel mechanism which allows L-plastin to self-regulate its calcium-binding affinity was developed.

0.2 - Acknowledgments

This thesis and research was accomplished due to continued support from Dr. Sergei Noskov and Dr. Hans Vogel. Over the past 2 years I have spent working on this master's I have received excellent mentorship from these supervisors! Thank-you for this fantastic opportunity and the continued support, I have grown significantly as a researcher and hope to carry this knowledge forward into any further academic work I may perform.

Additionally, I am thankful to my committee members, Dr. Ken Ng and Dr. Justin MacCallum. They have always made time to meet with me both in committee meetings and personal discussions, where they have provided valuable feedback and helpful suggestions.

As this has been my first attempt at research I also need to recognize all the support and mentorship I have received from other individuals within the U of C. Firstly my lab members who have worked directly with me throughout this project, with special thanks going out to Dr. Van Ngo and Dr. Hiroaki Ishida. These two researchers are hugely responsible for the success of this project. They always kept me focused, on track, and were always there to offer support, suggestions and help when needed. Additionally, I would like to recognize my academic mentors who were always willing to talk and provided the needed inspiration to undertake and complete this degree. Along with all the fantastic researchers that have worked with me in the lab, a few individuals who have been incredible academic mentors include Dr. Elke Lohmeier-Vogel, Dr. Israel Belenkie, Dr. Hristina Zhekova, Dr. Gareth Jones, and Dr. Igor Zdravkovic who were all extremely influential to my development as a researcher.

On a personal level, I have received a phenomenal amount of support from my family, especially my parents Dr. Ward Fanning and Mrs. Susan Fanning, and my siblings Mr. Vernon Merrick Fanning, Ms. Leah Haley Fanning, and Ms. Grace Maria Fanning. My friends have also

been a huge source of support by providing me with ongoing motivation, with special thanks going out to Mr. Kyle “the knowledgeable” Oliphant, Mr. Dan Van der Lann, Mrs. Claire Berenger, Dr. Anukul “Ace” Ghimire, Dr. Kevin Amaniampong, Mr. Jerry Flaws, Mr. Kyle Larson, Mr. Brad Steele, Mr. Matthew Laslo, Mr. Dion Galandy, Mr. Tim Popel, and Mrs. Anya Poulson. I would also like to recognize the Oliphant and Poulson families for their ongoing support.

One last thank-you goes out to the whole of the Calgary Saracens Rugby community. I would like to particularly thank my coaches Mr. Rob Amey, Mr. Kelly Wilson, and Mr. Andy Wilde for always understanding when my school came first and helping me grow as a rugby player and a person. Moreover, I would like to thank our team captain Mr. Matthew Whalley for his continuing support, inspiration, and positivity. With that said, all my teammates have been fun to play with, and always have my back, no matter what. To name just a few of the great people I played with I would like to thank Mr. Tyler Faith, Mr. Dylan Tickner, Mr. Ryan Tickner, Mr. Dan Salt, Mr. Trevor Wilde, Mr. Josh Portier, Mr. Gary Dellow, Mr. Grant Cromwell, Mr. Andy Fitzpatrick, Mr. Mark Mills, Mr. Tanner Nordstrom-Young, Mr. Christian Dudar, Mr. Cole McGregor, Mr. Mark Beatty, Mr. Mark Meidl, Mr. Gino Dominguez, Mr. Mitchell Steele, Mr. Fraser Flemons, Mr. Pádrdraige Bohane, Mr. Inaki Azpiazu, Mr. Nick VandenEnden, Mr. Leif Lind, Mr. Luke Church, Mr. Cody Tang, Mr. Matt Pronk, Mr. Evan Sieben, and Mr. Dave Roberts. I would also like to thank the women’s team, and the future Dr. Heather Boersma for showing me it is possible to be an academic rugby player and always being so optimistic about the future. **Up the Mighty Saracens!**

I cannot stress enough that this thesis would not have been possible without the continuing support I received throughout this work from all these people, so from the bottom of my heart
THANK-YOU!

0.3 - Table of Contents

Title Page	i
0.1 Abstract	ii
0.2 Acknowledgments	iii
0.3 Table of Contents	v
0.4 List of Figures and Tables	ix
0.5 List of Equations	xi
0.6 List of Symbols and Abbreviations	xiii
Chapter 1 – Introduction	1
1.1 Conformational Changes in Proteins	1
1.1.1 Calcium Signaling	1
1.1.2 Calcium-Binding Proteins and the EF-hand	3
1.1.3 Calmodulin: A Classic EF-hand Protein	6
1.1.4 Plastins: A Family of Actin Bundling EF-hand Proteins	8
1.1.5 L-Plastin: A Novel Model of Protein Regulation in an EF-hand	14
1.1.6 The Physiological Role of L-plastin	18
1.2 Significance of Thesis	20
1.2.1 Goals of the Thesis	20
1.2.2 Validation of the Drude Force Field (Chapter 3)	20
1.2.3 L-plastins Calcium Binding Regulation (Chapter 4)	21

1.2.4 Future Directions and Conclusions (Chapter 5)	21
Chapter 2 – Methods	23
2.1 - Computational Methods	23
2.1.1 Molecular Dynamics Challenges: Simulating Calcium Binding to Polarizable Proteins	23
2.1.2 CHARMM Classical Force Field	23
2.1.3 Drude Polarizable Force Field.....	25
2.1.4 Hydrogen Bonding Simulations and Analysis	27
2.1.5 Hydration Free-Energy of Side Chain Mimetics.....	31
2.1.6 Mapping of Ion-Dependent Conformational Switching.....	33
2.1.7 Annealing Simulation.....	33
2.2 – Experimental Methods	35
2.2.1 Need for Experimental Investigation.....	35
2.2.2 Bacterial Mutagenesis and Protein Purification	35
2.2.3 Isothermal Titration Calorimetry (ITC).....	35
2.2.4 Differential Scanning Calorimetry (DSC).....	38
2.2.5 Nuclear Magnetic Resonance Spectroscopy (NMR).....	39
2.2.6 Heteronuclear Single Quantum Coherence (HSQC) Experiments	40
Chapter 3 – Hydration Properties of Amino Acid Residues from Molecular Dynamic Simulations with Both Classical and Drude Polarizable Force Fields	43

3.1 – Abstract	43
3.2 – Introduction	44
3.3 – Methods	48
3.3.1 Hydration Free-Energy of Side-Chain Mimetics	48
3.3.2 Protein and Hydrogen-Bonding Simulations and Analysis.....	52
3.4 – Results	54
3.4.1 Hydration Free-Energy Values in the Drude Force Field Compared with the CHARMM36 Force Field	54
3.4.2 The Stability of the Drude Force Field vs the CHARMM36 Force Field.....	55
3.4.3 The Drude Force Field Shows Elongated Hydrogen Bond Lifetimes.....	57
3.4.4 The Drude Force Field Shows More Variation in the Lifetime of Hydrogen Bonds....	63
3.5 – Discussion	64
3.5.1 The Increase in Variation and Lifetime in the Drude Force Field May be Due to Polarizability.....	64
3.5.2 Explicitly Accounting for Polarizability may Result in More Dynamic Structures.....	65
3.5.3 Polarizability Can Account for Arg 10, with Multiple Hydrogen Bonds	65
3.6 – Conclusions	68
Chapter 4 – Atomic Level Characterization of Calcium-Induced Conformational Switching in Proteins	69
4.1 – Abstract	69

4.2 – Introduction	71
4.3 – Methods	77
4.3.1 Molecular Dynamics (MD) Simulations	77
4.3.2 Experiments	80
4.4 – Results	87
4.4.1 Structural States of L-plastin	87
4.4.2 Calcium-Binding Affinities and Mechanisms in WT and Mutant Forms of L-plastin	95
4.5 – Discussion and Conclusions	101
Chapter 5 – Conclusions and Future Directions	106
5.1 – Conclusions	106
5.1.1 Conclusions About the Drude Force Field	106
5.1.2 Conclusions About L-platin	106
5.2 – Future Directions	108
5.2.1 Docking Simulations	108
5.2.2 Replica Exchange with Solute Tempering (REST) Simulations	109
5.2.3 Developing L-plastin Simulations into a Markov State Model (MSM).....	112
5.2.4 Full L-plastin Simulations	113
5.2.5 Additional Mutants and Differential Scanning Calorimetry (DSC) Experiments	114
Bibliography	115

0.4 - List of Figures and Tables

Figure 1.1 – Both an open and closed EF-hand, and representative CaM-protein interactions	5
Figure 1.2 – A depiction of L-plastin’s structural components, and how it bundles F-actin.....	9
Figure 1.3 – Sequence alignment of plastins and fimbrins showing conserved regions within the N-terminus	13
Figure 1.4 – The NMR structure of L-plastin’s EF-hand domain in both its apo and holo conformations	16
Figure 2.1 – Hydrogen bonding geometric criteria.....	29
Table 3.1 - Calculated values for the free-energy of hydration	51
Figure 3.1 - Representative secondary structures run for 1ns in the Drude and C36 force fields	56
Figure 3.2 – The strict lifetime of protein-water hydrogen bonds	58
Figure 3.3 – The continuous lifetime of protein-water hydrogen bonds	60
Figure 3.4 – The strict lifetime of protein-protein hydrogen bonds	62
Table 4.1 - The sequence alignment and charge of the EF-hands and H5-tail in human plastins	73
Figure 4.1 - Structures of L-plastin used in MD simulations	75
Table 4.2 - The ITC data obtained for each L-plastin construct and the attributed mechanism...	82
Figure 4.2 - Representative ITC data.....	84
Table 4.3 - The DSC data obtained for each L-plastin protein construct	85
Figure 4.3 - RMSD of annealing simulations	88
Figure 4.4 - Probability of the positive sidechains found in L-plastin's H5-tail localizing near L-plastins two EF-hand motifs	91

Figure 4.5 –NMR data showing structural overlap of L-plastin and various mutant constructs .. 92

Figure 4.6 - Snapshots of sampled apo-like and holo-like structures, including a transition state94

Figure 4.7 - A comparison of ITC curves for all tested constructs of L-plastin 97

0.5 - List of Equations

eq. 2.1 – CHARMM potential energy function	24
eq. 2.2 – Non-bonded Van der Waal’s interaction function	24
eq. 2.3 – Coulomb’s law	25
eq. 2.4 – Drude particle distance equation	26
eq. 2.5 – Drude particle induced dipole moment equation	26
eq. 2.6 – Drude particle polarizability equation.....	26
eq. 2.7 – Strict (S) hydrogen-bond lifetime correlation function.....	30
eq. 2.8 – Continuous (C) hydrogen-bond lifetime correlation function	30
eq. 2.9 – Exponential function to determine hydrogen-bond lifetime	31
eq. 2.10 – Decoupled free-energy function used to determine solvation free-energy	32
eq. 2.11 – Gibbs free-energy function (1).....	36
eq. 2.12 – Gibbs free-energy function (2).....	36
eq. 2.13 – Conversion between K_d and K_a	37
eq. 2.14 – The thermal equivalent of Ohm’s law.....	38
eq. 3.1 – Repulsive and dispersive components of the solute-solvent interaction potential.....	48
eq. 3.2 – Decomposed repulsive component of the solute-solvent interaction potential.....	48
eq. 3.3 – Decomposed dispersive component of the solute-solvent interaction potential	49
eq. 5.1 – Decoupled hamiltonian used in regular replica exchange.....	111
eq. 5.2 – Boltzmann weighting factor (2)	111

eq. 5.3 – Scaled Hamiltonian used by replica exchange with solute tempering (REST) 111

0.6 - List of Symbols and Abbreviations

ABD – Actin Binding Domain

Apo – Ion-free

C36 – CHARMM36 force field

Ca²⁺ - Calcium ion

[Ca²⁺] - Calcium ion concentration

CaM – Calmodulin

CH – Calponin homology domain

CIB – Calcium-and-Integrin-Binding protein

Drude – Drude polarizable force field

DSC – Differential scanning calorimetry

EDTA - ethylenediaminetetraacetic acid

EF-WT – L-plastin protein construct containing the first 97 residues of the protein

EF-F90A – L-plastin protein construct with an alanine mutation at residue 90

EF-K96A – L-plastin protein construct with an alanine mutation at residue 96

EF-K97A – L-plastin protein construct with an alanine mutation at residue 97

EF-K96A-K97A – L-plastin protein construct with alanine mutations at residues 96 and 97

EF-R91A – L-plastin protein construct with an alanine mutation at residue 91

ER – Endoplasmic Reticulum

F-Actin – Filamentous actin

FEP – Free-energy Perturbation method

FF – Force field

FID – Free induction decay

FT – Fourier transform

ΔG – Gibb's free-energy

G-actin – Globular actin

ΔH – Enthalpy

ΔH_{melt} – Enthalpy of a proteins melting process

H-bond – Hydrogen bond

H5-tail – Helix-5-tail (found in the linking region of L-plastin)

Holo – Ion-bound

HSQC - Heteronuclear single quantum coherence

I-Plastin – Plastin 1 (human)

K_a – Association constant

K_d – Dissociation constant

LJ – Lennard-Jones potential

LRC – Long-range corrections

L-plastin – Plastin 2 (human)

ITC – Isothermal titration calorimetry

M – Markovian transition matrix

MC1 – Mechanism of L-plastin switching 1

MC2 – Mechanism of L-plastin switching 2

MC3 – Mechanism of L-plastin switching 3

MC4 – Mechanism of L-plastin switching 4

MD – Molecular dynamics

MSM – Markov state model

N – Stoichiometry

NMR – Nuclear magnetic resonance spectroscopy

NOE – Nuclear Overhauser effect

NVT – A canonical ensemble with a consistent number of atoms, volume, and temperature

PBC – Periodic boundary conditions

PCCA+ - Robust Perron cluster analysis

PMCA – Plasma membrane Calcium ATPase

QM – Quantum-mechanical

REM – Replica exchange method

REST – Replica exchange with solute tempering

RMSD – Root mean square deviation

RYR – Ryanodine receptor

ΔS – entropy

SCF – Self-consistent field

SERCA – Sarco/endoplasmic reticulum Ca^{2+} -ATPase

1SIBS – One set of identical binding sites

2SIBS – Two sets of independent binding sites

SR – Sarcoplasmic reticulum

STIM – Stromal interaction molecule

SWM4-NDP – Simple water model with four sites and negative Drude particle polarizability

τ_s^{PW} – The strict hydrogen bond lifetime of protein-water hydrogen bonds

$\langle \tau_s^{\text{PW}} \rangle$ – The average strict hydrogen bond lifetime of protein-water hydrogen bonds

τ_c^{PW} – The continuous hydrogen bond lifetime of protein-water hydrogen bonds

$\langle \tau_c^{\text{PW}} \rangle$ – The average continuous hydrogen bond lifetime of protein-water hydrogen bonds

τ_s^{PP} – The strict hydrogen bond lifetime of protein-protein hydrogen bonds

$\langle \tau_s^{\text{PP}} \rangle$ – The average strict hydrogen bond lifetime of protein-protein hydrogen bonds

T_m – Protein melting temperature

T_{onset} – The onset temperature of proteins melting process

T-plastin – Plastin 3 (human)

TCF – Time correlation function

TFP – Trifluoperazine

tICA – Time independent correlation analysis

TIP3 – Transferable intermolecular potential with 3 points (a water model)

TPT – Transition pathway theory

T-REMD – Tempering replica exchange MD

WCA – Weeks, Chandler, and Anderson's decomposition scheme

WHAM – Weighted histogram analysis method

WT – Wild-type

Chapter 1 – Introduction

1.1 - Conformational Changes in Proteins

1.1.1 – Calcium-Signalling

The mechanisms associated with ion-protein binding that can stimulate major protein structural changes are central to the ability of cells to transmit signals leading to various cellular functions such as muscle contraction, volume regulation, and cell motility [2-6]. Calcium ions (Ca^{2+}) are well known to be one of the major ion regulators of cellular function. Physiologically, Ca^{2+} plays a key role in maintaining heart rhythms, neurological functions, metabolism, transcription, cell movement, apoptosis, and controlling a large array of additional cellular tasks [5-17].

Physiologically, the cytosolic Ca^{2+} concentration ($[\text{Ca}^{2+}]$) in human cells is maintained at a low concentration ($\sim 10^{-7}$ M) [5, 6, 15, 18]. This low concentration is highly contrasted by the $[\text{Ca}^{2+}]$ found in the lumen of the endoplasmic reticulum (ER) and the extracellular space which are both maintained $\sim 10,000$ x higher than the cytosol ($\sim 10^{-3}$ M) [5, 6, 18]. This large concentration gradient is maintained by the cells use of various voltage gated ion channels in the cell membrane; the ryanodine receptor (RyR) in the ER/sarcoplasmic reticulum (SR); ATP driven Ca^{2+} pumps, such as sarco/endoplasmic reticulum Ca^{2+} -ATPase (SERCA); the plasma membrane Ca^{2+} -ATPase (PMCA); and $\text{Na}^+/\text{Ca}^{2+}$ exchangers [5, 6, 18].

Normally this large Ca^{2+} gradient is maintained as the Ca^{2+} channels remain closed and the cell maintains what is known as a “resting state” ($[\text{Ca}^{2+}] \approx 10^{-7}$ M) [5, 15, 18]. When voltage gated extracellular channels are stimulated and opened a large influx of Ca^{2+} enters the cytosol which further stimulates the release of Ca^{2+} from the ER into the cytosol [5, 15, 18]. This rapid and

substantial change in the cells $[Ca^{2+}]$ leads to the cell entering an “Activated State” ($[Ca^{2+}] \approx 10^{-5}M$) [5, 15, 19].

Once the cell is activated there are a wide variety of Ca^{2+} -binding proteins capable of eliciting a response, including membrane bound and soluble proteins [5-15, 18]. There are various physiological roles for these Ca^{2+} -binding proteins, with many capable of stimulating a further response through highly dynamic structures capable of shifting between structurally distinct ion-free (apo) and ion-bound (holo) conformations [11, 20]. Such dynamic structure, capable of changing in response to altered ion concentrations, enables these proteins to trigger physiological responses [11]. The wide range of these proteins found within a cell allows for varying response to an influx of Ca^{2+} [5, 15, 19, 21].

Among intracellular Ca^{2+} -binding proteins capable of conformational switching it is important to maintain an affinity that can be activated and deactivated within the physiological Ca^{2+} concentrations ($\sim 0.1\mu M - 10\mu M$) [5, 15, 18, 19]. Although the activation of many Ca^{2+} -binding proteins falls within this intracellular $[Ca^{2+}]$ there are examples outside of this. As the Ca^{2+} -concentration is 1000-1200x higher in the endoplasmic reticulum (ER) and extracellularly certain proteins exposed to these environments such as Stromal interaction molecule (STIM) proteins displays very high Ca^{2+} -binding affinities (with a K_d in the mM range) to allow these proteins to function within this different $[Ca^{2+}]$ [5, 22, 23]. Alternatively, some intracellular proteins such as calbindin D_{9K} and parvalbumin express nM Ca^{2+} affinities, and therefore are almost always bound to Ca^{2+} [23]. These high Ca^{2+} affinity proteins are known to help buffer Ca^{2+} levels within the cell rather than use Ca^{2+} to regulate cellular function [23]. This diverse range of Ca^{2+} affinities allows cells to recruit Ca^{2+} as a diverse cellular messenger within cells [23]. This

varied regulation indicates that in many ways accurately understanding a proteins affinity for Ca^{2+} is as important as the proteins mechanism.

This influx of Ca^{2+} can be further complicated by spatial and temporal effects [15, 21]. These effects allow cells to open specific channels for set durations to induce local and variable changes in $[\text{Ca}^{2+}]$ rather than causing a full Ca^{2+} response [6, 15, 21]. Various intensities and durations of Ca^{2+} signals are known as Ca^{2+} puffs, sparks, and waves and enable finer (and local) control over Ca^{2+} signalling [15, 21].

1.1.2 – Calcium-Binding Proteins and the EF-hand

Ca^{2+} -binding proteins can be classified into two groups: EF-hand proteins and non-EF-hand proteins [15]. Examples of non-EF-hand proteins include the protein families of annexins, cadherins, and the C2 domain family [15, 24-26]. These families all maintain different structures and functions [24-27]. Alternatively, EF-hand proteins comprise one superfamily [15, 28, 29]. Although the functions of each protein within this superfamily vary, EF-hand proteins are structurally similar, being identified by both a helix-loop-helix secondary structure and a well defined amino acid motif of the Ca^{2+} -binding loop [15, 29, 30].

The EF-hand helix-loop-helix secondary structure was first identified in parvalbumin describing its E and F helices (thus the name EF-hand), and the Ca^{2+} -binding loop [15, 29, 31]. The commonly observed helix-loop-helix secondary structure is often talked about in an “open” and “closed” state (figure 1.1 (a)/(b)) [28, 29]. In the closed state the two α -helices collapse side by side to form a hydrophobic core, burying many of the hydrophobic residues (figure 1.1 (b)) [28, 29, 32, 33]. Upon Ca^{2+} -binding to the Ca^{2+} -binding loop, the closed state is capable of transitioning into an open state (figure 1.1 (a)) [28, 29, 34, 35]. In the open state the two α -helices separate,

exposing the hydrophobic core, which further enables these proteins to bind a target (figure 1.1(a)) [28, 29]. Although the helix-loop-helix is the most common configuration of the EF-hand, variations of this structure are observed with only one of the two helices, or neither [36]. EF-hands are often found in pairs, joined by a small anti-parallel β -sheet between the Ca^{2+} -binding loop in each EF-hand [28, 29, 36].

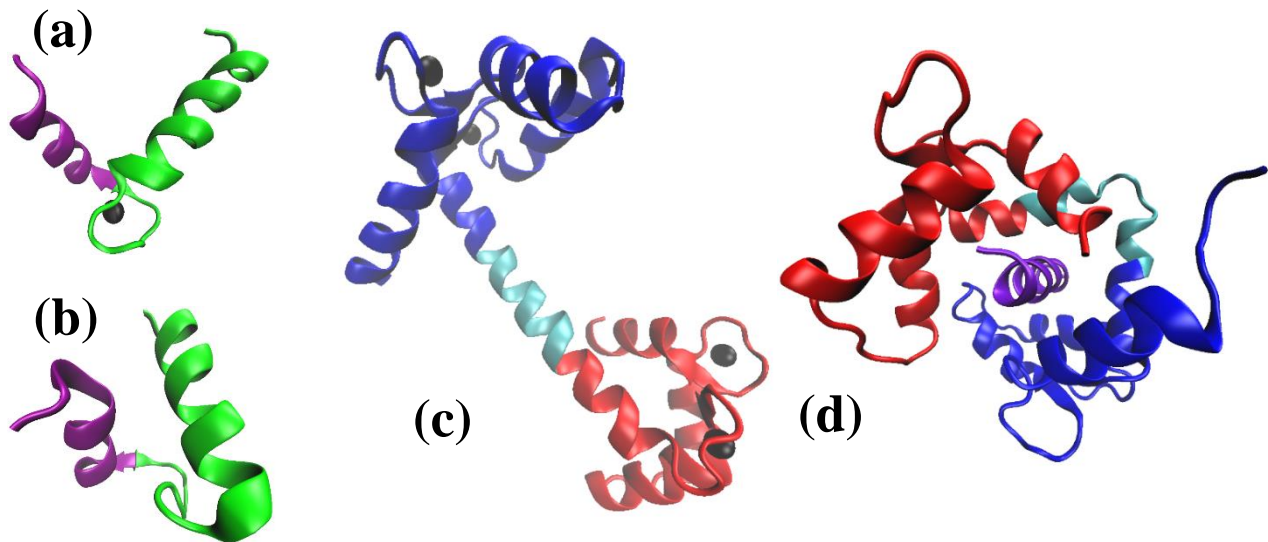


Figure 1.1 EF-hand motif two taken from L-plastin in both its **(a)** “open” (holo) and **(b)** “closed” (apo) states. Black spheres represent Ca^{2+} . **(c)** Rat CaM (structure 3CLN from PDB). Blue structure represents the N-terminal domain, red structure represents the C-terminal domain, and cyan structure represents the linker. **(d)** An example of a CaM-protein complex (structure 2LL6 from PDB). CaM (coloured the same as in **(b)**) is bound to iNOS (purple).

The specific amino acid motif which makes up the Ca^{2+} -binding loop is the other defining feature of EF-hands. The canonical sequence is 12 residues long and is often defined by a few key residues [15, 29, 36]. Within the loop residues 1(+X), 3(+Y), 5(+Z), 7(-Y), 9(-X), and 12(-Z) are responsible for coordinating the Ca^{2+} (here “(±X/Y/Z)” denote axis of coordination) [15, 29, 36]. Because these residues are directly responsible for coordinating ‘hard’ metal Ca^{2+} ions these residues contain many oxygen ligands and several negatively charged amino acids [29]. Residue 8 is conserved within this sequence as a hydrophobic residue (Ile, Val, or Leu) [29]. This hydrophobic residue 8 allows pairs of EF-hands to generate an anti-parallel β -sheet between the two EF-hand Ca^{2+} loops [29]. Another conserved residue is a Gly at position 6 [29]. This Gly 6 allows for the loop to make a 90° turn as only this amino acid it can facilitate an unusual main-chain conformation ($\phi \sim 60^\circ$ and $\psi \sim 20^\circ$) [29]. This canonical EF-hand motif is observed ubiquitously across eukaryotes, bacteria, and archaea [36].

1.1.3 – Calmodulin: A Classic EF-hand Protein

One of the most extensively studied EF-hand Ca^{2+} -binding proteins is calmodulin (CaM). CaM is found ubiquitously throughout all cells in the body and plays a crucial role in a plethora of Ca^{2+} regulatory interactions [11-15, 37-40]. This protein maintains 100% amino acid sequence conservation across vertebrates and is strictly conserved across all eukaryotic species [8, 15, 29, 37]. Additionally, CaM-like proteins have been found in bacteria [37, 41]. Inherited mutations that disrupt the ion-dependent dynamics of CaM or impact its interactions with other proteins have been reported as a root-cause in cardiac arrhythmias, sudden cardiac death, and skeletal muscle disorders [38-40]. CaM’s ability to respond to fine changes in the $[\text{Ca}^{2+}]$ has implicated this protein in a myriad of diseases including Parkinson’s, Alzheimer’s, and rheumatoid arthritis [11-13]. CaM has also been linked to memory consolidation [14]. Further, improper expression of wild-type

(WT) CaM has been shown to induce epilepsy, various neurological disorders, and diabetes leaving little doubt about its paramount role in proper physiological function [14, 39, 41, 42].

Structurally, CaM is highly α -helical in nature (figure 1.1 (c)) [29, 37]. CaM's four EF-hand domains are split into two, roughly symmetrical, lobes (75% sequence homology and 46% sequence identity), with each lobe containing two EF-hands (figure 1.1 (c)) [7, 11, 29, 30, 37, 43]. This bi-lobal structure enables CaM to freely rotate around a central α -helical linker when not bound to a complex (figure 1.1 (c)) [11, 28-30, 37, 43]. Although similar, each of the two lobes of CaM are shown to have different dynamics and Ca^{2+} affinities [29, 43]. The C-terminal domain of CaM has a ~10-fold greater affinity (K_d 0.1 μM - 6 μM) and shows slower dynamics than the N-terminal domain [11, 29, 43, 44]. By maintaining two Ca^{2+} -binding domains with different Ca^{2+} -affinities CaM can react selectively to a number of distinct [Ca^{2+}] [30].

The binding of Ca^{2+} to CaM is instigated by submicromolar increases in the cellular [Ca^{2+}] [11]. Once Ca^{2+} binds to one of CaM's four EF-hands the protein undergoes a conformational change which exposes the CaM protein-binding domain [11]. The structural shift then allows CaM to bind to over 400 known target proteins, further triggering a chain of events leading to an array of various cellular and physiological responses [11, 15, 28]. CaM is known to bind Ca^{2+} with a dissociation constant (K_d) ranging from 10^{-7} to 10^{-5} M [11, 43]. This conformational mechanism is reversible, by subsequent removal of Ca^{2+} -from the cell, and enables the regulation of a wide range of cellular functions [11].

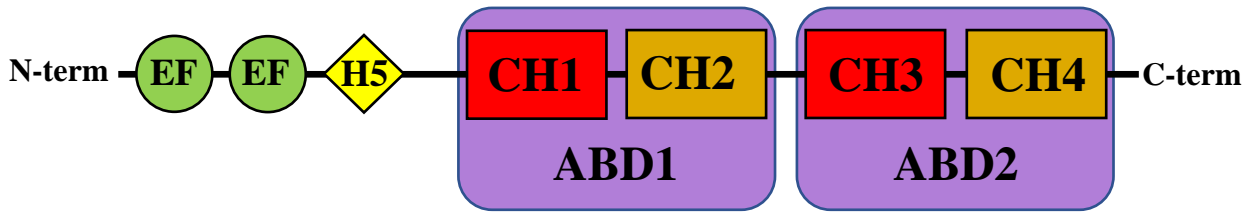
CaM contains a relatively high percentage of Met residues (with Met accounting for 5.3% in the N-terminus, and 6.8% in the C-terminus), compared to the average protein (~1.5%) [28, 43, 45, 46]. Although normally hydrophobic, of all the hydrophobic residues, Met displays the highest degree of flexibility, the least steric hinderance, and the lowest solvation energy [43, 45, 46]. The

polarizability of these highly-conserved Met residues is thought to contribute to the stabilization of the open conformation and stabilize CaM-protein binding, as they cover 46% of its hydrophobic binding pocket [7, 19, 28, 45, 46]. These residues are also thought to help promote the promiscuous binding observed by CaM through allowing flexibility of the binding pocket [7, 19, 28, 45, 46]. Moreover, a high percentage of Met residues have been found in other Ca²⁺-binding proteins capable of conformational change, including troponin C and α -spectrin, while Ca²⁺-binding proteins such as calbindin_{9k}, which express an absence of methionine residues, have shown an inability to undergo similar conformational switching [45, 46].

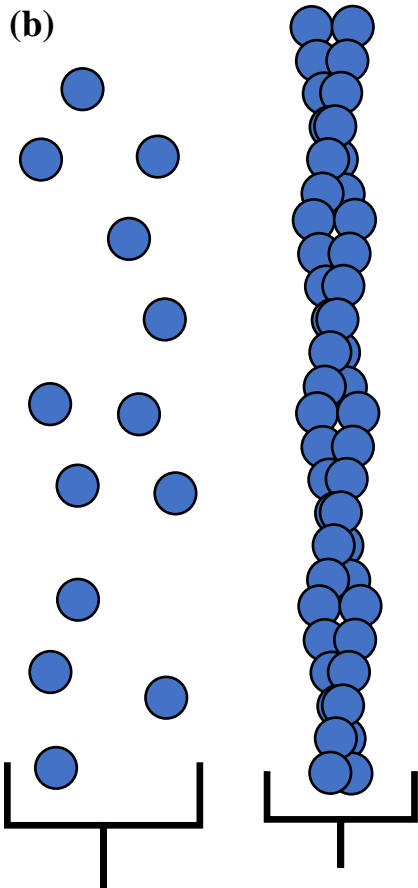
1.1.4 – Plastins: A Family of Actin Bundling EF-hand Proteins

The studies on CaM have contributed greatly to our understanding of the nature of EF-hand protein interactions, however CaM is a single protein in the much larger super family of EF-hand proteins [15, 28, 29]. Another group of EF-hand proteins, with a domain homologous to each lobe of CaM, is the family of plastins [47]. Plastins are a family of Ca²⁺-binding proteins responsible for actin bundling [48, 49]. Plastins are one of a handful of proteins capable of bundling actin; and cells use this array of actin bundling proteins to maintain a dynamic actin cytoskeleton [5, 22, 50]. Proteins such as plastins, fascin, and villin bundle filamentous actin (F-actin) tightly, solidifying column like structures, while less globular actin bundling proteins, such as α -actinin, bundle actin into networks of loosely packed actin often used to help stabilize cellular membranes [5, 22, 50].

(a)



(b)



G-Actin

F-Actin

(c)

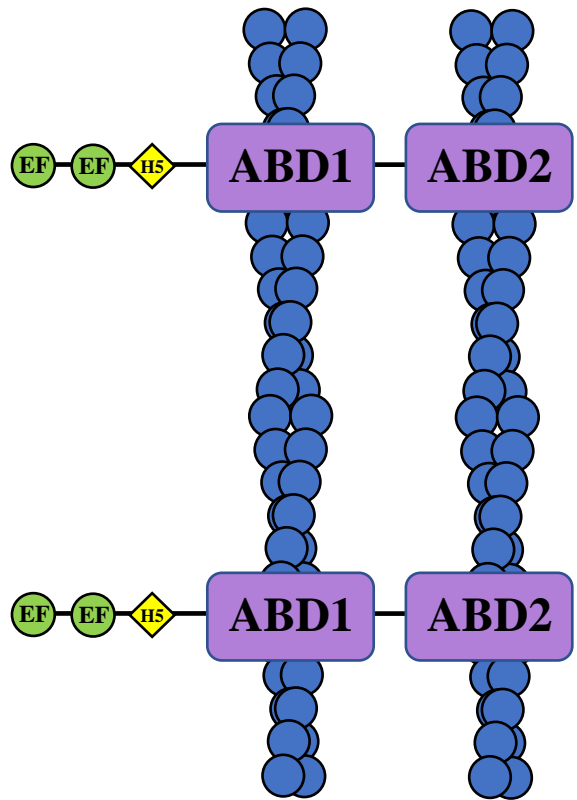


Figure 1.2. (a) A schematic of L-plastin's Structural domains. (b) A diagram of both G-actin (globular) and F-Actin (filamentous), and (c) a simple representation of how L-plastin's 2 ABD's bind F-actin filaments together.

In humans, there are three highly conserved isoforms of plastin: plastin-1, also known as I-plastin; plastin-2, also known as L-plastin; and plastin-3, also known as T-plastin [49]. Each human plastin homolog is localized to a specific cellular type [47, 49]. T-plastin is expressed in solid tissue cells, I-plastin is expressed in intestine and kidney cells, while L-plastin expression is normally only observed in hematopoietic cells [49, 51, 52]. These plastins are all structurally similar containing a Ca^{2+} -binding regulatory domain (comprised of two EF-hands) connected by a protein linker to two highly similar actin-binding domains (ABD's) (figure 1.2 (a)).

The EF-hand domain found in plastins is homologous to one lobe of CaM (34.7% identity) [47, 49, 50]. This well-preserved domain, regulated by two Ca^{2+} -binding sites, acts as a switch for the proteins function [48-50, 52, 53]. Unlike many Ca^{2+} -binding proteins, this domain deactivates the actin bundling function of this protein when Ca^{2+} binds [50, 53, 54]. Among plastin isoforms the affinity for Ca^{2+} varies, allowing for the precise activation and deactivation of actin-bundling in the cells which express the various isoforms [47, 49].

Although structurally similar, the Ca^{2+} affinity of plastins varies [48, 50, 53]. T-plastin maintains a Ca^{2+} affinity of $0.37\mu\text{M}$ - $10.6\mu\text{M}$ (K_d) [48]. L-plastin was thought to have a Ca^{2+} affinity of $0.7\mu\text{M}$ (K_d) [50]. While I-plastin has a Ca^{2+} affinity of $\sim 9\mu\text{M}$ (K_d) [53]. These varied affinities allow proteins some control to react to Ca^{2+} in a cell-specific manner, based on which plastin is expressed in the specific cell type [5].

The structure of the ABD's is homologous among plastins. The two ABD subunits found in each plastin can be further broken down into two calponin-homology (CH) subunits; ABD1 is composed of CH1 and CH2 while ABD2 is composed of CH3 and CH4 (figure 1.2 (a)) [49]. These two ABD's fold in a roughly hairpin anti-parallel manner where CH1 and CH4 make direct contact

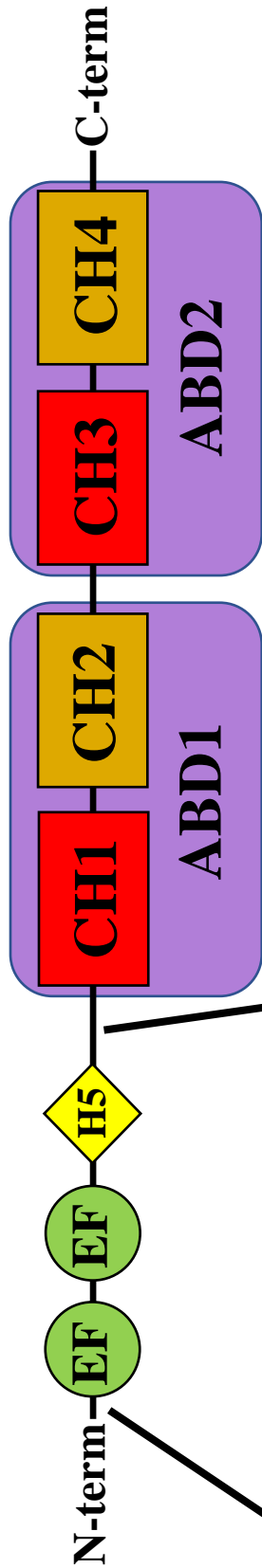
[49]. Having two ABD's side by side allows for this protein to bring F-actin together and bind them into higher order assemblies (figure 1.2 (c)) [5, 49].

Plastins specifically link F-actin into tightly packed parallel structures (filaments ~14nm apart) that rigidify the actin (figure 1.2 (c)) [5, 22]. This allows these structures to act as a supportive skeleton and are found in many structures that project from the cell [5]. Microvilli, stereocilia, and filopodia are all examples of such cellular projections [5].

Although both the EF-hand and the ABD domains are conserved in plastins, the protein sequence conservation of the ABD domains is higher than the EF-hand domain [47, 55]. The high-level of conservation within the ABD's indicates the conservation needed to bind two F-actin's simultaneously [47, 55]. Alternatively, the N-terminal "headpiece" region (or EF-hand domain) is much less conserved [55]. In many yeast and plants species the headpiece of fimbrin (another name for plastin) does not even bind Ca^{2+} (Figure 1.3) [55]. As this region of the protein functions as a regulator of the actin-bundling mechanics of plastins, the ability to react to local changes in Ca^{2+} (or not) is a highly important feature which clearly distinguishes the activation of many isoforms of plastins/fimbrins. Therefore, understanding how this regulation works, and varies among various species (and protein isoforms), is of significant interest.

Remarkably, within this "headpiece" region certain sections have been well conserved in the Ca^{2+} binding plastin homologues and lost in non- Ca^{2+} binding isoforms (Figure 1.3). Notably, in the Ca^{2+} -binding isoforms, the EF-hands are well preserved (Figure 1.3). Additionally, a highly charged region of protein which is found in the linker, connecting the "headpiece" to the ABDs, (later dubbed the H5-tail) appears to be consistently conserved in Ca^{2+} -binding homologues (Figure 1.3). The conservation of the EF-hands is understandable, as the EF-hand sequence

represents a well known Ca^{2+} -binding sites, however the role of the linker (H5-tail) plays in Ca^{2+} -binding isoforms of plastins remains a mystery.



```

LPL:  -----ARGSVSDEEMMELREAFKVPDIDGNGYISFNF  LNDLFKAACIPLPGYRVREITENIMATGDIIDODGRISFDFE  FIKIFHGLKSTD--VAKT--FRKAINKK-----
LPL:  -----MENSTTTISRRELEELQEA FNKIPIIDNSGYVSDYE  LQDLFKEASLPLPGYKRVREIVEKILLSVAIDSNKDKGISFEE  FVSLMQELKSKD--ISKT--FRKIINKR-----
TPL:  -----MDEMATTOISKDELDELKEAFKVPDINSNGFTCDYE  LHLELFKEANMPLPGYKRVREIIQKIMLDGIDRINKDKGISFDFE  FVYIFQEVKSSD--IAKT--FRKAINRK-----
FimA:  -----MTNFSESEISEFKASFNQFDENGDDGISALE  LQKILTKCGEKVTGVEVRDMIKEV-----PTDGNNGSIDFKE  FLQVMQKARQHSANASPA--FASAVKKV-----
Sac6p:  -----MNIIVKLQKRPILTQEDLFSITIEKFRAI  DLDDDKGWWEKQQALEEAVSKDGD-ATYDEARETLKHV-----GVDASGRVELDD  YVGLVAKLRESKTGAAPQTTFNVA PNSTPIVSTAA
Fim1:  -----MLALKLQKKYPELTNEEILTLTDQFNKLDVDGKGYLDOPT  TTIKAFEDSKK-GSYDEVREAIRREV-----NVDSSGRVEPED  FVGI FNVLKKGVGTEVVKKGRITIKGS-----
AtFim1: MSGYVGVVSDPWLQSQFTQVELRTLNSKYVSVKN-ONGKVTIED  LPPLFAKLKA-LSATFKEDEIKGMLGELGSDTSTDVSFEE  FLKIYLNLLSKA--AEKSGGHHK--NSSSFLKACT
AtFim5: MSYVGVLVSDPWLQSQFTQVELRTLKSKFVSNKT-OLGRFTVGD  LPEVFEKLKA-FNGTIDEDEIKSVLDKSYFNADDEVDFEF  FLRAFLSVQARG--VEKSGGSKG---ASSFLKTST

```

Figure 1.3 The sequence alignment of human L-plastin (LPL) with 7 other plastins and fimbrins: human I-plastin (IPL), human T-plastin (TPL), *Dictyostelium* fimbrin (FimA), *Saccharomyces* fimbrin (Sac6p), *Schizosaccharomyces* fimbrin (Fim1), *Arabidopsis* fimbrin (AtFim1), *Arabidopsis* fimbrin (AtFim5). Names highlighted in red indicate the protein does not bind Ca²⁺. The green highlighting depicts the alignment of the EF-hand regions. The yellow highlighting indicates the alignment of the H5-tail region

Although plastins are ultimately larger than Ca^{2+} -binding proteins such as CaM, the single N-terminal Ca^{2+} -binding domain is homologous to one lobe of CaM [47]. Therefore, by using a plastin's single Ca^{2+} -binding domain, the protein dynamics can be evaluated computationally with roughly half the number of protein atoms needed compared to a Ca^{2+} -binding protein such as CaM [11, 37, 47, 49]. Additionally, this small regulatory domain still represents a unique and physiologically relevant protein domain responsible for the activation and deactivation of actin-bundling, shaping the cytoskeleton of cells [47-54]. Thus, the relatively small single N-terminal domain of plastin makes it a good target for molecular dynamics simulations as it requires a reduced need for computational power resulting from the smaller model size, while still providing physiologically valid outcomes.

1.1.5 – L-Plastin: A Novel Model of Protein Regulation in an EF-hand

Importantly, Ishida and colleagues recently determined high resolution structures for the EF-hand domain of L-plastin in both the holo (Ca^{2+} -bound) and apo forms (figure 1.4) [50]. These structures show an interesting puzzle: the region of protein linking the EF-hand domain to the ABD's was observed to be the binding target of the EF-hand domain in the holo conformation (figure 1.4) [50]. This linking region observed in these studies comprises 14 residues of the 21-residue linker that connects the Ca^{2+} -binding domain to the ABDs [50]. Although, this linking region was observed to bind into the EF-hand and possess an α -helical structure in the holo form (as seen in figure 1.4), it was observed to be unstructured in the apo state [50]. Due to the α -helical nature of this region in the holo state, it will be referred to as the “Helix-5-tail” (H5-tail) (as it represents the 5th helix in the holo state) (figure 1.4). This H5-tail has also been called the “switch helix”, as it can turn the actin-bundling activity of L-plastin on-and-off [50]

The H5-tail is stabilized in an α -helical structure in the holo conformation by the hydrophobic core via several hydrophobic residues (notably Val86, Phe90, and Ile94) capable of coordinating this region [50]. Alternatively, the H5-tail appeared unstructured in the apo state [50]. In this work, although unstructured, the H5-tail was added onto the apo state in its helical structure to allow computational simulations to run in a reasonable timeframe (Figure 1.4).

Ion-binding induced stability is not unique to this protein and examples such as Calcium-and-Integrin-Binding protein (CIB) have shown so be mostly unstructured in the absence of either Ca^{2+} or Mg^{2+} , and conform to a helical structure upon binding either of these divalent ions [7].

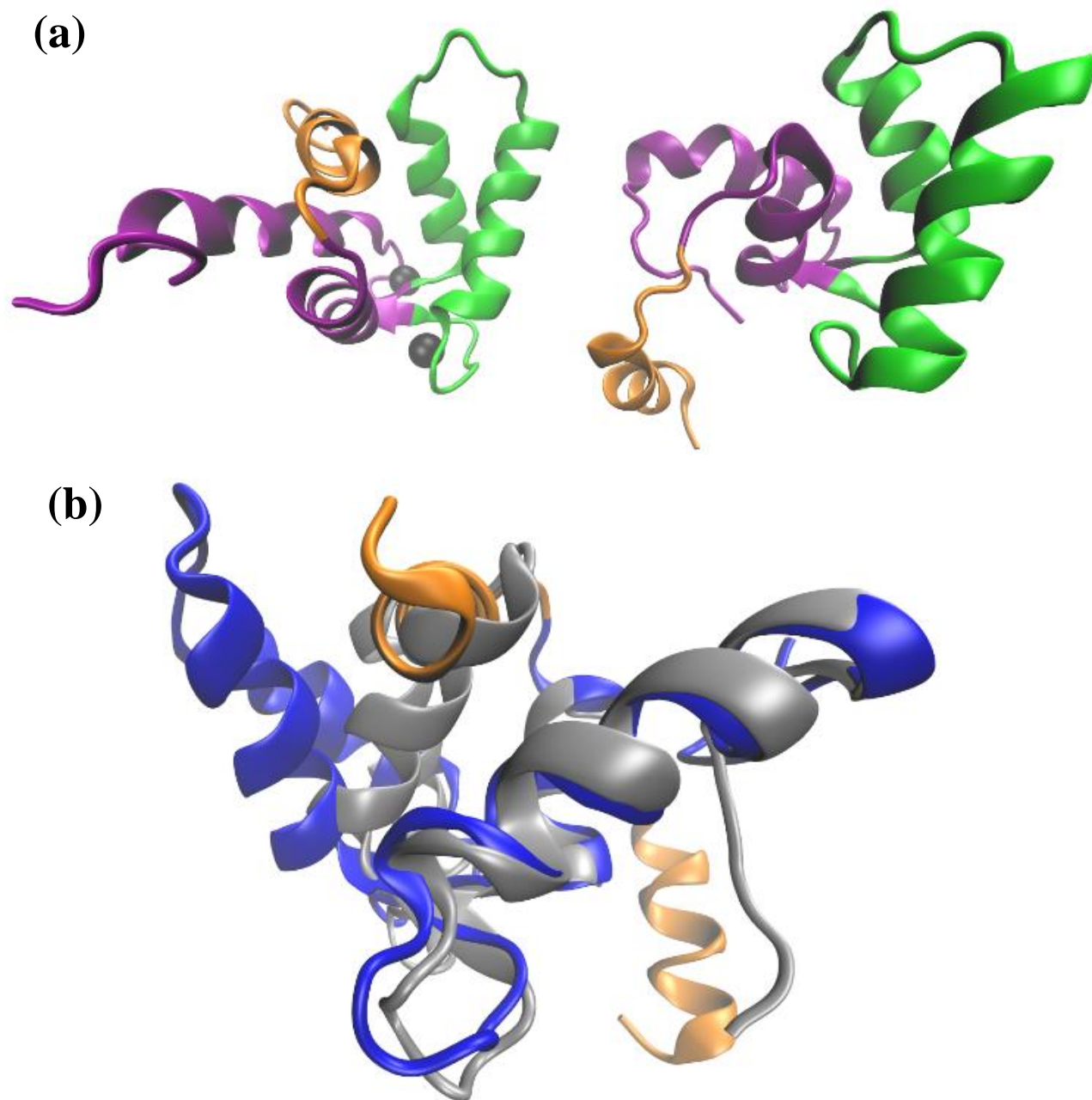


Figure 1.4 (a) The structure of L-plastins EF-hand domain in both holo (left) and apo (right) states. Purple and green structures represent both halves of the protein's binding pocket. The Orange α -helical structure represents the self-binding tail (H5-tail) observed by the Vogel lab. The α -helical H5-tail was added to the apo structure. Ca^{2+} are indicated by black spheres (only seen in the holo structure). **(b)** Apo L-plastin (grey) overlaid on holo L-plastin (blue). Orange again indicates the H5-tail.

Like in CaM, the EF-terminal domain of L-plastin undergoes a conformational shift after Ca^{2+} -binding to allow for the binding of target proteins [50]. Yet, L-plastin appears to show less binding promiscuity [50]. Similarly, to CaM, L-plastin contains a high number of methionine residues (~5% of the L-plastins EF-hand region) potentially stabilizing its conformational shift as discussed before [45, 50]. However, unlike CaM, the Met residues in L-plastin are located on the outside of the hydrophobic pocket [45, 50]. Therefore, where these residues are thought to enable promiscuous peptide and target protein binding in CaM, the L-plastin binding pocket appears to be designed to bind a single target (the H5-tail) [45, 50]. This is one example of how subtle differences in protein sequence and construction can lead to different mechanisms and cellular roles, and further stresses the importance of understanding the residue level mechanics of protein switches.

Interestingly both the EF-hand domain and the H5-tail of L-plastin were shown to have a strong effect on the proteins ability to bundle actin [50]. It was observed that when the H5-tail and the EF-hand domain were removed from WT L-plastin the two ABD's alone show a highly reduced actin-bundling activity [50]. Much of this activity is recovered when the H5-tail is returned, and full activity is only observed for the whole protein containing both EF-hands and the H5-tail (at low $[\text{Ca}^{2+}]$) [50]. This suggests that this EF-hand domain is not solely a regulatory domain, but also important for actin-bundling [50]. This has lead to the idea that the H5-tail and the EF-hand are used to help recruit actin filaments in a Ca^{2+} -free environment [50]. The protein is then able to self-regulate this function when the EF-hand domain is activated, by causing the EF-hand domain to bind, and inhibit, the H5-tail [50].

This self-binding H5-tail reveals that the binding action of L-plastin is distinguishable from CaM [50]. L-plastin appears to be selectively binding a region of unstructured protein attached to

the EF-hands (the H5-tail), where CaM has been shown to have promiscuous binding to a range of target proteins [50].

1.1.6 – The Physiological Role of L-Plastin

L-plastin is normally localized to hematopoietic cells. Within these cells, L-plastin has been shown to play an important role in macrophage functions such as bacterial host defence [49]. In leukocytes, L-plastin allows the rapid rearrangement of the actin cytoskeleton in order to respond to extra cellular stimuli [49]. L-plastin has been observed to be critical in the formation of cellular projections known as filopodia, which are observed in blood cells [50]. This homolog is further correlated with the development and efficacy of T-cell and B-cell leukocytes [49, 51, 52]. Additionally, neutrophils without L-plastin display reduced mobility and an inability to kill bacterial pathogens [49].

Interestingly, although L-plastin is found only in hematopoietic cells under normal conditions, its expression is highly upregulated in various non-hematopoietic tumours, making it a highly valuable biomarker for many forms of cancer [49]. L-plastin is thought to be a key indicator of a cell's ability to move throughout the body and has been also linked to tumour metastasis [48, 49]. In contrast, when T-plastin, normally found in solid tissues, is highly expressed in hematopoietic cell lines, these cells lose the capacity to move throughout the body [49]. Compared to T-plastin, L-plastin shows more sensitivity to Ca^{2+} , indicating a propensity for finer control over Ca^{2+} -induced conformational changes [48]. This increased control is thought to lead to a more rapid response to changes in $[Ca^{2+}]$ and ultimately allows for a more dynamic nature of the cells cytoskeleton, permitting cellular mobility [48].

Given that: (1) the EF-hand domain of L-plastin is ideally sized for computational study; (2) there are methods available to express, and isolate, this domain experimentally; (3) a novel self-binding tail mechanism was observed; and (4) L-plastins physiological relevance as a regulatory EF-hand, this domain was chosen for this research project.

1.2 – Significance of Thesis

1.2.1 – Goals of the Thesis

Although much is known about the initial and final states of Ca^{2+} -induced conformational shifts in EF-hand proteins, the exact dynamics of the process at a molecular level remain uncertain. It has been established that Ca^{2+} binding to the two negatively charged Ca^{2+} -binding sites within L-plastins N-terminal domain initiates the switching process [50]. Upon binding, the Ca^{2+} is too far away from most residues to have a significant electrostatic interaction, however the local electrostatic effects induced by Ca^{2+} on the binding site of the protein may propagate a global conformational change across the entire structure [56, 57]. Examples of global protein reorganization propelled by a local signal have previously been observed, and include the haemoglobin oxygen coupling, and various oligomeric enzymes, which can often reversibly associate and dissociate [56, 57]. The goal of this thesis is to determine the conformational mechanisms used by L-plastin's regulatory domain upon Ca^{2+} -binding through a combination of computational and experimental methods (for methods see Chapter 2).

1.2.2 – Validation of the Drude Force Field (Chapter 3)

Currently Ca^{2+} remains a difficult ion to assess in computational simulations as it is highly charged, relatively large, divalent, and polarizable. As this research project is centralized around the dynamics of L-plastin after binding to Ca^{2+} , therefore accurately portraying this ion was highly important for this research project. Here the novel Drude polarizable (Drude) force field (FF) was chosen to address this problem.

The novelty of the Drude (FF) necessitated some validation of its performance prior to its application in molecular dynamics (MD) simulations of L-plastin. These tests benchmarked some

key properties of the Drude FF that were previously unaddressed. In Chapter 3, the strength of hydrogen bonds (H-bonds), and the solvation free-energy of the Drude FF vs classical FF's are compared. The results indicated that the Drude FF provides a reasonable avenue to address polarizable systems as it represents the variation induced by polarizability better than classical FF. This validation justified using the Drude FF to simulate L-plastin's Ca²⁺-binding domain.

1.2.3 – L-plastin's Calcium Binding Regulation (Chapter 4)

This part of the project attempted to identify the Ca²⁺ induced conformational switching pathway of L-plastin, in a multifaceted manner. Initially, MD simulations were used with the Drude FF. This led to the identification of a novel molecular mechanism where the H5-tail was observed to regulate Ca²⁺-binding by occupying the empty Ca²⁺-binding sites found in the EF-hand in its apo state. These simulations were followed by experimental validation of the computational observations using isothermal titration calorimetry (ITC) and nuclear magnetic resonance (NMR) spectroscopy. Additionally, mutants were made in the H5-tail of the protein to determine experimentally whether the hypothesised interaction could be altered. The ITC and NMR results confirmed that L-plastin is capable of self regulating its Ca²⁺-binding using the attached H5-tail. These findings may have large ramifications for other ion-binding proteins.

1.2.4 – Future Directions and Conclusions (Chapter 5)

This research has provided new insights into protein dynamics and provided multiple avenues to continue with further investigation of the L-plastin headpiece. In adherence with the initial goal of fully elucidating the conformational dynamics of L-plastin, μ s time scale simulations of L-plastin have been generated and will be analyzed to help identify a complete pathway of switching. Further the discovery of L-plastins novel ability to self-regulate its Ca²⁺-binding

provides a unique opportunity to further understand this mechanism and look for other proteins capable of such dynamics. Finally, the ability to affect the activation of this protein potentially represents a valuable drug target, and therefore computational and experimental research will be undertaken into preliminary drug design.

Chapter 2 – Methods

2.1 - Computational Methods

2.1.1 - Molecular Dynamics Challenges: Simulating Calcium Binding to Polarizable Proteins

Molecular dynamics (MD) simulations require an accurate set of constraints to study any model system. Previous simulations involving Ca^{2+} have been problematic, since Ca^{2+} is highly charged, relatively large, divalent, and polarizable. Classical MD FFs (CHARMM, AMBER, GROMOS, etc.) represent each atom as a hard sphere with a fixed charge. This hard sphere representation is unable to account for the polarizable effects displayed by Ca^{2+} and, more importantly, protein atoms [20, 58, 59]. Although classical FFs can provide useful data in many systems, their inability to account for the polarizability of divalent ions leads to inaccurate results with errors in ion-protein interaction energies rising to 100-180 kcal/mol when compared to high-level ab-initio results [20, 58].

L-plastin represents a highly polarizable protein system which directly binds Ca^{2+} . Therefore, the accurate description of this binding process requires inclusion of electronic effects, such as induced dipole moments and the perturbation of macromolecular dipoles by cation binding/unbinding[20, 58]. This indicates that classical FFs may not be sufficient for assessment of Ca^{2+} binding in L-plastin.

2.1.2 – CHARMM Classical Force Field

Classical FFs represent each atom within a system as a hard sphere. These FF's rely on the addition of individual interaction energy terms that contribute to the atomic motion [60]. These terms include both bonded interactions, which help constrain the shape of molecules, and non-

bonded interactions, which provide an estimation of atomic forces between all atoms individually. The potential energy function for the CHARMM36 (C36) FF is given by eq. 2.1 [60].

$$U(\vec{R}) = \sum_{bonds} K_b (b - b_0)^2 + \sum_{UB} K_{UB} (S - S_0)^2 + \sum_{angle} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_\chi (1 + \cos(n\chi - \delta)) + \sum_{impropers} K_{imp} (\phi - \phi_0)^2 + \sum_{nonbond} \epsilon \left[\left(\frac{R_{min_{ij}}}{r} \right)^{12} - \left(\frac{R_{min_{ij}}}{r} \right)^6 \right] + \frac{q_i q_j}{\epsilon_1 r_{ij}} \quad (\text{eq. 2.1})$$

Although this additive approach cannot account for the polarizability of atoms and bonds, it can approximate dipole-dipole interactions through its non-bonded component [60], which comprises of Van der Waals interactions ($E_{\text{Van der Waals}}$) (eq. 2.2) and electrostatic interactions (E_{Coulomb}) (eq. 2.3) [60].

The Van der Waal interactions are divided further in two terms (eq. 2.2). The positive term describes Pauli repulsion between the atoms, which is a short-range repulsion force due to overlapping electronic orbitals. The negative term describes long-range attraction, arising from instantaneous dipole forces [60]. Together this equation represents the Van der Waal's interactions

$$E_{\text{Van der Wall}} = \epsilon \left[\left(\frac{R_{min_{ij}}}{r} \right)^{12} - \left(\frac{R_{min_{ij}}}{r} \right)^6 \right] \quad (\text{eq. 2.2})$$

The electrostatic interactions are described in the CHARMM FF by Coulomb's law, where the atomic charges are fixed at predetermined values (eq. 2.3). Although the CHARMM FF uses fixed charges, the electron distribution around the atoms is well known to vary in response to the external electronic field (a phenomenon known as polarizability) [61]. To account for this lack of polarizability, the CHARMM FF "enhances" the charge of bonded atoms to account for each atoms electronegativity [61]. This enhancement leads to unequal charge distribution in certain bonded interactions, to maintain the proper average charge observed [61]. This allows certain atoms to have a relatively low or high charge depending on their bonded interactions, and lets these bonded

interactions create “quasi-dipoles” between atoms [59, 60]. This ultimately creates molecular compounds with predefined permanent quasi-dipoles based on experimental data. Due to the relative charge difference created these predefined dipoles, in addition to charged atoms, can then be used to calculate electrostatic interactions using Coulombs law ($E_{Coulomb}$) (eq. 2.3) [60].

$$E_{Coulomb} = \frac{q_i q_j}{\epsilon_1 r_{ij}} \quad (\text{eq. 2.3})$$

Although these approximations work well with some models, the use of permanent dipoles to represent electrostatic interactions is a major simplification, which is inadequate in some cases (such as when dealing with Ca^{2+}). Therefore, a FF capable of addressing polarizability is necessary to accurately study L-plastin and related Ca^{2+} -binding proteins.

2.1.3 – Drude Polarizable Force Field

A new polarizable FF known as the Drude FF, provides a convenient solution to the problem of inaccurate assessment of polarizable effects. Building off the additive approximations used by the CHARMM FF, the Drude FF attaches an auxiliary pseudo- “Drude” particle to every atom (except for hydrogen atoms) by a harmonic spring [20, 58, 60]. For every heavy atom in the Drude FF, part of the atoms charge is transferred to its corresponding Drude particle, q_D [60]. Since part of the charge and mass for every heavy atom is transferred onto its corresponding Drude particle, a dipole (not a monopole) is created, while still maintaining the atoms original mass and charge [20, 58, 60].

This dipole allows each atom to respond to changes in the local electrical field (polarization). The harmonic spring used to attach each heavy atom to its representative Drude particle has a force constant, k , which enables variations in the polarizability for the heavy atoms within the system [60]. Once the system is placed in a uniform field, E , each Drude particle

oscillates around the surface of its corresponding heavy atom at a distance, d , calculated from eq. 2.4 [62]:

$$d = \frac{q_D E}{k} \quad (\text{eq. 2.4})$$

This displacement of the Drude particle from a central orbit allows for the calculation of the average induced dipole moment, μ , in response to the field strength according to eq. 2.5 [62]:

$$\mu = d * q_D = \frac{q_D^2 E}{k} \quad (\text{eq. 2.5})$$

Further, the polarizability of the system, α , can be described by eq. 2.6 [62]:

$$\alpha = \frac{\mu}{E} = \frac{q_D^2}{k} \quad (\text{eq. 2.6})$$

In 2015, the Noskov (University of Calgary) and Roux (University of Chicago) labs published new parameters for metalloproteins in the Drude FF which describe the interactions controlling the protein structural change upon Ca^{2+} binding more accurately than classical FFs [20, 62]. These new parameters for divalent ions and the use of the Drude FF resolve some of the problems associated with simulations involving Ca^{2+} and permit the modeling of systems, such as L-plastin, while accounting for polarizable effects. In the present work, the Drude FF was used in several MD simulations of WT and mutated L-plastin with Ca^{2+} to identify the effects of induced polarization, the role of polarizable dipoles in the stabilization of bound Ca^{2+} , and to provide additional testing of the Drude FF and parameters on a biologically relevant protein model.

While the Drude FF has performed well in numerous tests it has not undergone thorough comparison with classical FFs to describe secondary structure dynamics with water. Experimentally and theoretically, water dynamics within the first hydration shell have been shown to drive the folding and unfolding processes of proteins [20, 58]. To address this, an additional

ensemble of small ligand structures was generated to specifically test the free-energy of solvation and hydrogen bonding (H-bonding) dynamics with the Drude FF (Chapter 3). These structures were simulated with both the Drude FF and the C36 FF to (1) provide an in-depth and quantitative analysis of the H-bonding lifetimes observed between FF's and (2) probe how the solvation free-energy is affected by both the permanent and induced dipole/charge interactions. These simulations clearly demonstrated the Drude FFs ability to represent polarizability leads to obvious differences in these terms compared to classical FFs, and provides a measurable benchmark for the Drude FF. This work validated the use of the Dude FF in further computational studies of L-plastin.

2.1.4 – Hydrogen Bonding Simulations and Analysis

The Drude FF needed to be directly compared with classical FFs and experimental results to calculate the relative strength of H-bonds. The formation of H-bonds is essential for solvation processes and biological functions such as enzymatic catalysis and the stabilization of protein and DNA structures [63-66]. The strength of H-bonds has been shown experimentally to vary greatly ranging from ~4.9 kcal/mol for peptides in the gas phase to 0.5-1.5 kcal/mol when proteins are solvated in water [63]. This relative change in H-bond strength stems from an increase in the entropy of solvated systems due to the formation of H-bonds between protein and water atoms in addition to H-bonds formed between protein residues [63-65]. The ability of water to form H-bonds with proteins facilitates the breaking and reformation of H-bonds between protein atom pairs and decreases the large enthalpy cost of H-bond breaking for proteins in a non-solvated environment [63-65]. Ease of formation/disruption of H-bonds in an aqueous environment is central to proteins ability to fold properly, and within biologically relevant timescales [63-65].

Given the importance of H-bonds, the development of FFs and water models which can properly represent H-bond strength is critical.

In the present work the strength of H-bonds were evaluated between protein-protein and protein-water interactions with the polarizable SWM4-NDP [67] and non-polarizable TIP3 [68] water models for the Drude FF and CHARMM36 (C36) FF respectively. These water models are commonly used in calculations with the Drude FF and C36 FF [67, 68].

For analysis of the H-bond data a strictly geometric criterion for the definition of H-bonds developed by Reddy et al was used [1]. The same definition of H-bonds was used both for protein-protein and protein-water H-bonds and allowed for a frame by frame analysis of computational simulations to determine the rate at which the individual H-bonds were breaking and forming. This criterion defines a theoretical cone extending from each hydrogen (H_1). If an atom capable of acting as a H-bond acceptor (X_A) is (1) within 3.7\AA of the heavy atom attached to H_1 , and (2) θ_1 is between of $0^\circ - 30^\circ$ then H_1 can act as a H-bond donor (figure 2.1). Additionally, a similar cone is generated around each H-bond acceptor (O). For the H-bond donor (X_D) to can act as a donor, θ_2 must be between $80^\circ - 140^\circ$ (figure 2.1). Therefore, a H-bond is formed if both H_1 can act as a H-bond donor, and O can act as a H-bond acceptor. This criterion is represented in figure 2.1.

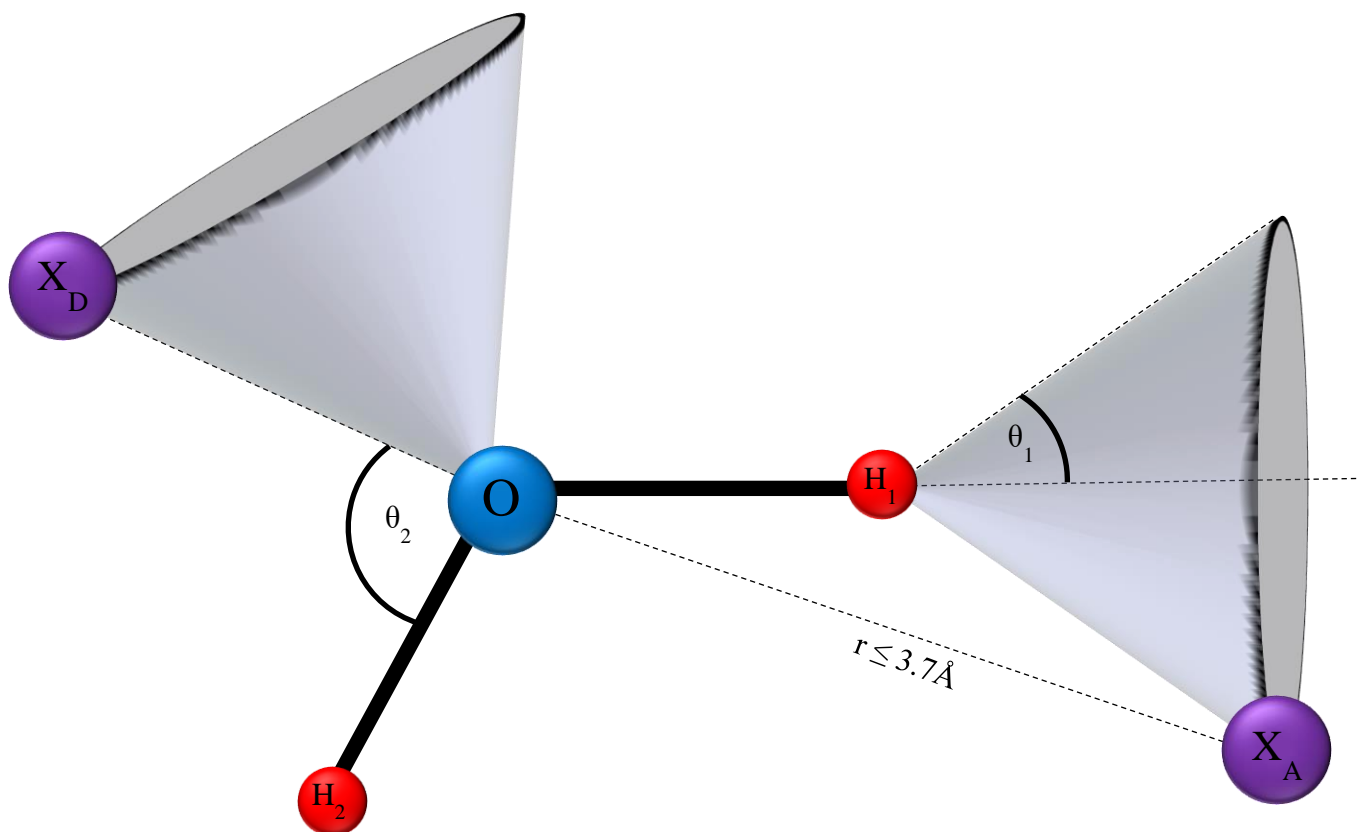


Figure 2.1 The geometric criteria used to define H-bonds. Figure is adapted from Reddy et al. [1]

Based on these criteria for formation and dissociation of H-bonds, the total probability of a H-bond being present at any time, t , was derived using two time correlation functions (TCFs) [64]: (1) the strict lifetime correlation function ($S(t)$) and (2) the continuous lifetime correlation function ($C(t)$), defined by eq. 2.7 and eq. 2.8, respectively:

$$S(t) = \frac{\langle h(0)H(t) \rangle}{\langle h \rangle} \quad (\text{eq. 2.7})$$

$$C(t) = \frac{\langle h(0)h(t) \rangle}{\langle h \rangle} \quad (\text{eq. 2.8})$$

The two TCFs depend on H-bond populations, $h(t)$ and $H(t)$. $H(t)$ equals 1 if any pair of atoms that are originally bound at time 0, t_0 , remains continuously bound until any future time, t . If a H-bond is broken at any time between t_0 and t , then $H(t)$ equals 0. Alternatively, $h(t)$ equals 1 if any pair of atoms that are originally h-bound at t_0 are also bound at t . If a H-bond is broken at any time after t_0 $h(t)$ equals 0 at that time, however if this bond is re-formed at t then the value of $h(t)$ returns to 1 at t .

Because many residues form multiple H-bonds these two TCF's calculate the probability of H-bonds being present at (t) . $S(t)$ calculates the probability that a H-bond with any residue will remain continuously present, without allowing for H-bonds to dissociate and reform. This means that the $S(t)$ function gives a strict definition of H-bonds. Alternatively, $C(t)$ defines the probability of H-bonds being present at t , allowing for H-bonds to break and reform. This means that $C(t)$ gives a continuous definition of H-bonds.

Because $C(t)$ allows for the reformation of H-bonds, and is independent of bond breaking, when this evaluates the formation of H-bonds between protein and water it can also act as an indicator of water diffusing away from a specific peptide. If water is maintained close to the

protein, there is an increased likelihood of H-bonds reforming. Thus, when $C(t)$ permanently reaches 0, this indicates that all the water initially surrounding a peptide has diffused away.

Using these TCFs as an indication of the probability of H-bonding, it is possible to fit this data to an exponential function (eq. 2.9) to obtain the average lifetime, τ , of the H-bonds. Because both $S(t)$ and $C(t)$ are defined based on the H-bonds present at t_0 , both TCFs give a diminishing probability of H-bonding, starting from 1 and eventually reaching 0. As the value of t is known for each corresponding probability, $F(t)$ ($S(t)$ or $C(t)$), it is possible to identify the τ of each H-bond within the system. This is possible through fitting the data to an exponential decay function (eq. 2.9). When applied to $S(t)$ this gives the average strict H-bond lifetime (τ_s), while for the $C(t)$ it gives the average continuous H-bond lifetime (τ_c).

$$F(t) = e^{\frac{-t}{\tau}} \quad (\text{eq. 2.9})$$

Many papers use a multiexponential function to fit the H-bond lifetime data claiming that such a function can yield individual TCF for various molecular motions (vibrational, rotational, etc.) [63-65]. However, in these papers, all calculated motions are averaged to give a single lifetime, indicating a single exponent is more accurate. Further this function saves computational time and leads to smaller errors in data fitting. Therefore, a single exponential function was used to estimate the life times of each residue. Additionally, because many of the data sets rapidly approached zero and stayed at zero indefinitely afterwards, data was used until $F(t) < 0.1$.

2.1.5 - Hydration Free-Energy of Side Chain Mimetics

Another essential component of any FF is the hydration free-energy of amino acids in water. This property has been extensively quantified for classical FFs, however this term in the Drude FF remains untested [69, 70]. The solvation free-energy essentially describes the tendency

of a system to associate and react with its surroundings [70]. In this way the free-energy of solvation is a good indication of how well a FF is capable of representing the interactions of amino acids with their surroundings [70]. Given the phenomenal impact that water has on biological and protein systems, with direct impacts on protein folding, enzyme activity, structural stability, etc., accurately portraying this property in a FF is of huge importance [5, 69, 70].

To identify the hydration free-energies of mimetics within the Drude FF a method defined by Peter Kollman and previously used by Deng and Roux to calculate the hydration free-energy for classical FFs was used [69, 70]. This well-defined method decouples the free-energy of each amino acid mimetic into separated interactions [69]. The interaction components include a short range, non-polar, contribution (ΔG_{np}) and an electrostatic contribution (ΔG_{Charge}) [69]. The non-polar contribution is further broken into repulsive (ΔG_{Rep}) and dispersive (ΔG_{Disp}) components, making the full equation represented by (eq. 2.10)[69].

$$G_{tot} = G_{Charge} + G_{np} = G_{Charge} + G_{Rep} + G_{Disp} \quad (\text{eq. 2.10})$$

These terms can be calculated directly using free-energy permutation (FEP) techniques [69]. This allows for both a full free-energy to be calculated, but importantly allows these separate interactions to be accounted for individually. This is especially important when considering the Drude FF due to its polarizable nature. Because of its polarizability the ΔG_{Charge} term may represent a much more variable interaction than seen by classical FFs.

A strength of using this method is the ability to directly compare the outcomes achieved using the Drude FF with both Classical FF and experimental values. This method has been previously used to identify the free-energy of Classical FF's, but additionally it is possible to obtain free-energy values of the individual amino acid mimetic values experimentally. This means that it

is possible to clearly identify which FF is able to more accurately portrait the experimental value. One caveat to consider here, is although it is possible to identify the free-energy value experimentally it is not possible to decouple the experimental value as it is done computationally. Therefore, relating various contributions between classical and polarizable FF's is possible, however it is hard to justify if the decoupled contributions are accurate.

2.1.6 - Mapping of Ion-Dependent Conformational Switching

L plastin is an example of a protein that undergoes ion-dependent conformational switching between two well-defined and characterized states via an unknown pathway. Here, this pathway started to be probed using a powerful arsenal of computational and experimental tools, including the novel Drude FF.

Mapping the transition pathway which connects two states requires the non-trivial, and computationally challenging, generation of intermediate geometries [71-74]. These intermediate geometries are then used to reconstruct the energy landscape of the transition pathway [71-73]. In the present work annealing simulations were used to generate a promising array of intermediate conformations from both the holo and apo forms of L-plastin.

2.1.7 - Annealing Simulation

Annealing simulations are commonly used as a starting point to generate intermediate structures in studies similar to the present work [74]. The annealing method involves running multiple simulations of each of the original well-defined states at systematically increasing temperatures [74]. The high temperature simulations allow the protein to rapidly overcome energy barriers which would normally trap the structure for an extended period of time in lower energy states. Thus, running the simulation at multiple increasing temperatures achieves a better sampling

of a highly diverse array of structural conformations of the protein that may not be observed over the same timeframe at a single lower temperature.

Although this method improves sampling it often leads to biologically improbable, denatured, conformations due to the elevated temperatures. To convert these conformations into biologically relevant structures the denatured systems are subsequently exposed to a lower temperature with the hope that the protein reassembles. The pathway is further used for identification of recognizable states that may provide additional structural configurations as the protein converts into a more stable state.

This method was used to sample both the holo and apo structures and to generate a large array of conformations, with the purpose of determining some key overlapping structures, stemming from the initial states. The overlapping structures could represent transition states between the apo and holo states, and lead to the identification of a full transition pathway in L-plastin.

Annealing simulations also provided an estimation of the temperatures at which the L-plastin begins to denature. This temperature was then used to establish a proper simulation protocol in future simulations that allowed sampling of the necessary high-energy conformations of L-plastin without leading to structural collapse.

The data generated by the annealing procedure was to initially identify interesting protein conformations. Further, frames from this simulation were used to start replica exchange with solute tempering (REST) simulations, in order to generate a more comprehensive landscape of the possible conformational changes in L-plastin [73].

2.2 - Experimental Methods

2.2.1 – Need for Experimental Investigation

Theoretical findings were further subjected to experimental validation. This validation is crucial for the evaluation of results based on novel computational methods, and codes, such as the ones used in the theoretical portion of this work. In addition, experimentation was necessary to assess the quality of the generated L-plastin conformational switching models.

2.2.2 – Bacterial Mutagenesis and Protein Purification

Several interesting states, which are stabilized by a few key amino acids, were identified from the annealing simulations. Therefore, a wild-type (WT) protein construct (EF-WT) and several mutant protein constructs targeting these key amino acids within the EF-hand domain of L-plastin were generated, where the identified amino acids were mutated to alanine. These protein constructs were then subjected to isothermal titration calorimetry (ITC) and nuclear magnetic resonance spectroscopy (NMR) (see below) to assess how each mutation impacted the energy profile and structure of the protein constructs. This provided an avenue to directly test the mechanisms determined by computational methods.

2.2.3 – Isothermal Titration Calorimetry (ITC)

Isothermal titration calorimetry (ITC), is a highly sensitive method that can measure the heat released during protein conformational dynamics, and further yields an energy profile of the process. In this method two cells of identical volume are maintained at a constant temperature [75]. The first of these cells (the control cell) holds a buffer, while the second cell (the sample cell) contains a protein sample (in this case, EF-WT or a mutated L-plastin protein construct) solvated in an equal volume of the same buffer [75]. A syringe, containing the same buffer with an added,

and a known, concentration of a ligand (Ca^{2+} in this case) is then inserted into the sample cell. As the experiment progresses, small volumes of the ligand are injected from the syringe into the sample cell. When Ca^{2+} binds to L-plastin, the protein undergoes the conformational switching process and energy is released and recorded as minor temperature changes by the ITC instrument [75].

Through multiple small injections, an overall energy profile of the switching process can be obtained, this profile can then be fitted to obtain an array of thermodynamic parameters which regulate the process [75]. An overall energy profile for the process of protein switching can be recorded and used to estimate the association constant (K_a) of the protein and ligand. The overall temperature change can additionally be used to calculate the enthalpy of binding (ΔH) [75].

Once the K_a of the protein is known the Gibb's free-energy (ΔG) can be calculated from eq. 2.11 where R is the universal gas constant, and T (in $^{\circ}\text{K}$) is the temperature the experiment was run at [76].

$$\Delta G = -RT \ln K_a \text{ (eq. 2.11)}$$

Further, as ΔG can be calculated, and the enthalpy is known, the entropy (ΔS) can be directly obtained from eq. 2.12 [76, 77].

$$\Delta G = \Delta H - T\Delta S \text{ (eq. 2.12)}$$

Using this data, changes between the binding profiles of multiple proteins can be evaluated. Thus, ITC can be used to construct the thermodynamic energy profile of binding for each protein construct. Additionally, although ITC records the K_a directly, the equilibrium dissociation constant, K_d , can also be obtained as the inverse of K_a (eq. 2.13).

$$K_d = \frac{1}{K_a} \text{ (eq. 2.13)}$$

After purified protein constructs were obtained, ITC was used to generate an energy profile of the conformational switching process for each of these protein constructs upon Ca^{2+} binding. The energy profile of the EF-WT protein was used as a benchmark for comparison to the mutant constructs. To ensure accurate, and comparable, energy profiles were obtained for all protein constructs, the starting concentration of Ca^{2+} was kept the same for all trials. To that end, each sample was buffer exchanged with a Ca^{2+} -free buffer using an Amicon Ultra Centrifugal Filter to over a 1,000,000x dilution until a “ Ca^{2+} -free” state was achieved. Once Ca^{2+} had been fully removed, the exact concentration of Ca^{2+} could be identified for every step of the energy profile as a solution of a buffer containing 1.2mM of Ca^{2+} was re-applied to the protein sample, in injections of 2 μL .

Although this method can very accurately obtain the energy released during protein dynamics, not all the energy is directly associated with the protein conformational dynamics. The majority of energy is released through conformational dynamics, such as helix formation and the opening of the hydrophobic core, however a small percentage of the energy released is associated with processes such as diffusion [75, 76]. Energy is also released when the ligand binds the protein regardless of whether conformational changes takes place [75, 76].

ITC provides a viable avenue for exploration of the energy profile of L-plastin. However, it is limited in that the changes in energy output can be attributed to many factors and the altered energy in these experiments may not correlate well with the simulation results. Therefore, in addition to ITC nuclear magnetic resonance spectroscopy (NMR) was used to probe the structural changes of the L-plastin protein and its mutant constructs.

2.2.4 Differential Scanning Calorimetry (DSC)

Differential scanning calorimetry (DSC) is a thermoanalytical technique method often used in biological systems to evaluate the stability of proteins. In a DSC instrument there are two identical cells: one cell (the reference cell) is filled with only a buffer, while the other cell (the sample cell) maintains an equal volume and is filled with the same buffer solvated with a known concentration of protein [78]. Once these cells are filled, they are heated at an equal rate, with the expectation that any differences in energy required to keep the two cells at a constant, and increasing, temperature are associated with the heat capacity (C_p) differences between the two cells, and therefore associated with the C_p of the protein [78]. This C_p is directly related to the heat flow into the cells and can be determined using the thermal equivalent of Ohm's law (eq 2.14)

$$q = \Delta T / R \text{ (eq 2.14)}$$

Where q is the sample's heat flow, ΔT is the temperature difference between the sample and the reference, and R is the resistance of the thermoelectric disk [78].

Although the C_p of the protein may slightly alter the solution over all temperatures, a peak in C_p is observed when the protein structure begins to melt. This is as during the melting process the protein is undergoing a phase transition process, and therefore an increased amount of energy is required to transition the protein from a structured to a denatured state. Therefore, to keep the two cells at a consistent temperature a relatively large influx of energy is required over this transition phase, which appears as a peak in the DSC experiment. The amount of energy required for this transition phase to take place can be used to directly calculate the enthalpy of this melting process (ΔH_{melt}). Additionally, the melting temperature itself provides useful information about the stability of the protein structure. If the protein melts at a higher temperature, this represents

that the structure is more stable, while if the protein melts at a low temperature, this represents that the protein is less stable. This method provided a means to directly evaluate the stability of the protein constructs of L-plastin generated, specifically in their holo state.

2.2.5 – Nuclear Magnetic Resonance Spectroscopy (NMR)

NMR machines (and MRI machines, the physically larger, but magnetically weaker, medical counterparts) can generate a uniform magnetic field within a cylinder (where the sample is placed). This field is created by a coil of electromagnets which surround the cylinder. The electromagnet coils are often made of a super-conducting metal cooled by liquid helium and can maintain large electrical currents indefinitely (thus maintaining the magnetic field), assuming the metal coil remains super-cooled [79].

NMR machines use this generated magnetic field to target atoms such as ^1H , ^{13}C , and ^{15}N . These atoms, and many more, all contain a nuclear spin (spin) and are important for NMR as the spins of these atoms aligns with the uniform magnetic field generated by NMR machines. Any atom with a spin can be used for NMR measurements, however the most commonly used atoms have nuclear spin of $\frac{1}{2}$, as these are the “easier” atoms to run spectroscopy on [79].

To generate a signal, the NMR machine applies a radio frequency pulse to the sample, which is initially aligned with the permanent magnetic field, this pulse rotates the alignment of nuclear spin states [79]. Once the radio frequency pulse ends, the now unaligned spin states of sensitive nuclei attempt to realign with the permanent magnetic field. During this realignment, a current is created in a perpendicular coil of wire, which creates a measurable, but weakening, sinusoidal wave of voltage known as a free induction decay (FID) [79]. This wave is subsequently subjected to a Fourier transform (FT) to yield a frequency-dependant spectrum [79]. Because each

atom experiences the effects of the magnetic fields to a different extent, depending on the atoms local environment, the signal frequencies of atoms can be distinguished from one another [79]. The local environment of each atom can then be directly linked to molecular structure. Thus, it is possible to identify which atoms belong to specific residues, and which are there neighboring residues [79].

2.2.6 – Heteronuclear Single Quantum Coherence (HSQC) Experiments

There are many ways to use NMR data to obtain structural data, all with different applications and benefits, however, in this work only heteronuclear single quantum coherence (HSQC) spectrums were used to evaluate the structure of the generated L-plastin protein constructs.

HSQC experiments are a 2D NMR experiment where the ^1H signal of a molecule is correlated with another heteronuclear atom, often ^{13}C or ^{15}N (in these experiments to ^{15}N) [79, 80]. This method relies on J-coupling between ^1H and ^{15}N [79, 80]. J-coupling only occurs between atoms connected by chemical bonds (usually with a detection limit of 3 bonds, however up to 5 bonds is sometimes seen) [79, 81]. In HSQC spectra however the 1J coupling is often used, therefore only ^1H directly attached to ^{15}N atoms show up in this spectrum [79-81]. Using a $^1\text{H},^{15}\text{N}$ HSQC spectrum is especially common for analyzing proteins since all amino acids (except for proline) have a peptide amide bond that will appear on this spectrum, meaning each residue gets an individual and unique peak which can be identified [79]. Once these peaks have been assigned to unique residues in the protein this experiment is relatively quick to run and is easily interpretable [79].

This method is often used in combination with other methods to determine structure. Alone HSQC spectrums can give a reasonable indication of how similar two protein structures are, but it does not provide enough structural data to obtain a 3D model [79]. Therefore, other NMR methods, such as nuclear Overhauser effects (NOE's), are used to obtain direct 3D structural data [79]. Methods such as NOE's are capable of directly determining distances between residues and are therefore capable of directly determining a 3D structure if enough distances are identified [79].

One other set back to using an HSQC is that although the natural abundance of ^1H is very high (>99%), the natural abundance of ^{15}N is low (0.368%), as is the natural abundance of ^{13}C (1.07%) [79, 81]. This low natural abundance creates problems identifying clear signals, as a low abundance creates a weak NMR signal, making peak identification difficult (or requiring significantly more time to clarify peaks). Because of this, the abundance of ^{15}N (or ^{13}C) needs to be increased by growing bacteria in isotopically enriched minimal media [79]. By using enriched media, the bacteria building the protein are guaranteed to use NMR sensitive atoms, which guarantees that all proteins maintain a high percentage of NMR sensitive atoms often leading to identifiable, and strong, signals [79]. $^{15}\text{NH}_4\text{Cl}$ was used in bacterial media as the source of ^{15}N . This source of ^{15}N was chosen as ammonium is the preferred source of nitrogen for *E. coli* (the bacterial host) and is the only inorganic source of nitrogen which these bacteria can be grown with, additionally this source of ^{15}N has already been shown to work when expressing the EF-WT protein construct [50, 82].

In this project the NMR structure of both apo and holo L-plastin were previously identified [50] and the 2-dimensional ^1H , ^{15}N -HSQC signal peak assignments for both the apo and holo states of the EF-WT L-plastin were already known [50]. Therefore, comparison of the HSQC peaks between the newly generated, and previously determined, EF-WT protein construct of L-plastin

could be used to ensure that the EF-WT protein structure generated was identical to the previously determined structure.

Given that the NMR peaks are only shifted due to changes in the local environment of the protein, it was possible to identify which areas of the protein were affected in mutation constructs by overlapping the EF-WT spectra with mutant protein HSQC spectra and observing which peaks had shifted.

As a change in the location of an NMR peak directly indicates a change in the local environment, changes in a peak location provides a reasonable identification of structural changes in proteins. This is not universally true, as it is possible for two structurally distinct states to provide a similar local environment for any given atom. With this said, if two structures differ minimally, the areas which show chemical peak shifts on the spectra likely represent the areas of structural change in the protein construct. In this way HSQC spectra can be used as an indication of the extent a protein structure is changing. A small number of shifted peaks usually indicates a localized environment change (pinpointing where the structure is changing), while generally a more major reorganization of all the peaks indicates a more complete structural overhaul. Similarly, larger shifts to individual peaks are often thought to indicate larger changes in the local environment of that residue, while smaller shifts imply that the local environment of that residue remained relatively unchanged. Peak shift size can thus be used as an indication of residues which have undergone the largest structural changes. All mutant protein constructs analyzed were shown to have a significant overlap with the EF-WT peak assignment and were considered structurally alike to the EF-WT construct, allowing an avenue to compare the EF-WT with all the mutated protein constructs.

Chapter 3 - Hydration Properties of Amino Acid Residues from Molecular Dynamic Simulations with Both Classical and Drude Polarizable Force Fields

3.1 Abstract

Recent development on a polarizable force field known as the “Drude” polarizable force field, based on the extension of an induced dipole model, has reached a milestone in the past few years by providing a complete set of parameters for protein simulations. The Drude polarizable force field enables stable simulations of protein with explicit (and fully polarizable) water, counterions, and some lipid types; and is further capable of reaching up to microsecond long simulations timescales. It is, however, not yet thoroughly compared with well-established classical force fields for modelling the secondary structures of small peptides in explicit solvent. The complex, and mutually-dependent, dynamics of inducible dipoles of hydrogen bond networks formed by water-water, protein-water, and protein-protein interactions is expected to have a major impact on the stability of protein structure. Here, a direct comparison between the current Drude polarizable force field, comparative classical force fields, and experimental values is provided for (1) the solvation free-energy of mimetics for all amino-acid side-chain equivalents, (2) the lifetime of hydrogen bonding in protein-water and protein-protein interactions, and (3) the comparative lifetime of the water on protein surfaces.

3.2 - Introduction

The development of better force fields (FFs) which more accurately represent biopolymer dynamics in the aqueous phase is an issue of paramount importance within computational structural biology. Classical FFs have provided significant insights onto the hydration of proteins and DNA, allowing short-(pico- to nano-second) and long-time scale (up to milliseconds) dynamics of biopolymer which rely on fixed-charge approximations to model induced polarization implicitly [59, 62]. The impressive performance achieved by classical FFs comes at the cost of an inability to induce polarizability relative to the environment, which impedes simulations in cases where charge transfer or induced polarization effects are of importance [59, 62]. With the current advancement of supercomputers, and effective algorithms, there is a significant demand on developing and applying accurate FFs that explicitly include electronic degrees of freedom [83-85]. While it can be challenging to study protein dynamics with quantum-mechanical (QM) calculations directly and still achieve microsecond time scales, there are multiple alternatives emerging as part of a concerted community effort to develop models which approximate polarization in the condensed phase (polarizable force-fields paradigm) [84, 86-89]. If a system undergoes a chemical reaction with significant charge transfer, bond-breaking, or bond-formation a QM approach is often unavoidable [90, 91], but using a QM approach comes at a significantly increased computational cost. However, there are many fundamental problems in biophysical chemistry where no new bonds are formed or broken, but local polarization effects are still essential for understanding the system's chemistry. For example, induced polarization effects play a significant role in strong electrostatic, but still non-covalent, interactions such as ion binding and subsequent transport by membrane proteins [20, 92], the ion-dependent stabilization of DNA [93, 94], ion-modulated stability of enzymatic sites [95, 96], and complex electrostatic responses at biological interfaces [97]. These are just a few of many central biochemical problems where the

explicit treatment of many-body effects and local electrostatic responses to a variety of stimuli (binding of an ion, change in local hydration, etc.) are expected to improve the accuracy of molecular simulations. Another canonical problem where induced polarization effects are expected to play a significant role, is in the dynamics of complex hydrogen-bonding networks. Preliminary studies of hydrogen-bond dynamics in complex water-alcohol and water-amide mixtures indicate a critical role of explicit polarization treatment in modelling of anomalous concentration-dependent thermodynamic and transport properties [96, 98, 99].

At least three polarizable FF development projects have already reached the stage required for MD simulations of an entirely polarizable protein embedded into an aqueous environment. The apparent need to explicitly model induced polarization often requires conservation of the Self-Consistent Field (SCF) regime at every integration step, which is often expensive (used by AMOEBA, FlucQ etc.) [62, 89]. This iterative approach to maintain the SCF has gained some significant success in studying materials, polarizable systems, and so on, however this approach requires at least 16x more computational cost than non-polarizable FF's [62]. This is due to each step of the simulation requiring roughly 16 evaluations of the atomic forces to converge on the desired criterion for the self-consistent field ($F_{\text{rms}} < 10^{-6}$ kcal/mol/Å) [62].

An alternative approach, which has gained significant momentum recently, is to introduce electron polarization via the addition of an auxiliary particle with a light mass which is coupled to a separate heat-bath (Drude particle). This model is known now as the “Drude Polarizable Force-Field” (Drude FF), and is referred sometimes as a Charge-On-a-Spring Model [100]. The physics of the model is based on Paul Drude’s treatment of electrons as light particles attached to massive nuclei via harmonic springs (1902) [101], with applications for modelling of the inter-molecular interactions in electron gas [102]. The wealth of QM data available for model compounds,

regarding electrostatic responses and interactions with various probes, enables the accurate fitting of charge separation between the main atom and the Drude particle, which is controlled by the polarizability parameter. As well this data provides the required benchmarks for fitting intramolecular parameters [98]. This relatively simple approach allows the Drude FF to create induced dipoles, and therefore enables polarizable effects to be modeled in the condensed phase. The Drude FF has been tested in applications with membrane systems, water-alcohols, metallo-proteins, and nucleic acids [62, 67, 84, 93, 99, 103-106]. Although this model does have known limitations on its treatment of polarizabilities [107], this approach is more computationally favorable as it requires only a $\sim 4x$ increase in computational cost to simulate polarizability. This increased cost stems from (1) simulating the additional Drude particles ($\sim 2x$ the particles), and (2) usually requiring a smaller timestep ($\sim 1/2$ that of classical FFs) to allow for system stability. One advantage of the Drude polarizable FF is that this model can be relatively easily ported to most of the popular MD software packages as it can be used with a dual-thermostat Langevin integrator [108] or other developed integrators based on the extended-Lagrangian formalism [109].

With the Drude FF coming of age, it is worthwhile exploring the similarities and added value this FF provides in the description of one of the most basic and fundamental phenomena in protein chemistry: the hydration of individual amino acid residues and various secondary structure elements. There is little doubt that the conformational dynamics of protein are complicated and affected by water binding patterns on the proteins surface. For example, soluble proteins often have a hydrophobic core and a hydrophilic surface with properties that can fluctuate significantly; occasionally even trapping water molecules within loop conformations for a few nanoseconds [32, 33]. Further, IR spectroscopy has shown that water dynamics on a proteins surfaces are roughly eight times slower than observed in bulk water [64, 65]. This frequently observed evidence

suggests that FF's must be able to accurately capture some crucial dynamics of water on protein surfaces. It is also evident that induced polarization is expected to play a role in the stabilization of water on protein surfaces, but it is unknown if this can be accurately captured with already existing additive FFs such as the CHARMM36 (C36) FF.

Here, a comprehensive analysis of small protein dynamics and thermodynamics in water is presented. First, solvation energies of individual amino acid mimetics were computed using the Drude FF and compared with both free-energy values previously obtained using classical FFs and experimental free-energy values. Secondly, four different globular proteins containing the most common secondary structure elements were used in MD simulations with the C36 and Drude FFs to generate a large set of hydrogen-bonding (H-bonding) pattern data used for analysis. This comprehensive data of the hydration free-energies for amino-acid residues, together with analysis of H-bonding data, allows for a detailed comparison between the approximations made to account for polarizable effects in the C36 and Drude FFs.

3.3 Methods

3.3.1 - Hydration Free-Energy of Side-Chain Mimetics

The values of the hydration free-energies for the molecular mimetics of the amino-acid residues calculated with the Drude polarizable FF are listed in Table 3.1. All of the values for hydration free-energies for the additive FF (C27-C36) are as reported by Deng and Roux[69]. For the sake of consistency with all other simulations reported here, the first stage of this study re-computed the free-energy of hydration for all 20 amino acid residue mimetics. Specifically, the free-energies of hydration were calculated *via* the Free-Energy Perturbation (FEP) method with the staged protocol of Deng and Roux[69]. The interaction free-energy component of the solvation free-energy ($\Delta G_{\text{int}}^{\text{solvent}}$ where *a* represents *site* or *bulk*) is comprised of three combined components: the electrostatic (ΔG_{dec}^a), dispersive Lennard-Jones (LJ) (ΔG_{dis}^a) and repulsive LJ (ΔG_{rep}^a) potentials [69]. The non-electrostatic LJ potential was separated into purely repulsive and attractive parts using Weeks, Chandler, and Andersen's (WCA) decomposition scheme[110, 111]. An in-depth discussion of the staged FEP simulations, and its applications to studies of hydration processes with polarizable FFs, can be found in Baker et al. [112]. Using the WCA representation combined with a traditional FEP approach, the repulsive and dispersive components of the solute-solvent interaction potential, $U^{\text{LJ}}(r)$, can be expressed as:

$$U^{\text{LJ}}(r) = \epsilon \left[\left(\frac{R_m}{r} \right)^{12} - 2 \left(\frac{R_m}{r} \right)^6 \right] \quad (\text{eq 3.1})$$

Which can be further decomposed into:

$$U^{\text{LJ}}(r) = U^{\text{rep}}(r) + \epsilon \quad (\text{when } r < R_m); U^{\text{LJ}}(r) = 0 \quad (\text{when } r \geq R_m) \quad (\text{eq 3.2})$$

$$U^{disp}(r) = -\varepsilon \text{ (when } r < R_{min} \text{)}; U^{LJ}(r) = U^{LJ}(r) \text{ (when } r \geq R_{min} \text{)} \quad (\text{eq 3.3})$$

where ε describes the dimension of energy, and R_{min} the dimension of length. When the separation, r , of two atoms is at R_{min} , the LJ potential reaches its well depth, $-\varepsilon$.

The staging in FEP simulations, turning off and on the solute-water interactions, was achieved by using three coupling (or staging) parameters (s , ξ , and λ) implemented in the PERT module of the program CHARMM. The repulsive component of the hydration free-energy was obtained with the staging parameter s set to 0.0, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0. The dispersive attraction was obtained with the coupling parameter ξ , which was varied from 0 to 1 by increments of 0.1. The electrostatic free-energy contribution was obtained using the coupling parameter λ , which was varied from 0 to 1 by increments of 0.05. FEP simulations corresponding to every value of the staging parameters were performed independently using protocols described previously [112, 113]. The solute self-free-energies were calculated in vacuo and then subtracted from the value calculated in water to remove all internal energy contributions for each solute. The vacuum free-energies were calculated (without solvent molecules) using the same coupling parameter schemes as in simulations with solvent but with a dual Langevin-Thermostat protocol described previously. The Langevin damping coefficient was set to 25.0 ps⁻¹ using a dual-thermostat formulation. [106] All FEP runs were performed for both forward and backward perturbation. The periodic boundary condition (PBC)-correction due to surface potential of SWM4-NDP water [105] was set to -12.5 kcal/mol/e (for cations) in keeping with previously reported simulations by Lamoureux et al. [106, 114] The simulation time for each of the FEP windows was run for 500 ps of equilibration, followed by 2.5 ns of forward and backward simulation time for every amino acid mimetic, both in water and in vacuum.

Weighted Histogram Analysis Method (WHAM)[115, 116] was used to compute the components of the hydration free-energy shown in Table 3.1. The coordinates were saved for every 0.5 ps, while WHAM parameters were saved at every step of the production. A long-range correction (LRC) was added to account for errors introduced by the truncation of LJ interactions. To calculate this LRC, a single MD simulation was run for every calculated value of the hydration free-energy. This MD simulation composed of a single solute molecule in a box of 250 SWM4-NDP water molecules which was run for 5000 ps in the NVT ensemble. During this simulation coordinates were saved every 1 ps for analysis, as described by Baker et al.[112]

Table 3.1. Excess free-energy values of amino acid side chains (molecular mimetics) computed with the Drude FF using a staged FEP protocol.

AA residue	Mimetic	ΔG_{disp}	ΔG_{rep}	ΔG_{elec}	LRC	ΔG_{total}	ΔG_{exp}	$\Delta G_{\text{additive}}$
LEU	iso-Butane	-11.3	13.6	0.4	-0.4	2.3	2.3	2.8
ILE	Butane	-11.7	14.5	-0.1	-0.4	2.3	2.1	2.7
VAL	Propane	-9.7	11.8	-0.4	-0.3	1.4	2	2.8
ALA	Ethane	-7.6	9.5	-0.2	-0.2	1.5	1.8	3.1
PHE	Toluene	-15.3	17.7	-2	-0.7	-0.3	-0.7	0.7
CYS	Methanethiol	-8.8	10.1	-2.2	-0.3	-1.2	-1.24	-0.2
MET	Ethylmethylsulfide	-12.8	14.9	-3	-0.5	-1.4	-1.49	0.9
THR	Ethanol	-8.6	11.3	-7.2	-0.3	-4.8	-4.9	-4.1
SER	Methanol	-7.2	8.4	-6.2	-0.2	-5.2	-5.08	-4.9
TRP	Indole	-16.6	18.4	-8.1	-0.7	-7	-5.9	-2.7
TYR	p-Cresol	-16.7	18.2	-6	-0.7	-5.2	-6.1	-4.1
ASN	Acetamide	-9.7	11.8	-11.5	-0.3	-9.7	-9.7	-7.1
GLN	Acetamide	-9.67	11.8	-11.5	-0.3	-9.67	-9.7	-7.1
HID	4-methyl imidazole	-14.1	16	-12.1	-0.5	-10.7	-10.2	-9.5
HIE	Imidazole	-12.3	13.1	-11	-0.4	-10.6	-10.2	-10.3
ARG	Butyl-guanidinium	-17.2	19.9	-45.9	-0.6	-56.3	-	-59.5
LYS	Guanidinium	-9	10.4	-56.1	-0.2	-67.4	-69.2	-64.9
GLU	Acetate	-11.9	9.7	-97.1	-0.4	-87.2	-80.6	-80.5
ASP	Acetate	-11.9	9.7	-97.1	-0.4	-87.2	-80.6	-80.5

* The air-water interface potential correction was added to all charged molecules using previously reported data (-12.45 kcal/mol per unit of charge |e|)[106]. The total change in the free-energy of amino acid residue hydration (ΔG_{total}) is the sum of the repulsive (ΔG_{rep}), dispersive (ΔG_{disp}), electrostatic (ΔG_{elec}), and a long-range correction (LRC). These values were corrected by the air-water interfacial potential for SWM4-NDP if required. 250 SWM4-NDP water molecules were used as solvent. All the free-energy values are in kcal/mol. The values for the additive FF and experimental values are from references. [69, 117, 118]

All simulation systems used to calculate the hydration free-energy contained 250 SWM4-NDP water molecules and one solute at the center of the simulation box. Positional constraints were applied on the solute's center of mass. The electrostatic interactions were calculated using Particle-Mesh Ewald (PME) summation, with a coupling parameter of 0.34, 6th order spline for mesh interpolation, and cubic periodic boundaries. Non-bonded pair lists were maintained with a cutoff of 16 Å, and a real space cutoff of 14 Å was used for LJ interactions, which were truncated via an atom-based energy switch algorithm from 10 to 12 Å with LRCs applied. All systems were simulated at 298.15°K in the NPT ensemble using a constant pressure of 1atm with a 1fs time integration time-step. The positions of Drude particles were propagated via an extended Lagrangian formalism through the assignment of a small mass (0.4amu), at a low temperature (0.1K), using a separate thermostat (VV2 scheme).

3.3.2 - Protein and Hydrogen-Bonding Simulations and Analysis

In the present work, the H-bonding patterns were evaluated between protein-protein and protein-water interactions with both the polarizable SWM4-NDP [67] and non-polarizable TIP3 [68] water models. To generate secondary-structure dependant data, four structures were simulated: 2 α -helical structures, and 2 β -hairpin structures. The α -helical structures were two 13-mers with sequences (1) SDELAKLLRLHAG, and (2) SDEEMMELREAFA. The 2 β -hairpin structures used were two 12-mer structures with sequences (3) V5PGV5 and (4) PIISEGNRNRHR. The structures (1) and (3) have previously been used to calculate the H-bond strength with the CHARMM26 FF [63]. Structures (2) and (4) were taken from the L-plastin to include more diverse side chains including a combination of charged, polar, and non-polar residues of various sizes. Hence, these systems offer additional insight into the difference between polarizable and non-polarizable FFs.

All structures were first minimized using the C36 FF and NAMD 2.11 for 1ps then equilibrated for 5ps at 300K using NAMD 2.11. These structures were subsequently solvated in a water box with an edge distance of 7.5Å from each protein (water boxes ranged from 1200 - 1600 water atoms) using CHARMM-GUI [119, 120]. These models were used as the starting point for simulations run using the C36 FF. These structures were also converted into to a Drude model using Drude prepper (in CHARMM-GUI) and used directly as starting point for simulations run with the Drude FF [119, 120].

Once the initial assemblies had been constructed and equilibrated, all structures were subjected to MD simulations using either the Drude or C36 FF. These simulations were run at 300K using NAMD 2.11. The C36 FF simulations were performed with a time-step of 2.5fs, while the Drude FF simulations were run with a timestep of 1fs (for stability).

Each of the systems reached RMSD plateau in the 50ns of MD simulations. Then each system was subjected to two additional 1ns simulations run at two temperatures (300K and 350K). The 1ns simulations were performed with structures being saved to a DCD file saved every 50 fs. The short time frame between structures allowed for better mapping and was necessary given the short and bifurcating nature of H-bonds.

3.4 - Results

3.4.1 - Hydration Free-Energy Values in the Drude Force Field Compared with the CHARMM36 Force Field

The summary of hydration free energies (Table 3.1) shows that the Drude FF provides more accurate hydration free-energy values than the C27-C36 FF. To compare the overall values between the Drude FF, the C36 FF, and experimental values, all residues except Arg (which has no experimental mimetic to compare with the values obtained by the FFs) were averaged in each FF and compared to the average experimental value (Table 3.1). Here the C26-C36 FF was shown to over estimate experimental hydration free-energy interactions by ~8.53% (or by 1.36 kcal/mol/residue on average) while the Drude FF underestimated experimental hydration free-energy interactions by 4.4% (or by -0.7 kcal/mol/residue on average).

Comparing the individual hydration free-energy values of each residue mimetic again showed the Drude FF to more accurately compute the hydration free-energy than the C27-C36 FF (Table 3.1). Of the 18 residues that can be directly compared between the Drude and C27-C36 FF, 15 show more accurate hydration free-energy values in the Drude FF compared to the C27-C36 FF. Only Residues HIE, GLU, and ASP appear to provide more accurate values in the C27-C36 FF. Another interesting observation is that every single residue mimetic in the C27-C36 FF overestimated the experimental values, while the Drude FF appears to more accurately represent a mean value by both over and under representing the experimental hydration free-energy values of various mimetics.

Two residues that need to be addressed in the Drude FF are the negatively charged residues GLU and ASP. These residues show the largest divergence from the experimental hydration free-energy values [20]. In the Drude FF the hydration free-energy values of these residues are under

estimated by 8.1%, however, due to the relatively large negative hydration free-energy value experimentally observed for these residues (-80.6 kcal/mol) this correlates to a large difference of -6.6 kcal/mol/residue. The overall parameterization for carboxylate group appears to cause problems in other applications of the Drude polarizable force-field and recent efforts to re-parameterize it are to be tested [121]. These large values likely stem from the large $\Delta G_{\text{electric}}$ value associated with these residues. These large discrepancies indicate that further work should be done to improve the hydration free-energy of these residues in the Drude FF.

3.4.2 - The Stability of the Drude Force Field vs the CHARMM36 Force Field

To ensure that the systems used to analyze H-bonds retained stable secondary structures, the RMSDs of all heavy atoms were monitored for the initial 50ns of MD run for all structures using both the C36 FF and the Drude FF. The structures taken from the final frame of simulation after this equilibration simulation were used to identify stability of the secondary structure elements by aligning these structures with the original structure and then calculating the root mean square deviation (RMSD) (Figure 3.1). Of the four systems tested, three showed a higher stability (a lower RMSD value) in the Drude FF, where only α -helical system 2 appeared to be more stable in the C36 FF. When all RMSD values were averaged the C36 FF showed only a 2.9% increase in overall RMSD. Visually, every structure except for α -helical system 1 in the C36 FF appeared to maintain its secondary structure (Figure 3.1).

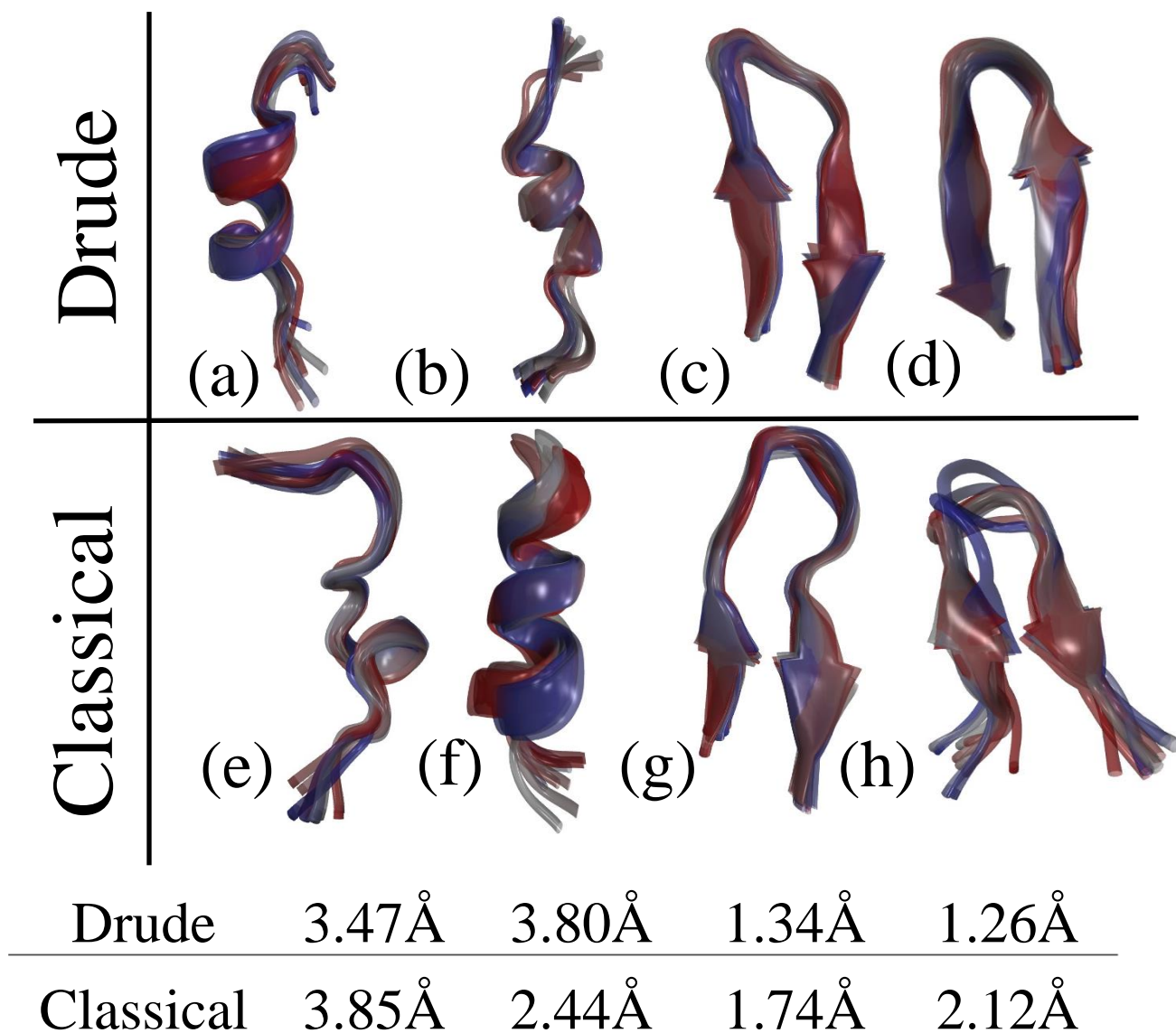


Figure 3.1. Representative structures of protein constructs from (a) α -helix system 1, (b) α -helix system 2, (c) β -sheet system 1, (d) and β -sheet system 2 run using the Drude FF; and (e) α -helix system 1, (f) α -helix system 2, (g) β -sheet system 1, (h) and β -sheet system 2 run using the C36 FF. These structures represent the 1ns of H-bond analysis taken after 50ns of equilibration. In each system a representative snapshot was taken every 0.1ns (2000 frames) and aligned. Red structures represent the beginning of the 1ns simulation, white structures represent the middle, and blue structures represent the end. The table of values represents the RMSD between the initial structure and the first structure after the 50ns equilibration.

3.4.3 – The Drude Force Field Shows Elongated Hydrogen Bond Lifetimes

The stability of secondary structures elements in small peptides depends on the intercalated hydrogen networks involving proteins-protein, protein-water, and water-water interactions [122, 123]. One of the most common metrics allowing direct comparison in model description of H-bonding patterns is hydrogen bond life-times. Accordingly, the lifetimes of all H-bonds involving protein residues and water for the α -helical and β -sheet structures were analyzed. System 2 was used to analyze both the α -helical and β -sheet secondary structures, as system 2 maintained an increased diversity of residues compared to system 1. To accurately measure the protein-water H-bond lifetimes the two previously discussed TCFs, $S_{PW}(t)$ and $C_{PW}(t)$, were identified and used to calculate the lifetimes of each H-bond (refer to section 2.1.4 – *Hydrogen Bonding Simulations and Analysis*).

In these simulations the strict lifetime of protein-water H-bonds (τ_s^{PW}) was shown to be shifted towards longer time-scales in the Drude FF than the C36 FF. To compare the τ_s^{PW} between FF's, the values of τ_s^{PW} were averaged for all residues to obtain the average τ_s^{PW} ($\langle \tau_s^{PW} \rangle$) (figure 3.2). This comparison showed the Drude FF exhibit longer $\langle \tau_s^{PW} \rangle$ values in every system, with an average increase of 49% (~0.05ps) compared with the C36 FF (figure 3.2).

The individual continuous lifetimes of protein-water H-bonds (τ_C^{PW}) were also averaged ($\langle \tau_C^{PW} \rangle$) to compare the occupancy of water in the shell around proteins (Figure 3.3). Again, this showed an increase in the $\langle \tau_C^{PW} \rangle$ of every system with the Drude FF compared to the C36 FF. An overall 8.5-fold (29ps) increase in the $\langle \tau_C^{PW} \rangle$ was observed when all residues were considered (figure 3.3). This substantial increase is mainly due to Arg 10 in the β -sheet protein (this will be discussed later). Without this residue included in the calculation of $\langle \tau_C^{PW} \rangle$ the Drude FF showed a more moderate increase of roughly 34% (roughly 1ps) compared with the C36 FF (Figure 3.3).

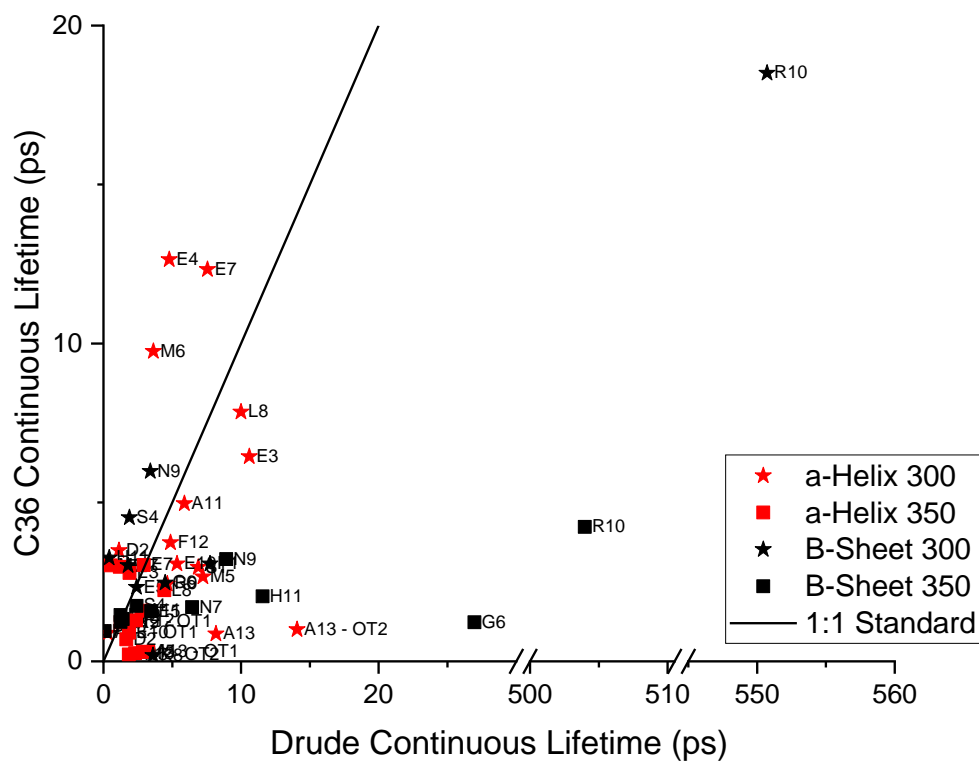


Figure 3.3 The continuous lifetime of protein-water H-bonds (τ_c^{PW}) for all protein residues. Black data points represent β -Sheet secondary structure, red data points represent α -helical secondary structures. The line represents the point where the τ_c^{PW} is the same in both the C36 and Drude FF.

The strict lifetimes of protein-protein H-bonds (τ_s^{PP}) also indicate a longer lifetime in the Drude FF. Using all identified protein-protein H-bonds, the τ_s^{PP} ($\langle \tau_s^{PP} \rangle$) showed an overall 17% (~ 0.07 ps) increase in the Drude FF compared to the C36 FF (Figure 3.4). This general analysis of $\langle \tau_s^{PP} \rangle$ does not take into consideration the specific protein-protein H-bonds formed. As the stability of H-bonds, and therefore the τ_s^{PP} , is dependant on the specific residues involved, it is critical to compare similar H-bonds between the FFs. To do this, the $\langle \tau_s^{PP} \rangle$ was also obtained using only the H-bonds which maintained the same H-bond donor and acceptor in both the C36 and Drude FF. This analysis of $\langle \tau_s^{PP} \rangle$ using only equivalent H-bonds still showed a 12% (~ 0.05 ps) increase using the Drude FF compared with the C36 FF.

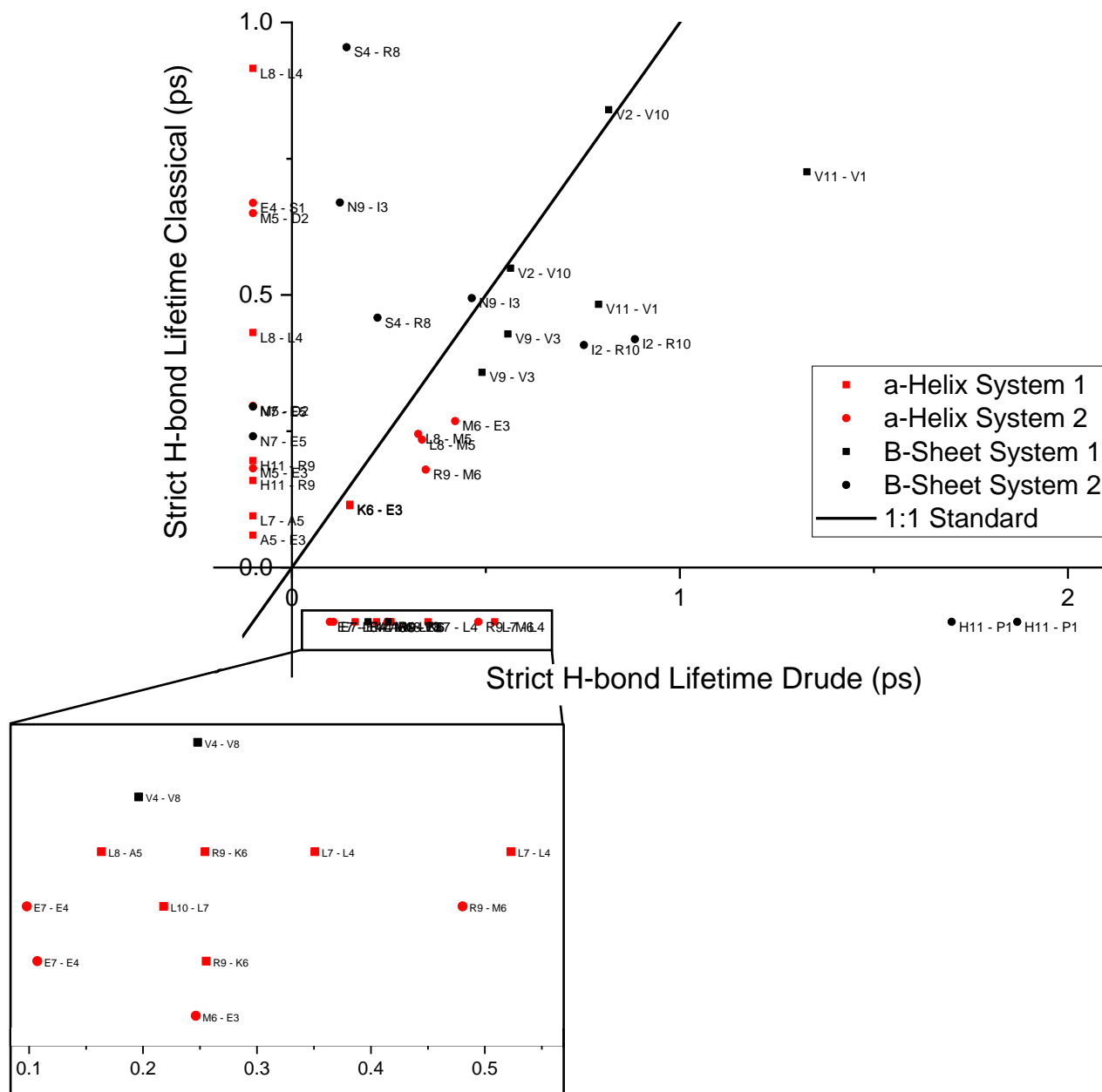


Figure 3.4 The strict lifetime of protein-protein H-bonds (τ_s^{PP}) for all analyzed protein-protein hydrogen bonds. The black line represents the point where the H-bond lifetimes are the same in both the C36 and Drude FF. All systems include data from both the simulation run at 300K and run at 350K. Across all simulations longer Lifetimes of the same bond correspond with the 300K simulation. Labels represent the H-bond being formed: the first residue is the H-bond donor, and the second residue is the H-bond acceptor. The Inset depicts all lifetimes only observed within the Drude FF with a lifetime less than 0.5ps. in this inset all y-axis values are -0.1, however, their vertical position of each points was altered to allow visibility

*Note: the same protein-protein H-bonds were not observed in all systems. To address this every H-bond which did not have an equivalent H-bond in the opposing FF had its lifetime arbitrarily set to -0.1 for the opposing FF. This describes all values with a lifetime less than 0.

3.4.4 - The Drude FF shows More Variation in the Lifetime of Hydrogen Bonds

In addition to longer lifetimes, the Drude FF also exhibits more variation in the length of H-bond lifetimes. By subtracting the shortest-lived H-bond from the longest-lived H-bond of each system a comparison of the range of τ_S^{PW} values between the Drude and C36 FFs could be obtained. The α -helix systems show a 284% (~0.44ps) increase in the variation of the τ_S^{PW} in the Drude FF compared with the C36 FF. The β -sheet show a 177% (0.2ps) increase in the variation of the τ_S^{PW} in the Drude FF compared with the C36 FF (Figure 3.2).

The τ_C^{PW} also showed an increase in the variation of H-bonds. The α -helix system showed a 11% increase in the τ_C^{PW} variation in the Drude FF compared with the C36 FF (figure 3.3). The β -sheet system showed a massive 2898% (532ps) increase in the variation of the τ_C^{PW} in the Drude FF compared with the C36 FF. Again, this is mostly due to Arg 10 (see below) causing a very large value. When Arg 10 is not considered, the variability of τ_C^{PW} of the β -sheet systems in the Drude FF still shows a 360% (~21ps) increase compared with the C36.

The variability in the τ_S^{PP} using the Drude FF again displayed higher variability than the C36 FF (figure 3.4). Comparing the τ_S^{PP} of only equivalent H-bond (as described above) the Drude FF showed 41% (~0.05ps) more variation than the C36 FF in α -helical systems, and 102% (~0.61ps) increased variation in β -sheet systems. Comparing τ_S^{PP} of all protein-protein H-bonds in the β -sheet systems the Drude FF showed 145% (~1ps) more variation than the C36 FF. The only exception to the increase in variation observed with the Drude FF is when comparing the ranges of all τ_S^{PP} values in the α -helix systems, where the C36 FF to show 101% (~0.43ps) more variation than the Drude FF.

3.5 – Discussion

3.5.1 - The Increase in Variation and Lifetime in the Drude Force Field May be Due to Polarizability

The increased variability observed in the lifetime of all analyzed H-bonds in the Drude FF may be directly linked to the FFs ability to represent polarizability. In the same way that polarizability can enhance H-bonds, polarizability also has the potential to decrease the strength of an interaction (or increase repulsive forces). The Drude FFs ability to model polarizable effects allows the relative strength of local electric fields to be evaluated by each particle. Because multiple residues are adjacent, the polarization of the Drude particles in one direction will enhance some interactions, while others are diminished. This could directly account for both increases and decreases in H-bond lifetimes and explain the increased variability in lifetime observed by H-bonds in the Drude FF.

To investigate this the induced dipole moment was calculated for the oxygen found on the backbone of every residue (H-bond acceptor). The most highly polarized residues in both systems correspond to the C-terminal residues. Interestingly, in the α -helical structures the C-terminal has to the largest τ_S^{PW} values, while C-terminal H-bonds are not observed in the β -sheet systems in the Drude FF. Although interesting, this analysis fails to consider both the direction of the generated dipoles, and the changing local electric fields observed by atoms. It is hard to quantify these values; however, it is important to note that the Drude FF is creating an environment with variable, local, polarizable effects.

The increased lifetime of H-bonds observed in the Drude FF could also be attributed to its polarizable nature. The hydration free-energy values in the C27-C36 FF were shown to be higher in the C27-C36 FF than the Drude FF, and yet the Drude FF still shows longer H-bond lifetimes.

The idea that polarizability strengthens interactions is not new, instantaneous polarization is well known to be responsible for VdW's interactions which helps strengthen interactions. This indicates that polarization could be strengthening the observed interactions. [124]

3.5.2 – Explicitly Accounting for Polarizability may Result in More Dynamic Structures

Mackerell et al identified secondary structures in the Drude FF to be more dynamic, and capable of forming and denaturing faster than in classical FFs [125], This may be directly related to this FF's ability to vary H-bond strength. As observed here, the Drude FF shows the ability to express both longer lived and more variable H-bond lifetimes than H-bonds in the C36 FF. This means that when a residue polarizes favourably, it could enhance the attractive interaction between residues enough to allow for the formation of a H-bond, potentially helping lead to the more rapid formation of secondary structures. Additionally, this would allow for H-bonds to break in a quicker time frame. If the polarizability is shifted in a non-favourable manner, this could destabilize H-bonds and propagate a more rapid denaturation of secondary structure. Variability therefore has the potential to increase the speed of protein secondary structure formation and denaturation. Given that this increased rate of secondary structure dynamics does not come at the cost of stability (as demonstrated in Figure 3.1), This increased rate of protein secondary structure dynamics may help mitigate the ~4x cost associated with the Drude FF compared with classical FFs.

3.5.3 - Polarizability Can Account for Arg 10, with Multiple Hydrogen Bonds

In the β -sheet structures, Arg 10 showed a τ_C^{PW} in the Drude FF almost 2 magnitudes larger than any other residue (Figure 3.3). Interestingly, in the Drude FF the τ_S^{PW} is also increased for Arg 10 by up to 214% in the 350K simulation compared to the C36 FF. This observed increase of the τ_S^{PW} in simulations using the Drude FF may stem from the polarizable effects this FF presents;

and when compounded by the ability to Arg to form multiple H-bonds, this may lead to the highly increased τ_C^{PW} value observed in simulations with the Drude FF. In simulations with the Drude FF Arg 10 is observed forming multiple protein-water H-bonds with the same water molecule. By forming more than one H-bond with a single water molecule synergy occurs. When one H-bonds breaks, rather than the water being able to diffuse away, the water molecule can be held in the area by the secondary H-bond. By keeping the water localized, the broken H-bond can reform. In this way, in order for the water to diffuse away from this residue it must break all H-bonds almost simultaneously. Examples of long-lasting τ_C^{PW} values have previously been observed by Bandyopadhyay using the CHARMM 22 all-atom FF, and was additionally associated with an Arg residue [64]. Using a classical FF, they observed the τ_C^{PW} of water surrounding Arg to last for 300ps or more. While still very large, this puts the τ_C^{PW} value of 551ps received with the Drude FF into perspective, as it shows an increase of 83% in the Drude FF. This is not surprising as it follows the reoccurring trend of longer lived lifetimes in the Drude FF.

Experimentally, water in the hydration shell of protein has been shown to have slow solvation dynamics that range from 100-1000ps compared to the bulk water. Battacharrya has shown the globular protein glutaminyl-tRNA synthetase to exhibit water dynamics as slow as 580ps, while still expressing a faster water dynamic component [65, 126]. This is another example showing that protein-water interactions can be maintained for extended periods in certain circumstances. This ultraslow component is thought to be directly related to a disruption of the H-bonding network, arising from water molecules ability to induce polarization in adjacent water molecules [126-128]. This ability of water to polarize can increase the dipole moment of water molecules from 1.85D to 2.6D in the solvation shell and lead to stronger H-bonding interactions [126-128]. This shows that the polarizability of both proteins and SWM4-NDP water molecules

within the Drude FF may be helping vary and extend the H-bond lifetimes observed in the Drude FF.

3.6 - Conclusions

Here the Drude FF is shown to more accurately display the hydration free-energy than the C36 FF when compared with experimental values. In addition, the Drude FF was shown to maintain H-bond interactions longer, and displays more variability in H-bond lifetime, while still maintaining structural integrity of the secondary structure elements of protein. This may be directly responsible for more dynamic structures within the Drude FF and allow the Drude FF a more rapid ability to simulate structural formation and denaturation. Given these trends this FF may be able to converge towards biologically relevant structures marginally faster, which may help mitigate the ~4x computational costs associated with using this FF, while providing the polarizability required to address many systems.

Additionally, this FF displays the ability to replicate the previous observed phenomenon where water can be caught interacting within the solvation shell of a protein. These long lived water interactions are occasionally seen experimentally, and FFs should be able to characterize long-lived water-protein interaction. As with the other interaction in the Drude FF the lifetime of this interaction appears to also be extended compared with classical FFs. This suggests that the Drude FF is enhancing the polarizable qualities of both water and protein.

Chapter 4 - Atomic Level Characterization of Calcium-Induced Conformational Switching **in Proteins**

4.1 Abstract

Many Ca^{2+} -binding proteins are of vital importance in numerous biological functions. They can act as biological switches underlay by conformational changes upon Ca^{2+} -binding that can trigger a cascade of cellular responses. However, the mechanisms that control these conformational switches, in tandem with Ca^{2+} -binding processes, are not well understood. To gain insights into such mechanisms Molecular Dynamics (MD) simulations, Markov-State modelling, and experimental methods were combined to map the thermodynamics and kinetics of Ca^{2+} induced conformational switching in a prototype-protein of L-plastin's regulatory EF-hand domain. In addition to being, arguably, the simplest Ca^{2+} -dependent switch the Ca^{2+} sensitivity of L-plastin's EF-hand domain affects actin bundling in hematopoietic cells, and it has recently been shown to have a direct link to T-cell motility and cancer metastasis. While a holo (Ca^{2+} -bound) state of L-plastin's EF-hand region has been completely solved, with a high-resolution NMR structure; the apo state displayed an unstructured protein tail which remains mechanically undetermined, in respects to its relative position and conformation in regard to the Ca^{2+} -binding EF-hand motifs. The so-called H5-tail was previously proposed to have some effects on the binding kinetics of Ca^{2+} and actin-bundling. In this report, the H5-tail is shown to intimately modulate the Ca^{2+} -binding by having direct, intermittent, contact with the EF-hand motifs. To unravel the conformational dynamics of L-plastin, both state-of-the-art MD simulations, with 120- μs of data, and biological studies using NMR and microcalorimetry were combined. Using these methods, it was possible to identify several distinguishable conformations of L-plastin revealing a complex transition pathway between the apo and holo states. As a proof-of-principles, MSM-driven structural insights were used to fine tune L-plastin Ca^{2+} -binding sensitivity via a few key positively

charged residues each resulting in a distinguishable ion-dependent switching mechanism. This paves the way to rational designs aiming at the altered motility and effectiveness of T-cell, and more generally Ca^{2+} -dependent switching.

4.2 - Introduction

The intimate coupling between protein conformational dynamics and selective solute/ion binding is a central piece of cellular machinery involved in electrical signaling, muscle contraction, volume regulation, and cell motility [2, 3, 129]. One of the most common examples of ion-control switching can be found in Ca^{2+} -binding proteins. Ca^{2+} -binding proteins are central to cellular function as these proteins are responsible for cascades of vital functions such as muscle contraction, replications, neuronal communications, and cell motility [2-4]. Among Ca^{2+} -binding proteins calmodulin (CaM) and L-plastin have highly dynamical structures capable of shifting between ion-bound (holo) and ion-free (apo) conformations due to submicromolar variations in the concentration of different cytosolic environments [8, 14, 130]. Their dynamical structures somehow enable precise control over interactions with other proteins, triggering various tasks [8, 14, 130] such as the mediations of gene transcription and DNA synthesis [130, 131], and are further linked with cardiac arrhythmias and neurological diseases [16, 17].

In humans, there are three highly conserved homologous isoforms of plastin (Table 4.1) specifically localized in various cellular types [47, 49]: I-plastin, L-plastin, and T-plastin. These plastins each contain two actin-binding domains (ABDs) attached to a single amino-terminal Ca^{2+} -binding domain, joined by a linker region. Each Ca^{2+} -binding domain expresses as a pair of EF-hand motifs, homologous to one lobe in CaM, displaying a high degree of α -helical structure. This Ca^{2+} -binding domain of plastins acts as a deactivation switch with unique, and specific, affinities for Ca^{2+} among different isoforms enabling precisely regulated actin-bundling mechanics.

Noticeably, the expression of L-plastin normally only occurs in hematopoietic cells, making L-plastin partially responsible for actin-bundling in the innate immune system [48, 49]. L-plastin is also correlated with the development and efficacy of T-cell and B-cell leukocytes [49,

51, 52], and allows leukocytes to rapidly rearrange their actin cytoskeleton in response to extra cellular stimuli. Specifically, neutrophils without the presence of L-plastin display reduced mobility and an inability to kill bacterial pathogens [48, 49]. Interestingly this protein has also been linked to the ability of many cancers to metastasize [48, 49].

Table 4.1. Sequences of both EF-hand motifs and the H5-tail in the three isoforms of human plastin with the net charge of each sequence (q). Note that in protein constructs the H5-tail was terminated by a C-terminus, while physiologically a conserved glutamic acid is found next in the sequence of all human plastins. Both terminal options maintain a charge of $-|e|$.

Isoform	1st EF-hand Motif	2nd EF-hand Motif	H5-Tail
L-plastin	22-DTDGNGYISFNE-33	62-DLDQDGRISFDE-73	85-DVAKTFRKAINKK-97
(q)	(-3 e)	(-4 e)	(+3 e)
T-plastin	25-DLNSNGFICDYE-36	65-DRNKDGKISFDE-76	87-DIAKTFRKAINRK-99
(q)	(-3 e)	(- e)	(+3 e)
I-plastin	24-DIDNSGYVSDYE-35	64-DSNKDGKISFEE-75	86-DISKTFRKIINKR-98
(q)	(-4 e)	(-2 e)	(+3 e)

Even though the Ca^{2+} -binding EF-hand motifs of L-plastin and CaM are homologous, their conformational dynamics and thermodynamics may be completely different. Their EF-hand motifs are indeed geometrically well-preserved, often described as a helix-loop-helix structure [130, 132]. Each pair of their EF-hand motifs helps stabilize the hydrophobic core of the incoming and outgoing α -helices [132]. It has been hypothesised for CaM that once Ca^{2+} binds to the EF-hand motifs, CaM exposes its hydrophobic core to enlarge the interaction surface for binding with target proteins [130, 132]. However, it has been shown that in the holo state of L-plastin the hydrophobic core of L-plastin is closed by its tail helix (dubbed as the H5-tail) that appears to significantly limit the exposure of the hydrophobic core (Figure 4.1 (a)) [50]. When reducing Ca^{2+} concentrations, L-plastin reversibly returns to its apo state with the H5-tail leaving the hydrophobic core, which re-enables the actin-bundling function of L-plastin. As a result, the presence or absence of Ca^{2+} helps reduce or enhance the ability of L-plastin to bundle actin, respectively, in a sharp contrast to the many roles of CaM. Up till now, the entire reversible transition from the apo to holo states of L-plastin has remained unknown, even though the structures of the holo state was resolved and structure of the apo state of L-plastin was partially resolved. In the apo state the H5-tail could not be resolved as it was unstructured and delocalized with respect to the hydrophobic core.

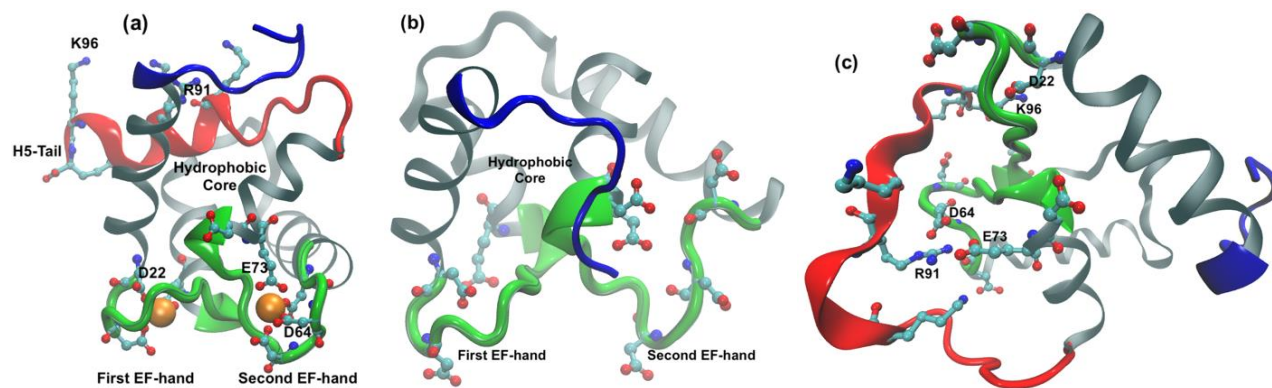


Figure 4.1. NMR structures of L-plastins in (a) holo (Ca²⁺-bound) and (b) apo states. The orange spheres are Ca²⁺ ions bound in the holo structure. The NMR data showed the structure of the H5-tail to be unstructured, therefore this region is missing in (b), but was artificially added in an arbitrary location in MD simulations. (c) A snapshot from MD simulations starting from the apo state structure shows the direct contacts of K96 and R91 with the EF-hand motifs. The H5-tail is highlighted in red while the other end is highlighted in blue.

To solve for the transition pathway between the apo and holo states of L-plastin, extensive molecular dynamics (MD) simulations were performed together with experiments to probe structural changes via NMR data and measure binding affinities via isothermal titration calorimetry (ITC) due to mutations. The 120- μ s MD simulations reveal that there are intermittently direct interactions between positively charged residues of the end tails, particularly H5-tail, and the negatively charged residues of the EF-hand motifs. The proposed mechanism of the apo-to-holo transition involves the self-regulation of two Ca^{2+} -binding events by means of a few positively charged residues of a proximal linker interacting with the EF-hand motifs. ITC data of mutations to the positively charged residues of the H5-tail show that there are two thermodynamically distinguishable binding events with a broad range of Ca^{2+} sensitivities in the various L-plastin constructs. So, it might be possible to fine tune the sensitivity of Ca^{2+} -binding in L-plastin as a way of controlling the motility and effectiveness of actin-bundling in T-cell.

4.3 - Methods

4.3.1 - Molecular Dynamics (MD) Simulations.

The NMR structures of apo and holo states were used [50] to build all-atom systems. Since the apo state does not have a reasonably high-resolution structure of its H5-tail, the α -helical H5-tail of the holo state structure was arbitrarily attached onto the apo state structure so that the tail oriented away from the rest of L-plastin. These atomic structures were first solvated with water, neutralized with CaCl_2 (2 Ca^{2+} and 1 Cl^-), minimized, and then equilibrated using CHARMM-GUI [119] and a classical CHARMM36 (C36) force field (FF) [133, 134]. Note that both the apo and holo systems maintained the same number of atoms (34315) and the same dimensions (69^3 \AA^3). These fully solvated systems were used for running alchemical free-energy simulations, replica exchange simulations, and briefly testing the classical FFs as well. Alchemical free-energy calculations were performed using NAMD [108] for F90A mutation to gain some insights into the roles of F90 for stabilizing the hydrophobic core of the L-Plastin. Essentially, a dual residue was constructed that has both side-chains of an original residue and mutated residue; then the interactions from this dual residue with the rest of the system were gradually changed to mimic an alchemical process. 18 intermediate states were used plus the two end-states, which were used for both backward and forward sampling. At each intermediate step 0.4ns was sampled with a time-step of 2fs. The total simulation times for the alchemical computation using 32 different dynamic holo structures was 0.512 μs .

To speed up sampling of some immediate states and potentially identify some probable transition pathways between the apo and holo states, tempering replica exchange MD (T-REMD) simulations were performed [135, 136]. The T-REMD simulations were started using the apo state structure with the added H5-tail. Since the relative position of the H5-tail was added manually, the

T-REMD simulations were run with 64 replicas for temperatures distributed between 275 to 400 K for 0.3 μ s/replica (or a total of 19 μ s) to thoroughly perturb the entire system. This perturbation helped to broadly generate a number of relevant conformations, which may not require Ca²⁺-binding but are sufficiently distant from the initial apo-like structure. The exchange rates among these replicas are between 19% and 31%, which were considered to be adequate for T-REMD sampling [135]. Since all the replicas seem to well diffuse in the temperature space, the final coordinates at the 0.3 μ s were taken from the 10th to 50th replicas to run 41 simulations in parallel ($T = 300$ K and $P = 1$ atm) and collected roughly 120 μ s of conformational data. One structure every 1 ns in these parallel simulations was saved for analysis.

To analyze this large number of conformations and find the most representative structures to describe a transition between the apo and holo state, a Markov State Model (MSM) was used [137-143] via PyEMMA package [144]. To do this, the distances between α -Carbon atoms of all residues in L-plastin except the four neighboring residues was chosen as a feature. Time independent correlation analysis (tICA) [141] was then used to rank the dynamics of L-plastin from the slowest to the fastest. The slowest motions were often considered as representative motions of the entire proteins. As a result, representative data could be extracted on the projected landscape of the first two slowest components in the tICA space via k-mean algorithm. 100 clusters were used to digitize this representative data, which was then used to construct a Markovian transition matrix (M) and build coarse-grain macrostates. This M matrix was also used to populate only 10,000 corresponding structures (instead of using 120 μ s data) for computing other observables via VMD software [145]. Note that this simulation strategy yields an effective implied timescale of 5-20ms for the first two tICA components estimated by the MSM model. At this point, however, even 120 μ s of conformational data appears to be incomplete in obtaining a convergent

set of coarse-grain macrostates (e.g., via Robust Perron Cluster Analysis (PCCA+) [146, 147]). One reason could be that too many distant apo state structures were sampled broadly, generated by T-REMD, having little connection with the apo-holo transition triggered by the binding of Ca^{2+} , which was not observed. To circumvent this, a hidden MSM was used [148] to group the 100 digitized clusters based on the M matrix into just three coarse-grain macrostates, which are reproducible if using different amounts of the data. Using these three macrostates, it was possible to identify all visually distinguishable conformations, and propose a transition pathway between the apo and holo states. The transition times were estimated using transition path theory (TPT), [149] which solves for a stationary flux between any states from the M matrix. The complete transition pathways (between any apo-like state and the holo state) will be reported separately with more thorough considerations of the FF and the polarizable and charge-transfer effects due to the binding of Ca^{2+} in the EF-hand motifs.

Finally, since Ca^{2+} -protein interactions modeled by the C36 FF have been shown to be overestimated by 20% compared with quantum calculations, particularly for highly negative charged binding pockets [20, 58], the EF-hand of L-plastin was also run using the Drude polarizable (Drude) FF, which has been shown to be more accurate than the C36 FF. Essentially, the Drude FF adds an auxiliary superlight particles to every atom (except hydrogen) in the C36 FF to mimic polarizable effects [58, 62, 84, 108, 150, 151]. To implement the Drude model, the non-polarizable fully solvated systems were converted via Drude prepper on CHARMM-GUI [119] and then subjected to a second round of minimization and equilibration. Every system was run for approximately 25 ns, giving a total of approximately 150 ns for both the holo systems and the apo systems, resulting in 300 ns of total simulation. To perturb the systems, annealing simulations were performed at six different temperatures ranging from 300 K to 550 K for every 50 K using both

the holo and apo states for about 30 ns/temperature. Note that the simulations using the Drude FF are about four times slower than using the C36 FF, thus more expensive to potentially capture some transitions from the apo to holo states. Since the Drude FF remains in its infancy, and not as efficient as C36, after obtaining many initial similar conformations of the apo state within 300ns in both FFs, the majority of simulation efforts were spent using the C36 FF to generate 120 μ s of diverse conformational changes of L-plastin in its apo state.

4.3.2 – Experiments

A bacterial vector was expressed with the first 97 residues of L-plastin, containing the N-terminal EF-hand domain and 12 residues of the linker region. This vector was used to directly produce the wild-type protein construct (EF-WT) (Figure 4.1(a)/(b)). This protein construct has an N-terminal His-tag and a TEV protease cleavage site. Using this EF-WT bacterial vector as a template, single alanine mutations were introduced at residues F90 and R91, generating the EF-F90A, and EF-R91A protein constructs respectively (Table 2). Additionally, a double alanine mutation was introduced at K96 and K97, generating the EF-K96A-K97A protein construct (Table 2). These mutations were introduced using Quick Change site-directed mutagenesis kits (Stratagene). The PCR products of these vectors were transformed into competent *E. coli* DH5 α for plasmid expression. The resultant plasmids were sequenced to confirm the accuracy of the mutations. These isolated plasmids were subsequently transformed in competent *E. coli* BL-21 (DE3) cells for protein expression. *E. coli* cells were grown in Luria Bertani medium with 100 μ g/ml of ampicillin at 37 °C. ¹⁵N-labelled proteins were prepared in M9 minimal medium supplemented with 0.5 g/L ¹⁵NH₄Cl. At an optical density of 0.5-0.7 (600 nm), the cultures were induced with 0.5–1.0 mM IPTG. After 4 hours, the bacterial cells were harvested by centrifugation. The cell pellet was re-suspended in the IMAC binding buffer (20 mM Tris-Cl, 0.1M NaCl, and 50

mM imidazole, pH 8.0) and lysed via French Press. The supernatant was applied onto an IMAC column (GE Healthcare). The column was washed extensively with the IMAC binding buffer and the His-tagged proteins were eluted with the elution buffer (20mM Tris-Cl, 0.1M NaCl, and 300mM imidazole, pH 8.0). The His-tag was then cleaved with TEV protease in the digestion buffer (20mM Tris-Cl, 0.1M NaCl, 0.5mM EDTA, and 1mM dithiothreitol) at 34 °C. The mixture was then loaded onto the complete column (GE Healthcare) to remove the TEV protease and the His-tag from the protein. Protein purification was verified using SDS-PAGE and Coomassie brilliant blue staining.

Table 4.2. Ranked sensitivity of L-plastin protein constructs in terms of K_{d1} computed by models using two sets of independent binding sites (2SIBS) and one set of identical binding sites (1SIBS); N is the stoichiometry of Ca^{2+} in a L-plastin construct for each site. Change of enthalpy, ΔH , is in kcal/mol; change of entropy, $\Delta S_{1\setminus}$ is in kcal/mol/deg. The relative sensitivity (RS) is computed by the ratio of the smallest K_d of one set to the smallest K_d of another set. Uncertainties of ΔH , ΔS , N_1 , estimated from different sets of data, are about 0.1, 0.002, and 0.4, respectively.

L-plastin Constructs	Fitting Model	K_{d1} (μM)	K_{d2} (μM)	$\Delta H_1/\Delta H_2$	$\Delta S_1/\Delta S_2$	N_1/N_2	RS	Mech.
EF-K96A-K97A	2SIBS	0.015 \pm 0.001	2.4 \pm 0.1	-5.4/-1.9	0.017/0.019	1.6/1.3	20	MC1
EF-WT	2SIBS	0.04 \pm 0.01	2.8 \pm 0.1	-4.6/-1.7	0.018/0.019	1.3/1.0	7.5	MC1
EF-H5-Removed	2SIBS	2.6 \pm 0.1	0.3 \pm 0.1	-4.2/-1.3	0.010/0.024	1.3/1.0	1	MC4
EF-R91A	1SIBS	0.7 \pm 0.1	N/A	-4.5	0.013	1.8	1/2	MC2
EF-F90A	2SIBS	5 \pm 0.1	5 \pm 0.1	-271/280	-0.9/1.0	1.1/1.0	1/17	MC3

To quantify the sensitivity of Ca^{2+} -binding in the L-plastin constructs (Table 4.2), isothermal titration calorimetry (ITC) experiments were performed using 40 μM of protein at 25°C with an initial injection of 2 μL of 1.2 mM CaCl_2 followed by 57 injections of 5 μL of 1.2 mM CaCl_2 for all tested proteins. Injections occurred over 10 seconds, and a 300-second time delay occurred between each injection to allow the system to equilibrate. All measurements were performed using a Microcal VP-ITC microcalorimeter (Figure 4.2). In all protein trials, Ca^{2+} was removed by adding 1mM ethylenediaminetetraacetic acid (EDTA) to the protein sample and then using an Amicon 3-kD Ultra Centrifugal Filter to buffer exchange the protein to a minimum of a 1,000,000x dilution. The buffer used for this dilution contained 20 mM HEPES and 100 mM KCl (pH 7.2) and was subjected to chelex[®] for over 3 days to remove Ca^{2+} . This buffer also acted as the solution in the buffer cell for all trials and was additionally used as the injectant after 1.2 mM of Ca^{2+} was re-applied. The apo state for both the EF-WT and mutants were confirmed to be Ca^{2+} -free using NMR (see below). From this Ca^{2+} -free state, all titration continued until all sites had been saturated with Ca^{2+} . All data was fit using either a one or a two-site model of Ca^{2+} -binding with the MicroCal Original software to obtain the stoichiometry (N), affinity (K), free enthalpy (ΔH), and entropy (ΔS) values (Table 4.2).

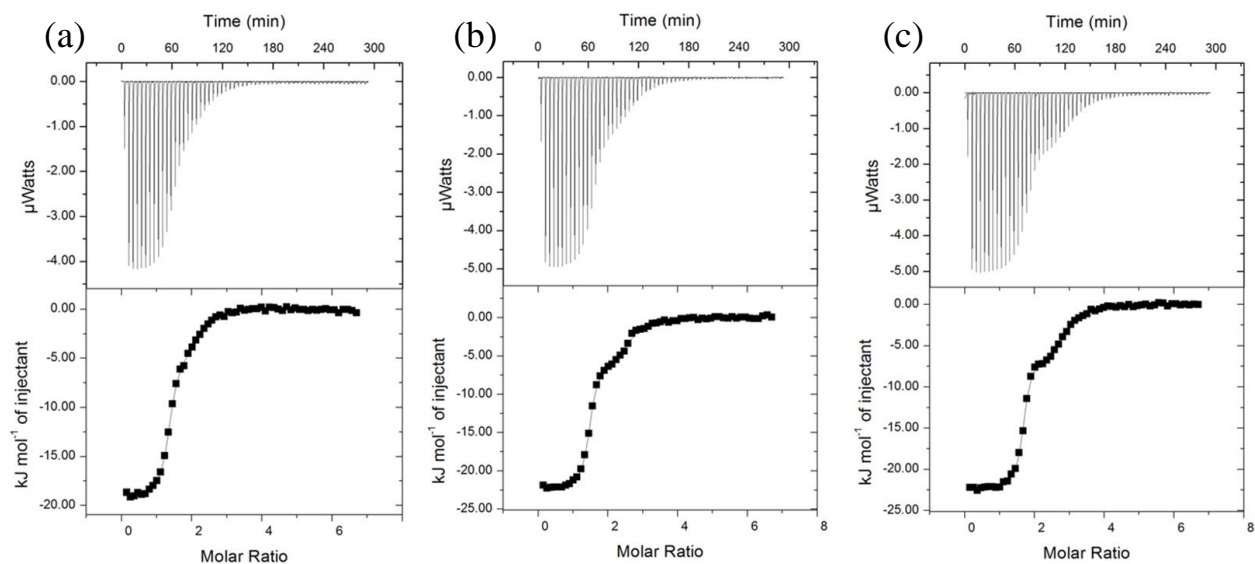


Figure 4.2. Isothermal titration calorimetry data of the (a) EF-WT protein construct, (b) the EF-K96A protein construct, and (c) the EF-K96A-K97A protein construct with Ca^{2+} . The baseline corrected ITC titrations are shown in the top panel. Derived binding isotherms were fitted with a 2-site model to obtain thermodynamic parameters (bottom panel).

Table 4. 3 Differential scanning calorimetry (DSC) values obtained for each of the protein constructs. Each value is the average of 3 separate runs. SD values are the \pm standard deviation

Protein Construct	T_m (°C) \pm SD	T_{onset} (°C) \pm SD	ΔH_{melt} \pm SD
EF-WT	102.6 \pm 0.30	83.1 \pm 3.02	68.3 \pm 1.59
EF-F90A	98.3 \pm 0.35	75.7 \pm 2.20	51.1 \pm 1.10
EF-R91A	102.6 \pm 0.30	83.7 \pm 3.18	67.0 \pm 2.16
EF-K96A-K97A	102.6 \pm 0.26	85.2 \pm 4.24	61.1 \pm 4.29

To confirm the stability of the protein constructs, DSC was performed on all protein constructs in their holo states. DSC experiments were performed using 40 μ M samples of Ca²⁺-free protein with 400 μ M of Ca²⁺ re-applied to ensure each protein construct was in its holo form. Ca²⁺-free protein samples were obtained the same way as in ITC (see above). The buffer used was the same as in ITC (see above). All DSC experiments were run using a MicroCal PEAQ-DSC instrument. Each sample was subjected to a temperature range of 20 to 130°C with a temperature increase of 1°C/minute. In all protein trials, a buffer-buffer baseline was performed previous to each run to use as a baseline. All data was fit to a two-state model with MicroCal Original software to obtain the protein melting temperature (T_m), the temperature of onset of protein melting (T_{onset}), and the enthalpy of the melting process (ΔH_{melt}) (Table 4.2).

To help confirm some states of the L-plastin constructs obtained from MD simulations, nuclear magnetic resonance (NMR) experiments were also performed under high (~ 500 μ M) and low (40 μ M) concentration conditions. These experiments were carried out at 25° degree Celsius on a Bruker Avance 700 MHz spectrometer. All experiments were performed in a solution of 10% D₂O/90% H₂O containing 90mM KCl, 18mM Hepes, 10 μ M DTT and 500 μ M DSS. 2-dimensional ¹H¹⁵N-HSQC experiments were performed for all ¹⁵N-Labeled protein.

4.4 - Results

4.4.1 - Structural States of L-plastin

While the entire holo state structure of L-plastin obtained from NMR can be ascertained from NMR data, there is an uncertainty about the positions and conformations of the H5-tail with respect to the hydrophobic core in the apo state structure (Figures 4.1(a-b)). If excluding the H5-tail, the root mean square deviation between the two NMR structures is only 5.5 Å, which is equivalent to the deviation of the dynamic structures simulated from the initial holo state (Figure 4.3). In the holo structure, Ishida et al observed that V86, F90, and I94 residues of this H5-tail act as anchors imbedding into the hydrophobic core. These residues apparently strengthen the hydrophobicity of the core in the holo structure with respect to the apo structure. To test a change of the hydrophobicity the free-energy change due to an alchemical transformation of F90A was estimated, $\Delta G = G_A - G_F$. The average free-energy change (see Methodology) using simulated holo structures in water medium is $\Delta G = -47 \pm 1$ kcal/mol, indicating a large hydrophobicity that F90 introduces to the core compared to its mutant.

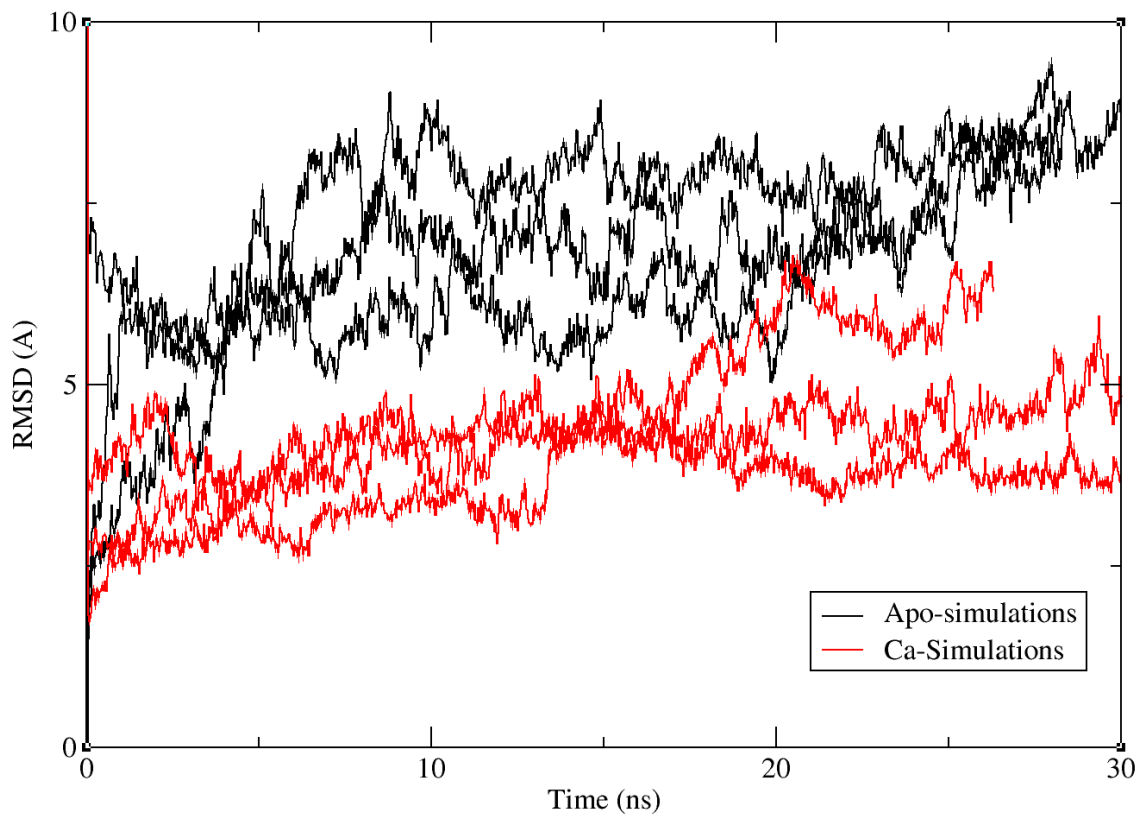


Figure 4.3. The change in RMSD of L-plastins EF-hand domain including the H5-tail during annealing simulations at temperatures below 450K for both the holo simulations (in Red) and the apo-simulations (in Black).

To sample apo state structures, the unstructured H5-tail was added onto to the NMR apo state structure (as an α -helix) and used to run long MD simulations (see Methodology). Interestingly, in the simulations of the apo state structures, in both non-polarizable and polarizable FFs, the H5-tail was observed to directly interact with the EF-hand motifs via its positively charged residues R91, K92, K96 and K97 (Figure 4.1 (c)). To estimate the probabilities of those sidechains being localized in each of the EF-hand motifs, representative data from the MSM was extracted, applied to 120 μ s of the conformational MD data, and then tracked the distances between the charged centers of those sidechains to the center-of-mass of each backbone EF-hand motif. (Figures 4.4 (a-b)) show that the positively charged sidechains frequently localize at least 12-25 Å away from the first EF-hand motif and 6.5-15 Å away from the second EF-hand motif in the apo state with the highest frequencies. Particularly, R91 and K96 can localize as close as at 6.5 and 3.5Å to the center of the second EF-hand motif, respectively. Figure 4.4(c) shows that in the apo state the sidechain of R91 has a large probability of forming hydrogen bonds with the sidechains of E73 and D64 (20-25%), which constitute the binding site for Ca^{2+} in the holo state as seen in Figure 4.1(a).

By overlapping the HSQC NMR spectrums observed for both the EF-WT and EF-K96A-K97A protein constructs in there apo and holo conformations it is possible to observe differences in the structures between these two constructs. The overlap of the holo HSQC NMR data suggests the holo structures of the EF-WT and EF-K96A-K97A constructs be similar. This is supported by the fact that the HSQC NMR data of these two constructs shows all residues distal from the H5-tail mutations to overlap in the spectrum (Figure 4.5 (b)). Although many residues are shifted within this spectrum, this can be attributed to the mutation of the two charged residues (K96 and K97) localized to hydrophobic core. Structural similarity between these two constructs in the holo

state is further supported by the observation that the T_m obtained by DSC are the same in both the EF-WT and EF-K96A-K97A protein constructs (Table 4.3). This suggests that the stability of these two constructs in the holo state is essentially the same.

Overlapping the apo HSQC NMR data of the EF-WT and EF-K96A-K97A constructs shows significantly more overlap between the two apo structures than observed in the holo structures, indicating that these two protein constructs express very similar structures (Figure 4.5 (a)). Although very similar, a few of the hypothesised key residues express peak shifts, which directly indicates that regions of the proteins EF-hand motifs are affected by terminal mutations to the H5-tail. The largest shifts observed in the apo states were identified at the second EF-hand motif, specifically at residues D64 and S70 (Figure 4.5(a)). This clear experimental evidence supports the observation from MD simulations that the H5-tail directly interacts with the EF-hand motifs in the apo state. As a result, the H5-tail should be capable of interfering with the binding of Ca^{2+} (see below).

Overlapping the apo state HSQC NMR data of the EF-WT and the EF-R91A protein constructs showed comparable results to what was observed when overlapping the EF-WT and EF-K96A-K97A constructs. Again, there was shown to be a large degree of structural overlap indicating highly similar structures (Figure 4.5 (c)). Although these structures were very similar, a few key shifts were observed in the spectra, notably within the EF-hand motifs, which directly indicates the H5-tail (and specifically R91) to be able to interact with the EF-hand motifs.

DSC values of the EF-R91A in its holo state also showed this construct to have the same T_m as the EF-WT construct (Table 4.3). This again indicates that the EF-R91A protein construct possesses the same stability as the EF-WT construct in its holo state, which points to a similar final structure in both constructs.

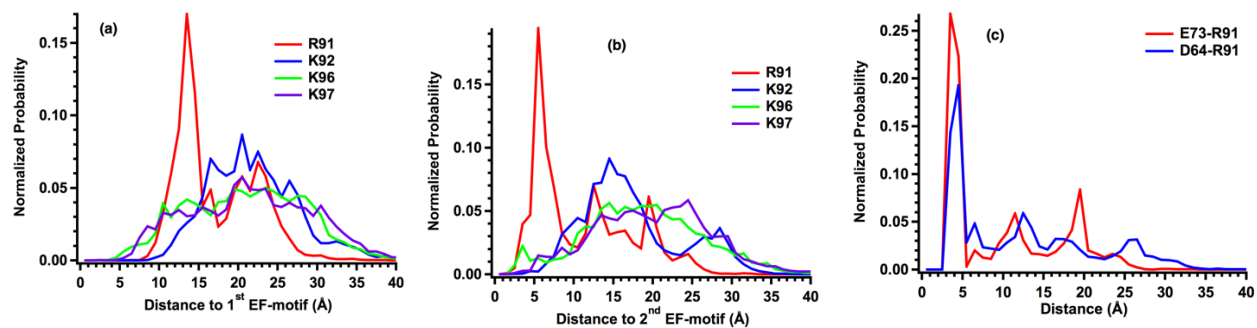


Figure 4.4. Probability of the positively charged sidechains being localized with respect to the EF-hand motifs. (a-b) The distance is computed for a charged center (namely, NZ for Lysine and CZ for Arginine) of a sidechain to the center-of-mass of a motif backbone. (c) Distance measured between the CD (or CG) atom of E73 (or D64) sidechain and the NZ atom of R91.

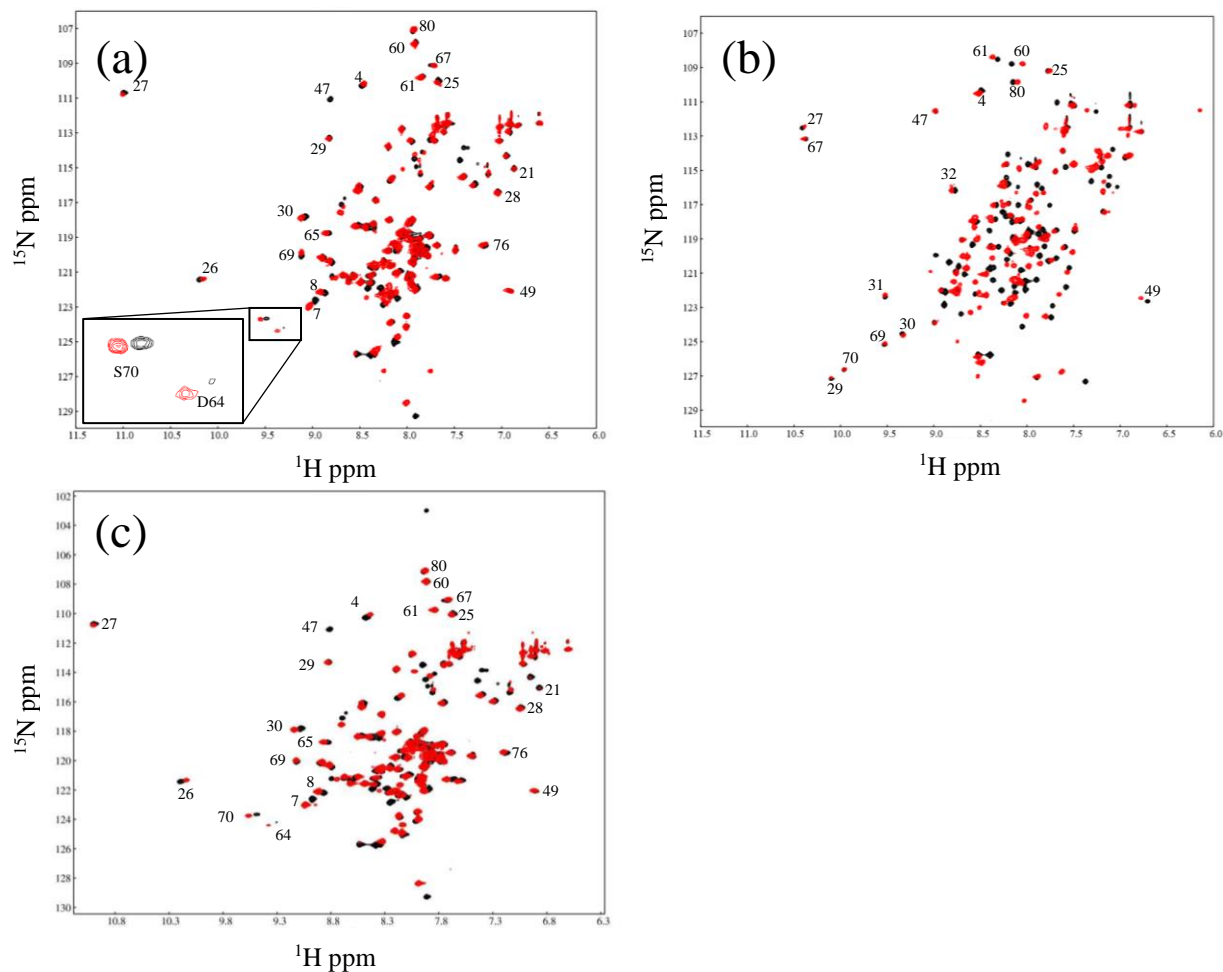


Figure 4.5. (a-b) An overlay of the EF-WT (in black) and EF-K96A-K97A mutant (in red) NMR data for the (a) apo and (b) holo structures. (c) An overlay of the EF-WT (in black) and EF-R90A mutant (in red) NMR data for the apo state. Numbers represent the residue number for overlapping (or similar) peak assignments. The inset in (b) magnifies the largest shifts in the apo structure.

A number of interesting intermediate structures of L-plastin were found in the 120- μ s data. Using a hidden MSM on this data set (see Methodology), these structures could be grouped into three coarse-grain macrostates. The smallest transition timescale was 7-11 μ s, while the largest was about 61 μ s. There is, however, no experimental measurements to compare with the computed transition timescales. Each macrostate expressed several visually distinguishable conformations (Figure 4.6). This usually indicates that the sampling data is not yet complete for describing the entire conformational space, but fortunately this data appears to sample the apo-holo transition pathway quite well. Despite not having the complete conformation space, this simulation strategy seemed to generate diverse apo-like and holo-like structures, which were populated and propagated from the T-REMD simulations without needing to sample Ca^{2+} -binding events. This set of the structures provided many insights into the conformational changes of L-plastin and Ca^{2+} -binding switch mechanisms. Figure 4.6 (a) shows how the H5-tail can interact with the EF-motifs differently. For example, the H5-tail can deform greatly ('overarching') to reach the first EF-hand motif, fold to poke the EF-hand motifs from the below, cross the two EF-hand motifs, lean on the second EF-hand motif while opening the hydrophobic core, or simply has the tip floating in the bulk water. This flexibility of the H5-tail implies that it can direct interfere with the binding of Ca^{2+} to the EF-hand motifs. As well, the diverse apo-like structures of L-plastin appear to have different exposures of the hydrophobic core in tandem with the 'opening' of the EF-hand motifs. For instance, the structures with the 'EF-hand beneath' and 'EF-hand leaning' appear to have more hydrophobic exposure than other mutants. Such structures may help L-plastin to recruit actin in T-cells.

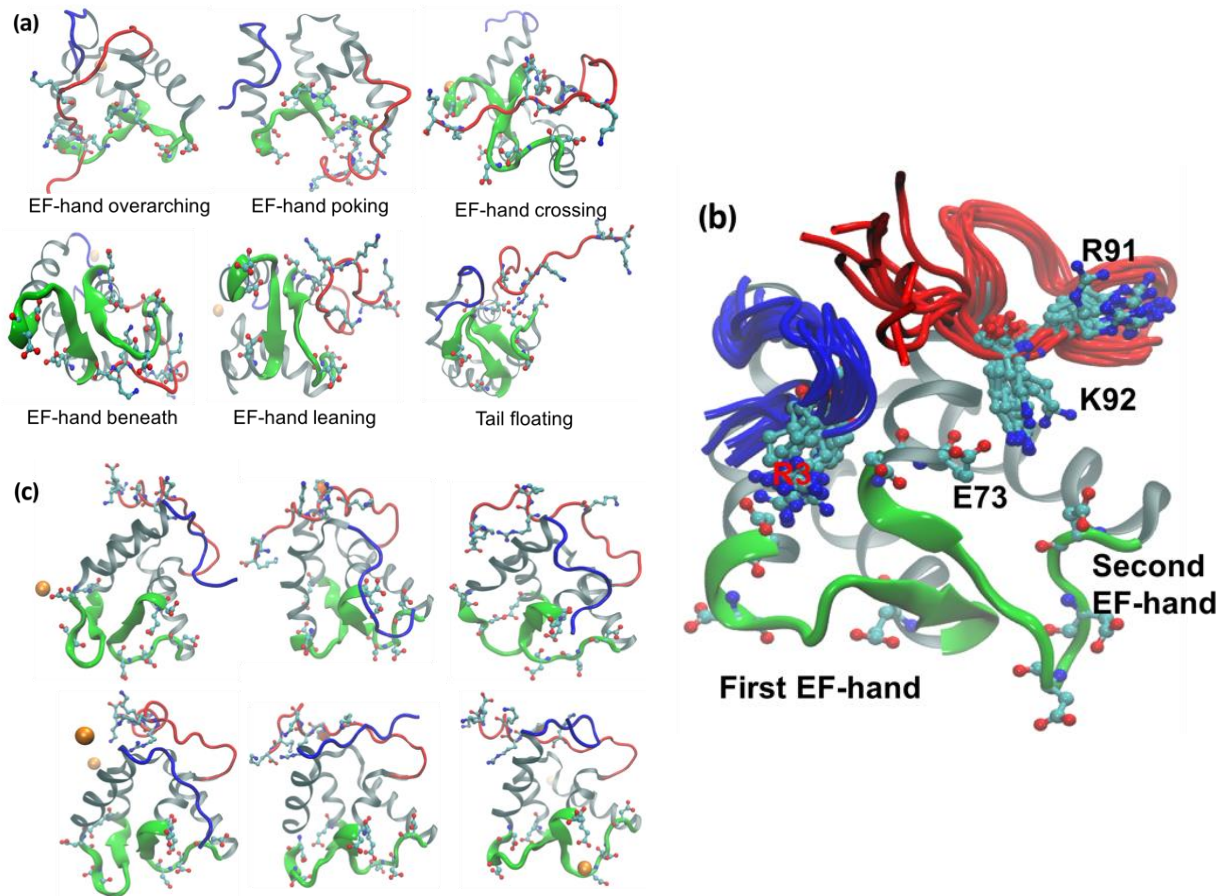


Figure 4.6. Snapshots of (a) sampled apo-like structures, (b) likely transition states, and (c) holo-like structures. The residues of the two end tails with direct interactions are G4S5 and I94N95.

Figure 4.6 (b) shows a likely transition state between the apo and holo state. In this transition state, the two end tails may interact directly with each other at residues G4S5 and I94N95, while the sidechains of R3 and K92 interact directly with the first and second EF-hand motifs, respectively. To transform from the apo state to the holo state, the H5-tail must swing up while the N-terminal tail has to move down toward the second EF-hand motif to shift the sidechains of R3 and K92 away from the Ca²⁺-binding sites. Only in these relative swinging motions of the two tails, would the H5-tail move to a position to close the hydrophobic core. Note that in these simulations, even though the RMSDs computed for the holo-like structures (Figure 4.6 (c)) with respect to the holo structure determined by NMR are about 5Å apart, this is similar to the deviations observed by the simulated dynamics of the holo state (Figure 4.3). During these simulations the binding of Ca²⁺ to the EF-hand motifs was not observed, this may be the reason why the F90 residue did not imbed in the hydrophobic core. In the reverse transition from the holo to apo state, one may expect the same transition state, at which the interactions of R3 and K92 with the two EF-hand motifs would knock Ca²⁺ out of the binding sites when the Ca²⁺ concentration is lowered.

4.4.2 – Calcium-Binding Affinities and Mechanisms in WT and Mutant Forms of L-plastin

To provide experimental evidence showing how the critical, positively charged, residues interfere with Ca²⁺-binding, the energy release during the Ca²⁺ titration was compared experimentally via ITC experiments of the EF-WT, EF-R91A, EF-F90A, EF-K96A-K97A, and the EF-WT construct with the deletion of the H5-tail (EF-H5-Removed) (Table 4.2 and Figure 4.7). A fitting model of two sets of independent binding sites (2SIBS) was used to determine equilibrium dissociation constants for the EF-WT, EF-K96A-K97A, EF-H5-Removed, and EF-F90A constructs as this model fit the data the best. Interestingly, the MD data provided deep insights into how Ca²⁺ and the L-plastin constructs undergo different interactions, even in the

same 2SIBS model. For the EF-R91A protein construct, a fitting model of a single set of identical binding sites (1SIBS) was used to obtain the best fit. Using these models and the MD data it was then possible to show different binding mechanisms among the protein constructs, which depend on how the H5-tail interacts with the EF-hand motifs.

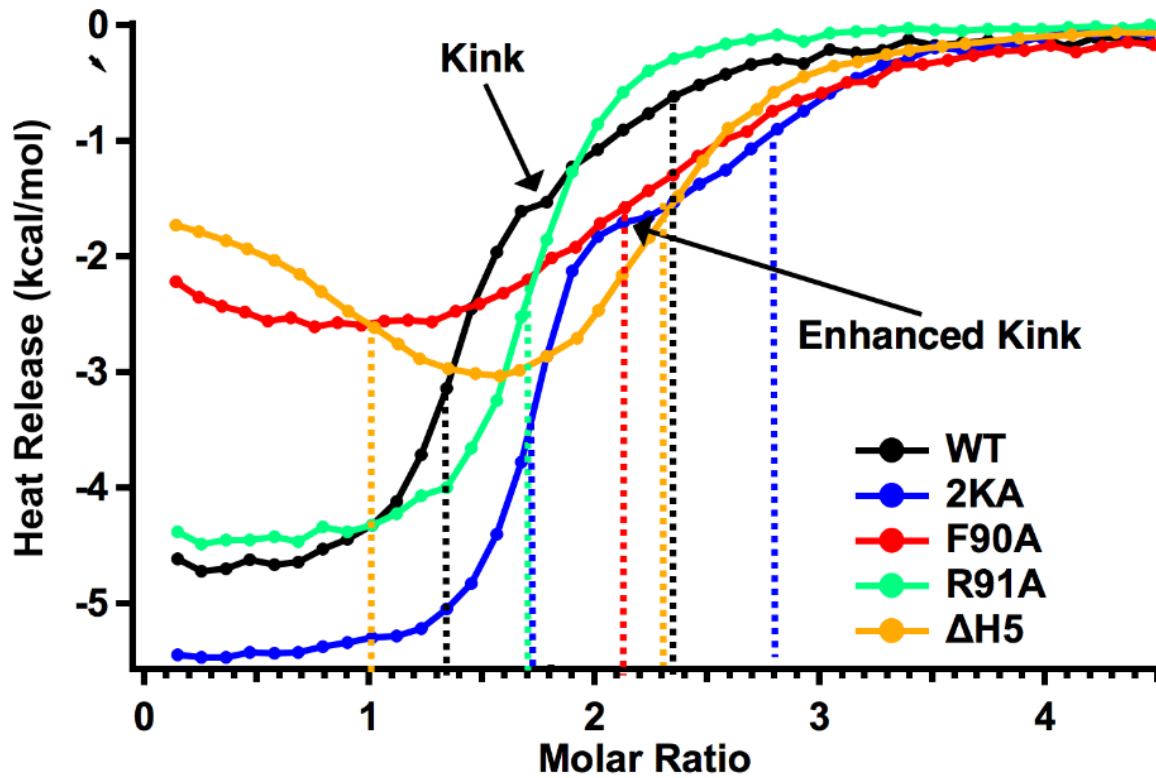


Figure 4.7. Isothermal titration calorimetry of L-plastin constructs. Any two dashed lines with the same color indicate N_1 and $N_1 + N_2$, which are given in Table 4.1. Here WT represents the EF-WT construct, 2KA represents the EF-K96A-K97A construct, F90A represents the EF-F90A construct, R91A represents the EF-R91A construct, and $\Delta H5$ represents the EF-H5-Removed protein construct.

The energy-release profiles of the EF-WT and EF-K96A-K97A constructs were observed to have kink patterns, suggesting two thermodynamically distinguishable binding events. The enhanced kink in the ITC profile of the EF-K96A-K97A mutant might suggest that the Ca^{2+} -binding triggers the conformations of both systems by a clearer two-step mechanism than in the EF-WT. This mechanism is regarded as MC1 for these constructs. Compared to the EF-WT, the EF-K96A-K97A mutant enhances the amount of energy released in the first binding event (ΔH_1) by roughly -0.8 kcal/mol and enhances the amount of energy released in the second binding event (ΔH_2) by roughly -0.2 kcal/mol. This indicates that there is less attraction between Ca^{2+} and the EF-WT than between Ca^{2+} and the EF-K96A-K97A protein construct. In other words, the EF-K96A-K97A mutant may reduce a barrier for facilitating Ca^{2+} -binding. Indeed, Table 4.2 shows that both $K_{d1}(\text{EF-K96A-K97A})$ and $K_{d2}(\text{EF-K96A-K97A})$ are smaller than $K_{d1}(\text{EF-WT})$ and $K_{d2}(\text{EF-WT})$, indicating the EF-K96A-K97A mutant is more sensitive to Ca^{2+} than the EF-WT. Particularly, $K_{d1}(\text{EF-K96A-K97A})$ is about 2.6 times smaller than $K_{d1}(\text{EF-WT})$. To explain this, MD data (Figure 4.3 (b)) shows that K96 and K97 of the EF-WT can locate within 5 Å close to the 2nd EF-motif for about 8-10% of the simulation time. Moreover, this interaction is directly observed by HSQC NMR data (Figure 4.5(a)). This suggests that the mutations of K96 and K97 would likely reduce the ability (by 8-10%) of the H5-tail to interact with both EF-hand motifs. Since the positively charged K96 and K97 residues are absent in the EF-K96A-K97A mutant, Ca^{2+} would have more access to the motifs by perhaps 8-10%. As a result, the stoichiometric numbers of the two independent binding sites are increased by about 0.3 in the EF-K96A-K97A mutant. So, in mechanism MC1, there are two independent binding sites in the EF-WT, which can interact with the K96 and K97 to interfere with the Ca^{2+} -binding. Here K_{d1} , N_1 and K_{d2} , N_2 are argued to be assigned to the 1st and 2nd EF-hand motifs, respectively.

For the EF-R91A mutant, 1SIBS is sufficient to fit the data, which looks like many CaM systems [152, 153]. This model returns the same $K_d(\text{EF-R91A}) = 0.7 \mu\text{M}$ for the two identical binding sites with a Ca^{2+} occupancy of approximately 1.8 (or 0.9 per site). It is interesting to see that the EF-R91A mutation transforms the EF-WT with two *independent* binding sites into approximately one sets of two *identical* binding sites (1SIBS) which bind Ca^{2+} simultaneously. To explain this transformation, Figure 4.4(c) shows that the R91 sidechain interacts the most with the 2nd EF-hand motif, particularly the negatively charged residues D64 and E73 for about 35-50% of time having distances $\leq 5.0 \text{ \AA}$. Again, this interaction is directly observed by HSQC NMR data (Figure 4.5 (c)). This interaction would help to distinguish the different binding processes in the EF-WT: Ca^{2+} must require some energy to knock the R91 sidechain out of the 2nd EF-hand motif, while having a direct access to the 1st EF-hand motif. As a result, Ca^{2+} would have a higher affinity to binding in the 1st EF-hand motif than to the 2nd EF-hand motif in the EF-WT. Because of this, K_{d1} , N_1 and K_{d2} , N_2 must correspond to the 1st and 2nd EF-hand motifs, respectively. When removing the sidechain by the EF-R91A mutation, it is likely that the H5-tail would remain interacting with the 2nd EF-motif by K96 (Figure 4.4 (b)) for about 10% of time. This interaction in the EF-R91A mutant would not critically hinder the binding of Ca^{2+} in the 2nd EF-hand motif, but effectively equate the net charges of the two motifs perceived by Ca^{2+} . As a result, the binding affinities of the EF-hand motifs in the R91A mutant may appear identical to Ca^{2+} , thus having the same K_d . This binding mechanism is clearly different from MC1, so it was named MC2.

Both MC1 and MC2 assume that the H5-tail of the EF-WT, EF-K96A-K97A and EF-R91A mutants fold to close the hydrophobic core flanked by the helices of the motifs, ending up in the same holo state. The combination of NMR and DSC data obtained here supports this

assumption (Table 4.2 and Table 4.3). As a result, the transition state for these constructs might look like the one shown in Figure 4.6 (b). To understand better how the binding of Ca^{2+} would trigger such a folding, the EF-F90A construct offers some insights. To fit the ITC data for the EF-F90A mutant, a 2SIBS must be used, which returns two separate binding steps, but both with the same large $K_d = 5 \mu\text{M}$. The first binding step requires about 1 Ca^{2+} /protein to saturate, and slightly deepen, the energy release (Figure 4.7). The R91, K96, and K97 of the EF-F90A mutant should still strongly interact with the EF-hand motifs, but now the mutated H5-tail may never fold into the hydrophobic core, because the F90A mutation, as estimated by the alchemical free-energy calculation, would remove -47 kcal/mol from the interaction of F90 with the hydrophobic. This is supported by the positive enthalpy $\Delta H_2 = 280 \text{ kcal/mol}$, which suggests the repulsive interaction upon adding more Ca^{2+} into the system. The difficulty of folding might keep the mutated H5-tail interacting with the two motifs even if there is a lot of Ca^{2+} near the end of the ITC curve ($N > 2.1 \text{ Ca}^{2+}$ /protein), thus reducing the binding affinities significantly as observed in the ITC experiments (Table 4.2). As a result, the holo state of the EF-F90A may be very different from the other L-plastin constructs. This is further supported by the fact that the DSC experiment of this protein construct showed a lower T_m than all other protein constructs (Table 4.3), indicating that the holo state of this protein construct is less stable than the others, potentially due to the H5-tails inability to fold into, and stabilize, the hydrophobic pocket. This mechanism is classified as MC3.

If the above arguments hold, it should be possible to explain the ITC curve of the EF-H5-Removed construct. Here, Ca^{2+} would feel the different net charges of the two EF-hand motifs. Consequently, the 2nd EF-hand motif would now have a stronger binding affinity than the first (as seen in Table 2). The addition of Ca^{2+} into this system would appear to help stabilize the EF-hand

motifs and the hydrophobic core, so the energy release would deepen at $N = 1.5 \text{ Ca}^{2+}/\text{protein}$. This mechanism is regarded as MC4. This EF-H5-Removed construct was chosen as the baseline to rank the sensitivity of the L-plastin constructs and mutants in terms of their smallest K_d (Table 4.2) because it has zero interaction between the H5-tail and the EF-hand motifs. By increasing or decreasing amount of the interactions, the binding mechanisms change, thus the Ca^{2+} sensitivity changes.

4.5 Discussion and Conclusions

In the entire simulations starting from the NMR apo state structure, no substantial binding of Ca^{2+} into the binding EF-hand motifs was observed. The binding of Ca^{2+} modeled by non-polarizable FF is known to be less accurate than polarizable FFs [20, 58]. Using a non-polarizable FF, however, allowed simulations to reach 120 μs roughly four times faster than the Drude polarizable FF. Note that the apo-holo transition of L-plastin triggered by Ca^{2+} -binding is part of the complete conformational space, which can be generated by classical FF if any polarizable and charge-transfer effects are not yet concerned. In any event, it is necessary to sample a diverse set of conformations even with the non-polarizable classical FF for future development purposes, which aim to effectively sample the effects of Ca^{2+} -binding. To broadly generate such a conformational space, T-REMD simulations were used to perturb the initial apo state structure. This approach allows sampling the conformations of L-plastin without necessarily requiring that Ca^{2+} must be bound to enable certain conformation states. This is because large kinetic energies given by high temperatures in the simulation approach can push the system out of local potential-energy minima to many other distant local minima, some of which may be similar to those that are facilitated by Ca^{2+} -binding events. Thus, to some extent this simulation approach helps to bypass the limitations of the classical FFs to gain some useful insights into the conformational dynamics and kinetics of L-plastin.

This report proposes a self-regulatory Ca^{2+} -binding mechanism of L-plastin via its end tails, particularly the H5-tail, and makes use of some experimental data to justify the outcomes of such a mechanism. These results clearly suggest the H5-tail is interacting with the second EF-hand motif via key positively charged residues, particularly R91 and K96 in the apo state. Among these residues, R91 forms stable hydrogen bonds with E73 and D64 to stabilize the second EF-hand

motif. The simulations also suggest that at the transition state the R3 sidechain near the N-terminus and the K92 sidechain of the H5-tail form hydrogen bonds with the negatively charged sidechains of the first and second EF-hand motifs, respectively. To transform into the holo state, these hydrogen bonds must be broken, perhaps, by the binding of Ca^{2+} ; so, the two end tails must pass by each other for the F90 residue of the H5-tail to properly insert into the hydrophobic core. Reversibly, when the concentrations of Ca^{2+} are depleted, the EF-hand motifs become less stable, and can no longer stabilize the hydrophobic core. Consequently, the H5-tail can move to the transition state (Figure 4.6 (c)) and have either, or both, the sidechains of R3 and K92 help kick Ca^{2+} out of the EF-hand motifs, thus returning the structure to its apo state. This proposed transition pathway of L-plastin is believed to be the first detailed molecular-based mechanism in the entire class of plastins and actin bundling proteins [154].

It is worth emphasizing that these MD simulations reveal S5 to be involved in the transition state of L-plastin, as there are interesting mutations, and chemical modifications, targeting S5. For instance, the S5A mutant becomes inactive, while the S5G mutant remains as active as the WT [155]. This can be explained by looking at the transition state of L-plastin, which has interactions between G4S5 and I94N95. Intuitively, the mutation S5A would likely change the interactions more than S5G because one Glycine residue is already part of the interactions. A critical chemical modification to S5 is phosphorylation, which is essential for the adhesion, migration, and remodeling of actin [156-159]. These simulations offer new valuable information of which residues S5 is interacting with, thus suggesting other residues (e.g., I94 and N95) that could be mutated to regulate the functions of phosphorylated L-plastin.

The ITC and MD results presented here show that the mutations to the H5-tail can create different Ca^{2+} -binding mechanisms. Based on how the H5-tail interacts with the EF-hand motifs,

it is possible to identify up to four Ca^{2+} -binding mechanisms (Table 4.2). In MC1, Ca^{2+} can easily bind to the 1st EF-hand motif, which has the strongest binding affinity, but must compete with the R91 and/or K96-K97 sidechains for binding into the 2nd EF-hand motif. In MC2, when the H5-tail does not have the R91 sidechain, the K96 sidechain may effectively equate the net charges of the two Ca^{2+} -binding motifs, allowing Ca^{2+} to bind the two EF-hand motifs equally. In MC3, when the H5-tail does not have the F90 sidechain, the H5-tail cannot fold properly in the hydrophobic core, the R91 and K96-K97 sidechains may keep interfering with the Ca^{2+} -binding, which appears to be un-able to trigger the transition into a holo state. In the last mechanism, MC4, when the H5-tail is deleted, Ca^{2+} may selectively bind to the EF-hand motifs with a larger net charge over the EF-hand motif with the smaller net charge and require a larger Ca^{2+} concentration to stabilize the holo state than observed in the EF-WT construct. In other words, the relative strengths of the binding affinities in the EF-H5-Removed construct are switched compared to the other L-plastin constructs.

These mechanisms help properly rank the Ca^{2+} sensitivities of L-plastin constructs. This ranking may imply the different effects of the mutations on the function of L-plastin in bundling actin. For example, at Ca^{2+} concentration of about 0.04 μM , WT L-plastin would likely transform into the holo state (inactive), thus significantly decreasing the ability of L-plastin to bundle actin. Alternatively, when an F90A mutation is inserted in the full L-plastin protein, MC3 would suggest that it cannot transform into a holo-like state capable of reducing actin bundling, and therefore this mutant would lose its ability to act as a Ca^{2+} switch. This further implies that it is possible to fine tune the Ca^{2+} sensitivity of L-plastin in a novel way. Table 4.2 shows an almost 400-fold change in the Ca^{2+} sensitivity among protein constructs. Since the Ca^{2+} sensitivity values directly regulate the level of actin-bundling in T-cell, further studies on the effects of the function and migration

speed of T-cells would be critically important to devise the best strategy to modify the effectiveness of T-cell.

Chapter 5 - Conclusions and Future Directions

5.1 – Conclusions

5.1.1 - Conclusions About the Drude Force Field

Initial simulations performed here using the Drude FF provided the necessary benchmarking for both the H-bonding and the solvation free-energy compared to classical additive FFs. This benchmarking was strongly needed to validate this FF and will help further the development of polarizable FFs in general. Once quantifiable benchmarks are in place this FF may receive more widespread use as its strengths and weaknesses are known.

Additionally, the use of this FF to simulate L-plastin represents one of the first cases of the Drude FF being used in a biologically relevant and revealing study. This again pointed towards the importance of polarizability in future biological models. The use of the Drude FF and its apparent success signifies that this FF, and the computational costs associated with modeling polarizability, are now feasible for protein systems (at least small protein system) and provides an avenue to address polarizability for additional studies.

5.1.2 - Conclusions About L-plastin

This research into L-plastin has directly lead to the discovery of a novel mechanism which allows L-plastin the capacity to self-regulate its Ca²⁺-binding, using an unstructured region of protein. This understanding of how L-plastin undergoes conformational switching is a crucial step towards understanding the full conformational dynamics of L-plastin. Further, the understanding of this novel mechanism, where unstructured protein is capable of modulating Ca²⁺-binding, provides a frame work for identifying the ability of other proteins to regulate ion-binding, and may help lead to similar regulatory mechanisms being observed in other ion-binding proteins.

Ca²⁺-binding proteins represent a large family of proteins with many important biological applications. A further understanding of the novel mechanism regulating the Ca²⁺-binding of L-plastin may have large implications within the whole family of proteins; especially, as it may indicate that previously unstudied regions of unstructured protein may have large implications on the function and regulation of proteins. In addition, this may represent large interest by the pharmaceutical industry due to L-plastin's ability to regulate cell motility, and thus potentially target metastasis and auto immune disorders.

5.2 - Future Directions

5.2.1 – Docking Simulations

The anti-psychotic drug Trifluoperazine (TFP) has been identified which is capable of binding directly to the hydrophobic binding region of the EF-hand domain of L-plastin through a weak interaction [50]. This interaction was further characterized using an NMR ^1H , ^{15}N HSQC spectrum to identify the residues which interact between TFP and L-plastin. This drug interaction directly competes with the H5-tail for binding into the hydrophobic binding region of L-plastins EF-hand domain and may represent a way to disrupt the inactivation of this protein by preventing the H5-tail from binding. As L-plastins function is directly linked to its ability to act as a switch protein, by preventing the proteins inactivation it may be possible to regulate the proteins function. By preventing inactivation this would push L-plastin to act more like T-plastin and may inhibit the motility of cells expressing L-plastin. The interaction between TFP and the EF-hands hydrophobic core is slightly stronger than the interaction between the H5-tail and the hydrophobic core, although it is not strong enough to work on the micromolar level that is needed for a drug. Hence further development would be needed to develop a more potent drug. Additionally, the selectivity of this compound compared to all three homologous of plastin needs to be characterized.

Using TFP as a starting compound Autodock Vina was used to dock TFP to the EF-hand domain L-plastin in its Ca^{2+} - bound state [160]. As this compound is shown to interact with residues found in the hydrophobic core, the H5-tail was removed from the Ca^{2+} bound state of L-plastin, leaving the hydrophobic core open to create a surface for compounds to dock with. Running docking simulations using the whole EF-hand protein surface was possible in this case given the small size of this protein domain. This generated an array of docked conformations of TFP with the EF-hand domain of L-plastin which were used to validate the docking method, by

confirming that the most stably docked structures aligned with the NMR data generated by Ishida et al [50, 160].

As TFP has been shown to bind to the hydrophobic binding region of L-plastins EF-hand this docking provided a benchmark to compare small molecules to, with the idea that when a database of molecules is screened in silico in the future, only the molecules that bind stronger than TFP are considered for further testing. Additionally, this docking confirmed that the TFP binds to the hydrophobic binding region suggesting that the docking software was working correctly as it identified the correct configuration of TFP binding with L-plastin's EF-hand domain (or at least one conformation that fit the NMR data). Now that a benchmark has been set for how strongly a compound needs to bind the hydrophobic binding region of L-plastin to disrupt binding of the H5-tail, a database of small compounds will be run through this same docking procedure. This will hopefully generate a small list of compounds from this database which can bind to this hydrophobic binding region more selectively and stronger than TFP. This refined list of compounds can then be experimentally tested to see if the computational docking was accurate. This is a first step in the process of designing a drug, potentially generating new drug leads for treating metastatic cancer, or auto-immune diseases, by regulating L-plastin's actin-bundling activity, and potentially cell mobility. L-plastin's role in cancer metastasis makes the ability to regulate this proteins activation a highly valuable drug target. This is especially important as >90% of human cancer deaths are due to metastasis [161, 162].

5.2.2 – Replica Exchange with Solute Tempering (REST) Simulations

Replica exchange with solute tempering (REST) is an algorithm developed by Lui and colleagues [73] which is a modification of classical replica exchange (REM) [163]. In REM, multiple simulations are run in parallel at different temperatures [163]. The algorithm allows

structural exchanges between each simulation (run at a different specific temperature) and its nearest neighbours as per the odd \leftrightarrow even rule, i.e., $1 \leftrightarrow 2, 3 \leftrightarrow 4, \dots (N-1) \leftrightarrow N$; with subsequent attempts for exchange as per the even \leftrightarrow odd rule, i.e., $2 \leftrightarrow 3, 4 \leftrightarrow 5, \dots (N-2) \leftrightarrow (N-1)$, etc. These exchanges occur based on a weighted probability function, which ensures that each system receives an appropriate amount of time in each temperature simulation.

The variability in temperature allows for an increased rate of conformational change to occur. The higher temperatures provide an opportunity for the protein to rapidly overcome large energy barriers. Alternatively, the lower temperature simulations allow unfavourable states, generated by the elevated temperature simulations, to recover to a more stable configuration. Thus, the low temperature simulations effectively sample energy wells separated by large barriers; while the high temperature simulations can effectively cross the high energy barriers separating wells. By shifting frames of simulations run in parallel between temperatures in this manner the entire system can maintain structural integrity, while still rapidly overcoming energy barriers, leading to accelerated conformational dynamics.

Although this method is effective, it is limited in that it can only be used on relatively small systems as the number of replicas needed to adequately sample a system is directly proportional to the system's degrees of freedom. Given that increasing the systems size leads to an increased number of degrees of freedom, solvating even medium size proteins can require a massive number of replicas for adequate sampling with this method. To address this Lui and colleagues proposed a modified version of REM known as replica exchange with solute tempering (REST) [73, 164].

REST differs from classical REM in that the potential energy of the entire system is divided into three distinct terms: the internal energy of the protein (E_P), the interaction energy between the

protein and water (E_{PW}), and the interaction between water molecules (E_{WW}) (eq. 5.1). This potential energy function is used directly to define the lowest temperature replica, E_0 .

$$E_0(X) = E_P(X) + E_{PW}(X) + E_{WW}(X) \quad (\text{eq. 5.1})$$

Although the energy of the lowest temperature system remains unchanged, REST makes use of the Boltzmann factor to rescale the energy potentials, E_m , of all replicas above the initial temperature. The Boltzmann weighting factor can be defined in terms of temperature, T (eq. 5.2).

$$\beta = \frac{1}{k_B T} \quad (\text{eq. 5.2})$$

The REST algorithm uses the Boltzmann factor directly to scale the E_m (eq. 5.3).

$$E_m(X) = E_P(X) + \left[\frac{\beta_0 + \beta_m}{2\beta_m} \right] E_{PW}(X) + \left[\frac{\beta_0}{\beta_m} \right] E_{WW}(X) \quad (\text{eq. 5.3})$$

β_0 is the Boltzmann factor calculated using the lowest simulated temperature, and β_m is calculated using the increasing temperature of each parallel system running, based on eq 5.2. Through this scaled interaction when the lowest temperature is calculated ($\beta_0 = \beta_m$) the systems original energy surface is recovered. Alternatively, when $\beta_0 < \beta_m$ a flatter energy surface is obtained allowing the system to pass over energy barriers more rapidly. Essentially this uses a new type of thermosetting where the energy landscape is altered based on temperature, while the system is kept in an isothermal bath.

The advantage of using REST over classical REM is that the E_{ww} term is removed algebraically from the weighted probability function dictating the exchanges between replicas [73]. This term is normally responsible for the poor scaling observed by traditional REM due to the relatively large number of water molecules needed to solvate systems [73, 164]. Therefore,

REST still provides the same sampling as classical REM while requiring less replicas to achieve adequate sampling.

Any method used to generate a significant array of intermediate states is computationally expensive, but the use of REST provides some computational reduction compared to REM. The modified Hamiltonian used in REST reduces the computational costs associated with running replica exchange 3 to 10-fold, depending on the amount of water used to solvate the protein [73]. Additionally, the REST algorithm provides improved sampling of a system while still maintaining structural integrity of the protein. This method is expected to generate enough intermediate states to produce a robust model of the conformational switching of L-plastin if run extensively.

5.2.3 – Developing L-plastin Simulations into a Markov State Model (MSM)

Microsecond timescale simulations of EF-WT, EF-K96A-K97A, and R91A-K96A-K97A protein constructs have been generated using a modified version of replica exchange known as replica exchange with solute tempering (REST). Although these simulations currently remain unanalyzed significant structural changes are seen from the apo state. This structural change indicates that the conformational switching process, which activate L-plastin, may be occurring in these simulations. These simulations will be analyzed to generate a Markov state model (MSM) allowing the mechanisms associated with L-plastin switching to be further studied. These simulations have already been identified to contain several sets of unique structures, which appear to represent a structural transition. Such simulations will hopefully give more detailed insight into the activation process of L-plastin.

A MSM represents a stochastic model of the structural conformational process between the holo and apo states [165]. This MSM is created by first identifying similar structural conformations

within each pool of intermediate structures generated from the REST simulations. Defining unique states is a non-trivial problem, difficult mainly due to the number of degrees of freedom that are present in the switching process, and is often done using the proteins RMSD [165]. Once the pools of intermediate states have been defined into a finite number of unique states, it is possible to look for similar structures between the apo and holo systems [165]. Because the REST simulations generate an array of intermediate states which stem from both apo and holo L-plastin if there are similar structural conformations in both pools of structures it might be possible to connect the two states in a defined pathway [165]. The completed MSM will identify all similar structures between the apo and holo states generated from the REST simulations, and create a network of intermediate states connecting these two conformations [165].

From the generated MSM substantial amounts of useful data can be extracted including the rate of the structural deformations, changes in free-energy associated with various intermediate states, and the size of energy barriers that separate states within this dynamic process [165]. This information is obtainable as enough frames will hopefully be generated to identify a probability of switching between individual states along the pathway from the apo to the holo state [165].

5.2.4 – Full L-plastin Simulations

Although the simulations of the EF-hand motif of L-plastin have been revealing in identifying the mechanisms controlling Ca²⁺-binding, to identify how these mechanisms activate actin bundling, simulations of the full protein may be necessary. The advantage of using the single N-terminal domain was its relatively small size, allowing fast and relatively inexpensive MD simulations to be performed. However, now that the mechanisms which regulate Ca²⁺-binding are understood, larger simulations of the whole protein should be performed to identify how these mechanisms activate the rest of the protein. In addition to gaining additional insight into how this

protein activates, these simulations may reveal additional, and novel, targets to block the activation of L-plastin.

5.2.5 – Additional Mutants and Differential Scanning Calorimetry (DSC) Experiments

In addition to running computational simulations, there is still more experimental work that can be done. To further validate the model of switching generated for L-plastin's EF-hand, differential scanning calorimetry (DSC) will be performed on the apo state of all protein mutants. This may should help identify that the starting state of each construct is equivalent, or if the H5-tail helps determine structural stability by altering DSC values, this data may help confirm the ranking of the stabilities of the protein constructs.

Although more experiments can be done on the mutants which have already been created, computational models have also identified residues in the N-terminal region of L-plastin's EF-hand which may contribute to the switching process. Much like the H5-tail in its apo state, this region of the EF-hand remains unstructured. Therefore, to fully clarify the involvement of the N-terminal tail in the switching process, mutants should be generated in the N-terminus, and studied using the same methods applied to the H5-tail. In this way a direct comparison can be drawn between the effects of these mutants, mutants in the H5-tail, and the wild-type protein to fully understand the contribution of the N-terminal and C-terminal regions to the conformational dynamics of the EF-hand.

Bibliography

1. Reddy, C.K., A. Das, and B. Jayaram, *Do water molecules mediate protein-DNA recognition?* J Mol Biol, 2001. **314**(3): p. 619-32.
2. Dudev, T. and C. Lim, *Ion selectivity strategies of sodium channel selectivity filters.* Acc Chem Res, 2014. **47**(12): p. 3580-7.
3. Henzler-Wildman, K. and D. Kern, *Dynamic personalities of proteins.* Nature, 2007. **450**(7172): p. 964-72.
4. Rembert, K.B., et al., *Molecular mechanisms of ion-specific effects on proteins.* J Am Chem Soc, 2012. **134**(24): p. 10039-46.
5. Karp, G., *Cell and molecular biology : concepts and experiments.* 6th ed. 2010, Hoboken, NJ: John Wiley.
6. Berne, R.M., B.M. Koeppen, and B.A. Stanton, *Berne & Levy physiology.* 6th ed. 2010, Philadelphia, PA: Mosby/Elsevier. xii, 836 p.
7. Yamniuk, A.P. and H.J. Vogel, *Calmodulin's flexibility allows for promiscuity in its interactions with target proteins and peptides.* Mol Biotechnol, 2004. **27**(1): p. 33-57.
8. Gifford, J.L., et al., *Comparing the calcium binding abilities of two soybean calmodulins: towards understanding the divergent nature of plant calmodulins.* Plant Cell, 2013. **25**(11): p. 4512-24.
9. Sun, X.R., et al., *Fast GCaMPs for improved tracking of neuronal activity.* Nat Commun, 2013. **4**: p. 2170.
10. Gabelli, S.B., et al., *Regulation of the NaV1.5 cytoplasmic domain by calmodulin.* Nat Commun, 2014. **5**: p. 5126.

11. Yanez, M., J. Gil-Longo, and M. Campos-Toimil, *Calcium binding proteins*. Adv Exp Med Biol, 2012. **740**: p. 461-82.
12. Murray, K.D., et al., *Differential regulation of brain-derived neurotrophic factor and type II calcium/calmodulin-dependent protein kinase messenger RNA expression in Alzheimer's disease*. Neuroscience, 1994. **60**(1): p. 37-48.
13. Martinez, J., et al., *Parkinson's disease-associated alpha-synuclein is a calmodulin substrate*. J Biol Chem, 2003. **278**(19): p. 17379-87.
14. Limback-Stokin, K., et al., *Nuclear calcium/calmodulin regulates memory consolidation*. J Neurosci, 2004. **24**(48): p. 10858-67.
15. Kretsinger, R.H., V.N. Uversky, and E.A. Permiakov, *Encyclopedia of metalloproteins*. 2013, New York: Springer Reference. 4 volumes (lvii, 2574 pages).
16. Makita, N., et al., *Novel calmodulin mutations associated with congenital arrhythmia susceptibility*. Circ Cardiovasc Genet, 2014. **7**(4): p. 466-74.
17. Heizmann, C.W. and K. Braun, *Changes in Ca(2+)-binding proteins in human neurodegenerative disorders*. Trends Neurosci, 1992. **15**(7): p. 259-64.
18. Lehninger, A.L., D.L. Nelson, and M.M. Cox, *Lehninger principles of biochemistry*. 6th ed. 2013, New York: W.H. Freeman.
19. Zhang, M. and H.J. Vogel, *Two-dimensional NMR Studies of Selenomethionyl Calmodulin*. Journal of Molecular Biology, 1994. **239**(4): p. 545-554.
20. Ngo, V., et al., *Quantum effects in cation interactions with first and second coordination shell ligands in metalloproteins*. J Chem Theory Comput, 2015. **11**(10): p. 4992-5001.
21. Berridge, M.J., M.D. Bootman, and H.L. Roderick, *Calcium signalling: dynamics, homeostasis and remodelling*. Nature Reviews Molecular Cell Biology, 2003. **4**: p. 517+.

22. Cooper, G.M. and R.E. Hausman, *The cell : a molecular approach*. 6th ed. 2013, Sunderland, MA: Sinauer Associates. xxv, 832 p.
23. Ishida, H. and H.J. Vogel, *EF-Hand Proteins*, in *Encyclopedia of Metalloproteins*, R.H. Kretsinger, V.N. Uversky, and E.A. Permyakov, Editors. 2013, Springer New York: New York, NY. p. 766-775.
24. Harris, T.J.C. and U. Tepass, *Adherens junctions: from molecules to morphogenesis*. Nature Reviews Molecular Cell Biology, 2010. **11**: p. 502+.
25. Tepass, U., et al., *Cadherins in embryonic and neural morphogenesis*. Nature Reviews Molecular Cell Biology, 2000. **1**: p. 91.
26. Gerke, V. and S.E. Moss, *Annexins: From Structure to Function*. Physiological Reviews, 2002. **82**(2): p. 331-371.
27. Sutton, R.B., et al., *Structure of the first C2 domain of synaptotagmin I: A novel Ca²⁺/phospholipid-binding fold*. Cell, 1995. **80**(6): p. 929-938.
28. McCarthy, M.R., et al., *Impact of methionine oxidation on calmodulin structural dynamics*. Biochemical and Biophysical Research Communications, 2015. **456**(2): p. 567-572.
29. Gifford, Jessica L., Michael P. Walsh, and Hans J. Vogel, *Structures and metal-ion-binding properties of the Ca²⁺-binding helix-loop-helix EF-hand motifs*. Biochemical Journal, 2007. **405**(2): p. 199-221.
30. Bhattacharya, S., C.G. Bunick, and W.J. Chazin, *Target selectivity in EF-hand calcium binding proteins*. Biochim Biophys Acta, 2004. **1742**(1-3): p. 69-79.
31. Kretsinger, R.H. and C.E. Nockolds, *Carp muscle calcium-binding protein. II. Structure determination and general description*. J Biol Chem, 1973. **248**(9): p. 3313-26.

32. Shea, J.E. and C.L. Brooks, *From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding*. Annual Review of Physical Chemistry, 2001. **52**: p. 499-535.
33. Kinoshita, M., *Importance of translational entropy of water in biological self-assembly processes like protein folding*. Int J Mol Sci, 2009. **10**(3): p. 1064-80.
34. Ambroggio, X.I. and B. Kuhlman, *Design of protein conformational switches*. Curr Opin Struct Biol, 2006. **16**(4): p. 525-30.
35. Ha, J.H. and S.N. Loh, *Protein conformational switches: from nature to design*. Chemistry, 2012. **18**(26): p. 7984-99.
36. Zhou, Y., et al., *Prediction of EF-hand calcium-binding proteins and analysis of bacterial EF-hand proteins*. Proteins: Structure, Function, and Bioinformatics, 2006. **65**(3): p. 643-655.
37. Means, A.R., J.S. Tash, and J.G. Chafouleas, *Physiological implications of the presence, distribution, and regulation of calmodulin in eukaryotic cells*. Physiol Rev, 1982. **62**(1): p. 1-39.
38. Nyegaard, M., et al., *Mutations in calmodulin cause ventricular tachycardia and sudden cardiac death*. Am J Hum Genet, 2012. **91**(4): p. 703-12.
39. Butler, L.S., et al., *Limbic epilepsy in transgenic mice carrying a Ca²⁺/calmodulin-dependent kinase II alpha-subunit mutation*. Proc Natl Acad Sci U S A, 1995. **92**(15): p. 6852-5.
40. Wang, B., K.M. Sullivan, and K. Beckingham, *Drosophila calmodulin mutants with specific defects in the musculature or in the nervous system*. Genetics, 2003. **165**(3): p. 1255-68.
41. Norris, V., et al., *Calcium in bacteria: a solution to which problem?* Mol Microbiol, 1991. **5**(4): p. 775-8.
42. Epstein, P.N., P.A. Overbeek, and A.R. Means, *Calmodulin-induced early-onset diabetes in transgenic mice*. Cell, 1989. **58**(6): p. 1067-73.

43. Lai, M., et al., *Modulation of Calmodulin Lobes by Different Targets: An Allosteric Model with Hemiconcerted Conformational Transitions*. PLOS Computational Biology, 2015. **11**(1): p. e1004063.
44. Linse, S., A. Helmersson, and S. Forsen, *Calcium binding to calmodulin and its globular domains*. J Biol Chem, 1991. **266**(13): p. 8050-4.
45. *Calmodulin and Signal Transduction*. 1998, 24-28 Oval Road, London NW1 7DX, UK: ACADEMIC PRESS. 473.
46. Zhang, M., et al., *The effect of Met-->Leu mutations on calmodulin's ability to activate cyclic nucleotide phosphodiesterase*. J Biol Chem, 1994. **269**(22): p. 15546-52.
47. de Arruda, M.V., et al., *Fimbrin is a homologue of the cytoplasmic phosphoprotein plastin and has domains homologous with calmodulin and actin gelation proteins*. J Cell Biol, 1990. **111**(3): p. 1069-79.
48. Miyakawa, T., et al., *Different Ca(2)(+)-sensitivities between the EF-hands of T- and L-plastins*. Biochem Biophys Res Commun, 2012. **429**(3-4): p. 137-41.
49. Shinomiya, H., *Plastin family of actin-bundling proteins: its functions in leukocytes, neurons, intestines, and cancer*. Int J Cell Biol, 2012. **2012**: p. 213492.
50. Ishida, H., et al., *The Calcium-Dependent Switch Helix of L-Plastin Regulates Actin Bundling*. Sci Rep, 2017. **7**: p. 40662.
51. Todd, E.M., L.E. Deady, and S.C. Morley, *The actin-bundling protein L-plastin is essential for marginal zone B cell development*. J Immunol, 2011. **187**(6): p. 3015-25.
52. Wang, C., et al., *Actin-bundling protein L-plastin regulates T cell activation*. J Immunol, 2010. **185**(12): p. 7487-97.

53. Lin, C.S., et al., *Identification of I-plastin, a human fimbrin isoform expressed in intestine and kidney*. Mol Cell Biol, 1994. **14**(4): p. 2457-67.
54. Namba, Y., et al., *Human T cell L-plastin bundles actin filaments in a calcium-dependent manner*. J Biochem, 1992. **112**(4): p. 503-7.
55. Staiger, C.J., et al., *Actin: A Dynamic Framework for Multiple Plant Cell Functions*. 2013: Springer Netherlands.
56. Nooren, I.M. and J.M. Thornton, *Structural characterisation and functional significance of transient protein-protein interactions*. J Mol Biol, 2003. **325**(5): p. 991-1018.
57. Traut, T.W., *Dissociation of enzyme oligomers: a mechanism for allosteric regulation*. Crit Rev Biochem Mol Biol, 1994. **29**(2): p. 125-63.
58. Li, H., et al., *Representation of Ion-Protein Interactions Using the Drude Polarizable Force-Field*. J Phys Chem B, 2015. **119**(29): p. 9401-16.
59. MacKerell, A.D., Jr., N. Banavali, and N. Foloppe, *Development and current status of the CHARMM force field for nucleic acids*. Biopolymers, 2000. **56**(4): p. 257-65.
60. Mackerell, A.D., Jr., *Empirical force fields for biological macromolecules: overview and issues*. J Comput Chem, 2004. **25**(13): p. 1584-604.
61. Alexander D. Mackerell, J., *CHARMM force fields, parameterization strategies and future/ongoing force field developments*. University of Maryland, Baltimore - School of Pharmacy, 2006.
62. Lamoureux, G. and B. Roux, *Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm*. Journal of Chemical Physics, 2003. **119**(6): p. 3025-3039.

63. Sheu, S.Y., et al., *Energetics of hydrogen bonds in peptides*. Proc Natl Acad Sci U S A, 2003. **100**(22): p. 12683-7.
64. Bandyopadhyay, S., S. Chakraborty, and B. Bagchi, *Secondary structure sensitivity of hydrogen bond lifetime dynamics in the protein hydration layer*. J Am Chem Soc, 2005. **127**(47): p. 16660-7.
65. Bandyopadhyay, S., et al., *Sensitivity of polar solvation dynamics to the secondary structures of aqueous proteins and the role of surface exposure of the probe*. J Am Chem Soc, 2005. **127**(11): p. 4071-5.
66. Bellissent-Funel, M.C., et al., *Water Determines the Structure and Dynamics of Proteins*. Chem Rev, 2016. **116**(13): p. 7673-97.
67. Lamoureux, G., A.D.M. Jr., and B.t. Roux, *A simple polarizable model of water based on classical Drude oscillators*. The Journal of Chemical Physics, 2003. **119**(10): p. 5185-5197.
68. Mark, P. and L. Nilsson, *Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K*. The Journal of Physical Chemistry A, 2001. **105**(43): p. 9954-9960.
69. Deng, Y.Q. and B. Roux, *Hydration of amino acid side chains: Nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules*. Journal of Physical Chemistry B, 2004. **108**(42): p. 16567-16576.
70. Kollman, P., *Free energy calculations: Applications to chemical and biochemical phenomena*. Chemical Reviews, 1993. **93**(7): p. 2395-2417.
71. Nguyen, T.H. and D.D. Minh, *Intermediate Thermodynamic States Contribute Equally to Free Energy Convergence: A Demonstration with Replica Exchange*. J Chem Theory Comput, 2016. **12**(5): p. 2154-61.

72. Jiang, W. and B. Roux, *Free Energy Perturbation Hamiltonian Replica-Exchange Molecular Dynamics (FEP/H-REMD) for Absolute Ligand Binding Free Energy Calculations*. J Chem Theory Comput, 2010. **6**(9): p. 2559-2565.
73. Liu, P., et al., *Replica exchange with solute tempering: a method for sampling biological systems in explicit water*. Proc Natl Acad Sci U S A, 2005. **102**(39): p. 13749-54.
74. Simons, K.T., et al., *Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions*. J Mol Biol, 1997. **268**(1): p. 209-25.
75. Lewis, E.A. and K.P. Murphy, *Isothermal Titration Calorimetry*, in *Protein-Ligand Interactions: Methods and Applications*, G. Ulrich Nienhaus, Editor. 2005, Humana Press: Totowa, NJ. p. 1-15.
76. Pierce, M.M., C.S. Raman, and B.T. Nall, *Isothermal Titration Calorimetry of Protein-Protein Interactions*. Methods, 1999. **19**(2): p. 213-221.
77. Dutta, A.K., J. Rösger, and K. Rajarathnam, *Using Isothermal Titration Calorimetry to Determine Thermodynamic Parameters of Protein-Glycosaminoglycan Interactions*. Methods in molecular biology (Clifton, N.J.), 2015. **1229**: p. 315-324.
78. Gill, P., T.T. Moghadam, and B. Ranjbar, *Differential Scanning Calorimetry Techniques: Applications in Biology and Nanoscience*. Journal of Biomolecular Techniques : JBT, 2010. **21**(4): p. 167-193.
79. Vogel, H.J., *BCEM 553 - Structural Biology: NMR lectures (1-12)*, J.K. Fanning, Editor. 2016.
80. Bodenhausen, G. and D.J. Ruben, *Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy*. Chemical Physics Letters, 1980. **69**(1): p. 185-189.

81. Mitchell, T.N. and B. Costisella, *NMR--from spectra to structures : an experimental approach*. 2nd rev. and expanded ed. 2007, Berlin ; New York: Springer. xiii, 207 p.
82. Muchmore, D.C., et al., *Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance*. *Methods Enzymol*, 1989. **177**: p. 44-73.
83. Zhekova, H.R., et al., *Selective ion binding and transport by membrane proteins - A computational perspective*. *Coordination Chemistry Reviews*, 2017. **345**: p. 108-136.
84. Lemkul, J.A., et al., *An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications*. *Chemical Reviews*, 2016. **116**(9): p. 4983-5013.
85. Albaugh, A., et al., *Advanced Potential Energy Surfaces for Molecular Simulation*. *Journal of Physical Chemistry B*, 2016. **120**(37): p. 9811-9832.
86. Gresh, N., et al., *Anisotropic, polarizable molecular mechanics studies of inter- and intramolecular interactions and ligand-macromolecule complexes. A bottom-up strategy*. *Journal of Chemical Theory and Computation*, 2007. **3**(6): p. 1960-1986.
87. Ponder, J.W., et al., *Current Status of the AMOEBA Polarizable Force Field*. *Journal of Physical Chemistry B*, 2010. **114**(8): p. 2549-2564.
88. Shi, Y., et al., *Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins*. *Journal of Chemical Theory and Computation*, 2013. **9**(9): p. 4046-4063.
89. Patel, S. and C.L. Brooks, *CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations*. *Journal of Computational Chemistry*, 2004. **25**(1): p. 1-15.
90. Christensen, A.S., et al., *Intermolecular interactions in the condensed phase: Evaluation of semi-empirical quantum mechanical methods*. *Journal of Chemical Physics*, 2017. **147**(16).

91. Christensen, A.S., et al., *Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications*. Chemical Reviews, 2016. **116**(9): p. 5301-5337.
92. Peng, X.D., et al., *Accurate Evaluation of Ion Conductivity of the Gramicidin A Channel Using a Polarizable Force Field without Any Corrections*. Journal of Chemical Theory and Computation, 2016. **12**(6): p. 2973-2982.
93. Lemkul, J.A. and A.D. MacKerell, *Polarizable Force Field for DNA Based on the Classical Drude Oscillator: II. Microsecond Molecular Dynamics Simulations of Duplex DNA*. Journal of Chemical Theory and Computation, 2017. **13**(5): p. 2072-2085.
94. Savelyev, A. and A.D. MacKerell, *Competition among Li⁺, Na⁺, K⁺, and Rb⁺ Monovalent Ions for DNA in Molecular Dynamics Simulations Using the Additive CHARMM36 and Drude Polarizable Force Fields*. Journal of Physical Chemistry B, 2015. **119**(12): p. 4428-4440.
95. Qiu, F., et al., *Molecular mechanism of Zn²⁺ inhibition of a voltage-gated proton channel*. Proceedings of the National Academy of Sciences of the United States of America, 2016. **113**(40): p. E5962-E5971.
96. Wheatley, R.W., et al., *Elucidating factors important for monovalent cation selectivity in enzymes: E-coli beta-galactosidase as a model*. Physical Chemistry Chemical Physics, 2015. **17**(16): p. 10899-10909.
97. Harder, E., A.D. MacKerell, and B. Roux, *Many-Body Polarization Effects and the Membrane Dipole Potential*. Journal of the American Chemical Society, 2009. **131**(8): p. 2760-+.

98. Harder, E., et al., *Atomic level anisotropy in the electrostatic modeling of lone pairs for a polarizable force field based on the classical Drude oscillator*. Journal of Chemical Theory and Computation, 2006. **2**(6): p. 1587-1597.
99. Noskov, S.Y., G. Lamoureux, and B. Roux, *Molecular dynamics study of hydration in ethanol-water mixtures using a polarizable force field*. Journal of Physical Chemistry B, 2005. **109**(14): p. 6705-6713.
100. Kunz, A.P.E. and W.F. van Gunsteren, *Development of a Nonlinear Classical Polarization Model for Liquid Water and Aqueous Solutions: COS/D*. Journal of Physical Chemistry A, 2009. **113**(43): p. 11570-11579.
101. Drude, P., R.A. Millikan, and C.R. Mann, *The Theory of Optics: by Paul Drude; Tr. from the German by C. Riborg Mann and Robert A. Millikan*. 1902, New York; London; Bombay: Longmans, Green, and Co.
102. Waldman, M. and R.G. Gordon, *Generalized Electron Gas-Drude Model Theory for Ion-Molecule Forces*. Journal of Chemical Physics, 1979. **71**(3): p. 1353-1358.
103. Drude, P., et al., *The Theory of Optics: by Paul Drude; Tr. from the German by C. Riborg Mann and Robert A. Millikan*. 1902, New York; London; Bombay: Longmans, Green, and Co.
104. Lamoureux, G., et al., *Polarizable model of chloroform based on classical Drude oscillators*. Chemical Physics Letters, 2009. **468**(4-6): p. 270-274.
105. Lamoureux, G., et al., *A polarizable model of water for molecular dynamics simulations of biomolecules*. Chemical Physics Letters, 2006. **418**(1,3): p. 245-249.
106. Lamoureux, G. and B. Roux, *Absolute hydration free energy scale for alkali and halide ions established from simulations with a polarizable force field*. Journal of Physical Chemistry B, 2006. **110**(7): p. 3308-3322.

107. Riahi, S., B. Roux, and C.N. Rowley, *QM/MM molecular dynamics simulations of the hydration of Mg(II) and Zn(II) ions*. Canadian Journal of Chemistry, 2013. **91**(7): p. 552-558.
108. Jiang, W., et al., *High-Performance Scalable Molecular Dynamics Simulations of a Polarizable Force Field Based on Classical Drude Oscillators in NAMD*. Journal of Physical Chemistry Letters, 2011. **2**(2): p. 87-92.
109. Lemkul, J.A., et al., *Implementation of Extended Lagrangian Dynamics in GROMACS for Polarizable Simulations Using the Classical Drude Oscillator Model*. Journal of Computational Chemistry, 2015. **36**(19): p. 1473-1479.
110. Andersen, H.C., D. Chandler, and J.D. Weeks, *Roles of Repulsive and Attractive Forces in Liquids - Optimized Random Phase Approximation*. Journal of Chemical Physics, 1972. **56**(8): p. 3812-+.
111. Weeks, J.D., D. Chandler, and H.C. Andersen, *Perturbation Theory of Thermodynamic Properties of Simple Liquids*. Journal of Chemical Physics, 1971. **55**(11): p. 5422-+.
112. Baker, C.M., et al., *Accurate Calculation of Hydration Free Energies using Pair-Specific Lennard-Jones Parameters in the CHARMM Drude Polarizable Force Field*. Journal of Chemical Theory and Computation, 2010. **6**(4): p. 1181-1198.
113. Zhao, C.F., D.A. Caplan, and S.Y. Noskov, *Evaluations of the Absolute and Relative Free Energies for Antidepressant Binding to the Amino Acid Membrane Transporter LeuT with Free Energy Simulations*. Journal of Chemical Theory and Computation, 2010. **6**(6): p. 1900-1914.
114. Yu, H.B., et al., *Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field*. Journal of Chemical Theory and Computation, 2010. **6**(3): p. 774-786.

115. Roux, B., *The Calculation of the Potential of Mean Force Using Computer-Simulations*. Computer Physics Communications, 1995. **91**(1-3): p. 275-282.
116. Kumar, S., et al., *The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .I. The Method*. Journal of Computational Chemistry, 1992. **13**(8): p. 1011-1021.
117. Sitkoff, D., K.A. Sharp, and B. Honig, *Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models*. Journal of Physical Chemistry, 1994. **98**(7): p. 1978-1988.
118. Chambers, C.C., et al., *Model for aqueous solvation based on class IV atomic charges and first solvation shell effects*. Journal of Physical Chemistry, 1996. **100**(40): p. 16385-16398.
119. Jo, S., et al., *CHARMM-GUI: a web-based graphical user interface for CHARMM*. J Comput Chem, 2008. **29**(11): p. 1859-65.
120. Lee, J., et al., *CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field*. Journal of Chemical Theory and Computation, 2016. **12**(1): p. 405-413.
121. Lin, F.Y., et al., *Polarizable Force Field for Molecular Ions Based on the Classical Drude Oscillator*. J Chem Inf Model, 2018. **58**(5): p. 993-1004.
122. Tarek, M. and D.J. Tobias, *Role of protein-water hydrogen bond dynamics in the protein dynamical transition*. Phys Rev Lett, 2002. **88**(13): p. 138101.
123. Baker, E.N. and R.E. Hubbard, *Hydrogen bonding in globular proteins*. Prog Biophys Mol Biol, 1984. **44**(2): p. 97-179.
124. Knight, R.D., *Physics for scientists and engineers : a strategic approach*. 3rd ed. 2013, Boston: Pearson.
125. Mackerell, A., *Overview of the Classical Drude Oscillator Polarizable Force Field for Biomolecules*. 2017, University of Maryland.

126. Bhattacharyya, K., *Solvation dynamics and proton transfer in supramolecular assemblies*. Acc Chem Res, 2003. **36**(2): p. 95-101.
127. Rick, S.W., S.J. Stuart, and B.J. Berne, *Dynamical fluctuating charge force fields: Application to liquid water*. The Journal of Chemical Physics, 1994. **101**(7): p. 6141-6156.
128. Nandi, N. and B. Bagchi, *Anomalous Dielectric Relaxation of Aqueous Protein Solutions*. The Journal of Physical Chemistry A, 1998. **102**(43): p. 8217-8221.
129. Rembert, K.B., et al., *Molecular mechanisms of ion-specific effects on proteins*. Journal of the American Chemical Society, 2012. **134**(24): p. 10039-10046.
130. Gifford, Jessica L., Michael P. Walsh, and Hans J. Vogel, *Structures and metal-ion-binding properties of the Ca²⁺-binding helix-loop-helix EF-hand motifs*. Biochemical Journal, 2007. **405**(2): p. 199-221.
131. Toutenhoofd, S.L. and E.E. Strehler, *The calmodulin multigene family as a unique case of genetic redundancy: multiple levels of regulation to provide spatial and temporal control of calmodulin pools?* Cell Calcium, 2000. **28**(2): p. 83-96.
132. Lewit-Bentley, A. and S. Rety, *EF-hand calcium-binding proteins*. Curr Opin Struct Biol, 2000. **10**(6): p. 637-43.
133. Fang, L., et al., *Extension of the CHARMM General Force Field to Linked Nitrogen-Containing Heteroaromatic Rings*. Biophysical Journal, 2015. **108**(2): p. 160a-160a.
134. Yu, W.B., et al., *Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations*. Journal of Computational Chemistry, 2012. **33**(31): p. 2451-2468.
135. Sugita, Y. and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding*. Chemical Physics Letters, 1999. **314**(1-2): p. 141-151.

136. Bernardi, R.C., M.C. Melo, and K. Schulten, *Enhanced sampling techniques in molecular dynamics simulations of biological systems*. *Biochim Biophys Acta*, 2015. **1850**(5): p. 872-7.
137. Beauchamp KA, B.G., Lane TJ, Maibaum L, Haque I, and P. VS, *MSMBuilder2: Modeling Conformational Dynamics at the Picosecond to Millisecond Scale*. *J Chem Theory Comput*, 2011. **7**: p. 3412-3419.
138. Pande, V.S., K. Beauchamp, and G.R. Bowman, *Everything you wanted to know about Markov State Models but were afraid to ask*. *Methods*, 2010. **52**(1): p. 99-105.
139. Prinz, J.H., et al., *Markov models of molecular kinetics: generation and validation*. *J Chem Phys*, 2011. **134**(17): p. 174105.
140. Schwantes, C.R. and V.S. Pande, *Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9*. *Journal of Chemical Theory and Computation*, 2013. **9**(4): p. 2000-2009.
141. Schwantes, C.R. and V.S. Pande, *Modeling Molecular Kinetics with tICA and the Kernel Trick*. *Journal of Chemical Theory and Computation*, 2015. **11**(2): p. 600-608.
142. Suarez, E., J.L. Adelman, and D.M. Zuckerman, *Accurate Estimation of Protein Folding and Unfolding Times: Beyond Markov State Models*. *Journal of Chemical Theory and Computation*, 2016. **12**(8): p. 3473-3481.
143. Wu, H., et al., *Statistically optimal analysis of state-discretized trajectory data from multiple thermodynamic states*. *J Chem Phys*, 2014. **141**(21): p. 214106.
144. Senne, M., et al., *EMMA: A Software Package for Markov Model Building and Analysis*. *J Chem Theory Comput*, 2012. **8**(7): p. 2223-38.
145. Humphrey, W., A. Dalke, and K. Schulten, *VMD- Visual Molecular Dynamics*. *J. Molec. Graphics*, 1996. **14**: p. 33-38.

146. Deuffhard, P. and M. Weber, *Robust Perron cluster analysis in conformation dynamics*. Linear Algebra and Its Applications, 2005. **398**: p. 161-184.
147. Weber, M. and S. Kube, *Robust Perron Cluster Analysis for various applications in computational life science*. Computational Life Sciences, Proceedings, 2005. **3695**: p. 57-66.
148. Scherer, M.K., et al., *PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models*. Journal of Chemical Theory and Computation, 2015. **11**(11): p. 5525-5542.
149. Metzner, P., C. Schutte, and E. Vanden-Eijnden, *Transition Path Theory for Markov Jump Processes*. Multiscale Modeling & Simulation, 2009. **7**(3): p. 1192-1219.
150. Lamoureux, G., J.A.D. MacKerell, and B. Roux, *A simple polarizable model of water based on classical Drude oscillators*. The Journal of Chemical Physics, 2003. **119**(10): p. 5185-5197.
151. Lopes, P.E.M., et al., *Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator*. Journal of chemical theory and computation, 2013. **9**(12): p. 5430-5449.
152. Keeler, C., et al., *An Explicit Formulation Approach for the Analysis of Calcium Binding to EF-Hand Proteins Using Isothermal Titration Calorimetry*. Biophysical Journal, 2013. **105**(12): p. 2843-2853.
153. Zhang, M., et al., *Structural Basis for Calmodulin as a Dynamic Calcium Sensor*. Structure, 2012. **20**(5): p. 911-923.
154. Stossel, T.P., et al., *Filamins as integrators of cell mechanics and signalling*. Nature Reviews Molecular Cell Biology, 2001. **2**(2): p. 138-145.

155. Janji, B., et al., *Phosphorylation on Ser5 increases the F-actin-binding activity of L-plastin and promotes its targeting to sites of actin assembly in cells*. Journal of Cell Science, 2006. **119**(9): p. 1947-1960.
156. Lommel, M.J., et al., *L-plastin Ser5 phosphorylation in breast cancer cells and in vitro is mediated by RSK downstream of the ERK/MAPK pathway*. Faseb Journal, 2016. **30**(3): p. 1218-1233.
157. Messier, J.M., et al., *Fimbrin Localized to an Insoluble Cytoskeletal Fraction Is Constitutively Phosphorylated on Its Headpiece Domain in Adherent Macrophages*. Cell Motility and the Cytoskeleton, 1993. **25**(3): p. 223-233.
158. Shinomiya, H., et al., *Complete Primary Structure and Phosphorylation Site of the 65-Kda Macrophage Protein Phosphorylated by Stimulation with Bacterial Lipopolysaccharide*. Journal of Immunology, 1995. **154**(7): p. 3471-3478.
159. Lin, C.S., A. Lau, and T.F. Lue, *Analysis and mapping of plastin phosphorylation*. DNA and Cell Biology, 1998. **17**(12): p. 1041-1046.
160. Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading*. J Comput Chem, 2010. **31**(2): p. 455-61.
161. Spano, D., et al., *Molecular networks that regulate cancer metastasis*. Seminars in Cancer Biology, 2012. **22**(3): p. 234-249.
162. Mehlen, P. and A. Puisieux, *Metastasis: a question of life or death*. Nature Reviews Cancer, 2006. **6**: p. 449+.
163. Sugita, Y. and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding*. Chemical Physics Letters, 1999. **314**(1): p. 141-151.

164. Wang, L., R.A. Friesner, and B.J. Berne, *Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2)*. J Phys Chem B, 2011. **115**(30): p. 9431-8.
165. McGibbon, R.T., C.R. Schwantes, and V.S. Pande, *Statistical model selection for Markov models of biomolecular dynamics*. J Phys Chem B, 2014. **118**(24): p. 6475-81.