ABSTRACT

Title of dissertation: Examining the role of water and hydrophobicity in folding, aggregation, and allostery

Gregory S. Custer
Doctor of Philosophy, 2018

Dissertation directed by: Dr. Silvina Matysiak, Assistant Professor
Fischell Department of Bioengineering

Solvation and hydrophobicity drive many critical processes in nature, playing an important role in the folding of proteins, aggregation of surfactants into micelles, and in the disorder to order transitions that occur in some allosteric proteins upon ligand binding. Understanding how solvation and hydrophobicity affect these processes at a molecular level is important to finding new ways to use these processes, but it can be difficult to characterize these molecular details using experimental methods. Molecular dynamics (MD) simulations have proven useful in exploring details and thermodynamic conditions inaccessible in experiment, as MD captures the time evolution of the system at a molecular level. The phenomena which can be studied with an MD simulation depend on the mathematical model employed. Atomistic models provide the most detail for a simulation, but due to the computational costs required are not typically used to study phenomena which require system sizes larger than several hundred thousand atoms and time scales greater than several μs. Coarse-grained (CG) models reduce the complexity of the system being

studied, enabling the exploration of phenomena that occur at longer time scales. We have developed CG models to study protein folding and surfactant aggregation. Our CG surfactant model uses a three-body potential to account for hydrogen bonding without an explicit electrostatic potential, reducing the computational cost of the model. With our surfactant model we studied the stability of non-ionic micelles at extremes of temperature, capturing a window of thermal stability with destabilization of the micelles at both high and low temperatures. We observed changes in structure and solvation of the micelle at low temperatures, with a shift in enthalpy of solvation water providing the driving force for destabilization. Solvation and hydrophobicity are also critical in the folding and stability of proteins. With a modified version of our surfactant model we characterized the folding landscape of a designed sequence which folds to a helical bundle in water. We found two competing folded states which differ by rotation of a helix and trade between hydrophobic packing and solvation of protein's core. Changes in hydrophobic packing can also be involved in the disorder to order transitions that occur upon liganding binding in an allosteric protein, such as the *E. Coli* biotin ligase/repressor (BirA), in which ligand binding promotes dimerization. We have used atomistic simulations of BirA mutants in collaboration with an experimental group to identify structural changes, accompanied by changes in solvation, at both the dimer interface and ligand binding regions for distal mutations which impact the functionality of BirA.

Examining the role of water and hydrophobicity in folding, aggregation, and allostery

by

Gregory Scott Custer

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Dr. Silvina Matysiak, Chair/Advisor
Dr. Dorothy Beckett
Dr. Payel Das
Dr. Edward Eisenstein
Dr. Jeffery Klauda

Acknowledgments

I would like to thank my wife, for her constant support and encouragement in my studies, as she has been immeasurably helpful in my studies and a constant source of enjoyment in my life.

I would like to thank my advisor, Dr. Silvina Matysiak, for her help and encouragement throughout graduate school, for the patience she has shown in the delays in my projects, and the many things she has taught me over past six years. She has been an excellent advisor, and I am happy to have been a part of her lab. I would also like to thank Dr. Payel Das, for the insight she has provided on our projects and for the frequent help and advice she has given me in my research over the years. I would like to thank Dr. Dorothy Beckett, for allowing me to work on such an interesting project, and for her excellent help and discussions. I would like to thank Dr. Edward Eisenstein, for the time I was able to spend in his lab, for making him self available to be on committee, and the discussions he has provided on my projects in the past. I would also like to thank Dr. Jeffery Klauda, for taking the time to be on my committee.

I would like to thank my fellow lab members, who have provided me with continuous support over the years: Hongcheng Xu, Abhilash Sahoo, Sai Ganesan, Anu Nagarajan, Sudi Jawahery, Christopher Look, and many others who have made graduates school enjoyable and provided countless insights and discussions on my projects.

I would like to thank the students I have collaborated with in the Beckett

group, Jingheng Wang and Chenlu He, for their patience, thoughtful discussions, and excellent analysis that has made working with them enjoyable.

I would also like to thank my family, for their continuous encouragement, especially my parents, Scott and Connie Custer, for their excellent example, frequent advice, and constant support. I would like to thank my sisters, Melanie and Kate, for their uplifting attitudes and encouraging phone calls.

Table of Contents

Acknowledgements ii

List of Tables vi

List of Figures vii

List of Abbreviations ix

1 Introduction 1
1.1 Outline of Thesis . . . 5

2 Role of solvation in thermal stability of a non-ionic micelle 8
2.1 Overview . . . 8
2.2 Introduction . . . 8
2.3 Methods . . . 12
2.3.1 Model . . . 12
2.3.2 Parametrization of Model . . . 15
2.3.3 Graphics and visualization . . . 23
2.3.4 Identification of micelles . . . 23
2.3.5 Micelle radius . . . 25
2.3.6 Water Tetrahedral Order Parameter . . . 25
2.3.7 Water energy . . . 26
2.3.8 Mean contacts for surfactants leaving the micelle . . . 26
2.4 Results and Discussion . . . 28
2.4.0.1 Structural properties of surfactant and micelles . . . 28
2.4.0.2 Dynamics of surfactants leaving the micelle . . . 33
2.4.0.3 Changes in solvation of the micelle . . . 38
2.5 Conclusion . . . 42

3 Interplay between Conformational Heterogeneity and Hydration in the Folding Landscape of a Designed 3-helix Bundle 45
3.1 Overview . . . 45
3.2 Introduction . . . 45

3.3 Methods . . . 49
3.3.1 Model . . . 49
3.3.2 Simulation Setup . . . 52
3.3.3 Identification of 3-helix bundle . . . 54
3.3.4 Quantification of waters in the core . . . 56
3.4 Results and Discussion . . . 57
3.4.0.1 Folding landscape of a designed three-helix bundle . 57
3.4.0.2 Shifts in solvation with protein folding . . . 59
3.4.0.3 Solvation of folded substates . . . 61
3.5 Conclusion . . . 68
3.6 Acknowledgements . . . 69

4 Role of intramolecular contacts, hydrophobicity, and solvation in the allosteric communication of BirA 70
4.1 Overview . . . 70
4.2 Introduction . . . 70
4.3 Methods . . . 74
4.3.1 Simulation Setup . . . 74
4.3.2 Visualization and analysis . . . 77
4.4 Results and Discussion . . . 79
4.4.1 Comparison of apo and holo $BirA^{wt}$ . . . 79
4.4.2 Structure and dynamics of holoBirA alanine substitution variants . . . 82
4.4.2.1 Changes in pairwise interactions . . . 84
4.4.2.2 Structure and dynamics of the dimer interface and ligand binding surfaces . . . 86
4.4.2.3 Altered solvation and hydrophobic packing . . . 90
4.4.3 Shifts in residue contacts and solvation in BirA superrepressors 95
4.4.3.1 Structure of the dimer interface . . . 96
4.4.3.2 Shifts in electrostatics and solvation . . . 100
4.5 Conclusion . . . 104

5 Thesis Summary 107
5.0.1 Future Work . . . 110

Bibliography 111

List of Tables

2.1 Bonded parameters for CG surfactant 14
2.2 Non-bonded parameters for CG surfactant 15

3.1 Bonded parameters 50
3.2 Bonded parameters for CG protein 50
3.3 Non-bonded parameters for CG protein 52

List of Figures

2.1 Model $C_{12}E_5$ surfactant . . . 13
2.2 Water RDF comparison for CG and AA surfactant . . . 18
2.3 Contact probability maps for CG and AA surfactant . . . 20
2.4 Initial micelle structures and cluster size vs time . . . 22
2.5 Micelle formation . . . 28
2.6 Structural properties of monomer surfactant and micelles . . . 31
2.7 Surfactant–Micelle contacts during transition to solvent - Head–Head and Head–Tail contacts . . . 34
2.8 Surfactant–Micelle contacts during transition to solvent - Comparison at 300 K . . . 35
2.9 Surfactant–Micelle contacts during transition to solvent - Tail–Tail and Tail–Head contacts . . . 37
2.10 Water contacts and density within the micelle . . . 39
2.11 Surfactant solvation water energy and tetrahedral order . . . 41

3.1 Model for CG protein . . . 53
3.2 Radius of gyration time series for folding of CG protein . . . 55
3.3 Raw and normalized distribution of water densities in CG protein core 58
3.4 PMF for folding landscape of CG protein with representative structures for each state . . . 60
3.5 Solvation of CG protein . . . 62
3.6 Properties of helical rotation states in CG protein . . . 64
3.7 Time series of water density in core of CG protein . . . 65

4.1 BirA structure and mechanism . . . 72
4.2 BirA equilibration . . . 76
4.3 Comparison of apo and holoBirA . . . 80
4.4 Solvation of apo and holoBirA^{wt} . . . 82
4.5 Location of BirA alanine substitutions^{wt} . . . 83
4.6 Pairwise energy difference maps for ligand binding region of holoBirA alanine substitution variants . . . 85
4.7 Pairwise energy difference maps for the dimerization surface of holoBirA alanine substitution variants . . . 87

4.8 Structural changes from alanine substitutions in holoBirA 88
4.9 RMSD of central domain and ligand binding loops in BirA variants . 91
4.10 Solvation of holoBirA alanine substitution variants 92
4.11 Solvation of holoBirA ABL, BBL, and ligand for alanine substitution variants 94
4.12 Location of mutations in BirA superrepressors 95
4.13 Pairwise energy difference maps for holoBirA superrepressor variants 98
4.14 Pairwise energy difference maps for holoBirA superrepressor variant M310L 99
4.15 Helical structure in dimer interface loops of BirA superrepressor variants 99
4.16 Residues belonging to an electrostatic network in BirA 101
4.17 Changes in solvation of residues in BirA 103

List of Abbreviations

AA	all-atom
ABL	adenylate binding loop
BBL	biotin binding loop
BCCP	biotin carboxyl carrier protein
bio-5'-AMP	biotinoyl-5'-adenylate
bioOH-5'-AMP	biotinol-5'-AMP
CG	coarse-grained
CMC	critical micelle concentration
Escherichia coli	*E. coli*
HP	hydrophobic-polar
LJ	Lennard-Jones
mW	monatomic water
MD	molecular dynamics
NMR	nuclear magnetic resonance
RG	radius of gyration
RDF	radial distribution function
RMSD	root-mean-square deviation
RMSF	root-mean-square fluctuation
SAXS	small angle X-ray scattering
SW	Stillinger-Weber
wt	wild-type

Chapter 1: Introduction

Folding and aggregation of macromolecules in water are critical steps found in many biological processes. For most proteins, folding to a unique 3-dimensional structure is a necessary step to allow functionality. Aggregation of misfolded proteins is a hallmark of many common human diseases such as Alzheimer's, Parkinson's, and Huntington's disease [1, 2]. Aggregation processes are required for formation of lipid bilayers and micelles, which are important to both biology and industry. In particular, aggregation of surfactant molecules into micelles is critical for their use in industry as detergents, dispersants, emulsifiers, and drug delivery vehicles [3–6]. A driving force behind macromolecular folding and aggregation is the effect of solvating hydrophobic and hydrophilic solutes with water. While it can be difficult to examine the involvement of water in these processes at a molecular level using experimental techniques, great strides have been made in studying them through molecular dynamics (MD) simulations. MD is the simulation of the motion of atoms using computational models. The computational models which are developed for MD are called force fields, and define the interaction energies between atoms using mathematical equations referred to as potentials. Using these force fields, the motion of a system of atoms can be simulated by calculating interactions between the atoms

and integrating Newton's laws of motion. These MD simulations have been able to provide much insight into the molecular details of processes which may be difficult or impossible to study experimentally.

Many biological systems have been studied using traditional all-atom (AA) force fields, which model each atom in a system explicitly and attempt to replicate interactions between atoms as accurately as possible. However, for larger systems or longer time scales that may be necessary to study biological systems, coarse-grained (CG) models are frequently used. The cutoff for what system size or time scale is feasible with atomistic simulations depends both on the phenomena being studied and the computational resources available. Common limits for system size of AA simulations are in the hundreds of thousands of atoms and time scales for AA simulations do not typically go beyond several μs. CG models decrease the complexity of the system being studied, typically by grouping multiple atoms into single interaction sites and/or by using simplified models to describe interactions between atoms. CG models are not capable of the accuracy found in AA force fields, but are capable of studying processes on larger time scales or with larger systems. Additionally, there is a wide range in the level of "coarse-graining" found in commonly used CG models. United-atom models are very similar to AA models, having similar force fields and modeling all atoms but hydrogen. On the other end of the scale, many lattice models use single interaction sites, or beads, to represent an entire residue in a protein and restrict the simulation to 2-dimensional space. The choice of model depends on the time-scale of the phenomena being studied, the level of detail the researcher wishes to observe, and the computational resources

which are available.

Folding and aggregation are typically slow processes. Even with simplest known protein fold, the 35-residue villin headpiece 3-helix bundle [7], *ab initio* folding using an AA model takes on the order of 1 μs [8,9] of simulation time. This simulation time would be for a single folding event; characterizing the folding landscape of a protein requires many folding and unfolding events. Simulating aggregation of surfactants into micelles requires large systems representing hundreds of thousands of atoms, and multiple μs of simulation time may be required to observe complete formation [10]. Therefore, macromolecular folding and aggregation are common targets of CG models. A CG model improves the ability to sample a folding landscape by both reducing the computational complexity of the simulation and by smoothing energy barriers in the folding landscape, allowing more frequent transitions between different free energy states to be observed. As water often represents the bulk of the computational costs in atomistic simulations, one approach for study of large systems is to employ coarse-grained models that consider the effects of water implicitly, such as the generalized-Born [11], Gō model [12], HP [13], BLN [14], and other protein models [15–18]. While these models have been instrumental in studying many processes, the lack of explicit details concerning interactions with water make them unsuitable for studying the role of solvation in macromolecular folding and aggregation. Another method of reducing the computational cost due to water is to group multiple water molecules into a single interaction site [19,20]. However, this makes it difficult to represent hydrogen bonding, accurately capture the water phase diagram, and to examine interactions between solute and solvent in detail.

Folding and unfolding are also found in the disorder to order transitions which occur upon binding of a ligand to some allosteric proteins. Allostery is a term used to describe a common phenomena found in proteins, when binding of a ligand or other molecule at one site on a protein has and effect on other parts of the protein. Allostery is a common form of regulation found in proteins. It typically occurs with binding of a ligand to one portion a protein promoting or discouraging binding at another site, thus altering the functionality of the protein [21–23]. Experimental evidence is commonly used to identify allosteric proteins, and to identify residues which are critical to allosteric communication. However, it is difficult to identify the molecular mechanisms behind allostery using experimental evidence alone. Changes in a protein on an atomistic level are typically examined using X-ray crystallography. However, due to the necessarily static nature of crystal structures, they are not able to provide much insight into the dynamics of the protein. There are other techniques which can be used to investigate a more dynamical structure, such as small angle X-ray scattering (SAXS) and nuclear magnetic resonance spectroscopy (NMR), but these typically involve some loss of spatial resolution (for SAXS) or are limited to smaller proteins or domains (for NMR). MD simulations are particularly well-suited to examining fluctuations within a protein on an atomistic level. MD has been used with success to look at communication between residues in proteins, and to identify key residues for allostery [24, 25]. The ease with which mutations can be made in MD simulations and the detailed dynamics which can be observed make MD a useful tool in evaluating the structural changes which occur in protein mutants which have been evaluated experimentally.

There are four objectives of this work:

- To develop a CG model which can be used to study the effects of solvation on macromolecular folding and aggregation processes
- To investigate the role of solvation in the folding landscape of a three-helix bundle
- To characterize the roles of water and hydrophobicity in non-ionic surfactant micelle formation and destabilization at extremes of temperature
- To explore the importance of intramolecular contacts, particularly between hydrophobic regions, to allosteric switching in the *Escherichia coli* protein BirA

1.1 Outline of Thesis

This work is split into four chapters, the first of which encompasses this introduction and outline. In Chapter 2, we introduce a CG model that we have developed and used to study aggregation of non-ionic surfactant in explicit water. Specifically, we have used this model for exploration of aggregation in n-alkyl polyethylene glycol surfactants (C_nE_m). Each heavy atom of the surfactant is represented with a single bead. While interactions involving the nonpolar beads are represented using a typical two-body potential, interactions between polar beads incorporate a three-body potential that allows the model to mimic hydrogen bonding without the use of an electrostatic potential. This model is designed to interact with the mW explicit

water model, which makes use of the same three-body potential for polar interactions [26]. We have parameterized this model for the surfactant $C_{12}E_5$. The model has been used to explore micelle formation at a range of temperatures, observing a region of thermodynamic stability for micelle formation with both cold and heat destabilization of the micelles. The causes of the destabilization of micelles at high and low temperatures are characterized in a study of the breakdown of pre-formed micelles. In particular, we focus on the role of solvation in micelle destabilization at low temperature.

In Chapter 3, we extend the CG model described in Chapter 2 for use in the study of protein folding. Our model uses two beads to represent each amino acid, one for the backbone and one for the sidechain. Three residue types, hydrophobic, neutral, and turn, are used, differing in the size and presence of their sidechain bead. Backbone beads are polar, interacting with both each other and water using the a three-body potential. The three-body interactions in the backbone of the protein allow our model to emulate the hydrogen bonds which form along the axis of an α-helix without an electrostatic potential. When using a realistic pattern of hydrophobic and neutral residues we are able to fold α-helices. We have used this model to characterize the folding landscape of a designed three-helix bundle. The sequence is able to fold quickly to a helical bundle, and explore a wide range of conformations. Multiple folded substates are observed, with movement between the substates driven by a tradeoff between hydrophobic packing and water expulsion from the core. Our work indicates that while burial of hydrophobic residues can drive folding, this burial must be balanced with the overall solvation of the protein,

and competing states can be generated by designing a sequence in which hydrophobic residues cannot pack efficiently.

In Chapter 4, we investigate allostery in the *E. coli* protein BirA using MD simulations with an AA force field. In addition to simulations of wild-type (wt) BirA, we performed simulations of variants of BirA chosen based on the experimental work of the Beckett group. The computational results we obtained for wt BirA and its variants correlated strongly with those experimental results. We characterized the structure and dynamics of the protein, focusing on the allosterically coupled dimer interface (DI) and ligand binding (LB) regions. We were able to observe changes in the dynamics of loops in the DI and LB regions which help to explain the coupling observed between these regions in the experimental results of the Beckett group. In particular, we identify changes in the hydrophobic packing of both the LB region and DI which are associated with changes in the functionality of the protein. Distant mutations which cause changes in dimerization experimentally are shown to have a large effect on the packing and stability of the dimerization surface. The relationship between the fold and stability of the DI and the strength of dimerization found experimentally does not appear to be simple, with mutations in different regions strengthening or weakening dimerization through different molecular mechanisms.

Chapter 2: Role of solvation in thermal stability of a non-ionic micelle

2.1 Overview

This chapter is based on the author's publication in preparation for submission: Role of solvation in thermal stability of a non-ionic micelle; Custer, G.S., Matysiak, S, Das, P.;

2.2 Introduction

Micellization of surfactants is a critical process with many biological, industrial, and medical uses [3–6]. Micelles are one of many microstructures that can be formed from surfactants, depending on the thermodynamic conditions, surfactant concentration, and other properties of the solution [27, 28]. In a polar solvent, micelles are organized surfactant aggregates with a hydrophobic core and a hydrophilic shell. Their ability to carry nonpolar solutes through a polar solvent makes micelles useful as detergents, dispersants, emulsifiers, and drug-delivery vehicles [3–6]. These uses may require micelles to be stable at extremes of temperature or pressure. For example, when surfactants are used as dispersants for deep-water oil spills they need

to be stable at both low temperatures and high pressures. A primary driving force behind aggregation of surfactants into micelles is the effect of solvating hydrophobic and hydrophilic solutes with water [29–31]. Solvation of the hydrophobic tails of surfactant is minimized by their placement at the core of the micelle. Understanding the role solvation plays in micellization of surfactants at a variety of thermodynamic conditions is critical to their use in many industrial applications.

Micellization and properties of micelles have been studied by many experimental techniques including NMR [32–35], calorimetry [31, 36], light scattering [37, 38], small angle neutron scattering [39], sedimentation equilibrium [40], and HPLC [41, 42]. Experimental studies have characterized well the critical micelle concentration (CMC) of many surfactants, the concentration above which surfactants spontaneously aggregate into micelles, as well as other physical properties such as aggregation number, size, and shape [27, 28, 31, 40, 43–45]. The CMC of surfactants varies with temperature, with the CMC-temperature relationship typically appearing parabolic for nonionic surfactants, with a minimum in CMC occurring at some optimal temperature for micelle formation. When temperature is increased or decreased from this optimal temperature, the CMC increases and a higher surfactant concentration is required for micelle formation [31, 43, 44]. Thus at a given surfactant concentration it is possible to observe micelle destabilization at both low and high temperatures. While the existence of an increase in CMC at low temperatures is known, the molecular details of micelle destabilization at low temperatures have not been characterized. Solvation of micelles has also been studied experimentally, with evidence indicating that while micelles have a dry inner core, the polar

shell of the micelle is highly solvated and a considerable portion of the hydrophobic surface is in contact with water [41, 42, 46, 47]. Experiments have also identified a dehydration of micelles that occurs with increasing temperature [41, 42, 46]. Experiments are limited, however, in examining transient properties of micelles, during their formation or destabilization. It is also difficult to experimentally to capture a detailed solvation profile of the micelle.

Many of the drawbacks of experimental study of micelles can be avoided with computational methods, though these come with their own limitations. Molecular dynamics simulations have been used to provide a more detailed view of the physical properties of many micelles [10, 48–56]. The primarly limitation to what can be studied with computational models is accessing the time and length scales required to study the phenomena of interest. Both formation of micelles and destabilization are slow processes, and can require simulation lengths of 1μs or greater to study [10]. Combined with the large system size required to study the dilute surfactant concentrations at which micelles form, frequently in the hundreds of thousands of atoms, using atomistic simulations to characterize properties of micelles can be challenging. Therefore, aggregation processes such as micellization are frequent targets of coarse-grained modeling. As water typically represents the bulk of the computational costs in simulation, one approach for the study of aggregation is to treat water implicitly. Several models utilizing implicit solvent have been developed for the the study of micellization and properties of micelles [55, 57, 58]. While these models have provided much insight into the properties of micelles, the lack of explicit solvent interactions makes them unsuitable for examining the role of solvation in

micelle thermal stability. An alternative to reducing the computational burden of water is to group multiple water beads into a single interaction site [10, 50, 53, 59]. However, in these models the inability to represent the effects of hydrogen bonding makes it difficult to accurately capture the water phase diagram, and thus to model changes in solvation and water structure with temperature.

In this study, we have developed an off-lattice coarse-grained model of surfactants designed for the study of micelle solvation and thermal stability. Specifically, we have designed this model for exploration of aggregation of n-alkyl polyethylene glycol surfactants (C_nE_m), a common type of nonionic surfactant used in medicine, industry, experiments, and simulations [60]. To reduce computational costs while preserving the unique structural properties of water, we incorporate the single particle mW water model [26]. This model makes use of the three-body Stillinger–Weber (SW) potential [61] to emulate the hydrogen bond network found in water by enforcing a tetrahedral angle between neighboring water beads. Without any electrostatic interactions, this model is able to accurately capture the phase diagram of water [26]. This SW potential is also used to represent interactions between potential hydrogen bonding partners within the surfactant and between the surfactant and water. Each heavy atom in the surfactant is represented by a single bead, with four bead types used to represent the different carbon and oxygen atoms within the surfactant. We have previously developed homopolymer, heteropolymer, and protein models which incorporate the mW water model to study the role of solvation in protein folding [62–64]. The model we present here applies a similar approach to study the solvation and stability of the nonionic micelle $C_{12}E_5$ across a range of temperatures.

We characterize properties of $C_{12}E_5$ micelles in water and evaluate how changes in the structure and solvation of the micelle contribute to its destabilization at low and high temperatures.

2.3 Methods

2.3.1 Model

In the study presented here, we have represented each heavy atom of a surfactant molecule with one CG bead. Four different bead types, shown in Figure 2.1A, are used to represent the non-ionic surfactant pentaethylene glycol monododecyl ether ($C_{12}E_5$). The beads differ by both the element which they represent and the hydrophobicity of the domain to which they belong. The terms "head" and "tail" are used here to refer to ether domain (first 17 atoms of the molecule) and alkyl domain (last 11 carbon atoms of the molecule), respectively. The Ch bead represents carbons in the tail, while Cp represents carbons in the head. Ether oxygens are represented by the O bead, with the terminal hydroxyl group represented by the OH bead. The model incorporates the mW model to represent water [26]. Due to their similarity, OH and water beads are treated identically. Mass for each bead is based on its atomic counterpart, including hydrogen where appropriate. The mW bead has a mass of 18.015 g/mol [26], the Ch and Cp beads have a mass of 14.01 g/mol (carbon with two hydrogen atoms), and the O bead has a mass of 16 g/mol.

Parametrization of this CG model was performed by comparing properties of the CG system to simulations of surfactant molecules with the all-atom (AA)

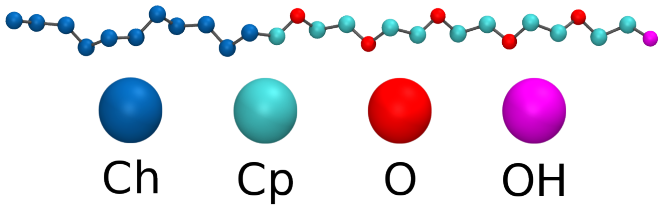

Figure 2.1: $C_{12}E_5$, with atoms color-coded by CG bead type.

force-field AMBER99, which has been validated extensively using experimental observations [65]. For the AA comparison, a two randomly distributed surfactant system in explicit water was simulated using the simulator GROMACS [66–69] with the SPC/E water model [70]. These simulations were performed at temperatures of 275, 300, and 350 K. The results were compared to simulations of a two surfactant system using our CG model. Parameters within the CG model were adjusted so that interactions between a surfactant and itself, between two surfactants, and between surfactant and water closely matched the interactions observed in the AA model.

Bonds and angles between beads are represented using simple harmonic potentials. Values used for all bond and angle parameters in our system can be found in Table 3.1. As all heavy atoms are represented explicitly, bond lengths and angles between atoms were taken directly from the AMBER99 force field [65].

Non-bonded interactions involving nonpolar beads (Ch and Cp) were represented by a 9-6 Lennard-Jones potential, with parameters for interactions between all pairs of beads shown in Table 3.3.1. The Stillinger-Weber (SW) potential was used to represent non-bonded interactions between polar beads (O and OH/mW). Again, due to their similarity the model treats OH and mW beads identically. The

Bonds		
Bead types	r_0 (Å)	K (kcal/Å^2)
C-C	1.526	248
C-O	1.410	256
Angles		
Bead types	θ_0 (°)	K (kcal/rad^2)
C-C-C	109.5	12
C-C-O	109.5	12
C-O-C	109.5	12

Table 2.1: Bonded parameters

three-body angular penalty found in the SW potential is used to enforce the tetrahedral hydrogen bond network found in water [26]. The SW potential has been used in the past to account for hydrogen bonding in water, polymer-water, and peptide-water systems [26, 62–64]. Here we use this potential to emulate water-water, water-surfactant, and surfactant-surfactant hydrogen bonding. For the interactions represented by the SW potential, initial parameters in the potential were set as found in Molinero and Moore [26]. Parameters controlling the potential well depth(ϵ), bead size (σ), and strength of the three-body interaction (λ) were modified for each interaction type involving Cp, Ch, or O beads (shown in Table 3.3.1). Since ether oxygens (represented by the O beads) are unable to act as hydrogen bond donors, the strength of the three-body interaction was reduced by 1/6th for interactions with one O bead and two mW beads, and by 1/3rd for interactions involving two O beads and one mW beads. Interactions between three O beads excluded the three-body portion of the potential, as it would not be possible to form

Lennard-Jones			
Bead types	ϵ (kcal/mol)	σ (Å)	
Ch-Ch	0.204	4.0	
Ch-Cp	0.224	4.0	
Cp-Cp	0.204	4.0	
Ch-O	0.255	3.55	
Ch-OH/mW	0.30	3.7	
Cp-O	0.289	3.55	
Cp-OH/mW	0.40	3.7	
Stillinger-Weber			
Bead types	ϵ (kcal/mol)	σ (Å)	λ
mW/OH-mW/OH-mW/OH[a]	6.189	2.3925	23.15
mW/OH-mW/OH-O	5.250	2.25	19.29
mW/OH-O-O	5.250	2.25	15.43
O-O-O	0.6375	2.1	0.0

[a] Parameters for mW bead taken from Molinero, et al. [26]

Table 2.2: Non-bonded parameters

hydrogen bonds.

2.3.2 Parametrization of Model

Parametrization of this CG model was performed by comparing properties of the CG system to simulation of surfactant molecules with the all-atom (AA) force-field AMBER99 [65]. AA simulations were performed using the simulator GROMACS [66–69] to simulate a two surfactant system in explicit water, with the SPC/E water model [70]. The AA simulations were performed at a pressure of 1 atm and at temperatures of 300, 350, and 400 K. Velocities were intialized using a Boltzmann distribution at the appropriate temperature. The V-rescale al-

gorithm [71], with a time constant of 0.1 ps, was used to regulate the temperature of water and surfactant independently. Pressure was maintained isotropically with the Parrinello-Rahman barostat [72], using a time constant of 2 ps and a compressibility of 4.5×10^{-5} bar^{-1}. Bond lengths were constrained using the LINCS algorithm [67].

Since all heavy atoms are represented explicitly in the described CG model, bond lengths and angles between atoms were derived directly from the AMBER99 force field [65] and parametrization was only necessary for the non-bonded interactions. Non-bonded interactions involving nonpolar beads (Ch and Cp) were represented by the following 9-6 Lennard-Jones (LJ) potential (Eq. 2.1):

$$E_{ij} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{9} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.1}$$

where ϵ_{ij} is the depth of the energy minimimum, σ_{ij} is the distance at which the potential equals zero, r_{ij} is the distance between atoms, and E_{ij} is the total non-bonded energy. All of these values are specific to the pair of atoms i and j.

Non-bonded interactions between polar beads (O, OH, and mW) were represented by the following SW potential (Eqs. 2.2, 2.3, 2.4):

$$E_{ijk} = \sum_{i}\sum_{j<i}\phi_2(r_{ij}) + \sum_{i}\sum_{j\neq i}\sum_{k>j}\phi_3(r_{ij}, r_{ik}, \theta_{ijk}) \tag{2.2}$$

$$\phi_2(r_{ij}) = A_{ij}\epsilon_{ij}\left[B_{ij}\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{p_{ij}} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{q_{ij}}\right] exp\left(\frac{\sigma_{ij}}{r_{ij} - a_{ij}\sigma_{ij}}\right) \tag{2.3}$$

$$\phi_3(r_{ij}, r_{ik}, \theta_{ijk}) = \lambda_{ijk}\epsilon_{ijk}\left[cos\theta_{ijk} - cos\theta_{0,ijk}\right]^2 exp\left(\frac{\gamma_{ij}\sigma_{ij}}{r_{ij} - a_{ij}\sigma_{ij}}\right) exp\left(\frac{\gamma_{ik}\sigma_{ik}}{r_{ik} - a_{ik}\sigma_{ik}}\right) \tag{2.4}$$

where E_{ijk} is the total non-bonded energy, ϕ_2 is the two-body portion of the potential, and ϕ_3 is the three-body portion of the potential. Within this potential, σ and

ϵ serve the same purpose as in the Lennard-Jones potential.

For the LJ potential, only the σ and ϵ values needed to be adjusted for each pair of beads. For the SW potential, the σ, ϵ, and λ values were adjusted. In the SW potential, λ controls the strength of the three-body interaction. Since ether oxygens are unable to act as hydrogen bond donors, the strength of the three-body interaction was reduced by change λ for three-body interactions involving O beads. For interactions with one O bead and two mW beads λ was reduced by 1/6th, and by 1/3rd for interactions involving two O beads and one mW beads. Interactions between three O beads excluded the three-body portion of the potential by setting λ to 0, as it would not be possible to form hydrogen bonds. Other parameters in the SW potential were kept identical to those found in the mW model [26]. The radial distribution function (RDF) of water (mW bead) around Ch, Cp, and O beads was used to adjust σ and ϵ for interactions with water. These parameters were adjusted to fit the RDF calculated for the CG simulation to that of the AA simulation, using data from the 300K simulation. Final results of this fitting are shown in Figure 2.2, with the RDF for the CG model comparing very well to the AA RDF. For each bead type we capture the first hydration shell, the depletion zone following this peak, and the second hydration shell.

For interactions within the surfactant, σ was set based on the minimum distance observed between a pair of bead types from the AA simulations. Contact probability maps were then compared between AA and CG simulations, with ϵ being adjusted until the general distribution in the maps matched. Figures 2.3A and 2.3B show contact probability maps for atoms within a single surfactant at 400K

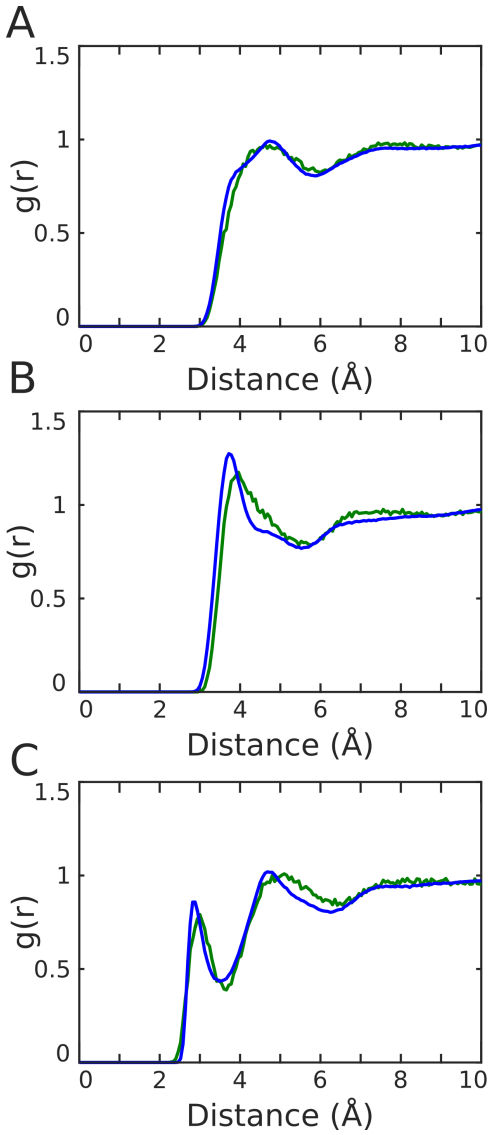

Figure 2.2: Radial distribution function at 300K of water around A) carbon in the surfactant tail (Ch), B) carbon in the surfactant bead (Cp), and C) ether oxygens (O). Data from all-atom simulation is in blue, while data from simulation the CG model is in green.

for AA and CG, respectively. Although the overall contact probability is slightly lower in our model, the overall distribution in the contact probability is captured well. The observed difference in contact probability may be due to differences in flexibility of the chain between the models. Both the AA (Figure 2.3A) and CG (Figure 2.3B) plots show that the surfactant is folding such that the polar head-group is contacting both itself and the hydrophobic tail. Contact probability maps for contacts between two different surfactants at 400K are shown in Figures 2.3C and 2.3D for AA and CG, respectively. These contacts are captured very well by our model, with both the magnitude and trends of the plot being similar in AA and CG simulations.

Simulation Setup

Simulations of the surfactant $C_{12}E_5$ were performed using the LAMMPS molecular dynamics package [73]. Surfactants were placed in a cubic box and solvated with CG water. Surfactant concentration for the simulations was 16mM (approximately 0.65% by weight). This concentration was chosen as it is above the experimental critical micelle concentration at 300K and 1 atm (0.062 mM) [74], but below the concentration needed to reach more complex mesophases (as high as 30 wt%) [27, 28]. Initial test simulations indicated that this concentration is high enough above the CMC that we can observe micelle formation on a feasible timescale while still being far below the transition to more complex mesophases. Integration was performed using the Verlet integrator with a timestep of 8 fs [75]. Velocities were initialized

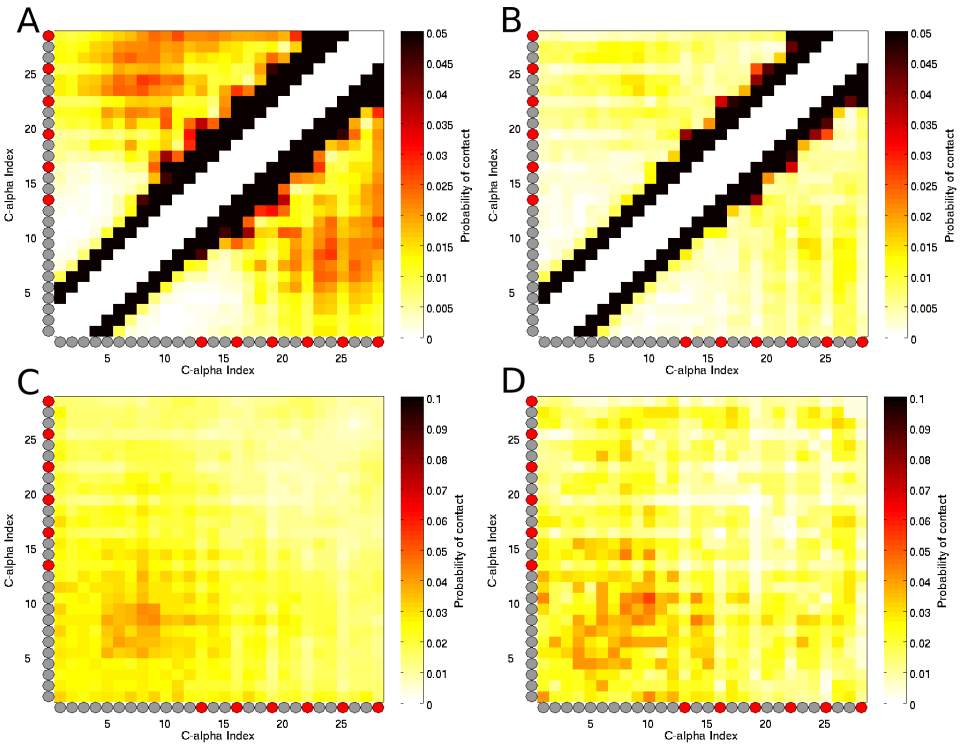

Figure 2.3: Contact probability maps for AA (plots A and C) and CG (plots B and D) simulation of surfactant at 400 K. X and Y axes of the plots represent heavy atom indices along a single surfactant, with circles indicating whether an atom is carbon (gray) or oxygen (red). Probabilities given are for contacts within a single surfactant for A and B, and for contacts between two interacting surfactants for C and D.

using a Boltzmann distribution at the appropriate temperature. Temperature and pressure were regulated using the Nosé-Hoover thermostat/barostat [76, 77], with temperature of the water and surfactant maintained independently. Pressure was set to 1 atm, with temperature varying depending on the simulation. Damping time was 100 fs for temperature coupling and 1 ps for pressure coupling.

Two initial CG simulations of micelle formation were performed at a temperature of 350 K to generate formed micelles for use in further simulations. Simulations of micelle formation had a system size of 160 surfactants in water at a concentration of 16 mM. The systems were initialized with surfactants randomly distributed throughout the simulation box. Following placement of the surfactants in the box, 543160 water beads were added to fill the remaining space in the box. A large, stable micelle formed in each simulation. To compare the window of stability for micelle formation, CG simulations of micelle formation were also performed at 275, 300, and 500 K.

CG simulations of micelle breakdown were initialized from pre-formed micelles containing 50 surfactants, taken from the simulations of micelle formation at 350 K. For each simulation, a single micelle was placed in a cubic box, sized such that the overall surfactant concentration was 16 mM. Following placement of the micelle in the box, approximately 172360 water beads were added to fill the remaining space in the box. Two simulations were performed at each temperature, each starting from a different initial micelle, shown in Figures 2.4A and 2.4B.

Simulations were performed at temperatures of 275, 300, 350, and 500 K. Simulation time varied depending on the temperature, with simulations being extended

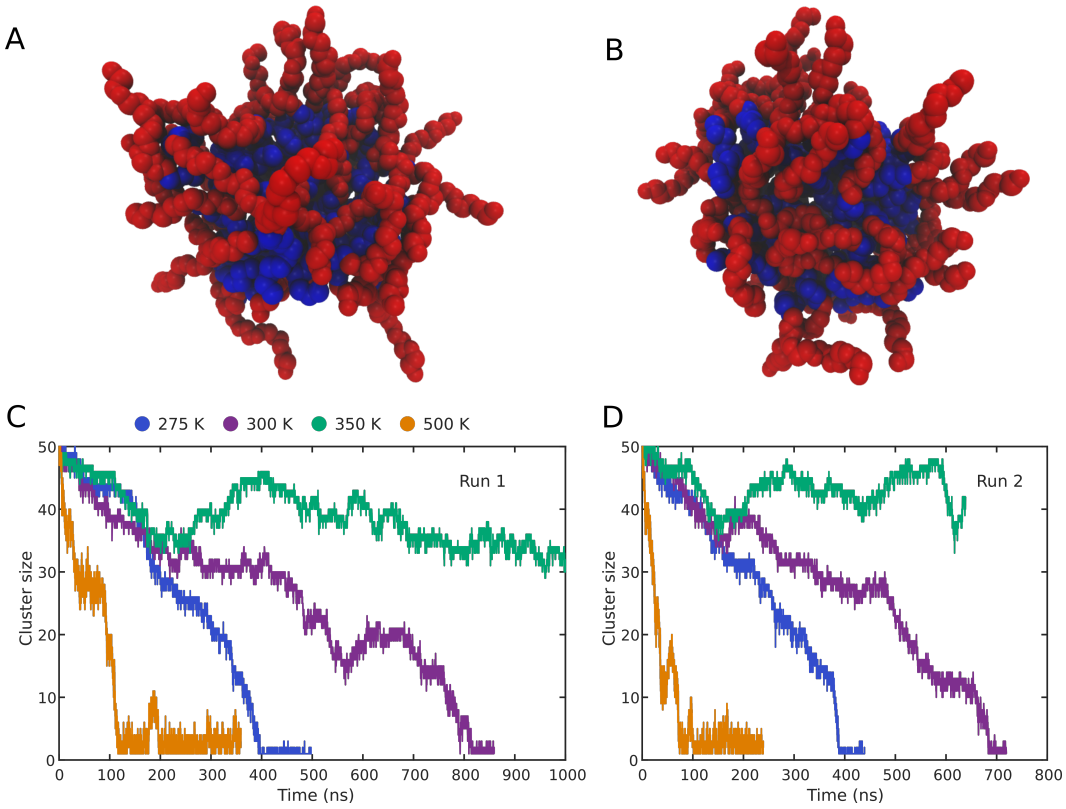

Figure 2.4: A and B) Structures of the two initial micelles used for simulation of micelle breakdown. Nonpolar domains are shown in blue, polar in red. C and D) Maximum cluster size vs time for simulation of micelle breakdown at 275, 300, 350, and 500 K.

until the micelle completely dissipated in the solution or it became apparent that the micelle was stable.

Simulations of a single CG surfactant in water were performed at temperatures of 275, 350, and 500 K. The surfactant was placed in a cubic box and solvated with 5203 water beads. Each simulation was run for 10 ns.

Simulations were also performed of pure CG water for calculation of bulk water properties. A cubic box with edge lengths of 30 Å was filled with 895 water beads. Simulations were performed at 275, 300, 350, and 500 K. Each simulation was run for 20 ns.

2.3.3 Graphics and visualization

Images of molecular structures in this work were generated using VMD [78]. Unless otherwise noted, red colors have been used for beads belonging to the head groups of the surfactant and blue colors for the tails.

2.3.4 Identification of micelles

Micelles were identified using a modified version of the GROMACS tool *g_clustsize* [66–69]. Surfactants were grouped such that all surfactants within a cluster have a minimum number of contacts, N, to other surfactants within that cluster. Surfactants are considered in contact when the distance between them is below a cutoff, R_{cut}. Both N and R_{cut} were optimized to identify clusters accurately. The optimal value we identified for N was 7 contacts. R_{cut} was found to be optimal at 4.5 Å.

This value corresponds approximately to the first minima in the radial distribution function between beads of different surfactants.

Micelle size in this work refers to the number of surfactants within a micelle, as identified by the clustering algorithm described above. As many micelle properties vary with size, a minimum micelle size of 35 and maximum micelle size of 50 was chosen for analysis. Micelle properties studied showed little variance with size across this range. Additionally, this range in micelle size was represented at each temperature.

In addition to identifying micelles, we must also identify surfactants participating in the core of a micelle. As the core of a micelle is predominantly made up of interacting hydrophobic domains, interactions between the Ch beads of surfactants were used to identify the core. To identify these clusters of Ch beads which make up the core, the previously described clustering algorithm was run using only Ch beads. The clusters identified in this analysis were considered to be the core of their corresponding micelles. As with clustering of surfactants into micelles, the number of contacts, N, required for participation in a core cluster was set to 7 and the cutoff, R_{cut} for contacts was 4.5 Å. Surfactants that are within the micelle but not identified as being within the core of the micelle are considered to be on the surface.

Classification of positions within micelle

Positions of beads within the micelle were measured relative to the center of mass of the micelle and then normalized. Normalization of distances to the center

of the micelle was based on an ellipsoid model for the micelle. Axes of the ellipsoid were set to the principal axes of the micelle, with the length of each axis of the ellipsoid set based on the radius of gyration of the micelle around the principal axes. For each bead, i, having a vector extending from the center of the mass to the bead, r_i, the normalized distance to the center would be calculated as:

$$r_{norm,i} = \frac{|r_i|}{\sqrt{\sum_{n=1}^{3} \left(\frac{r_i \cdot R_{g,n}}{|r_i|}\right)^2}} \tag{2.5}$$

where $R_{g,n}$ is the radius of gyration around the nth principal axis. Thus, the normalization factor for each distance to a bead is the radius of the ellipsoid which passes through that bead.

2.3.5 Micelle radius

Mean radii of micelle structrues were calculated using the convex hull of the micelle. A convex hull was calculated using all surfactants in the micelle, and the mean radius of the micelle was taken to be the average distance between the center of mass of the micelle and the surfactant beads on the surface of the convex hull.

2.3.6 Water Tetrahedral Order Parameter

The orientation of water beads in bulk and in the solvation shell around surfactant was characterized using an order parameter, q, for each water bead calculated as [79, 80]:

$$q = 1 - \frac{3}{8} \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left(cos\psi_{jk} + \frac{1}{3}\right)^2 \tag{2.6}$$

where ψ_{jk} is the angle between vectors extending from the given water bead to two of its four nearest neighbors, j and k. This order parameter scales between 0 for an ideal gas and 1 for a perfect tetrahedral network [79, 80].

2.3.7 Water energy

Total potential energy given for bulk and solvation water represents the total pairwise interaction energy for each water bead, as calculated within the simulation. Solvation water includes all water beads within 4 Åof each surfactant bead.

2.3.8 Mean contacts for surfactants leaving the micelle

Contacts between of surfactants transitioning from the micelle to solvent and surfactants within the micelle were quantified. For this analysis, we first identified surfactants which were part of the micelle and moved to completely to the solvent. As the goal of the analysis was to examine the type of contacts a surfactant made as it left the micelle, only surfactants which began in the micelle core and completely left the micelle were counted. For each surfactant in the simulation, the entire trajectory was analyzed. A transition to solvent would be counted at a time point for a surfactant if the following criteria were met at that time point: 1) The micelle was made up of at least 35 surfactants. 2) At the time point being checked, the surfactant had to be out of the micelle 3) For at least 80% of the 2 ns prior to that time point the surfactant was classified as part of the micelle's core. 4) For at least 80% of the 2 ns after the time point the surfactant was classified as not being part

of the micelle's core. 5) Within 4 ns following that time point, the surfactant had to spend at least 80% of a 2 ns time period completely out of the micelle. If all of these conditions were met at a time point, the surfactant was considered to have transitioned from the micelle to the solvent. The requirements were chosen to limit transitions to surfactants which started as part of the micelle and fully leave the micelle. The use of a percentage of the time in or out to classify transitions makes this analysis insensitive to brief contacts by a leaving surfactant with the micelle, while still ensuring that it is no longer part of the micelle.

After identifying surfactants that leave the micelle and the time which they leave the micelle, referred to here as t_0, contacts between transitioning surfactants and the micelle were counted. Contacts were classified based on the contacting region of the surfactant, splitting the surfactant into head and tail. For example, any contacts between beads in the head group of a leaving surfactant and the tail of the micelle are counted as head-tail contacts. Contacts were counted when the distance between two beads was less than or equal to 4.5 Å. Contacts were counted over time for the 2 ns preceding the surfactant leaving as well as the 2 ns after the surfactant left. The time series for each surfactant transition at a given temperature were aligned, and the mean number of contacts at each time point relative to t_0 was calculated. The resulting mean time series was normalized for each temperature by dividing by the maximum value observed at 350 K for that contact type. Thus, in each of the resulting plots the maximum value observed at 350 K will be 1.

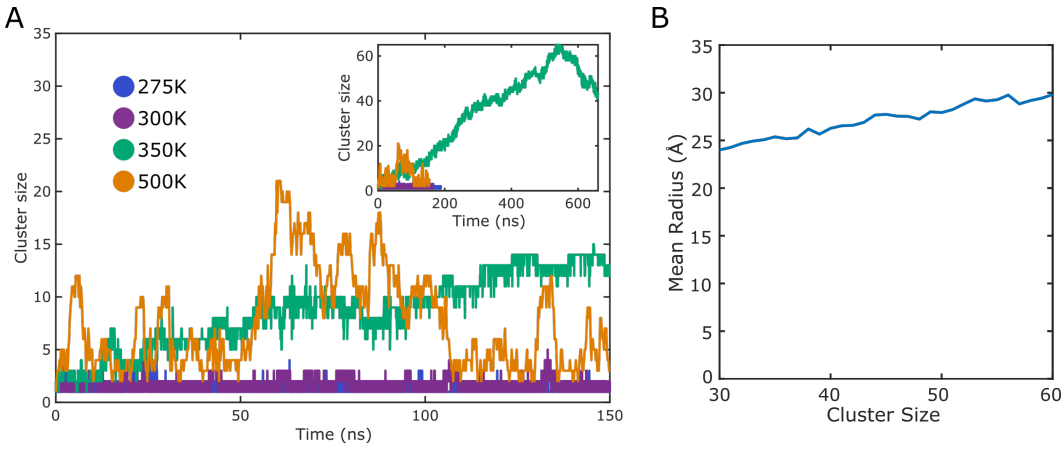

Figure 2.5: A) Maximum cluster size vs time for single simulation of micelle formation at 275, 300, 350, and 500 K. Inset includes full time series at 350 K. B) Average micelle radius for cluster sizes 30–60 at 350 K

2.4 Results and Discussion

2.4.0.1 Structural properties of surfactant and micelles

Initial simulations of micelle formation captured a window of thermal stability (Figure 2.5A), with micelle formation occuring at 350 K, but no stable micelles forming at 275, 300, or 500 K. Maximum micelle size at 350 K is reached in approximately 500 ns, though large fluctuations in size are still occurring at that time, making complete characterization of micelle formation impractical. Based on the micelle sizes observed, micelles containing 50 surfactant appear to be stable at 350 K. Thus micelles of size 50 were used as the initial structures for simulation of pre-formed micelles. Our micelle size is not far from that observed by others, as computational studies have indicated an average micelle size of 55-56 for $C_{12}E_5$ at

similar concentrations [10, 81]. While we observe roughly spherical micelles with average radius of approximately 27 Åfor a micelle size of 50 (Figure 2.5B), experimental work has found cylindrical micelles for $C_{12}E_5$ at this concentration [82]. However, for $C_{12}E_6$ spherical micelles are observed experimentally with a radius around 25 Å, [82] indicating that our results are similar when allowing for slight differences in hydrophobicity and polarity due to the coarse-grained nature of our model. These differences should not impair our ability to capture qualitative changes in micellization with temperature.

Initial structures for the pre-formed micelles used in simulation are shown in Figures 2.4A and 2.4B. A window of thermal stability is observed from the maximum cluster size time series (Figures 2.4C- 2.4D). Maximum stability is observed at 350 K, while high temperature destabilization is observed at 500 K and low temperature destabilization occurs at 275 and 300 K. It is also apparent that destabilization occurs more rapidly at 500 K compared to that at low temperature (275 and 300 K). At 275 K, the micelle breaks apart slowly but steadily. As temperature increases to 300 K, the micelle takes longer to break apart, but remains unstable. This window of thermal stability is identical to that observed for micelle formation (Figure 2.5A).

High temperature destabilization is due to an entropic effect, as a micelle is an ordered structure [31]. At low temperatures micelle breakdown is not as intuitive, but might be expected to occur due to changes in properties of water and its associated interactions with surfactant at low temperature, as occurs in cold denaturation of proteins [83,84]. An increase in the CMC of many surfactants at low temperature has been observed experimentally, indicating that cold destabilization

does occur for micelles [31,43,44]. To our knowledge, cold destabilization of micelles in simulation has not been previously studied. Since cold destabilization is driven by changes in solvation of the surfactant, it requires a water model that accurately captures the phase diagram of water, as the mW water model does [26]. Models which group multiple water molecules together or use implicit solvent are not able to accurately account for the structure of water in solution, and thus cannot capture cold denaturation.

To investigate the molecular factors underlying the micelle destabilization at lower temperatures, we first examined structural properties of the surfactants and micelle. The distribution of radius of gyration (RG) at each temperature for monomeric surfactant is shown in Figure 2.6A. As temperature increases, a second peak with decreased RG becomes accessible, with the surfactants folding into a hairpin-like structure in which the hydrophobic tail is in contact with the polar head. This folding can be observed in the angular distribution between the head group of a surfactant within the micelle and a vector between the center of the micelle and that same head group (Figure 2.6B). At lower temperatures (275 and 300 K), surfactants within the micelle remain extended, with a peak at low angles between the polar head and a vector to the micelle center, while at higher temperatures (350 and 500 K) the distribution shifts to larger angles. Thus, we observe folding of polar head groups for individual surfactants against the micelle as temperature increases. This may contribute to the increased stabiliity of the micelle by both reducing the overall envelope of the micelle and reducing the portion of the hydrophobic core accessible to solvent.

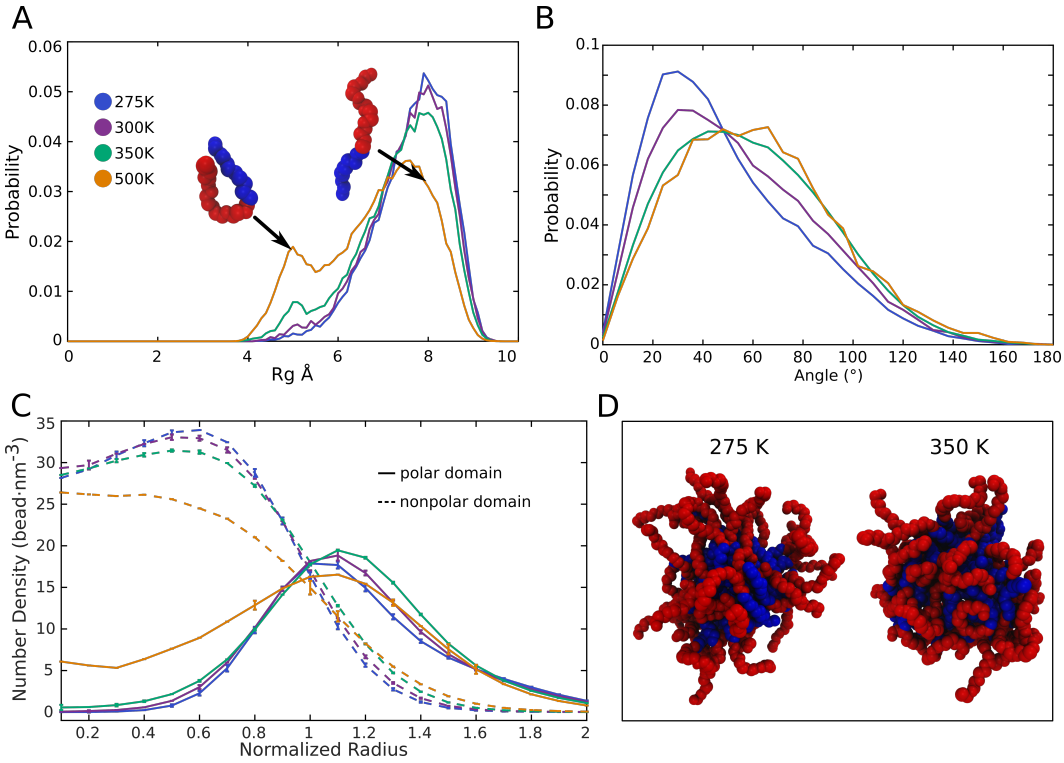

Figure 2.6: Properties of surfactant and micelles at varying temperatures. A) Distribution of radius of gyration for monomer surfactant. Example of surfactants with RG of 5 and 8 Å, shown with nonpolar domain in blue and polar domain in red. B) Distribution of angles between the polar domain of surfactants in the micelle shell and line between the center of the polar domain and the center of the micelle. C) Average density at varying positions within the micelle for the polar domain (solid line) and nonpolar domain (dashed line) beads. D) Example micelle structures from 275 K and 350 K simulations. Coloring as in A.

Changes in the distribution of tail and head group beads within the micelle at different temperatures are also observed. Figure 2.6C gives the density of beads belonging to the head and tails of surfactants within the micelle as a function of position within the micelle. As expected, beads from the nonpolar tails are found primarily in the core of the micelle, and polar heads found primarily in the outer shell of the micelle. We will refer to the inner portion of the micelle primarily composed of tail beads as the core, stretching out to a normalized distance of approximately 1.0 units from the center of the micelle. The region with a higher head group density we will refer to as the shell of the micelle, covering the region from 1.0 to the micelle edge at approximately 1.8-2.0 units from the center of the micelle. In the interface between the core and shell of the micelle, from 0.75-1.25 units, we find both tail and head group beads. At high temperatures, the micelle becomes more disordered, with the density of the polar head groups in the micelle core increasing. As temperatures decreases below 350 K, we observe a drop in density of the head groups in the outer shell of the micelle. This is consistent with the polar domains remaining extended. As indicated by Figure 2.6B, the polar head groups fold against the surface of the micelle at higher temperature, increasing their density near the micelle core. We also observe an increase in the density of the nonpolar core of the micelle at temperatures below 350 K. Combining these structural properties of the micelle gives us an overall picture of its organization. At 275 K the extended conformation indicated by the radius of gyration, the alignment of the polar head groups with the radius of the micelle, and the decreased density of polar headgroups in outer shell of the micelle indicates that, while the micelle has a defined hydrophobic core at 275

K, the surfactants extend straight out from center the micelle and leave much of the core exposed. Figure 2.6D contains example micelle structures from 275 and 350 K, which illustrate the difference in organization. While there are still some extended surfactants in the micelle at 350 K, the folding of polar head groups against the surface reduces exposure of the micelle core. Additionally, the reduction in contacts between surfactants in the outer shell of the micelle at 275 K will decrease the energetic penalty for individual surfactants to leave the micelle.

2.4.0.2 Dynamics of surfactants leaving the micelle

We further investigate the kinetics of the process in which an individual surfactant leaves the micelle by looking at contacts between the micelle and surfactants that are transitioning from the micelle to the solvent. Time series for mean contacts during a transition from the micelle to solvent, centered around the time when the surfactant leaves the micelle, are shown in Figure 2.7. Each plot classifies contacts based on the contact type, with contacts between the head of the transitioning surfactant and heads of the micelle (head-head) shown in Figure 2.7A, and between the head of the transitioning surfactant and tails of the micelle (head-tail) shown in Figure 2.7B. For clarity, only data from 275 and 350 K is shown. Transition contacts for 300 K can be found in Figure 2.8, and generally do not show the large deviations from 350 K observed at 275 K. Results for transitions at 500 K are not provided due to the short-lived nature of the micelle at that temperature. Prior to the surfactant leaving the micelle, head-head contacts (Figure 2.7A) are reduced at

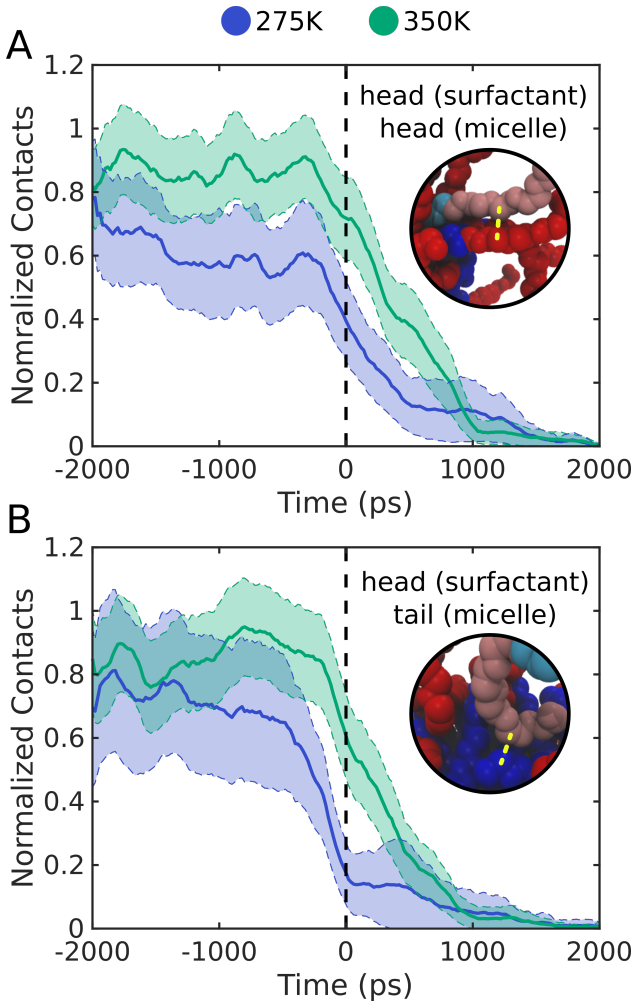

Figure 2.7: Time series for contacts between surfactants leaving the micelle and surfactants within the micelle. Contacts between A) polar head of leaving surfactant and polar heads in micelle and B) polar head and nonpolar tails. Inset images illustrate example contacts with micelle tail in blue, leaving surfactant tail in cyan, micelle head in red, leaving surfactant head in pink, and example contacts indicated by yellow dashed lines. Values normalized based on the maximum observed at 350 K. All data smoothed using a 500 ps window. 95% confidence intervals shown as shaded areas.

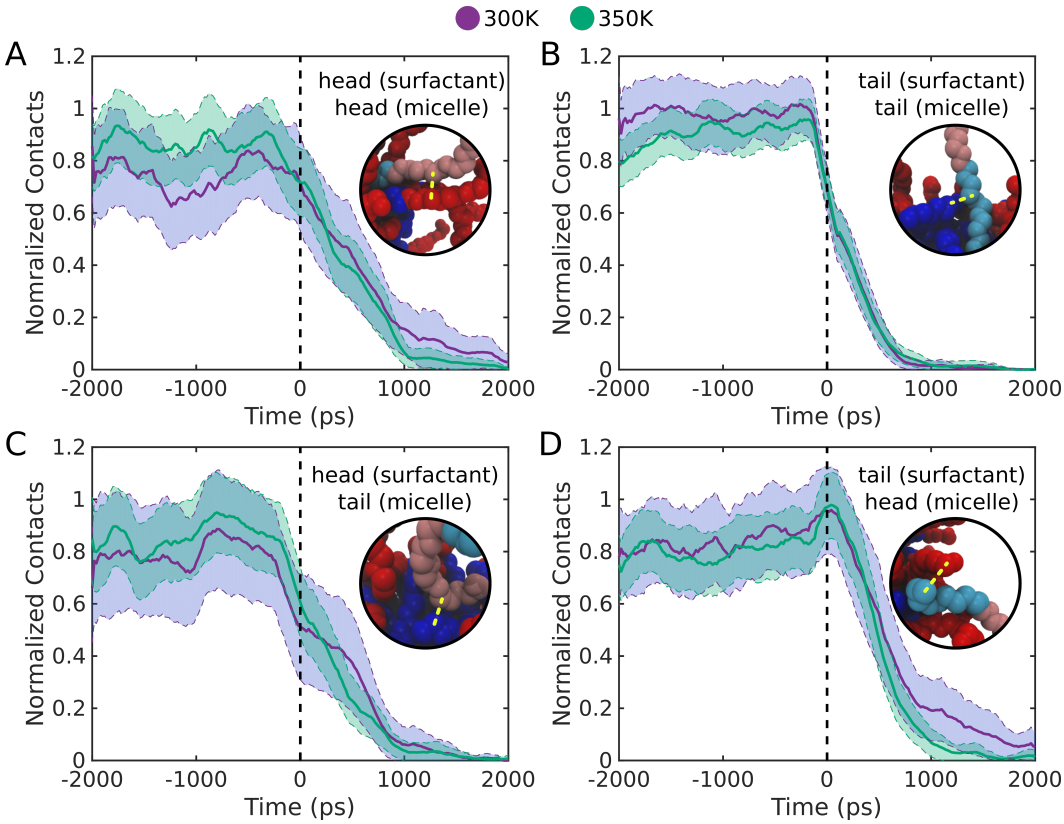

Figure 2.8: Time series for contacts between surfactants leaving the micelle and surfactants within the micelle. Contacts between A) polar head of leaving surfactant and polar heads in micelle, B) polar head and nonpolar tails, C) nonpolar tail and nonpolar tails, and D) nonpolar tail and polar heads. Inset images illustrate example contacts with micelle tail in blue, leaving surfactant tail in cyan, micelle head in red, leaving surfactant head in pink, and example contacts indicated by yellow dashed lines. All data smoothed using a 500 ps window. 95% confidence intervals shown as shaded areas.

275 K relative to 350 K. These contacts begin dropping rapidly at all temperatures slightly before the surfactant leaves the micelle, and are significantly fewer at 275 K than 350 K during the transition. Even before the head-head contacts begin to drop, around time 0 ps, they are reduced at 275 K relative to 350 K, an effect of the reduced density of the head groups in the outer shell of the micelle observed in Figure 2.6C. We also observe a large difference in the time series for head-tail contacts. At 275 K these contacts drop almost to their minimum value before the surfactant is considered to have left the micelle, while at 350 K many of these contacts are still present during the transition. Again, these contacts are reduced at 275 K relative to 350 K before the transition. As the tails of surfactants are in the micelle core and head groups primarily in the outer shell, contacts with the head groups prior to the transition are increased at 350 K due to the folding of the head groups against the core at 350 K, shown in Figure 2.6B. Time series for tail-tail and tail-head contacts, found in Figure 2.9, show little difference between 275 and 350 K. Tail-tail contacts would be expected to break first as the surfactant transitions, so it is no surprise that they change similarly at each temperature during a transition. Contacts between the transitioning surfactant tail and micelle heads would be expected to be found primarily as surfactant passes out of the shell of the micelle, so it appears that at this point in the transition there is little difference with temperature. The differences in head-head and head-tail contacts give some insight into the stability of the micelle at 350 K. More favorable head-head contacts need to be broken for the micelle to escape at 350 K. At 275 K the quick breaking of the head-tail contacts would be expected if the surfactant moves straight out of the

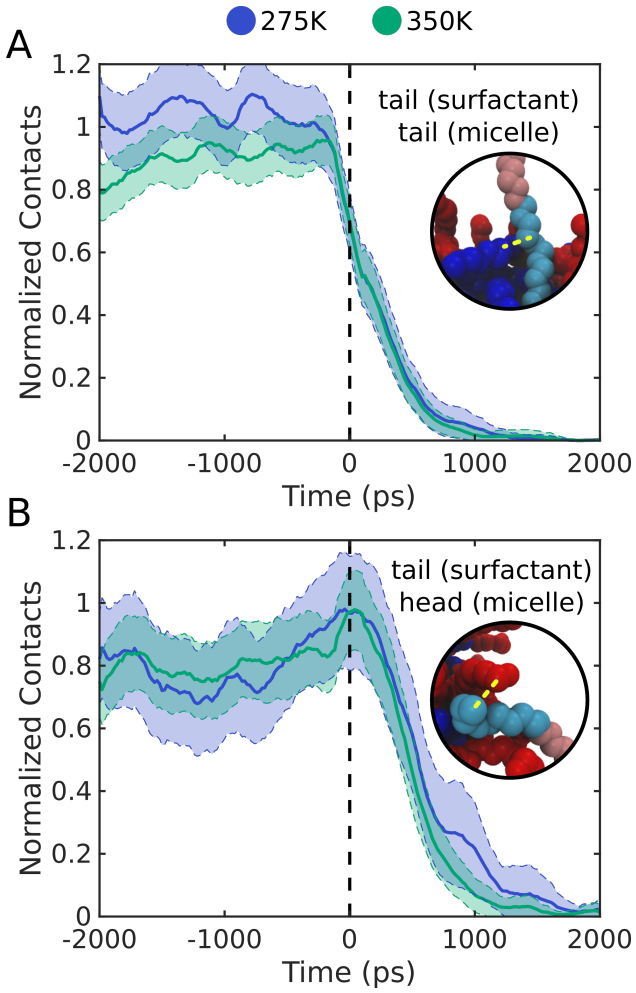

Figure 2.9: Time series for contacts between surfactants leaving the micelle and surfactants within the micelle. Contacts between A) Nonpolar tail of leaving surfactant and nonpolar tails in micelle and B) nonpolar tail and polar heads. Inset images illustrate example contacts with micelle tail in blue, leaving surfactant tail in cyan, micelle head in red, leaving surfactant head in pink, and example contacts indicated by yellow dashed lines. All data smoothed using a 500 ps window. 95% confidence intervals shown as shaded areas.

micelle during a tranistion. The lingering head-tail contacts at 350 K indicate that the contacts between the head groups folded against the micelle and the core of the micelle are retained for some time during the transition, and the surfactant does not necessarily move straight away from the micelle. This provides an opportunity for the surfactant to remain with the micelle instead of transitioning out.

2.4.0.3 Changes in solvation of the micelle

The effects of temperature on micelle solvation were investigated by comparing the density of water at varying positions within the micelle for each temperature, shown in Figure 2.10A. Based on comparisons to Figure 2.6C, we can see that the inner core of the micelle, stretching out to a normalized distance of approximately 0.8-1.0 units, is essentially dry, with solvation beginning around the edge of the micelle core. Extending out from micelle's core, water density increases with distance until it reaches approximately bulk density at the outer edge of the micelle. This density profile is consistent with published experimental and computational results [41,42,46,85]. We observe a shift in solvation at a normalized distance of 1.0-1.6 from the center of the micelle, with a decrease in water density accompanying the increase in temperature. Dehydration of micelles with an increase in temperature has been observed both experimentally and computationally [41,42,46,85]. Our observations here agree with these results, as we observe a steady change in hydration with temperature, with the dehydration effects being most prominent nearer the interface between the nonpolar tails and polar head groups. To further characterize hydration

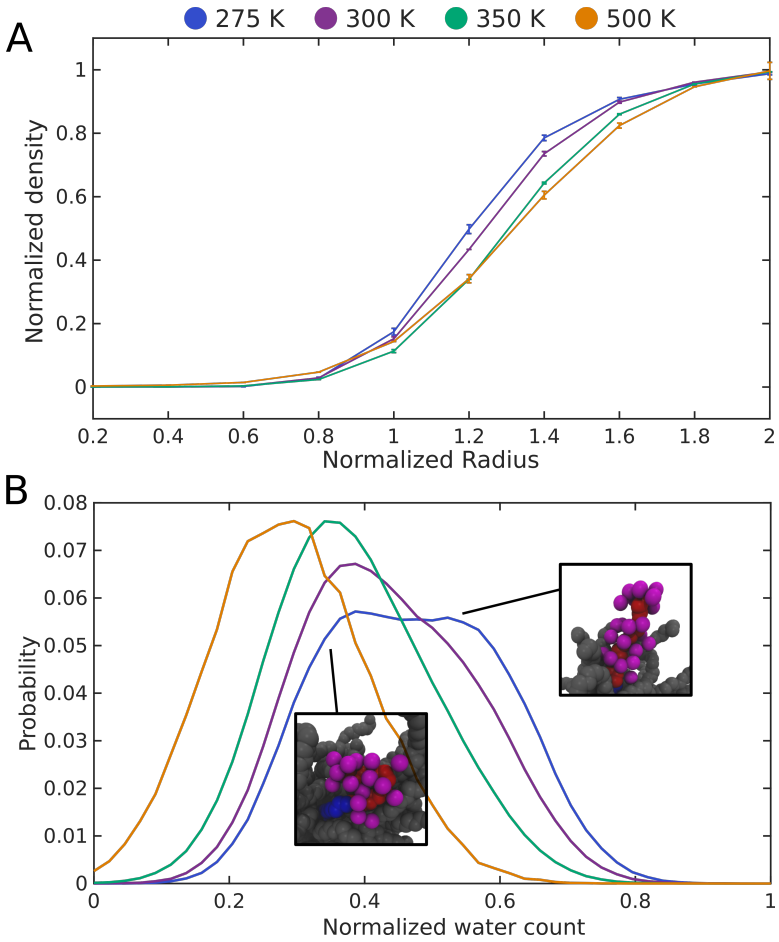

Figure 2.10: Number distribution of water contacts per surfactant with A) the polar domain of surfactants which are participating in the micelle core. Contacts used a cutoff distance of 4 Å. Number of water contacts normalized by the maximum number of contacts observed. B) Average water density at varying positions within the micelle.

in the micelle's shell region, we quantified the number of water contacts with the head group of each surfactant in the micelle. The distribution of these contacts at each temperature is shown in Figure 2.10B. At higher temperatures (350 and 500 K) we find a single peak at lower solvation, while at lower temperatures a second, higher solvation state of the head group arises. Structurally, the higher solvation peak at 275 K corresponds to a fully extended head group oriented away from the micelle core, while the lower peak in solvation corresponds to head groups which are laying against the surface of the micelle. As shown in Figure 2.6B, at higher temperature (350 and 500 K) the angular distribution of the head groups with the micelle radius is increased. We see here that this increase in angle corresponds with a decrease in solvation of the head group.

Deeper insights into the observed changes in surfactant solvation with temperature can be obtained from analysis of the solvation water order parameter and total potential energy (Figure 2.11). The total potential energy of water here refers to the pairwise interaction energy for each water bead summed over all interactions for that water bead. To remove any potential effects from nearby surfactants, the total potential energy and the order parameter of water in the first solvation shell around a monomer surfactant were calculated using simulations of a single surfactant in water, at 275 and 350 K. For the same reason, values from bulk water are taken from simulations of pure water at 275 and 350 K. At 350 K, we find higher total potential energy, shown in Figure 2.11A, for surfactant solvation water than bulk water, as might be expected. At 275 K, however, the total potential energy is very similar for solvation and bulk water. The tetrahedral order parameter of

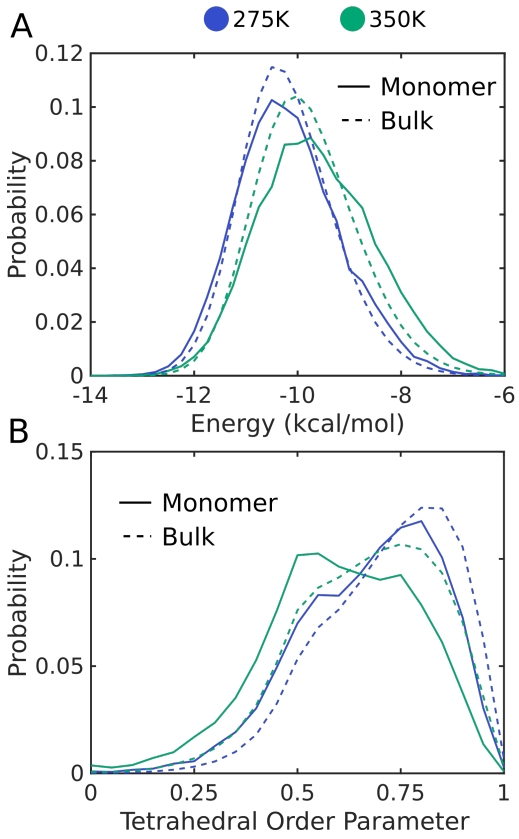

Figure 2.11: Properties of water around surfactant at 275K and 350K. A) Distribution of total potential energy for water beads in first solvation shell of micelle (solid line) and single surfactant (dashed line). B) Tetrahedral order parameter for water in first solvation shell around single surfactant (solid line) and bulk water (dashed line).

water in the first solvation shell around a single surfactant, shown in Figure 2.11B, was compared to the order parameter for bulk water and used to measure the effect of solvating surfactant on the order of water. At 350 K solvation water is significantly less ordered than bulk order. With the decrease in temperature to 275 K, both bulk and solvation water become more ordered, as expected. However, the difference in order between bulk and solvation water is less at 275 K than at 350 K, indicating that the order of solvation water has increased relative to bulk water. Accompanying this increase in order of the solvation water relative to bulk water is a decrease in the total potential energy of solvation water relative to bulk. The increase in order around monomer surfactant at low temperature indicates that solvation of the surfactant is less favorable entropically at low temperature, while the increase in total potential energy for solvation water indicates that solvation is more favorable enthalpically at low temperature. Thus the destabilization of the micelle at low temperature is driven by the enthalpic gain of surfactant solvation. This is in agreement with thermodynamic properties of micelle formation determined experimentally, which indicate that micelle formation is driven entropically at low temperature while micelle destabilization is driven enthalpically [86, 87]

2.5 Conclusion

In conclusion, we have developed a CG model of nonionic surfactant and used this model to characterize the destabilization of pre-formed micelles at extremes of temperature. Our model is capable of capturing both micelle formation and desta-

bilization, with simulations of formation and destabilization indicating the same temperature of maximum stability for micellization. We obtain micelles with the expected dry hydrophobic core surrounded by a solvated polar shell. At high temperatures we observe a rapid disordering and destabilization of the micelle. At low temperatures we note changes in both the overall structure of the micelle as well as the individual surfactants which make up the micelle. At 350 K the polar head groups of surfactant fold against the surface of the micelle, reducing the solvation of the hydrophobic core, while at low temperatures surfactants within the micelle remain extended, increasing the water density within the shell of the micelle as well as the solvation of hydrophobic groups. Surfactants that transition from the micelle to solvent at 275 K appear to move straight out of the micelle, rapidly losing head-head and head-tail contacts with the micelle. By contrast, surfactants leaving the micelle at 350 K have lingering head-head and head-tail contacts, which may contribute to the retention of surfactants by the micelle. Destabilization of the micelle at low temperature is enthalpically driven, as a decrease in enthalpy for solvated monomer surfactants at 275 K leads to a decrease in the hydrophobic effect, as solvation of monomer surfactant becomes more enthalpically favorable at 275 K relative to bulk than at 350 K.

In its current form, the model presented here is easily extensible to any surfactants of the form C_nE_m. In principle the model could be extended to any nonionic surfactant, though surfactants of another form may require additional parametrization. As the current form of the model lacks charged beads, it would require modification for the simulation of ionic surfactant. Additionally, the lack of representation

for the directionality of hydrogen bonds means that the model would not be capable of capturing a dipole moment. Future work with the model will involve extension to charged residues, as well as testing with more complex mesophases.

Chapter 3: Interplay between Conformational Heterogeneity and Hydration in the Folding Landscape of a Designed 3-helix Bundle

3.1 Overview

This chapter is based on the author's publication: Interplay between Conformational Heterogeneity and Hydration in the Folding Landscape of a Designed Three-Helix Bundle; Custer, G.S., Das, P., Matysiak, S; J. Phys. Chem. B, 2017, 121 (13) 2731-2738

3.2 Introduction

Proteins are the one of the most important types of macromolecules in biological systems. They consist of chains of amino acids which typically fold into a unique structure and have numerous potential uses, including catalysis of chemical reactions, transportation of other molecules, signaling within cells, and providing structure for cells. Due to their ability to be modified and the many different sequences that can be combined, there is great potential for the use of designed proteins in medicine and industry. However, the relationships between sequence, fold,

and function in proteins are not simple and have the subject of constant research for decades [88–91].

Water plays a fundamental role in determining the structure and function of proteins. The water molecules constituting the first hydration shell of a protein surface are of particular importance, as dry proteins lack biological activity. Hydration water is actively engaged in enzyme catalysis, protein-protein and protein-DNA interactions, and protein folding. Both experimental and computational studies have shown that hydration water is necessary to facilitate large motions within a protein [92–94]. Water expulsion and hydrophobic core formation are amongst the final steps of protein folding. However, buried waters are frequently found in protein structures [95–98]. In thermophilic proteins, increased internal water content has been associated with enhanced stability [99]. Nearly-native, functional states of biological proteins such as SH3 domain have been identified with a solvated hydrophobic core [100,101], though solvation of the core is often associated with denaturation [98].

The role of water in shaping the protein folding landscape has been investigated by using both experiments, such as x-ray [102], neutron diffraction [102], NMR [97, 103], and femtosecond fluorescence [104], as well as by computational methods [99, 101, 105–113]. A practical difficulty of studying the role of water in protein folding by fully atomistic MD is the large computational cost associated. Even for small proteins, such as the 35-residue villin headpiece, folding times are frequently longer than 1 μs [114]. An alternative approach is to employ coarse-grained models that consider the effect of water on the folding landscape in an implicit manner, such as the Generalized-Born [11], Gō model [12], HP [13], BLN [14] and other

protein models [15–18]. These models have been instrumental in studying protein folding, but the lack of explicit details representing protein-water interactions make it difficult to use them to study the role of solvent in folding. A number of CG water models [19, 20] group multiple water molecules into a single interaction site which, while reducing computational requirements, makes it difficult to incorporate hydrogen bonding, examine interactions between water and protein in detail, and accurately capture the water phase diagram.

To circumvent these issues, in the present study we have developed an off-lattice coarse-grained model of a solvated protein which makes use of a water model in which each water molecule is represented with a single interaction center or "bead". The water model used is known as the mW water model [26], and uses the Stillinger-Weber (SW) potential to enforce a tetrahedral angle between neighboring water molecules. This model is thus able to accurately emulate the hydrogen bond network present in water and is capable of reproducing the water phase diagram [26]. The protein model used here is comparable in the method of coarse-graining to the BLN, MARTINI, and other protein models [14, 15, 115–119]. As in the BLN model, we make use of three bead types for the protein, but each residue is split into a backbone and sidechain bead, as with other two-bead models [15,115,117–119]. Backbone dipole interactions have been shown to drive formation of secondary structure [120], but to reduce computational requirements CG models frequently do not include explicity hydrogen bonding or dipoles. Here we adopt the use of the three-body SW potential to accounting for hydrogen bonding, as in the mW water model [26]. In previous studies, we have developed interaction parameters of hydrophobic ho-

mopolymer and hydrophobic-polar heteropolymer, which were used in conjunction with the mW water model to study the pressure-temperature dependence of the stability landscape [62,63]. In the current study, we have adopted a similar strategy to investigate the interplay between the conformational heterogeneity and explicit hydration of a designed three-helix bundle protein.

Helix bundles are one of the simplest protein folds found in nature, making them frequent targets for studies of protein folding. They can be formed from very few residues, with the villin headpiece 3-helix bundle folding in solution with as few as 35 residues [7], and are among the fastest folding structures [114]. Folding of several 3-helix bundles has been characterized both experimentally and computationally, including the villin headpiece [7,121,122], fragment B of protein A [123,124], and albumin binding domain [9,125]. These bundles typically fold to a single native state, with hydrophobic residues oriented inwards [9, 122, 124], but helical bundles with more flexibility have been observed [126,127]. As helical bundles are relatively simple folds found commonly in nature, they represent a useful starting point for models of protein folding. The goal of this study is to characterize the the role of water in folding of a designed helical bundle by simulating the solvated protein with our CG model.

3.3 Methods

3.3.1 Model

In the present study, we have developed a CG model similar in its treatment of polar interactions to the CG surfactant model presented in Chapter 2. In this model we have represented each amino acid within a helix using two beads, one for backbone and one for sidechain (see Figure 3.1A). Residues connecting helices (i.e turn residues) are represented with a single backbone bead without sidechain (see Figure 3.1A). Four different bead types are used in our model, backbone (BB), hydrophobic (HB), neutral (NB), and water (mW). Bond and angle interactions between the beads are represented using simple harmonic potentials. Values used for all bond and angle parameters in our system can be found in Table 3.1. The size and mass for each bead type are based on their atomistic counterparts: BB is based on backbone atoms of protein, HB is based on the sidechain of isoleucine, and the neutral bead is based on the sidechain of alanine. Bonds between the backbone beads have a length equal to the approximate distance between adjacent C_α atoms in a protein, *i.e.* 3.8 Å [128]. Bonds between backbone and hydrophobic beads are set to the average distance between the C_α carbon and the center of mass of the sidechain atoms in isoleucine, *i.e.* 1.97 Å. This distance was calculated from a set of non-redundant structures taken from the Protein Data Bank using the Pisces server [129]. Bonds between backbone and neutral beads are set to the approximate length of the C_α-C_β bond in alanine, 1.53 Å [130]. Angles between backbone beads

Table 3.1: Bonded parameters

Bonds		
Bead types	r_0 (Å)	K (kcal/Å^2)
BB-BB	3.80	200.0
HB-BB	1.97	200.0
NB-BB	1.53	200.0
Angles		
Bead types	θ_0 (°)	K (kcal/rad^2)
BB-BB-BB	105	4.0
BB-BB-HB	123	4.0
BB-BB-NB	123	4.0
Dihedrals		
Bead types	A (kcal)	B (kcal)	C (kcal)
BB-BB-BB-BB	0.31	0.920471	1.037159

Table 3.2: Bonded parameters

are set to the expected angle between C_α atoms of adjacent residues, 105° [131], and angles between backbone and sidechain beads set to the expected angle for C_α-C_α-C_β, 123° [132].

The three residue types, shown in Figure 3.1A, used in this protein model were hydrophobic (H), neutral (N), and turn (T). A dihedral potential with three minima was used to penalize unrealistic dihedral angle formation along the backbone. This dihedral potential has been applied to the backbone dihedrals in order to bias the structure towards sheet, right-hand helix, or left-hand helix [133]. In this potential, shown in Figure 3.1B, the minima corresponding to right-hand helix and sheet are set at equal depth and the minimum for left-hand helix is given a higher energy.

The dihedral potential is applied across bonds between backbone beads of H and N residues in our protein, but not across the backbone beads of T residues. Parameters used for the dihedral potential can be found in Table 3.1. Non-bonded interactions involving hydrophobic or neutral beads are represented with a 9-6 Lennard-Jones potential, with parameters for each possible bead pair shown in Table 3.3.1. The SW potential was used for non-bonded interactions between backbone beads and mW water beads, in a similar manner as with the CG surfactant model. This SW potential has been successfully used in the past to emulate intra-polymer, polymer-water, and water-water hydrogen bonding [26, 62, 63]. The previous polymer models which made use of the SW potential used one to two bead types (fully hydrophobic to hydrophobic-polar) in simulating a linear polymer in explicit water [62, 63]. Here we have expanded the model to four total bead types, with three beads used for the protein with sidechains. The three-body angular penalty present in the SW potential acts in this context to impose the hydrogen bond networks naturally found in water [26]. For the interactions represented using the SW potential, only the parameters controlling the potential well depth (ϵ), bead size (σ), and target angle (θ) were modified (see Table 3.3.1). All other parameters in the SW potential were set as found in Molinero2009 and Moore [26]. For non-bonded interactions between backbone and water, θ is set to 109.5°, allowing backbone beads to participate in the hydrogen bond network with water. While the previous polymer models, as well as the CG surfactant model described in Chapter 2 used a θ of 109.5° for all polar interactions [63], here we need to distinguish between polar interactions within the backbone and between backbone and water because polar interactions

Lennard-Jones			
Bead types	ϵ (kcal/mol)	σ (Å)	
HB-HB	1.90	6.78	
HB-BB	0.05	6.14	
HB-NB	0.05	5.09	
HB-mW	0.05	4.59	
NB-NB	0.05	3.40	
NB-BB	0.05	4.45	
NB-mW	0.05	3.95	
Stillinger-Weber			
Bead types	ϵ (kcal/mol)	σ (Å)	θ (°)
mW-mW[a]	6.189	2.3925	109.5
mW-BB	7.86	3.43	109.5
BB-BB	9.70	4.46	180

[a] Parameters for mW bead taken from Molinero2009, et al. [26]

Table 3.3: Non-bonded parameters

within the backbone do not form a tetrahedral network. In an α-helix the angle between nearest nonbonded C_α atoms should be close to 180°. Thus, to account for the hydrogen bonding within an α-helix, we set θ between backbone beads to 180°.

3.3.2 Simulation Setup

Molecular dynamics simulations were performed using the model described above to simulate a protein sequence designed to fold into a 3-helix bundle, an example of which is shown in Figure 3.1C. This protein sequence consists of three identical segments of N and H residues alternating in a pattern known to fold into a helix [134, 135]. These helical segments are connected with a turn consisting of 4 T residues. We use a 12-residue sequence derived from West, *et al.* for each

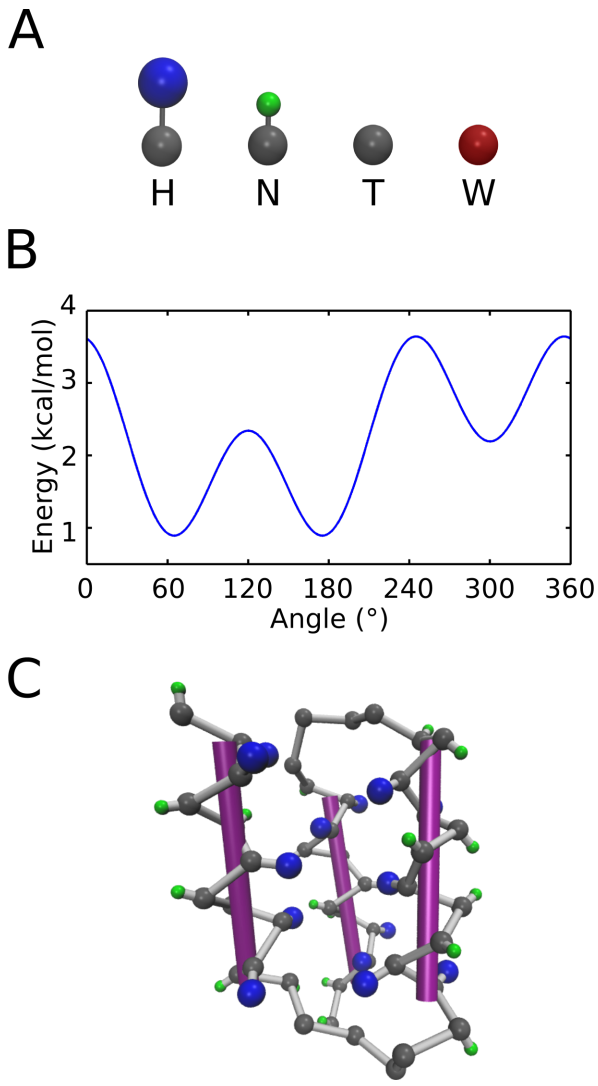

Figure 3.1: A) Residue types in the model shown in CPK representation, where H is hydrophobic, N is neutral, T is turn, and W is water. Backbone beads shown in gray, hydrophobic in blue, and neutral in green. B) The dihedral potential applied to the helical portion of the backbone. C) 3-helix bundle shown in CPK representation, with rods added to represent the helices.

helix [135]: NNHHNNHNNHHN (referred to as HX). This sequence aligns the hydrophobic H residues along one side of the helix and caps either end of the helix with neutral N residues. Combined with the turns, the full sequence used is: HX-TTTT-HX-TTTT-HX. Simulations were performed with LAMMPS [73] using the Verlet integrator and a timestep of 8 fs. All simulations begin with the protein in an extended random coil configuration, as illustrated in the top left of Figure 3.2. The protein was placed in a cubic box containing 16985 mW water beads. Velocities were initialized using a Boltzmann distribution set at the temperature being used for the simulation. Temperatures and pressures were maintained using the Nosé-Hoover thermostat/barostat [76, 77] with temperature at 300 K and pressure at 1 atm. Damping time was 100 fs for temperature coupling and 1 ps for pressure coupling. Unless otherwise stated, all results reported in this manuscript are estimated from three different simulations performed at 300 K, each starting from a different extended random coil protein structure and run for 300 ns.

3.3.3 Identification of 3-helix bundle

The folding landscape of the three-helix bundle is quantified using the following two reaction coordinates: a helicity score (F_H) and an inter-helix alignment score (F_A). The helicity score, F_H, is calculated as:

$$F_H = \left(1 - \frac{RMSD_1}{RMSD_{Max}}\right)\left(1 - \frac{RMSD_2}{RMSD_{Max}}\right)\left(1 - \frac{RMSD_3}{RMSD_{Max}}\right) \tag{3.1}$$

where $RMSD_N$ refers to the root-mean squared deviation (RMSD) of helix segment N (1, 2, or 3) from an ideal helix, and $RMSD_{Max}$ refers to the largest observed

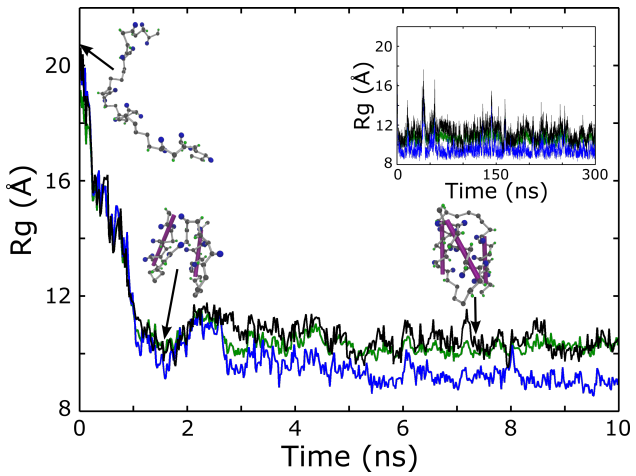

Figure 3.2: Radius of gyration over time of the whole protein (black), hydrophobic beads (blue), and neutral beads (green) for the first 10 ns of a simulation at 300 K. Data for the complete 300 ns long simulation shown in inset. Structures are colored as in Figure 3.1. Structures sampled at different times during the trajectory are shown with arrows indicating their associated time point.

RMSD from an ideal helix. $RMSD_{Max}$ was taken from a simulation performed at 400 K, allowing sufficient unfolding of the protein. As a result, Equation 3.1 accounts for the deviation of normalized RMSD for each helix ranging from 0 to 1, where a value equal to 1 represents a perfect helix. By multiplying these RMSD deviations for the individual helices with each other (Equation 3.1) we obtain a single metric that does not downplay deviations from helicity in any individual helix. This score, F_H allows us to measure the secondary structure of our protein. The second reaction coordinate is the inter-helix alignment score, F_A, given by:

$$F_A = \left(1 - \frac{\theta_{12}}{180^\circ}\right)\left(1 - \frac{\theta_{13}}{180^\circ}\right)\left(1 - \frac{\theta_{23}}{180^\circ}\right) \tag{3.2}$$

where θ_{NM} refers to the angle between vectors aligned to helices N (1 or 2) and M (2 or 3). This equation produces a normalized angle for each pair of helices which

ranges from 0 to 1, where 1 indicates that the helices are perfectly aligned (with an inter-helix angle of 0). In an ideal three-helix bundle all of the helices would be perfectly aligned against each other. Again, multiplying the inter-helix alignment values for the three possible pairs together allows us to obtain a single metric. This metric provides us an indirect measure of the tertiary structure of the helix bundle, though it is not a direct measure of the fit to ideal tertiary structure. We have extensively tested both of these reaction coordinates to characterize an ideal helix bundle and found that $F_H \geq 0.8$ and $F_A \geq 0.76$ are needed for a structure to be considered a 3-helix bundle.

3.3.4 Quantification of waters in the core

We define the core of our protein as the space within a convex hull surrounding the backbone beads in the three helices. After calculating the convex hull, we consider all waters within the hull to be within the core of the protein. To determine the number density of water in the core, we divide the number of waters in the core by the volume of the convex hull defining the core. Density is then normalized by rescaling the raw values such that 99.8 percent of water densities for conformations near ideal helix ($F_H \geq 0.65$ and $F_A \geq 0.76$) fall between 0 and 1. For all plots shown with normalized water density, a normalized density of 0 corresponds to a water concentration of 6.64 M and a normalized density of 1 corresponds to a concentration of 25.54 M, with values between 0 and 1 scaled linearly. The full distribution of normalized and non-normalized water densities for helical conformations is shown

in Figure 3.3. A convex hull around backbone beads works well for defining the core when the peptide is in a compact conformation but not when the peptide is in an extended conformation. As both the dry and wet states observed in our simulations are compact, we are able to use this method to quantify the hydration state of the protein core.

3.4 Results and Discussion

3.4.0.1 Folding landscape of a designed three-helix bundle

Collapse of the protein is evident from the evolution of radius of gyration (Rg) of the protein, as shown in Figure 3.2. An initial drop in Rg for both hydrophobic and neutral beads is seen, followed by a further drop in Rg of the hydrophobic beads. In all runs at 300 K, the protein typically remains collapsed with hydrophobic groups buried to some extent, even when the protein is in an extended conformation (Figure 3.2 inset). Figure 3.4A illustrates the potential of mean force at 300 K as a function of F_A and F_H. Two minima are observed at regions with high values of F_A and F_H. The minimum labeled S1 includes the fully folded bundle, an example of which is shown in Figure 3.4B. Minimum S2 corresponds to structures in which the helical strands are aligned, but helicity is lacking, typically on one of the helices. The unfolded region of the conformational space will be referred to as S3, and contains many different conformations. Two primary folding pathways, shown in Figure 3.4B, are observed at 300 K. In the S3-S2-S1 pathway, shown in green in Figure 3.4B, the helical segments align against each other prior to the complete

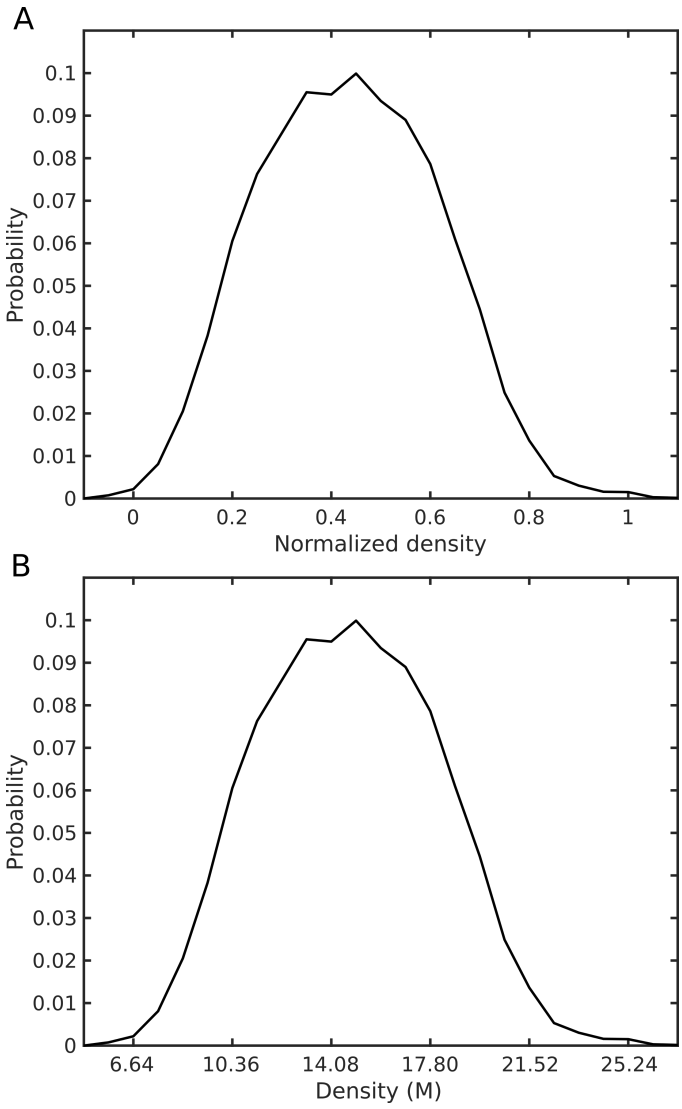

Figure 3.3: Distribution of water density in the core of the protein in state S1. A) Water density normalized such that 99.8 percent of water density values in S1 and S2 fall between 0 and 1. B) Water density before normalization.

formation of helices. In the S3-S1 pathway, shown in blue in Figure 3.4B, helices form fully as or before the helices align. The S3-S2-S1 pathway is observed most frequently in our simulations, implying inter-helix tertiary interactions are needed to stabilize individual helices. In both of these pathways, some helical content is present at an early stage of folding, indicating that folding occurs in a manner consistent with the framework or diffusion-collision models of folding, which would expect secondary structure to form before arrangement into the native tertiary structure [136–140]. However, the dominant presence of the collapsed protein with buried hydrophobic residues, even in the unfolded state, is consistent with the hydrophobic collapse model, in which the hydrophobic collapse of the protein reduces the conformational space sampled [140–144]. It appears this three-helix bundle folds in a manner incorporating characteristics of each model.

3.4.0.2 Shifts in solvation with protein folding

As the protein folds, accompanied by hydrophobic core formation, we would expect water expulsion from the protein core as a critical step. Figure 3.5A depicts the PMF plot shown in Figure 3.4A, with contour lines instead of color indicating the free energy of each state. Figure 3.5A is scaled to focus on S1 and S2, and is colored according to the corresponding average water density in the core of the protein. Comparison of the density of water corresponding to each state of the protein suggests a slight drying of the protein as it folds from S2 to S1. The relatively dehydrated nature of the S1 structures compared to S2 is also evident from the water

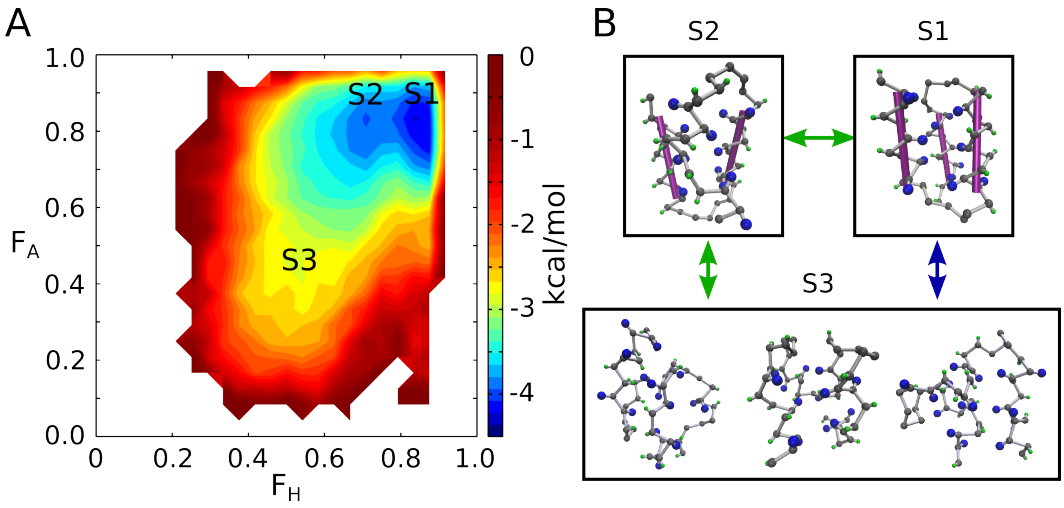

Figure 3.4: A) Potential of mean force for 3-helix bundle simulation at 300 K. states are numbered. F_A refers to the inter-helix alignment score, and F_H to the helicity score. B) Structures corresponding to each numbered minima from the plot in A, with arrows denoting possible pathways between minima. Arrow colors distinguish the two possible pathways for moving from S3 to S1. Protein drawn in CPK representation with backbone beads colored gray, hydrophobic beads blue, and neutral beads green. Purple rods have been added to highlight formed helices.

density distribution of the two states (Figure 3.5B), which shows a shift to decreased water density as the protein moves into S1. However, the water density distributions also show that there is a considerable population of structures with wet core in both S1 and S2 states. The fraction of protein structures with a normalized water density above 0.6 is identical in both states. This finding suggests that the folded protein populates structures with a wide range of core solvation. The presence of internal water revealed in this study is consistent with what has been seen in thermophilic proteins [99]. A near-native state with a solvated core, similar to what we observe with S2, has also been seen for a SH3 domain protein in NMR [100,101]. Thus while we typically expect drying of the core to accompany folding of a protein, complete drying of the core is not a necessity.

3.4.0.3 Solvation of folded substates

To explain this broad range of solvation in the folded structure, we examined the water density in the protein core over time. A sample of this time series for one trajectory is shown in Figure 3.6A (full time sequence for each simulation can be found in Figure 3.7). Within S1, transitions between wet and dry core are often observed. Similar nano-scale dewetting transitions have also been observed at the inter-peptide interface [145–147] as well as within the hydrophobic core of biological proteins [110, 148]. Associated changes in the packing of the protein core are also analyzed. We found that, within state S1, rotation of an individual helix around its central axis can occur, such that sometimes the hydrophobic face of the helix

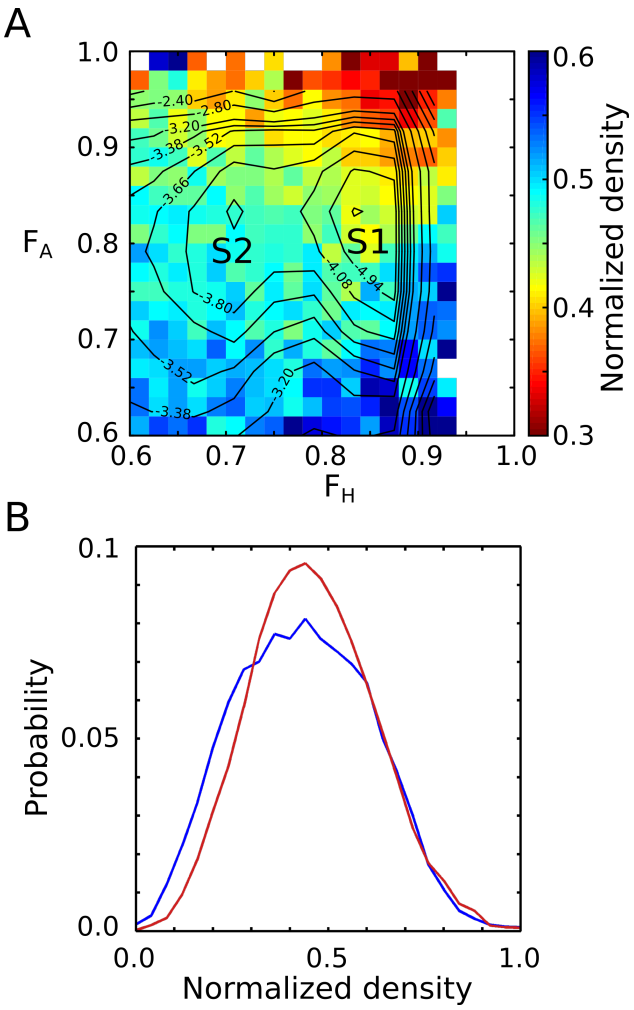

Figure 3.5: A) Upper right portion of the PMF plot shown in Figure 3.4, colored by average normalized density of water in the core of the protein. Minima from Figure 3.4 labeled. B) Distribution of water density in the core of the protein for S1 (blue) and S2 (red). Water is normalized such that 99.8 percent of water density values in states S1 and S2 fall between 0 and 1.

was oriented inwards and at other times the neutral face was oriented inwards. If we consider each helix to have two primary orientations, the hydrophobic face in or out, then there are four possible configurations for the folded bundle: 3 hydrophobic faces in (3-in), 2 hydrophobic faces in (2-in), 1 hydrophobic face in (1-in), and no hydrophobic faces in. The distributions of these configurations for the folded state (S1) were: 7.1 ± 2.0% 1-in, 55.4 ± 7.0% 2-in, and 37.4 ± 7.4% 3-in. The configuration with no hydrophobic domains in was very rare, appearing in only 0.1% structures. The time series in Figure 3.6A is shaded according to the number of helices in, and shows that transitions between 2-in and 3-in are also associated with a shift in solvation of the protein core. Distributions of water density for 1-in, 2-in, and 3-in (Figure 3.6B) show that the 1-in and 2-in configurations have considerably lower water density in the core than 3-in. Therefore, there appear to be distinct wet and dry states for the protein. The 1-in and 2-in density distributions for S1 (Figure 3.6B) are nearly identical, indicating that only one of the three helices needs rotate such that the hydrophobic face is outwards (2-in) to form a drier core, as no further reduction in solvation is seen with the 1-in conformation. Thus, the folded ensemble of the designed three-helix bundle sequence exhibits a diversity of structures in which the solvation state of the hydrophobic core is strongly correlated with the helix rotation. Both designed and biological helical bundles that require rotation of the helices about their central axes for activity are known to exist. For example, the transmembrane HAMP domain is a 4-helix bundle that exhibits rotation of helices about their axes [126]. Another 4-helix bundle, the artificial oxygen transport protein HP7, requires rotation of the helices for heme binding [127].

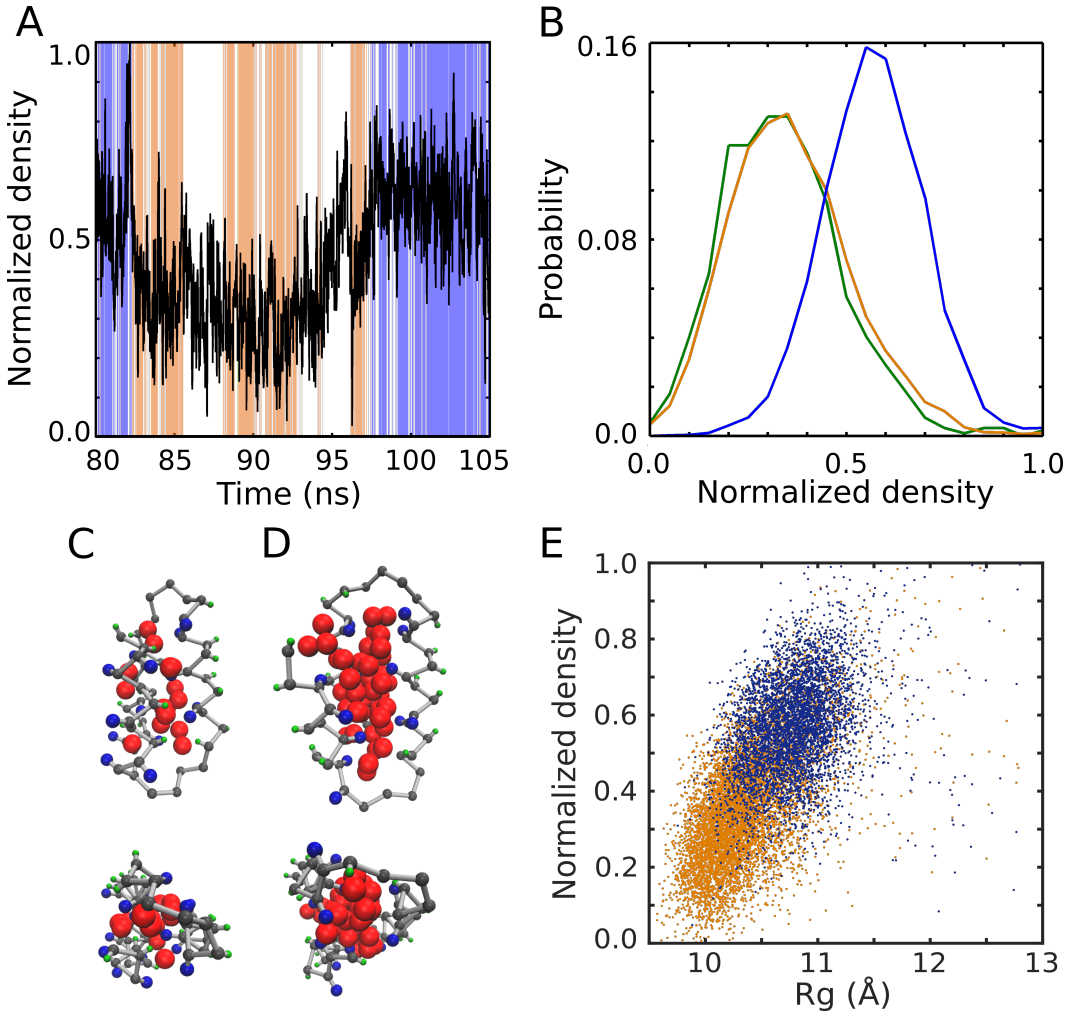

Figure 3.6: A) Normalized density of water in the core of the protein vs time for a portion of the trajectory at 300 K. Slices shaded by number of hydrophobic domains oriented inwards, with 2-in blue and 3-in orange. Unshaded slices not in S1. B) Distribution of water density in the core of the protein for 1 helix oriented with hydrophobic inwards (green), 2-in (orange), 3-in (blue). Water density in A and B normalized such that 99.8 percent of water density values in S1 and S2 fall between 0 and 1. C) Side and top views of protein in state 1 with 2 helices oriented with hydrophobic inwards and D) with 3 helices oriented inwards. Protein bead colors for C and D are as in Figure 3.1. Water beads (red) are those identified as within the core of the protein. E) Normalized density of water in the protein core plotted against radius of gyration for structures with 2 hydrophobic domains in (orange) and 3 hydrophobic domains (blue).

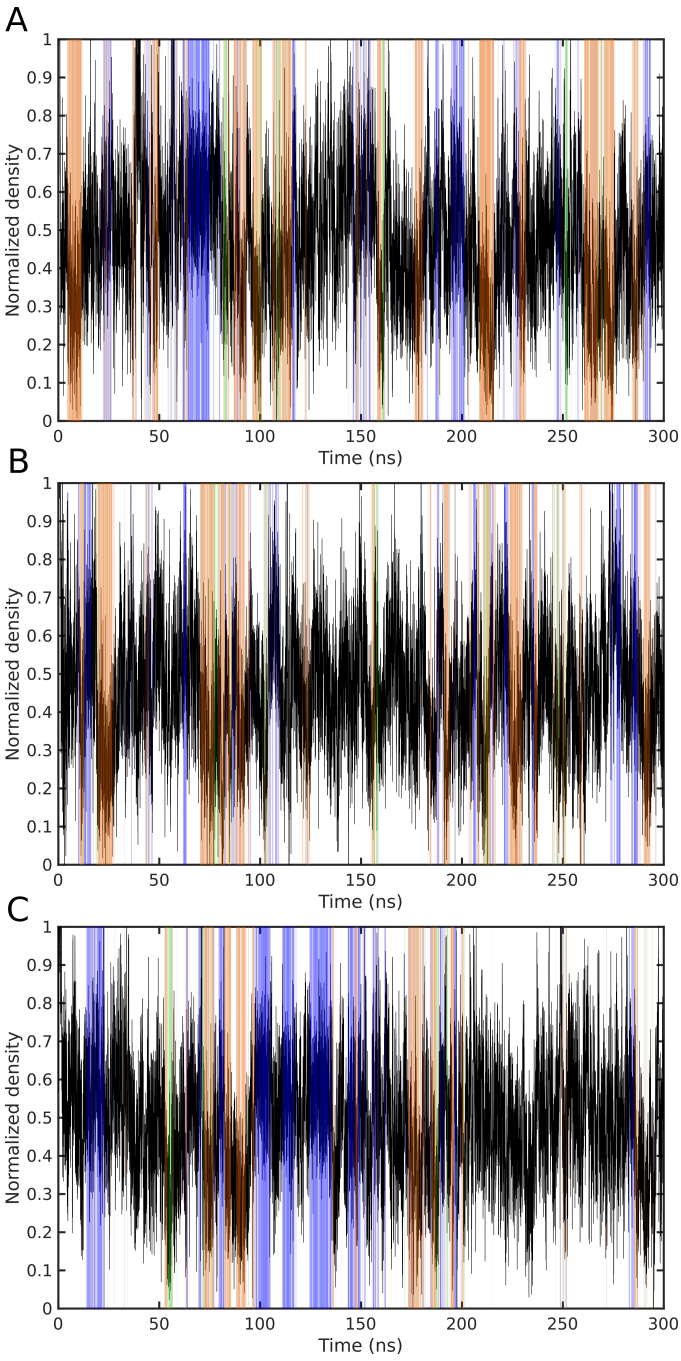

Figure 3.7: A), B), and C) Normalized density of water in the core of the protein vs time for three different trajectories at 300 K. Slices shaded by number of hydrophobic domains oriented inwards, with 1-in green, 2-in blue, and 3-in orange. Unshaded slices not in S1.

Structurally, the 2-in configuration (Figure 3.6C) appears to have much tighter packing than that found in 3-in (Figure 3.6D). Based on observed structures, the protein appears to be unable to pack all of the hydrophobic residues in the core without some swelling and influx of water, as a result of the large size of the residues. It is possible that the difficulty in burying the hydrophobic residues without some influx of water is due to the simplistic spherical representation of the sidechains. In a fully atomistic representation of the isoleucine residue, on which the hydrophobic beads are based, there would be flexibility within the sidechains which could allow the residues to pack against each other more efficiently. However, there are larger, less flexible nonpolar sidechains such as phenylalanine and tryptophan that might also cause the spatial arrangement issues observed with our CG hydrophobic beads. The radius of gyration (Rg) of the protein plotted against water density in the core (Figure 3.6E) agrees with this observation. We found a strong relationship between water in the core and Rg for both the 2-in and 3-in structures, with Rg increasing with solvation. Overall, the Rg for the 2-in structure is less than that of 3-in. Taken together, our results imply that while the sequence explored in this study does fold to a three-helix bundle, it is not the three-helix bundle that might be expected. Instead of a dry three-helix bundle with fully-buried hydrophobic residues, the protein model folds into two molten globule [149–151] states, one with a dry core and 2-in and a second with a wet core and 3-in. The existence of similar dry and wet molten globules on the folding landscape of real proteins has been suggested by both experiments and simulations [150–156].

We can further explore the formation of the 2-in and 3-in structures by exam-

ining the changes in solvation of the protein as it folds. In the S2 state we observe a generally folded but fully solvated structure. Some water expulsion occurs in the transition from S2 to S1, regardless of whether the protein folds to the 3-in or 2-in structures, which may be due to the disorder found in one helix for the S2 state. No preference is seen for 2-in or 3-in when transitioning from S2 to S1, indicating that the protein does not need to pass through one conformation to reach the other. The 2-in structure appears in 57.9% of transitions, and 3-in in 35.2% of transitions, very similar to the final observed fractions of 2-in and 3-in in state S1. Conversion from 3-in to 2-in is accompanied by an expulsion of water, and 2-in to 3-in with in influx of water. To our knowledge, this coupling of water content in the core to rotation of the helices has not been observed previously. The result implies an interesting balance of hydrophobic burial and solvation, as the 3-in structure shows increased hydrophobic burial, while the 2-in structure exhibits an overall reduction in solvent-accessible surface area and core hydration. This alternating population of two distinct states is similar to that seen in the HP7 protein designed by Koder, *et al*, in which rotation of the helices buried charged residues, but also resulted in a conformation conducive to heme binding [127]. Packing of hydrophobic residues in the core of a helical bundle is known to play a large role in its overall stability. In the design of a more stable 4-helix Rop bundle [157], as well as another designed 4-helix bundle [158], alternating large and small hydrophobic residues were found to improve helical stability. Stability of a single substate in our structure might also be improved by the inclusion of interacting charged or polar residues on the surface of the helix, which are common in natural and designed helical bundles [89]

3.5 Conclusion

In summary, we have developed and used a CG model of solvated protein to directly explore the folding/unfolding landscape of a three-helix bundle protein. This model is able to capture the hydrophobic collapse, helix formation, and folding into a bundle as commonly seen in biological 3-helix bundle sequences. The three-helix bundle which forms is not the expected bundle with fully-buried hydrophobic residues and a dry core. Instead, the designed sequence populates two molten globule states, one with a dry core and another with a wet core. In the dry molten globule state, one of the helices exposes its hydrophobic face to the solvent. In contrast, all three helices have hydrophobic faces oriented inwards in the wet molten globule state. Frequent transition between these two states is observed, which reveals a nano-scale dewetting transition accompanied with a single helix rotation. It appears that the sub-optimal packing of hydrophobic core in our designed protein model gives rise to enhanced flexibility allowing helix flipping. This method of destabilization could be used to design proteins with enhanced flexibility as needed for function.

In its current form, there are several limitations of the model presented here. A drawback of using a single angular penalty to account for hydrogen bonds within the protein is that the model can only study a single type of secondary structure. Here we set θ of the SW potential to 180° for BB-BB interactions so that α-helices can form, but formation of β-sheets would require a different angle. Changing θ should allow folding of β-sheets, but the model in its current form is not suitable for secondary structure prediction or for studying sequences which switch between

sheets and helices. Additionally, as the model does not have explicit hydrogen bonding, there is no directionality associated with the hydrogen bonds. Thus the model is unsuitable for studying phenomena in which the dipole moment of secondary structure elements is important, such as peptide-membrane binding [119]. The model also currently lacks any representation of charged residues, but it is not expected that their addition would present any major hurdles.

In the future, we plan to improve the model described here to explore folding of other designed sequences into helical bundles, as well as extend the model to support other types of folds.

3.6 Acknowledgements

P.D. acknowledges support from the IBM Blue Gene Science Program and support from IBM research team led by Satya Nitta. Acknowledgment is made to the donors of The American Chemical Society Petroleum Research Fund for partial support of this research (S.M., G.S.C).

Chapter 4: Role of intramolecular contacts, hydrophobicity, and solvation in the allosteric communication of BirA

4.1 Overview

This chapter is based in part on the author's contributions to: Long distance modulation of disorder-to-order transitions in protein allostery; Wang, J., Custer, G., Beckett, D., Matysiak, S.; *Biochem.* 2017. 56 (34) 4478–4488, as well as: Superrepression through Altered CorepressorActivated Protein:Protein Interactions; He, C., Custer, G., Wang, J., Matysiak, S., Beckett, D.; *Biochem.* 2018. 57 (7) 1119–1129.

4.2 Introduction

Regulation of biological processes is necessary for organisms to respond to changes in environment, available resources, and the needs of the organism. A common method of communicating these changes is known as allosteric regulation, in which the binding of an effector molecule to some site on a protein alters the function of a second site. The binding site of the effector molecule is frequently far from the second site, with the act of effector binding communicated through the

protein. Due to the importance of allosteric regulation in biological systems it is the subject of much research, but the molecular details behind the transmission of an allosteric signal within a protein are still not fully understood.

The *Escherichia coli* biotin ligase/repressor (BirA), shown in dimer form in Figure 4.1A, is a bifunctional protein, which can both adenylate biotin and repress biotin synthesis. BirA forms its product, biotinoyl-5'-adenylate (bio-5'-AMP) following binding of biotin and ATP. After adenylation of biotin there are two pathways possible for the BirA·bio-5'-AMP complex (holoBirA) (Figure 4.1B). If biotin carboxyl carrier protein (BCCP) is available BirA can form a heterodimer with BCCP and transfer biotin to BCCP [160–162]. If BCCP is absent, holoBirA will homodimerize and can then bind to the biotin operator sequence (bioO), repressing synthesis of biotin. Homodimerization of BirA is required for binding to bioO, and is promoted by the binding of bio-5'-AMP [163,164]. Experimental evidence has shown that apoBirA and BirA·biotin form a monomer in solution, while BirA·bio-5'-AMP readily forms a dimer [163], with BirA·bio-5'-AMP having a dimerization energy 4 kcal/mol less than that of apoBirA [165], indicating communication of ligand binding to the dimer interface. Structures of apo and BirA bound to a bio-5'-AMP analog, biotinol-5'-AMP (bioOH-5'-AMP), indicate that ligand binding drives disorder to order transitions in both the ligand binding and dimer interface loops of the protein (loops labeled in Figure 4.1A). In the ligand binding region, the biotin binding loop [residues 116-124 (BBL)] and the adenylate binding loop [residues 211-222 (ABL)] become ordered. In the dimer interface residues 140-146 and 193-199 order upon ligand binding.

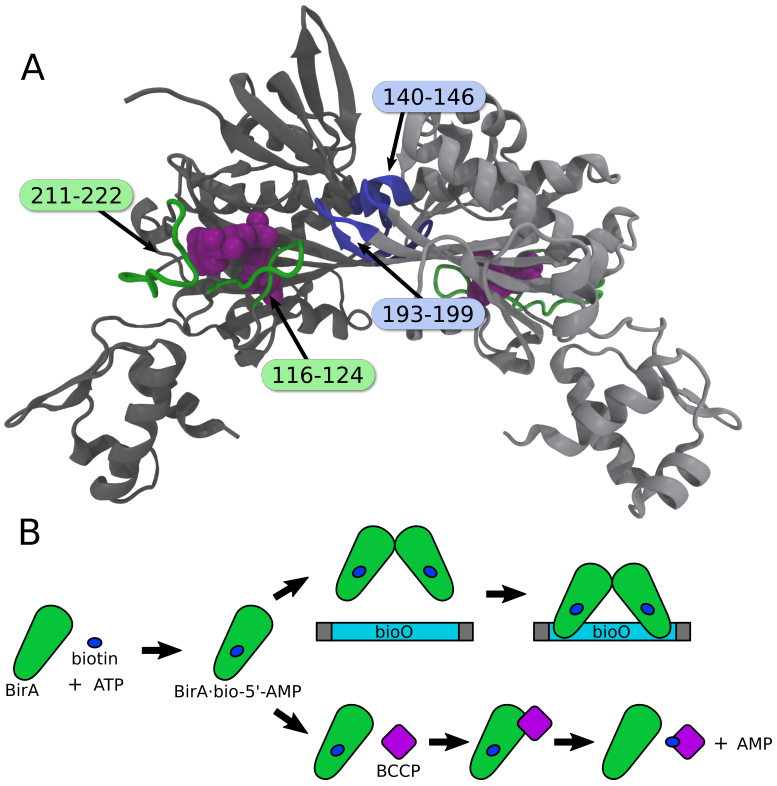

Figure 4.1: A) BirA·bioOH-5'-AMP dimer. Dimer interface loops labeled and highlighted in blue, ligand binding loops in green. Ligand highlighted in purple. One chain is shown in light gray, and the other in dark gray. Structure from PDB entry 2EWN [159] B) General BirA mechanism, based on Wood, *et al.* 2006. [159]

Disorder to order transitions are found in many allosterically regulated systems [166–169]. In human glucokinase, an enzyme critical for regulating glucose levels, binding of glucose triggers ordering of a domain, leading to the active structure of the enzyme [169]. Tetracycline repressor relies on flexibility of its DNA binding domains to access the conformations needed for DNA binding. Ligand binding increases the order of those domains, preventing them from accessing the conformations needed for DNA binding [168]. Cholera toxin requires a more complex order-disorder-order transition of the CTA1 domain to facilitate movement of the toxin through the cell and activation of the toxin [167]. Thus disorder to order transitions allow for many different mechanisms by which a system can be regulated.

In the past, structural changes in a protein on an atomistic level have typically been examined through X-ray crystallography. Due to the necessarily static nature of crystal structures, it can be difficult to use these structures to study processes in proteins which are inherently dynamic. Dynamic processes in proteins are typically studied with other techniques, such as NMR spectroscopy, SAXS, and AUC. These techniques, however, typically come with limits on size or spatial resolution. MD simulations have proven to be well-suited to exploring the changes which occur in allosteric transitions and examining the conformational dynamics of a protein on an atomistic level. Simulations have provided insight into the conformational dynamics involved in allostery for the repressor NikR [24], changes which occur in disorder to order transitions in both Sortase A and phosphorylation of smooth muscle myosin [170]. Insights into the role of solvation in the allosteric transitions of hemoglobin, first identified in experiment [171], have also been obtained through

simulation [172]. Simulations are able to characterize at an atomistic level processes that have been identified experimentally and guide future experimental work.

Our work here uses all-atom molecular dynamics simulations of BirA to examine the allosteric transitions that occur upon ligand binding. We focus on variants of BirA that have been characterized by our experimental collaborators in the Beckett group and identified as altering the dimerization and ligand binding affinities of BirA. We characterize structural changes which occur upon ligand binding for loops in the ligand binding region and dimer interface, and examine changes in flexibility and solvation which occur in the protein when mutating residues both in the ligand binding and interfacial regions as well as in distant regions.

4.3 Methods

4.3.1 Simulation Setup

MD simulations were performed using the coordinates of a monomer (chain A) for the BirA dimer of Protein Data Bank (PDB entry 2EWN) [159]. This structure, which is of the BirA dimer bound to the corepressor analogue biotinol-5'-AMP (btnOH-AMP) was chosen because it is the most complete of the available $BirA^{wt}$ structures, with all residues modeled in the dimerization and ligand binding surfaces. For simulations of apoBirA, the ligand was removed prior to simulation. For simulations of holoBirA, btnOH-AMP was used as the ligand. Amino acid substitutions in the protein sequence were made in PyMOL [173]. For simulations, the protein model was placed in a rhombic dodecahedral box with walls extending 1 nm

past the protein and solvated with 20300 SPC/E [174] water molecules. For BirA and its variants, Na+ and Cl- ions were added to the system as needed to render the system neutral, replacing water molecules when added. Prior to production runs, the energy of the system was minimized using the steepest descent method, followed by NVT and NPT equilibration runs of 100 ps each, using position restraints with a force constant of 1000 kJ mol^{-1} nm^{-2} on the protein. Production runs were performed without position restraints, using an NPT ensemble with a temperature of 300 K and a pressure of 1 bar. These runs were carried out for 1 μs, with the last 500 ns of the simulation used for all analysis as, on the basis of the root-mean-square deviation (RMSD) of backbone positions in the central domain relative to starting structure, all simulated variants equilibrate within the first 500 ns of simulation time (Figure 4.2 shows the variant with the longest equilibration time, wt apo).

All MD simulations were conducted using the GROMACS 4.6 simulator [66–69] and the OPLS-AA force field [175]. Parameters for btnOH-AMP were constructed by analogy to existing molecule/functional group force fields, as done by others [176]. Simulations were run using a time step of 2 fs, with neighbor list updates every five steps. Temperatures of protein and water were maintained independently using the V-rescale algorithm [71] with a time constant of 0.1 ps. Ions, where present, were grouped with water for temperature coupling, while btnOH-AMP was grouped with protein. Isotropic pressure coupling in NPT simulations used the Parrinello–Rahman barostat [72] with a time constant of 2 ps and a compressibility of 4.5×10^{-5} bar^{-1}. Bond lengths were constrained using the LINCS algorithm [177].

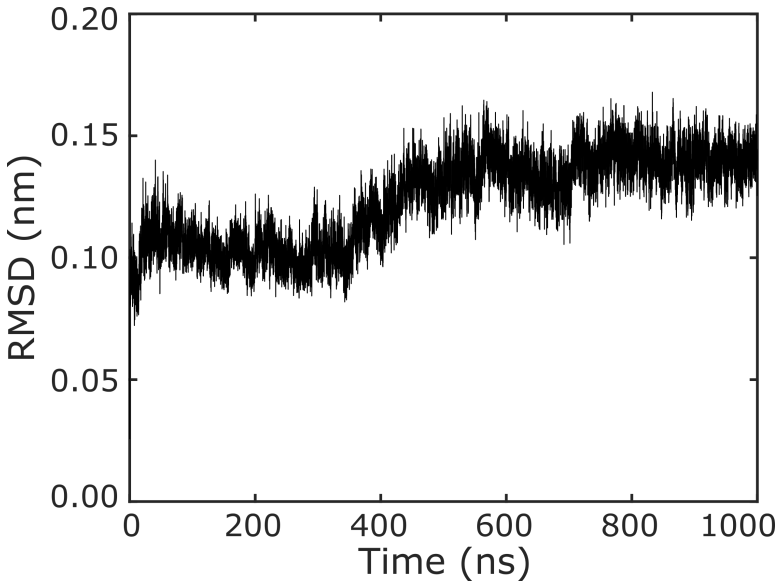

Figure 4.2: Root-mean-square-deviation (RMSD) time series for the C, CA, and N atoms of the central domain of apoBirA^{wt} simulation (residues 80-115,127-139, 147-192, 200-210, and 224- 269), after aligning to C, CA, and N atoms of the same residues in the starting simulation structure

4.3.2 Visualization and analysis

MD structure images were rendered using VMD [78]. To determine representative structures given for the dimer interface of BirA variants, a pairwise RMSD matrix was calculated, using the GROMACS tool *g_cluster* [66–69], for residues 140-146 and 193-199 in the dimer interface using all structures from the last 500 ns of each trajectory. The structure with the minimum dimer interface RMSD relative to all other structures was then taken as the representative structure. As shown, each representative structure has been aligned to the backbone of the holoBirA^{wt} representative structure, including residues 135-155 and 189-204.

For pairwise energy difference plots, the Coulombic and Lennard-Jones energies between each pair of residues were output over time using GROMACS [66–69]. The Coulombic and Lennard-Jones energies for each pair were then summed and averaged for the last 500 ns of each trajectory. The energy shown in the difference map is the average energy for each pair minus the average energy for that pair in the holoBirA simulation. As the average energy for all pairs shown was negative (attractive), more negative values would indicate stronger interaction. In the difference maps, this means that negative values are found for a pair when their interaction is stronger in holoBirA^{wt} and positive values when the interaction is weaker.

Helicity in the simulation of residues 140-152 was quantified by measuring the per-residue average RMSD of an ideal helix (RMSDhx). As both 3_{10}- and α-helices form in the segment, both were considered in this analysis. Ideal helices were constructed using heavy backbone atoms only (N, CA, C, and O), setting the

ϕ and ψ angles to -49°and -26°, respectively, for the α-helix and to -57°and -47°, respectively, for the 3_{10}-helix [178]. A five residue ideal α-helix and 3_{10}-helix were used to calculate RMSDα and RMSD3_{10}, respectively. For each residue, i, in the range of residues 140-152, the five-residue ideal helices were aligned to a five-residue segment centered on residue i, and RMSD was calculated between the structures. The smaller of either RMSDα or RMSD3_{10} was taken as the RMSDhx for that residue in each structure. The RMSD shown is the average per-residue RMSDhx across the last 500 ns of each trajectory.

Principal component analysis was performed on the last 500 ns of each trajectory. The covariance matrix was calculated for the main-chain backbone atoms (N, CA, and C) of residues 140-146 and 193-199 in the dimer interface, using the GROMACS program *g_covar*. Root-mean-square fluctuations of these residues along the principal eigenvectors of the covariance matrix were calculated using the GROMACS program *g_anaeig* [66–69].

Water contact distributions for the dimer interface (residues 140–146 and 193–199) and ligand binding regions (residues 116–124 and 211–222) were calculated using a 3.5Ådistance cutoff for contacts, with distances being calculated between all atoms of protein (including hydrogen) and oxygen atoms only of water. When a water molecule made multiple contacts with a region simultaneously, it was only counted once. Thus, the number of water contacts given represents the total number of water molecules in contacts for a region. Per residue water contact counts were calculated in the same manner. Generation of the surface representation for per residue water contact images was done in VMD [78] with a 1.4Å probe radius. The

difference in water contacts shown for each variant is the difference in the per residue mean number of water contacts for the variant and the per residue mean number of water contacts for holoBirA^{wt}. Coloring has been scaled such that blue represents an increase of 4 water contacts for the residue, white no change in water contacts, and red a decrease of 4 water contacts.

4.4 Results and Discussion

4.4.1 Comparison of apo and holo BirA^{wt}

For all BirA variants studied simulations reached equilibrium within 500 ns and all analysis was performed on the final 500 ns of each 1 μs trajectory. Simulations of apo and holoBirA^{wt} were first compared to existing experimental data to verify the accuracy of the model used for simulation. Crystal structures of BirA indicate the interfacial loops (residues 140-146 and 193-199) become more ordered upon ligand binding. The difference in pairwise interaction energies for residues in the dimer interface was compared for apo and holoBirA^{wt} simulations (Figure 4.3A). Energies represented in this plot are apo energy minus holo, thus interactions which are stronger in apo are negative and those stronger in holo are positive. In general interactions in the dimer interface are weaker in the interfacial loops, especially in loop 140-146, which leads into an α-helix which extends down to the ligand binding surface of the protein. A sample structure illustrating the corresponding change interactions for apo relative to holo is shown in Figure 4.3C. We can further characterize this structural change by examining the RMSD to an ideal helix (Figure 4.3E),

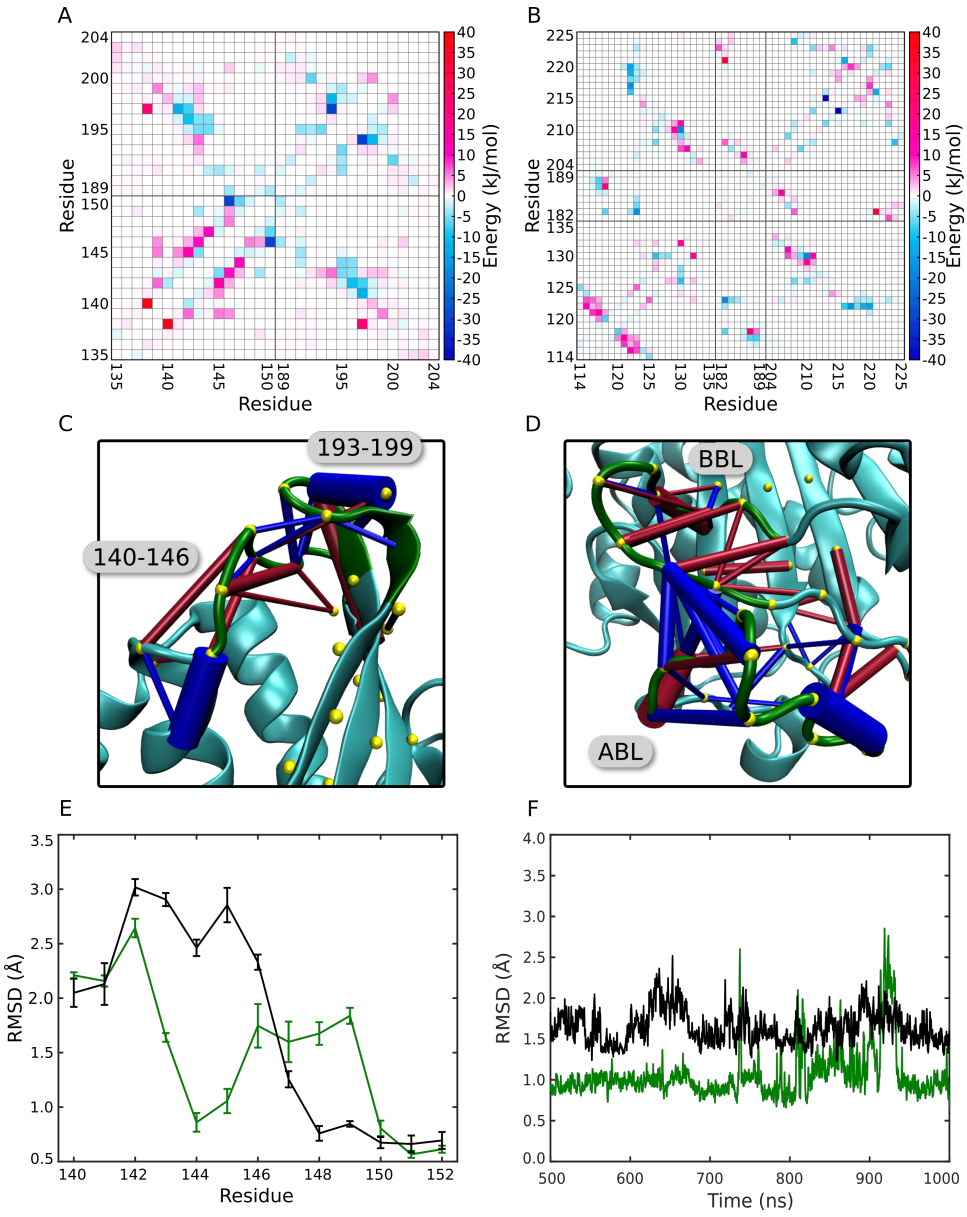

Figure 4.3: Difference average interaction energy map for residue pairs in the A) dimerization surface and B) ligand binding region of apoBirA^{wt}. C)&D) Median structure of apoBirA focused on the C) dimer interface and D) ligand binding region, with rods illustrating pairwise interaction differences from A) and B) respectively with magnitudes greater than 5 kJ/mol. Red rods represent interactions weaker than holo, blue rods stronger, with the thickness of the rod indicating magnitude of the difference. E) Per-residue average RMSD to ideal helix for apoBirA^{wt} (black) and holoBirA^{wt} (green). Error bars represent 95% confidence intervals. F) RMSD time series for C, CA, and N atoms of residues 211-223, after aligning to C, CA, and N atoms of the central domain (residues 80-115, 127-139, 147-192, 200-210, and 224-269). Values smoothed using a sliding average with a 0.5 ns window. Variants shown are apoBirA^{wt} (black) and holoBirA^{wt} (green).

where we see there is a short helical turn around residues 144-145 of the holo simulation (indicated by the low RMSD to ideal helix), while in apo the helix is absent before residue 147. This is consistent with crystal structures, where in apo this helix ends before it reaches residue 146, while in the holo structure the helix extends slightly to include residues 142–145, with a small kink at residue 146 [159, 179]. We note that the extension of the helix in holo appears to come with some weakening of the structure around residues 148–149 relative to apo. This is also captured in the decreased strength of the 1–4 interaction between residues 146 and 150 in Figure 4.3A for holo, where the 29 kj/mol increase in energy would represent a significant weakening of the backbone hydrogen bond. Difference in pairwise interaction energies for the ligand binding region is shown in Figure 4.3B, with changes in packing seen throughout this region for apo compared to holo (Figure 4.3D). Backbone RMSD for the ABL (Figure 4.3F) relative to the starting structure is increased in apo compared to holo, indicating greater structural changes in this loop. The ligand binding pocket of BirA contains many hydrophobic residues and repacking of the ligand binding region is expected in the absence of ligand. Water contacts with hydrophobic sidechains found in the ligand binding pocket, shown in Figure 4.4, increase significantly in the absence ligand, indicating that without the ligand the hydrophobic residues in the pocket are not able to pack effectively. Overall, simulations of apo and holoBirA^{wt} agree very well with experimental results, capturing both the structural changes observed in the dimer interface and the disorder found in the ligand binding loops [159, 179].

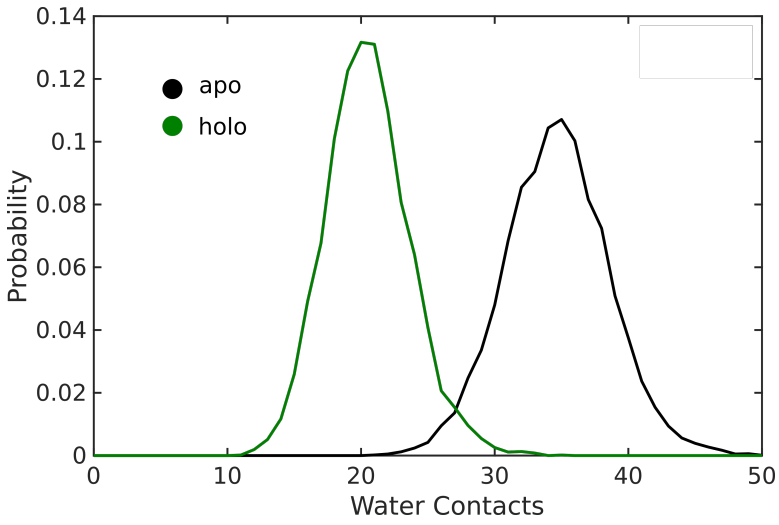

Figure 4.4: Distribution of water contacts for hydrophobic sidechains in the ligand binding pocket (residues W123, F124, M211, V214, V218, V219, and W223) for simulations of apo and holoBirA^{wt}

4.4.2 Structure and dynamics of holoBirA alanine substitution variants

Dimerization and ligand binding free energies of wt and alanine substitution variants of BirA have been determined experimentally by our collaborators in the Beckett group, using isothermal titration calorimetry for ligand binding and sedimentation equilibrium for dimerization [180–184]. In addition to identifying single alanine mutations that alter dimerization and ligand binding, they have identified double-alanine variants of BirA which indicate coupling between residues in the dimerization and ligand binding regions of BirA [184]. Here we have used MD simulations to characterize the structural and dynamic coupling between two of these residues, P143 and M211. P143 is found on the dimerization surface of BirA

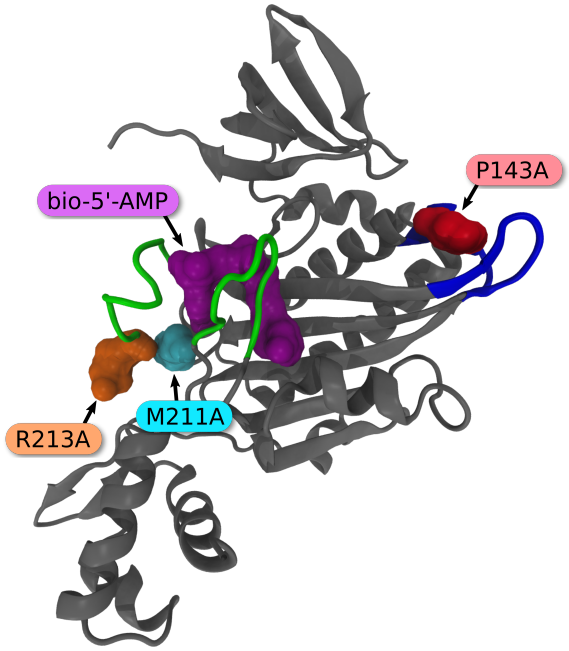

Figure 4.5: Structure of BirA^{wt} monomer (PDB entry 2EWN [159]) highlighting residues modified in variants in the dimer interface, P143 (red) and ligand binding regions, M211 (cyan), and R213 (orange). Ligand highlighted in purple, dimer interface loops in blue, ligand binding loops in green.

(Figure 4.5), in one of the two interfacial loops which become ordered upon ligand binding, while M211 is found in the back of the ligand binding pocket on the ABL. While alanine substitution of M211 has been observed experimentally to increase the Gibbs free energy of bio-5'-AMP binding by over 2 kcal/mol [182, 184], substitution of P143 has no observable effect [183]. However, when a double-alanine substitution is performed with these residues, binding affinity for bio-5'-AMP is partially recovered, with a coupling free energy of -0.7 $\pm$ 0.3 kcal/mol, indicating coupling between these two residues [184]. Alanine substitutions of P143 and M211 both have a negative effect on dimerization of BirA, as determined experimentally

by sedimentation equilibrium [180–182]. However, coupling is not observed in the effects on dimerization for these residues, with changes in Gibbs free energy of dimerization being additive [184].

4.4.2.1 Changes in pairwise interactions

We have examined the changes in interactions within ligand binding region for simulations of holoBirA variants using maps of the pairwise interaction energy between residues (Figure 4.6). As with the comparison between wt apo and holo-BirA, maps shown here represent the difference in interaction energy between each variant and holoBirA^{wt} for each residue pair. To aid with visualizing the change in interactions, example structures are given with rods highlighting pairwise interactions which change by more than 5 kJ/mol. Difference energy maps for BirA^{M211A} indicate large changes and losses in interactions both between and within the ABL and BBL (Figure 4.6A&D), consistent with the reduction in ligand binding affinity found for BirA^{M211A}. In particular, many interactions are lost between the ABL and residue F124, a hydrophobic residue which stretches between the ABL and BBL. These lost interactions are not surprising, given the coupling in bio-5'-AMP binding that has been observed for residues F124 and M211 [182]. Little change in interaction energies are observed on the ligand binding surface for BirA^{P143A} (Figure 4.6B&E), which does not alter the ligand binding affinity relative to BirA^{wt}. When combined, the P143A and M211A substitutions give interaction energies in the ligand binding surface more similar to BirA^{wt} than those observed for the M211A substitution alone

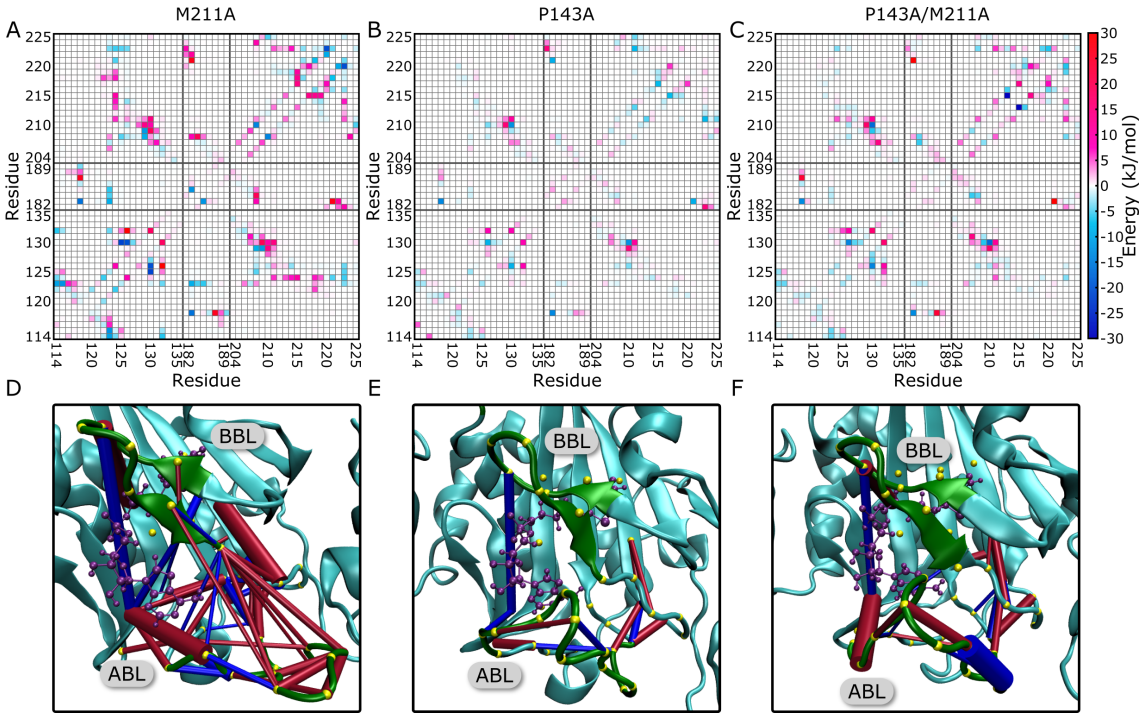

Figure 4.6: Difference average interaction energy maps (top panels) and corresponding structures (bottom panels) for residue pairs in the ligand binding of (A and D) holoBirA^{M211A}, (B and E) holoBirA^{P143A}, and (C and F) holoBirA$^{P143A/M211A}$. Median structure for each variant used, with rods representing corresponding differences in pairwise interaction energy, where the magnitude of that difference exceeds 5 kJ/mol. Blue rods represent interactions stronger in the variant and red weaker interactions. Thickness of the rod correlates with magnitude of the difference.

(Figure 4.6C&F), in agreement with the experimental observation that P143A in combination with M211A is able to partially recover the loss in bio-5'-AMP binding affinity found for the single-alanine M211 substitution [184].

While M211A has been shown experimentally to have weaker dimerization [182], the difference energy map for the dimer interface of $BirA^{M211A}$ appears very similar to that of $holoBirA^{wt}$ (Figure 4.7A&D). The difference energy map for the dimer interface of $BirA^{P143A}$ (Figure 4.7B&E), however, shows a loss of interactions within the 140–146 loop of the dimer interface, consistent with its experimental decrease in dimerization free energy [181]. When combined with the M211A substitution, we observe a loss of many interactions both in and between the 140–146 and 193–199 interfacial loops(Figure 4.7C&F). Thus while we are not able to capture the affect of M211A alone on the dimerization surface, we do observe its effect in combination with P143A.

4.4.2.2 Structure and dynamics of the dimer interface and ligand binding surfaces

To gain further insight into the changes which occur in these single-alanine BirA variants we examined structural properties of the dimer interface and ligand binding regions (Figure 4.8). Comparing the median structure for $holoBirA^{wt}$ and $BirA^{P143A/M211A}$, we observe increased separation between the 140–146 and 193–199 loops in $BirA^{P143A/M211A}$, with little interactions occurring between the loops which pack against each other in the wt structure (Figure 4.8A). As mentioned previously,

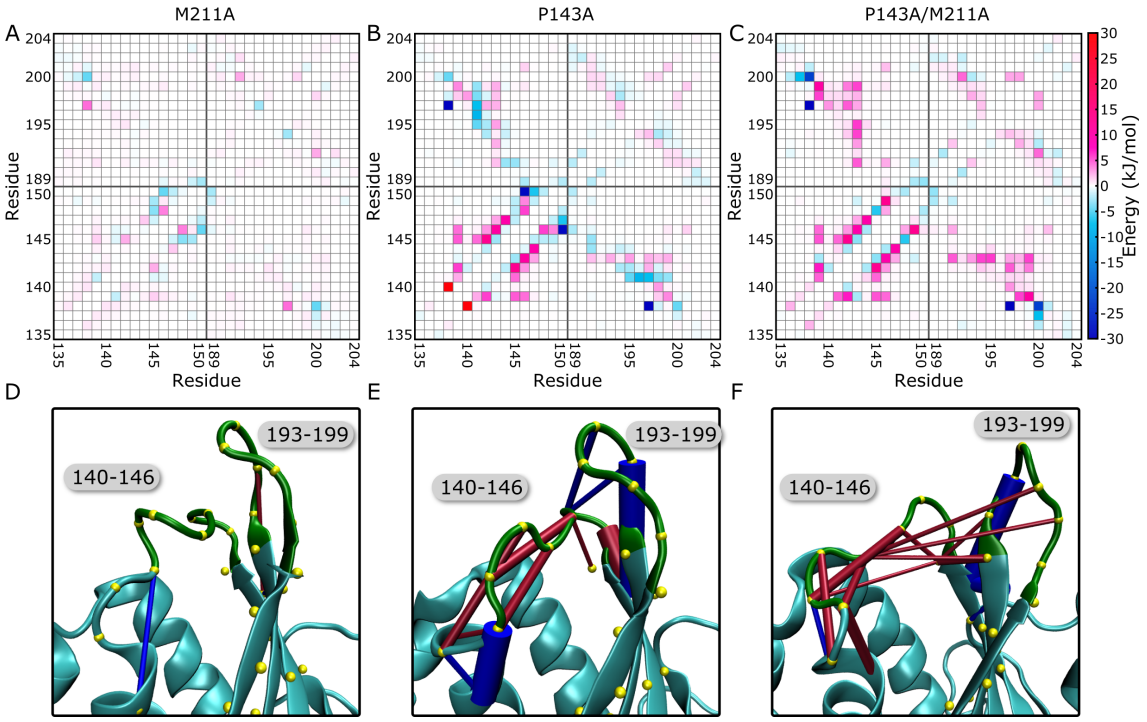

Figure 4.7: Difference average interaction energy maps (top panels) and corresponding structures (bottom panels) for residue pairs in the dimerization surface of (A and D) holoBirA^{M211A}, (B and E) holoBirA^{P143A}, and (C and F) holoBirA$^{P143A/M211A}$. Median structure for each variant used, with rods representing corresponding differences in pairwise interaction energy, where the magnitude of that difference exceeds 5 kJ/mol. Blue rods represent interactions stronger in the variant, red weaker interactions. Thickness of the rod correlates with magnitude of the difference.

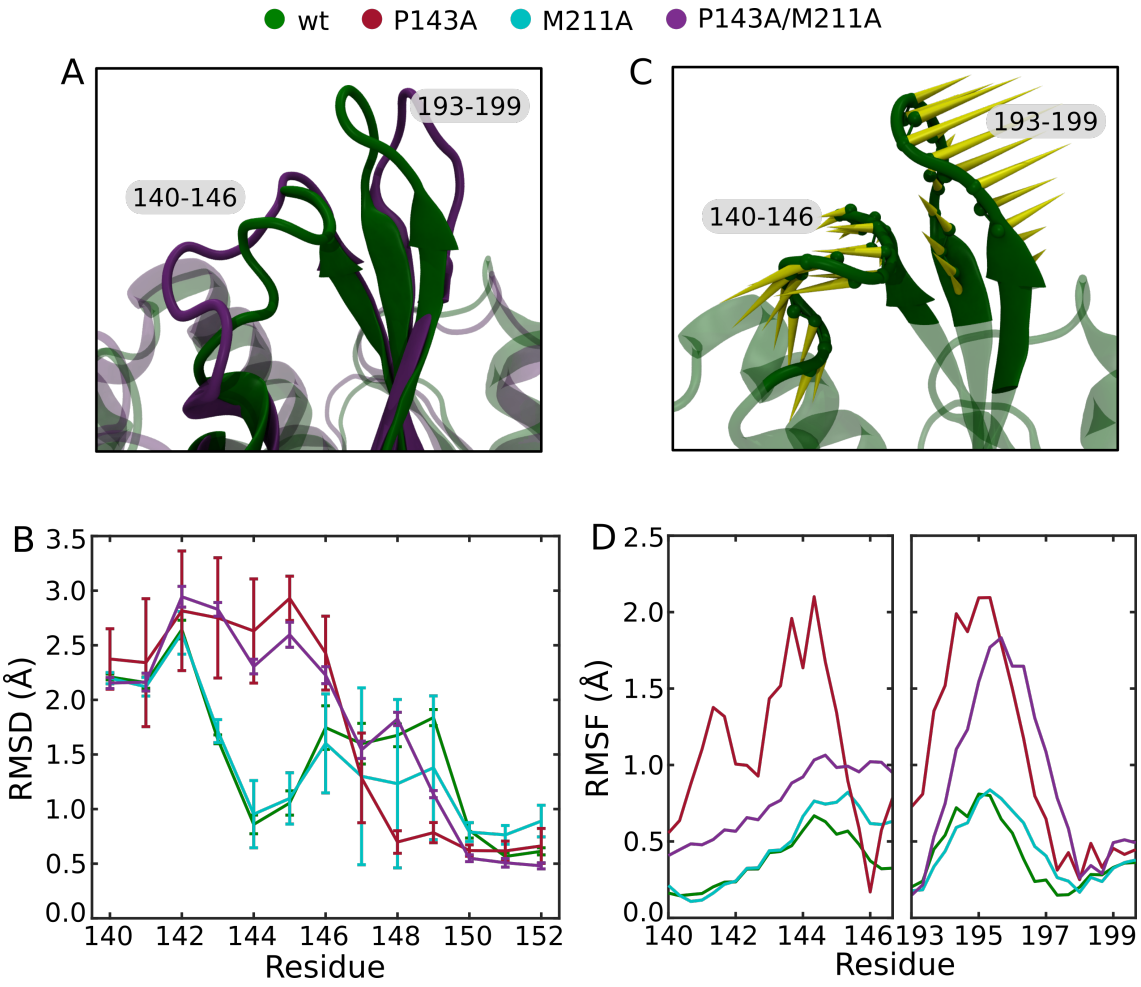

Figure 4.8: Changes in structure and dynamics of BirA dimerization loops for alanine substitution variants. A) Dimerization surface loops 140–146 and 193–199 for holoBirA^{wt} and holoBirA$^{P143A/M211A}$. B) Per-residue average RMSD from an ideal helix for residues 140–152. error bars represent 95% confidence intervals. C) Dimer interface loops of the median holoBirA^{wt} structure shown with the principal eigenvector (yellow) of the covariance matrix for residues 140–146 and 193–199. D) root-mean-square fluctuation (RMSF) of residues 140–146 and 193–199 (C, CA, and N atoms) along the principal eigenvector

crystal structures have shown that the helix following loop 140–146 becomes extended upon bio-5'-AMP binding [159, 179]. Examining this helical segment here we see that, while the helical extension is observed for holoBirA^{wt} and BirA^{M211A}, it is absent in BirA^{P143A} and BirA$^{P143A/M211A}$ (Figure 4.8B). As with apoBirA^{wt}, this loss of helicity is associated with the loss of pairwise interactions found in the 140–146 loop (Figures 4.7B & C).

Simulation results also indicate that alanine substitutions in the ligand binding or dimerization surfaces are capable of altering the dynamics of the 140–146 and 193–199 loops. Principal component analysis of the motions of these interfacial loops indicate that approximately 60% of the motion of these loops is captured by the first principal component, illustrated in Figure 4.8C. The motions are primarily directed such that the loops, folded against each other in the median holoBirA^{wt} structure, oscillate towards and away from each other during the simulation. This motion is minor for holoBirA^{wt} and BirA^{M211A}, as indicated by the root-mean square fluctuations along the principal axis (Figure 4.8D). Motion increases significantly for the BirA^{P143A} and BirA$^{P143A/M211A}$ variants. It is also interesting to note that, motion in the BirA$^{P143A/M211A}$ variant is decreased relative to BirA^{P143A}. In BirA^{P143A}, the loops oscillate between closed (loops folded against each other) and open (loops not interacting) conformations, leading to frequent large motions of the loops. In BirA$^{P143A/M211A}$ the loops remain separated, rarely interacting, which contributes to the decreased overall motion of the loops.

While we do not capture changes in the dimerization surface of BirA^{M211A}, we do capture changes in the structure of the BBL, shown in the RMSD time series for

the M211A variant, which varies significantly from the time series for holoBirA^{wt} and the other variants (Figure 4.9A). The BBL is known to play a role in dimerization [185], and the median structure we observe for BirA^{M211A} differs from that of holoBirA^{wt}(Figure 4.9C). Thus it is possible that the M211A substitution alone alters dimerization in ways that we cannot directly characterize using simulations of monomeric BirA. The magnitude of the changes in RMSD can be put into context by comparing to the RMSD of the central domain of BirA (residues 80–115, 127–139, 147–192, 200–210, and 224–269) for each variant, shown in Figure 4.9B. The RMSD values of the central domain range from approximately 0.7 to 1.1 Å, with all variants being structurally similar. The R213A substitution is used here as a control, as it is on the ABL in close proximity to M211, but its substitution affects neither ligand binding affinity nor dimerization [181]. We observe an RMSD for the BBL of the BirA^{R213A} variant closer to that of holoBirA^{wt} than the M211A variant. Packing of the adenylate binding loop in simulations of holoBirA variants was analyzed by examining the RMSD time series (Figure 4.9D). While we observe an RMSD for BirA^{R213A} similar to that of holoBirA^{wt}, BirA^{M211A} exhibits a large RMSD, approaching 8Å. However, we note that addition of P143A, to the BirA^{M211A} variant results in an RMSD for the ABL similar to that of holoBirA^{wt}.

4.4.2.3 Altered solvation and hydrophobic packing

Hydrophobicity is known to play a large role in binding of bio-5'-AMP by BirA, with a network of hydrophobic residues in the ABL and BBL clustering around the

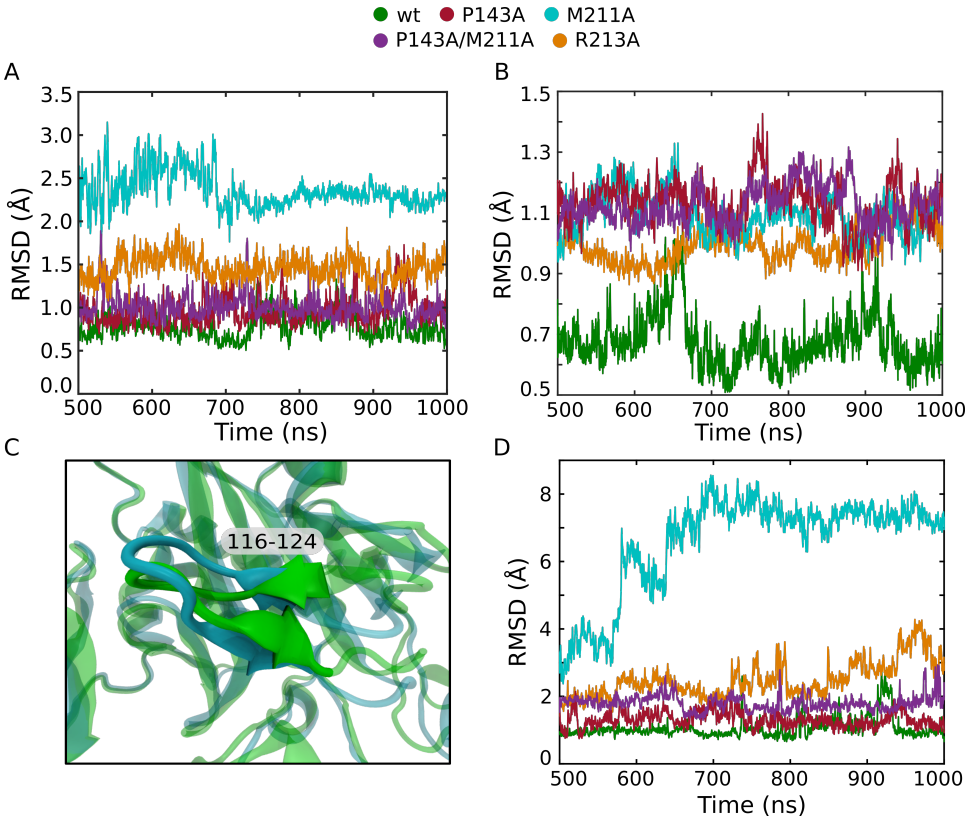

Figure 4.9: A) RMSD time series for BBL residues 116–124 (C, CA, and N atoms). B) RMSD time series for C, CA, and N atoms of the central domain (residues 80–115, 127–139, 147–192, 200–210, and 224–269) C) Overlay of the BBL after alignment of the central domain for median structures of holoBirA^{wt} and BirA^{M211A}. D) RMSD time series for C, CA, and N atoms of the ABL (residues 211–223). All values were smoothed using a sliding average with a 0.5 ns window.

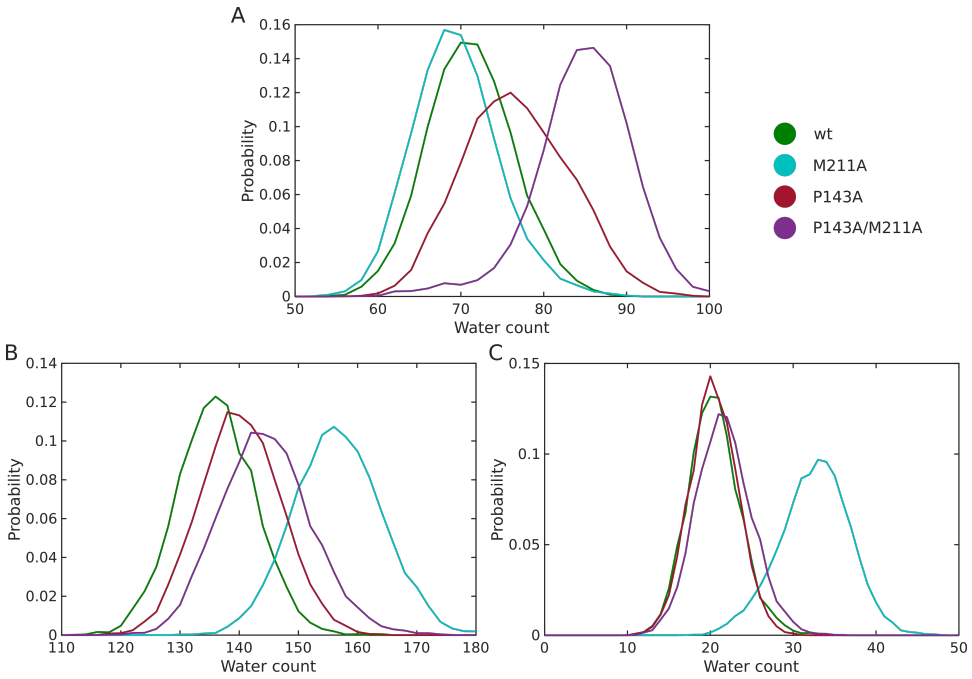

Figure 4.10: Distribution of water contacts for variants of holoBirA. Contacts with A) residues in the dimer interface (residues 140–146 and 193–199), B) residues in the BBL (residues 116 to 124) and ABL (residues 211–222), and C) hydrophobic sidechains in the ligand binding pocket (residues W123, F124, M211, V214, V218, V219, and W223)

ligand in holoBirA^{wt} [182]. We have investigated the changes in solvation of the interfacial and ligand binding loops of BirA in response to the alanine substitutions by counting. Water contacts for the dimer interface of BirA, shown in Figure 4.10A, are similar for holoBirA^{wt} and holoBirA^{M211A}. Some increase in solvation is observed for the P143A variant which, as indicated by the RMSF of the interfacial loops, exhibits increased flexibility of the interfacial loops. The flexibility of the dimer interface loops in BirA^{P143A} leads to a broad distribution of water contacts for the variant. Addition of the M211A substitution to BirA^{P143A} increases solvation further, with 25% more water contacts than holoBirA^{wt}. This is consistent with the

lack of interactions between the dimer interface loops observed for BirA$^{P143A/M211A}$.

Examining water contacts in the ABL and BBL (Figure 4.10B), we observe some (16%) increase in solvation for BirA^{M211A} relative to holoBirA^{wt}, which might be expected given the involvement of M211 in the hydrophobic clusters surrounding the bio-5'-AMP binding site. We observe a reduction in solvation of the ligand binding residues upon addition of the P143A substitution to BirA^{M211A}. The impact of the alanine substitution at M211 is clearer when comparing solvation of the hydrophobic sidechains involved in ligand binding, shown in Figure 4.10C. We observe a nearly 75% increase in water contacts for these hydrophobic residues relative to holoBirA^{wt}, revealing the drastic changes in hydrophobic packing which occur with the M211 alanine substitution. We observe a shift in solvation of similar proportions for both the BBL (Figure 4.11A) and the ABL (Figure 4.11B), though total water contacts with hydrophobic sidechains in the BBL is lower as there are fewer hydrophobic sidechains than in the ABL. This shift in water contacts is even observed for the ligand (Figure 4.11C). It is interesting to note that this increase in water contacts gives the ligand bound BirA^{M211A} a similar level of solvation as found in apoBirA^{wt} (Figure 4.4). Addition of the alanine substitution at P143 nearly brings solvation of the hydrophobic sidechains down to the level of holoBirA^{wt}, indicating a significant improvement in hydrophobic packing, illustrating the large effect the coupling between these residues can have on the disorder to order transition found in the ligand binding region.

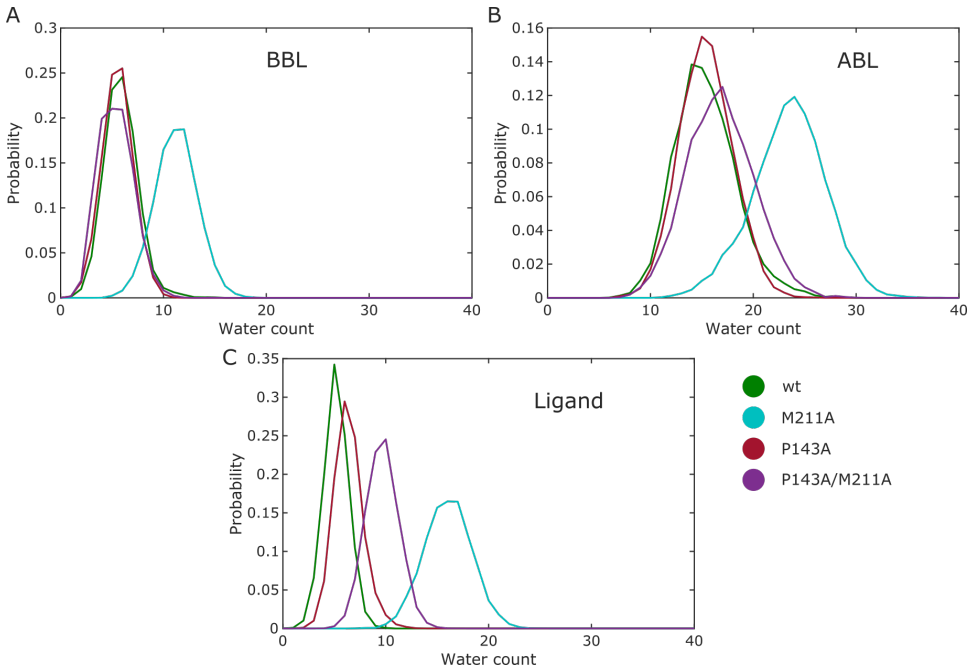

Figure 4.11: Distribution of water contacts for variants of holoBirA. Contacts with A) hydrophobic sidechains in the BBL (W123 and F124), B) hydrophobic sidechains in the ABL (M211, V214, V218, V219, and W223), and C) the ligand

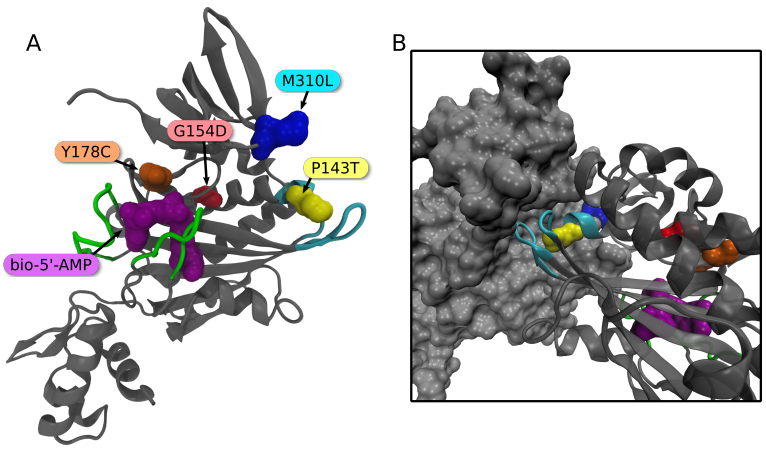

Figure 4.12: A) Structure of BirA^{wt} monomer (PDB entry 2EWN [159]) highlighting residues modified in "superrepressor" variants, G154 (red), M310 (blue), P143T (yellow), and Y178C (orange). Ligand highlighted in purple, dimer interface loops in blue, ligand binding loops in green. B) Dimer interfacial region of BirA^{wt} dimer (PDB entry 2EWN [159]), with same residues and loops colored. Chain A, shown in surface representation, is in a similar orientation to the monomer shown in A.

4.4.3 Shifts in residue contacts and solvation in BirA superrepressors

Single amino acid substitution variants of BirA which exhibit repression of biotin synthesis at lower concentration than BirA^{wt} have been identified by a combination of *in vivo* genetic screens followed by *in vitro* measurement of DNA binding affinity [186]. We have investigated by MD simulations the molecular level changes in structure and solvation which occur in four of these variants, BirA^{G154D}, BirA^{M310L}, BirA^{P143T}, and BirA^{Y178C} (Figure 4.12A). BirA binds to the biotin operator in dimer form, with dimerization made favorable by binding to its product bio-5'-AMP. Thus repression of biotin synthesis would be increased by improved binding of BirA to DNA, improved favorability of dimerization, or enhanced affinity

for biotin. As measurements of BioO affinity for DNA in a wide range of holoBirA variants exhibited little change relative to holoBirA^{wt} [187], DNA binding affinity is not expected to be the cause of enhanced repression. Biotin binding affinities for these BirA "superrepressor" variants, measured by isothermal titration calorimetry, have been determined by our experimental collaborators and indicate no change in binding affinity, indicating that changes in biotin binding are not responsible for the altered repression [188]. All but one of the variants, however, exhibited dimerization free energies more favorable than holoBirA^{wt}, as determined by sedimentation equilibrium [188]. The strongest enhancement of dimerization was observed for BirA^{G154D} [188], which also had the greatest effect on repression of biotin synthesis [186]. BirA^{M310L} and BirA^{P143T} variants also had lower dimerization free energies, while BirA^{Y178C} demonstrated a sharp increase in Gibbs free energy of dimerization [188]. Two of the residues which we examine here, G154 and Y178, are distant from dimer interface, shown in Figure 4.12. Residue M310 is near the dimer interface, but does not directly participate, and P143 is on the interior of dimer interface loop 140–146. The BirA^{P143T} variant is particularly interesting, as we have already noted that an alanine substitution at this position impairs dimerization.

4.4.3.1 Structure of the dimer interface

Simulations of each ligand bound superrepressor variant were performed in the same manner as the alanine substitution variants. We investigate the effects of the substitutions on interactions within the dimer interface by comparing difference

energy maps for the dimerization surface, shown in Figure 4.13, along with corresponding structural representations of these interactions. Three of the variants, BirA^{G154D}, BirA^{P143T}, and BirA^{Y178C}, exhibited very similar interaction energies to holoBirA^{wt}. Interestingly, all three of these variants had decreased interactions for residue R138 with residues D197, and Q200. These interactions were increased relative to holoBirA^{wt} for the alanine substition variants which impaired dimerization. While BirA^{Y178C} exhibits decreased dimerization free energy [188], we do not see any changes that would explain this in the dimer interface loops, with interactions appearing very similar to the P143T variant which enhances dimerization. Large shifts in interactions are observed for BirA^{M310L} (Figure 4.14), with changes that at first glance appear very similar to those found in BirA$^{P143A/M211A}$ (Figure 4.7C). Both variants have decreased interactions within the 140–146 loop, particularly for the 1–4 interactions of 142–144, where we would expect the helical extension found in BirA^{wt}. However, while interactions between the 140–146 and 193–199 loops are largely absent in BirA$^{P143A/M211A}$, some of these interactions are strengthened relative to BirA^{wt} for the M310L variant.

Structural changes in the dimer interface are further investigated by examining changes in the helical extension at residues 142–145 (Figure 4.15). While the BirA^{G154D}, BirA^{P143T}, and BirA^{Y178C} variants all exhibit an extension of the helix comparable to that observed in holoBirA^{wt}, the helical extension does not occur in BirA^{M310L} (Figure 4.15A). However, while BirA^{M310L} does exhibit an enhancement of the helicity relative to wt for residue 147–150, similar to that observed for BirA^{P143A} and BirA$^{P143A/M211A}$, it does not sample the same open conformation seen

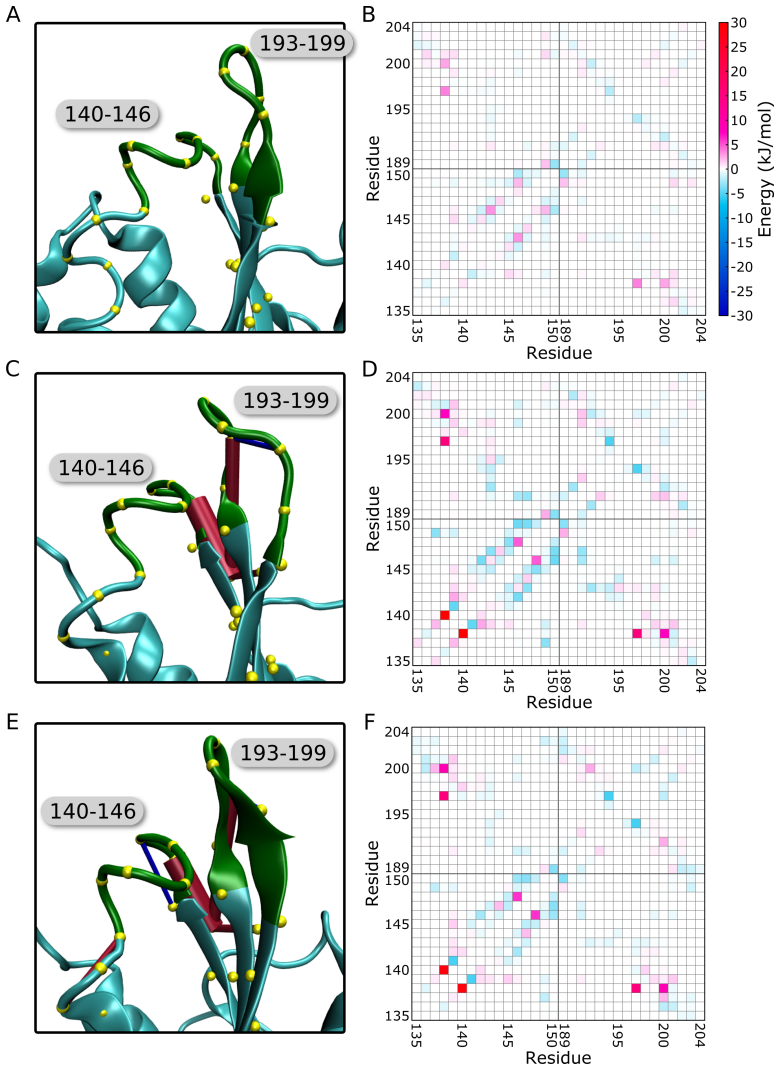

Figure 4.13: Difference average interaction energy maps (right panels) and corresponding structures (left panels) for residue pairs in the dimerization surfaces of A) and B) holoBirA^{G154D}, C) and D) holoBirA^{P143T}, and E) and F) holoBirA^{Y178C}. Median structure for each variant used, with rods representing corresponding differences in pairwise interaction energy, where the magnitude of that difference exceeds 5 kJ/mol. Blue rods represent interactions stronger in the variant, red weaker interactions. Thickness of the rod correlates with magnitude of the difference.

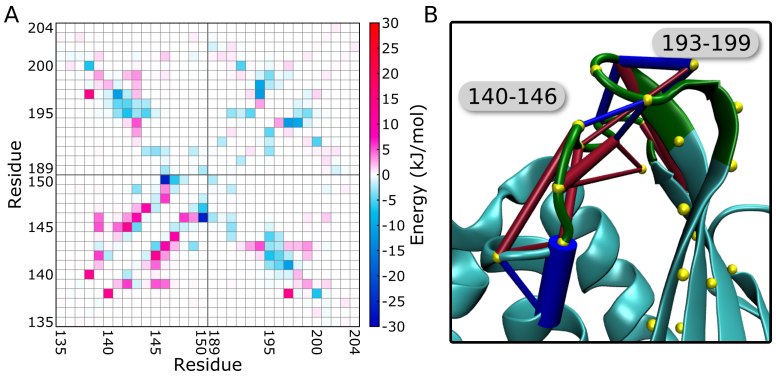

Figure 4.14: Difference average interaction energy maps (A) and corresponding structure (B) for residue pairs in the dimerization surfaces holoBirA^{M310L}. Median structure used, with rods representing corresponding differences in pairwise interaction energy, where the magnitude of that difference exceeds 5 kJ/mol. Blue rods represent interactions stronger in the M310L, red weaker interactions. Thickness of the rod correlates with magnitude of the difference.

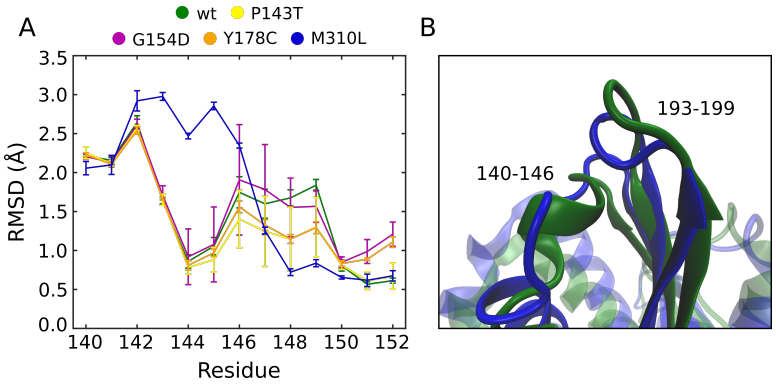

Figure 4.15: Changes in structure of BirA dimerization loops for superrepressor variants. A) Per-residue average RMSD from an ideal helix for residues 140–152. error bars represent 95% confidence intervals. B) Dimerization surface loops 140–146 and 193–199 for holoBirA^{wt} and holoBirA^{M310L}.

in the variants which impaired dimerization. Instead the 193–199 loop folds closer to the 140–146 loop(Figure 4.15B), explaining the stronger interactions observed between the loops in Figure 4.13B. This may contribute to the enhanced dimerization observed for BirA^{M310L}, and indicates that extension of the helix is not necessary for dimerization, with multiple folds of the loops being acceptable. For the other variants, it appears that their effects on dimerization may not come from a direct change of 140–146 and 193–199 interfacial loops.

4.4.3.2 Shifts in electrostatics and solvation

Investigation into structural changes between experimental apo- and holoBirA^{wt} crystal structures by our collaborators have identified a network of electrostatic residues which connects the central and C-terminal domains of BirA [188]. Residues within this network are shown in Figure 4.16. The network observed in simulation of holoBirA^{wt} is essentially the same as that observed in the experimental structure [188]. Involvement of residues R118, R119, and R121 from the BBL means that many connections in the network only form when the BBL becomes structured upon ligand binding. Our collaborator has also identified connections in this network that are altered for the simulated BirA^{G154D}, BirA^{P143T}, and BirA^{Y178C} variants [188]. The addition of the charged aspartic acid in the G154 variant alters many interactions in the network. Comparing BirA^{Y178C}, which impairs dimerization, and BirA^{G154D}, which enhances it, as well as holoBirA^{wt}, there are several changes in the electrostatic network. BirA^{G154D} loses the interactions of E313 with

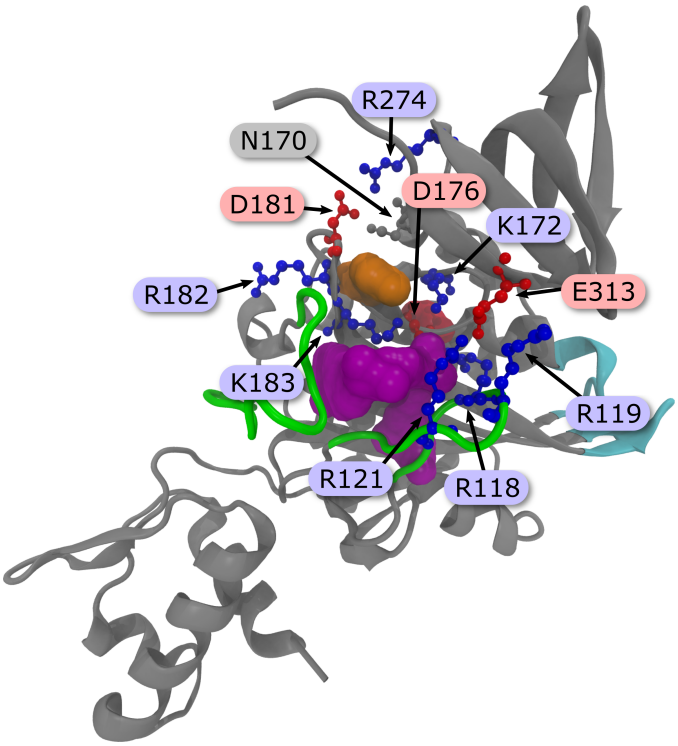

Figure 4.16: Structure of BirA (PDB entry 2EWN [159]) with residues participating in an electrostatic network shown in CPK representation and labeled. Positively charged residues shown in blue, negatively charge residues in red, and the neutral N270 in gray. Y178, also participating in the network, shown in orange surface representation, with G154 in red and bioOH-5'-AMP in purple. Dimer interface loops colored cyan and ligand binding loops green.

R121 and R118 which are found in wt, but gains a strong salt bridge with R119, based on the mean interaction energy found in simulation. BirA^{Y178C} loses some interactions within the central domain, but the interactions of E313 with R121 and R118 become even stronger than holoBirA^{wt}.

To further explore how the superrepressor variants effect dimerization, we quantified solvation changes across the protein, focusing on changes in the dimerization surface. Solvation has been shown to play a large role in dimerization of BirA, as measurements of holoBirA dimerization in D_2O improved free energy of dimerization by -1.5 kcal/mol over dimerization in H_2O [189]. To examine changes in solvation of the dimerization interface, we have calculated the difference in mean water contacts per residue for each variant relative to holoBirA^{wt}. We have used these values to color surface and secondary structure representations of the median structures for each variant, allowing us to compare solvation throughout the dimerization surface and C-terminal domain (Figure 4.17) The introduction of the charge G154D substitution between the C-terminal and central domains results in a large increase in solvation in the residues at the surface of the cleft in which the interfacial 140–146 and 193–199 loops of the second chain dock during dimerization (Figure 4.17A and B). The increase in solvation centers around the strong E313–R119 interaction in BirA^{G154D}. This shift in solvation and increased opening between the C-terminal and central domains may facilitate docking of the many charged residues which make up the interfacial loops, allowing them to remain solvated in the dimer. By contrast, while there is some increased solvation deep in the dimer interface of BirA^{Y178C}, there is slightly decreased solvation around the

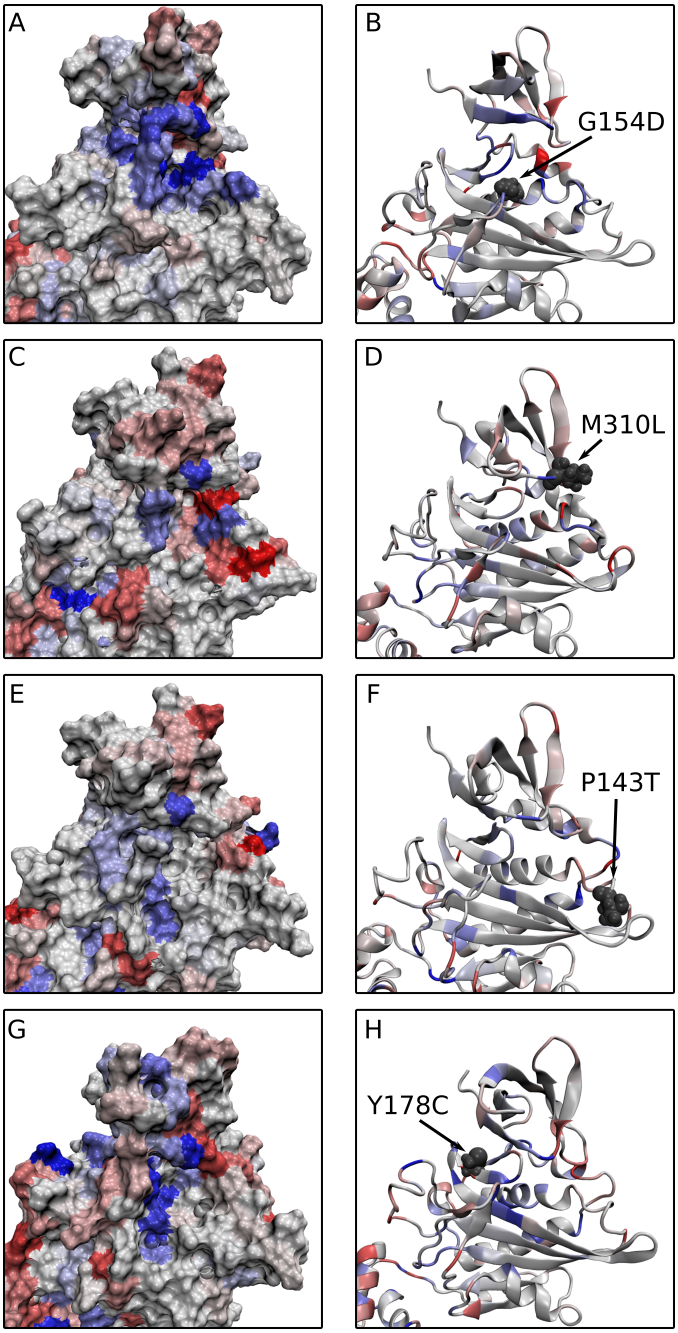

Figure 4.17: Surface (on left) and cartoon (on right) plots of (A and B) holoBirA^{G154D}, (C and D) holoBirA^{M310L}, (E and F) holoBirA^{P143T}, and (G and H) holoBirA^{Y178C}. Residues are colored by their difference in mean number of water contacts compared to holoBirA^{wt}, with blue indicating higher solvation than wt and red lower solvation. Scale of colors is set such that a decrease of 4 water contacts is full red, no change in solvation is white, and an increase of 4 water contacts is full blue. Modified residues shown in gray for cartoon plots.

resdues with bridge the BBL and C-terminal domain and (Figure 4.17G and H), likely due to the stronger interactions observed in the electrostatic network there. This tightening of the interactions between the C-terminal and central domains may contribute to the decreased dimerization free energy observed experimentally for this variant [188]. A number of changes are observed for the BirA^{M310L} variant, with drying observed both locally around the substitution and between the dimer interface loops (Figure 4.17C and D). The drying observed between the interfacial loops agrees well with the observed tightening of the loop conformation. BirA^{P143T} exhibits solvation more similar to holo, though it too has increased solvation around the E313–R119 interaction, similar to that observed for BirA^{G154D}, though smaller in magnitude (Figure 4.17. These observed changes in solvation in the dimerization interface which accompany the changes in dimerization free energy are consistent with the role desolvation of the interface is known to play in forming the dimer [189].

4.5 Conclusion

We have shown here that simulations can provide valuable insight into the structural and dynamic changes which occur in proteins in response to mutation of key residues. We observed excellent agreement between apo- and holoBirA^{wt} simulations and experimental results [159, 165, 179], capturing disorder and increased solvation in the ligand binding loops (Figure 4.4), particularly for key hydrophobic residues which cluster around the ligand [182]. The strong consistency with experimental results validates the computational approach to investigating changes in

structure and solvation.

Simulations of single and double alanine substitions reveal structural changes in the dimer interface for the BirA^{P143A} variant which are amplified when combined with the distant M211A substitution(Figures 4.7 and 4.8), consistent with experimentally determined dimerization free energies [181, 182, 184]. That we are unable to directly capture changes in the BirA^{M211A} could indicate that the changes which occur with the single alanine substitution at M211 happen beyond the time scale we study here, or that the substitution does not directly alter the dimerization interface and are not manifested when simulating a single monomer. Along with the coupling of disorder in the dimer interface to the M211 alanine substitution in the ABL, we observe a recovery of structure in the ABL coupled to the P143A substitution in the dimer interface(Figure 4.9, as indicated by experimental results [184]. Thus it is clear, both by experiment and simulation, that coupling of the dimerization and ligand binding surfaces is bidirectional in BirA. This coupling is also reflected in solvation of the dimer interface (Figure 4.10A) and ligand binding loops (Figure 4.10B). The drastic change in solvation observed in BirA^{M211A}, along with the resulting loss of interactions between F124 and the ABL are consistent with the coupling in bio-5'-AMP binding between these residues observed experimentally [182], and emphasize the importance hydrophobic packing can play in disorder to order transitions.

Structural changes observed in simulations of BirA variants which exhibit increased repression of biotin synthesis do not appear to impact dimerization through a single pathway, instead similar changes in dimerization free energy involve very

different changes in structure in solvation. The BirA^{M310L} reduces free energy of dimerization [188], yet adopts a conformation for the 140–146 loop of the dimer interface more similar to BirA$^{P143A/M211A}$ than holoBirA^{wt} (Figure 4.15). In contrast to the BirA$^{P143A/M211A}$ variant, however, BirA^{M310L} exhibits strong interactions between the 140–146 and 193–199 interfacial loops, indicating that packing of the loops can occur in multiple ways. We also observe changes in solvation and structure of the C-terminal domain which are in agreement with the role experimental results have indicated it plays in allosteric communication within BirA [181,184,188]. Introduction of the G154D substitution, which had both the strongest effect on repression of BioO and enhancement of dimerization free energy [186,188], causes a rearrangement of the electrostatic network linking the C-terminal and central domains as well as an increase in solvation of the BirA dimerization surface between the domains (Figure 4.17A and B). Similarly, changes in solvation and salt bridges appear to play a role in the negative impact the M310L substitution has on dimerization free energy (Figure 4.17G and H). It is clear that communication within BirA, like other allosteric proteins [190,191], involves many possible pathways. While we cannot provide a complete explanation for how each mutation impacts the function of BirA, our results indicate that computational studies can be a valuable tool in combination with experiment for building a better understanding of allosteric communication.

Chapter 5: Thesis Summary

The overall goal of this work was to develop and use computational models to investigate the role of solvation and hydrophobicity in a) aggregation of surfactant into micelles, b) protein folding, and c) protein allostery.

The development of a coarse-grained model for surfactant and its application to micellization was described in Chapter 2. Our model is able to accurately capture interactions of the surfactant with water without an explicit electrostatic potential, significantly reducing the computational requirements of the model. This allows the model to capture both micelle formation and breakdown. We use the model to examine micelle destabilization, with a focus on low temperature destabilization of the micelle. We observe changes in both the overall structure of the micelle and the structure of surfactants within the micelle which cause an increase in water density within the micelle at low temperatures. Decreased contacts between the head and tail groups of surfactants leaving the micelle and tail groups within the micelle accompany the changes in structure at low temperature, making it easier for surfactants to break contact with the micelle. In addition to changes in accessible conformations for surfactants at low temperature, the destabilization of micelles at low temperature is driven by a downward shift in the enthalpy of solvation water

for monomer surfactant at low temperature. Effectively, what we see is a reduction of the hydrophobic effect at low temperatures, as is observed in cold unfolding of proteins [83, 84, 192].

In Chapter 3, the coarse-grained model described in Chapter 2 was modified for application to proteins. We then used the model to examine the folding of a designed helical bundle, characterizing the conformational landscape of the sequence. Our model was able to capture *ab initio* helical folding, and revealed two competing states in the folding landscape of the bundle. The helical flipping observed in our designed sequence is driven by a balance between hydrophobic packing and overall solvation of the protein, and occurs due to suboptimal packing of hydrophobic residues in the helix core. This designed instability in the protein core could be used as a tool to build proteins with enhanced flexibility as needed for function.

The role of hydrophobic packing and intramolecular contacts in allostery for the *E. coli* protein BirA was studied in Chapter 4. We used all-atom simulations of BirA to examine the structural changes in BirA mutants identified by our experimental collaborators as being relevant to allostery in BirA. Simulations showed strong agreement with experimental results, capturing structural changes at both the dimer interface and ligand binding surfaces in response to distant mutations. For single and double alanine substitutions which impair dimerization, we observed a decrease in contacts between two loops in the dimer interface, while a substitution which enhanced dimerization caused an increase in contact between these same loops. Accompanying these changes in contacts is a shift in the solvation of the dimer interface loops, with mutations which impair dimerization exhibiting an in-

crease in solvation of the dimer interface. Variants involving distant residues which enhance or impair dimerization trigger large series of changes in intramolecular contacts and shifts in solvation of the dimerization interface. Overall it appears there are many overlapping pathways by which the ligand binding and dimerization of BirA can be affected.

Combined, the work we have presented here emphasizes the importance of solvation to many critical biological processes, in which the hydrophobic effect is often used to drive the formation of structures. For nonionic micelles, the increased favorability of hydrophobic solvation at low temperatures leads to destabilization. Similarly, poor hydrophobic packing in the ligand binding pocket of BirA, whether caused by the absence of ligand or the substitution of a large hydrophobic residue, can cause disorder in the ligand binding loops and a loss of functionality. In folding of our designed helical sequence, poor hydrophobic packing gives rise to competing folding states which vary in solvation. States which result in an increase in solvation can be useful for functionalilty, as we observe for $BirA^{G154D}$, where an increase in solvation between the C-domain and central domains of the protein, may improve the accessibilty of the binding site for the dimer interfacial loops. We hope that the insights we have provided here on the effects of solvation in folding, aggregation and allostery prove useful for the design and use of protein and surfactant systems in the future.

5.0.1 Future Work

In the future, we plan to use the CG surfactant model described here to characterize the effects of pressure on preformed micelles. Increased pressure is expected to destabilize micelles in a manner similar to low temperature. We will also be expanding the existing model to accomodate charged surfactants. The CG protein model will be expanded to allow formation of sheets, allowing us to characterize the energy landscape of additional protein folds. We also plan to extend our representation of the sidechain beyond the single bead used here, improving the accuracy of our model and allowing us to study more realistic sequences. In our work on allosteric communication within BirA, we are also currently working on using simulations to map potential communication networks through BirA. We also plan to examine additional variants of BirA, including variants involving the electrostatic network connecting the C-domain and central domains.

Bibliography

[1] Claudio Soto. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat Rev Neurosci*, 4(1):49–60, January 2003.

[2] Fabrizio Chiti and Christopher M. Dobson. Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry*, 75(1):333–366, 2006. PMID: 16756495.

[3] Calum J Drummond and Celesta Fong. Surfactant self-assembly objects as novel drug delivery vehicles. *Current Opinion in Colloid & Interface Science*, 4(6):449 – 456, 1999.

[4] Alexander Patist, James R. Kanicky, Pavan K. Shukla, and Dinesh O. Shah. Importance of micellar kinetics in relation to technological processes. *Journal of Colloid and Interface Science*, 245(1):1 – 15, 2002.

[5] Laurier L. Schramm, Elaine N. Stasiuk, and D. Gerrard Marangoni. Surfactants and their applications. *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.*, 99:3–48, 2003.

[6] Ajay Singh, Jonathan D. Van Hamme, and Owen P. Ward. Surfactants in microbiology and biotechnology: Part 2. application aspects. *Biotechnology Advances*, 25(1):99 – 121, 2007.

[7] C. James McKnight, Paul T. Matsudaira, and Peter S. Kim. Nmr structure of the 35-residue villin headpiece subdomain. *Nat Struct Mol Biol*, 4(3):180–184, March 1997.

[8] Yong Duan, Lu Wang, and Peter A. Kollman. The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation. *Proceedings of the National Academy of Sciences*, 95(17):9897–9902, 1998.

[9] Hongxing Lei, Chun Wu, Haiguang Liu, and Yong Duan. Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations. *Proceedings of the National Academy of Sciences*, 104(12):4925–4930, 2007.

[10] Maria Velinova, Durba Sengupta, Alia V. Tadjer, and Siewert-Jan Marrink. Sphere-to-rod transitions of nonionic surfactant micelles in aqueous solution modeled by molecular dynamics simulations. *Langmuir*, 27(23):14071–14077, 2011. PMID: 21981373.

[11] W. Clark Still, Anna Tempczyk, Ronald C. Hawley, and Thomas Hendrickson. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, 112(16):6127–6129, 1990.

[12] Yuzo Ueda, Hiroshi Taketomi, and Nobuhiro G. Studies on protein folding, unfolding, and fluctuations by computer simulation. ii. a. three-dimensional lattice model of lysozyme. *Biopolymers*, 17(6):1531–1548, 1978.

[13] Ken A Dill. Theory for the folding and stability of globular proteins. *Biochemistry*, 24(6):1501–1509, 1985.

[14] JD Honeycutt and D Thirumalai. Metastability of the folded states of globular proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 87(9):3526–3529, 1990.

[15] Michael Levitt and Arieh Warshel. Computer simulation of protein folding. *Nature*, 253(5494):694–698, 1975.

[16] Anne Voegler Smith and Carol K Hall. α-helix formation: Discontinuous molecular dynamics on an intermediate-resolution protein model. *Proteins*, 44(3):344–360, 2001.

[17] Feng Ding, Sergey V Buldyrev, and Nikolay V Dokholyan. Folding trp-cage to nmr resolution native structure using a coarse-grained protein model. *Biophys. J.*, 88(1):147–155, 2005.

[18] Giovanni Bellesia and Joan-Emma Shea. Self-assembly of β-sheet forming peptides into chiral fibrillar aggregates. *J. Chem. Phys.*, 126(24):245104, 2007.

[19] Siewert J. Marrink, H Jelger Risselada, Serge Yefimov, D Peter Tieleman, and Alex H. de Vries. The martini force field: coarse grained model for biomolecular simulations. *J Phys Chem B*, 111(27):7812–7824, Jul 2007.

[20] Zhe Wu, Qiang Cui, and Arun Yethiraj. A new coarse-grained model for water: the importance of electrostatic interactions. *J Phys Chem B*, 114(32):10524–10529, Aug 2010.

[21] Jacque Monod, Jeffries Wyman, and Jean-Pierre Changeux. On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology*, 12(1):88 – 118, 1965.

[22] D. E. Koshland, G. Nmethy, and D. Filmer. Comparison of experimental binding data and theoretical models in proteins containing subunits*. *Biochemistry*, 5(1):365–385, 1966. PMID: 5938952.

[23] Qiang Cui and Martin Karplus. Allostery and cooperativity revisited. *Protein Science*, 17(8):1295–1307, 2008.

[24] Michael J. Bradley, Peter T. Chivers, and Nathan A. Baker. Molecular dynamics simulation of the escherichia coli nikr protein: Equilibrium conformational fluctuations reveal interdomain allosteric communication pathways. *Journal of Molecular Biology*, 378(5):1155 – 1173, 2008.

[25] Andre A. S. T. Ribeiro and Vanessa Ortiz. Determination of signaling pathways in proteins through network theory: Importance of the topology. *Journal of Chemical Theory and Computation*, 10(4):1762–1769, 2014.

[26] Valeria Molinero and Emily B. Moore. Water modeled as an intermediate element between carbon and silicon. *The Journal of Physical Chemistry B*, 113(13):4008–4016, 2009.

[27] D. John Mitchell, Gordon J. T. Tiddy, Loraine Waring, Theresa Bostock, and Malcolm P. McDonald. Phase behaviour of polyoxyethylene surfactants with water. mesophase structures and partial miscibility (cloud points). *J. Chem. Soc., Faraday Trans. 1*, 79:975–1000, 1983.

[28] Reinhard Strey, Reinhard Schomacker, Didier Roux, Frederic Nallet, and Ulf Olsson. Dilute lamellar and l3 phases in the binary water-c12e5 system. *J. Chem. Soc., Faraday Trans.*, 86:2253–2261, 1990.

[29] Sudhakar Puvvada and Daniel Blankschtein. Molecularthermodynamic approach to predict micellization, phase behavior and phase separation of micellar solutions. i. application to nonionic surfactants. *The Journal of Chemical Physics*, 92(6):3710–3724, 1990.

[30] D. J. Lee. Enthalpy-entropy compensation in ionic micelle formation. *Colloid and Polymer Science*, 273(6):539–543, Jun 1995.

[31] Stefan Paula, Willy Sues, Juergen Tuchtenhagen, and Alfred Blume. Thermodynamics of micelle formation as a function of temperature: A high sensitivity titration calorimetry study. *The Journal of Physical Chemistry*, 99(30):11742–11751, 1995.

[32] Olle Söderman, Harald Walderhaug, Ulf Henriksson, and Peter Stilbs. Nmr relaxation in isotropic surfactant systems. a 2 h, 13 c, and 14 n nmr study of the micellar (l 1) and cubic (i 1) phases in the dodecyltrimethylammonium chloride/water system. *Journal of Physical Chemistry*, 89(17):3693–3701, 1985.

[33] Haakan Wennerstroem, Bjoern Lindman, Olle Soederman, Torbjoern Drakenberg, and Jarl B Rosenholm. Carbon-13 magnetic relaxation in micellar solutions. influence of aggregate motion on t1. *Journal of the American Chemical Society*, 101(23):6860–6864, 1979.

[34] Olle Sderman, Peter Stilbs, and William S. Price. Nmr studies of surfactants. *Concepts in Magnetic Resonance Part A*, 23A(2):121–135, 2004.

[35] Per Gunnar Nilsson and Bjoern Lindman. Mixed micelles of nonionic and ionic surfactants. a nuclear magnetic resonance self-diffusion and proton relaxation study. *The Journal of Physical Chemistry*, 88(22):5391–5397, 1984.

[36] Gerd Olofsson. Microtitration calorimetric study of the micellization of three poly(oxyethylene)glycol dodecyl ethers. *The Journal of Physical Chemistry*, 89(8):1473–1477, 1985.

[37] M. Lesemann, A. Martin, L. Belkoura, D. Woermann, and E. Hoinkis. Static light scattering and small angle neutron scattering experiments with a nonionic surfactant/deuterium oxide system approaching the liquid/liquid coexistence curve: Study of the system c12e5/d2o. *Berichte der Bunsengesellschaft fr physikalische Chemie*, 101(2):228–235, 1997.

[38] J. Brunner-Popela, R. Mittelbach, R. Strey, K.-V. Schubert, E. W. Kaler, and O. Glatter. Small-angle scattering of interacting particles. iii. d2o-c12e5 mixtures and microemulsions with n-octane. *The Journal of Chemical Physics*, 110(21):10623–10632, 1999.

[39] Dalila Bendedouch, Sow Hsin Chen, and W. C. Koehler. Structure of ionic micelles from small angle neutron scattering. *The Journal of Physical Chemistry*, 87(1):153–159, 1983.

[40] Charles Tanford, Yasuhiko Nozaki, and Michael F. Rohde. Size and shape of globular micelles formed in aqueous solution by n-alkyl polyoxyethylene ethers. *The Journal of Physical Chemistry*, 81(16):1555–1560, 1977.

[41] Laurence S. Romsted and Jihu Yao. Arenediazonium salts: new probes of the interfacial compositions of association colloids. 4.1-3 estimation of the hydration numbers of aqueous hexaethylene glycol monododecyl ether, c12e6, micelles by chemical trapping. *Langmuir*, 12(10):2425–2432, 1996.

[42] Laurence S. Romsted and Jihu Yao. Arenediazonium salts: new probes of the interfacial compositions of association colloids. 5.14 determination of hydration numbers and radial distributions of terminal hydroxyl groups in mixed nonionic cmen micelles by chemical trapping. *Langmuir*, 15(2):326–336, 1999.

[43] Li-Jen Chen, Shi-Yow Lin, Chiung-Chang Huang, and En-Ming Chen. Temperature dependence of critical micelle concentration of polyoxyethylenated non-ionic surfactants. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 135(1):175 – 181, 1998.

[44] Mohamed Aoudia and Raoul Zana. Aggregation behavior of sugar surfactants in aqueous solutions: Effects of temperature and the addition of nonionic polymers. *Journal of Colloid and Interface Science*, 206(1):158 – 167, 1998.

[45] S.K. Mehta, K.K. Bhasin, Renu Chauhan, and Shilpee Dham. Effect of temperature on critical micelle concentration and thermodynamic behavior of dodecyldimethylethylammonium bromide and dodecyltrimethylammonium chloride in aqueous media. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 255(1):153 – 157, 2005.

[46] Mikael Jonströmer, Bengt Jönsson, and Björn Lindman. Self-diffusion in nonionic surfactant-water systems. *The Journal of Physical Chemistry*, 95(8):3293–3300, 1991.

[47] Joshua A. Long, Blake M. Rankin, and Dor Ben-Amotz. Micelle structure and hydrophobic hydration. *Journal of the American Chemical Society*, 137(33):10809–10815, 2015. PMID: 26222042.

[48] S. J. Marrink, D. P. Tieleman, and A. E. Mark. Molecular dynamics simulation of the kinetics of spontaneous micelle formation. *The Journal of Physical Chemistry B*, 104(51):12165–12173, 2000.

[49] Peter J. Bond, Jonathan M. Cuthbertson, Sundeep S. Deol, and Mark S. P. Sansom. Md simulations of spontaneous membrane protein/detergent micelle formation. *Journal of the American Chemical Society*, 126(49):15948–15949, 2004. PMID: 15584713.

[50] W. Shinoda, R. DeVane, and M. L. Klein. Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants. *Molecular Simulation*, 33(1-2):27–36, 2007.

[51] Wataru Shinoda, Russell DeVane, and Michael L. Klein. Coarse-grained molecular modeling of non-ionic surfactant self-assembly. *Soft Matter*, 4:2454–2462, 2008.

[52] Shekhar Garde, Lu Yang, Jonathan S. Dordick, and Michael E. Paulaitis. Molecular dynamics simulation of c8e5 micelle in explicit water: structure and hydrophobic solvation thermodynamics. *Molecular Physics*, 100(14):2299–2306, 2002.

[53] Subas Dhakal and Radhakrishna Sureshkumar. Topology, length scales, and energetics of surfactant micelles. *The Journal of Chemical Physics*, 143(2):024905, 2015.

[54] Michael A. Johnston, William C. Swope, Kirk E. Jordan, Patrick B. Warren, Massimo G. Noro, David J. Bray, and Richard L. Anderson. Toward a standard protocol for micelle simulation. *The Journal of Physical Chemistry B*, 120(26):6337–6351, 2016. PMID: 27096611.

[55] Andrew P. Santos and Athanassios Z. Panagiotopoulos. Determination of the critical micelle concentration in simulations of surfactant systems. *The Journal of Chemical Physics*, 144(4):044709, 2016.

[56] Sadegh Faramarzi, Brittany Bonnett, Carl A. Scaggs, Ashley Hoffmaster, Danielle Grodi, Erica Harvey, and Blake Mertz. Molecular dynamics simulations as a tool for accurate determination of surfactant micelle properties. *Langmuir*, 33(38):9934–9943, 2017. PMID: 28836794.

[57] Themis Lazaridis, Buddhadeb Mallik, and Yong Chen. Implicit solvent simulations of dpc micelle formation. *The Journal of Physical Chemistry B*, 109(31):15098–15106, 2005. PMID: 16852911.

[58] Arben Jusufi, Antti-Pekka Hynninen, and Athanassios Z. Panagiotopoulos. Implicit solvent models for micellization of ionic surfactants. *The Journal of Physical Chemistry B*, 112(44):13783–13792, 2008. PMID: 18844395.

[59] Semen O. Yesylevskyy, Lars V. Schäfer, Durba Sengupta, and Siewert J. Marrink. Polarizable water model for the coarse-grained martini force field. *PLoS Comput Biol*, 6(6):e1000810, Jun 2010.

[60] Simon Schrdle, Glenn Hefter, Werner Kunz, , and Richard Buchner. Effects of nonionic surfactant c12e5 on the cooperative dynamics of water. *Langmuir*, 22(3):924–932, 2006. PMID: 16430250.

[61] Frank H. Stillinger and Thomas A. Weber. Computer simulation of local order in condensed phases of silicon. *Phys. Rev. B*, 31:5262–5271, Apr 1985.

[62] Payel Das and Silvina Matysiak. Direct characterization of hydrophobic hydration during cold and pressure denaturation. *J Phys Chem B*, 116(18):5342–5348, May 2012.

[63] Silvina Matysiak and Payel Das. Effects of sequence and solvation on the temperature-pressure conformational landscape of proteinlike heteropolymers. *Phys Rev Lett*, 111(5):058103, Aug 2013.

[64] Gregory S. Custer, Payel Das, and Silvina Matysiak. Interplay between conformational heterogeneity and hydration in the folding landscape of a designed three-helix bundle. *The Journal of Physical Chemistry B*, 121(13):2731–2738, 2017. PMID: 28282142.

[65] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins*, 78(8):1950–1958, Jun 2010.

[66] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman J C Berendsen. Gromacs: fast, flexible, and free. *J Comput Chem*, 26(16):1701–1718, Dec 2005.

[67] Berk Hess, Carsten Kutzner, David van der Spoel, and Erik Lindahl. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008.

[68] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1-3):43–56, 1995.

[69] Erik Lindahl, Berk Hess, and David van der Spoel. Gromacs 3.0: a package for molecular simulation and trajectory analysis. *Molecular Modeling Annual*, 7(8):306–317, 2001.

[70] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *The Journal of Physical Chemistry*, 91(24):6269–6271, 1987.

[71] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *J Chem Phys*, 126(1):014101, Jan 2007.

[72] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, 1981.

[73] Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, 117(1):1 – 19, 1995.

[74] M Shick. *Nonionic Surfactants*, volume 23 of *Surfactant Science Series*. Physical Chemistry, New York, 1987.

[75] Loup Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98–103, Jul 1967.

[76] Shuichi Nosé. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81(1):511–519, 1984.

[77] William G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, Mar 1985.

[78] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[79] P-L Chau and AJ Hardwick. A new order parameter for tetrahedral configurations. *Molecular Physics*, 93(3):511–518, 1998.

[80] Jeffrey R. Errington and Pablo G. Debenedetti. Relationship between structural order and the anomalies of liquid water. *Nature*, 409:318, January 2001.

[81] Zaid A Al-Anber, Josep Bonet i Avalos, M Antonio Floriano, and Allan D Mackie. Sphere-to-rod transitions of micelles in model nonionic surfactant solutions. *The Journal of chemical physics*, 118(8):3816–3826, 2003.

[82] Otto Glatter, Gerhard Fritz, Helmut Lindner, Judith Brunner-Popela, Rainer Mittelbach, Reinhard Strey, and Stefan U. Egelhaaf. Nonionic micelles near the critical point: micellar growth and attractive interaction. *Langmuir*, 16(23):8692–8701, 2000.

[83] Peter L. Privalov. Cold denaturation of protein. *Critical Reviews in Biochemistry and Molecular Biology*, 25(4):281–306, 1990.

[84] Cristiano L Dias, Tapio Ala-Nissila, Jirasak Wong-ekkabut, Ilpo Vattulainen, Martin Grant, and Mikko Karttunen. The hydrophobic effect and its role in cold denaturation. *Cryobiology*, 60(1):91–99, Feb 2010.

[85] Fabio Sterpone, Carlo Pierleoni, Giuseppe Briganti, and Massimo Marchi. Molecular dynamics study of temperature dehydration of a c12e6 spherical micelle. *Langmuir*, 20(11):4311–4314, 2004. PMID: 15969130.

[86] Li-Jen Chen, Shi-Yow Lin, and Chiung-Chang Huang. Effect of hydrophobic chain length of surfactants on enthalpyentropy compensation of micellization. *The Journal of Physical Chemistry B*, 102(22):4350–4356, 1998.

[87] Gohsuke Sugihara and Mihoko Hisatomi. Enthalpyentropy compensation phenomenon observed for different surfactants in aqueous solution. *Journal of Colloid and Interface Science*, 219(1):31 – 36, 1999.

[88] Ken A Dill. Dominant forces in protein folding. *Biochemistry*, 29(31):7133–7155, 1990.

[89] James W Bryson, Stephen F Betz, Helen S Lu, Daniel J Suich, Hongxing X Zhou, Karyn T O'neil, and William F DeGrado. Protein design: a hierarchic approach. *Science*, 270(5238):935–941, 1995.

[90] Christopher M. Dobson. Protein folding and misfolding. *Nature*, 426(6968):884–890, December 2003.

[91] Ken A. Dill and Justin L. MacCallum. The protein-folding problem, 50 years on. *Science*, 338(6110):1042–1046, 2012.

[92] Dennis Vitkup, Dagmar Ringe, Gregory A. Petsko, and Martin Karplus. Solvent mobility and the protein 'glass' transition. *Nat. Struct. Mol. Biol.*, 7(1):34–38, January 2000.

[93] Jeremy C. Smith, Franci Merzel, Ana-Nicoleta Bondar, Alexander Tournier, and Stefan Fischer. Structure, dynamics and reactions of protein hydration water. *Philos. Trans. R. Soc. Lond. B*, 359(1448):1181–1190, 2004.

[94] Tomotaka Oroguchi and Masayoshi Nakasako. Changes in hydration structure are necessary for collective motions of a multi-domain protein. *Sci. Rep.*, 6:26302, May 2016.

[95] J. T. Edsall and H. A. McKenzie. Water and proteins. ii. the location and dynamics of water in protein systems and its relation to their stability and properties. *Adv. Biophys.*, 16:53–183, 1983.

[96] A. A. Rashin, M. Iofin, and B. Honig. Internal cavities and buried waters in globular proteins. *Biochemistry*, 25(12):3619–3625, Jun 1986.

[97] J. A. Ernst, R. T. Clubb, H. X. Zhou, A. M. Gronenborn, and G. M. Clore. Demonstration of positionally disordered water within a protein hydrophobic cavity by nmr. *Science*, 267(5205):1813–1817, Mar 1995.

[98] A. Caflisch and M. Karplus. Molecular dynamics simulation of protein denaturation: solvation of the hydrophobic cores and secondary structure of barnase. *Proc. Natl. Acad. Sci. U.S.A.*, 91(5):1746–1750, Mar 1994.

[99] Debashree Chakraborty, Antoine Taly, and Fabio Sterpone. Stay wet, stay stable? how internal water helps the stability of thermophilic proteins. *J. Phys. Chem. B*, 119(40):12760–12770, Oct 2015.

[100] Ouwen Zhang and Julie D. Forman-Kay. Nmr studies of unfolded states of an sh3 domain in aqueous solution and denaturing conditions. *Biochemistry*, 36(13):3959–3970, 1997.

[101] Margaret S. Cheung, Angel E. García, and José N. Onuchic. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. U.S.A.*, 99(2):685–690, Jan 2002.

[102] Hugh Savage and Alexander Wlodawer. Determination of water structure around biomolecules using x-ray and neutron diffraction methods. *Methods Enzymol.*, 127:162–183, 1986.

[103] Gottfried Otting, Edvards Liepinsh, and Kurt Wuthrich. Protein hydration in aqueous solution. *Science*, 254(5034):974–980, Nov 15 1991.

[104] Samir Kumar Pal, Jorge Peon, and Ahmed H. Zewail. Biological water at the protein surface: Dynamical solvation probed directly with femtosecond resolution. *Proc. Natl. Acad. Sci. U.S.A.*, 99(4):1763–1768, 2002.

[105] Garegin A. Papoian, Johan Ulander, Michael P. Eastwood, Zaida Luthey-Schulten, and Peter G. Wolynes. Water in protein structure prediction. *Proc. Natl. Acad. Sci. U.S.A.*, 101(10):3352–3357, 2004.

[106] Jos Nelson Onuchic and Peter G Wolynes. Theory of protein folding. *Curr. Opin. Struc. Biol.*, 14(1):70–75, 2004.

[107] Lu Yang, Jonathan S. Dordick, and Shekhar Garde. Hydration of enzyme in nonaqueous media is consistent with solvent dependence of its activity. *Biophys. J.*, 87(2):812–821, 2004.

[108] David M. Huang and David Chandler. Temperature and length scale dependence of hydrophobic effects and their possible implications for protein folding. *Proc. Natl. Acad. Sci. U.S.A.*, 97(15):8324–8327, 2000.

[109] Fabio Sterpone, Guillaume Stirnemann, James T. Hynes, and Damien Laage. Water hydrogen-bond dynamics around amino acids: The key role of hydrophilic hydrogen-bond acceptor groups. *J. Phys. Chem. B*, 114(5):2083–2089, 2010.

[110] Ruhong Zhou, Xuhui Huang, Claudio J. Margulis, and Bruce J. Berne. Hydrophobic collapse in multidomain protein folding. *Science*, 305(5690):1605–1609, 2004.

[111] D. Thirumalai, Govardhan Reddy, and John E. Straub. Role of water in protein aggregation and amyloid polymorphism. *Accounts Chem. Res.*, 45(1):83–92, 2012.

[112] Alexander N. Drozdov, Alan Grossfield, and Rohit V. Pappu. Role of solvent in determining conformational preferences of alanine dipeptide in water. *J. Am. Chem. Soc.*, 126(8):2574–2581, 2004.

[113] Jayendran C. Rasaiah, Shekhar Garde, and Gerhard Hummer. Water in nonpolar confinement: From nanotubes to proteins and beyond. *Annu. Rev. Phys. Chem.*, 59(1):713–740, 2008.

[114] Jan Kubelka, James Hofrichter, and William A Eaton. The protein folding 'speed limit'. *Curr Opin Struct Biol*, 14(1):76–88, Feb 2004.

[115] I Bahar and Robert L Jernigan. Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. *J. Mol. Biol.*, 266(1):195–214, 1997.

[116] Luca Monticelli, Senthil K Kandasamy, Xavier Periole, Ronald G Larson, D. Peter Tieleman, and Siewert-Jan Marrink. The martini coarse-grained force field: Extension to proteins. *J Chem Theory Comput*, 4(5):819–834, May 2008.

[117] Amy Y Shih, Anton Arkhipov, Peter L Freddolino, and Klaus Schulten. Coarse grained protein- lipid model with application to lipoprotein particles. *J. Phys. Chem. B*, 110(8):3674–3684, 2006.

[118] Arnab Mukherjee and Biman Bagchi. Contact pair dynamics during folding of two small proteins: Chicken villin head piece and the alzheimer protein β-amyloid. *J. Chem. Phys.*, 120(3):1602–1612, 2004.

[119] Sai J Ganesan, Hongcheng Xu, and Silvina Matysiak. Effect of lipid head group interactions on membrane properties and membrane-induced cationic β-hairpin folding. *Phys. Chem. Chem. Phys.*, 18(27):17836–17850, 2016.

[120] Sai J. Ganesan and S. Matysiak. Role of backbone dipole interactions in the formation of secondary and supersecondary structures of proteins. *J. Chem. Theory Comput.*, 10(6):2569–2576, Jun 2014.

[121] Daniel L Ensign, Peter M Kasson, and Vijay S Pande. Heterogeneity even at the speed limit of folding: large-scale molecular dynamics study of a fast-folding variant of the villin headpiece. *J Mol Biol*, 374(3):806–816, Nov 2007.

[122] Peter L. Freddolino and Klaus Schulten. Common structural transitions in explicit-solvent simulations of villin headpiece folding. *Biophys. J.*, 97(8):2338–2347, 2009.

[123] E. M. Boczko and C. L. Brooks, III. First-principles calculation of the folding free energy of a three-helix bundle protein. *Science*, 269(5222):393–396, Jul 1995.

[124] Angel E. García and Jos N. Onuchic. Folding a protein in a computer: An atomic description of the folding/unfolding of protein a. *Proc. Natl. Acad. Sci. U.S.A.*, 100(24):13898–13903, 2003.

[125] P. J. Kraulis, P. Jonasson, P. A. Nygren, M. Uhlén, L. Jendeberg, B. Nilsson, and J. Kördel. The serum albumin-binding domain of streptococcal protein g is a three-helical bundle: a heteronuclear nmr study. *FEBS Lett.*, 378(2):190–194, Jan 1996.

[126] Michael Hulko, Franziska Berndt, Markus Gruber, Jürgen U. Linder, Vincent Truffault, Anita Schultz, Jörg Martin, Joachim E. Schultz, Andrei N. Lupas, and Murray Coles. The hamp domain structure implies helix rotation in transmembrane signaling. *Cell*, 126(5):929–940, Sep 2006.

[127] Ronald L. Koder, J L Ross Anderson, Lee A. Solomon, Konda S. Reddy, Christopher C. Moser, and P Leslie Dutton. Design and engineering of an o(2) transport protein. *Nature*, 458(7236):305–309, Mar 2009.

[128] Anil Korkut and Wayne A. Hendrickson. A force field for virtual atom molecular mechanics of proteins. *Proceedings of the National Academy of Sciences*, 106(37):15667–15672, 2009.

[129] Guoli Wang and Roland L. Dunbrack. Pisces: a protein sequence culling server. *Bioinformatics*, 19(12):1589–1591, 2003.

[130] M. S. Lehmann, T. F. Koetzle, and W. C. Hamilton. Precision neutron diffraction structure determination of protein and nucleic acid components. i. crystal and molecular structure of the amino acid l-alanine. *Journal of the American Chemical Society*, 94(8):2657–2660, 1972. PMID: 5017418.

[131] Scott Brown, Nicolas J. Fawzi, and Teresa Head-Gordon. Coarse-grained sequences for protein folding and design. *Proceedings of the National Academy of Sciences*, 100(19):10712–10717, 2003.

[132] F. Fogolari, G. Esposito, P. Viglino, and S. Cattarinussi. Modeling of polypeptide chains as c alpha chains, c alpha chains with c beta, and c alpha chains with ellipsoidal lateral chains. *Biophysical Journal*, 70(3):1183–1197, 1996.

[133] Z. Guo and D. Thirumalai. Kinetics and thermodynamics of folding of ade novodesigned four-helix bundle protein. *Journal of Molecular Biology*, 263(2):323 – 343, 1996.

[134] Satwik Kamtekar, Jarad M Schiffer, Huayu Xiong, Jennifer M Babik, and Michael H Hecht. Protein design by binary patterning of polar and nonpolar amino acids. *Science*, 262(5140):1680–1685, 1993.

[135] Michael W. West and Michael H. Hecht. Binary patterning of polar and nonpolar amino acids in the sequences and structures of native proteins. *Protein Science*, 4(10):2032–2039, 1995.

[136] O. B. Ptitsyn and A. A. Rashin. A model of myoglobin self-organization. *Biophys. Chem.*, 3(1):1–20, Feb 1975.

[137] M. Karplus and D. L. Weaver. Protein-folding dynamics. *Nature*, 260(5550):404–406, Apr 1976.

[138] P. S. Kim and R. L. Baldwin. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.*, 51:459–489, 1982.

[139] M. Karplus and D. L. Weaver. Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci.*, 3(4):650–668, Apr 1994.

[140] Valerie Daggett and Alan R. Fersht. Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.*, 28(1):18–25, Jan 2003.

[141] J. A. Schellman. The stability of hydrogen-bonded peptide structures in aqueous solution. *C. R. Trav. Lab. Carlsberg Chim.*, 29(14-15):230–259, 1955.

[142] W. Kauzmann. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.*, 14:1–63, 1959.

[143] Charles. Tanford. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J. Am. Chem. Soc.*, 84(22):4240–4247, November 1962.

[144] R. L. Baldwin. How does protein folding get started? *Trends Biochem. Sci.*, 14(7):291–294, Jul 1989.

[145] Lan Hua, Xuhui Huang, Pu Liu, Ruhong Zhou, and Bruce J. Berne. Nanoscale dewetting transition in protein complex folding. *J. Phys. Chem. B*, 111(30):9069–9077, 2007.

[146] Pu Liu, Xuhui Huang, Ruhong Zhou, and B. J. Berne. Observation of a dewetting transition in the collapse of the melittin tetramer. *Nature*, 437(7055):159–162, September 2005.

[147] Srivathsan Vembanur, Amish J. Patel, Sapna Sarupria, and Shekhar Garde. On the thermodynamics and kinetics of hydrophobic interactions at interfaces. *J. Phys. Chem. B*, 117(35):10261–10270, 2013.

[148] Payel Das, Divya Kapoor, Kevin T. Halloran, Ruhong Zhou, and C. Robert Matthews. Interplay between drying and stability of a tim barrel protein: A combined simulationexperimental study. *Journal of the American Chemical Society*, 135(5):1882–1890, 2013. PMID: 23293932.

[149] Evgeny I. Shakhnovich and Alexey V. Finkelstein. Theory of cooperative transitions in protein molecules. i. why denaturation of globular protein is a first-order phase transition. *Biopolymers*, 28(10):1667–1680, 1989.

[150] Thomas Kiefhaber, Alexander Labhardt, and Robert Baldwin. Direct nmr evidence for an intermediate preceding the rate-limiting stepin the unfolding of ribonuclease a. *Nature*, 375(6531):513–515, June 1995.

[151] Santosh Kumar Jha and Jayant B. Udgaonkar. Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. *Proc. Natl. Acad. Sci. U.S.A.*, 106(30):12289–12294, 2009.

[152] S D Hoeltzli and C Frieden. Stopped-flow nmr spectroscopy: real-time unfolding studies of 6-19f-tryptophan-labeled escherichia coli dihydrofolate reductase. *Proc. Natl. Acad. Sci. U.S.A.*, 92(20):9318–9322, September 1995.

[153] Andreas Reiner, Peter Henklein, and Thomas Kiefhaber. An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. *Proc. Natl. Acad. Sci. U.S.A.*, 107(11):4955–4960, 2010.

[154] SaswataSankar Sarkar, JayantB Udgaonkar, and Guruswamy Krishnamoorthy. Unfolding of a small protein proceeds via dry and wet globules and a solvated transition state. *Biophys. J.*, 105(10):2392–2402, September 2013.

[155] Lan Hua, Ruhong Zhou, D. Thirumalai, and B. J. Berne. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proc. Natl. Acad. Sci. U.S.A.*, 105(44):16928–16933, 2008.

[156] Amrita Dasgupta, Jayant B. Udgaonkar, and Payel Das. Multistage unfolding of an sh3 domain: An initial urea-filled dry molten globule precedes a wet molten globule with non-native structure. *J. Phys. Chem. B*, 118(24):6380–6392, 2014.

[157] Mary Munson, Lynne Regan, Ronan O'Brien, and Julian M Sturtevant. Redesigning the hydrophobic core of a four-helix-bundle protein. *Protein Science*, 3(11):2015–2022, 1994.

[158] Christian E Schafmeister, LJ Miercke, and Robert M Stroud. Structure at 2.5 a of a designed peptide that maintains solubility of membrane proteins. *Science*, 262(5134):734–738, 1993.

[159] Zachary A. Wood, Larry H. Weaver, Patrick H. Brown, Dorothy Beckett, and Brian W. Matthews. Co-repressor induced order and biotin repressor dimerization: A case for divergent followed by convergent evolution. *Journal of Molecular Biology*, 357(2):509–523, 2006.

[160] David F. Barker and Allan M. Campbell. Genetic and biochemical characterization of the bira gene and its product: Evidence for a direct role of biotin holoenzyme synthetase in repression of the biotin operon in escherichia coli. *Journal of Molecular Biology*, 146(4):469–492, 1981.

[161] M.A. Eisenberg, O. Prakash, and S.C. Hsiung. Purification and properties of the biotin repressor. a bifunctional protein. *Journal of Biological Chemistry*, 257(24):15167–15173, 1982.

[162] A. Chapman-Smith and J.E. Cronan Jr. The enzymatic biotinylation of proteins: A post-translational modification of exceptional specificity. *Trends in Biochemical Sciences*, 24(9):359–363, 1999.

[163] E. Eisenstein and D. Beckett. Dimerization of the escherichia coli biotin repressor: Corepressor function in protein assembly. *Biochemistry*, 38(40):13077–13084, 1999.

[164] E.D. Streaker and D. Beckett. Coupling of protein assembly and dna binding: Biotin repressor dimerization precedes biotin operator binding. *Journal of Molecular Biology*, 325(5):937–948, 2003.

[165] Emily D Streaker, Aditi Gupta, and Dorothy Beckett. The biotin repressor: thermodynamic coupling of corepressor binding, protein assembly, and sequence-specific dna binding. *Biochemistry*, 41(48):14263–14271, Dec 2002.

[166] Vincent J. Hilser and E. Brad Thompson. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proceedings of the National Academy of Sciences*, 104(20):8311–8315, 2007.

[167] Ravi S. Ampapathi, Andrea L. Creath, Dianne I. Lou, John W. Craft, Steven R. Blanke, and Glen B. Legge. Orderdisorderorder transitions mediate the activation of cholera toxin. *Journal of Molecular Biology*, 377(3):748 – 760, 2008.

[168] Sean E. Reichheld, Zhou Yu, and Alan R. Davidson. The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor. *Proceedings of the National Academy of Sciences*, 106(52):22263–22268, 2009.

[169] Mioara Larion, Roberto Kopke Salinas, Lei Bruschweiler-Li, Brian G. Miller, and Rafael Brschweiler. Order-disorder transitions govern kinetic cooperativity and allostery of monomeric human glucokinase. *PLoS Biology*, 10(12):e1001452, November 2012.

[170] L. Michel Espinoza-Fonseca, David Kast, and David D. Thomas. Molecular dynamics simulations reveal a disorder-to-order transition on phosphorylation of smooth muscle myosin. *Biophysical Journal*, 93(6):2083 – 2090, 2007.

[171] MF Colombo, DC Rau, and VA Parsegian. Protein solvation in allosteric regulation: a water effect on hemoglobin. *Science*, 256(5057):655–659, 1992.

[172] Yaoqi Zhou, Hongyi Zhou, and Martin Karplus. Cooperativity in scapharca dimeric hemoglobin: Simulation of binding intermediates and elucidation of the role of interfacial water. *Journal of Molecular Biology*, 326(2):593 – 606, 2003.

[173] Schrodinger, LLC. The PyMOL molecular graphics system, version 1.3r1. August 2010.

[174] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *The Journal of Physical Chemistry*, 91(24):6269–6271, 1987.

[175] William L. Jorgensen, David S. Maxwell, , and Julian Tirado-Rive. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, 1996.

[176] Justin A Lemkul, William J Allen, and David R Bevan. Practical considerations for building gromos-compatible small-molecule topologies. *Journal of chemical information and modeling*, 50(12):2221–2235, 2010.

[177] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. Lincs: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.

[178] Roger Armen, Darwin OV Alonso, and Valerie Daggett. The role of α-, 310-, and π-helix in helix coil transitions. *Protein Science*, 12(6):1145–1157, 2003.

[179] K P Wilson, L M Shewchuk, R G Brennan, A J Otsuka, and B W Matthews. Escherichia coli biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and dna-binding domains. *Proceedings of the National Academy of Sciences*, 89(19):9257–9261, 1992.

[180] Huaying Zhao, Emily Streaker, Weilan Pan, and Dorothy Beckett. Protein-protein interactions dominate the assembly thermodynamics of a transcription repression complex. *Biochemistry*, 46(47):13667–13676, 2007.

[181] Poorni R. Adikaram and Dorothy Beckett. Functional versatility of a single protein surface in two protein:protein interactions. *Journal of Molecular Biology*, 419(34):223 – 233, 2012.

[182] Christopher Eginton, Saranga Naganathan, and Dorothy Beckett. Sequence-function relationships in folding upon binding. *Protein Sci*, 24(2):200–211, Feb 2015.

[183] Christopher Eginton, William J. Cressman, Sharrol Bachas, Herschel Wade, and Dorothy Beckett. Allosteric coupling via distant disorder-to-order transitions. *Journal of Molecular Biology*, 427(8):1695–1704, 2015.

[184] Jingheng Wang, Gregory Custer, Dorothy Beckett, and Silvina Matysiak. Long distance modulation of disorder-to-order transitions in protein allostery. *Biochemistry*, 56(34):4478–4488, 2017.

[185] Keehwan Kwon, Emily D Streaker, Shreyesh Ruparelia, and Dorothy Beckett. Multiple disordered loops function in corepressor-induced dimerization of the biotin repressor. *Journal of molecular biology*, 304(5):821–833, 2000.

[186] Vandana Chakravartty and John E. Cronan. Altered regulation of escherichia coli biotin biosynthesis in bira superrepressor mutant strains. *Journal of Bacteriology*, 194(5):1113–1126, 2012.

[187] Poorni R. Adikaram and Dorothy Beckett. Protein:protein interactions in control of a transcriptional switch. *Journal of Molecular Biology*, 425(22):4584 – 4594, 2013.

[188] Chenlu He, Gregory Custer, Jingheng Wang, Silvina Matysiak, and Dorothy Beckett. Superrepression through altered corepressoractivated protein:protein interactions. *Biochemistry*, 57(7):1119–1129, 2018. PMID: 29355305.

[189] Christopher Eginton and Dorothy Beckett. A large solvent isotope effect on protein association thermodynamics. *Biochemistry*, 52(38):6595–6600, Sep 2013.

[190] Lee A. Freiburger, Oliver M. Baettig, Tara Sprules, Albert M. Berghuis, Karine Auclair, and Anthony K. Mittermaier. Competing allosteric mechanisms modulate substrate binding in a dimeric enzyme. *Nature Structural &Amp; Molecular Biology*, 18:288, January 2011.

[191] Hesam N. Motlagh, James O. Wrabl, Jing Li, and Vincent J. Hilser. The ensemble nature of allostery. *Nature*, 508:331, April 2014.

[192] Monika Davidovic, Carlos Mattea, Johan Qvist, and Bertil Halle. Protein cold denaturation as seen from the solvent. *J Am Chem Soc*, 131(3):1025–1036, Jan 2009.