MOLECULAR SIMULATIONS OF ALTERNATE FRAME FOLDING IN ENGINEERED PROTEIN-BASED SWITCHES

by

Brandon Michael Mills

B. Phil., University of Pittsburgh, 2009

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Brandon M. Mills

It was defended on

April 24, 2012

and approved by

Rob D. Coalson, Professor, Chemistry

Sean A. Garrett-Roe, Assistant Professor, Chemistry

Thesis Advisor: Lillian T. Chong, Associate Professor, Chemistry

ACKNOWLEDGEMENTS

This document is submitted for the approval of the comprehensive examination committee: Profs. Rob Coalson, Seth Horne, and Lillian Chong, whom I thank for their service. I especially thank Prof. Chong for her service as my research advisor, and my fellow members of the Chong lab for insightful discussions.

COMPUTATIONAL STUDIES OF MUTUALLY EXCLUSIVE FOLDING IN A TWO-DOMAIN MOLECULAR SWITCH

Brandon Michael Mills, M. S.

University of Pittsburgh, 2014

Living organisms take advantage of proteins in order to carry out most of the biological tasks that keep them alive. Although most proteins do not drastically change shape, some behave as conformational switches: in response to an outside signal from its environment, the protein will shape-change to either an "on" or "off" position. These properties of conformational switches have motivated recent efforts towards the conversion of regular ligand binding proteins into novel switches for use as optical sensors and therapeutics. Here we seek to examine one such design that exhibits an intermolecular tug-of-war between two alternate frames of folding that can be made sensitive to calcium. The challenges of elucidating structural and mechanistic details from a partially unfolded protein have led us to consider coarse-grained simulation techniques. We plan to demonstrate that results from these simulations are in agreement with experimental data and can provide novel insight into the mechanisms of switching in this class of engineered proteins.

TABLE OF CONTENTS

1.0 INTRODUCTION 1

1.1 OVERVIEW 1

1.2 OBJECTIVE 3

2.0 BACKGROUND 4

2.1 THE ALTERNATE FRAME FOLDING DESIGN STRATEGY 4

2.2 OUR MODEL SYSTEM: THE CALBINDIN ALTERNATE FRAME FOLDING CONSTRUCT (CALBINDIN-AFF) 6

3.0 SIMULATION APPROACH 8

3.1 THE PROTEIN MODEL 8

3.2 PARAMETERIZATION 10

3.3 SIMULATION DETAILS 11

3.4 THE WEIGHTED ENSEMBLE PATH SAMPLING ALGORITHM 12

4.0 PROJECT PROPOSAL 16

4.1 PRELIMINARY STUDIES 17

4.1.1 Feasibility of simulating alternate frame folding. 17

4.1.2 Feasibility of simulating cooperative effects. 19

4.1.3 Agreement of weighted ensemble and brute force simulations. 20

4.2 RESEARCH PLAN AND TIMELINE 22

4.2.1 Parameterize the protein model to reproduce experimental folding free energy of calbindin (Months 1-3). 22

4.2.2 Test the ability of our protein model to capture the effect of permutation on its thermodynamic stability (Month 4). 24

4.2.3 Simulate alternate frame folding of the calbindin construct (Month 5). ... 25

4.2.4 Model the binding of calcium ions to calbindin and its circular permutant (Months 6-8). 26

4.2.5 Model the binding of calcium ions to the E65Q-1 and E65Q-2 mutants of the calbindin construct (Months 9-10). 28

4.3 EXPECTED OUTCOMES 29

4.4 POTENTIAL PROBLEMS & ALTERNATIVE STRATEGIES 30

4.4.1 Lack of favorable nonnative interactions. 30

4.4.2 Insufficient sampling for convergence of holo calbindin simulations. 30

4.4.3 Discrepant experimental and theoretical stabilization of calbindin by calcium. 31

5.0 FUTURE DIRECTIONS 32

BIBLIOGRAPHY 34

LIST OF FIGURES

Figure 1: The two types of mutually exclusive folding. 2

Figure 2: Calbindin. 7

Figure 3: Schematic diagram of a weighted ensemble simulation. 13

Figure 4: Potential of mean force surfaces for calbindin-AFF constructs. 18

Figure 5. Foldedness of the C-terminal EF hand in Calbindin-AFF and in isolation. 19

Figure 6: Comparison of weighted ensemble and brute force simulation results. 21

1.0 INTRODUCTION

1.1 OVERVIEW

Nature has produced numerous examples of biosensors that allow cells to respond appropriately to their environment [1, 2]. Among these, allosteric proteins stand out as effectual, autonomous sensors, capable of rapidly switching themselves on and off in response to their signal of interest [3]. Emboldened by the possibility of creating new sensor molecules, bioengineers have recently endeavored to introduce novel allostery into proteins through various strategies of structural modification [4-7]. One design paradigm put forth by Prof. Stewart Loh (SUNY Upstate Medical University) is particularly interesting in that it seeks to exploit the specificity inherent to protein folding [8]. This hypothesis (dubbed mutually exclusive folding) proposes that certain modifications of non-allosteric proteins may contrive an intramolecular, thermodynamic tug-of-war between two folded states that are mutually exclusive. This situation can be engineered in one of two ways: through domain insertion, where a "guest" protein is inserted into a "host" such that folding of one disrupts the folding of the other [Figure 1A] [8-11], or alternate frame folding, where a protein is fused with its circular permutant in a region of overlapping sequence causing the folding of the permutant to displace the non-permutant at the overlap site and vice versa [Figure 1B] [12-15].

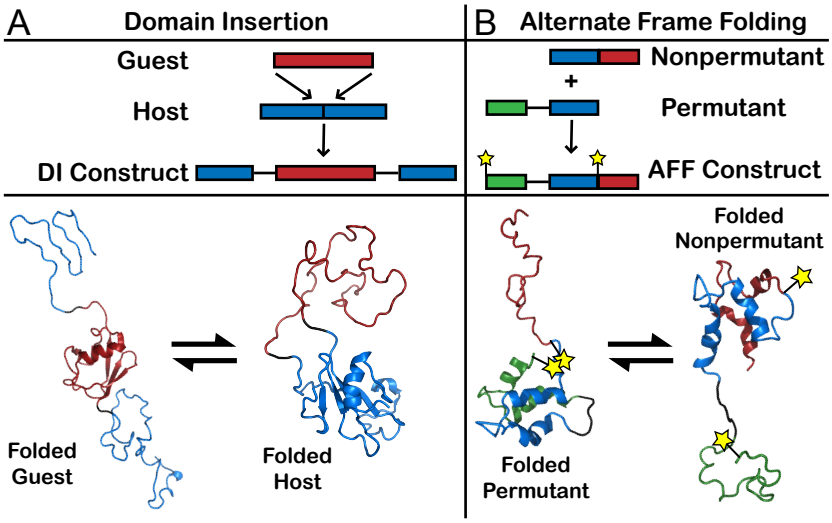

Figure 1: The two types of mutually exclusive folding. In the case of domain insertion (A), the folding of one domain exerts an expansive or compressive force on the other. Under alternate frame folding (B), sterics prevent both frames from folding at the same time. Attaching pyrene moieties (stars) allows for optical signaling of the fold state via excimer fluorescence.

Trial implementations of these basic design paradigms have shown promise. Early model systems have either produced modest success or have served as case studies for some of the pitfalls that may arise. However, the creation of new allosteric proteins still requires significant trial-and-error, partly due to the challenges of working with proteins that are, by design, partially unstructured [11]. The dearth of high-resolution structural or mechanistic data for these models further hinders our ability to understand the inner workings of these prototypes, and thus to make rational modifications to existing design strategies. On the other hand, molecular simulation techniques allow for the direct visualization of protein conformational changes, and therefore may provide insight into engineered allosteric switching events that cannot be obtained through experiment alone.

1.2 OBJECTIVE

Our research objective will be to explore the mechanism of alternate frame folding at residue-level detail. Towards this end, we will employ molecular simulation techniques to examine a calcium sensor designed by applying the alternate frame folding (AFF) paradigm, dubbed calbindin-AFF. Experiments such as circular dichroism, NMR and fluorescence spectroscopic methods as well as thiol-disulfide exchange chemistry have provided interesting thermodynamic and kinetic data on the switching of calbindin-AFF [12-14], but little is known about the molecular mechanism of switching. We will simulate this protein at the temperature of experiments, validate the simulations with experimental data, and then examine the molecular details of the observed switching mechanism.

2.0 BACKGROUND

2.1 THE ALTERNATE FRAME FOLDING DESIGN STRATEGY

In surveying the proteome, we find that the number of targets that can be specifically bound by proteins is very broad, from as small as a single-atom ion (like calcium [16]) to as large as another protein (like the barnase-barstar complex [17]). How can we take a protein-ligand interaction and use it to make a sensor for that ligand? Any method we could surmise would require that the protein-ligand interaction beget some sort of change that can be monitored. For glucose oxidase, the enzyme used in blood glucose monitoring by millions of diabetics everyday and the center of a $5 billion industry [18], that change is the production of hydrogen peroxide, measured with electrochemical methods. Many proteins, however, do not produce observable changes in response to ligand binding. It would appear advantageous to use some of these proteins as starting points in sensor design, as nature has already done much of our work for us: the challenge becomes coaxing these proteins that already excel at binding their target into informing us that they have done so. This is what the alternate frame folding design strategies seeks to achieve.

Alternate frame folding is predicated on a number of established ideas about proteins. All proteins undergo conformational changes, some of which are discernable using spectroscopic methods. Folding and unfolding events are both the most common and the most dramatic of

protein conformational changes; thus, contriving these types of conformational changes in response to binding would seem to be a promising endeavor towards the production of new sensors. Indeed, such an observable transition has been engineered into the SH3 domain of Fyn tyrosine kinase by removal of amino acids from its C-terminus [19]. However, any method that looks to utilize unfolded proteins as detectors must overcome the challenges of aggregation and oligmerization, both of which would be detrimental to switch activity.

Another concept that forms the basis for alternate frame folding is that of circular permutation of protein topology. By cleaving one terminal end of a protein and reattaching it to the alternate terminus, a new protein is produced with the same amino acids now in a slightly different order; for example, a sequence of residues ABCDEF may be permuted to DEFABC. Many circularly permuted proteins can fold to produce similar structure as proteins with nonpermuted topology [20-22], the net effect of which is to alter the location of the N- and C-termini. Some circular permutants arise naturally through evolutionary mechanisms [23] or even via posttranslational modification [24], though engineered circular permutants have been made to explore the relationship between topology and other properties of proteins [25-27]. Often a flexible linker sequence is included, as the N- and C- termini of the original protein may not be near each other. While structurally the nonpermuted and permuted folds may be very similar, often the permuted fold will be slightly less thermodynamically stable.

Alternate frame folding seeks to duplicate one terminal sequence of residues at the alternate terminus, making both a permuted and nonpermuted fold state available to the protein [Figure 1B]. These fold states will be mutually exclusive, as they must compete over the sequence of residues they have in common (the blue residues in Figure 1B). The result is an intramolecular tug-of-war between two highly similar folded states, where unfolding and folding

can occur to a large but equal degree. This tug-of-war is likely to be a fair fight, with the nonpermuted form favored slightly due to the thermodynamic penalty common to circularly permuted folds.

By appending an unstructured region, at most about half as large as the original protein, alternate frame folding causes an immense conformational change, but this change is not naturally sensitive for a ligand. However, if a protein's ligand binding site is part of the duplicated region, we can then perform knockout mutations to preventing binding in one folding frame. Thus, the switch equilibrium becomes sensitive to the presence of the target ligand. Other mutations can be performed to each frame if adjustments to the apo switch equilibrium are desired.

2.2 OUR MODEL SYSTEM: THE CALBINDIN ALTERNATE FRAME FOLDING CONSTRUCT (CALBINDIN-AFF)

We have chosen as our model system a calcium-dependent switch created from calbindin D_{9k}, a 75-residue protein that plays a role in calcium transport in mammals [**Error! Reference source ot found.**] [28]. Calbindin is part of a large family of proteins that exhibit the EF-hand calcium-binding motif, itself being entirely comprised of two adjoined EF-hands [16]. Unlike its related signaling brethren including calmodulin, calbindin does not undergo a conformational change in response to calcium binding. However, through the application of the alternate frame folding design strategy [Figure 1B], calbindin has been converted into a calcium sensor (dubbed calbindin-AFF) whose switching can be monitored via fluorescence using either pyrene or boron-dipyrromethene (BODIPY) [12, 13]. These experiments have shown that switching occurs

on the timescale of seconds and is sensitive to micromolar concentrations of calcium. CD spectroscopy data shows that the binding of calcium results in zero net folding, suggesting that the unfolding of one frame and the folding of the other occur in concert. In addition, NMR HSQC spectra indicate that the native and permuted folds of the calbindin-AFF construct are similar in structure to calbindin and each other [14]. This experimental data and others allow for comparison to the thermodynamics and kinetics that will be observed in simulations.

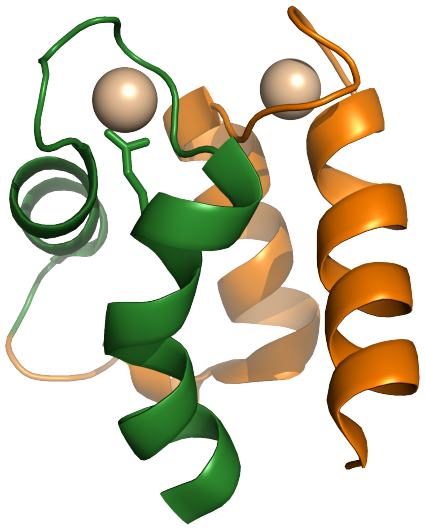

Figure 2: Calbindin. The two EF hands of calbindin, residues 1-43 (orange) and residues 44-75 (green), each bind a single calcium ion (brown). Calbindin-AFF is made by duplicating residues 44-75 and affixing them to the calbindin N-terminus with a six-residue linker. Also shown is the side chain of Glu65, an important residue in calcium binding. Mutation of this residue to glutamine reduces calcium affinity at that site. Coordinates taken from X-ray crystal structure 3ICB [29].

3.0 SIMULATION APPROACH

Although all-atom simulations offer the most detailed views of protein dynamics, they would be computationally prohibitive to studying the seconds-scale conformational switching of calbindin-AFF. However, coarse-grained simulations at residue-level structural detail and using a Gō-type force field have been used to examine similar conformational switching events [30]. These models unify each residue's many atoms into one single "pseudo-atom," and allow the pseudo-atoms to repel or attract one another based on their contacts in the native state.

3.1 THE PROTEIN MODEL

Each protein has been modeled at the residue level, using a single pseudo-atom at the C_α position to represent each amino acid. The folded state model for the calbindin protein was generated from the X-ray crystal structure 3ICB [29]. Models for each protein's unfolded state were taken from statistical coil conformations generated by the Unfolded State Server [31]. Additionally, a model for a fully folded circular permutant of calbindin as well as models of the Calbindin-AFF switch designs that are folded with respect to each possible folding frame were generated using the MODELLER homology modeling software [32, 33]. The current models utilize amino acid sequences that have been examined experimentally [12], though other circular permutants and calbindin-AFF constructs may also be modeled in identical fashion.

The conformational energetics of these protein models are governed by a Gō-type potential energy function similar to that described by others [34, 35]. This energetic model differentiates in its treatment of bonded vs. nonbonded residue pairs, with bonded pairs governed by standard molecular mechanics terms:

$$E_{bonded} = \sum_{bonds} k_{bond}\left(r - r_{eq}\right)^2 + \sum_{angles} k_{angle}\left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} V_1\left[1+\cos\left(\varphi - \varphi_1\right)\right] + V_3\left[1+\cos\left(3\varphi - \varphi_3\right)\right]$$

where r, θ, and φ are pseudo-bond lengths, pseudo-angles, and pseudo-dihedrals, respectively; V_1 and V_3 are potential barriers for the dihedral terms. Equilibrium bond lengths r_{eq}, angles θ_{eq}, and dihedral phase angles, φ_1 and φ_3, were taken from the crystal structure mentioned above. The force constants k_{bond} and k_{angle} were set to 20 kcal/mol/Å and 10 kcal/mol/rad, respectively. The values for V_1 and V_3 are adjustable parameters of our simulations that are usually adjusted to reproduce experimental protein stability data. For regions where there is no detailed structural information from experiment, such as the linking region of a circular permutant, V_1 and V_3 are set to zero, allowing free rotation.

Assigning energies to nonbonded residue pair interactions is slightly more complicated, as there are two different interaction potentials. Prior to simulation, the folded crystal structure for calbindin was examined and all nonbonded residue pairs were assigned to one of two sets: native or nonnative. The interaction between two residues was only considered to be native if any of their heavy atoms are within 5.5 Å of each other in the crystal structure. Additionally, a minimum residue separation of 4 was required for a pair interaction to be considered native. Although no experimentally determined structures are available for the calbindin circular permutant or the AFF constructs, it is reasonable to assume that the new residue ordering that

arises from permutation will not significantly alter the folded state; this allows us to apply the native pair contact lists from calbindin to the residues of the circular permutant and AFF constructs. Native contacts were modeled using a Lennard-Jones-like potential:

$$E_{ij}^{native} = \varepsilon^{native}\left[5\left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}}\right) - 6\left(\frac{\sigma_{ij}^{10}}{r_{ij}^{10}}\right)\right]$$

where ε^{native} is the energy well depth for the interaction, r_{ij} is the distance between residues i and j in the simulation, and σ_{ij} is the distance between residues i and j in the corresponding crystal structure of the native state. In addition to V_1 and V_3 in the bonded potential, ε^{native} is adjusted to reproduce experimental protein stability data (see M1.3: Parameterization).

Non-native contacts were modeled using a purely repulsive potential:

$$E_{ij}^{non-native} = \varepsilon^{non-native}\left[\frac{\sigma_{ij}^{12}}{r_{ij}^{12}}\right]$$

where σ_{ij} and $\varepsilon^{non\text{-}native}$ are set to 4.0 Å and 0.60 kcal/mol, respectively.

3.2 PARAMETERIZATION

Our three energetic parameters, ε^{native}, V_1 and V_3, determine the conformational thermodynamics of our modeled proteins, with higher values encouraging more native-like conformations. Additionally, the balance between bonded and nonbonded interactions influences folding cooperativity, and an optimal ratio for ε^{native} : V_1 : V_3 of 0.60 kcal/mol : 0.50 kcal/mol : 0.25 kcal/mol is used [35], effectively reducing our adjustable parameters from three to one. In preliminary work, we utilized experimental T_m values and adjusted our parameter until

simulations exhibit melting temperature behavior (equal populations of folded and unfolded states). T_m values of 85 °C and 100 °C for apo and holo calbindin, respectively, were used [36].

Simulating at a much higher temperature than the experiments is not ideal, but parameterizing to match experimental folding free energy values for calbindin at 20 °C (where ΔG_{fold} = -3.9 kcal/mol) is much more challenging than reproducing melting temperature behavior. In addition, ΔG_{fold} values may go as high as -8.3 kcal/mol in the case of calbindin in the presence of calcium. Therefore, we will employ the weighted ensemble algorithm in order to meet this challenge.

3.3 SIMULATION DETAILS

Simulations are performed using a standard Brownian Dynamics algorithm developed by Ermak and McCammon [37] with hydrodynamic interactions [38]; the time evolution of such a simulation can be formulated as the following equation:

$$\boldsymbol{r_i}(t + \Delta t) = \boldsymbol{r_i}(t) + \sum_j \left(\frac{\Delta t}{k_b T} \Delta \mathbf{v}_{ij} \right) + \boldsymbol{R}_i$$

where $\Delta \mathbf{v}_{ij}$ is the velocity perturbation of pseudo-atom i that results from the force on pseudo-atom j, and $\boldsymbol{R}_i$ is the random displacement applied to pseudo-atom i. The $\Delta \mathbf{v}_{ij}$ term is the product of a diffusion tensor with the force on pseudo-atom j. When i=j, this diffusion tensor comes from the Stokes-Einstein equation. However, to determine the velocity perturbation on one pseudo-atom that results from force acting on another pseudo-atom, we use an approximation of the Oseen tensor developed by Rotne, Prager [39], and Yamikawa [40]:

$$\Delta \mathbf{v}_{ii} = \left[\frac{k_B T}{6\pi\eta_s a}\right] \boldsymbol{I} \cdot \boldsymbol{F}_i$$

$$\Delta \mathbf{v}_{ij} = \left[\frac{k_B T}{8\pi\eta_s}\right]\left\{\left(\frac{1}{r_{ij}}\right)\left[\left(\frac{1+2a^2}{3r_{ij}^2}\right)\boldsymbol{I} + \left(\frac{1-2a^2}{r_{ij}^2}\right)\left(\frac{\vec{r}_{ij}\cdot\vec{r}_{ij}}{r_{ij}^2}\right)\right]\right\} \cdot \boldsymbol{F}_j \text{ for } r_{ij} \geq 2a$$

$$\Delta \mathbf{v}_{ij} = \left[\frac{k_B T}{8\pi\eta_s}\right]\left\{\left(\frac{1}{r_{ij}}\right)\left[\left(\frac{r_{ij}}{2a}\right)\left(\frac{8}{3} - \frac{3r_{ij}}{4a}\right)\boldsymbol{I} + \left(\frac{r_{ij}}{4a}\right)\left(\frac{\vec{r}_{ij}\cdot\vec{r}_{ij}}{r_{ij}^2}\right)\right]\right\} \cdot \boldsymbol{F}_j \text{ for } r_{ij} < 2a$$

where η_s is the dynamic viscosity of the solvent (water), **a** is the pseudo-atom hydrodynamic radius, $\boldsymbol{I}$ is a unit 3x3 matrix (1 for diagonal elements, 0 for all others), $\vec{r}_{ij}$ is the position vector connecting pseudo-atoms *i* and *j* while r_{ij} is the length of this vector, and $\boldsymbol{F}_j$ is the 3x1 force vector for pseudo-atom j.

For alpha carbon models, a hydrodynamic radius of 5.3 Å was found to give the best agreement with fully atomic models with regards to the translation diffusion coefficient [38] as computed by HYDROPRO [41, 42].

3.4 THE WEIGHTED ENSEMBLE PATH SAMPLING ALGORITHM

To reduce the computational effort required to sample fast-but-rare events such as protein folding and unfolding at 20 °C, we will employ the weighted ensemble path sampling algorithm [43]. The weighted ensemble algorithm works via "statistical ratcheting:" by replicating trajectories as they make small progress up the relevant free energy barrier and halting those that regress, we can produce more attempted transitions and also favor those trajectories that are making progress. In order to obtain relevant thermodynamic and kinetic information from such simulations, the statistical weight of trajectory is tracked throughout a weighted ensemble

simulation: simulations that are replicated will divide their initial probability among all of the new simulations, while those that are halted have their probability assigned to another trajectory in a similar region of the progress coordinate space. Thus, seemingly accelerated kinetics for barrier transition are counterbalanced by the small probability carried by a successfully transitioned trajectory. Similarly for thermodynamics, rather than tracking the movement of particles through progress coordinate space to measure a potential of mean force, we instead track the movement of probability carried by particles.

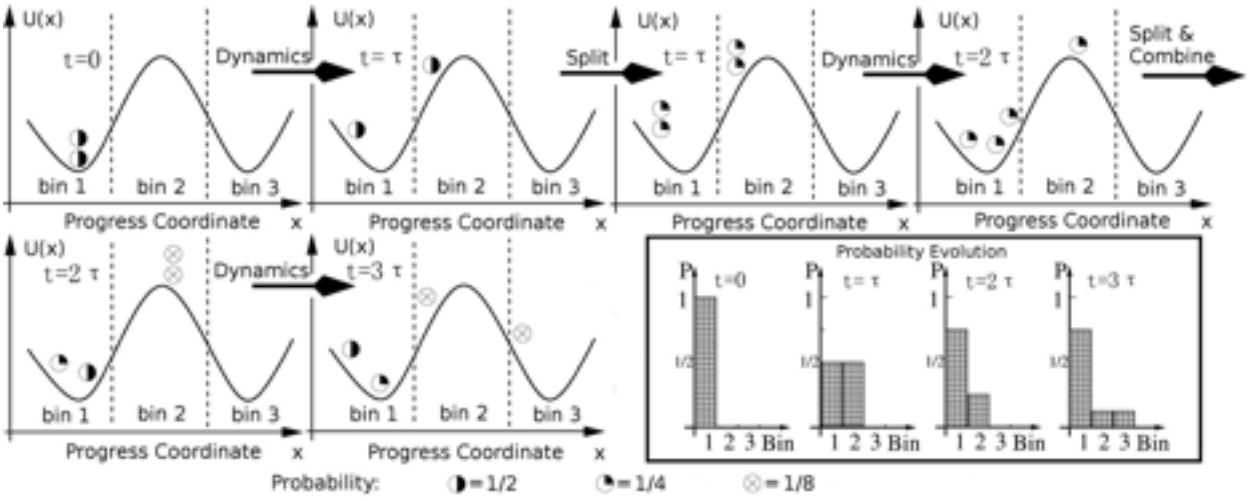

Figure 3: Schematic diagram of a weighted ensemble simulation. Figure reproduced from [44]. This figure represents a possible weighted ensemble simulation with three bins and two trajectories per bin. After one **τ** of unbiased dynamics, one trajectory has passed a bin boundary, leaving only one per bin. Therefore, for each initial trajectory, another trajectory with identical coordinates and velocities is replicated; these new replicas bring the particles per bin back up to the desired two and now carry half of the probability originally carried by their "parents." After another **τ**, a trajectory may regress and overpopulate a bin. To bring the number of particles back down to two, one of the trajectories will be randomly halted and its probability added at random to another trajectory from the bin. After the third, one of the four trajectories has carried 1/8 of the total probability into the 3rd final bin. This demonstrates the statistical ratcheting effect: weighted ensemble generates many more attempts to cross a barrier, but a trajectory that succeeds in crossing it will lose probability along the way such that the true rate of

barrier crossing can be obtained by taking rate of a weighted ensemble transition and multiplying it by its final probability.

The first step in setting up a weighted ensemble simulation is to define a coordinate that reasonably approximates progress over an energetic barrier (e.g., for folding and unfolding transitions, RMSD to a native structure or the fraction of native contacts). That progress coordinate is then divided into bins such that a small fraction of simulations in one bin can be expected to progress to another bin after a short period of unbiased dynamics τ (thus, a simulating attempting to traverse a large barrier will benefit from a finer bin spacing than a small barrier), and a desired number of particles per bin is specified (more particles per bin allows for a greater exploration of the conformational space within that region of progress coordinate space). It is important to note that these considerations do not bias the simulation dynamics but rather adjust the extent of sampling for each region of the progress coordinate space. Careful choices of these parameters will allow weighted ensemble simulations to converge as fast as possible. In these simulations, convergence can be most easily defined as the point when the rate of probability entering and leaving each particular bin becomes equal, i.e. the probability distribution along the progress coordinate becomes static. We will use as our progress coordinate the fraction of native contacts, as formation of these contacts is the sole nonbonded attractive force in our minimal model. We will develop a protocol to determine proper bin spacing and particles per bin in the first month of this project.

After these settings are determined, the final step before simulation is to seed the progress coordinate with starting structures and probabilities. This can be as simple as starting from a single structure or as complicated as supplying an ensemble of structures for each bin and specifying variable starting probability on each bin. The closer your starting probability distribution is to the "true" probability distribution, the faster simulations will converge. We will

seed our first probability distributions for 20 °C simulations using results obtained from 85 °C simulations.

4.0 PROJECT PROPOSAL

The aims of this project are to 1) demonstrate that the behavior of the calbindin alternate frame folding construct (calbindin-AFF) observed using the aforementioned simulation protocol is supported by available experimental data, and 2) examine the molecular details of the observed switching mechanism.

Since much of the experimental data we will be using for comparison was collected at 20 °C, that will also be our simulation temperature. These simulations follow up on successful feasibility tests conducted at the melting temperature of isolated calbindin, 85 °C. Proper thermodynamic parameters for each folding frame of calbindin-AFF will be determined through simulations of the switch components: calbindin and its circular permutant. We will first model switching of the wild type calbindin-AFF construct in the absence of calcium; then, we will model calcium-induced switching of two mutant constructs, E65Q-1 and E65Q-2, which only possess intact calcium binding sites in their permuted and nonpermuted frames, respectively.

4.1 PRELIMINARY STUDIES

4.1.1 Feasibility of simulating alternate frame folding.

To determine if our simulation model was capable of capturing switching events in alternate frame folding proteins, we performed simulations of the calbindin-AFF protein at the melting temperature of "apo" calbindin, 85 °C. Stability parameters for contacts within each folding frame were first determined by simulating calbindin and its circular permutant at 85 °C with different stability parameters until each reproduced approximately equal populations of unfolded and folded conformations.. This process was repeated to parameterize melting temperatures of 100 °C for the calcium-bound (or "holo") forms of calbindin and its circular permutant as well. We then conducted three simulations of calbindin-AFF, all at 85 °C: 1) applying the apo parameters to both folding frames, representing the calbindin-AFF case with no mutations and no calcium present, 2) applying the apo parameter to the permuted frame and the holo parameter to the nonpermuted frame, representing the E65Q-2 case where calcium is present and an E65Q mutation prevents calcium from binding to the permuted frame, and 3) applying the holo parameter to the permuted frame and the apo parameter to the nonpermuted frame, representing the E65Q-1 case where calcium is present and an E65Q mutation in the nonpermuted frame prevents calcium binding. Each frame is about equally likely to fold in the wild type construct [Figure 4A], while the mutants exhibit a shifted switching equilibrium in the expected directions i.e. towards the nonpermuted frame for E65Q-2 [Figure 4B] and towards the permuted frame for E65Q-1 [Figure 4C]. Most strikingly, we find that in all three cases that folding is indeed mutually exclusive, in agreement with circular dichroism and NMR HSQC spectroscopy data [12, 14]. Additionally, this data suggests that switching events in both directions begin with

global unfolding, though this is likely an artifact of the high simulation temperature. Experiments suggest that switching at 20 °C involves partial unfolding but not global unfolding [13], and parameterization of our model to reproduce experimental stabilities at 20 °C can be expected to make the unfolded state a great deal less favorable.

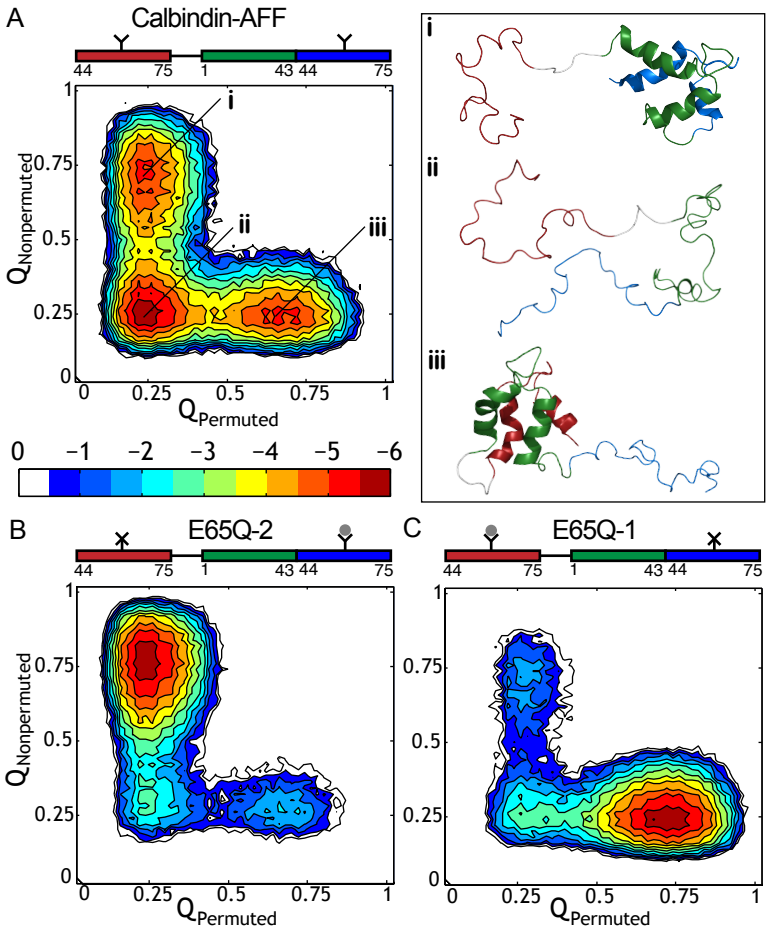

Figure 4: Potential of mean force (PMF) surfaces for calbindin-AFF constructs as a function of the fraction of native contacts in the nonpermuted ($Q_{Nonpermuted}$) and permuted frames ($Q_{Permuted}$). A) wild-type construct in the absence of calcium ions. B) E65Q-2 construct with calcium stabilizing the nonpermuted frame. C) E65Q-1 construct with calcium stabilizing the permuted frame. Representative conformations of the major states sampled are provided in A) (i, ii, and iii). Each PMF is based on five independent 4-μs brute force simulations at the T_m of apo-calbindin (85°C). Contours are drawn at intervals of the available thermal energy, 0.5RT.

4.1.2 Feasibility of simulating cooperative effects.

In this project, we hope to model calcium binding-induced folding of calbindin using explicit ions. Since that will require the stabilizing effect of binding to transmit to some degree through the protein, it would be beneficial to show that folding cooperativity is a consequence of our model. Some evidence of cooperativity has already been demonstrated: simulations of truncated barnase fragments showed a surprising agreement between the average fraction of native contacts observed [35] and experimental fluorescence intensities [45]; additionally, we showed previously that the binding of barstar to barnase stabilized the entire domain while also destabilizing an attached ubiquitin molecule [30]. We chose to further explore cooperativity of folding by simulating the 32-residue EF-hand of the calbindin N-terminus that is duplicated and appended to the calbindin C-terminus in calbindin-AFF. We hypothesized that the isolated polypeptide would be unable to fold in the absence of the rest of the calbindin protein, as seen in CD spectroscopy experiments [36].

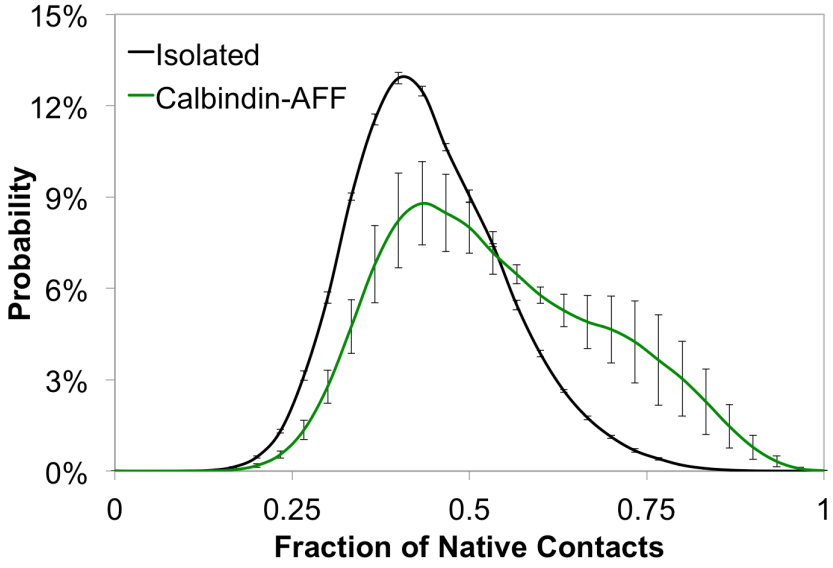

Figure 5. Foldedness of the C-terminal EF hand in Calbindin-AFF (green) and in isolation (black) at 85 °C. In order to make a fair comparison, this plot examines only the internal contacts of the C-terminal

polypeptide. Each data set shown is based on five independent 4-μs simulations. Error bars are standard deviations (n=5).

Simulations of this polypeptide support the hypothesis: it is unable to fold in isolation (Figure 5 black), but once attached to calbindin becomes capable of folding (Figure 5 green), though it must compete with the calbindin N-terminal EF-hand to do so. This data suggests that our model qualitatively captures the cooperativity of protein folding.

4.1.3 Agreement of weighted ensemble and brute force simulations.

In order to confirm that the weighted ensemble path sampling algorithm does not alter the resulting thermodynamics of a system, we conducted a simulation of calbindin at 85°C using the algorithm. We plotted 95% confidence intervals of bin probability against the calbindin fraction of native contacts every 0.1 μs until they became, on average, equal in width to the 95% confidence intervals from a set of eight 20-μs brute force simulations, achieving this result after 0.7 μs [Figure 6]. The brute force average falls within these weighted ensemble confidence intervals for more than 90% of all histogram bins, a strong indication that the weighted ensemble algorithm does not alter observed thermodynamics.

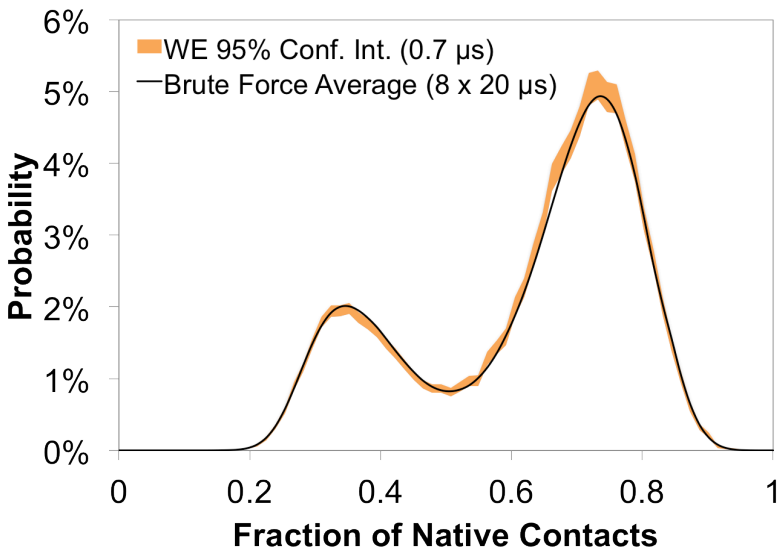

Figure 6: Comparison of weighted ensemble and brute force simulation results for calbindin at 85 °C. A) Histogram of calbindin fraction of native contacts, plotted as the average of eight 20-μs brute force simulations for brute force (black line) and 95% confidence intervals from a 0.7-μs weighted ensemble simulation (orange band). The average widths of 95% confidence intervals from brute force simulations (not shown) and weighted ensemble are very nearly equal. Confidence intervals were calculated from Monte Carlo bootstrapping for weighted ensemble and from the standard deviation (n=8) for brute force.

Our weighted ensemble simulation achieved the same average confidence interval widths as a set of brute force simulations but required less than 1% as many CPU hours. This result is encouraging but also surprising because weighted ensemble efficiency improvement is expected to have an exponential dependence on the height of the transition barriers; the trough separating the folding and unfolded states is modestly populated (near Fraction of Native Contacts = 0.5 in Figure 6), indicating that the energetic barrier separating the folded and unfolded states is small. Planned weighted ensemble simulations for 20 °C, where the folding and unfolding barriers will be a great deal larger [12], can be expected to even further outpace their brute force counterparts.

4.2 RESEARCH PLAN AND TIMELINE

4.2.1 Parameterize the protein model to reproduce experimental folding free energy of calbindin (Months 1-3).

Accurately capturing the thermodynamics of calbindin is an important step towards studying calbindin-AFF switching. In our simulations, the primary adjustable parameter that determines the stability of calbindin is the well depth of the attractive potential used to model native contacts between pairs of residues. Therefore, we will adjust this parameter to reproduce the experimental folding free energy ΔG_{fold} of calbindin. In order to give our simulations the best chance of reproducing experimental results, we will simulate at 20 °C, the temperature at which much of the experimental data for this system was collected [12, 13]. To validate these simulations, we will compare our computed rate constants of switching to those measured by experiment.

While previous brute force simulations have been able to reproduce a $\Delta G_{fold} = 0$ at 85 °C, reproducing the $\Delta G_{fold} = -3.9$ kcal/mol at 20 °C would be a prohibitive task for brute force simulations: significantly populating the unfolded state would require greater simulation time (by about a factor of 500), most of which would be spent waiting for rare unfolding events. Therefore, we will employ the "weighted ensemble" path sampling algorithm, which facilitates the sampling of fast-but-rare events (see Section 3.5 for details of this algorithm).

We will first develop the necessary software tools to conduct on-the-fly parameterization of the well depth using weighted ensemble simulations; these tools will allow our simulations to periodically check for convergence of the free energy profile and automatically adjust the well depth parameter to bring ΔG_{fold} closer to its desired value. These tools will be applied in later

phases of this project and will also be of broader use to any project that involves parameterizing coarse-grained simulations.

In parallel with this software development, we will begin to determine proper weighted ensemble settings for this system, including our bin boundaries along the progress coordinate (fraction of native contacts), number of simulations per bin, and time between resampling events. While these settings do not influence any thermodynamic or kinetic properties of the system, careful choice of these settings will allow our simulations to converge with as little computational effort as possible.

Previous studies suggest that the theoretical limit on computational efficiency improvement that can be achieved using weighted ensemble increases exponentially with the height of the transition barrier [46]. Our parameterization simulations will allow us to build a data set for comparing our observed efficiency improvements against these theoretical limits for a few different barrier heights. Plotting the total simulation time required for convergence against the observed rate of unfolding k_{unfold} and extrapolating will help approximate the simulation time required to converge simulations for even larger barriers; in addition, comparing this simulation time against the observed transition event durations will inform us about any correlation between these two, which has been hypothesized [46].

A back-of-the-envelope calculation of the total amount of data needed for the parameterization can be generated assuming 1) we will need to sequentially try about 10 different parameter values, 2) convergence of each parameter test will require 0.7 μs of data, the same as was required for the $\Delta G_{fold} = 0$ case. This value assumes that we will be able to achieve the theoretical limit of weighted ensemble efficiency improvement for this system (about a factor of 1600 over brute force). Using weighted ensemble along with our coarse-grained structural

model and Gō-type force field on a system the size of calbindin, we can generate about 2 μs of data per day per each 2.66-GHz core used. As we plan to use at least eight cores, this suggests that parameterization can be completed in a matter of hours using very conservative computational resources. While we hope to achieve efficiency near the theoretical limit, it would require only a tiny fraction of this limit to make parameterization within one month a feasible goal.

4.2.2 Test the ability of our protein model to capture the effect of permutation on its thermodynamic stability (Month 4).

Circular permutation of calbindin at residue 44 produces a surprising result: while most circular permutations result in a less stable protein fold, permuted calbindin is more stable than native calbindin, with a ΔG_{fold} = -5.5 kcal/mol at 20 °C [12]. Correctly modeling this increased stability will be important as we prepare to model the tug-of-war between the native and permuted folding frames in calbindin-AFF. We can test the ability of simulations to predict this increase in stability by applying the well depth parameter appropriate for native calbindin to the circular permutant. If our theoretical folding free energy is significantly similar to that achieved through experiment, it will be a strong indicator that these simulations can be used to predict the impact of circular permutation on a protein.

It will be necessary to accurately reproduce the thermodynamics of the circular permutant before simulating the calbindin-AFF switch. Therefore, if the simulated folding free energy value for the circular permutant is significantly different from experiment, we will parameterize the circular permutant to reproduce its experimental folding free energy in the same manner as described for calbindin. It can be expected that this parameterization will require less simulation

time than the calbindin parameterization, since the parameter determined for calbindin serves as a reasonable initial guess.

4.2.3 Simulate alternate frame folding of the calbindin construct (Month 5).

Once we have obtained well depth parameters for the components that comprise calbindin-AFF (calbindin and its circular permutant), we can apply these parameters to their respective regions of the calbindin-AFF protein. This offers us our best chance to accurately capture the thermodynamics of the folding tug-of-war in calbindin-AFF while using a minimal structural and energetic model. Thanks to the clever usage of thiol-disulfide exchange chemistry, an experimental approximation for the relative stability of the two folding frames is available for comparison with simulation results [12].

Circular dichroism spectroscopy reveals that there is no net folding or unfolding of calbindin-AFF when switching occurs, therefore it is hypothesized that the unfolding of one frame and the folding of the other occurs in a concerted fashion [13]. In addition, switching of calbindin-AFF has been indirectly observed by monitoring the fluorescence activity of pyrene and BODIPY probes affixed to the protein, thus providing us with rates of switching in the forward and reverse directions [12, 13]. By simulating calbindin-AFF, we may obtain switching rates along with an extent of switch concertedness to compare to experiments, and also present a mechanism for switching at 20 °C.

Simulation of calbindin-AFF requires no new parameterization, and, being a single simulation rather than a series, should take only a fraction of the computational effort, though it is worth noting that calbindin-AFF is about 50% larger than calbindin or its circular permutant.

4.2.4 Model the binding of calcium ions to calbindin and its circular permutant (Months 6-8).

In order for calbindin-AFF to function as a calcium sensor, the tug-of-war between the native and permuted folds must be significantly shifted by the addition of calcium. To prevent calcium from stabilizing both frames, the binding site of one frame is inactivated: a glutamate-to-glutamine mutation is performed in either the native folding frame or the permuted frame, greatly reducing that frame's calcium affinity [12]. The calbindin-AFF variant with the mutation in the native frame is referred to as E65Q-1, while the permuted frame mutant is referred to as E65Q-2. Because these mutants are expected to switch in response to calcium, much of the available experimental data for calbindin-AFF is for these two mutant versions. Since we would like our comparisons of simulation and experiment to be as direct as possible, we would prefer to conduct tests of these mutants in both the presence and absence of calcium. Our first step, then, is to correctly capture the thermodynamic effects of calcium on calbindin and its circular permutant.

However, simulating the effect of calcium presents a few significant computational challenges, most importantly the extent of stabilization of calbindin by calcium. Adding 100 μM calcium to a 12 μM sample of calbindin decreased ΔG_{fold} by 4.4 kcal/mol at 20 °C (resulting in a K_{fold} of 1.5 x 10^6) [12]. If these conditions were reproduced *in silico*, it would require an astronomical amount of time to simulate a handful of unfolding transitions using brute force. However, the theoretical limit of efficiency for weighted ensemble simulations rises exponentially with the barrier height [46], and recent weighted ensemble results suggest that sampling conformations and transitions of such immense rarity using modest computing resources is feasible: weighted ensemble simulations of binding of the P53 N-terminal peptide

with MDM2 using a slightly more detailed coarse-grained model have produced hundreds of unique binding transitions using about 34,000 times less simulation data than would have been required with brute force simulations [Wang, Zwier and Chong, In Preparation].

We will try two different approaches to explore the effects of calcium binding, each of which will be parameterized to reproduce experimental calbindin stability in the presence of 100 μM calcium: thermodynamic re-parameterization, where we adjust the contact well depth parameter of the protein contacts to implicitly model calcium, and an explicit ion model, where we include calcium pseudo-atoms with an attractive potential to their binding site contacts, and adjust the strength of these contacts to capture the experimental dissociation constants. Under the thermodynamic re-parameterization approach, there is an assumption that increased stability from calcium binding is propagated evenly throughout the entire protein. With the explicit ion model, any stabilization beyond the binding site is not explicit in the energy function; however, previous studies of this energetic model have shown that a binding event at one part of a protein can alter the stability at a distant location [30].

We will first attempt thermodynamic re-parameterization of calbindin and its circular permutant at 20 °C, using data from the original calbindin parameterization to approximate the simulation time that will be required. This simulation will provide a more challenging test of weighted ensemble's ability to increase path sampling efficiency; therefore, we will first conduct our parameterization using 96 processors, generating 200 μs of data per day. Simulation convergence for our initial guess parameter, along with the original calbindin data set, will give us a much better idea of the resources necessary to achieve parameterization than is currently available.

The explicit ion approach presents a few advantages over the thermodynamic reparameterization approach. The binding of calcium may cause important changes to the protein folding and unfolding mechanisms that can't be see with an implicit approach. Explicit ions allow us to compare simulation results against experiment in a novel way: by parameterization to reproduce the dissociation constants of calcium binding, we will be able to compare the resulting thermodynamics of calbindin against experiments; this will tell us how well our model is able to capture the change in protein stability that results from ligand binding. Additionally, explicit ions allow us to test the effects of varying the calcium concentration. While the computational effort required should be comparable between the explicit ion and thermodynamic reparameterization, we will need to implement new adjustments to the energetic model to ensure that the binding of calcium is cooperative (the dissociation constants for binding are ~1 μM and ~0.1 μM [47]).

4.2.5 Model the binding of calcium ions to the E65Q-1 and E65Q-2 mutants of the calbindin construct (Months 9-10).

Following the reproduction of experimental stabilities of calbindin and its circular permutant in calcium, our next step would be to simulate the E65Q-1 and E65Q-2 proteins in calcium. The E65Q-1 and E65Q-2 proteins are expected to predominately adopt the permuted and nonpermuted folds, respectively, as that is where their intact calcium binding sites can be found. We will use simulations to further explore the mechanism of switching in both directions, including comparisons of experimental and *in silico* kinetics.

Experiments have shown that the E65Q mutation does not appreciably change the stability of calbindin [12], therefore the behavior of E65Q-1 and E65Q-2 in the absence of calcium is assumed to be identical to each other as well as the native calbindin-AFF. Since

accurately describing thermodynamics is important for modeling the switching tug-of-war in these proteins, we will use the explicit ion approach only of it can demonstrate reasonable reproduction of the calbindin and circular permutant stabilities in the presence of calcium. We will apply either the calcium-binding site contact well depth parameter (in the explicit ion case) or the native contact well depth parameter (in the re-parameterization case) to the folding frame that possesses the intact calcium binding site.

4.3 EXPECTED OUTCOMES

These simulations will provide molecular insights into the switching mechanism of alternate frame folding that have been sorely lacking from experiments. Comparisons with experiments will allow us to gauge to what degree of accuracy such a simplified structural and energetic model is capable of determining the thermodynamic and kinetic properties of protein switches. We hope to demonstrate that weighted ensemble allows for greatly enhanced sampling in coarse-grained protein simulations, and gauge to what extent the improvement over brute force is dependent on the height of the energetic barriers and kinetics related to folding and unfolding transitions. Through the implementation of coarse-grained modeling and weighted ensemble, we hope to show that simulation methods can provide novel insights into the behavior of protein switches that are both supported by experimental data and attainable with a modest amount of time and computational resources.

4.4 POTENTIAL PROBLEMS & ALTERNATIVE STRATEGIES

4.4.1 Lack of favorable nonnative interactions.

The proposed energetic model assumes a minimally frustrated energetic landscape, while in actuality there are a number of favorable interactions that are not native to the folded state, including electrostatic, hydrophobic, and polar interactions. Models that neglect such interactions may generate folding mechanisms that are inconsistent with experiments [48] or more detailed simulation methods [49, 50]. However, we believe it is valuable to first extensively explore the simplest residue-level model that may reproduce experimental results, and therefore consider the inclusion of nonnative attractive forces an interesting future direction. Additionally, we hope to demonstrate that any mechanisms revealed under our simplified energetic model are reasonable by comparing their kinetics with experimental data.

4.4.2 Insufficient sampling for convergence of holo calbindin simulations.

As previously stated, accurately capturing the folding equilibrium of calbindin in the presence of calcium is a considerable challenge that should be made easier with weighted ensemble. However, it still may require an impractical amount of time and computational resources to converge simulations of this equilibrium. If that is the case, we will model the switching of E65Q-1 and E65Q-2 using an arbitrarily high contact well depth in the holo folding frame. As we will not be able to converge the thermodynamics of such switches, we will instead simulate with the goal of converging the switching rate from the apo folding frame to the holo frame. This will still allow us to compare with experimental switching kinetics for both switches.

4.4.3 Discrepant experimental and theoretical stabilization of calbindin by calcium.

Reproduction of the apo stability of calbindin and the calcium dissociation constant may not result in accurate holo stability for calbindin due to our lack of an electrostatic energy term and/or side chains. In this case, we will first use the thermodynamic reparameterization approach described previously, but we will also attempt to reproduce this holo stability through the inclusion of electrostatics and/or coarse-grained side chains.

5.0 FUTURE DIRECTIONS

The design and testing of the first alternate frame folding switches has produced plenty of questions about how they work and what can be done to improve or modify them. While we will begin to provide some computational insights into these questions with this project, there are a number of open questions that may be amenable to study with coarse-grained molecular simulations.

Some adjustable parameters in the design of alternate frame folding proteins, such as the site of circular permutation and the native stability of each folding frame, are likely to play an important role in switching activity. Calbindin, being a 75-residue protein, has 74 different circularly permuted topologies that may be explored, and therefore 148 different potential alternate frame folding designs based solely on the concept of duplication-and-ligation (since performing this operation from N-terminus to C-terminus is not identical to performing it from C-terminus to N-terminus, in contrast with circular permutation). We may be able to systematically explore all of these topologies and examine how switching behavior changes in each case. Additionally, by artificially adjusting the stabilities of the folding frames for the original calbindin-AFF design, we may be able to test the hypothesis that intrinsic disorder in a protein "primes" it to produce maximal allosteric coupling [51]. Exploring these modifications may indicate that some can improve the range of detection or response rate of the switch; with experimental agreement of simulations in this project will lend credence to these predictions.

We have chosen purposefully to use a minimal model to examine calbindin-AFF switching, and therefore one natural question will be how modifications to this model alter our results. The rationale behind the minimal model includes the fact that it is the least computationally expensive to examine, but depending on the efficiency improvements achieved by weighted ensemble for the minimal model, we may find that a multi-scale approach involving the addition of more detailed models will be vital to understanding the switching mechanism or properly modeling the effect of explicit calcium ions.

One of the most exciting possibilities of this work will depend on the degree of agreement between these coarse-grained simulations and experimental data. If this work can produce even qualitative agreement with experiments, and do so using only modest computational resources, it may suggest that new alternate frame folding proteins can be proposed and tested using this simulation protocol as a starting point, requiring only an appropriate protein structure and thermodynamic data (and possibly thermodynamic data on any circular permutants). Such a computational screening method for new alternate frame folders, rooted in experimental data that is already available in the literature for many proteins, could lead experimentalists to design much more promising switch prototypes on their first attempt.

BIBLIOGRAPHY

1. Hunter, T., *Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling.* Cell, 1995. **80**(2): p. 225-36.

2. Catterall, W.A., *Structure and function of voltage-gated ion channels.* Annu Rev Biochem, 1995. **64**: p. 493-531.

3. Pearl, L.H. and C. Prodromou, *Structure and mechanism of the Hsp90 molecular chaperone machinery.* Annu Rev Biochem, 2006. **75**: p. 271-94.

4. Ostermeier, M., *Engineering allosteric protein switches by domain insertion.* Protein Eng Des Sel, 2005. **18**(8): p. 359-64.

5. Ambroggio, X.I. and B. Kuhlman, *Design of protein conformational switches.* Curr Opin Struct Biol, 2006. **16**(4): p. 525-30.

6. Stratton, M.M. and S.N. Loh, *Converting a protein into a switch for biosensing and functional regulation.* Protein Sci, 2011. **20**(1): p. 19-29.

7. Vallee-Belisle, A. and K.W. Plaxco, *Structure-switching biosensors: inspired by Nature.* Curr Opin Struct Biol, 2010. **20**(4): p. 518-26.

8. Radley, T.L., et al., *Allosteric switching by mutually exclusive folding of protein domains.* J Mol Biol, 2003. **332**(3): p. 529-36.

9. Ha, J.H., et al., *Modular enzyme design: regulation by mutually exclusive protein folding.* J Mol Biol, 2006. **357**(4): p. 1058-62.

10. Cutler, T.A. and S.N. Loh, *Thermodynamic analysis of an antagonistic folding-unfolding equilibrium between two protein domains.* J Mol Biol, 2007. **371**(2): p. 308-16.

11. Cutler, T.A., et al., *Effect of interdomain linker length on an antagonistic folding-unfolding equilibrium between two protein domains.* J Mol Biol, 2009. **386**(3): p. 854-68.

12. Stratton, M.M., D.M. Mitrea, and S.N. Loh, *A Ca2+-sensing molecular switch based on alternate frame protein folding.* ACS Chem Biol, 2008. **3**(11): p. 723-32.

13. Stratton, M.M. and S.N. Loh, *On the mechanism of protein fold-switching by a molecular sensor.* Proteins, 2010. **78**(16): p. 3260-9.

14. Stratton, M.M., et al., *Structural characterization of two alternate conformations in a calbindin Dk-based molecular switch.* Biochemistry, 2011. **50**(25): p. 5583-9.

15. Mitrea, D.M., L.S. Parsons, and S.N. Loh, *Engineering an artificial zymogen by alternate frame protein folding.* Proc Natl Acad Sci U S A, 2010. **107**(7): p. 2824-9.

16. Lewit-Bentley, A. and S. Rety, *EF-hand calcium-binding proteins.* Curr Opin Struct Biol, 2000. **10**(6): p. 637-43.

17. Buckle, A.M., G. Schreiber, and A.R. Fersht, *Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-A resolution.* Biochemistry, 1994. **33**(30): p. 8878-89.

18. Goodsell, D.S., *Glucose Oxidase.* PDB Molecule of the Month, 2006.

19. Kohn, J.E. and K.W. Plaxco, *Engineering a signal transduction mechanism for protein-based biosensors.* Proc Natl Acad Sci U S A, 2005. **102**(31): p. 10841-5.

20. Baird, G.S., D.A. Zacharias, and R.Y. Tsien, *Circular permutation and receptor insertion within green fluorescent proteins.* Proc Natl Acad Sci U S A, 1999. **96**(20): p. 11241-6.

21. Ivarsson, Y., et al., *Folding and misfolding in a naturally occurring circularly permuted PDZ domain.* J Biol Chem, 2008. **283**(14): p. 8954-60.

22. Viguera, A.R., L. Serrano, and M. Wilmanns, *Different folding transition states may result in the same native structure.* Nat Struct Biol, 1996. **3**(10): p. 874-80.

23. Vogel, C. and V. Morea, *Duplication, divergence and formation of novel protein topologies.* Bioessays, 2006. **28**(10): p. 973-8.

24. Carrington, D.M., A. Auffret, and D.E. Hanke, *Polypeptide ligation occurs during post-translational modification of concanavalin A.* Nature, 1985. **313**(5997): p. 64-7.

25. Luger, K., et al., *Correct folding of circularly permuted variants of a beta alpha barrel enzyme in vivo.* Science, 1989. **243**(4888): p. 206-10.

26. Qian, Z. and S. Lutz, *Improving the catalytic activity of Candida antarctica lipase B by circular permutation.* J Am Chem Soc, 2005. **127**(39): p. 13466-7.

27. Whitehead, T.A., L.M. Bergeron, and D.S. Clark, *Tying up the loose ends: circular permutation decreases the proteolytic susceptibility of recombinant proteins.* Protein Eng Des Sel, 2009. **22**(10): p. 607-13.

28. Feher, J.J., C.S. Fullmer, and R.H. Wasserman, *Role of facilitated diffusion of calcium by calbindin in intestinal calcium absorption.* Am J Physiol, 1992. **262**(2 Pt 1): p. C517-26.

29. Szebenyi, D.M. and K. Moffat, *The refined structure of vitamin D-dependent calcium-binding protein from bovine intestine. Molecular details, ion binding, and implications for the structure of other calcium-binding proteins.* J Biol Chem, 1986. **261**(19): p. 8761-77.

30. Mills, B.M. and L.T. Chong, *Molecular simulations of mutually exclusive folding in a two-domain protein switch.* Biophys J, 2011. **100**(3): p. 756-64.

31. Jha, A.K., et al., *Statistical coil model of the unfolded state: resolving the reconciliation problem.* Proc Natl Acad Sci U S A, 2005. **102**(37): p. 13099-104.

32. Eswar, N., et al., *Comparative protein structure modeling using MODELLER.* Curr Protoc Protein Sci, 2007. **Chapter 2**: p. Unit 2 9.

33. Marti-Renom, M.A., et al., *Comparative protein structure modeling of genes and genomes.* Annu Rev Biophys Biomol Struct, 2000. **29**: p. 291-325.

34. Chavez, L.L., J.N. Onuchic, and C. Clementi, *Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates.* J Am Chem Soc, 2004. **126**(27): p. 8426-32.

35. Elcock, A.H., *Molecular simulations of cotranslational protein folding: fragment stabilities, folding cooperativity, and trapping in the ribosome.* PLoS Comput Biol, 2006. **2**(7): p. e98.

36. Julenius, K., et al., *Coupling of ligand binding and dimerization of helix-loop-helix peptides: spectroscopic and sedimentation analyses of calbindin D9k EF-hands.* Proteins, 2002. **47**(3): p. 323-33.

37. Ermak, D.L. and J.A. Mccammon, *Brownian Dynamics with Hydrodynamic Interactions.* Journal of Chemical Physics, 1978. **69**(4): p. 1352-1360.

38. Frembgen-Kesner, T. and A.H. Elcock, *Striking Effects of Hydrodynamic Interactions on the Simulated Diffusion and Folding of Proteins.* Journal of Chemical Theory and Computation, 2009. **5**(2): p. 242-256.

39. Rotne, J. and S. Prager, *Variational Treatment of Hydrodynamic Interaction in Polymers.* Journal of Chemical Physics, 1969. **50**(11): p. 4831-&.

40. Yamakawa, H., *Transport Properties of Polymer Chains in Dilute Solution - Hydrodynamic Interaction.* Journal of Chemical Physics, 1970. **53**(1): p. 436-&.

41. de la Torre, J.G., M.L. Huertas, and B. Carrasco, *Calculation of hydrodynamic properties of globular proteins from their atomic-level structure.* Biophys J, 2000. **78**(2): p. 719-730.

42. de la Torre, J.G., *Hydration from hydrodynamics. General considerations and applications of bead modelling to globular proteins.* Biophys Chem, 2001. **93**(2-3): p. 159-170.

43. Huber, G.A. and S. Kim, *Weighted-ensemble Brownian dynamics simulations for protein association reactions.* Biophys J, 1996. **70**(1): p. 97-110.

44. Zhang, B.W., D. Jasnow, and D.M. Zuckerman, *Efficient and verified simulation of a path ensemble for conformational change in a united-residue model of calmodulin.* Proc Natl Acad Sci U S A, 2007. **104**(46): p. 18043-8.

45. Neira, J.L. and A.R. Fersht, *Exploring the folding funnel of a polypeptide chain by biophysical studies on protein fragments.* J Mol Biol, 1999. **285**(3): p. 1309-33.

46. Zwier, M.C., J.W. Kaus, and L.T. Chong, *Efficient Explicit-Solvent Molecular Dynamics Simulations of Molecular Association Kinetics: Methane/Methane, Na(+)/Cl(-), Methane/Benzene, and K(+)/18-Crown-6 Ether.* Journal of Chemical Theory and Computation, 2011. **7**(4): p. 1189-1197.

47. Kragelund, B.B., et al., *Hydrophobic core substitutions in calbindin D9k: effects on Ca2+ binding and dissociation.* Biochemistry, 1998. **37**(25): p. 8926-37.

48. Daggett, V. and A. Fersht, *The present view of the mechanism of protein folding.* Nat Rev Mol Cell Biol, 2003. **4**(6): p. 497-502.

49. Clementi, C., *Coarse-grained models of protein folding: toy models or predictive tools?* Curr Opin Struct Biol, 2008. **18**(1): p. 10-5.

50. Koga, N. and S. Takada, *Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model.* J Mol Biol, 2001. **313**(1): p. 171-80.

51. Hilser, V.J. and E.B. Thompson, *Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins.* Proc Natl Acad Sci U S A, 2007. **104**(20): p. 8311-5.