PROTEIN ASSOCIATION INVOLVING INTRINSIC DISORDERED PROTEINS AND ITS REGULATION ON SYNAPTIC PLASTICITY THROUGH MODULATING CA^{2+} BINDING

A Dissertation Presented to

the Faculty of the Department of Physics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Pengzhi Zhang

August 2016

PROTEIN ASSOCIATION INVOLVING INTRINSIC DISORDERED PROTEINS AND ITS REGULATION ON MEMORY AND LEARNING THROUGH MODULATING CA^{2+} BINDING

Pengzhi Zhang

Approved:

Dr. Margaret S. Cheung, Chairperson

Dr. Kevin E. Bassler

Dr. Gemunu Gunaratne

Dr. Vassiliy Lubchenko

Dr. M. Neal Waxham

Dean, College of Natural Sciences and Mathematics

ACKNOWLEDGEMENTS

I would like to give my sincerest gratitude to my advisor Dr. Margaret Shun Cheung. Her instructions and criticism guide me all the way on the path to a Ph.D. Her research on modeling of proteins in cellular environment interested me to make the decision of joining her lab. I would like to thank her for giving me the interesting project and for the continuous encouragement, as well as her generous financial support during the period. Without her, this thesis will never exist.

I would like to thank my current and former lab mates. Thanks to the following for sharing their ideas / giving me advice: Dr. Dirar Homouz, Dr. Antonios Samiotakis, Dr. Balamurugan Desinghu, Dr. Guoxiong Su, Fabio Zegarra Choque, Jianfa Chen, Mohammadmehdi Ezzatabadipour, Jacob Ezerski, Andrei Gasic, Rodney Helm, Jonathan Hankin, Basilio Cieza Huaman, especially Dr. Qian Wang and Dr. Swarnendu Tripathi. Dr. Qian Wang instructed me hand by hand and generously gave me help after leaving the group. Dr. Swarnendu Tripathi is a wonderful person to work with and he gave me a lot of ideas on modelling and analyses. Moreover, I was fortunate to mentor undergraduate student Rodney Helm and he helped me with simulations of calmodulin and its mutant targets as well as the following data analysis.

I would like to thank Dr. M. Neal Waxham and Dr. Yin Liu for their collaboration and advice. In addition, I would like to thank my committee members for sharing their experiences and for giving me honest and critical comments on my research every year, as well as on my career goal. These comments helped me identify my shortcomings in all aspects including research projects, methods, and presentation skills.

I need to thank generous financial support from Department of Physics, including the Teaching Assistantship. Thank the support from Dr. Cheung's National Institute of Health grant. I also want to thank computational resources from Center for Advanced Computing and Data Systems at University of Houston, from Center for Theoretical and Biological Physics at Rice University, and from Extreme Science and Engineering Discovery Environment.

Finally I owe my deepest gratitude to my family. Thanks for their endless encouragement and support to get me through ups and downs in my life. My special thanks go to my wife Qin. Her deep and unconditional love and support enable me to focus on my research.

PROTEIN ASSOCIATION INVOLVING INTRINSIC DISORDERED PROTEINS AND ITS REGULATION ON SYNAPTIC PLASTICITY THROUGH MODULATING CA^{2+} BINDING

An Abstract of a Dissertation

Presented to

the Faculty of the Department of Physics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Pengzhi Zhang

August 2016

ABSTRACT

Protein-protein association is essential for biological functions in all living organisms, including synaptic plasticity that relates to learning and memory formation, which involves the regulation of Ca^{2+} signals. The ubiquitin Ca^{2+} signaling protein calmodulin (CaM) plays a central role in encoding Ca^{2+} signals by interacting with a diversity of CaM binding target proteins (CaMBTs). Therefore, the mechanism of how CaM can recognize CaMBTs has aroused a broad and lasting interest. Experimental investigations on binding of CaM and CaMBTs uncover thermodynamic and kinetic properties of protein interactions but they lack molecular details to explain some macroscopic phenomena. Moreover, in many cases it is implausible to determine the molecular mechanisms. Analytical models usually use simplified models but overlook the structural flexibility of the proteins. In this study, firstly I implemented a sequence-based dihedral angle potential for capturing the characteristics of intrinsically disorder unstructured CaMBTs. By applying molecular simulations on association of CaM and CaMBTs, I revealed that molecular recognition requires mutual conformational changes of both CaM and CaMBT. Secondly, among the CaMBTs, CaM-dependent kinase II (CaMKII) and neurogranin (Ng) play an essential role in synaptic plasticity and they are experimentally shown to have opposing effects on Ca^{2+} affinity for CaM, but the molecular mechanism is unknown. Advancing coarse-grained molecular simulations employing the model, I found that the interaction between Ng peptide and the C-domain of CaM (cCaM) disrupts the intra-molecular interaction between the two Ca^{2+} binding loops. Next, I performed steered molecular dynamics simulations on atomistic models from several reconstructed coarse-grained structures to compute the changes in Ca^{2+} affinity for CaM with and without Ng peptide using Jarzynski's equality. I discovered the molecular underpinnings of lowered affinity of Ca^{2+} for CaM in the presence of Ng peptide by showing that the N-terminal acidic region of Ng pries open the β-sheet structure between the Ca^{2+} binding loops at cCaM, enabling Ca^{2+} release. In contrast, CaMKII increases Ca^{2+} affinity for cCaM by stabilizing them. This study allows opportunities to connecting the molecular regulations in atomistic detail to the understanding of cellular process cascade of learning and memory formation.

Table of Contents

Abbreviations

ApoCaM: calmodulin without calcium ions

Ca^{2+}: calcium ion

CaM: calmodulin

CaMBT: calmodulin binding target

CaMKI: Ca^{2+}-dependent calmodulin kinase I

CaMKII: Ca^{2+}-dependent calmodulin kinase II

CD: circular dichroism

CI2: chymotrypsin inhibitor 2

HoloCaM: calmodulin with four calcium ions loaded

IDP: intrinsically disordered protein / peptide

KB: Karanicolas-Brooks dihedral angle potential

LTD: long term depression

LTP: long term potentiation

Ng: neurogranin

$Ng_{13\text{-}49}$: neurogranin peptide (residue 13 to 49).

NgIQ: IQ motif part of neurogranin

NMDA: N-Methyl-D-aspartate

NMR: nuclear magnetic resonance

PCA: principal component analysis

PDB: protein data bank

PME: particle mesh Ewald

PPI: protein protein interaction

PRE: paramagnetic relaxation enhancement

REMD: replica exchange molecular dynamics

RMSD: root mean square deviation

RMSF: root mean square fluctuation

SB: structure-based

SCM: side-chain C_α model

SMD: steered molecular dynamics

US: umbrella sampling

WHAM: weighted histogram analysis method

Chapter 1

Introduction

1.1 Protein-protein association and recognition

1.1.1 Introduction for protein structures and functions

Proteins are the main workhorses in the cells. A protein is made up of a chain of subunits called amino acids. Each amino acid has a central carbon atom (C_α) connecting an amino group (NH2-), an acidic group (-COOH) and a side-chain. The side-chain determines the type of the amino acid, including hydrophobic, polar uncharged and polar charged. The amino acids are covalently connected by stiff peptide bonds (-NH-CO-) and form a polypeptide chain as shown in Fig. 1.1a. While the unique sequence of the amino acids forms the primary structure, an important functional action for a protein is to fold into its native or a compact near-native tertiary structure (Fig. 1.1c), in some cases a quaternary structure by forming a complex with other proteins (Fig. 1.1d), except for intrinsically disordered proteins (IDPs), which do not maintain a stable tertiary structure under physiological conditions. As the protein starts to fold, parts of the polypeptide chain display distinct features of some local structural conformations between neighboring amino acids, known as the secondary structures. Depending on how the hydrogen bonds form between the backbones, there are two most common secondary structures: α-helix and β-sheet, as shown in Fig. 1.1b. The tertiary structure is the overall three-dimensional structure of the protein when the secondary structures are arranged through the interaction between

the side-chains, or the interaction with the living cell environment. The tertiary structure is the structure that a protein usually maintain in the cell environment and functional. Subunit proteins interact with each other and form complex structures or quaternary structures. Tertiary and quaternary structures are determined by various experimental techniques such as X-Ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy (cryo-EM) are deposited to databases. The structures can be accessed from protein data bank (PDB).

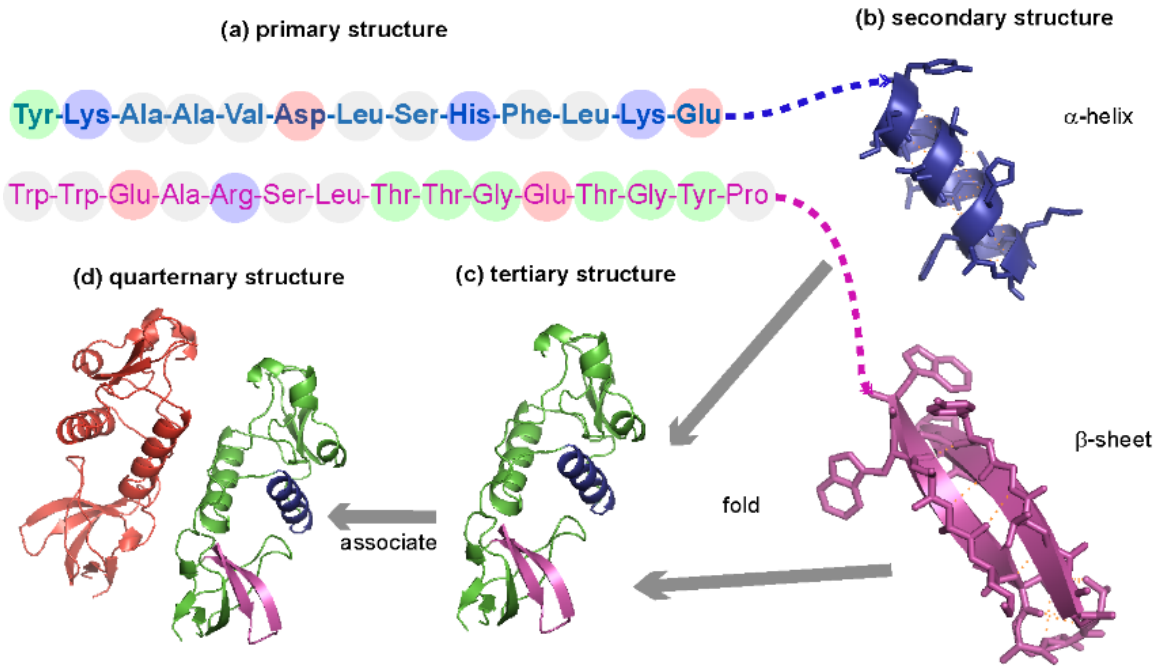

Figure 1.1 Protein structures. (a) Primary structure of the a-helix and b-sheet part from the FYN SH3 domain (PDB code: 1EFN). The 3-letter code of the sequences of a-helix and b-sheet are provided in blue and magenta, respectively. The string of colored beads encircling the 3-letter code represent the primary structure. The beads are colored according to polarity of the amino acids: red represents negatively charged, blue represents positively charged, white represents nonpolar, and green represents polar uncharged amino acids. (b) Cartoon representation of the secondary structure of α-helix and anti-parallel β-sheet. The hydrogen bonds are shown in orange dashed lines. (c) Cartoon representation of the folded native structure of the FYN SH3 domain. The corresponding α-helix and β-sheet are colored according to (b). (d) Cartoon representation of the crystalized homodimer. Protein figures are generated by Pymol (Schrodinger, 2015).

Proteins fold from micro-seconds to hours, depending on the length and topology of the protein (Kubelka et al., 2004). In 1968, Cyrus Levinthal (Levinthal, 1968) noticed that due to a huge number of degrees of freedom in an unfolded polypeptide chain, the time is enormous for a random search of all possible configurations assuming the sampling is at a rapid rate of picosecond scale. This is known as Levithal's paradox. Levinthal thus proposed that the folding must be through meta-stable intermediate states. Zwanzig (Zwanzig et al., 1992) argued that the conformational search is not random and the energy landscape of protein folding is not flat. While the protein in a non-native state explores probable conformations along many different series of traps, formation of the native contacts results in a decrease in the overall average energy and directs the following search more likely to the native state. In other words, there is a strong energetic bias towards the native basin that competes with the entropy. This is known as "the principle of minimal frustration", proposed by Bryngelson and Wolynes in 1987 (Bryngelson and Wolynes, 1987), capturing the roughness or ruggedness of the energy landscape. The nature selects sequences that favors a native states instead of those resemble the energetically frustrated glassy states, which would make the search for the native state endless. Later Onuchic et al. (Onuchic et al., 1997) proposed a "funnel" like energy landscape of a protein so that folding occurs through organizing an ensemble of structures rather than randomly searching through only a few uniquely defined structural intermediates or traps. It suggests that the most realistic model of a protein folding is to follow a sequential manner towards the native structure.

Globular proteins in living cells have evolved to have a smooth energy landscape while energetic frustration are not fully ruled out (Onuchic and Wolynes, 2004). Remarkably, the energetic frustration left during evolution can contribute to the biological function of the protein. In reality, flexible and disordered proteins preserve a rugged energy landscape, allowing them to be present multiple conformations. In Chapter 3, I will discuss how the nonspecific interactions can better describe the transition states of the ordered protein chymotrypsin inhibitor 2 (CI2) and reproduce the residual structures of the intrinsically disordered protein neurogranin.

1.1.2 Introduction to calmodulin

My computational study on protein structural flexibility and protein-protein interactions surrounds the ubiquitous Ca^{2+} signaling protein calmodulin (CaM). CaM is a small two-domain flexible protein composed of 148 amino acids. CaM exists in the cytosol where it participates in signaling pathways that regulate many crucial cellular processes including cell growth, apoptosis, muscle contraction, and short and long term memory formation. It constitutes more than 0.1% of the total protein content in the cell (1 to 10 μM). CaM is expressed at higher levels in rapidly growing cells, such as the cell undergoing division (Chin and Means, 2000). CaM deciphers the information encoded in the transient-Ca^{2+} flux and interacts with a vast selection of more than 300 binding targets (CaMBTs). What emerges from this is the remarkable conformational flexibility of CaM, which exists in highly dynamic structures in both of the Ca^{2+} free and bound forms (Slaughter et al., 2005; Anthis et al., 2011) and will adopt distinct conformations when bound to CaMBTs.

One of the factors contributed to the conformational flexibility of CaM is the flexible central linker connecting the two domains of CaM (Fig. 1.2 (A)). For the 30 structures of Ca^{2+}-free CaM (apoCaM) determined by Nuclear Magnetic Resonance (NMR) experiments, the average root mean square deviation (RMSD) of the C_α atoms is about 8.9 Å. However, if the structures are aligned according to N-domain of apoCaM (nCaM, residue 1-76), the average RMSD of C_α atoms in nCaM is only 2.1 Å (Fig. 1.2B); on the other hand, the C-domain of apoCaM (cCaM, residue 82-148) is also very stable and the average RMSD of the C_α atoms in cCaM is 1.2 Å if aligned to cCaM (Fig. 1.2C). This indicates that the structural changes in folded apoCaM are from the extent of unwinding of the central helix linker (residue 77 to 81).

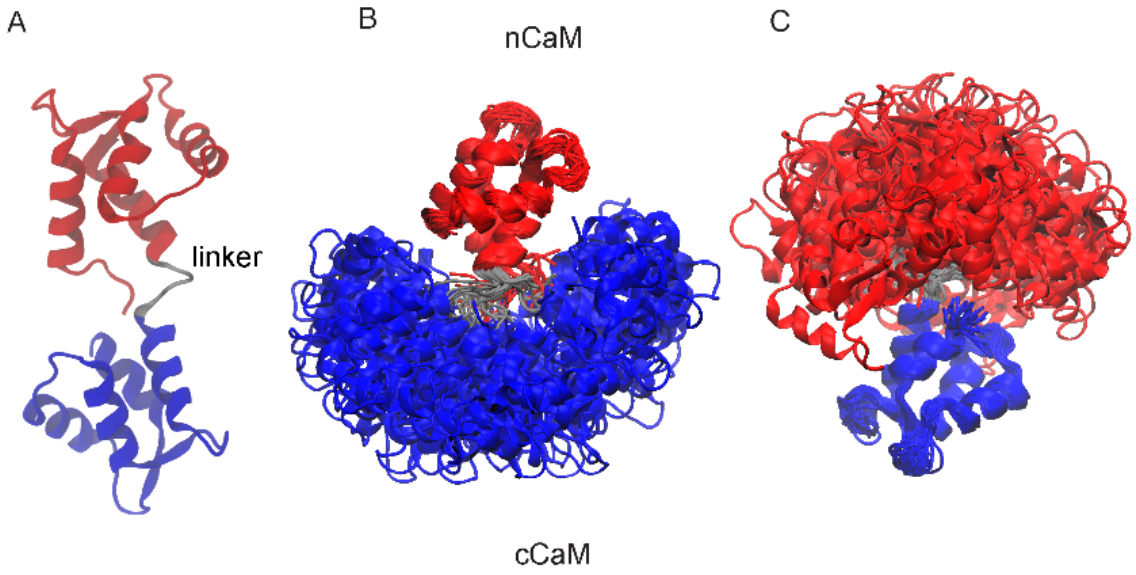

Figure 1.2 Cartoon representations of the Ca^{2+}-free CaM. (A) Solution structure of apoCaM (PDB ID: 1CFD). The CaM is colored according to the two globular domains and the central helix linker. N-terminal domain (nCaM, residue 1 to 76) is in red; C-terminal domain (cCaM, residue 82 to 148) is in blue and the central helix linker (residue 77 to 82) is in silver. (B) and (C) superimposed solution structures of apoCaM (PDB ID: 1DMO). An ensemble of 30 models were determined. The structures are aligned according to nCaM and cCaM, respectively.

CaM consists of eight α-helices. Each two helices and the inter-helix linker forms a motif called EF-hand motif, which is capable of binding metal ions including Ca^{2+} and Mg^{2+}. Each EF-hand motif includes a twelve-residue binding loop. Upon binding with Ca^{2+} ions, the helices within the EF-hand, which are initially at a closed anti-parallel position (Fig. 1.2(A)), open up (Fig. 1.3) and stabilize the Ca^{2+} binding. Up to four Ca^{2+} ions can be loaded meanwhile CaM undergoes global structural changes and adopts an open conformation (Fig. 1.3).

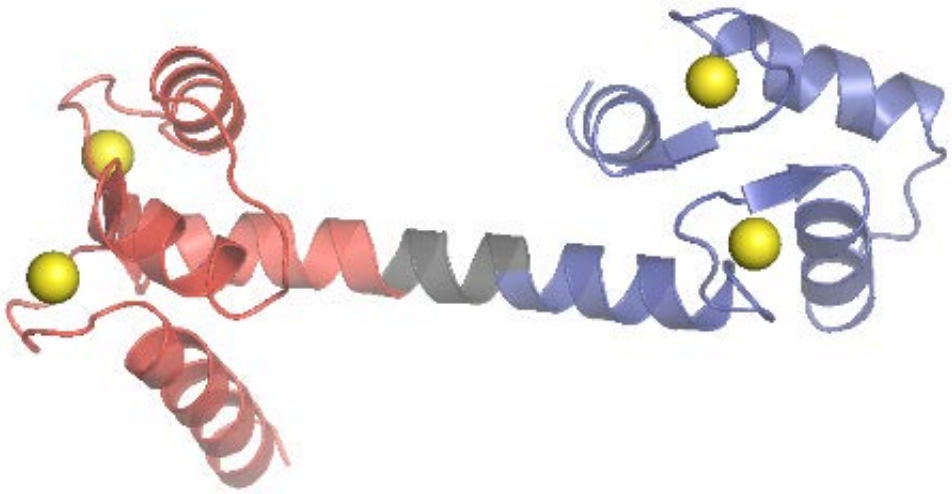

Figure 1.3 Cartoon representation of the Ca^{2+}-loaded CaM (PDB ID: 1CLL). CaM is colored in red, gray and blue according to nCaM (residues 1-76), central linker (residue 77-82) and cCaM (residue 75-148), respectively. Each domain is loaded with 2 Ca^{2+} ions (yellow beads). Note that although in the crystalized structure, the central linker is α-helical, in the solution structure, the central linker is highly flexible and non-helical.

In each of the two domains of CaM, a small anti-parallel β-sheet structure is formed between a pair of EF-hands, which stabilizes the globular domain, shown in Fig. 1.4 for cCaM. This construct of EF-hand pair (Kretsinger, 1996) is called EF-hand β-scaffold, which is proposed to contribute to the cooperativity of Ca^{2+} binding (Grabarek, 2006).

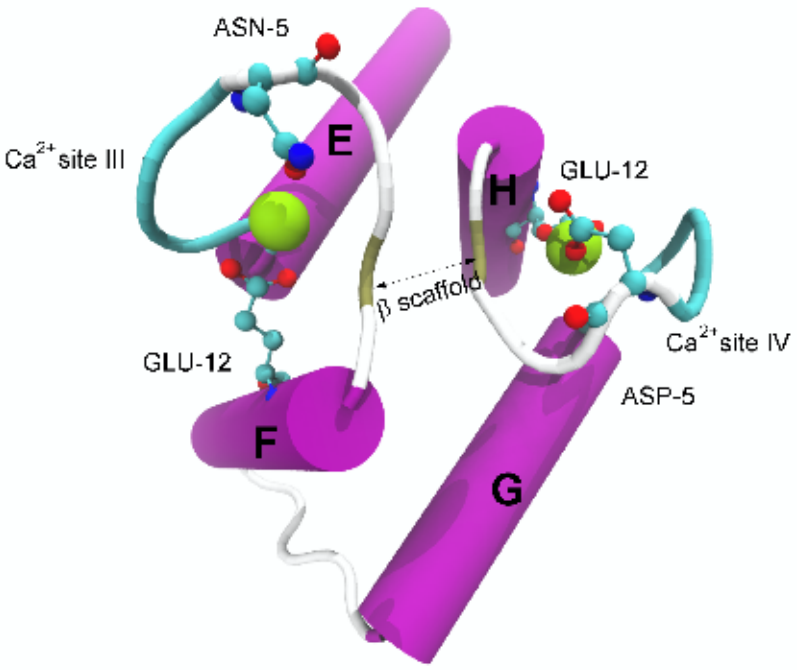

Figure 1.4 Illustration of the EF-hand β-scaffold. The structure is from C-terminus of the crystal structure of holoCaM (PDB code: 1CLL). The Ca^{2+} ions are shown in yellow beads. Each Ca^{2+} binding loop consists of 12 amino acids and can be labeled from N-term to C-term with number 1 to 12. The fifth (+Z coordinating ligand, ASN-5 or ASP-5) and twelfth (-Z coordinating ligand, GLU-12) residues in the Ca^{2+} binding loops are shown in ball-and-stick representation. β-sheet structure is formed between the central two residues from the two Ca^{2+} binding loops. The EF-hand pair with the β-sheet is called "EF-hand β-scaffold".

CaM can bind and activate more the 300 target proteins in various forms depending on the targets (Hoeflich and Ikura, 2002; Yamniuk and Vogel, 2004; Tidow and Nissen, 2013). For example, Fig 1.5 illustrates the canonical binding between CaM and a target peptide, where holoCaM binds to a CaMBT by wrapping its two domains around the peptide forming a helical structure (Heidorn et al., 1989; Meador et al., 1993). The binding affinity is largely increased in the presence of Ca^{2+} ions (Xia and Storm, 2005). The separations between some specific hydrophobic residues along the sequence of the CaMBT are evolutionarily conserved (Rhoads and Friedberg, 1997; Yamniuk and Vogel, 2004). The patterns of those separations are categorized as CaM binding motifs. These conserved motifs of the CaMBTs that display impressive variability in amino acid sequence classify the targets. How CaM can recognize and select its binding targets remains an open question. In the next part, I will review current views of possible protein-protein recognition mechanisms.

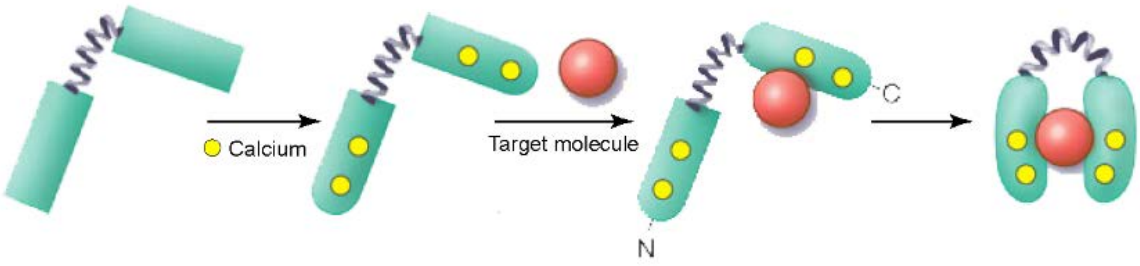

Figure 1.5 Schematic illustration of a canonical binding of a CaMBT to CaM. The CaM is represented by two domains in cyan connected by a spring. The CaMBT is represented by a red sphere. The Ca^{2+} ions are represented by yellow beads. The figure is adapted from (Smock and Gierasch, 2009). Up to 4 Ca^{2+} ions bind to each of EF hand domains and facilitate binding of the CaM to the CaMBT.

1.1.3 Protein-protein recognition mechanisms

Several models have been proposed for protein-protein recognition dating back to the late 19th century. In 1894, a “lock-and-key” model was proposed by Fisher (Fischer, 1894) neglecting the conformational changes of the proteins. In this mechanism, only proteins with compensated shapes can recognize each other (Fig. 1.6 (A)) and the attractive electrostatic interaction is the driving force for the association. This model can successfully explain many experiments on binding between an enzyme protein and its small binding substrate in the years when a protein structure could not be determined. However, it failed on explaining that the reaction rate between ribose-5-phosphate and 5’-nucleotidase is 1/100 the rate between adenylic acid and 5’-nucleotidase (Heppel and Hilmoe, 1951). It predicts equal rates for both. Therefore, including the structural flexibility of the proteins, Koshland (Koshland, 1958) modified the “lock-and-key” mechanism by assuming that the substrate can adopt its conformation according to the shape of the ligand upon binding in 1958, called the “induced-fit” mechanism as shown in Fig. 1.6(B). In the “induced fit” mechanism, the global minimum of the energy landscape moves to another state, or the bound state, which may be forbidden in the unbound state. The “induced fit” theory gave an explanation for the previous phenomenon: the conformation of 5’-nucleotidase is changed to the correct orientation by the binding of adenylic acid so that the reaction can happen, while the conformation of 5’-nucleotidase cannot be changed to the correct orientation by the binding of 5’-nucleotidase.

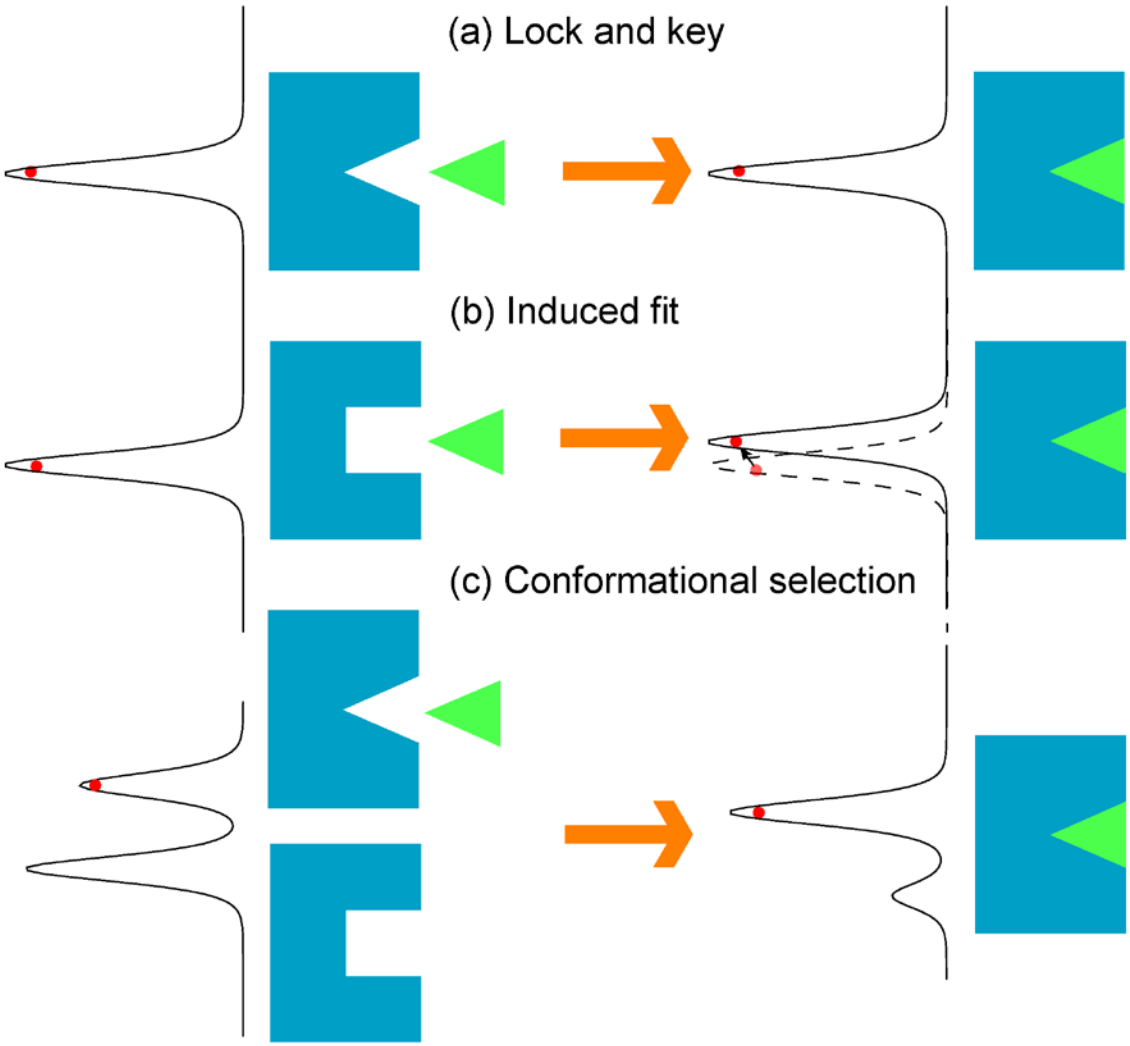

Figure 1.6 Schematic representation of protein recognition mechanisms. The substrate is in blue and the ligand is a green triangle. The red dots marks the states of the substrate in its own energy landscape. In models (A) and (B) the substrate can only adopt a single structure. In model (C) the substrate can adopt two structures. The ligand is considered to be small and rigid in all of the models.

Both of the two early mechanisms assumed that the unbound substrate only adopts a defined stable structure. However, according to the landscape theory of the protein folding (Onuchic et al., 1997), a protein may dynamically adopt multiple conformations in the solution. Thus, the "conformational selection" or "population shift" model was proposed (Ma et al., 1999). In this model, as shown in Fig 1.6(C), the binding of the substrate shifts one conformation to another.

However newer theories, termed extended conformational selection (Tsai et al., 1999), mutually-induced fit (Williamson, 2000), fly-casting (Shoemaker et al., 2000), or folding and binding (Levy et al., 2004), have begun to integrate molecule dynamics to describe folding and binding involving flexible molecules, especially for the intrinsically disordered proteins (IDPs) (Wright and Dyson, 1999; Dunker et al., 2001; Uversky, 2002) whose folding and binding are concomitant. Specially, the term mutually-induced fit, or co-folding, was introduced to describe the conformational changes of both the protein and the RNA components (Williamson, 2000). This mechanism may also be applied to protein-protein interactions. In Chapter 2, from my computer simulations, my colleagues and I proposed a mechanism of mutually and conformational induced fit for recognition between CaM and its target proteins.

1.1.4 Thermodynamics and kinetics of protein-protein association

Typically, there is structural information about the unbound and bound state proteins, as well as essential physical properties including the binding affinity and the association rate. The binding affinity reflects the strength of interaction between the binding partners. The affinity of binding between object **A** and object **B** (**A** + **B** ↔ **AB)**, is defined as K_A=

[**AB**]/[**A**][**B**], when the system is in equilibrium. The square brackets mean concentration in the solution.

The binding affinity of a ligand and the receptor is measured by experimental techniques including fluorescence spectroscopy, surface plasmon resonance (Deutch and Felderho, 1973) and atomic-force microscopy (Ratcliff and Erie, 2001). However, it is usually difficult for experimentalists to study the dynamics with structure information as well as the binding pathways where one or more intermediate states exist.

Fortunately, computer simulations can be employed to study the dynamics in multiple aspects in different resolutions, including the prediction of complex structure, the binding affinity and the binding rate. The binding affinity can be calculated using equilibrium or non-equilibrium methods or Monte Carlo simulations (Wolynes and Deutch, 1976). The equilibrium method is to compute the fraction of the complex concentration as in the definition. Methods like free energy perturbation (FEP) are widely utilized for accurate estimation of the binding affinity, where only the unbound and bound states are required. However, in some cases when the reaction barrier is high, the exchange between bound and unbound state is difficult to achieve an equilibrium, and the transition states as well as the binding pathway cannot be captured. In the latter case, Monte Carlo sampling may endure the same shortcoming. Umbrella sampling (US, please see the Appendix 6.5 for details) (Roux, 1995; Kastner, 2011) is used to improve the sampling of the system. A biased potential is applied in the many distributed windows along the reaction coordinate to overcome high energy barrier(s) so that the sampling at each window is equally good. The bias can be removed using Weighted Histogram Analysis Method (WHAM). Usually

the computational cost for US simulations is high. On the other hand, non-equilibrium methods such as steered molecular dynamics (SMD) are employed to explore the binding pathways as well as to compute the binding affinity. In SMD, an external force is applied to one or more of the binding targets to overcome the barrier(s) and accelerate the simulation. In Chapter 4, I applied SMD for calculation of binding affinity of Ca^{2+} for cCaM by using the Jarzynski's equality, where the free energy difference between the bound and unbound states was estimated as an ensemble average of exponentials of the work done during the unbinding process over a large number of pathways. This method depicts the pathways and dynamics at greatly reduced simulation time. However, it falls short of accuracy when the free energy change is much greater than k_BT demanding usually more than 1 million trajectories (Ytreberg and Zuckerman, 2004). Accurate binding affinity can be used for parameterization of the coarse-grained protein models.

Another aspect of the protein-protein interactions is the kinetic binding rates k_a or k_{on} which describes how fast the two proteins associate. There are plenty of experimental techniques developed for studying kinetics of protein-protein association. In 1923, the stopped-flow technique was introduced to reduce the observable halftimes to milliseconds (Hartridge and Roughton, 1923) and later it was integrated to fluorescence experiments. In 1940, Chance (Chance, 1940) was able to use stopped-flow fluorescence experiments to measure the rapid association rate constants up to 10^7 M^{-1} s^{-1}. Advancement of techniques including surface plasmon resonance (SPR) can measure association constants over a wide range of 10^3 to 10^9 M^{-1} s^{-1} (Liedberg et al., 1983). However, the incapability of capturing

temporal transient structures makes it difficult to study the pathway and mechanism of association.

Early theory developed by Michaelis and Menten of protein-protein association kinetics divides the reaction dynamics into diffusional stage and catalyzing stage. At the diffusional stage, the rate $k_{diffusion}$ of forming a transient complex **A·B** is solely dependent on diffusion of the proteins in the solution, the electrostatic interactions between them steer their binding; at the catalyzing stage, the transient complex is assumed to rapidly form a product **AB** at a rate k_{cat}.

$$A + B \underset{k_{rev}}{\overset{k_{diffusion}}{\leftrightarrows}} A \cdot B \xrightarrow{k_{cat}} AB$$

The effective on-rate is thus written as $k_{on} = k_{\mathrm{diffusion}} \frac{k_{\mathrm{cat}}}{k_{\mathrm{cat}}+k_{\mathrm{rev}}}$, where k_{rev} is the rate of dissociation of the transient complex **A·B**. The formula was derived based on rapid equilibrium approximation $k_{cat} << k_{rev,}$ i.e. a reaction controlled process. However, it has been shown the relation stands without this approximation (Briggs and Haldane, 1925).

The upper limit association rate depends on $k_{diffusion}$, when $k_{cat} >> k_{rev}$, $k_{on} \sim k_{diffusion}$. Therefore, the diffusion controlled association arouses interest in both theory and computational studies. Roughly the diffusion-controlled and reaction controlled association regimes fall on high and low end of the spectrum of observed association rate constants (Schreiber et al., 2009), as shown in Fig. 1.7. It is widely recognized that the diffusion-controlled association does not involve large conformational changes while the reaction-controlled association usually involves global conformational changes including

domain movement (Schreiber et al., 2009). This allows analytical estimation and computational calculation of diffusion-controlled rates.

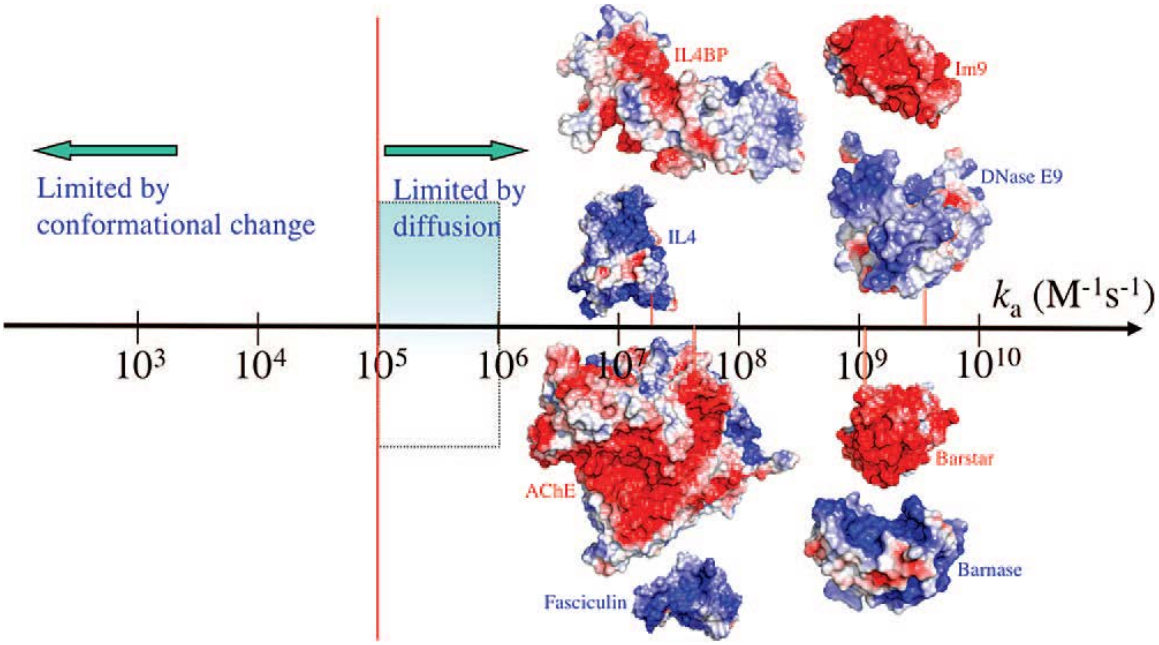

Figure 1.7 Spectrum of association rate constants. The vertical line separates the diffusion controlled and reaction controlled regime. The shaded region represents absence of long-range interactions between the binding partners. Representative protein complex in diffusion regime are provided (red lines mark the association rate constants). Figure is adapted from ref. (Schreiber et al., 2009).

1.1.5 Computational model for association kinetics between CaM and CaMBT

Firstly, assuming there is no change in the conformations of the proteins, the diffusional rate can be computed analytically first presented by Smoluchowski (Smoluchowski, 1916). Basic understanding of the diffusion picture can be arrived by considering the primitive model, neglecting all of the interactions between the two reactants. If the two reactants are approximated by two hard spheres, and spherical symmetry of reaction is satisfied, the steady-state Smoluchowski diffusion equation is written as,

$$\partial\rho(\mathbf{r},t)/\partial t = -\nabla\cdot\mathbf{D}(\mathbf{r})\cdot[-\nabla+\beta\mathbf{F}]\rho(\mathbf{r},t) - (k_i/4\pi\sigma^2)S(r)\rho(\mathbf{r},t) = 0 \quad \text{(Eqn. 1.1)}$$

where ρ(**r**,t) is the probability density of a pair of reactants separated by vector **r** at time t are unreacted (which implies when r = σ , they react and the density becomes 0); **D**(**r**) is the relative diffusion coefficient tensor, and in the primitive model, hydrodynamics is not included thus only a translational diffusion coefficient D_0 is considered; $\beta = 1/k_BT$, k_B is Boltzmann constant and T is temperature; force **F** is the gradient of a potential of mean force. The constant k_i is the intrinsic rate constant for reaction from separation σ, which is 0 in this case. S(r) is the radial δ function.

From the subsequently obtained diffusion equation,

$$\frac{1}{r^2}\frac{\partial}{\partial r}r^2 J(r) = 0 \quad \text{(Eqn. 1.2)}$$

$$J(r) = -D\frac{\partial\rho(r)}{\partial r} \quad \text{(Eqn. 1.3)}$$

The flux through any sphere that contains the reactants should be the same,

$$j(r=b>\sigma) = J(r=b)*4\pi b^2 \quad \text{(Eqn. 1.4)}$$

$$k^0{}_D(b)*\rho(b) = j(b) \quad \text{(Eqn. 1.5)}$$

From Eqns. (1.2-1.5), it arrives at Smoluchowski's solution,

$$k^0{}_D(b) = 4\pi D_0 b \quad \text{(Eqn. 1.6)}$$

Here b is the radius of a chosen sphere, which is greater than the distance between the two hard spheres. This primitive model gives insight on the diffusion rate. Analytical solutions for the rate constants usually exist focusing on the influence of one or more of the factors, such as the electrostatic interactions (Iakoucheva et al., 2002) or the hydrodynamics (Heppel and Hilmore, 1951; Lakey and Raggett, 1998; Colomer and Means, 2000). However, it is growingly more difficult, if not impossible, to consider multiple factors that affect the system simultaneously. Moreover, it is not possible to obtain analytical solutions at the atomistic or even residual level for understanding the recognition of association. One alternative method is to find numerical solution (Zientara et al., 1982) to the diffusion equation (Eqn. 1.1), whose computational demand grows with system size.

Another alternative method which is more convenient is Brownian dynamics trajectory simulation (Northrup et al., 1984). The method can easily include the hydrodynamic interactions (Ermak and Mccammon, 1978) and other arbitrary inter-molecular forces, e.g. short range solvent potential energy effects and dynamical correlations in the caging process (Northrup and Hynes, 1979; Northrup and Hynes, 1979) can be taken in account. The algorithm is widely used in the coarse-grained molecular simulations to capture the large time-scale of association. Recently, it was successfully employed in all-atomistic simulations of protein association, while still treating the protein as rigid body (Huber and McCammon, 2010; Mohan et al., 2016). As shown by Northrup, Allison and McCammon (Northrup et al., 1984), the association can be divided into the pure diffusional regime ($r > b$) as shown in Fig. 1.8 and the inner shell where intermolecular interactions play a more important role. In the pure diffusional regime, the association rate

constant is simply Smoluchowski's solution $k_D^0(b)$. In the interaction regime, the portability β_∞ for the two reactants to associate and react is $\beta_\infty = \beta / [1-(1-\beta)\Omega]$, where β is the probability that reactants achieving separation b will subsequently have at least one collision rather than escape to the outer boundary r = q and Ω is the probability that a reactant at r = q will eventually return to r = b rather escape to infinity. If a collision is considered to guarantee a reaction event, the association rate constant can be expressed as a combination of the two stages.

$$k = k_D(b)[\beta / (1-(1-\beta)\Omega)] \qquad \text{(Eqn. 1.7)}$$

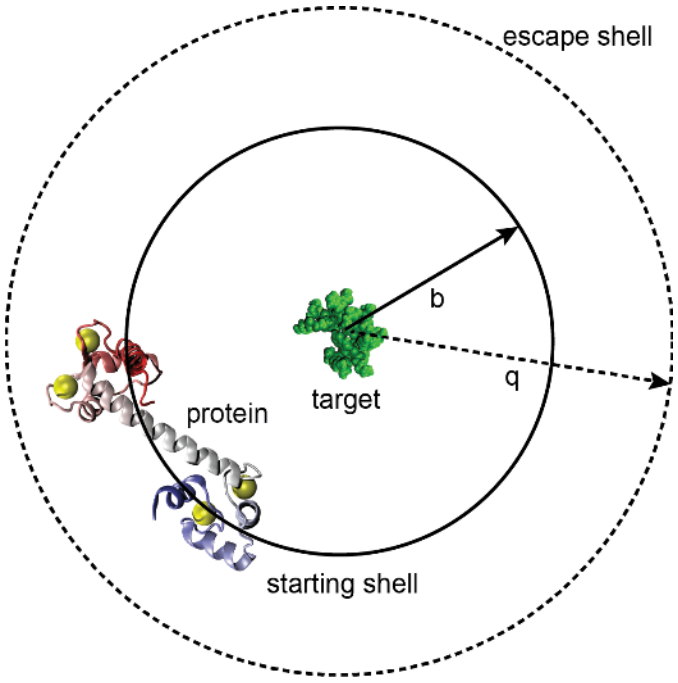

Figure 1.8 Schematic representation of the NAM algorithm of computing association rates. The b shell is the distance at which the protein is distributed around the target. r = b is chosen sufficiently large so that the interaction between the protein and the target is centrosymmetric and the flux going through is isotropic. r = q is the longest distance in the simulation to track the position of the protein. q >> b, so that the probability of the protein coming back is significantly lower than that at r = b. In the simulations, q = 5b.

The Brownian dynamics trajectory simulations were adopted for studying the binding kinetics between CaM and two of the binding target peptide CaMKI and CaMKII. The two peptides bind CaM anti-parallelly in a canonical mode and form α-helical structure. Although they share common total charges and similar size (CaMKI has 22 residues and CaMKII has 21 residues), the association rate constant between CaMKI and CaM is more than 2 times larger than that between CaMKII and CaM. This defies the prediction of pure diffusional process (Wang et al., 2013).

In order to explore the kinetics and pathways the association of CaM and CaMBT at a large timescale (100 μs to ms), I used a coarse-grained sidechain-C_α protein model, which is described in details in Appendix 6.1. In Chapter 2, the computed association rate constants computed from the Brownian dynamics trajectory simulations of CaM and the two CaMBTs individually matches the same pattern. Study of the structural information among the successful binding trajectories leads to discovery of new recognition mechanism between CaM and CaMBTs.

1.2 Protein-protein association involving intrinsically disordered protein

1.2.1 Introduction to intrinsically disordered proteins

Since the determination of the first three-dimensional structure of myoglobin in 1958 (Kendrew et al., 1958), it has been very successful to use three-dimensional structure of a protein to explain its function. It is long believed that a specific folded three-dimensional structure is necessary for its function, known as the “structure-function” paradigm. Unfortunately, the role of unstructured proteins has been overlooked since the

lack of stable three-dimensional structure under physiological conditions (Uversky et al., 2000). Although the unstructured proteins or part of the proteins were discovered as early as in 1970s by both X-ray crystallography and NMR experiments, the biological importance of them was not appreciated until Wright, Dyson (Chang et al., 1997) and Uversky et al. classified the IDPs (Uversky et al., 2000; Uversky, 2002). Using bioinformatics and exploring all the available sequences of known disordered proteins, Uversky and Dunker found that the disorder comes from the sequence of the amino-acids. Sequences leading to intrinsic disorder are low in hydrophobic and high in charged in nature. The prediction of disorder is based on the mean hydrophobicity <H> and mean net charge <R> of the amino-acids sequence.

Interestingly, the degree of disorder increases in evolution: about 9%~37% of the genomes in bacteria were estimated to be disordered and about 35%~51% in eukaryote (Dunker et al., 2000). The increased disorder indicates that perhaps in the eukaryote it requires more complex functions of the proteins in signaling, regulation and control (Dunker and Obradovic, 2001) than in the prokaryote. It was identified that among all the signaling proteins, more than 60% of the proteins contain long disordered regions (greater than or equal to 30 amino-acids) (Iakoucheva et al., 2002; Uversky et al., 2005). This implies IDPs' unique advantage in cell signaling than the ordered proteins.

It is proposed that the advantage of IDPs lie in their capability of adopting multiple conformations across energy barriers within thermal agitation. On the one hand, IDPs can switch from different states very quickly and allow fast response required in signaling. IDPs are usually hub proteins in the signaling protein networks (Dunker et al., 2005).

Moreover, the electrostatic steering effects as well as great capturing radius of IDPs can gain kinetic advantages for forming complex structures with their target proteins (Huang and Liu, 2009; Ganguly et al., 2012). Arguably the extending structure ensemble of IDPs would decrease the diffusion coefficient and thus slow down the diffusional stage of association. I will review computational models for studying protein protein interactions (PPIs) involving IDPs.

1.2.2 Computational models for intrinsically disordered proteins in protein protein interactions

IDPs mostly fold via binding with other binding partners, including proteins and deoxyribonucleic acids (DNAs) among others. The lack of stable structures can facilitate binding to several different targets of high associate rates and moderate or low affinity. IDPs usually persist residual structure that samples the structure in the bound complex (Fuxreiter et al., 2004). The conformational propensities were found to influence the mechanism of binding and folding: folding of the IDP after binding refers to “induced fit”; binding to a pre-existed state of the IDP refers to “conformational selection”; or any combination of the two (Arai et al., 2015). It raises difficulty in modelling the IDP in association using a structure based model because the IDP can adopt distinct conformations upon binding with different targets and reserve preformed residual structure in the unbound state.

Experimentally circular dichroism (CD) and chemical shifts from NMR have been widely used to measure the residual structure of the IDPs, which makes it possible to develop a universal technique for modeling IDPs. Although protein models based on a

single structure (in one of the bound complexes) (Wang et al., 2011; Ganguly et al., 2012) or two structures have been used by a few groups for studying the binding mechanisms of IDP association, inevitably they would do poorly in reconstructing the unfolded ensemble of the IDP which dynamically maintains the transient secondary structures. Atomistic protein models are still generally costly despite the advance in the computer speed. Moreover, all the current force fields are developed according to ordered proteins and unnecessarily they are equally applicable to intrinsically disordered proteins. Nevertheless, techniques such as discrete molecular dynamics (Szollosi et al., 2014) or multi-scale protein models (Lee and Chen, 2016) have been actively developed to enhance the sampling.

In my model, based on the above properties of IDP, a statistical dihedral angle potential (Karanicolas and Brooks, 2002) was implemented as well as the nonspecific electrostatic interactions, van der Waals interactions to sample the nonspecific unbound state broadly. Note that the statistical dihedral angle potential and solvent mediated interactions are both derived from folded stable structures from Protein Data Bank (PDB). This shortcoming may result in inaccuracy in the calculation. Applying this protein model to neurogranin, I was able to reconstruct the residual structure measured by chemical shifts. The bound complex formed by Ng peptide and apoCaM was also determined from my coarse-grained simulations. This provide a fundamental structure basis for studying the reciprocal relationship between CaMBT, CaM, and the Ca^{2+}.

1.3 Protein protein interactions regulate Ca^{2+} signaling for synaptic plasticity

1.3.1 Memory formation, synaptic plasticity and Ca^{2+} signaling

One remarkable feature of the mammalian brain is its capacity to encode and integrate new information through maintaining memories. Learning and memory formation are mediated by plasticity of the synapses that connect neurons. The synaptic plasticity includes a long-lasting increase in the efficacy of the glutamatergic synaptic transmission, namely long-term potentiation (LTP). It is well-established that the strength and number of synaptic connections can undergo rapid and extensive changes after sensory alterations and learning throughout life (Whitlock et al., 2006).

Activity of the Ca^{2+}-CaM-dependent kinase II (CaMKII) and the glutamate receptor, N-Methyl-D-aspartate (NMDA) receptor, are required for LTP induction. When NMDA receptors are activated, the external signals encoded in Ca^{2+} flux are transmitted to Ca^{2+} sensor proteins, e.g. CaM, resulting in activation of CaMKII, as shown in Fig. 1.9. Activation of CaMKII during LTP further increases the size of the glutamatergic synapses and promotes learning and memory.

Another aspect of the Ca^{2+} signaling is the availability of CaM. Interestingly, a kind of CaMBT, neurogranin (Ng), known as a substrate of the protein kinase C. Ng is highly expressed near the membrane of the neurons. Ng does not have any enzymatic activity, however, reduction of Ng concentration in the frontal cortex and hippocampus was observed in Alzheimer's diseases (Chang et al., 1997; Davidsson and Blennow, 1998). Other experiments of Ng knock-out mice show depress in the activation of CaMKII (Pak

et al., 2000). Those studies suggest that Ng plays a dual role in regulating Ca^{2+} signal transduction by binding with CaM: (I) localizing the available CaM to the membrane (2) decreasing the Ca^{2+} affinity for CaM; thus at low Ca^{2+} concentration in the resting cell, CaM does not activate CaMKII.

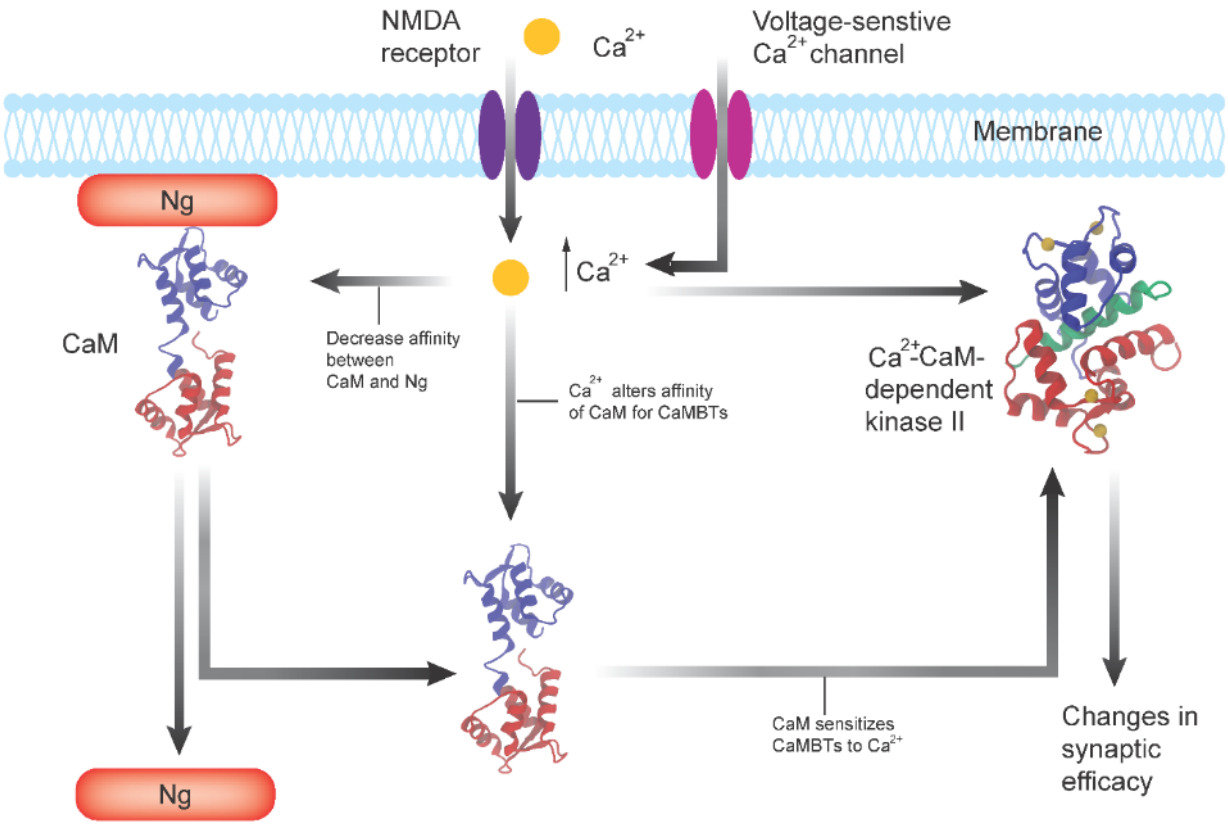

Figure 1.9 Overview of the effects of CaM-dependent Ca^{2+} signaling and effects of CaMBTs on changes in synaptic plasticity. Many of the effects of intracellular Ca^{2+} on synaptic plasticity are mediated through CaM-regulated proteins. Increase in intracellular Ca^{2+}, generated through the activity of NMDA (N-methyl-D-aspartate) receptors or voltage-sensitive Ca^{2+} channels, results in the release of CaM that is bound to Ng. CaM mediates the Ca^{2+} stimulation of CaMKII which is required for changes in synaptic plasticity. The structures of apoCaM (PDB ID: 1CFD) and Ca^{2+}-CaM-CaMKII peptide (PDB ID: 1CDM) are provided. CaM is colored: red → nCaM (residue 1 to 76), grey → central linker (residue 77 to 82) and blue → cCaM (residue 83 to 148); the CaMKII peptide is colored in green. Picture is reproduced from Figure 1 in ref. (Xia and Storm, 2005).

1.3.2 My computational model for studying Ca^{2+} affinity for CaM

To explore how the binding of CaMBTs to CaM affects Ca^{2+} affinity, I calculated the Ca^{2+} affinity to cCaM in the presence or absence of the CaMBTs from molecular simulations. Since the binding between the Ca^{2+} and the CaM involves the formation of ionic bonds between Ca^{2+} and oxygen-donors from the coordination residues in the Ca^{2+} binding loops (Fig. 1.4), I used the all-atomistic protein models with explicit water molecules. For this study, as mentioned in 1.1.3, the rough and fast estimate of Ca^{2+} affinity is calculated from non-equilibrium paths from the Ca^{2+}-bound to Ca^{2+}-unbound state based on Jarzynski's equality (Jarzynski, 1997).

1.4 Thesis Organization

The following chapters are organized as follows:

In Chapter 2, I studied the protein-protein interaction on CaM and two binding targets CaMKI and CaMKII. I reported a mechanism of mutually and conformational induced fit for their recognition.

In Chapter 3, I studied the coarse-grained computational modelling of intrinsically disordered proteins. I reported that the KB dihedral angle potential better represents the transition state of small globular protein CI2 and it reconstructs the residual structure ensemble of the IDP, Ng.

In Chapter 4, I studied the regulation of Ca^{2+} binding stability to calmodulin by interaction between calmodulin and the target proteins. I used multi-scale molecular simulations to show that structurally why CaMKII peptide and Ng peptide have opposing effects on Ca^{2+} affinity for CaM.

In Chapter 5, I summarized the main conclusions of this thesis.

Chapter 6 is appendices with details about the coarse-grained models, other analysis tools, data sets used in the previous chapters, and tables of the partial charges.

Chapter 2

Protein recognition and selection through conformational and mutually induced fit

The majority of the work presented in this paper is published in Proceedings of the National Academy of Sciences, 110(51): 20545-20550 (2013) (P. Zhang is the co-first author of this paper).

Part of the work is published in Journal of Molecular Recognition, 28, 74–86 (2015) (P. Zhang is the third author of this paper).

2.1 Introduction

The ubiquitous protein calmodulin (CaM) interacts with a vast selection of binding targets (CaMBTs); however, the molecular mechanisms that underlie target selectivity are not known despite an enormous wealth of structural information (Yamniuk and Vogel, 2004; Tidow and Nissen, 2013). What emerges from this is the remarkable conformational flexibility of CaM, which exists in highly dynamic structures in the Ca^{2+} free and bound forms (Slaughter et al., 2005; Homouz et al., 2009; Anthis et al., 2011; Wang et al., 2011) and will adopt distinct conformations when bound to protein targets. These distinct modes of binding are encoded by CaM-recognition motifs of targets that display impressive variability in amino acid sequence and are often partially or largely disordered in the absence of CaM. These data indicate that CaM-CaMBT interactions lie at the opposite end of the spectrum from the classic “lock and key” mechanism (Fischer, 1894) initially proposed for rigid ligand binding to proteins and require adopting more dynamic models such as “induced fit” (Koshland, 1958) or “conformational selection” (Monod et al., 1965). The “induced fit” mechanism posits that productive binding occurs because the rigid ligand can alter the conformation of the enzyme and that the final conformation exists only in the

presence of ligand. “Conformational selection” assumes that the enzyme naturally samples a variety of conformational states and that the ligand binds to one, or a small subset, of these states. However newer theories, termed extended conformational selection (Tsai et al., 1999), mutually-induced fit (Williamson, 2000), fly-casting (Shoemaker et al., 2000), or folding and binding (Levy et al., 2004), have begun to integrate molecule dynamics to describe folding and binding involving flexible molecules in protein-protein interactions, especially for the intrinsically disordered proteins (Wright and Dyson, 1999; Dunker et al., 2001; Uversky, 2002) for which folding and binding are concomitant.

The goal of the present study is to provide mechanistic insights at a molecular level into the time-dependent conformational adjustments between CaM and CaMBT for the binding mechanism. Results show that although CaM visits conformations similar to a target-bound structure in the absence of target, targets are incapable of binding these conformations. Instead, the binding occurs through a mechanism where both target and CaM undergo a conformational “search” for the natively bound conformation and provide an experimental foundation for defining the mechanism of mutually-induced fit. My role in this project is to carry out the molecular simulations and analyze the association data.

To establish experimental constraints for the coarse-grained simulations, Dr. Neal Waxham’s group measured the association rates of CaM with peptides representing the CaM-binding domains of Ca^{2+}-CaM-dependent kinase I (CaMKI) and Ca^{2+}-CaM-dependent kinase II (CaMKII) by stopped-flow fluorescence techniques. These peptides represent two distinct motif classes of CaM-binding molecules and high-resolution nuclear magnetic resonance (NMR) and crystal structures of the bound complexes are available

(Meador et al., 1993; Gifford et al., 2011). Motif classes are defined by the spacing of hydrophobic residues that make stabilizing contacts between CaMBTs and CaM. CaMKI contains two motifs, 1-14 and 1-5-10, while CaMKII contains only the 1-5-10 motif (Table 2.1). Given that the initial contact of both peptides interaction with CaM is diffusion-controlled, the differences in on-rates must involve transitions between initial contact and the final complex formation. To investigate the molecular basis behind such phenomena, QW and I employed a coarse-grained model and QW, ST and I performed Brownian dynamics simulation of the association between CaM and each of the two CaMBTs. Important to note is that the Hamiltonian of CaM is identical in all simulations so that differences revealed in interactions must be due to the distinct amino acid sequence differences between CaMKI and CaMKII.

Table 2.1 Residue indices and the annotation of the hydrophobic motifs for CaMKI and CaMKII. As indicated below, CaMKI has two hydrophobic motifs 1-5-10 and 1-14, whereas CaMKII has one 1-5-10.

CaMKI			CaMKII		
Motif spacing	Residue index	Residue name	Motif spacing	Residue index	Residue name
	299	ALA		292	LYS
	300	LYS		293	PHE
	301	SER		294	ASN
	302	LYS		295	ALA
1	303	TRP		296	ARG
	304	LYS		297	ARG
	305	GLN		298	LYS
	306	ALA	1	299	LEU
5	307	PHE		300	LYS
	308	ASN		301	GLY
	309	ALA		302	ALA
	310	THR	5	303	ILE
	311	ALA		304	LEU
10	312	VAL		305	THR
	313	VAL		306	THR
	314	ARG		307	MET
	315	HIS	10	308	LEU
14	316	MET		309	ALA
	317	ARG		310	THR
	318	LYS		311	ARG
	319	LEU		312	ASN
	320	GLN			

Initial computational studies on the association rates of proteins were typically based on the diffusion of rigid objects in which the structural flexibility of a protein was ignored (Elcock et al., 1999; Gabdoulline and Wade, 2001; Alsallaq and Zhou, 2007). Several computational approaches were developed using a structure-based protein model that incorporated the structural flexibility of a protein in the events of coupled folding and binding (Levy et al., 2004; Turjanski et al., 2008). Others extended this structure-based

model approach by introducing a set of experimentally determined structures in the Hamiltonian (Best et al., 2005; Okazaki et al., 2006; Whitford et al., 2007; Lu and Wang, 2008; Okazaki and Takada, 2008; Weinkam et al., 2012) (such as multiple-basin structure-based models) or by including additional non-specific interactions (Ravikumar et al., 2012) to investigate the transitions between two or more distinct conformations of a protein (e.g. calcium-bound and unbound conformations of CaM (Zuckerman, 2004; Tripathi and Portman, 2009)). All of these studies required *a priori* knowledge of specific structures of bound complexes; the mechanism of target recognition and selection among structurally flexible proteins is still elusive. In this study, QW and I overcame the limitation of existing computational models by having a Hamiltonian for CaM that is transferrable in the model and provides no bias among its conformations in the bound state. In addition, ST, QW, and I computed the association rates of CaM and CaMBTs using an experimentally constrained coarse-grained molecular simulation and investigated the molecular mechanisms in post-collisional events. A new mechanism was provided of "mutually induced-fit" for the understanding on how intrinsically disordered proteins achieve protein recognition and selection.

2.2 Materials and Methods

2.2.1 *In vitro* experiments

All the experiments were performed by Neal M. Waxham's group. Details of the experiments can be found in reference.

2.2.2 Coarse-grained modelling of the proteins

I studied the association of Ca^{2+}-bound calmodulin (CaM) and two of the CaM binding targets (CaMBTs) derived from CaM Kinase I (CaMKI) and CaM Kinase II (CaMKII) that are similar in length, as well as the structures in the bound state. QW, ST and I performed coarse-grained molecular simulations to calculate the association rates of the two CaMBTs with CaM motivated by the fluorescence experiments as described above.

A CaM protein composed of 148 amino acids (Bhattacharya et al., 2004) interacts with over 300 target proteins (Yamniuk and Vogel, 2004). It is a key regulator of cellular processes responsible for a wide variety of biological functions related to Ca^{2+}-signal transduction (Choi and Husain, 2006). It is shown that a highly malleable structure of CaM, where the N-domain (nCaM, residues 1-74) and C-domain (cCaM, residues 75-148) are connected by a flexible linker, allows a wide range of motion across the two domains (Crivici and Ikura, 1995; Hoeflich and Ikura, 2002; Bhattacharya et al., 2004; Yamniuk and Vogel, 2004; Slaughter et al., 2005; Homouz et al., 2009; Price et al., 2010). Several hydrophobic residues on CaMBTs are hydrophobic motifs so that the spacing of the conserved amino acids can be used to classify CaMBTs. CaMKI has two hydrophobic motifs (1-5-10 and 1-14) and CaMKII has only one (1-5-10) (Please see the Table 2.1 for the amino acid sequences of the peptides). These CaMBTs, which are characterized as intrinsically disordered peptides (IDP), lack the structural information in solution. However, they form short helices when bound to CaM.

The simulation was executed using a coarse-grained protein model that allows the exploration of conformational changes during the course of association. A side chain C_α

model (SCM) (Cheung et al., 2003) that includes two beads per amino acid (except glycine) was adopted to represent CaM and each of the two CaMBTs (CaMKI and CaMKII). Despite its low resolution, this model captures the polymeric behavior of a protein and retains the chemical characteristics of side chains. The modeling of CaM and four of its embedded calcium ions can be found in previous work (Wang et al., 2011). The sequences of CaMKI and CaMKII peptides were taken from the experimental studies in reference (Marlow et al., 2010) and (Forest et al., 2008), respectively.

The Hamiltonian is described in the Appendix 6.1. For CaM φ_{ijkl}^{0} is calculated from an extended form without a target (PDB ID: 1CLL), and $k_{\phi}^{1} = 40\ kcal/mol$ and $k_{\phi}^{3} = 20\ kcal/mol$. With these parameters, the average radius of gyration of a model CaM in an unbound state is ~21.2 Å, close to the experimental value 21.3 Å measured by the X-ray scattering experiments (Heidorn et al., 1989). The distribution of the radius of gyration of CaM (Fig. 2.1) shows two peaks corresponding to the extended state and the compact state at the ratio of 9.25 to 1, close to 9 to 1 that was observed experimentally by the paramagnetic relaxation enhancement method (Anthis et al., 2011).

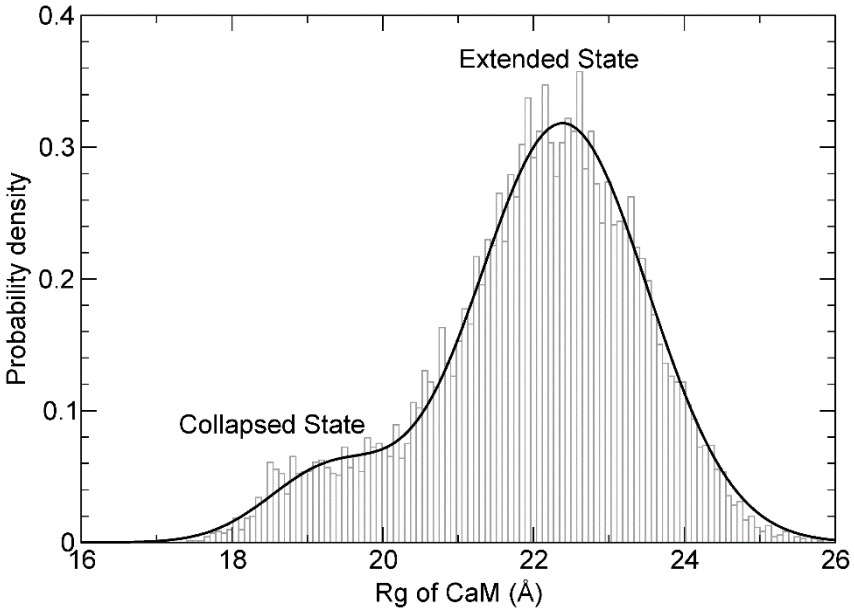

Figure 2.1 Distribution of the radius of gyration of holo-CaM. Two populations that represent the collapsed and extended states are observed. The histogram is fitted to bimodal normal distribution in black solid curve $f\left(Rg\right) = 0.21e^{-0.75(Rg-19.3)^2} + 1.21e^{-0.40(Rg-22.4)^2}$, in which the ratio of the areas beneath the curves that represents the collapsed state and the extended state is about 1:9.

2.2.3 Initialization of the kinetics simulation

A constrained molecular simulation was used to produce initial configurations at temperature $k_BT/\varepsilon = 1.1$. The distance between the center of mass of CaM and the center of mass of the binding target was restrained by a harmonic force at 15 σ (where the interactions between two proteins are isotropic at this distance, σ is 3.8 Å) with a spring constant 66.7 ε/σ^2. A total of 1000 statistically different initial configurations of CaM and CaMBT were obtained for the association simulation.

2.2.4 Association simulations

Coarse-grained molecular simulations were performed using an in-house version of AMBER10 (Case et al., 2008) to investigate the association process through the Brownian equations of motion. The viscosity of the system was set to 0.001 Pa· s in aqueous solution. QW and I started with an algorithm developed by McCammon group (Northrup et al., 1984). The association rate *k* was calculated by Eqn. (2.1).

$$k = 4\pi Db\left[\frac{\beta}{1-(1-\beta)\Omega}\right] \qquad \text{Eqn (2.1)}$$

D is the diffusion coefficient of the system, which is the sum of the diffusion coefficients of the target and CaM. In my calculations, I used the experimental measurement of *D* (Slaughter et al., 2004) for CaM (7.8×10^{-7} cm^2/s). For the targets, I approximated the diffusion coefficient to be two times that of CaM based on the ratio of their sizes in terms of their radius of gyrations, since experimental values were not available. β is the probability of successful association events. CaM was randomly distributed in a spherical surface that is 15 σ away from the center of mass of a binding target. The interactions between CaM and the target are isotropic beyond this distance; therefore, b=15 σ. Ω is the returning rate, which was set to 0.2 in the simulation (Northrup et al., 1988; Northrup and Erickson, 1992; Gabdoulline and Wade, 1997). In other words, the simulation stopped when the distance between CaM and a binding target exceeded 75 σ. Else, the simulation reached the maximum time of 240,000,000 time steps.

2.2.5 Summary of simulation methods

The association process between CaM and CaMBT (CaMKI or CaMKII) was investigated by Brownian Dynamics (BD). The rate of association was computed by an algorithm developed by Northrup, Allison and McCammon (NAM) (Northrup et al., 1984). Further discussion on the NAM algorithm and diffusion coefficient is provided in the *Supplement*. The viscosity of the system was set to be equal to the aqueous solution 1.0×10^{-3} Pa·s at the ztemperature$k_B\,T/\varepsilon = 1.1$, where $\varepsilon = 0.6$ kcal/mol and k_B is the Boltzmann constant. In the BD simulation of CaM-CaMBT association, an initial configuration of CaM was randomly distributed in a spherical surface at b = 15 σ (σ is a reduced unit length in the coarse-grained model, which equals to 3.8 Å) away from a CaMBT. The simulation of a trajectory would be stopped when the distance between center of mass of CaM and the center of mass of the CaMBT exceeded q = 75 σ (see Fig. 1.8). Otherwise a total time of 240,000τ was carried out for each trajectory where an integration time step was 0.0005 τ. In an overdamped limit (Veitshans et al., 1997), τ was mapped to the real time between 0.01-0.37 ns as described in the *Supplement*. For each system, 1000 association trajectories were performed. Criteria for a successful association event are discussed in the next section.

2.2.6 Analyses

Definition of successful and unsuccessful associations. The definition of a successful association event of a trajectory from the simulations was guided by the experimentally measured association rates (Wang et al., 2013), shown by previous study. Here, I summarize the method: in order to determine a successful association process between CaM and a CaMBT, I monitored the formation of side-chain contacts between the

residue Lys75 (of CaM) and any amino acids from the CaMBT, which was denoted as Z_{75}. ST and I used a cut-off distance of 2σ (σ is the reduced unit of length in the coarse-grained model, which equals to 3.8 Å) to define a contact. The choice of Z_{75} was motivated from the stopped-flow fluorescence study used to determine association rates (Wang et al., 2013) in which the reporter dye was positioned on residue Lys75 at the N-terminal end of the linker region of CaM after mutating it to Cys75.

In the simulation a successful association event was determined once the threshold of Z_{75} reached a value of 9 (Wang et al., 2013). This is a value where the ratio of the association rates between CaM-CaMKI and CaM-CaMKII closely matches with that measured experimentally. QW and I set the maximum time of 240,000 τ to ensure that the portion of the trajectories that fail to reach the b-shell or the q-shell is less than 3% of the total trajectories. I calculated the estimated error in Table 2.2 and I concluded that the statistics from the sampling and the justification of the termination of a simulation was sound.

Contact analyses for the N- and C-lobes of CaM: Definition of Z_n and Z_c. First, ST defined a parameter Z to calculate the number of intermolecular side chain contacts between the CaM and the CaMBTs. ST used a cut-off distance of 2σ to define a contact when he calculated Z. He divided the total number of intermolecular contacts Z in two components: Z_n and Z_c. Z_n is defined as the number of intermolecular contacts between nCaM (counted from residue 1 to residue 75 of CaM) and CaMBT. Similarly, Z_c is defined as the number of contacts between cCaM (counted from residue 76 to residue 148 of CaM) and CaMBT. For successful trajectories Z, Z_n, and Z_c were normalized by their

corresponding maximum values observed from our simulation, labeled as $\bar{Z}$, $\overline{Z_n}$, and $\overline{Z_c}$, respectively.

2.2.7 Atomistic molecular dynamics simulations

The atomistic initial structures were reconstructed from sidechain-C_α configurations of the final complexes in the Brownian dynamics trajectory simulation (Wang et al., 2013), using SCAAL method (Samiotakis et al., 2010). All molecular dynamics simulations were performed using version 4.6.5 of the GROMACS molecular dynamics package. The proteins were modeled with two force fields, AMBER-99SB-ILDN force field (Lindorff-Larsen et al., 2010) and CHARMM27 force field with CMAP corrections for protein and ions (Bjelkmar et al., 2010). The rigid three-site TIP3P model (Jorgensen et al., 1983) was used to simulate water molecules. The systems were neutralized by Na^+ and Cl^- ions, maintaining an ionic strength of 100 mM. The systems were minimized using steepest descent method before a short 0.1 ns NVT equilibration at 300 K and a following 1 ns NPT equilibration with proteins fixed. The proteins were afterwards released and were further equilibrated for 5 ns. Another 10 ns NPT simulation was carried out for data collection. All NPT simulations maintained a constant pressure of 1 bar and temperature of 300 K using the Parrinello-Rahman barostat (Parrinello and Rahman, 1981). The bond lengths in proteins are constrained using the LINCS algorithm of Hess (Hess et al., 1997). The equations of motion were integrated using a 2 fs time steps. Snapshots were saved for analysis once every 1 ps. Periodic boundary conditions were employed to mimic the macroscopic setting for electrolytes. Long-range electrostatic interactions between periodic images were treated using the particle mesh Ewald approach (Darden et al., 1993), with a grid size of 0.16 nm, fourth-order cubic interpolation and a

tolerance of 10^{-5}. Neighbor lists were updated every 5 time steps. A cutoff of 10 Å was used for van der Waals interactions and real space Coulomb interactions as well as for updating neighbor lists.

2.3 Results and Discussion

2.3.1 Association rates of CaM-CaMBT

The association rates of CaM-CaMBT k_{on} was measured by Waxham's group. k_{on} was determined by the slope of k_{obs} values as a function of peptide concentration (Fig. 2.2A). On rates were 37.88 x 10^7 and 15.40 x 10^7 $M^{-1}s^{-1}$ for CaMKI and CaMKII peptides, respectively. This difference in the on-rates cannot be accounted for from differences in the diffusion rates, as the peptides are not significantly different in size (2.5 and 2.3 kDa, respectively) and diffuse more quickly than the larger CaM molecule (16.8 kDa). The experimentally measured association rates of CaMKI and CaMKII with CaM were used to guide the determination of the computed association rates for the two targets from our simulations (Table 2.2). The definition of a successful association lies in the threshold of the number of nonspecific contacts (Z_{75}).

The intermolecular contacts Z_{75} is defined in the following. In the stopped-flow fluorescent experiment, the dye that was attached to Cys75 of CaM probes the environmental changes upon association with the CaMBTs. Analog to this, I used the number of intermolecular contacts (Z_{75}) between CaM and CaMBTs to computationally assess the trajectories for the calculation of the association rates. Z_{75} is defined as the total number of side-chain contacts from a target interacting with the amino acid at position 75

of CaM. A distance of 2 σ is used as the cut-off to define a contact between the side-chain bead of this residue of CaM and any side-chain bead of CaMBT.

Table 2.2 Calculated association rates obtained by simulations. The association rates k_a were computed from Eqn. 2.1. β is the probability of successful association, N is the number of successful association out of 1000 simulations. Given mean of the probability of each trajectory to be successful binding event (β), the number of successful trajectories follows Poisson distribution. The mean and propagating standard deviation of association rates were calculated. Standard error, defined by $\sigma/\sqrt{N}$, is provided. Both methods gave similar results that the standard error of k_a is small when Z_{75} is set to 9. This result indicates that the association rate calculations are statistically meaningful when the threshold is set to be Z_{75}=9.

Threshold Z_{75}	CaM-CaMKI			CaM-CaMKII		
	N	β	k_a (10^7 $M^{-1}s^{-1}$)	N	β	k_a (10^7 $M^{-1}s^{-1}$)
5	512	0.512±0.001	572.68±0.71	517	0.517±0.001	577.63±0.71
6	359	0.359±0.001	415.64±0.64	402	0.402±0.001	460.88±0.66
7	237	0.237±0.001	282.29±0.55	230	0.230±0.001	274.41±0.54
8	148	0.148±0.000	180.07±0.45	114	0.114±0.000	139.85±0.40
9	45	0.045±0.000	56.14±0.26	18	0.018±0.000	22.61±0.17
10	2	0.002±0.000	2.52±0.06	1	0.001±0.000	1.26±0.04

When Z_{75} is less than 5, the computed association rates (k_a) for both CaMKI and CaMKII are similar. It indicates that the basic mechanism of association involving diffusion is the same for both targets at the beginning of association. As the threshold of Z_{75} grows to 9, the computed rates for the two targets differ by a factor of 2, which closely matches with the experimental data (k_a of CaM-CaMKI is 56.14×10^7 $M^{-1}s^{-1}$ and k_a of CaM-CaMKII is 22.61×10^7 $M^{-1}s^{-1}$). Therefore ST used the threshold of Z_{75} = 9 as the criterion to define the end of the association process in the simulations.

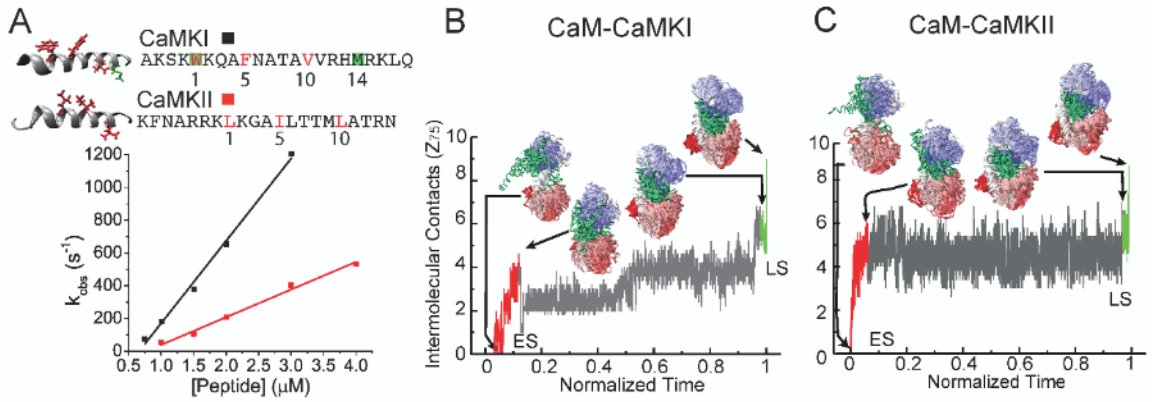

Figure 2.2 Association of calmodulin (CaM) and calmodulin binding target (CaMBT). (A) Hydrophobic motifs are highlighted in CaMKI and CaMKII peptide structures and sequences. Rates for association, k_{obs}, of CaMKI (black) and CaMKII (red) peptides with CaM were measured and plotted as a function of peptide concentration. A linear function was used to fit the data where the resulting slope represents the on rate, k_{on}. On rates were determined to be 37.88 x 10^7 and 15.40 x 10^7 $M^{-1}s^{-1}$ for CaMKI and CaMKII peptides, respectively. (B) shows the two major transitions in typical trajectories of the association between CaM and CaMKI and (C) shows that between CaM and CaMKII. The number of nonspecific contacts between the 75th residue on CaM and the target (please see the supplement for the definition) is plotted as a function of the normalized time. The superimposed structures of the CaM-CaMBT complexes are shown for the first transition (early stage, ES, in red). Similarly, the complex structures are shown for the second transition (late stage, LS, in green). CaM from N- to C- termini is colored from red to blue. CaMBT is colored in green.

2.3.2 Early and late stages of association

A successful association of CaM and CaMBTs can be described by two major transitions in a Z_{75} vs time plot (Fig. 2.2B, 2.2C and Fig. 2.3, 2.4): the early stage transition (ES) and the late stage transition (LS). ES denotes the first sharp transition shown in red (Fig. 2.2B and 2.2C for CaMKI and CaMKII, respectively). Post initial contact, the onset of association between CaM and CaMBTs involves the initial formation of a “transient complex”, at the ES transition. At this stage, the targets preferentially bind to the C-domain

of CaM (cCaM) and this initial encounter can be modeled as a diffusion-controlled association. This initial event is followed by a second structural transition or the late stage (LS) transition encompassing the last 2% of a trajectory, which is shown in green in Fig. 2.2B and 2.2C, where the N-domain of CaM (nCaM) collapses towards, and wraps around both CaMKI and CaMKII.

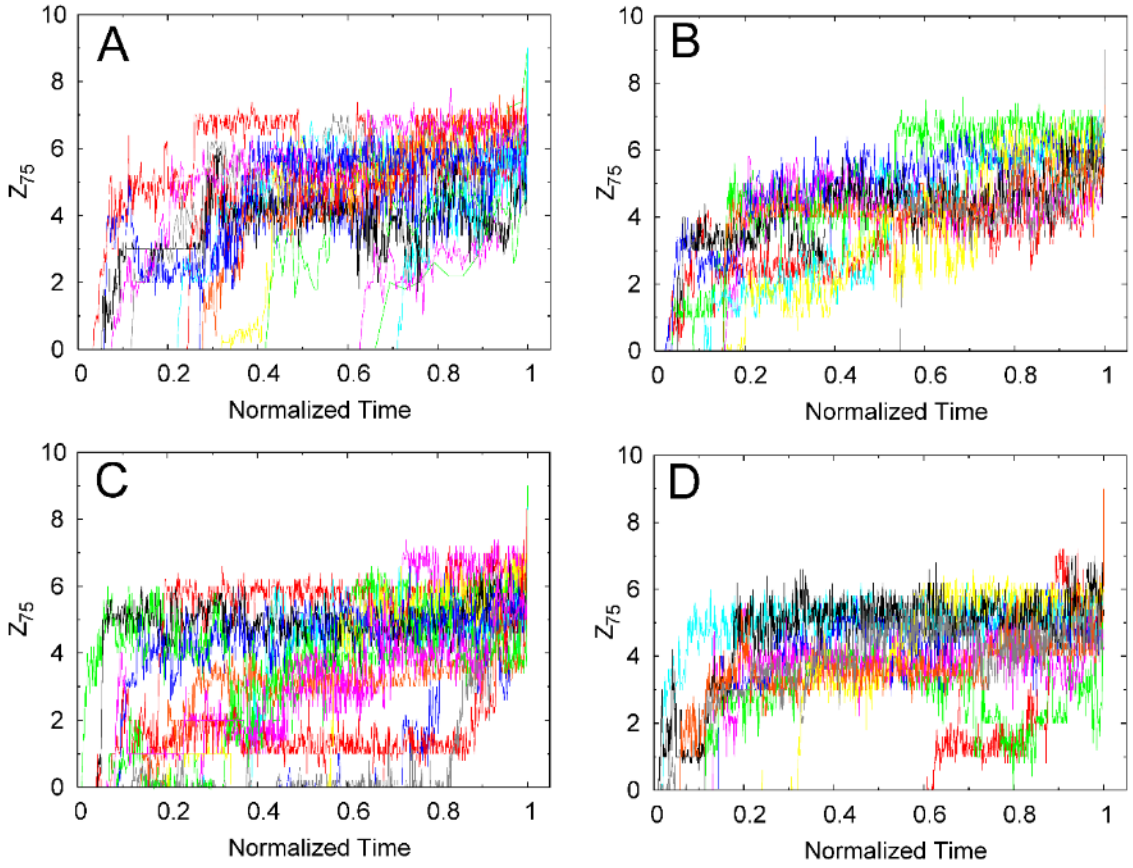

Figure 2.3 All association trajectories of calmodulin (CaM) and CaMKI. The number of intermolecular contacts between the 75th residue on CaM and the target (Z_{75}) is plotted as a function of the normalized time for all of the successful association trajectories. Two major transitions were detected in all successful trajectories between CaM and CaMKI.

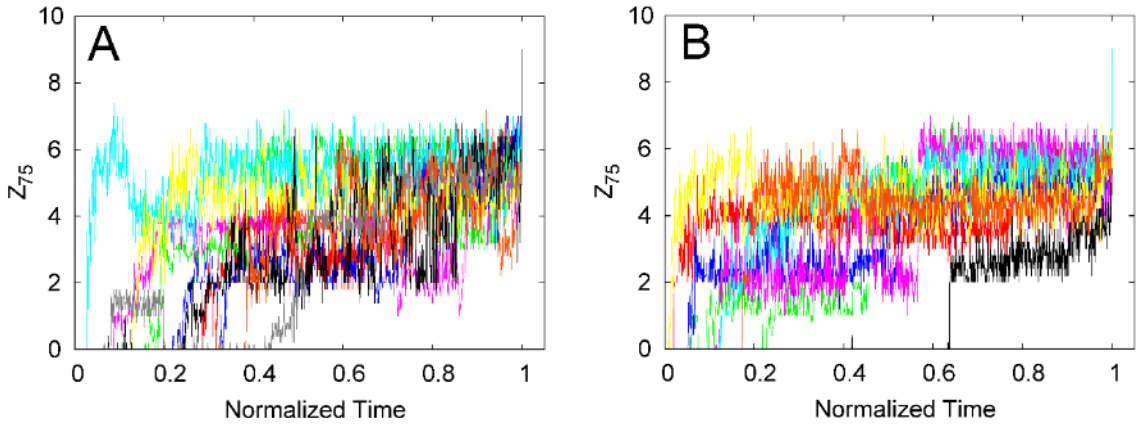

Figure 2.4 All association trajectories of calmodulin (CaM) and CaMKII. Two major transitions were detected in all successful in all trajectories of the association between CaM and CaMKI and. The number of intermolecular contacts between the 75th residue on CaM and the target (Z_{75}) is plotted as a function of the normalized time for all of the successful association trajectories. Two major transitions were detected in all successful trajectories of the association between CaM and CaMKII.

Table 2.3 Structural assessment of CaM and the target at the early and late stages of association.

CaM-CaMKI				
	Δ^a (CaM)	S^b (CaM)	$D^c_{75\text{-target}}$ (σ)	CO^d (CaMKI)
ES1	0.43±0.01	0.56±0.02	6.45±0.24	0.25±0.01
ES2	0.32±0.01	0.34±0.02	3.37±0.17	0.25±0.01
LS1	0.27±0.01	0.26±0.01	1.82±0.05	0.25±0.01
LS2	0.27±0.01	0.24±0.01	1.52±0.03	0.26±0.01
CaM-CaMKII				
	Δ (CaM)	S (CaM)	$D_{75\text{-target}}$ (σ)	CO (CaMKII)
ES1	0.45±0.01	0.60±0.02	6.28±0.36	0.25±0.01
ES2	0.28±0.02	0.28±0.03	3.20±0.29	0.28±0.02
LS1	0.20±0.01	0.15±0.02	1.84±0.11	0.32±0.02
LS2	0.20±0.01	0.15±0.02	1.53±0.06	0.33±0.01

[a]Asphericity Δ of CaM (Homouz et al., 2008) is defined between 0 and 1. 0 refers to the shape of a sphere and 1 refers to rod.

[b]Shape parameter S of CaM (Homouz et al., 2008) ranging between -0.25 and 2 is defined as following: S < 0 is oblate, S = 0 is spherical, and S > 0 is prolate.

[c]$D_{75\text{-target}}$ is the distance between C_β of 75th residue from CaM and the center of mass of the CaMBT. σ is the reduced unit of length in the coarse-grained model, which equals to 3.8 Å.

[d]CO is the contact order of the target. During the association with CaM the targets may select distinct conformations upon forming the bound complex. Therefore, in order to characterize different conformations of the target I calculate the CO, which is defined as the average sequence distance between all pairs of contacting residues normalized by the total sequence length (Plaxco et al., 1998) $\mathrm{CO} = \frac{1}{\mathrm{L}\cdot\mathrm{N}}\sum_{\{i,j\}}^{\mathrm{N}} \Delta S_{ij}$,where L is the total sequence length, N is the number of contacts, ΔS_{ij} is the sequence separation between contacting residues *i* and *j*. A contact between a pair of amino acids *i* and *j* is formed when either the distance between C_α bead of amino acid *i* and C_α bead of amino acid *j* or the distance between C_β bead of amino acid *i* and C_β bead of amino acid *j* is less than 2σ (σ = 3.8 Å).

To further explore the CaM-CaMBT association process, I analyzed the different stages of association by calculating different parameters in Table 2.3. At the beginning of the early stage (ES1), the target and CaM is still far apart (the distance between residue 75 of CaM and the center of mass of the target, $D_{75\text{-target}} > 6\ \sigma$ in Table 2.3). Initially both CaMKI and CaMKII bind to cCaM, as shown in the probability of intermolecular contact map at ES1 between CaM and CaMBT (see Fig. 2.5 (A, E)), while the shape of CaM is rather extended (S~0.6 in Table 2.3). For both the CaMBTs the relative contact order (CO) is same (CO = 0.25 in Table 2.3) at ES1. At the end of the early stage (ES2), $D_{75\text{-target}}$ decreases significantly to ~ 3 σ accompanied by a significant decrease in the shape parameter of CaM at S ~ 0.3. As CaM becomes less extended, the CaMBTs interact with both cCaM and nCaM, shown by the enhanced contact probability between the target and Helix B and C, as well as Helix F and G in Fig. 2.5 (B, F) for CaM-CaMKI and CaM-CaMKII, respectively. Moreover, the CO of CaMKI remains unchanged, whereas for CaMKII the CO increases to 0.28 (Table 2.3).

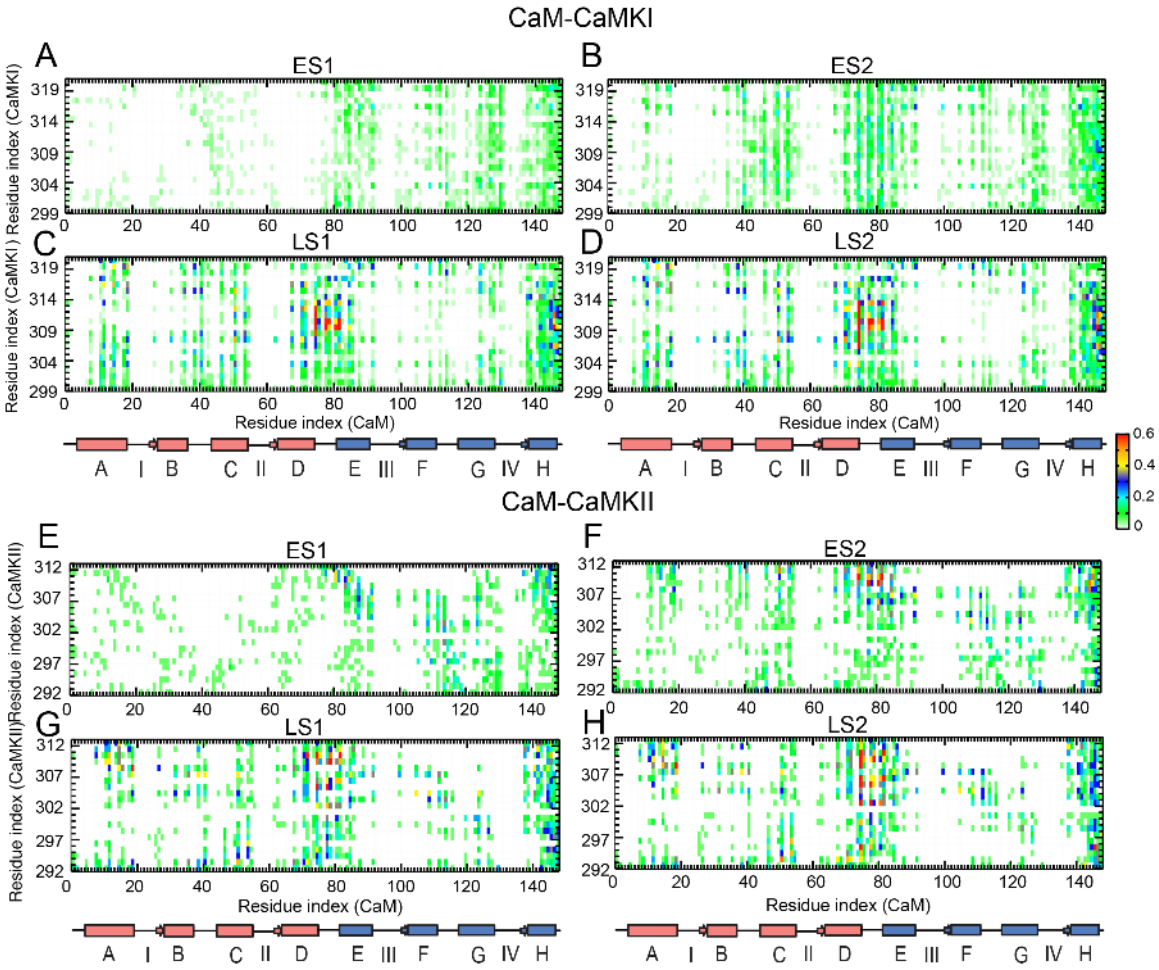

Figure 2.5 Probability of the intermolecular contacts between CaM and CaMBTs in an encounter complex at the different stages of association. (A, B, C, and D) represent the probability of contact formation between the side-chain residues from CaM and the side-chain residues from CaMKI at ES1, ES2, LS1, and LS2, respectively. (E, F, G, and H) represent those between CaM and CaMKII at ES1, ES2, LS1, and LS2, respectively. In the color bar green (red) indicates low (high) contact probability. Schematic representations of the secondary structures of CaM are provided, colored in red (nCaM) and blue (cCaM). (The figure was created by Dr. ST.)

Between the two major transitions, from ES2 to the beginning of LS1, I observed that $D_{75\text{-target}}$ significantly decreases to ~1.8 σ for both CaM-CaMKI and CaM-CaMKII (Table 2.3). The position of a target moves closer to the binding pocket, as the shape of CaM becomes a spheroid. There exist noticeable differences in the structural characteristics of CaM and CaMBTs when transitioning from ES) to LS: CaM from CaM-CaMKII experiences a greater change in shape (from S = 0.28 to S = 0.15) than the CaM from CaM-CaMKI (from S = 0.34 to S = 0.26). Also, the CO of CaMKI still remains unchanged and CO of CaMKII increases to 0.32. Between ES2 and LS1 a significant amount of contacts to the cCaM are lost (especially from helix F and G of CaM) while those to the nCaM increase (Fig. 2.5 C, 2.5 G). At the end of LS (LS2), I observed a further decrease in $D_{75\text{-target}}$ to ~ 1.5 σ for both systems without further changes in the overall shape of CaM, as well as the probability of contact formation (Figs. 2.5 D and 2.5 H). The CO of CaMBTs increases to 0.26 and 0.33 for CaMKI and CaMKII, respectively. The change in CO of CaMBTs at the different stages of association further infers that although both CaMKI and CaMKII adopt similar conformation at the initial stages (ES1 and ES2) of binding, significantly distinct conformations are apparent near the end of association (LS1 and LS2). This study showed that upon binding with CaM, CaMKII adopts both helical and bent helical conformations, whereas CaMKI exhibits mostly helical structure. The overall helicity (H)(Wang et al., 2011) of CaMKI and CaMKII throughout the association, however, is not significantly different (H_{CaMKI} and its standard error ranges from 0.41 ± 0.04 to 0.46 ± 0.04 and H_{CaMKII} and its standard error ranges from 0.34 ± 0.04 to 0.42 ± 0.06). This is because H only captures the contact formation of the helical pitches, while

CO entails both the contact formation of the helical pitches and the contact formation at a greater separation in the sequence of a peptide.

Based on the analysis presented here, *the sequence variation in the two targets must dictate the distribution of structures, and that the unique distribution of conformations influences the overall rate of the final complex formation.*

2.3.3 Structural changes in CaM and CaMBTs during the course of association

To investigate the above hypothesis, ST and I next analyzed the structural variation of the targets upon association measured by the root mean square fluctuations (RMSF) of CaMKI and CaMKII at the different stages of association (Figs. 2.6 A and 2.6 B, respectively). When the association progresses from ES to LS, the amplitude of the structural fluctuation at the middle segment of CaMKI diminishes, whereas for CaMKII the residues at the C-terminus diminish, indicating unique sequence-dependent differences between the two targets dictate the final complex formation at the LS.

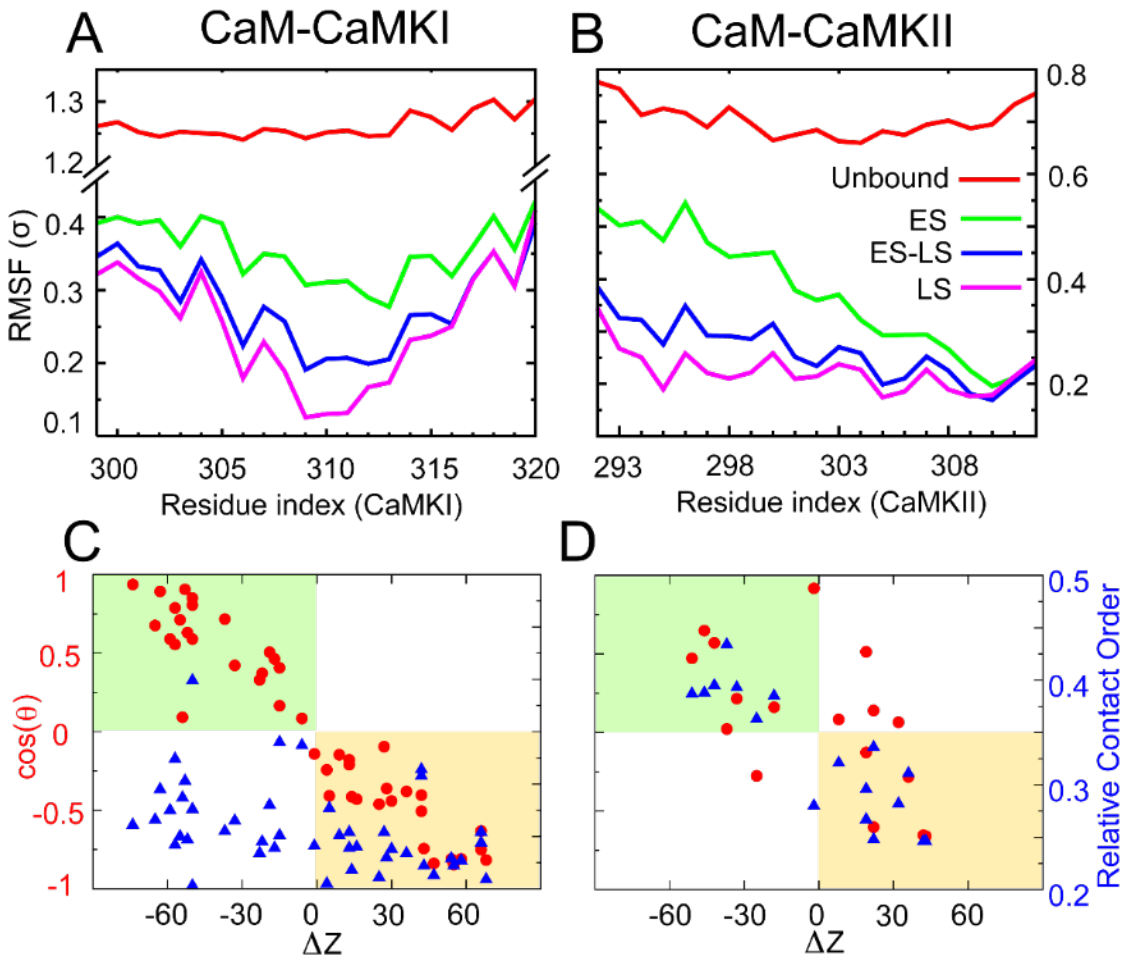

Figure 2.6 Conformational change and orientation of the targets during association. (A, B) Root mean square fluctuation (RMSF) of CaMKI and CaMKII were calculated using the structures from the unbound ensemble (red line), the early stage transition (ES, green line), the conformational adjusting stage (ES to LS, blue line) and the late stage transition (LS, pink line) as defined from Fig. 2.2 using successfully associated trajectories. (C, D) The structural characteristics and the orientations of a target in the final functional complex of CaM-CaMKI and CaMKII, respectively.

To understand the role of the hydrophobic motifs of the target during the association with CaM, parameter Q_I was defined as the number of the native contacts from the NMR structure of the bound form (CaM-CaMKI, PDB ID: 2L7L; CaM-CaMKII, PDB: 1CDM) between CaM and the hydrophobic motifs from the targets (see Table 2.1). From our simulation a native contact is defined between the side-chain of the residue *i* from CaM and *j* from the residues in the hydrophobic motifs if the distance is within 1.3 σ_{ij}^{o} where σ_{ij}^{o} is the corresponding distance from the native bound complex of both the systems. The native pairs of the two systems are given as follows. In the brackets on the left are the indices of contacting residues from CaM and on the right are the indices of hydrophobic motifs from CaMBTs.

For CaM-CaMKI: (36, 312) (51, 316) (54, 316) (55, 316) (71, 312) (71, 316) (72, 312) (74, 316) (75, 312) (79, 307) (81, 307) (84, 307) (85, 307) (88, 307) (92, 303) (100, 303) (105, 303) (124, 303) (128, 303) (136, 303) (141, 303) (141, 307) (144, 303) (145, 303) (145, 307)

For CaM-CaMKII: (15, 308) (19, 308) (68, 308) (71, 308) (72, 308) (74, 308) (84, 308) (92, 299) (92, 303) (105, 299) (109, 299) (124, 299) (127, 299) (128, 299) (136, 299) (141, 299) (141, 303) (144, 299) (144, 303) (145, 303)

ST then computationally assessed the roles of hydrophobic motifs of the CaMBTs (Table 2.1) during association by plotting the 3-D probability distributions as a function of both the native (Q_I) and intermolecular interactions (Z_{75}) between CaM and CaMBTs, and also the shape of CaM (S) in Fig. 2.7. The native intermolecular contacts are defined only for 1-14 and 1-5-10 motifs of CaMKI and 1-5-10 motif of CaMKII. Both native and

nonnative contact formation provide a complementary analysis to the mechanism of the CaM-CaMBT association in response to the shape changes in CaM. Upon association, CaMKI forms a greater number of native contacts corresponding to both 1-14 and 1-5-10 motifs with CaM (Figs. 2.7 A and 2.7 B, respectively) than CaMKII that has only the 1-5-10 motif (Fig. 2.7 C). These unique features are circled at $Q_I = 6$ between ES ($Z_{75} \sim 3$) and LS ($Z_{75} \sim 7$) in in Fig. 2.7. Also evident in these 3-D probability distributions is that as CaM forms contacts with a CaMBT, its shape changes from an extended ellipsoid to become more spherical, as discussed in the previous section with details in Table 2.3. In order to satisfy the final interaction to the 1-5-10-14 binding contacts on CaMKI, the central helix of CaM has to unwind to a lesser extent than for CaMKII (with only 1-5-10 binding motif), evident by the CaM/CaMKII complex ending up in a more extended (less spherical) conformation. These results are consistent with the strucural unwinding of the central helix of CaM revealed through NMR and crystallography (Meador et al., 1993; Gifford et al., 2011).

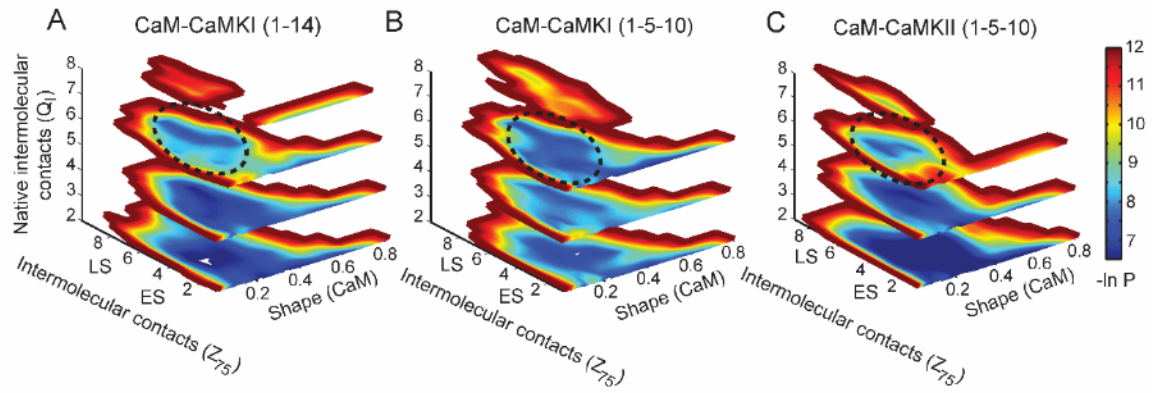

Figure 2.7 Interplay between the structural changes in CaM and the CaM-CaMBT interactions. 3-D distributions are plotted in log-scale -ln P, where P is the probability as a function of the number of selected native intermolecular contacts between CaM and the target (Q_I), the number of intermolecular contact formation between the 75th residue on CaM and the target (Z_{75}) and the shape parameter of CaM (S). Q_I captures the formation of intermolecular contacts between a CaM and a target that carries a hydrophobic motif such as 1-5-10 (CaMKI and CaMKII) and 1-14 (CaMKI). S is the shape parameter which describe a spheroid (S > 0, prolate; S < 0, oblate; S = 0, sphere). (A-C) are plotted with a distinct set of Q_I: (A) 1-14 motif of CaMKI; (B) 1-5-10 motif of CaMKI; (C) 1-5-10 motif of CaMKII. The early stage (ES) and late stage (LS) of association are indicated along the Z_{75} axis at $Z_{75} \sim 3$ and $Z_{75} \sim 7$, respectively in each plot. (A) and (B) were plotted from the same set of simulation data. In the colorbar red and blue indicate low and high probability, respectively.

Next, the targets CaMKI and CaMKII may associate with CaM either in a normal or inverse position (Barth et al., 1998). I analyzed the orientation of the target against the major axis of CaM and their secondary structures after forming functional complexes. I used a phase diagram of cos(θ) and ΔZ to analyze the orientation of bound targets relative to CaM in Figs. 2.6 C and 2.6 D for CaM-CaMKI and CaM-CaMKII, respectively.

θ is defined as the angle between the vector pointing from the center of mass of cCaM (from residue 75 to 148) to that of nCaM (from residue 1 to 74) and the vector pointing from the geometrical center of the last four to the first four amino acids of CaMBT (definition of θ is seen in Fig. 2.8).

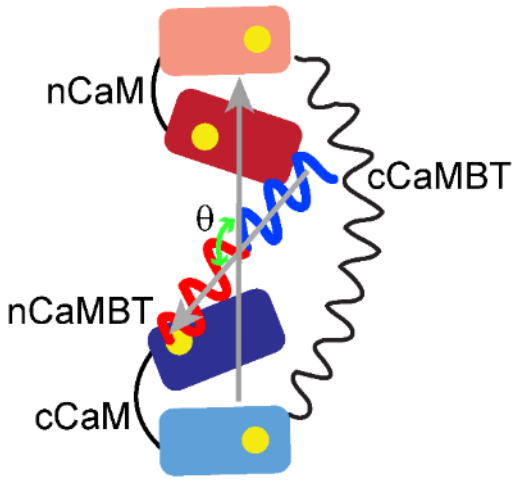

Figure 2.8 Schematic representation of the alignment between CaM and CaMBT by angle θ. The orientation of CaM is defined by a vector pointing from center of mass of cCaM (residue 75 to 148) to the center of mass of nCaM (residue 1 to 74). The orientation of CaMBT is defined by a vector pointing from the geometrical center of the last four amino acids of CaMBT to the geometrical center of the first four amino acids of CaMBT. cos(θ) is defined by the inner product of these two vectors. CaM is colored from red (nCaM) to blue (cCaM). CaMBT is colored from red (nCaMBT) to blue (cCaMBT). Ca^{2+} ions are represented by yellow spheres.

ΔZ is defined as the difference in the number of intermolecular contacts of a target at its two ends to the CaM domains given as, $\Delta Z = Z_{norm} - Z_{inv}$, where Z_{norm} is the sum of

the number of contacts between cCaM and nCaMBT and the number of contacts between nCaM and cCaMBT. nCaMBT for CaMKI is counted from residue 299 to 309 and for CaMKII it is from residue 292 to 302. cCaMBT for CaMKI is counted from residue 310 to 320 and for CaMKII it is from residue 303 to 312. Z_{inv} is the sum of the number of contacts between cCaM and cCaMBT and the number of contacts between nCaM and nCaMBT. Fig. 2.9 gives an illustrative representation of the complex when $\Delta Z > 0$ (normal binding) and $\Delta Z < 0$ (inverse binding).

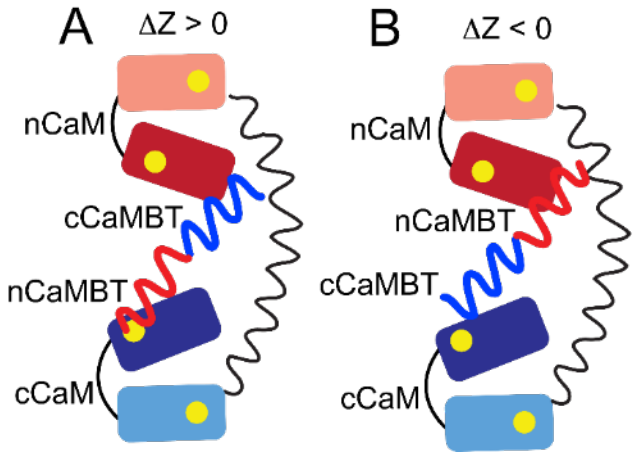

Figure 2.9 Illustration of the orientation of a binding target relative to CaM by ΔZ. (A) and (B) show the orientation of a binding target in a complex when $\Delta Z > 0$ (normal binding) or $\Delta Z < 0$ (inverse binding), respectively. Each CaM and CaMBT is colored from blue to red from the N-terminus to the C-terminus. The Ca^{2+} ions are represented in yellow spheres.

There are several dominant binding orientations including normal, inverse, and others shown in Table 2.4. Nearly all CaMKI peptides bind in either normal (49%) or inverse orientations (49%), with equal probability of each; for CaMKII, it can bind in normal (30%), inverse (40%) and other orientations (30%). Specifically, when I used CO to describe the content of secondary structures of these targets, the secondary structure of CaMKI in a bound complex is helical in either a normal or an inverse position (Fig. 2.6 C).

In contrast, CaMKII binding trajectories showed that nearly 40% of the bound complex adopt bent helical structures with a large CO value (Fig. 2.6 D and Table 2.3).

Table 2.4 Percentage of the targets in several binding orientations in the CaM-CaMKI and CaM-CaMKII complexes. The shaded area in orange (green) in Fig. 2.6 (C, D) contains the data from the trajectories that end in a normal (inverse) binding orientation. Data from other quadrants in white represent a conformation that is neither in an inverse nor a normal binding orientation.

Binding modes	Normal (%)	Inverse (%)	Other (%)
CaM-CaMKI	49	49	2
CaM-CaMKII	30	40	30

2.3.4 Principal component analysis on the CaM-CaMBT complexes

Such structural divergence between targets upon initial binding is expected to induce variation in the structural changes of CaM as the two proteins progress along dissimilar mutually induced fit trajectories towards bound complexes. To quantify these differences, I performed principal component analysis (PCA) for all of the successful trajectories to reveal the major conformational adjustment of CaM upon target binding between the early stage and the late stage transitions. Principal component 1 (PC1) reveals the most dominant interdomain motion of CaM dictated by interactions with target. Principal component 2 (PC2) indicates the second largest conformational change of CaM that involves the coupling between the intradomain and the interdomain motions. A few typical trajectories were projected on the potential of mean force (PMF) and these plots reveal a distinct pattern caused by the unique binding orientations between CaM-CaMKI and CaM-CaMKII (Fig. 2.10).

I observed that PC2 of CaM-CaMKI distinguished the trajectories of association leading to a normal or inverse binding orientation (Figs. 2.10 A and 2.10 C). When CaMKI starts in an inverse mode of binding (Fig. 2.10 A), there are strong and broad anti-correlations among the contacts across interdomain and intradomain CaM motions (Fig. 2.11 B). At the late stage of the trajectory when CaMKI flips to a normal orientation, the strained interaction among interdomain motion diminishes (Fig. 2.11C). Fig. 2.10C shows another typical trajectory of CaM with a CaMKI peptide that starts at a normal binding position and then ends at an inverse binding position. The intradomain and interdomain movement of CaM are strongly anticorrelated (Fig. 2.11E) even after they form a complex.

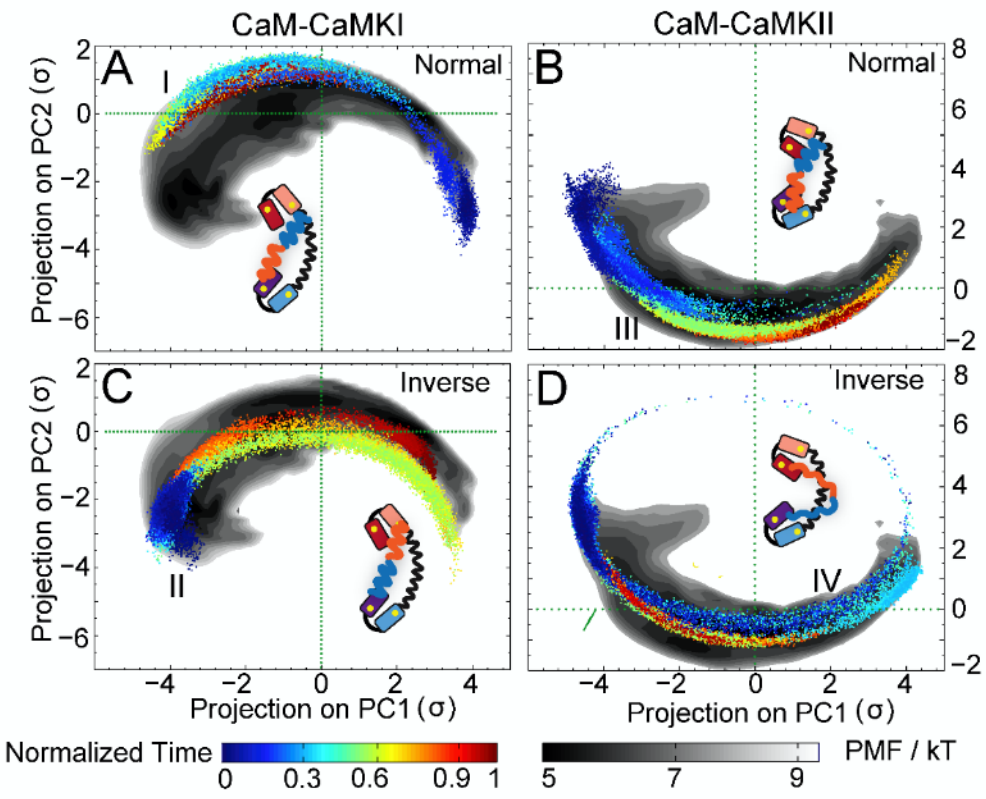

Figure 2.10 Principal component analysis (PCA) on the trajectories between the early stage and the late stage of the association. The first principal component (PC1) characterizes the inter-domain movement of calmodulin and the second principal component (PC2) denotes the combined intra- and inter-domain movements of calmodulin. The potential of mean force (PMF, in gray scale) is obtained by projecting structures from all the trajectories along directions PC1 and PC2. Representative trajectories are shown for CaM-CaMKI (A and C) and CaM-CaMKII (B and D) that lead to normal and inverse binding, respectively. Trajectories were colored by the normalized time and projected on the surface of PMF along PC1 and PC2. A schematic complex of CaM-CaMBT as well as two green guidelines are provided in each panel. σ is 3.8 Å.

A similar pattern from the PCA is observed for the binding of CaM to CaMKII. Fig. 2.10B shows a typical trajectory traveling through the lower region of the PMF that starts with an inverse binding position and ends in a normal binding position. From PC1 (Fig. 2.12B) and PC2 (Fig. 2.12C), I note that before the formation of a functional complex, the intradomain and interdomain anti-correlated motions in CaM are diminished significantly as the time evolves. For another trajectory that starts in a normal orientation and ends at an inverse binding (Fig. 2.10D), the patterns differ at PC2 where the anticorrelation among the interdomain contacts of CaM grows instead of shrinks (Fig. 2.12F lower triangle). MSC and I observed distinct types of antagonistic motions between the two CaM domains in response to the orientation of CaMKII upon binding. Overall, CaM-CaMKII has fewer trajectories that lead to a normal binding position in a functional complex than CaM-CaMKI. This may offer an explanation for the molecular mechanism of the slower measured association rates of CaMKII over CaMKI.

CaM-CaMKI. PCA was performed for all the successful trajectories to unveil the conformational adjustment of CaM between the early stage and the late stage transitions. The data for the potential of mean force (PMF, Fig. 2.10 in gray scale) was collected from ES2 and LS1 from all of the trajectories. PMF is projected on the largest two principal components (PC1 and PC2). A typical trajectory of CaM-CaMKI association ended with a normal or inverse binding orientation of CaMKI is superimposed on the PMF in Figs. 2.10A and 2.10C, respectively.

Fig. 2.10A shows a typical binding trajectory of CaM-CaMKI that starts from an inverse binding position and ends up at the normal binding one. The positive value of pc1

corresponds to the CaM motion toward an inverse binding of the CaMKI target, while the negative value of pc1 corresponds to a normal binding one. As shown by the correlation matrix elements in the two orange boxes in Fig. 2.11B, there are anti-correlated motion among the interdomain contacts at the off-diagonal of the correlation matrix map: (a) Helix G and Helix H from cCaM and Helix A from nCaM, and (b) the linker between Helix F and Helix G from cCaM (Linker FG) and the linker between Helix B and Helix C from nCaM (Linker BC). Particularly, (b) points to the regions where the CaM directly interacts with a target. At the late stage of the trajectory, PC1 sweeps back to pc1 = -4.1 σ, where the correlation between interdomain contacts diminishes (Fig. 2.11C) as the CaM motion is then mediated by a normal binding of the target.

In Fig. 2.10A, PC2 represents the motion of CaM-CaMKI where interdomain movement of CaM is coupled by the intradomain movement. PC2 distinguishes the trajectories of association leading to an inverse or normal binding of the CaMKI target. The trajectory starts at pc2 = -2 σ (i.e. 7.6 Å) when CaM is interacting with an inverse CaMKI target. There are strong and broad anti-correlations among the contacts across interdomain and intradomain CaM (Fig. 2.11E). As time evolves, PC2 reaches to 1 σ where CaM interacts with a normally positioned CaMKI target. The correlation between the contacts at intradomain and interdomain CaM diminish (Fig. 2.11F).

Fig. 2.10C shows another typical trajectory of CaM that starts at a normal binding position and ends at an inverse binding position of the CaMKI target. The range of motion along PC1 is similar to that of Fig. 2.10A. However, their difference rises at PC2, which the trajectory leading to inverse binding converges to -1.4cσ. The intradomain and

interdomain movement of CaM is strongly anticorrelated (Fig. 2.11E). It indicates a collectively non-vanishing motion of the entire CaM is to accommodate an inversely bound CaMKI target.

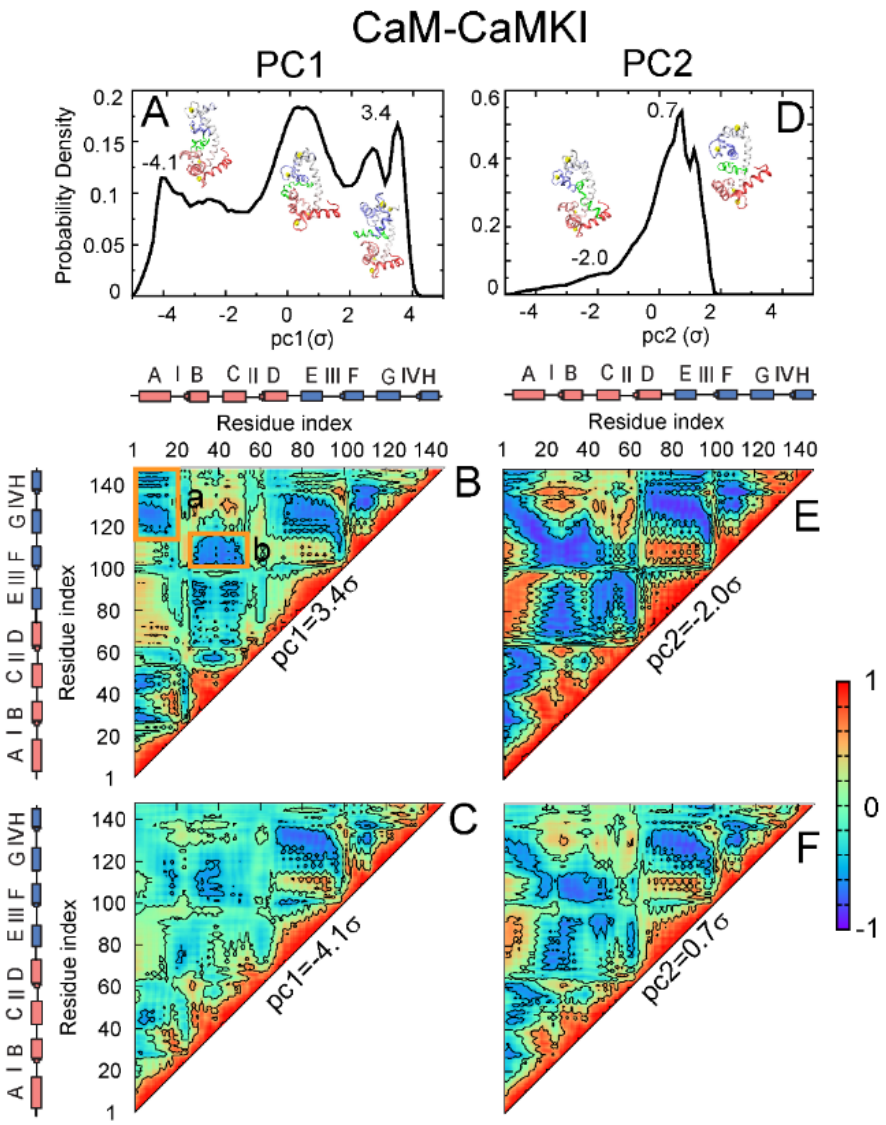

Figure 2.11 Cross-correlation matrices of CaM calculated from the complexes of CaM-CaMKI taken from the trajectories between the early and the late stages of association. (A) and (D) show the distributions of the principal components one (PC1) and two (PC2), respectively. (B) and (C) represent the correlation matrices of CaM at the projected eigenvalues pc1 = 3.4 σ and -4.1 σ, respectively. (E) and (F) represent the correlation matrices for CaM at the projected eigenvalues pc2 = -2.0 σ and 0.7 σ, respectively. σ is 3.8 Å. The orange rectangles (a and b) in (B) indicate the interdomain correlations. In the color bar for (B), (C), (E), and (F) red is +1 and indicates strong correlations, blue is -1 and stands for anticorrelations and green indicates no correlations. Representative structures of CaM-CaMKI complexes from simulations are shown for each peak of PC1 and PC2 in (A) and (D), respectively. In these structures CaM is colored from red (N-terminus) to blue (C-terminus) and CaMKI is shown in green. The schematic secondary structures of CaM are provided at the outside of the cross-correlation matrices. Red blocks are the helices from the N-domain. Blue blocks are the helices from the C-domain. A-H are the labels of the helices and I-IV are those of the Ca^{2+} binding loops.

CaM-CaMKII. Compared to CaM-CaMKI a similar pattern is observed for the binding of CaM and CaMKII in Fig. 2.10B and 2.10D for the normal and inverse binding, respectively. Fig. 2.10B shows a typical trajectory that travels through the lower strap of the PMF and ends in a normal binding position. For PC1, the trajectory starts around pc1 = -4.5 σ (Fig. 2.12A), where there exists a broad anti-correlation across the interdomain and intradomain contacts in CaM (Fig. 2.12B). PC1 then sweeps to pc1 = 4.1 σ as CaM interacts with a normal-positioned CaMKII target. The anticorrelation among the contacts at the interdomain and intradomain CaM rapidly diminished (Fig. 2.12C). For PC2, the anticorrelation among interdomain contacts as well as those among intradomain contacts weakens as time evolves (from Figs. 2.12E to 2.12F upper triangle). In contrast, for another trajectory that leads to an inverse binding of CaMKII (Fig. 2.12D), it follows a similar pattern in PC1, but not in PC2. For PC2, the anticorrelation among the interdomain contacts of CaM grows instead (Fig. 2.12F lower triangle). It indicates that there exists a collective motion between nCaM and cCaM in response to the movement of a CaMKII target in an inverse binding position.

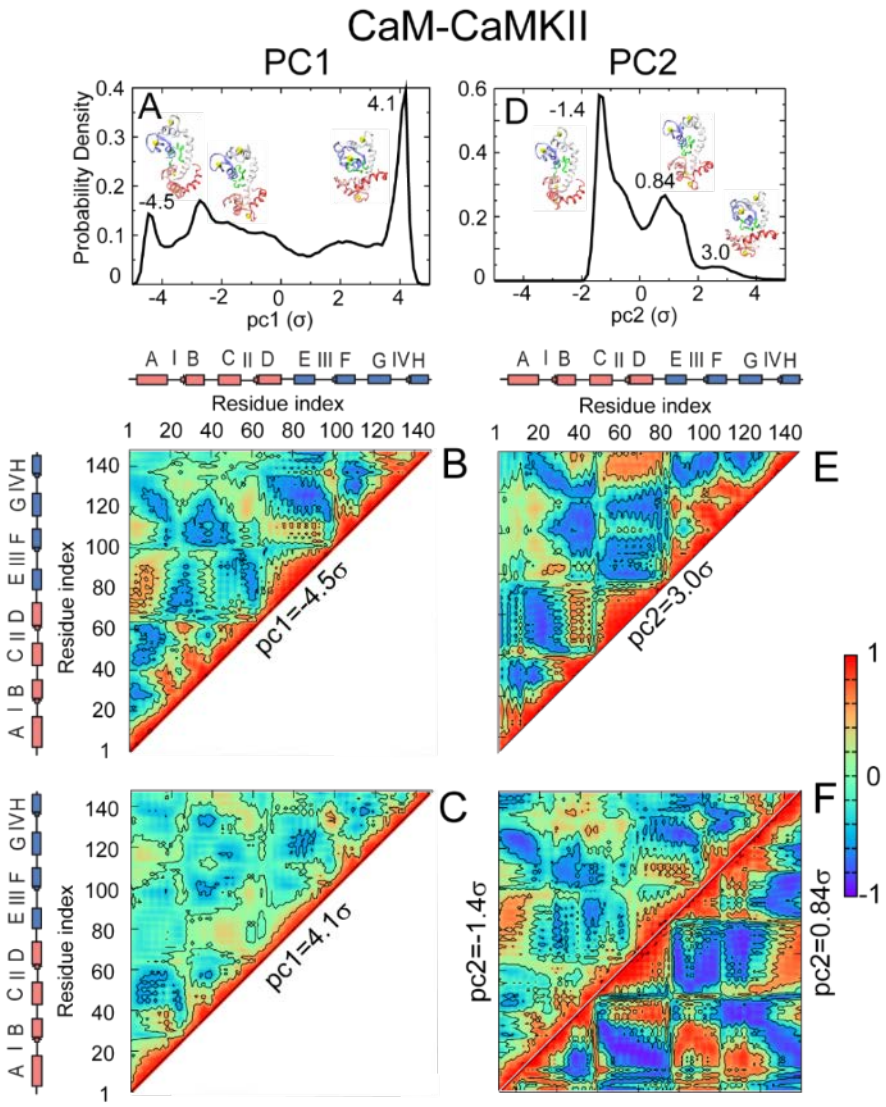

Figure 2.12 Cross-correlation matrices of CaM calculated from the complexes of CaM-CaMKII taken from the trajectories between the early and the late stages of association. (A) and (D) show the distributions of principal component one (PC1) and two (PC2), respectively. (B) and (C) represent the correlation matrices of CaM at the projected eigenvalues pc1 = -4.5 σ and 4.1 σ, respectively. (E) represents the correlation matrix of CaM at the projected values pc2 = 3.0 σ. (F) represents the correlation matrices of CaM at the projected eigenvalues pc2 = -1.4 σ and 0.84σ, in the upper and lower triangles, respectively. Representative structures of CaM-CaMKII complex from simulations are shown for each peak of PC1 and PC2 in (A) and (D), respectively. The notations used on this figure are the same as the one in Fig. 2.11.

2.3.5 Predicted structures of final complex are stable

It is important to note that the structures were determined from the kinetic trajectories. These structures are not necessarily at the global minimum or an energetically trapped state (Schreiber et al., 2009). They were not chosen based on their binding stability. However, I was interested in comparing these conformations with the experimentally determined ones. I reconstructed these coarse-grained complexes into all-atomistic protein models (Homouz et al., 2008; Samiotakis et al., 2010). I ran all-atomistic molecular dynamics simulations with explicit solvents on these structures using both CHARMM and AMBER forces fields. I averaged over the mean-square fluctuation (MSF) for all of these complexes and compared them to the experimental B-factor values. I also ran all-atomistic molecular dynamics simulations on the experimentally determined structures as an initial condition. I computed that MSF as a control. I analyzed these values for both CaMKI (PDB: 1MXE from x-ray as an alternative because the native complex 2L7L was determined from NMR) and CaMKII (PDB: 1CDM) in Fig. 2.13 and Fig. 2.14, respectively. I computed the correlation coefficients between the MSFs and the B-factors. The correlation coefficients (CC) were provided in the supplementary materials (Figs. 2.13 and 2.14). CC values are greater than or close to 0.3 for both the reconstructed protein models and the controls. They both represent the experimental B-values qualitatively well. The amplitude of the $\overline{MSF}$ values of the reconstructed proteins are overall greater than the controls, meaning that the structural distribution in the ensemble of reconstructed proteins is broader than those in the global minimum of a binding energy landscape. Given that both profiles on the MSF plots are qualitatively similar to the experimental B-values, I speculate that the ensemble of the

reconstructed proteins is not so far away from the global minimum of the binding energy landscape.

For CaM-CaMKI, the native complex structure (2L7L) is solved by NMR and does not provide B factor. The B factor values for C_α atoms were taken from an alternative X-ray structure (1MXE) of CaM-CaMKI. The sequence identity between the CaM sequence from 2L7L and 1MXE is 98%. The Root Mean Square Deviation (RMSD) of C_α atoms between the two is 1.52 Å. In 1MXE.pdb, two crystal structures are provided and the RMSD of C_α atoms between them are 0.35 Å. Only one set of the B-factors from the file were used in this analysis.

$\overline{MSF}$ was calculated for C_α atoms from atomistic simulations using the initial configurations reconstructed from 45 coarse-grained structures. A control simulation was carried out using the native complex structure (PDB ID: 2L7L) and MSF was calculated accordingly for C_α atoms in Fig. 2.13.

For CaM-CaMKII, B factor for CaM-CaMKII is available in native complex (PDB code: 1CDM). The B factor values for C_α atoms were collected. $\overline{MSF}$ was calculated for C_α atoms from atomistic simulations using the initial configurations reconstructed from 18 coarse-grained structures. A control simulation was carried out using the native complex structure (1CDM) and MSF was calculated accordingly for C_α atoms in Fig. 2.14.

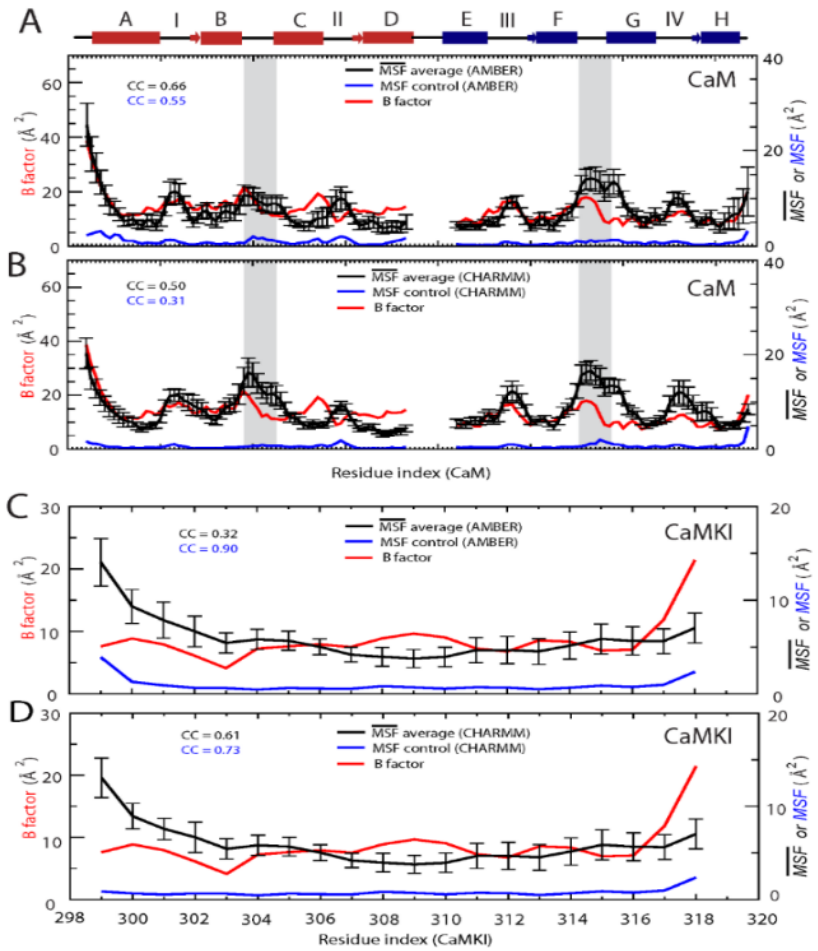

Figure 2.13 Mean square fluctuations from all-atomistic simulations and the B-factor experiments for CaM and CaMKI from the X-ray structure (PDB ID: 1MXE). (A) $\overline{MSF}$ (MSF) of CaM from molecular dynamics simulation using the AMBER99SB-ILDN force field (black from the reconstructed protein representations and blue from the control) and the B-factor (red curve). (B) $\overline{MSF}$ (MSF) of CaM from molecular dynamics simulations using the CHARMM27 force field and the B-factor (red curve). (C) $\overline{MSF}$ (MSF) of CaMKI from molecular dynamics simulation using the AMBER99SB-ILDN force field (black from the reconstructed protein representations and blue from the control) and the B-factor (red curve). (D) $\overline{MSF}$ (MSF) of CaMKI from molecular dynamics simulation using the CHARMM27 force field (black from the reconstructed protein representations and blue from the control) and B-factor (red curve). The correlation coefficient (CC) of average $\overline{MSF}$ and control MSF with respect to the B-factor from experiments are shown in each of the plot for both CaM and CaMKI. The $\overline{MSF}$ of CaM (or CaMKI) was calculated for the functional complexes of CaM-CaMKI from all the successful trajectories. The control MSF of CaM (or CaMKI) was calculated from the NMR structure of CaM-CaMKI complex (PDB ID: 2L7L). The $\overline{MSF}$ (MSF) of residue 74-83 from the central-linker of CaM are not shown in (A) and (B) and were not considered for the calculation of the CC. The shaded regions in (A) and (B) indicate the BC and FG helix-linker of CaM.

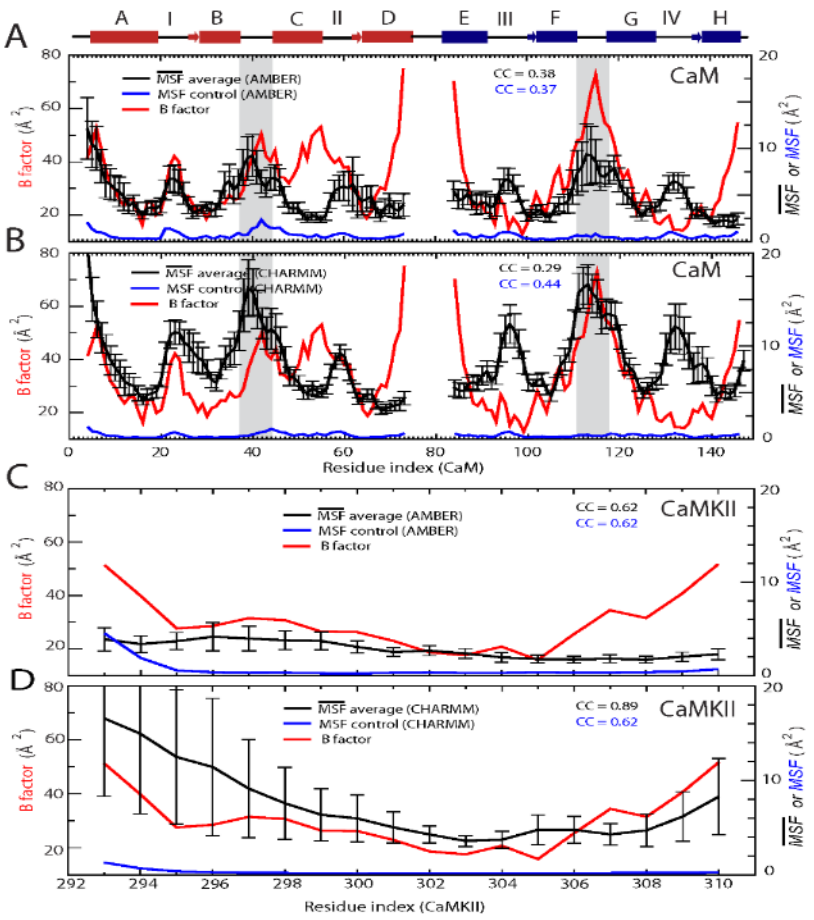

Figure 2.14 Mean square fluctuations from all-atomistic simulations and the B-factor experiments for CaM and CaMKII from the X-ray structure (PDB ID: 1CDM). (A) $\overline{MSF}(MSF)$ of CaM from molecular dynamics simulation using the AMBER99SB-ILDN force field (black from the reconstructed protein representations and blue from the control) and the B-factor (red curve). (B) $\overline{MSF}(MSF)$ of CaM from molecular dynamics simulations using the CHARMM27 force field and the B-factor (red curve). (C) $\overline{MSF}(MSF)$ of CaMKII from molecular dynamics simulation using the AMBER99SB-ILDN force field (black from the reconstructed protein representations and blue from the control) and the B-factor (red curve). (D) $\overline{MSF}(MSF)$ of CaMKII from molecular dynamics simulation using the CHARMM27 force field (black from the reconstructed protein representations and blue from the control) and B-factor (red curve). The correlation coefficient (CC) of average $\overline{MSF}$ and control MSF with respect to the B-factor from experiments is shown in each of the plot for both CaM and CaMKII. The $\overline{MSF}$ (or MSF) of CaM (or CaMKII) was calculated for the functional complexes of CaM-CaMKII from all the successful trajectories. The control MSF of CaM (or CaMKII) was calculated from the NMR structure of CaM-CaMKII complex (PDB ID: 1CDM). The $\overline{MSF}$ (or MSF) of residue 74-83 from the central-linker of CaM are not shown in (A) and (B) and were not considered for the calculation of the CC. The shaded regions in (A) and (B) indicate the BC and FG helix-linker of CaM.

2.3.6 Protein recognition through conformational and mutually induced fit

I observed that at the onset of an association between CaM and CaMBTs, the formation of a transient complex is initiated, and then a newly formed energy landscape emerges that accounts for conformational adjustments required to form the partially associated complex (Fig. 2.15A). The distinction between CaM-CaMKI and CaM-CaMKII occurs at the second stage (Fig. 4B), where differences in the structural fluctuation of the targets induce different responses in the structure of CaM. These fluctuations involve both interdomain and intradomain movement as the target binding pockets in CaM adjusts to the unique structurally changing targets, while CaM collapses from a normally more extended conformation. *The target recognition happens during the process of "mutually induced fit" because both CaM and the target must undergo significant structural changes to access new conformations that rarely existed prior to binding.*

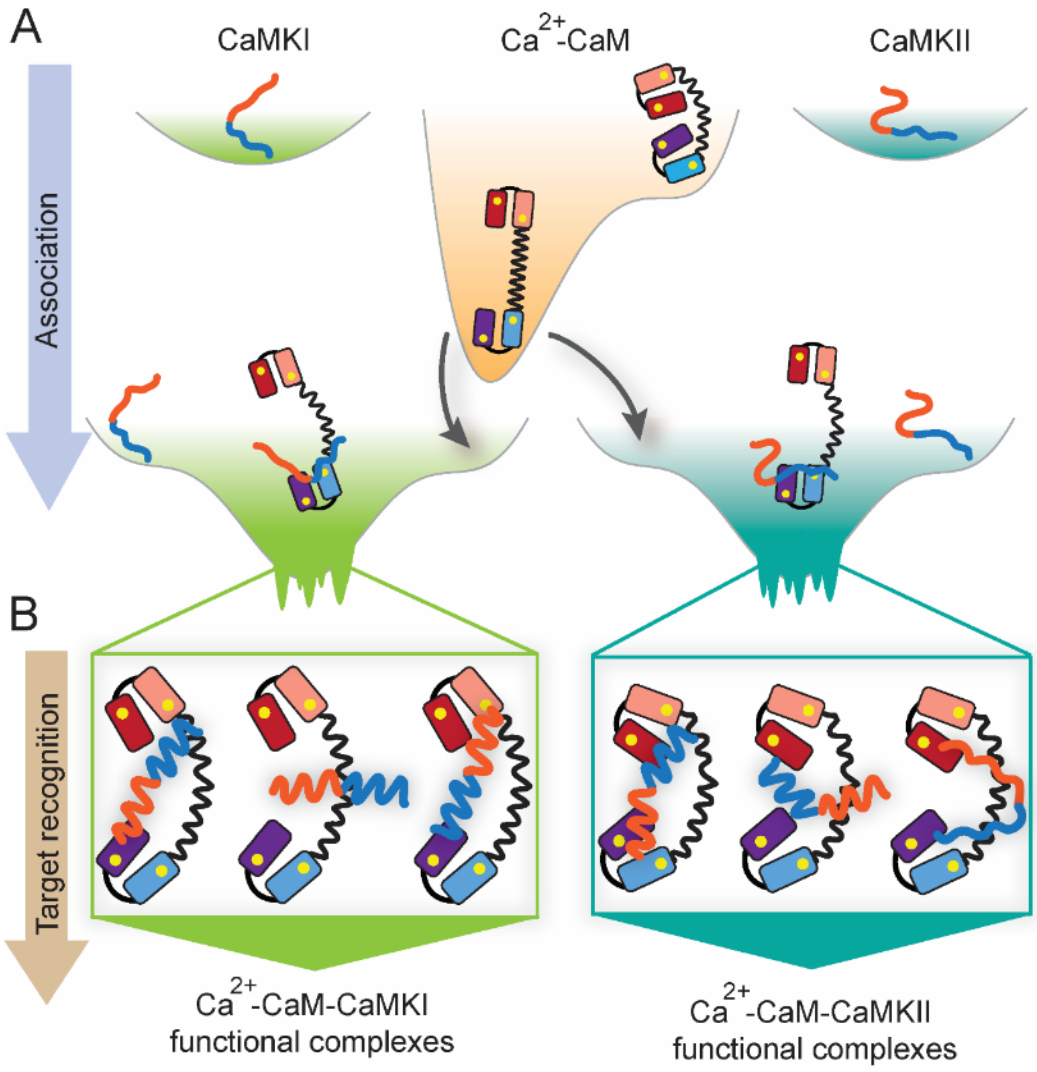

Figure 2.15 Schematic energy landscape of calmodulin-target recognition through mutually induced fit. (A) Prior to association, CaMKI and CaMKII targets each remain unstructured in the unbound state, while Ca^{2+}-CaM samples both the extended and the collapsed conformations that are different than the conformation in the bound complex. Once a target binds to one of the domain of Ca^{2+}-CaM in the extended conformation they form a transient complex. (B) Both the target (CaMKI or CaMKII) and CaM undergo significant conformational changes and form an encountered complex. Each target renders distinct orientations and structures that induce subtle differences in the correlated motion within Ca^{2+}-CaM. Recognition occurs during the "conformational and mutually-induced fit" process prior to the formation of the functional complexes of Ca^{2+}-CaM-CaMKI and Ca^{2+}-CaM-CaMKII. (The figure was created by Dr. ST)

To further understand the nature of intermolecular interactions (hydrophobic and electrostatics) in the mechanism of the conformational and mutually induced fit, I calculated the average number of contacts for each residue of CaMKI and CaMKII with CaM at the different stages of association. The results clearly indicate that most of the charged residues of CaMBTs (labeled with '*' in Fig. 2.16) have significant interactions with CaM at ES1. Nonetheless, at the later stages of association, especially at ES2 and LS1, I found that the increase of contacts from hydrophobic residues of the CaMBTs is relatively higher than those from the charged residues as the CaMBTs approach the binding pocket of CaM to form a functional complex at LS2. From this study, I can identify key rate-limiting steps that dictate the formation of CaM and CaM target binding. CaMKI has two hydrophobic motifs (1-5-10 and 1-14) while CaMKII has only one (1-5-10), and my findings have revealed how such seemingly subtle differences may induce a differential response as CaM undergoes structural adjustments to form a canonical functional complex.

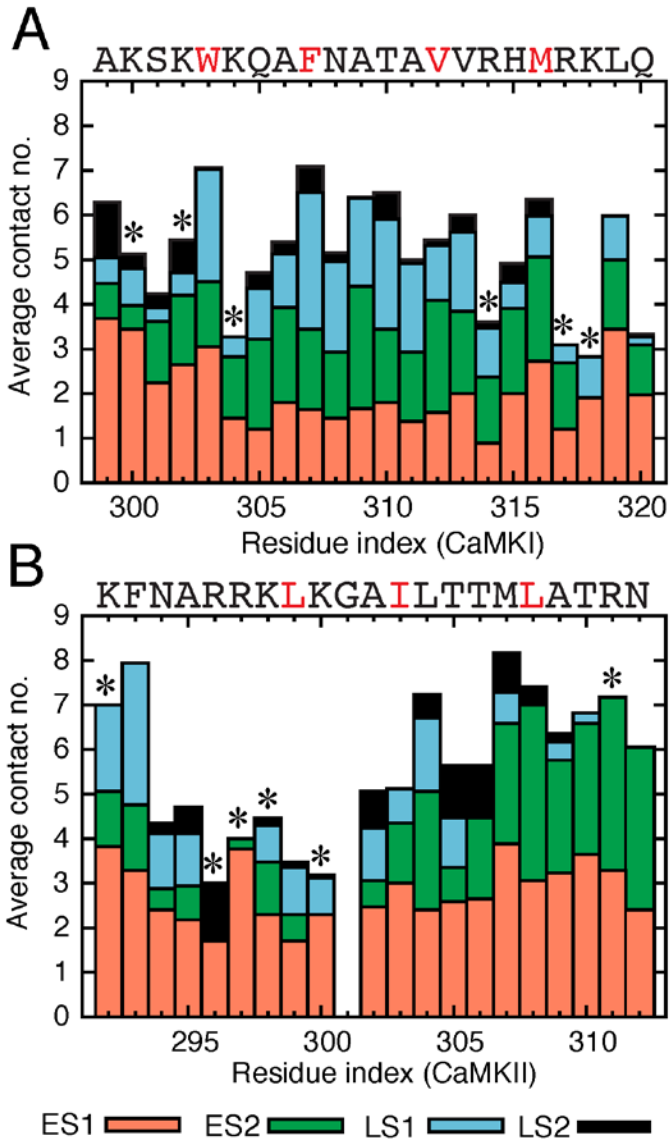

Figure 2.16 Average number of contacts of CaMBTs during their association with CaM. (A) Average number of contacts of CaMKI with CaM are shown at the different stages of association [ES1 (red), ES2 (green), LS1 (blue) and LS2 (black)]. (B) Average number of contacts of CaMKII with CaM are shown at the different stages of association. For CaMKI and CaMKII the charged residues are indicated by the (*) symbol in (A) and (B), respectively. Definition of ES1, ES2, LS1, and LS2 can be found in the main text. (The figure was created by Dr. ST).

2.3.7 Binding frustration in the course of conformational and mutually induced fit

The binding frustration is caused by the sequence of conformational changes required in both binding partners before the formation of a stable complex. The extent of conformational changes in CaM is induced by the specific and unique amino acid sequence and composition of the CaMBT and was termed "conformational and mutually induced fit" (Wang et al., 2013). There are two main stages, the early stage and the late stage, of the association when trajectories were diagnosed with Z_{75} as a function of normalized time (the parameters $\bar{Z}$ and Z_{75} are defined in the *Materials and Methods* section). The binding frustration along the association process were shown by computing Z_{75} at several $\bar{Z}$ for CaM-CaMKI and for CaM-CaMKII and ST showed that most of the binding frustration is resolved before the late stage when the CaM-CaMBT forms a functional complex (Tripathi et al., 2015).

MSC, NW, ST, and I speculated that the association rate for CaMKII with CaM can increase by a simple mutation that lessens its binding frustration. I focused on the contact formation between CaM and CaMKII, as well as CaM and CaMKI, at the onset of association because it dictates the subsequent frustration in the binding route. I computed the intermolecular contact probability of the residues from the CaMBTs before and after backtracking at low Z. I evaluated their differences (ΔP_i) on each residue *i* seen in Fig. 2.18. A residue from CaMBT with a negative ΔP must first form and then break contacts with the residues from CaM. For CaMKI, S301 is the most frustrated residue. For CaMKII, R297 is the most frustrated residue. They share the same analogous position (labeled with an asterisk) counting towards the N-terminus from the first hydrophobic residues that form

the anchors of the binding motifs (colored in green). I speculate that by mutating the arginine 297 in CaMKII to a serine (like in CaMKI) or alanine, its binding frustration might be lessened and hence its rate of association increases. This interesting hypothesis was later proven in my computational studies of binding between CaM and CaMKII mutant R297A (data are not shown).

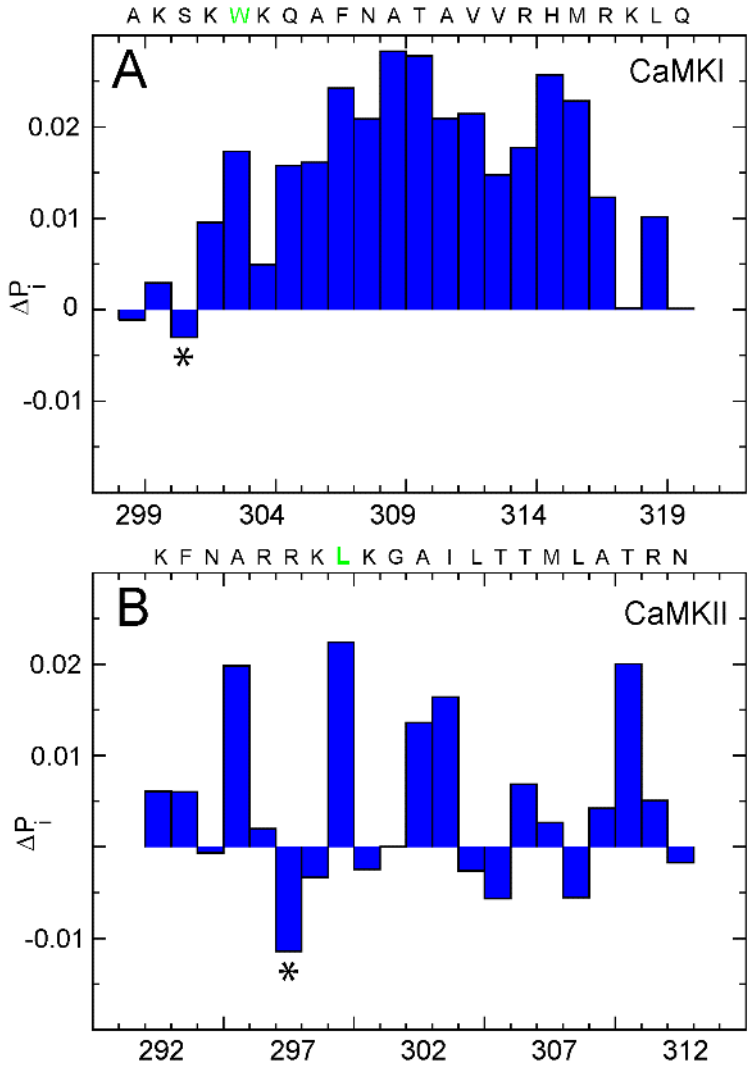

Figure 2.18 Difference in the intermolecular contact probability of the residues from the CaMBTs along the binding route at the early stage of association. (A) Difference in the contact probability (ΔP_i) for each residue of CaMKI, at $\bar{Z} \sim 0.3$ and $\bar{Z} \sim 0.6$. (B) ΔP_i for each residue of CaMKII, at $\bar{Z} \sim 0.2$ and $\bar{Z} \sim 0.3$. The sequences of CaMBTs are seen on top of each plot. The first hydrophobic residues from the binding motif of the CaMBTs are seen in green (W303 from CaMKI and L299 from CaMKII).

Chapter 3

Computational modelling of intrinsically disordered proteins

3.1 Introduction

Intrinsically disordered proteins (IDPs) / regions (IDRs) have aroused broad and lasting interest since the discovery in late 20th century. Proteins containing disordered regions of over 30 residues in length occupy more than a third of the eukaryotic proteins (Ward et al., 2004). More importantly, they are often involved in development of diseases, especially cancer, where more than 60% of the proteins in the pathways contain IDRs which are over 30 amino-acid long. It is essential to determine the unbound state ensemble of the IDPs, which determines the recognition mechanism of the IDPs and the binding partners.

Currently, there are several computational models on IDP for determining the unbound conformational ensemble, as well as interaction between the binding targets. Usually no native structure is present for IDPs, therefore structure-based models, including all-atomistic and coarse-grained, are limited to those IDPs with a determined bound structure in complex with their binding targets. All-atomistic models have been very popular for reproducing the non-specific structure ensemble. However, the parameters used by various force-fields are based on ordered proteins and fall short to describe the disordered states. Moreover, all-atomistic simulations demand unreachable computational costs for large complex proteins. In the present study, I report how a sequence-based dihedral angle potential performs on a small globular two-state protein chymotrypsin

inhibitor 2 (CI2) and a mostly intrinsically disordered protein neurogranin (Ng). The sequence-based dihedral angle potential was deduced from determined structures by X-ray crystallography by Karanicolas and Brooks (Karanicolas and Brooks, 2002), which we will mention it as KB dihedral potential in the context. The KB potential was able to produce the folding transition states of protein L and protein G with a coarse-grained one-bead (C_α) protein model (Karanicolas and Brooks, 2002). The KB dihedral angle is popular in the simulation of protein association involving intrinsically disordered proteins together with other different energy terms (Ganguly and Chen, 2011; Ganguly et al., 2012; Kim et al., 2014). In this study, I implemented the KB dihedral potential in the sidechain-C_α model which has been previously used to study protein-protein association (Wang et al., 2013). From the coarse-grained molecular simulations, the KB dihedral potential presents a better description of the transition state of CI2 and the unbound ensemble of Ng.

3.2 Materials and Methods

The virtual dihedral angle between four adjacent α-carbons depends on the backbone dihedral angles of the two middle residues. Using Ramachandran plots obtained from a survey of the protein structures determined by X-ray crystallography from protein data bank (PDB), a probability distribution for the virtual dihedral angle was obtained for each of the 400 possible ordered pairs of amino acid residues. The probability distribution can be related to the energy function of the virtual dihedral angle. The virtual dihedral potential presents two minima located at approximately +45° and −135°, corresponding to local α-helical and β-strand geometry, respectively. These "statistical" potentials are modeled as a cosine series consisting of four terms $E_{dihedral}^{ij} = \varepsilon_{KB} \sum_{n=1}^{4} K_{ij,n}\left[1 + cos\left(n\varphi_{ij} - \delta_{ij,n}\right)\right]$,

where φ_i is the dihedral angle i, δ_{ij} stands for four angles and $K_{D,ij}$ is the force constant, both of which are obtained from PDB. Although the KB dihedral angle potential was obtained from ordered proteins, it has been shown to give considerable close description of the protein, especially for the nonnative state, compared to a transferrable topology based model (Cavalli et al., 2005).

3.2.1 Modeling of CI2

CI2 is a small globular protein that is extensively studied by both experiments and molecular simulations. The folding characteristics of the protein, as well as the transition state structures are well documented. Therefore, I used CI2 as a test protein for the KB dihedral angle potential. The crystal structure of CI2 is shown in Fig 3.1.

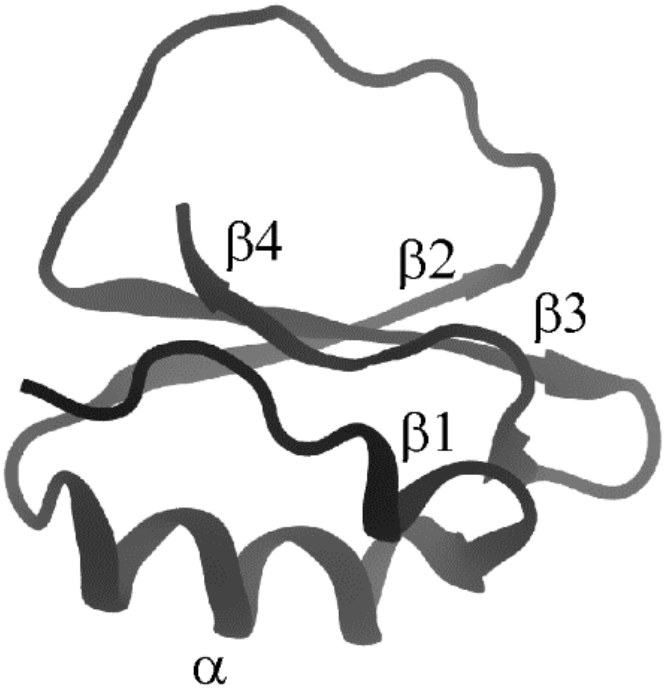

Figure 3.1 Cartoon representation of the chymotrypsin inhibitor 2 (CI2). The structure is from crystallography (PDBID: 1YPA). The main secondary structures are annotated as α, β1-β4. The figure is prepared using VMD (Humphrey et al., 1996).

I adopted a side-chain-C_α coarse-grained model developed by Cheung et al (Cheung et al., 2003) for the proteins. In this model, each amino acid (except glycine, which is represented by a C_α bead) is represented by two beads: the C_α bead is located at the C_α position to represent the backbone atoms; the side-chain bead is located at the center of mass of the side-chain atoms. The side-chain-C_α model enables us to capture large structural fluctuations of the proteins at a long timescale that warrants a wide search of the phase space.

The total potential energy E of CI2 is given by,

$$E = E_{bond} + E_{angle} + E_{dihedral} + E_{chirality} + E_{HB} + E_{LJ} + E_{rep} \qquad \text{Eqn. (3.1)}$$

The bonded interactions include the bond energy (E_{bond}), the bond-angle energy (E_{angle}), the dihedral-angle energy ($E_{dihedral}$), and the L-isomer restraint of chirality ($E_{chirality}$). The formula for E_{bond}, E_{angle}, and $E_{chirality}$ can be found the Appendix 6.1. The nonbonded hydrogen bonding interactions E_{HB} and van der Waals interactions E_{LJ} take the same forms as in Appendix 6.1, but only the contacts found in the native structure (PDB ID: 1YPA) were included. For the rest of the nonbonded pairs, a repulsive interaction is E_{rep} assigned. The depth of the interaction potential $\varepsilon_{ij} = \varepsilon = 0.6$ kcal/mol.

To compare the structure-based dihedral angle potential, I employed a backbone dihedral angle potential that is sequence-dependent and topology-independent, KB dihedral angle potential (please find details of the two types of dihedral angle potentials in the Appendix 6.2). The KB dihedral angle potential was initially designed for C_α model, which uses different force constants for the energy terms for the bonded interactions and a

different form of the LJ potential (Karanicolas and Brooks, 2002). Therefore, the strength for the dihedral angle potential may not match our sidechain-C_α model. The barrier of the KB potential was adjusted to generate the residual structure in the unbound state.

3.2.2 Modeling of Ng

Ng is a 78 amino-acid protein that is partly ordered in solution. The fragment G25-A42 (see the *sample preparation* section) has a fraction of 22% and 28% residual structure calculated from the C_α and the H_α chemical shifts in the nuclear magnetic resonance (NMR) experiment (Ran et al., 2003), respectively. The rest part of the peptide remains unstructured. In order to compare with results from several experiments (Ran et al., 2003; Hoffman et al., 2014), the full sequence Ng were used in the simulations. The only available experimental measurement for residual structure of Ng is for the full-length protein Ng from mouse (Ran et al., 2003), therefore, the sequence is used in the modeling and parameterization of Ng, as seen below (the part having residual structure involving the IQ domain is in magenta),

```
1                 10                  20                  30                  40                  50
MDCCTESACSKPDDDILDIPLDDPGANAAAAKIQASFRGHMARKKIKSGE
                  60                  70
CGRKGPGPGGPGGAGGARGGAGGGPSGD 78
```

This intact Ng is used in replica exchange molecular dynamics (REMD) simulations for calculation of average helicity of the residual structure and the model free order parameter (S^2).

The Hamiltonian of Ng mainly takes the form in the Appendix 6.1. Here I provide the description of each term. The bonded interactions include the bond energy (E_{bond}), the

bond-angle energy (E_{angle}), the dihedral-angle energy ($E_{dihedral}$), and the L-isomer restraint of chirality ($E_{chirality}$). The formula for E_{bond}, E_{angle}, and $E_{chirality}$ can be found in previous work (Wang et al., 2013). The equilibrium bond length, bond angle and side-chain chirality parameters of the missing segment were obtained from the structures predicted by the Sparks-X protein structure prediction server (Yang et al., 2011), since these terms are mainly constrained by chemical rules and do not vary by the conformations.

The first dihedral angle potential is a sequence-based model. The statistical potential is modeled as a 4-term cosine series,

$$E_{dihedral}^{ijkl} = \varepsilon_{KB} \sum_{n=1}^{4} K_{jk,n}\left[1 + cos\left(n\varphi_{ijkl} - \delta_{jk,n}\right)\right] \qquad \text{Eqn. (3.2)}$$

where φ_{ijkl} is the dihedral angle formed by four consecutive C_α beads *i, j, k, l* with beads *j* and *k* in the middle, $K_{jk,n}$ and $\delta_{jk,n}$ are statistically determined constants. ε_{KB} is a factor to adjust the barrier of the KB dihedral angle potential in relative to other interactions in the current model.

The second dihedral angle potential is a structure-based model, or SB model, where the structure of a specific segment of Ng was used as the reference (please see Table 3.1 for the residues in this segment). It is composed of two-term cosine-series,

$$E_{dihedral}^{ijkl} = \varepsilon_{SB} \sum_{n=1,3} k_{\varphi}^{n} \left[1 - \cos\left(n \times \left(\varphi_{ijkl} - \varphi_{ijkl}^{0}\right)\right)\right] \qquad \text{Eqn. (3.3)}$$

where *i, j, k,* and *l* are four consecutive C_α beads. φ_{ijkl} is the dihedral angle formed by those four beads. The equilibration position of the dihedral angle φ_{ijkl}^{0} for this specific segment was taken from the crystal structure (PDB code: 4E50 for Ng). $k_{\varphi}^{1} = 2k_{\varphi}^{3} = 2\,\varepsilon$.

ε_{SB} was used to adjust the barrier of the dihedral angle potential. For the remainder of the unstructured segment $k_{\varphi}^{1,3} = 0$.

Table 3.1 The composition of the three dihedral potentials for modeling Ng.

fragment / models	from G25 to A42	Others
SB	structure-based	0
Hybrid	structure-based	sequence-based
KB	sequence-based	sequence-based

The third dihedral potential is the hybrid of the two models, or Hybrid model, as seen in Table 3.1. I replaced the dihedral angle potential of the unstructured segment from the SB model with a KB model.

Determination of partial charges. I added partial charges to the intact protein Ng using a multi-scale method developed by Cheung group (Wang et al., 2011). First I ran Replica Exchange Molecular Dynamics (REMD) simulations without electro-static interactions (please find the details of REMD in the Appendices *6.4*) and obtained the free energy surface F(Δ, χ) as a function of asphericity Δ and overlap function χ. Δ measures the shape of the protein: it is like a rod or sphere when Δ = 1 or 0, respectively. χ shows how different the overall structure is from the reference structure. I selected about 400-600 frames from the REMD simulations of Ng through importance sampling. The partial charges were averaged over all the sampled configurations computed from the semi-empirical quantum chemistry program MOPAC (Stewart, 2007). I repeated the same process for all the three models seen in Table 3.1 and the three sets of charges I obtained were highly similar (correlation coefficient ~ 1.00). I therefore used only one set of charges

for all three models. The partial charges of Ng protein are provided in Table A4 in the Appendices.

3.2.3 Replica Exchange Molecular Dynamics simulations

In order to study dynamics of the proteins, I applied REMD to enhance the sampling with structure-based or sequence-based dihedral angle potential models. Especially for the Ng protein, REMD is required to sample the rugged energy landscape because of its intrinsic disorder.

For each of the models for CI2, 12 replicas were distributed at temperatures T = 0.40, 0.42, 0.44, 0.48, 0.49, 0.50, 0.51, 0.52, 0.54, 0.56, 0.58, 0.60 ε/k_B.

For each of the models for Ng, 24 replicas were distributed at temperatures T = 0.80, 0.82, 0.83, 0.85, 0.87, 0.88, 0.90, 0.93, 0.97, 1.00, 1.03, 1.07, 1.10, 1.13, 1.17, 1.20, 1.23, 1.27, 1.30, 1.33, 1.40, 1.47, 1.57, 1.67 ε/k_B to produce ample exchanges between the replicas (ε is the reduced energy unit = 0.6 kcal/mol, and k_B is the Boltzmann constant). The average acceptance ratio for exchange among the neighboring replicas ranges from 0.57 to 0.77. An exchange is attempted between neighboring replicas for every 100,000 time steps. A total of 6 sets of different initial configurations were used in each REMD simulation to explore more conformation space, making up 1200 exchange attempts for each model.

3.3 Results and Discussions

3.3.1 The statistical dihedral potential depicts transition states of CI2

The protein CI2 presents a two-state transition where there is only a single free

energy barrier for folding ($\Delta G = G_T - G_U$, where T stands for transition state and U stands for unfolded state). With the implementation of the KB dihedral potential into the coarse-grained model, I computed the heat capacity and ΔG from the REMD simulations. The barrier of the KB dihedral angle potential is rescaled by a factor ε_{KB} from 0.1 to 2.0 to explore the dependence of ΔG on ε_{KB}. In Fig. 3.2, an example of the folding thermodynamics is provided at ε_{KB} = 0.5. Indeed, only a single transition between folded and unfolded states is observed. At the folding temperature (T_F), multiple folding and unfolding events were observed indicating sufficient sampling. ΔG is about 2.3 kcal/mol.

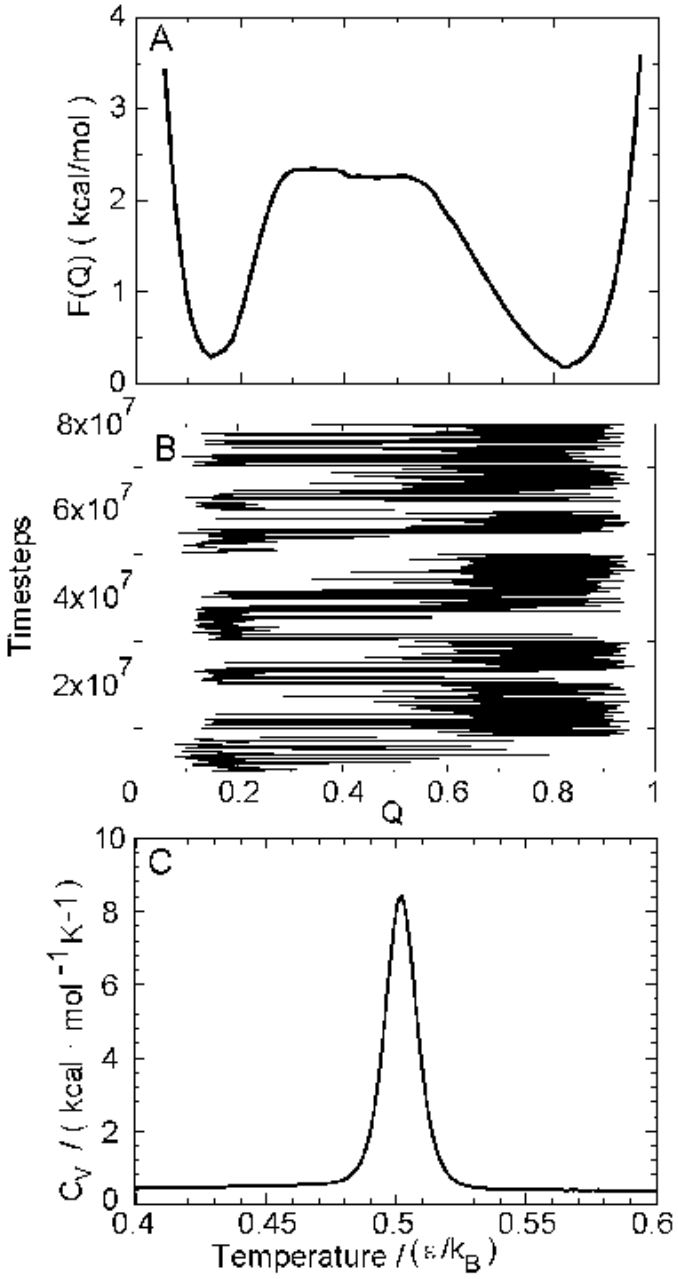

Figure 3.2 Folding thermodynamics of CI2 using KB dihedral angle potential. (A) Free energy F(Q) as a function of the reaction coordinate, fraction of native contacts Q at the folding temperature $T_f = 0.5\ \varepsilon/k_B$. (B) Q evolves in a typical sample simulation near the folding temperature. The two-state behaviour of CI2 is evident. (C) Heat capacity (C_V) as a function of temperature. C_V peaks at the folding temperature.

ΔG was calculated with different ε_{KB} by identifying the T_f from the C_V plots (Fig. 3.3) and F(Q). ΔG is about 2.0 kcal/mol with non-scaled KB dihedral angle potential (ε_{KB} = 1.0); 2.3 kcal/mol with ε_{KB} = 0.1 or 0.5; the barrier decreases to 1.8 kcal/mol when ε_{KB} = 2.0. This indicates a nonlinear dependence of ΔG on ε_{KB}. I speculate that when the dihedral angle is weak, it lowers the energy barrier by increasing the entropy; when the dihedral angle potential is too strong, because the potential is topology-independent, the non-native interactions are the main factor that increases the probability of misfolding or unfolding.

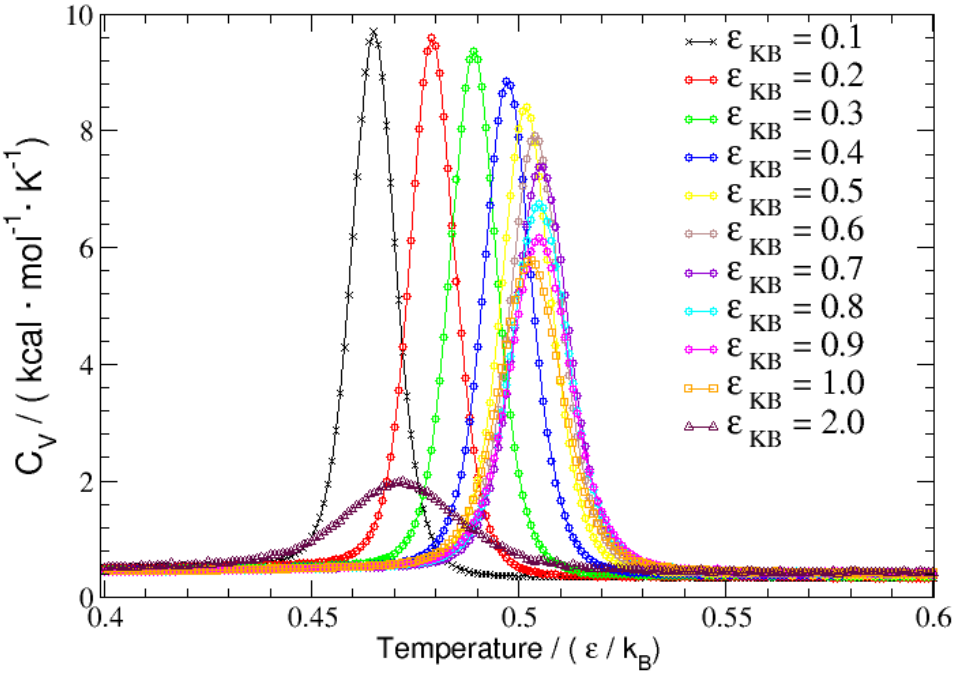

Figure 3.3 Heat capacity of CI2 with different strengths of KB dihedral angle potential. **ε_{KB}** is defined as the scaling factor of the dihedral angle potential energy. **ε_{KB}** = 1.0 (non-scaled) refers to the dihedral angle potential obtained from the MMTSB server. All of the curves are from REMD simulations and prepared using WHAM.

To note that with all of the dihedral angle potentials, including the topology based compare group (data not shown), the calculated ΔG was at most half of the ΔG measured from experiments. ΔG was measured to be 7.0 ~ 8.5 kcal/mol in either thermal denaturation

or chemical denaturation experiments at temperature 25 ℃ for wild type CI2 (Jackson and Fersht, 1991; Harpaz et al., 1994). This could be the limitation of using coarse-grained simulations to estimate absolute free energy barrier.

The folding temperature (T_f, located at the C_V peak) does not monotonically depend on the barrier of the KB dihedral potential (ε_{KB}). Instead, as seen in Fig. 3.4, the relation between T_f and ε_{KB} described by an inverted U-shaped curve. I speculate that on the one hand, with low barrier of dihedral energies, the protein is quite flexible and is easier to fold and unfold at lower temperature; on the other hand, with very high barrier of the nonnative dihedral angle potential, the contribution of variance of dihedral energy to the variance of total energy is very small and hence leads to low folding temperature. Indeed, T_f for ε_{KB} = 2.0 is 0.472 ε / k_B, close to 0.465 ε / k_B when ε_{KB} = 0.1. For C_V, it's not surprising that as the protein becomes more rigid (or ε_{KB} goes up), the C_V curve becomes broader and the peak value decays, indicating loss of folding cooperativity (Fig. 3.4). The folding temperature reaches a plateau when ε_{KB} falls into the range [0.6, 0.9].

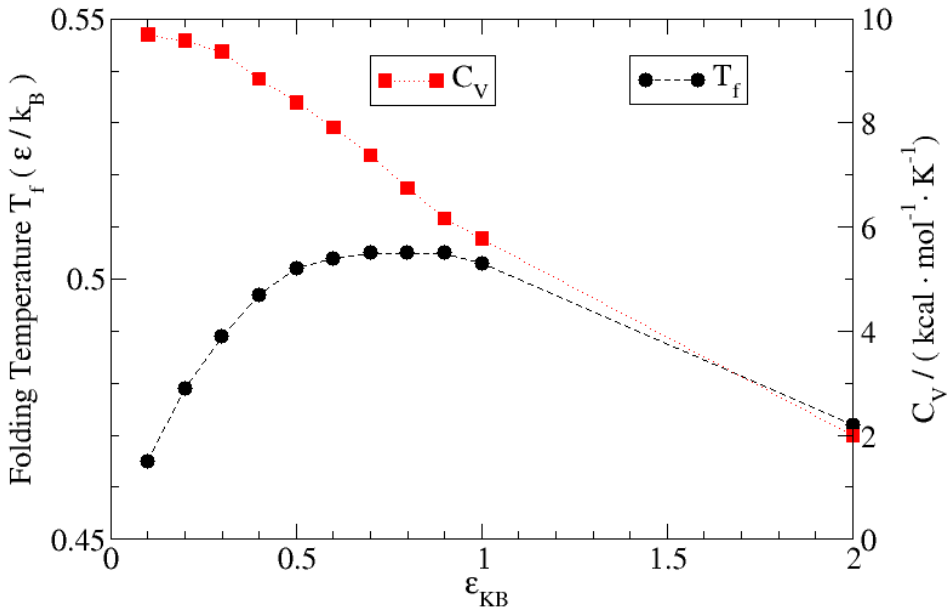

Figure 3.4 Folding temperature and peak value of the heat capacity as a function of ε_{KB}. T_f is in black solid dots and C_V is in red solid squares.

Free energy as a function of fraction of native contacts (Q) was plotted at the folding temperatures. As shown in Fig. 3.5, the barrier height between folded and unfolded states ΔG generally decreases with increase of the KB potential. When the KB potential is weak ($\varepsilon_{KB} <= 0.6$), a broad transition state covering Q = 0.3 to Q = 0.6 exists in common, which matches with experiment fact that CI2 is a two-state protein; two transition states are observed when $\varepsilon_{KB} > 0.6$. In the extreme case as $\varepsilon_{KB} = 2.0$, the second transition is minor and the folded state becomes a broad region including more nonnative contents (Q ~ [0.6, 0.9]).

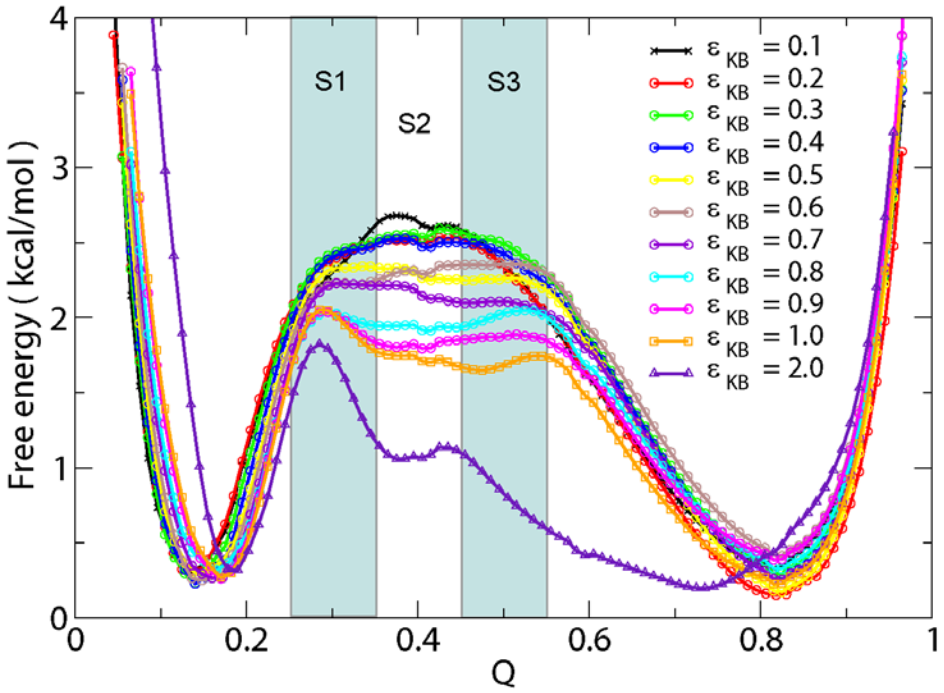

Figure 3.5 Free energy profile of CI2 using order parameter Q. The free energy were computed at the corresponding folding temperature of each ε_{KB}. All of the curves are from REMD simulations and prepared by WHAM. Three regions were marked for clustering analyses as S1, S2 and S3 between folded and unfolded states. S1 is [0.25, 0.35], S2 is [0.35, 0.45] and S3 is [0.45, 0.55].

In order to find out the strength of KB potential that best simulates the property of transition state of CI2, I employed φ value analyses. $\phi = \frac{\Delta G^{T}_{mut-wt} - \Delta G^{U}_{mut-wt}}{\Delta G^{F}_{mut-wt} - \Delta G^{U}_{mut-wt}}$ provides a quantitative indication about which part of the protein forms native structure and thus plays a more important role in the transition state. I calculated φ-value from the coarse-grained molecular simulations using the free energy perturbation method (F.E.P) as introduced in this study (Nymeyer et al., 2000). In this method, the mutation of a specific residue in the experiment is mimicked by weakening the interaction between this residue and other residues to a certain percentage (1-λ) and thus the free energy change can be calculated. The bond φ-values are barely dependent on λ (data are not shown), therefore, later on

residue φ-values would be calculated using the data only from λ= 0.1. In Table 3.2, I provide the correlation between φ-values computed from simulation and those from the experiments. ε_{KB} = 0.2 or 0.5 with correlation of φ-values computed in either the whole broad transition region S [0.3, 0.6], or separated transition regions S1 [0.25, 0.35] or S3 [0.45, 0.55] show better correlation with experimental values than others. For these two ε_{KB}, the correlation diagrams as well as linear fitting are provided in Fig. 3.6.

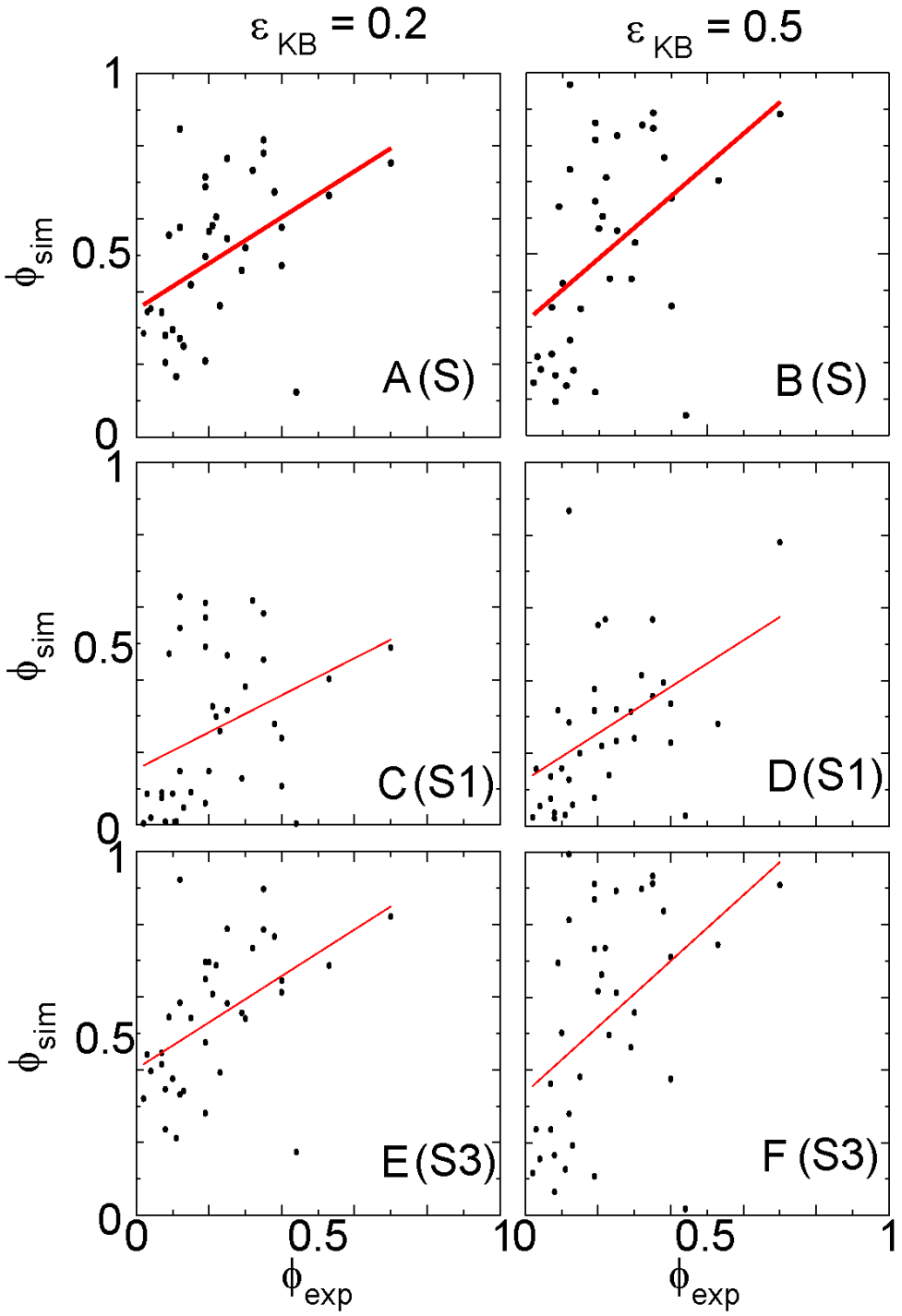

Figure 3.6 Comparison between residue-φ values from experiment and those calculated at different transition regions from simulation with ε_{KB} = 0.2 and 0.5. Correlation between those two data sets are provided by linear fitting which is in red solid line. Those points (φ > 1 or φ < 0) are neglected since they are not well interpreted. S1, S3 and S refers to regions (0.25, 0.35), (0.45, 0.55) and (0.25, 0.55), respectively.

Table 3.2 Correlation coefficient from linear fitting between computed and experimental φ-values. S1, S3 and S refers to regions (0.25, 0.35), (0.45, 0.55) and (0.25, 0.55), respectively.

ε_{KB}	S1	S3	S = S1+S2+S3
0.1	0.20	0.46	0.44
0.2	0.35	0.49	0.47
0.3	0.32	0.46	0.45
0.4	0.37	0.44	0.45
0.5	0.46	0.45	0.47
0.6	0.31	0.45	0.41
0.7	0.40	0.45	0.44
0.8	0.27	0.45	0.42
0.9	0.37	0.40	0.39
1.0	0.40	0.44	0.42
2.0	0.34	0.36	0.34
Control (SB)	N/A		0.19

To better understand the effects of the KB potential on the transition state, I used four representative barriers of the KB potential for bond-φ-value analyses in region S. In Fig. 3.7, φ-value maps are plotted for ε_{KB} = 0.2, 0.5, 1.0, and 2.0. There is no apparent changes in the structure of CI2 in the transition region in all cases, except the flexible C-terminal end.

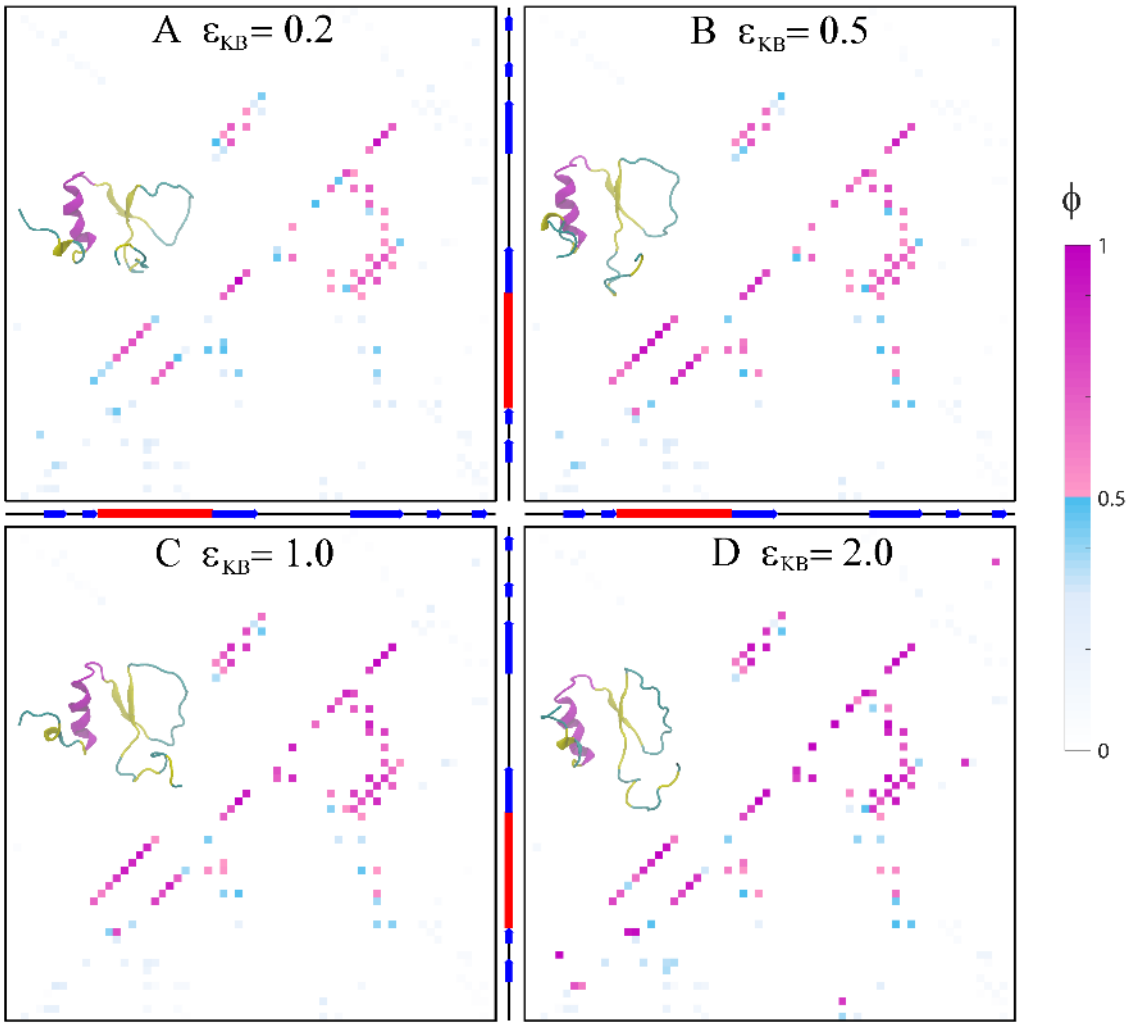

Figure 3.7 Bond φ-values of CI2 with different barriers of KB potential. ε_{KB} = 0.2, 0.5, 1.0 and 2.0 in A-D, respectively. The secondary structures of CI2 are provided along the axes. φ-values of backbone-backbone contacts and sidechain-sidechain contacts are in the upper and lower triangles, respectively. Structures of CI2 are colored according to the secondary structure (α-helix: magenta; β-sheet/β-strand: yellow; rest: cyan).

3.3.2 The statistical dihedral potential reproduces NMR results on Ng

From the previous analyses, I have learned that the KB dihedral angle potential depicts the transition state of the ordered protein CI2. Next, I will explore how well it can describe the unbound states of intrinsically disordered protein Ng. As described in the

method section, I used three different dihedral angle potential models. For each of them, a range of strengths of dihedral angel potential were used. For the structure-based (SB) model, ε_{SB} = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, and 0.7; for the Hybrid model, ε_{SB} = 0.3 and 0.4 for the part G25-A23, ε_{KB} = 1.8, 2.0, and 2.2 for the rest; for the KB model, ε_{KB} = 1.0, 1.5, 1.8, 2.0, 2.2, 2.5, and 3.0 were investigated.

NMR experiment (Ran et al., 2003) showed that, the fragment G25-A42 of Ng protein contains helical populations of 22% and 28% determined from C_α and H_α chemical shifts at 25 ℃, respectively. This temperature corresponds to 1.1 ε / k_B in the coarse-grained simulation by matching the average radius of gyration (Rg) of apoCaM and distribution of Rg with those from the paramagnetic relaxation enhancement experiments. In order to match the helicity of the fragment G25-A42 of the Ng protein, REMD simulations were carried out with the three dihedral angle potentials (Structure-Based, Hybrid and Sequence-Based, please see Table 3.1 for definitions and more details).

Fig. 3.8 shows the helicity (please find the definition in the Appendices 6.11) of the fragment G25-A42 decreases with temperature in all the cases. At the temperature 1.1 ε / k_B, for the SB model, the helicity matches with experimental value when ε_{SB} = 0.3 (Fig. 3.8A); for the Hybrid model, the helical fraction is best matched when ε_{SB} = 0.3, and ε_{KB} does not make any influence in the range from 1.8 to 2.2 (Fig. 3.8B); for the KB model, the helical fraction is best matched when ε_{KB} is in the range from 1.5 to 2.2 (Fig. 3.8C).

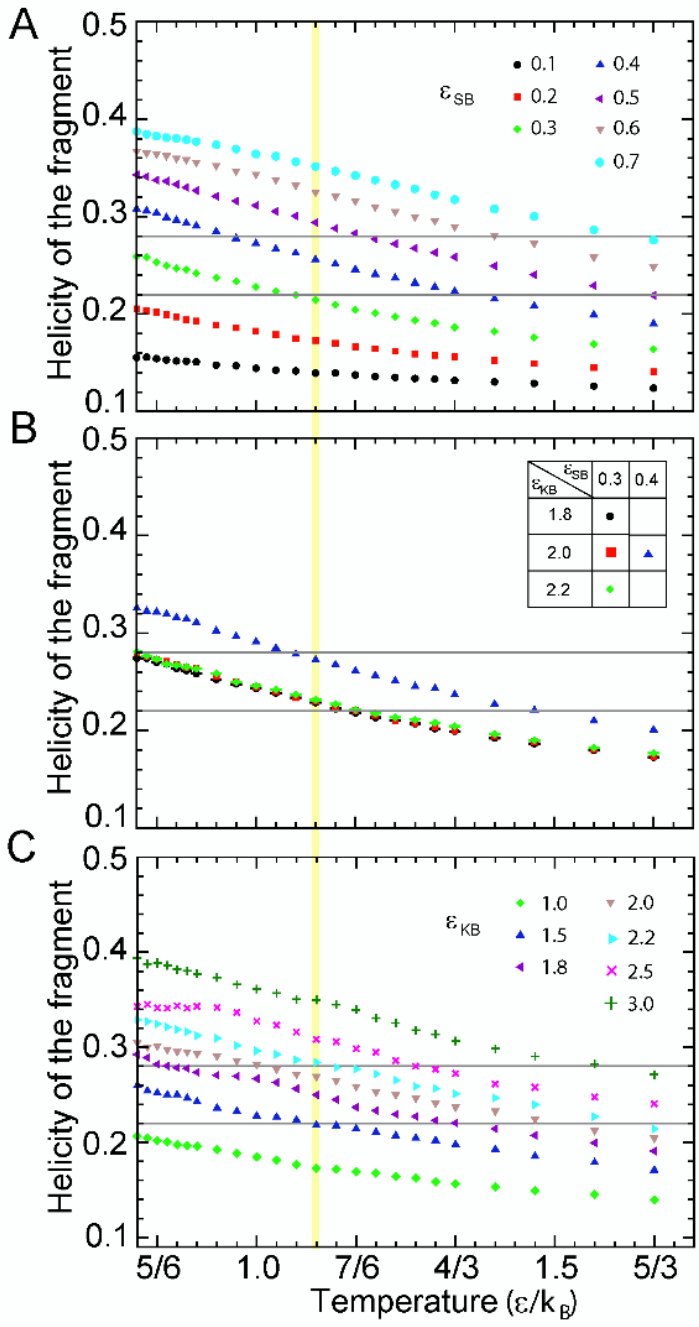

Figure 3.8 Calculated helicity of the fragment of Ng protein. The helicity of the fragment G25-A42 is calculated from REMD simulations of the intact Ng protein with the coarse-grained side-chain C_α model. Three models of dihedral angle potential are used and are shown in three panels. (A) Structure-Based (SB) potential; (B) Hybrid potential; (C) Karanicolas-Brooks (KB) statistical potential. The two gray lines mark the upper and lower limit of the helicity of the segment from the experiment (Ran et al., 2003); the yellow shade marks the corresponding reduced temperature 1.1 ε / k_B. Helicity is computed according to the pseudo dihedral angles between four consecutive C_α beads.

To determine which best represents the properties of Ng among all the three models, I further investigated them at the residue level. Therefore, for those models with parameters that match the overall helicity (Fig. 3.8), I computed the model free order parameter S^2 (see the computational definition in Appendices 6.9) for the backbone beads from my simulations, and compared with data from nuclear magnetic resonance (NMR) experiments (Ran et al., 2003). The computed S^2 (Table 3.3) correlates positively with the experimental values in all cases. Interestingly, for KB model, the correlation is overall higher than the SB and Hybrid model, especially for $\varepsilon_{KB} = 1.8$. Since the KB dihedral angle potential has no bias to any specific structure, it enables sampling of a broad scope of conformations; whereas SB and Hybrid model have full or partial bias to a specific structure, limiting the flexibility to explore more conformations of intrinsically disordered protein Ng. Therefore, for modeling intrinsically disordered protein Ng, I adopted the sequence-based KB dihedral potential model and $\varepsilon_{KB} = 1.8$, which reproduced the average helicity of the fragment as well as best correlation with experimental values of NMR model free order parameter, was used in further study of apoCaM-Ng $_{13\text{-}49}$ binding simulations.

Table 3.3 Correlation between computed and experimental derived model free order parameter S^2.

Models	SB		Hybrid	KB	
	0.3	0.4		1.8	2.0
Correlation Coefficient	0.40	0.73	0.29	0.78	0.64

Chapter 4

Effect of binding targets on Ca^{2+} binding to calmodulin

The work presented in this paper is submitted.

Pengzhi Zhang is the first author of the paper.

4.1 Introduction

Calcium (Ca^{2+}) is exquisitely utilized by the cells for transporting external stimuli to the cells through the gradient of extracellular (~ 1000 μM) and intracellular (~ 0.1 μM) concentration (Gifford et al., 2007). In eukaryotic cells, protein calmodulin (CaM) encodes Ca^{2+} signals (Yamniuk and Vogel, 2004) by binding Ca^{2+} with its four calcium binding loops. CaM activates numerous downstream CaM-binding enzymes that regulate cellular functions including apoptosis (Soderling, 1999; Berchtold and Villalobo, 2014), muscle contraction (Sparrow et al., 1981; Walsh, 1994) and synaptic plasticity in memory formation (Colbran and Brown, 2004; Xia and Storm, 2005). As part of changes in synaptic plasticity in neurons, CaM activates Ca^{2+}/CaM-dependent protein kinase II (CaMKII) at high Ca^{2+} level (Lisman et al., 2002) while the availability of CaM is controlled by a group of post-synaptic proteins such as neurogranin (Ng) at a lower Ca^{2+} level (Gerendasy et al., 1994; Diez-Guerra, 2010). Both CaMKII (Yamauchi, 2005) and Ng (Pak et al., 2000) play a key role in learning and memory formation. However, how the Ca^{2+} affinity for CaM is regulated by Ng and CaMKII at the molecular level is not well understood. I performed the first computational simulation study to explore the molecular underpinnings of Ca^{2+} release associated with synaptic plasticity that is critical to learning and memory formation (Llinas

et al., 1991; Moriguchi et al., 2006; Pang et al., 2010; Zhong and Gerges, 2010) of the differential binding process of Ng and CaMKII to CaM,

CaM has two domains connected by a flexible linker that allows different conformations of CaM (Stigler et al., 2011). The CaM can bind to the targets in various forms depending on the target (Hoeflich and Ikura, 2002; Yamniuk and Vogel, 2004; Tidow and Nissen, 2013). For example, in a canonical form of CaM-target complex, calcium-bound CaM (Ca^{2+}-CaM) binds to a CaMKII peptide by wrapping its two domains around the CaMKII peptide which forms a helical structure (Meador et al., 1993). The binding affinity is largely increased in the presence of Ca^{2+} ions (Xia and Storm, 2005). In this canonical binding conformation, the separation in the sequence between hydrophobic residues of a binding motif from CaM binding targets (CaMBTs) are evolutionarily conserved (Yamniuk and Vogel, 2004) (Rhoads and Friedberg, 1997). The binding of particular category of CaMBTs can drastically increase CaM's affinity for Ca^{2+} by slowing down the dissociation rates of Ca^{2+} from CaM (Yamniuk and Vogel, 2004); this is known as the "reciprocal binding" of target and Ca^{2+} to CaM (Gaertner et al., 2004; Putkey et al., 2008; Swulius and Waxham, 2008; Theoharis et al., 2008; Hoffman et al., 2014).

There is another group of CaMBTs such as Ng that bind to CaM in the absence of Ca^{2+} ions. Ng significantly raises the dissociation rate of Ca^{2+} from calcium-binding loop III and IV in cCaM up to 60-fold with little effect on Ca^{2+} association at low Ca^{2+} concentration (Gaertner et al., 2004; Kubota et al., 2008). Ng has a special kind of binding motif, IQ motif (Jurado et al., 1999; Bahler and Rhoads, 2002) (see the sequence of IQ motif colored in magenta in Fig. 4.1), composed of hydrophobic and basic amino acids. Ng preferably binds to C-domain of CaM (cCaM) (Ran et al., 2003) without a defined complex

structure, excepted when CaM is covalently linked with Ng (Kumar et al., 2013). Previous experimental study revealed that the IQ motif alone does not represent the biophysical characteristics of the intact Ng protein (Hoffman et al., 2014). Inclusion of the acidic region in the N-terminal prior to the IQ motif (Ng $_{13\text{-}49}$, sequence is shown in Fig. 4.1) reproduces the Ng-mediated affinity of Ca^{2+} to Ca^{2+}-free CaM (apoCaM), as well as the intermolecular interactions between Ng and apoCaM (Hoffman et al., 2014). However, the molecular mechanism of how Ca^{2+} affinity for CaM is decreased by association of Ng remains elusive.

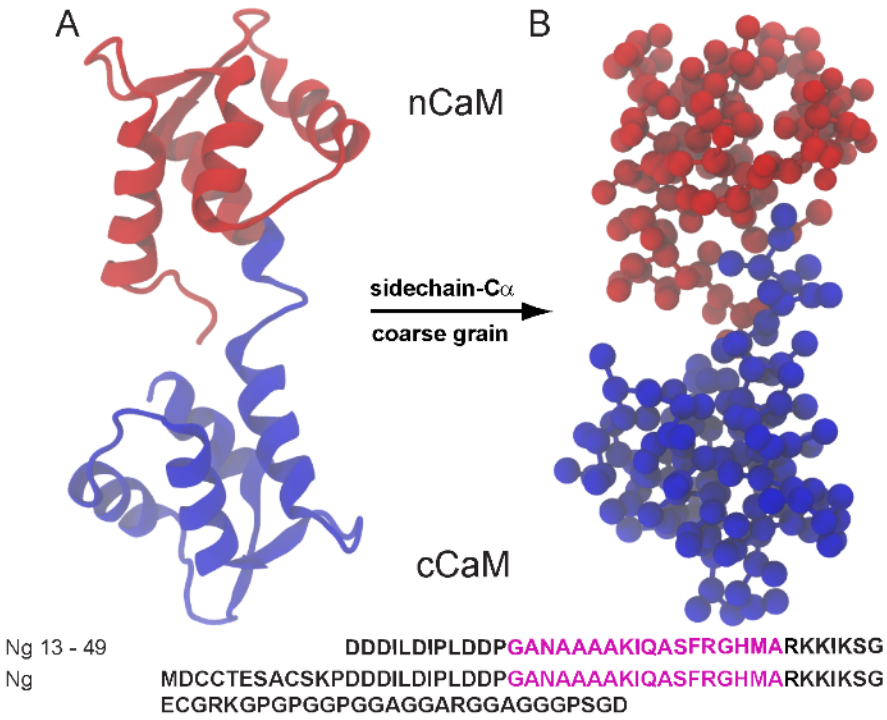

Figure 4.1 Atomistic and coarse-grained representations of the apoCaM and sequence of Neurogranin. (A) The atomistic representation of apoCaM (PDB code: 1CFD) is shown in ribbon representation. ApoCaM is colored from red (nCaM residue 1-74) to blue (cCaM residue 75-148). (B) The coarse-grained representation of apoCaM is shown in balls and sticks, and the colors have the same definitions as in (A). The sequences of the Ng 13-49 from human and the full length Ng protein from rat are provided in bottom panel. The sequence that presents residual structure (including the IQ motif) (Ran et al., 2003) is colored in magenta. The IQ domains share the consensus motif IQXXXRGXXXR, where X stands for any amino acid.

In my prior work, I successfully modeled the interactions between CaM and CaMKII peptide (Wang et al., 2013; Tripathi et al., 2015) with the available complex structure. The main obstacle to computationally investigate CaM-Ng binding is a lack of experimentally solved structures due to weak binding. In this work, I developed *in silico* models to comprehensively understand the binding of Ng to CaM with limited guidance from experiments. I modeled the binding of apoCaM and Ng $_{13\text{-}49}$, which includes both the acidic region and the IQ motif (Hoffman et al., 2014).

Ng is an intrinsically disordered protein/peptide (IDP) that exists in aqueous solution predominantly in random form with the presence of partial helical residual structures (Ran et al., 2003). For performing the computer simulations on apoCaM and Ng $_{13\text{-}49}$ that lacks experimentally defined structures in both the unbound the bound state, I improved the modeling of the target peptide from my prior work (Wang et al., 2013) with a sequence-based dihedral angle potential developed by Brooks's group (Karanicolas and Brooks, 2002). My simulation results revealed a diverse distribution of complex structures of apoCaM-Ng $_{13\text{-}49}$. I selected a cluster of possible candidates for apoCaM-Ng $_{13\text{-}49}$ complex structures by comparing the "apparent chemical shifts" from the coarse-grained simulations and those from the NMR experiments that tracked the change in chemical shifts of apoCaM upon binding of Ng $_{13\text{-}49}$ (Hoffman et al., 2014).

I hypothesized that the conformation of the calcium-binding loop from the c-domain of CaM (cCaM) in a bound CaM-CaMBT complex dictates cCaM's affinity for Ca^{2+}. To validate this hypothesis, I performed fast-growth steered molecular dynamics simulations and applied Jarzynski's equality method (Jarzynski, 1997) to calculate the change in the free energy of the Ca^{2+} release from all-atomistic CaM-CaMBT reconstructed

from selected coarse-grained models. As a control, the same simulations were performed on an extended conformation of Ca^{2+}-CaM without a CaMBT (PDB ID: 1CLL). I found that the presence of Ng destabilizes the Ca^{2+} binding to cCaM by forming a bent structure with an acidic region that pries open the two Ca^{2+} binding loops in cCaM. On the contrary, CaMKII stabilizes the Ca^{2+} binding to cCaM by adopting a canonical binding conformation where a helical CaMKII peptide is wrapped around by CaM involving a robust scaffold of paired Ca^{2+} binding loops from the cCaM.

4.2 Materials and Methods

4.2.1 Sample preparation

Ng $_{13\text{-}49}$ is used to study the interaction with apoCaM (PDB ID: 1CFD) that allows the comparison of the results from computer simulations with the experimentally measured dissociation constant of apoCaM and Ng $_{13\text{-}49}$, the affinity of Ca^{2+} to CaM, and the changes in the chemical shifts of apoCaM upon Ng $_{13\text{-}49}$ binding (Hoffman et al., 2014). The sequence is shown below,

```
                20                  30                  40
13 DDDILDIPLDDPGANAAAAKIQASFRGHMARKKIKSG 49
```

4.2.2 Coarse-grained protein or peptide models

Hamiltonian of Ng peptide. As seen in Chapter 3, a sequence-based dihedral angle potential was used to sample non-specific conformations for Ng peptide.

Hamiltonian for apoCaM. I used apoCaM for this study. The Hamiltonian for the apoCaM takes the same form of those for the Ng $_{13\text{-}49}$ or Ng except the dihedral angle

potential. The solution structure of apoCaM adopts an extended conformation (Kuboniwa et al., 1995). I therefore incorporate a variant of structure-based model in its dihedral term by following Eqn. A2 in the Appendices where the equilibrium angles are from native unbound structure (PDB code: 1CFD). The strength for dihedral angle potential is set to 2.5ε so that distribution of radius of gyration (Rg) of apoCaM shows two distinct populations corresponding to open and closed conformation of apoCaM at a ratio of 9.5:0.5, respectively as measured in paramagnetic relaxation enhancement experiment (Anthis et al., 2011).

Hamiltonian for ApoCaM-Ng system. According to experimental measurement of Hoffman et al. (Hoffman et al., 2014), peptide Ng $_{13\text{-}49}$ can best represent the full-length protein to interact with calmodulin. Therefore, Ng $_{13\text{-}49}$ is used to study the binding interaction with apoCaM. The total potential energy of apoCaM-Ng system consists of three terms,

$$E^{CaM-Ng} = E^{CaM} + E^{Ng} + E^{inter} \quad \text{Eqn. (4.1)}$$

$$E^{inter} = E_{elec} + \lambda(E_{HB} + E_{LJ}) \quad \text{Eqn. (4.2)}$$

where the three types of nonbonded inter-molecular interactions in Eqn 4.2 take the same form as intra-molecular interactions in Eqns A3–A5, respectively. In addition, I rescaled the strength for the E_{HB} and E_{LJ} in Eqn 4.2 by a factor of λ to reproduce the experimentally measured dissociation constant. The dissociation constant K_d of apoCaM and Ng $_{13\text{-}49}$ was calculated by $K_d = e^{-\Delta PMF/k_BT}$ (M). ΔPMF, the change in the potential of mean force (PMF) between bound and unbound states of apoCaM-Ng$_{13\text{-}49}$, is the reweighed stability

from the umbrella sampling simulations. I used WHAM (Roux, 1995; Grossfield, 2013) to reweight the ΔPMF along the distance between the center of mass of apoCaM and that of Ng $_{13\text{-}49}$ (d_{COM}), as shown in Fig. 4.2. k_B is Boltzmann constant, and T is the temperature 1.1 ε/k_B. The dissociation constant (K_d = 680 nM) of apoCaM and Ng $_{13\text{-}49}$ peptide was measured by calorimetry experiments at salt concentration of 0.15 M at 25 ℃, pH = 7.2 (Hoffman et al., 2014). Using λ = 1.4 achieves the optimal estimate of K_d (K_d = 119 nM, Table 4.2), which is in the same order of magnitude with the experiment.

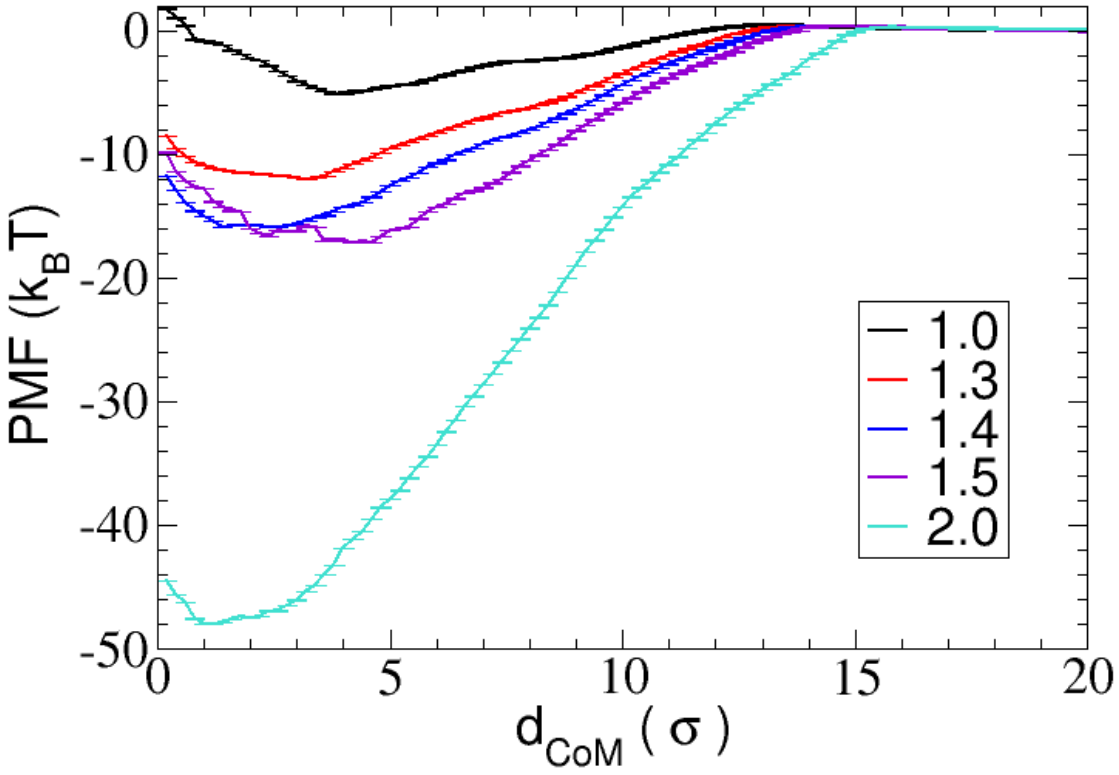

Figure 4.2 Reweighted potential of mean force of apoCaM and Ng $_{13\text{-}49}$. The PMF was reweighted from umbrella sampling simulations using WHAM at varying scaling factors of the inter-molecular nonbonded interaction λ (excluding electrostatic interactions). KB statistical dihedral angle potential was employed. d_{COM} is the distance between center of mass of apoCaM and center of mass of Ng $_{13\text{-}49}$. σ equals 3.8 Å.

Table 4.1 Dissociation constant K_d of apoCaM and Ng $_{13-49}$ with several values of strength of non-electrostatic intermolecular hydrogen bonding and van der Waals interactions λ. λ is shown in Eqn. 4.2. Experimentally measured K_d = 680 nM (Hoffman et al., 2014).

Strength λ	ΔPMF (k_BT)	K_d (nM)
1.0	5.12	6.0×10^{6}
1.3	12.04	5.9×10^{3}
1.4	15.94	119
1.5	17.16	35.4
2.0	48.01	1.4×10^{-11}

4.2.3 Details of the coarse-grained simulations

Umbrella sampling (US) method was used to enhance sampling of rare events in exploration of the thermodynamic properties of the apoCaM-Ng system. To generate the initial structures for umbrella sampling simulations at each window, dissociation molecular dynamics simulations were first performed for the apoCaM-Ng $_{13-49}$ complex starting from the bound form (PDB code: 4E50, in coarse-grained form, missing atoms were added) at a high temperature of 1.33 ε / k_B (ε is reduced energy unit, which equals 0.6 kcal/mol, and k_B is Boltzmann constant). A total of 5 sets of initial configurations for the following umbrella sampling simulations were generated from the dissociation trajectory.

The distance between the center of mass of apoCaM and the center of mass of Ng $_{13-49}$ was restrained by a harmonic force with spring constant 66.7 ε / σ^2 (σ is reduced length unit, which equals to 3.8 Å). The equilibrium positions of the harmonic force range from 0.2 σ to 20.0 σ with a bin size 0.2 σ, making up a total of 100 windows. For each window, 10,000,000 time-steps of constrained molecular dynamic simulation were carried out.

In order to generate an ensemble of structures for calculation of partial charges, US simulations were first performed without charge-charge interactions. After I obtained the

partial charges for apoCaM and Ng $_{13\text{-}49}$, US simulations were carried out with charge-charge interactions to study how the strength of non-electrostatic intermolecular interactions influences the binding affinity and to determine the optimal strength by comparing with experiments (Hoffman et al., 2014). Using the optimal strength of non-electrostatic intermolecular interactions, US simulations were carried up at appropriate experimental conditions to study the binding thermodynamic properties of apoCaM-Ng $_{13\text{-}49}$ system. Each set of US simulations were performed from 5 different initial structures at the temperature of 1.1 ε/k_B. Thermodynamic properties and errors were calculated after reweighting using WHAM.

4.2.4 Selection of initial structures for free energy calculation

For apoCaM-Ng $_{13\text{-}49}$ complex, because the complex of Ng and apoCaM is instable in solution and no structure is not available, I made use of coarse-grained molecular simulations, which are capable to sample efficiently a broad ensemble of complex structures. I employed the experimental measurements as a guide to strategically select four structures from the main cluster (cluster 1) of coarse-grained complex structures. According to the crystal structure of apoCaM-NgIQ (PDB ID: 4E50), the EF hands from cCaM are open (EF hand angles = 101°) and the EF hands from nCaM are closed (EF hand angles = 129°) (Table 4.1). The EF hand angles are computed according to inner product of the vectors that define the orientation of the helices in the EF hand motif. The vectors are defined from the center of mass of the C_α atom in the first four residues and the center of mass of the C_α atoms in the last four residues in a helix of the EF hand motif. Moreover, NgIQ only interacts with cCaM and NMR experiments (Hoffman et al., 2014) showed that Ng mainly interacts with cCaM. I set the criteria that (i)

angles of EF hands from cCaM must be within (85° to 105°) and angles of EF hands from nCaM must be greater than 125°; (ii) cCaM has more interactions with Ng $_{13\text{-}49}$ than nCaM to select four structures and reconstructed them with atomistic details for further analyses. For holoCaM, I assessed the affinity of Ca^{2+} for CaM in the absence of the CaMBT as a reference. The crystal structure of holoCaM (PDB ID: 1CLL) was used. For CaM-CaMKII complex, the crystal structure of CaM-CaMKII complex (PDB ID: 1CDM) served for comparative studies.

Table 4.2 EF-hand angles in different forms of CaM.

	EF 1	EF 2	EF 3	EF 4
apoCaM (PDB: 1CFD)	131°	123°	127°	133°
holoCaM (PDB: 1CLL)	87°	84°	105°	96°
apoCaM-NgIQ (PDB: 4E50)	129°	129°	100°	101°
CaM-CaMKII (PDB: 1CDM)	84°	83°	98°	94°

4.2.5 Reconstruction of apoCaM-Ng $_{13\text{-}49}$ complex

The atomistic initial structures of apoCaM-Ng $_{13\text{-}49}$ were reconstructed based on the sidechain-C_α configurations of the complex from the coarse-grained molecular simulations as shown in Fig. 4.3, using SCAAL method (Samiotakis et al., 2010), Four Ca^{2+} ions were added and their positions were estimated as the center of mass of sidechain beads of the corresponding Ca^{2+} co-ordination residues and were optimized by energy minimization using conjugate gradient and line search algorithm provided by molecular dynamics software package NAMD.

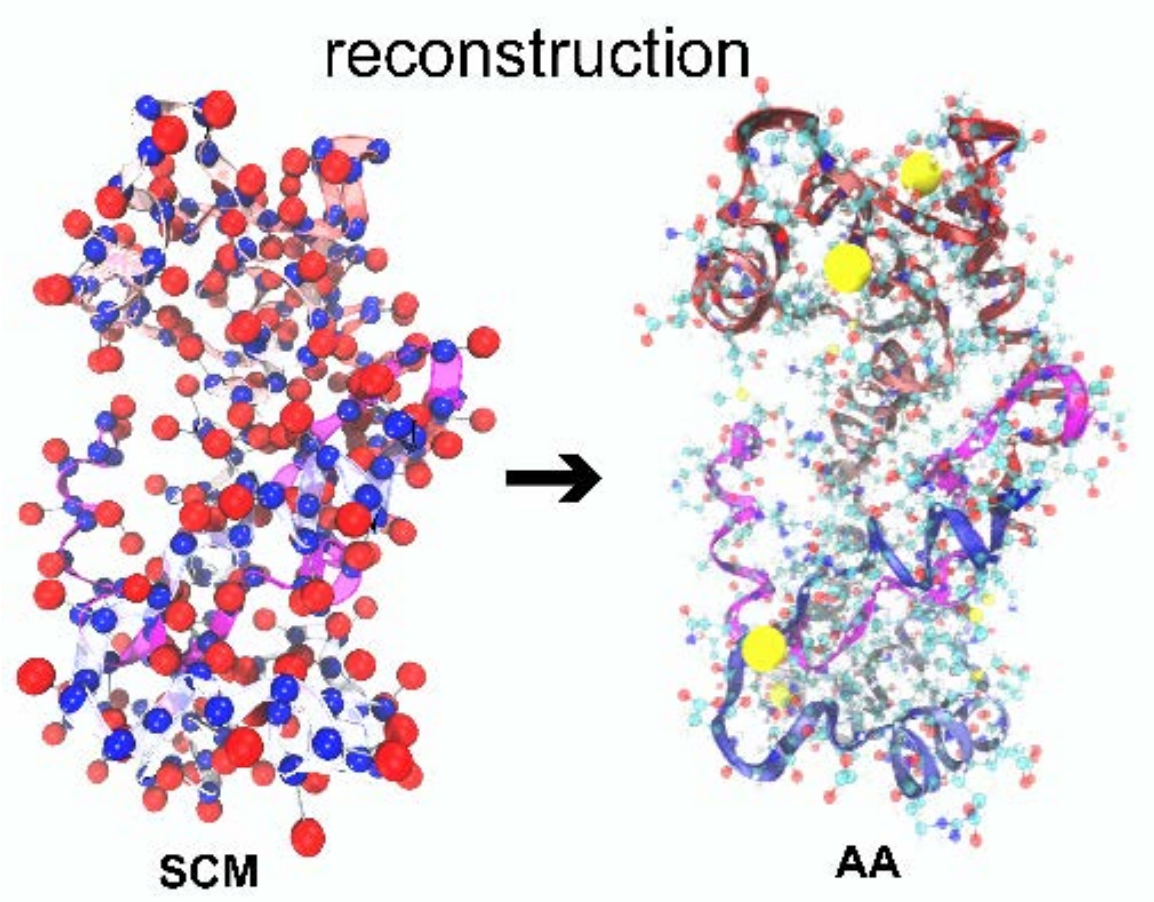

Figure 4.3 Illustration of protein reconstruction from coarse-grained configuration to atomistic configuration for the CaM-Ng $_{13\text{-}49}$ complex. The backbone of CaM and Ng $_{13\text{-}49}$ is shown in ribbon representation. CaM is colored from red (N-Domain) to blue (C-Domain); Ng $_{13\text{-}49}$ is shown in magenta. (left) The Sidechain-C_α model (SCM), where the backbone atoms are represented by the C_α beads (blue balls) and sidechain atoms are represented by the sidechain beads (red balls). (right) All-Atom (AA) configuration of the CaM-Ng complex with Ca^{2+} (yellow balls) added to the Ca^{2+} binding loops. The other atoms are colored according to the atom names (oxygen atoms are in red, carbon in cyan, hydrogen in white, nitrogen in blue, etc.).

4.2.6 Fast-growth method for steered molecular dynamics simulations

Classical molecular dynamics simulations were carried out using GROMACS molecular dynamics package (version 5.0.4). The proteins were modeled with AMBER-99SB-ILDN force field (Lindorff-Larsen et al., 2010). The rigid three-site TIP3P model (Jorgensen et al., 1983) was used to simulate water molecules. The systems were

neutralized by Na^+ and Cl^- ions, maintaining an ionic strength of 100 mM. The systems were minimized using steepest descent method before a short 0.1 ns NVT equilibration at 300K and a following 1 ns NPT equilibration with proteins fixed. The side-chain atoms of the proteins were afterwards released and were further equilibrated for another 1 ns. All NPT simulations maintained a constant pressure of 1 bar and temperature of 300 K using the Parrinello-Rahman barostat (Parrinello and Rahman, 1981). The bond lengths in proteins were constrained using the LINCS algorithm of Hess (Hess et al., 1997). The equations of motion were integrated using a 2-fs time steps. The proteins were placed at least 1 nm away from the edges of the cubic box and the corresponding Ca2+ binding loops were at the center of the box. The size of the box was approximately 10*10*10 nm^3. Periodic boundary conditions were employed to mimic the macroscopic settings for electrolytes. Electrostatic interactions between periodic images were treated using the particle mesh Ewald approach (Darden et al., 1993), with a grid size of 0.16 nm, fourth-order cubic interpolation and a tolerance of 10^{-5}. A cutoff of 10 Å was used for van der Waals interactions and real space Coulomb interactions as well as for updating neighbor lists.

For each of the initial structures including 1 structure for each of holoCaM and holoCaM-CaMKII and 4 selected structure of CaM-Ng_{13-49} , I pulled the Ca^{2+} from site III and IV for 4000 times independently to 4 nm away where the interaction between the Ca^{2+} binding loop and the corresponding Ca^{2+} ion is negligible. The backbone heavy atoms of the proteins were fixed while the side-chains were flexible. The force constant k and speed of pulling v of the reference position are described in the next section. The direction of the pulling force was randomly assigned and points away from the center of mass of CaM to

avoid clashes between CaM and the Ca^{2+}. The pulling angle was selected if the angle Ω between the pulling vector and the vector connecting center of mass of CaM and center of mass of the Ca^{2+} binding loop is within 90 degrees; otherwise leads to a large work (as seen in Fig. 4.4). The displacement of the Ca^{2+} ion and the pulling forces are printed every 20 fs for calculation of the work as shown in Eqn. 4.3. The coordinates and velocities of the system are saved every 1 ps.

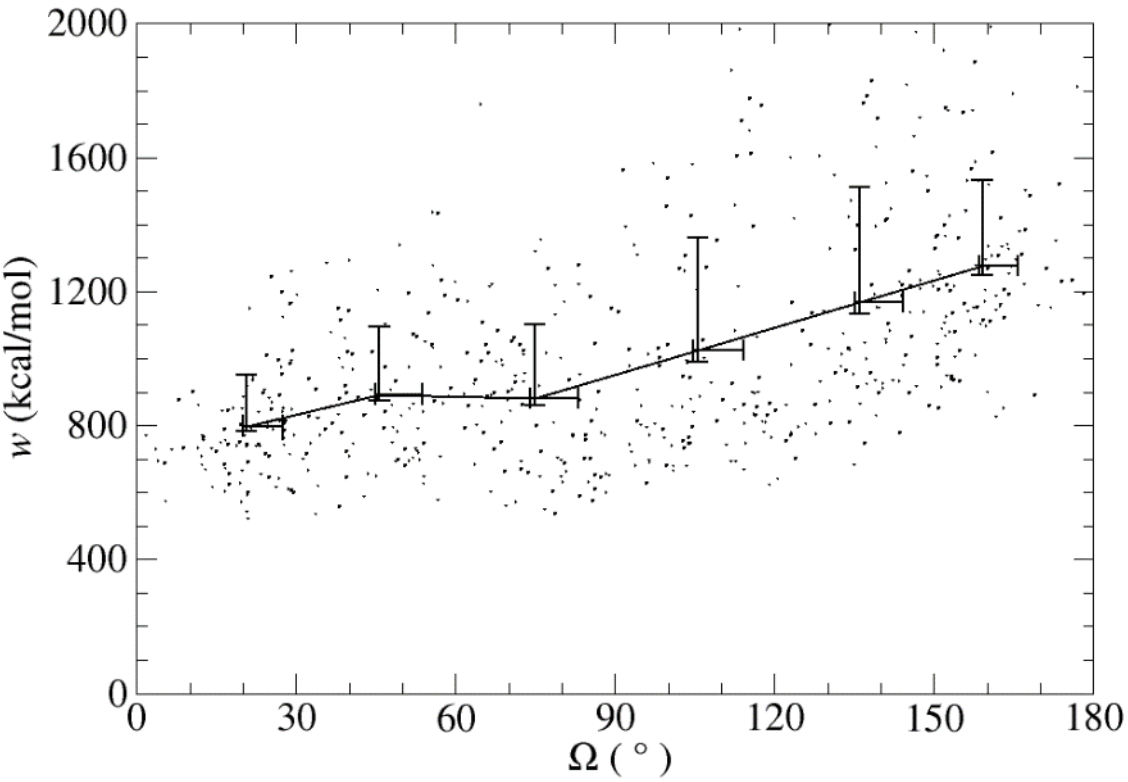

Figure 4.4 Work needed for pulling Ca^{2+} at different pulling angles. At each pulling angle zone Ω in [0, 30], [30,60], [60,90], [90,120], [120,150], [150,180], the Ca^{2+} were pulled from site IV of holoCaM for 100 times. The average work, deviation (upper bar) as well as standard error (lower bar) are provided. The pulling speed v = 0.01 nm/ps and the spring constant of the pulling force k = 1000 kJ/mol/nm^2.

The setup of the pulling simulations is illustrated in Fig. 4.5. The Ca^{2+} is constrained to a reference dummy bead which moves along $\vec{x}$ direction at a speed of v = $|\vec{v}|$ = 0.01

nm/ps. The force constant of the spring k is set to 1000 kJ/mol/nm^2 to guarantee that the Ca^{2+} strictly follows the reference position. Therefore, the pulling force is calculated as,

$$\vec{f}(t) = -k(\vec{x}(t) - \vec{v}t - \vec{x}_0) \qquad \text{Eqn. (4.3)}$$

where $\vec{x}$ ($\vec{x}_0$) is the instantaneous (initial) displacement of the Ca^{2+} from the center of mass of the Ca^{2+} binding loop.

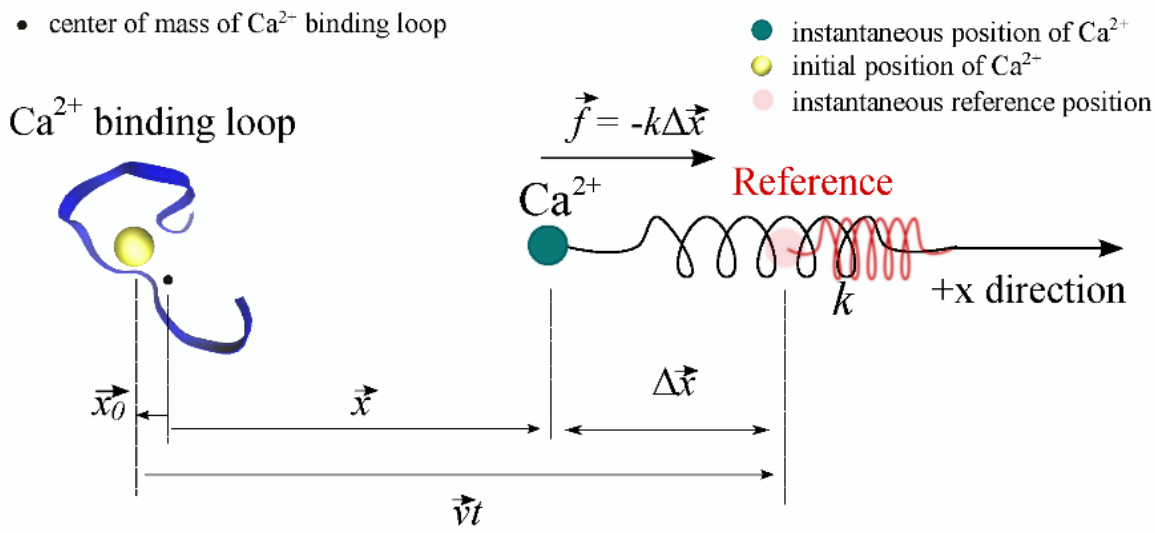

Figure 4.5 Schematic illustration of pulling Ca^{2+} from one of the calcium-binding loop of CaM. The rest of the system including CaM and the solvent molecules is not shown in this illustration for better visualization.

4.2.7 Justification of parameters used in steered molecular dynamics simulations

The accuracy of the calculation of free energy using steered molecular dynamics relies on the pulling speed v and the spring constant of the external force k. Therefore, I explored different combinations of spring constant and pulling speed to rationalize the parameter setting. I took it as an example to pull Ca^{2+} from site III of holoCaM, using k = 100, 1000, 10000, and 100000 kJ/mol/nm^2; v = 0.001, 0.01, and 0.02 nm/ps. The simulation time is 4 ns, 0.4 ns, and 0.2 ns to ensure that Ca^{2+} is pulled the same distance. In Fig. 4.6(A),

with fairly slow pulling speed, using k = 1000, 10000, and 10000 kJ/mol/nm^2, the Ca^{2+} follows the motion of the dummy bead. For k = 100 kJ/mol/nm^2, the pulling force is so weak during the entire simulation to pull the Ca^{2+} out. For k = 1000 kJ/mol/nm^2, there exists a lag in the beginning of the simulation, indicating strong favored interaction between CaM and Ca^{2+} at site III. However, the tails that deviate from the straight line indicate that the thermal fluctuations in the unbound state of Ca^{2+} dominates its motion. By increasing v to 0.01 or 0.02 nm/ps, k = 1000, 10000, and 100000 kJ/mol/nm^2 meet the stiff spring approximation, shown in Figs 4.6(B) and 4.6(C). It is known that for the same amount of simulation time, fewer slow pulling trajectories are more efficient than more fast pulling trajectories, therefore, slower pulling speed, v = 0.01 nm/ps, was used.

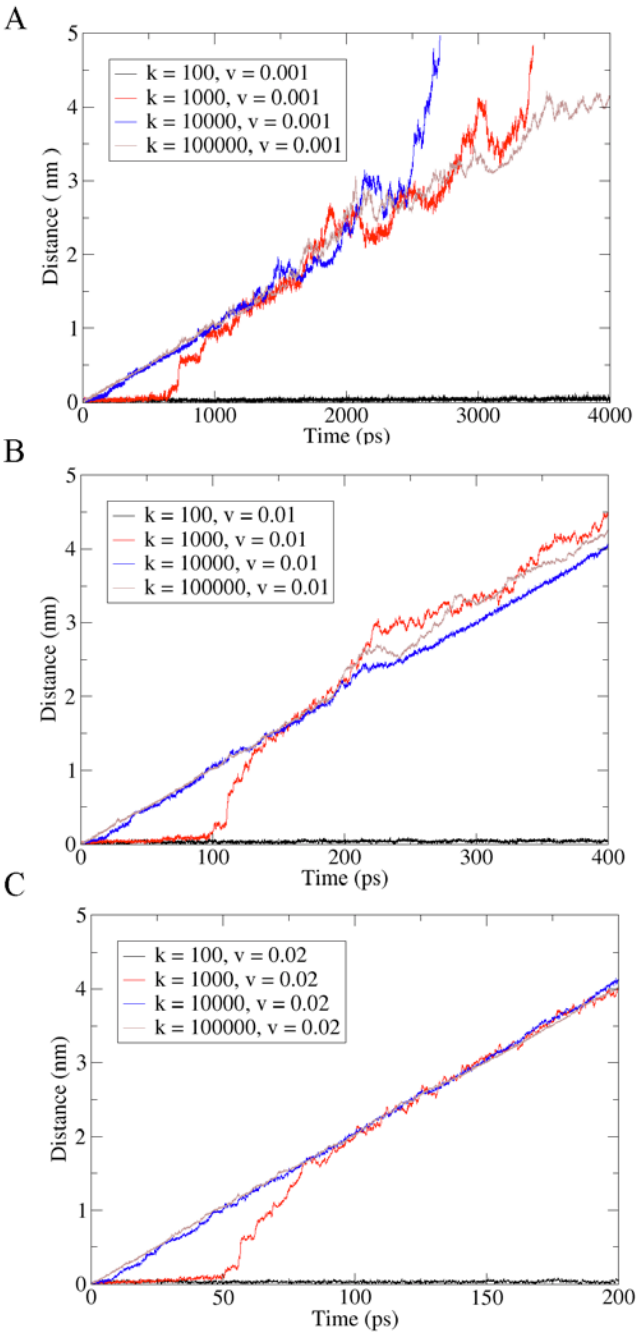

Figure 4.6 The displacement of the Ca^{2+} from the Ca^{2+} binding loop in pulling simulations of Ca^{2+} from the calcium-binding loop III of holoCaM. (A) (B) (C) show the case v = 0.001, v = 0.01, and v = 0.02 nm/ps.

In Fig. 4.7, the typical force profiles lead to the same conclusion. At v = 0.01 nm/ps, k = 10000 and 100000 kJ/mol/nm^2 did not show apparent difference in the force profiles, indicating a converged behavior of the dissociation of Ca^{2+} in this parameter range. Therefore, k = 10000 kJ/mol/nm^2 were used for the pulling simulations.

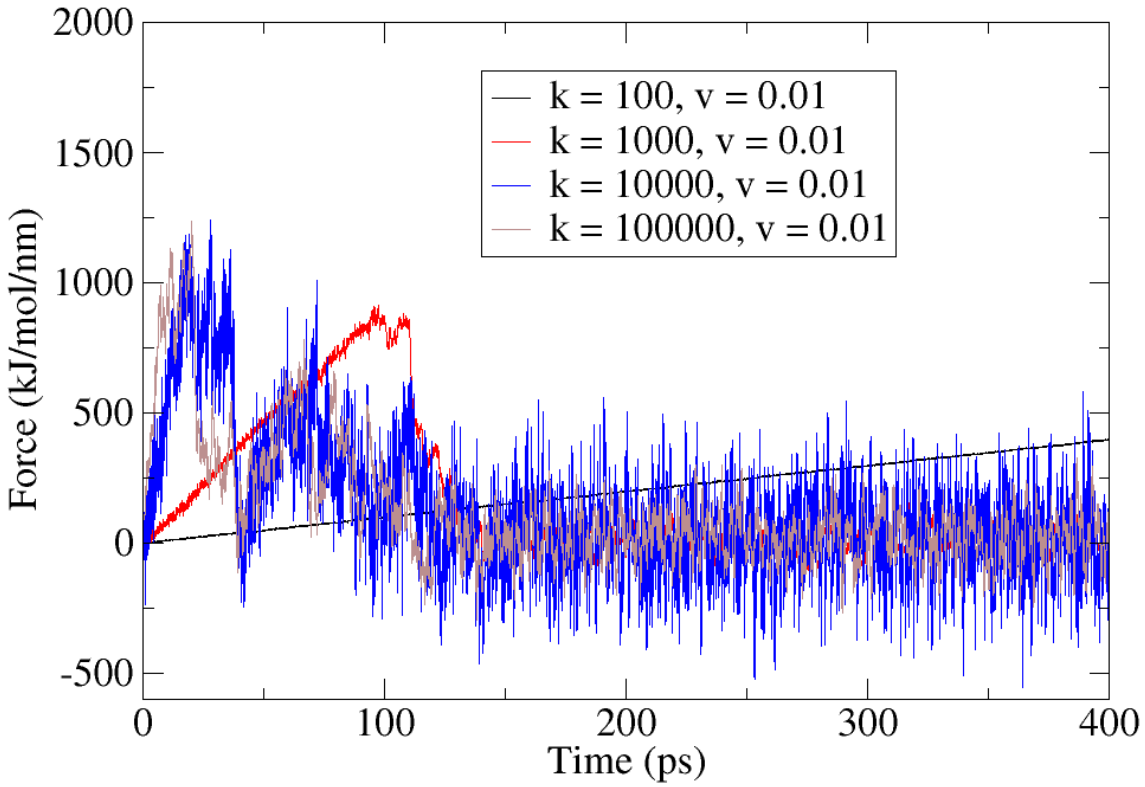

Figure 4.7 The profile of external force in pulling simulations of Ca^{2+} from the calcium-binding loop III of holoCaM. v = 0.01 nm/ps.

4.2.8 Free energy calculation using Jarzynski's equality

For each complex structure, the coordinates as well as velocities are rotated to a random orientation with respect to the pulling direction. Ca^{2+} ions were pulled 100 times toward randomly assigned directions for calculating of the work *w*. The free energy

difference ΔG(CaM-Ng) between a calcium-free complex of CaM-Ng ($G_{unbound}$) and a calcium-bound complex of CaM-Ng (G_{bound}) was calculated according to Jarzynski's equality (Jarzynski, 1997),

$$\Delta G(\text{CaM–Ng}) = G_{bound} - G_{unbound} = k_B T ln \langle e^{-\frac{W}{k_B T}} \rangle \approx k_B T ln \frac{\sum_{n=1}^{M} e^{-w_n/k_B T}}{M}$$

Eqn. (4.4)

where k_B is Boltzmann constant, T is temperature, M is number of pulling simulations (M=100), w_n is the work in n^{th} pulling calculated as

$$w_n(\text{B} \rightarrow \text{U}) = \int_0^t \overrightarrow{f_n}(\tau) \cdot \vec{v}_n d\tau \qquad \text{Eqn. (4.5)}$$

The free energy changes are compared with that from independent simulations of pulling Ca^{2+} from the holoCaM (PDB ID: 1CLL) ΔG(CaM). The difference ΔΔG = ΔG(CaM-CaMBT) - ΔG(CaM) allows us to evaluate the influence of CaMBT on Ca^{2+} binding to CaM. ΔΔG = ΔG(CaM-CaMBT) - ΔG(CaM). B and U stands for bound and unbound states of the Ca^{2+} ion, respectively. If ΔΔG > 0, it means that the CaMBT destabilizes the bound state and thus decreases the Ca^{2+} affinity, as illustrated in Fig. 4.8.

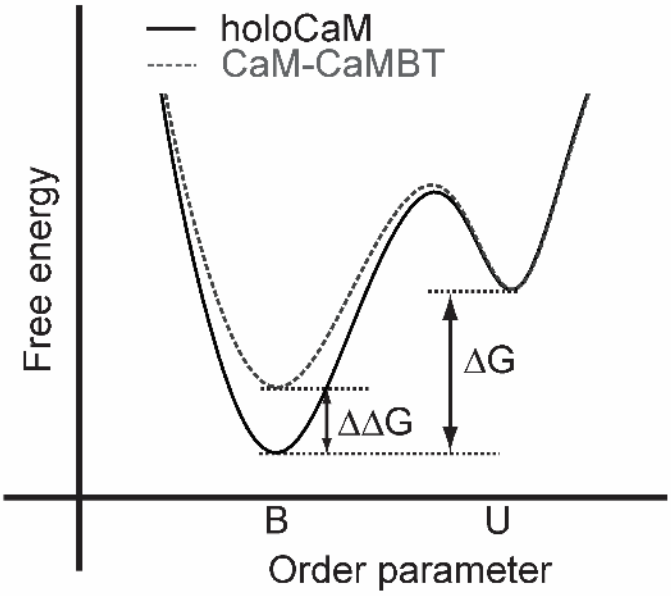

Figure 4.8 Illustration of the definitions of free energy ΔG and ΔΔG. ΔΔG=ΔG(CaM-CaMBT)-ΔG(CaM). B and U stands for bound and unbound states of the Ca^{2+} ion, respectively. If ΔΔG>0, it means that Ng destabilizes the bound state and thus decreases the Ca^{2+} affinity.

4.3 Results

4.3.1 Weak binding between apoCaM and Ng $_{13-49}$

Because there is a lack of defined bound complex for apoCaM-Ng $_{13-49}$, I first aimed to acquire the ensemble of bound complexes by use of coarse-grained simulations with experimental input of the dissociation constant, a measure of the thermal stability of the complex. The binding is fairly weak compared to that between CaMKII peptide and CaM ($K_d = 5*10^{-4}$ nM) (Waxham et al., 1998).

To explore intermolecular contacts between apoCaM and Ng $_{13\text{-}49}$, I plotted the two dimensional reweighted free energy surface as a function of d_{COM} and the number of intermolecular contacts, Z, (see the definition in the Appendices 6.12) using WHAM at λ = 1.4 in Fig. 4.9. At $d_{COM} \geq 15\ \sigma$ (σ = 3.8 Å), the apoCaM and Ng $_{13\text{-}49}$ do not interact with each other (Z ~ 0). At the close proximity of the two where d_{COM} is less than 4 σ and Z is greater than 300, the energy landscape without a global minimum is abundant with shallow basins separated by low barriers at the order of 1 ~ 2 k_BT. The ruggedness of the binding landscape allows the two partners to undergo transient changes in conformations that likely attribute to their weak binding.

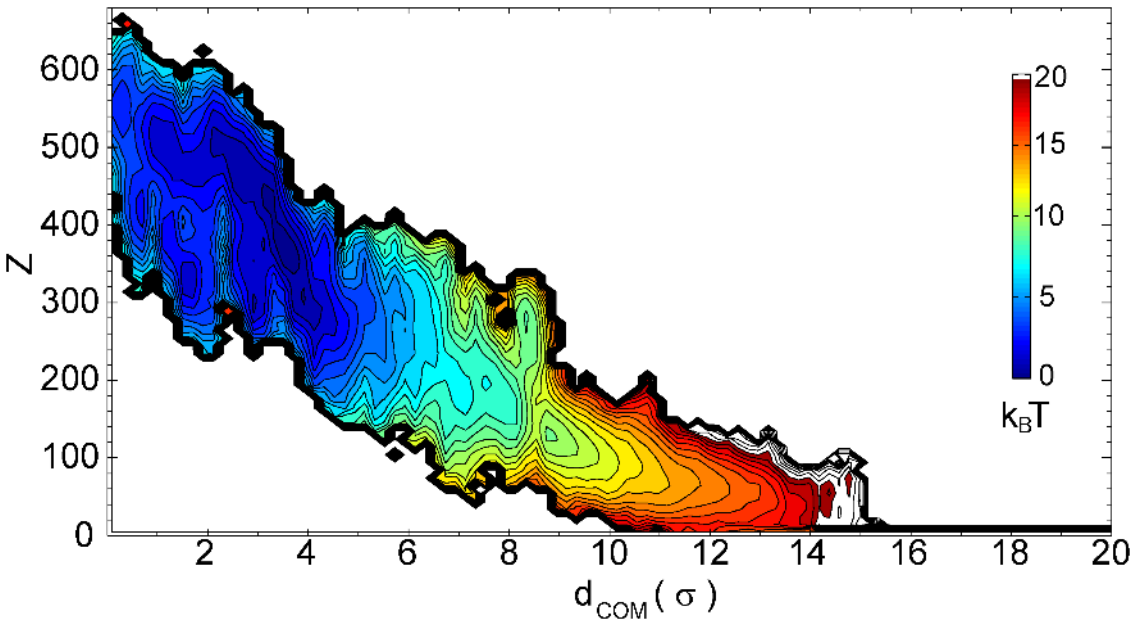

Figure 4.9 Reweighted potential of mean force as a function of center of mass distance between apoCaM and Ng $_{13\text{-}49}$ (d_{COM}) and the number of intermolecular contacts (Z). The coarse-grained molecular simulations were performed at pH = 6.3 and ionic strength = 0.1 M in accord to the conditions of the nuclear magnetic resonance experiments. λ = 1.4, σ = 3.8 Å. The color is scaled in k_BT.

4.3.2 Apparent chemical shift characterizes the weakly-bound apoCaM-Ng $_{13-49}$

ApoCaM samples an ensemble of varying conformations with Ng $_{13-49}$ through non-specific intramolecular contacts. To further narrow down the most probable complex structure of apoCaM-Ng $_{13-49}$, I employed NMR measurement as a guide that characterizes the backbone chemical shifts of apoCaM upon Ng binding. Because of the low-resolution binary data, a signal cannot guarantee a direct contact between CaM and Ng. It nevertheless reveals information on the structural changes of apoCaM at a residue level upon binding with Ng $_{13-49}$ (Hoffman et al., 2014).

By comparing with the spectra of chemical shifts of apoCaM in the absence and presence of Ng $_{13-49}$, the experiment showed that the major structural changes occur in cCaM as well as Helices B and C in nCaM, shown in Fig. 4.10 (I). I generated an ensemble of the transient complexes by utilizing my coarse-grained simulations for a wide sampling of the conformational space of apoCaM-Ng $_{13-49}$. From the reweighed ensemble of apoCaM-Ng $_{13-49}$ structures from the US simulations using WHAM, I then employed an importance sampling method (Samiotakis et al., 2010) to select a subset of 23,722 structures from 5 million conformations for clustering (see 6.13 and 6.14 from the Appendix for importance sampling and clustering method, respectively). The population of the major four clusters (comprising 86% of the total) and correlation coefficient of computationally obtained “apparent chemical shifts” with the experimental NMR data (Hoffman et al., 2014) are seen in Table 4.3.

Table 4.3 The population of major clusters from computer simulations and the correlation coefficient of their computational "apparent chemical shifts" with the data from NMR experiments (Hoffman et al., 2014).

Cluster	1	2	3	4
Population	33%	26%	15%	12%
Correlation Coefficient	0.36	0.07	-0.1	0.07

The apparent chemical shift Δ (see the *Materials and Methods* section for definition) computed from the most dominant cluster 1 (33%) better matches with the experimental data with a relevant correlation coefficient of 0.36. cCaM is mostly stabilized after binding with Ng $_{13\text{-}49}$ and helices B and C are partially stabilized (Fig. 4.10 (II)). In cluster 2, the whole apoCaM shows changes in Δ, implicating interactions throughout the apoCaM and Ng $_{13\text{-}49}$ (Fig. 4.10 (III)). In cluster 3, Ca^{2+} binding loop 1, helix B and BC linker from nCaM, central linker, helices E and G from cCaM are stabilized by interacting with Ng $_{13\text{-}49}$ (Fig. 4.10 (IV)). In cluster 4, flexible parts including Ca^{2+} binding loops 1 and 3, BC linker, central linker and FG linker show changes in Δ (Fig. 4.10 (V)). It is important to note that although the apparent chemical shifts might indicate intermolecular interactions, it is not necessarily true in all cases. For example, although clusters 3 and 4 display broad apparent chemical shifts in nCaM, nCaM makes less direct contacts with Ng $_{13\text{-}49}$ than cCaM (shown by $\Delta Z > 0$ in Table 4.4; see Appendix 6.12 for definition of ΔZ). The analysis that compared to the NMR chemical shifts shows the most interactions between apoCaM and Ng $_{13\text{-}49}$, particularly in cCaM that is structurally more flexible than nCaM.

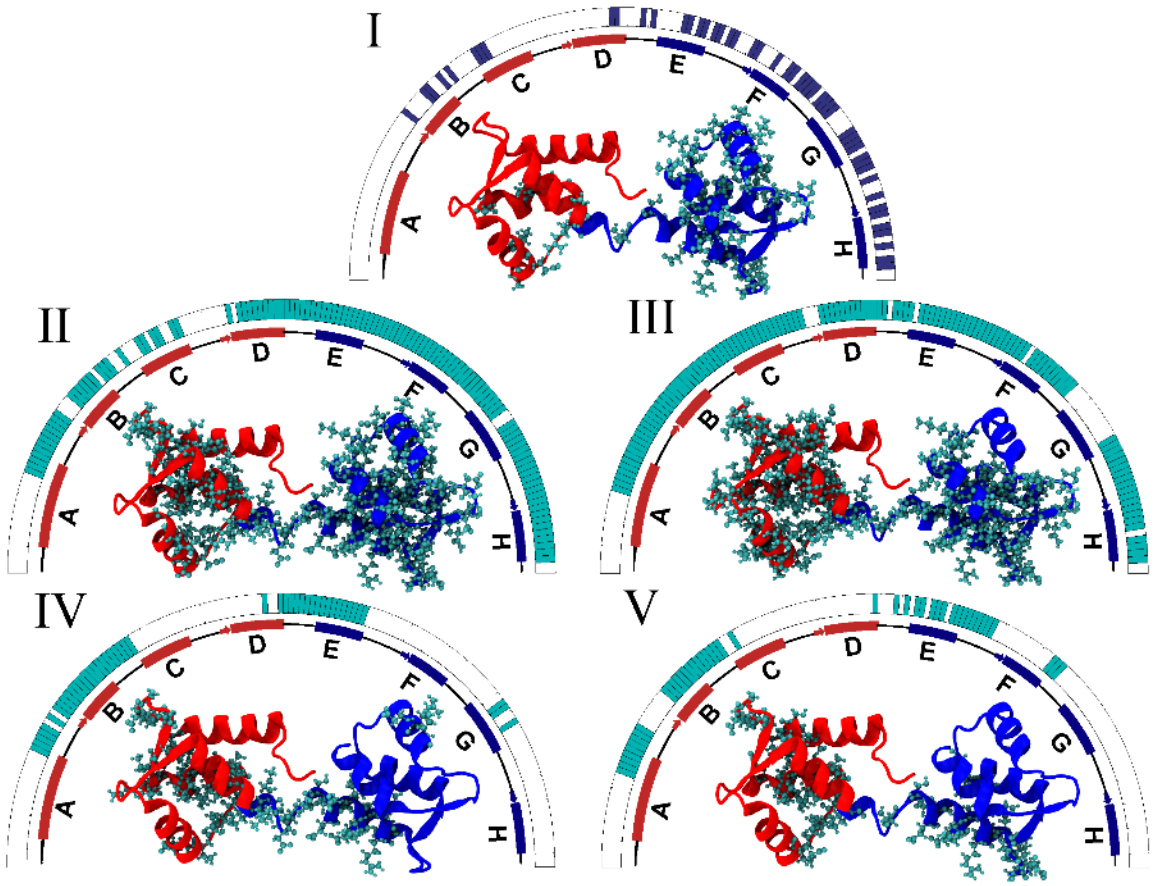

Figure 4.8 Changes in chemical shifts for apoCaM upon Ng $_{13\text{-}49\ binding}$ from the NMR experiment and the "apparent chemical shifts" (Δ) for the clustered structures from computer simulations. (I) in the outside half-circles, blue bars indicate residues with significant changes in chemical shifts from the experiments. The residues that experience chemical shifts are shown in ball-stick representation projected on the apoCaM structure (PDB ID: 1CFD) that are colored from red (N-terminus) to blue (C-terminus). In analog, I computed the apparent chemical shifts" (please see detailed definition in the *Materials and Methods* session) of $\Delta = 1$ from the computer simulations for the four major clustered data for cluster 1, 2, 3, and 4 in (II), (III), (IV), and (V), respectively.

Table 4.4 Properties of major clusters and unbound ensemble.

Cluster	1	2	3	4	unbound
Binding energy	-84.4±0.2	-101.9±0.1	-114.9±0.2	-111.9±0.3	-0.0±0.0
Z	351.9±0.7	414.4±0.5	460.9±0.8	445.4±1.0	0.0±0.0
$\Delta Z = Z_C - Z_N$	37.8±1.4	28.8±1.1	133.3±2.0	80.0±2.5	0.0±0.0
Rg(apoCaM) (σ)	4.832±0.002	4.851±0.002	4.742±0.003	4.798±0.003	4.774±0.000
D_{CoM} (σ)	2.82±0.01	1.89±0.01	2.21±0.02	2.28±0.02	20.00±0.00

σ = 3.8 Å.

4.3.3 Molecular mechanism of apoCaM-Ng $_{13\text{-}49}$ binding

The computed apparent chemical shifts Δ in apoCaM from the most dominant cluster, cluster 1, show the most relevant correlation with the experimental measurements among others. Therefore, I next looked at the contact formation from cluster 1 for better understanding of the molecular mechanism of binding. In the apoCaM-Ng $_{13\text{-}49}$ complexes, contacts within apoCaM (shown in green triangles in Fig. 4.11c) break upon Ng $_{13\text{-}49}$ binding. In addition, the inter-domain contacts between helix A from nCaM and inter-domain linker, helices G and H from cCaM (shown in red box in Fig. 4.11c) are lost to allow interaction with Ng $_{13\text{-}49}$ (shown in blue box in Fig. 4.11a). In contrast, after binding with apoCaM, the Ng $_{13\text{-}49}$ itself loses contacts between IQ motif and the acidic region (Fig. 4.11b, rectangle in dotted line); subsequently, probability of forming α-helix structure in the IQ motif increases (Fig. 4.11b, rectangle in solid line). Although there is no bias to the formation of α-helix structure, my model is able to capture the secondary structure formation of IQ motif after binding with apoCaM.

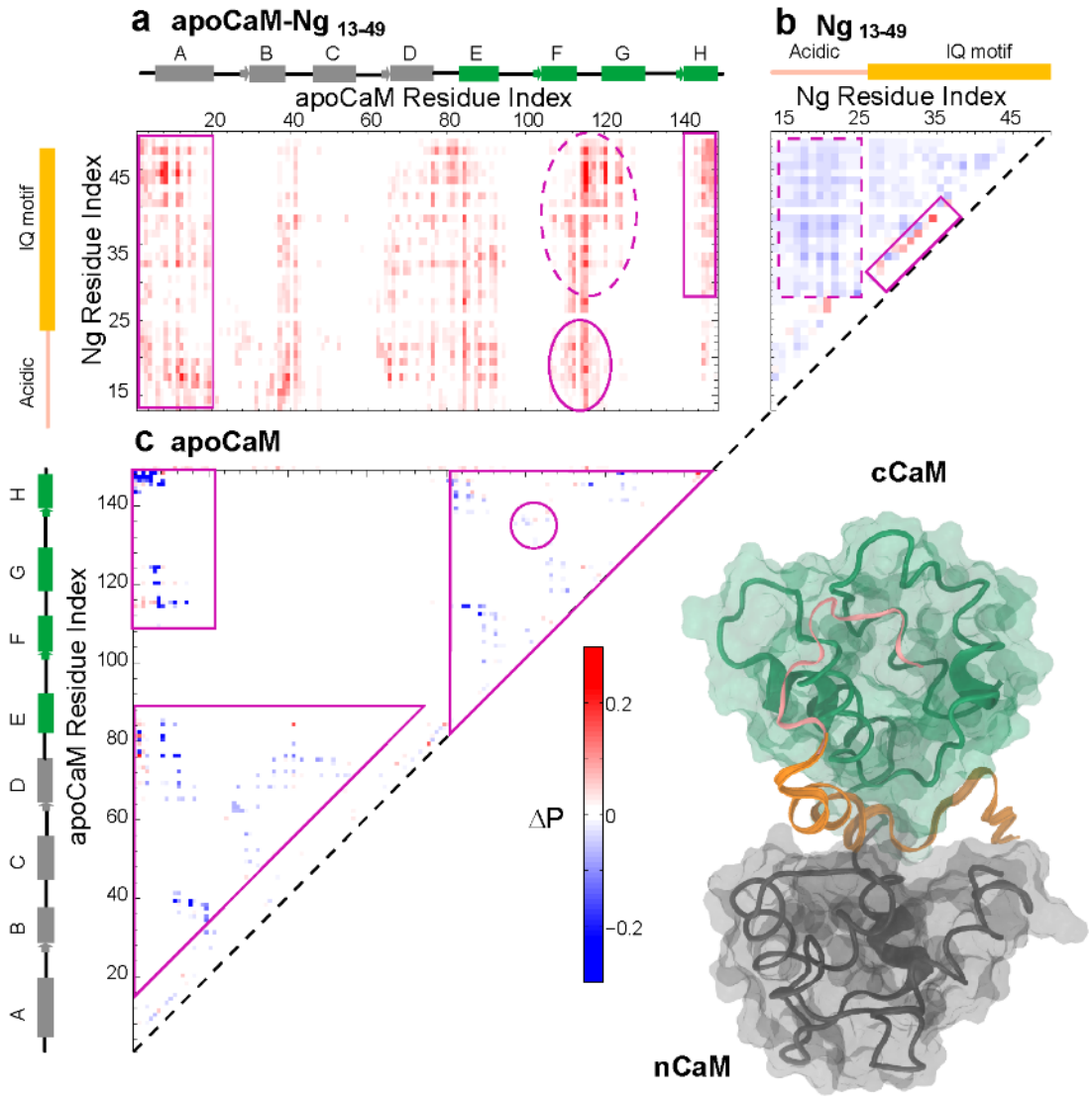

Figure 4.11 Difference in probability of contact formation between the bound ensemble and the unbound ensemble for the most dominant cluster (cluster 1). The difference maps for apoCaM intramolecular (a), Ng $_{13\text{-}49}$ intramolecular (b) and apoCaM-Ng $_{13\text{-}49}$ intermolecular (c) contacts are provided. The schematic representation of the helices of apoCaM and Ng $_{13\text{-}49}$ are provided along the axes. The representative structure, which has minimal root mean square deviation from the averaged structure, is in ribbon representation and colored according to the secondary structures: nCaM is in gray, cCaM is in green, acidic region of Ng $_{13\text{-}49}$ is in pink and IQ motif of Ng $_{13\text{-}49}$ is in orange. The regions of particular interest are encircled by ellipses, rectangles, circles and triangles. The color bar represents the difference in probability of contact formation ΔP between the bound and the unbound conformations where positive ΔP indicates an increase of contact probabilities in the bound ensemble.

For the intermolecular interaction, IQ motif of Ng $_{13\text{-}49}$ increase interaction with Helix A, B/C Helix linker, and Helix D from nCaM, and inter-domain linker, Helix E and Helix H, and the F/G Helix linker from cCaM (Fig. 4.11a, ellipse in dotted line). These regions from CaM also participate in canonical binding of holoCaM (Wang et al., 2013). While the IQ motif of Ng $_{13\text{-}49}$ forms a helical structure and interacts with cCaM of high probability, the acidic region of Ng $_{13\text{-}49}$ remains unstructured (Fig. 4.11b, no change in contact formation along diagonal for the acidic region). The acidic region interacts with both nCaM and cCaM of high probability, especially with F/G helix linker and the central linker (Fig. 4.11a, ellipse in solid line).

In Fig. 4.12, I provided superimposed structures of CaM-Ng $_{13\text{-}49}$ complex during the association, at d_{COM} = 20.0, 10.0, and 2.8 σ, representing unbound, encounter and bound states, respectively. From unbound state (Fig. 4.12 a) to the encounter complex (Fig. 4.12 b), there is no apparent change in the structure of CaM, and IQ motif of the Ng $_{13\text{-}49}$ mainly interacts with cCaM. This transition does not involve major global conformational change and is likely to be diffusion-controlled. From encounter (Fig. 4.12 b) to bound (Fig. 4.12 c), Ng $_{13\text{-}49}$ mainly gained interactions with nCaM and large conformational change of CaM is observed. Specially, the EF-hand β-scaffold in cCaM becomes destabilized caused by insertion of acidic region of Ng $_{13\text{-}49}$.

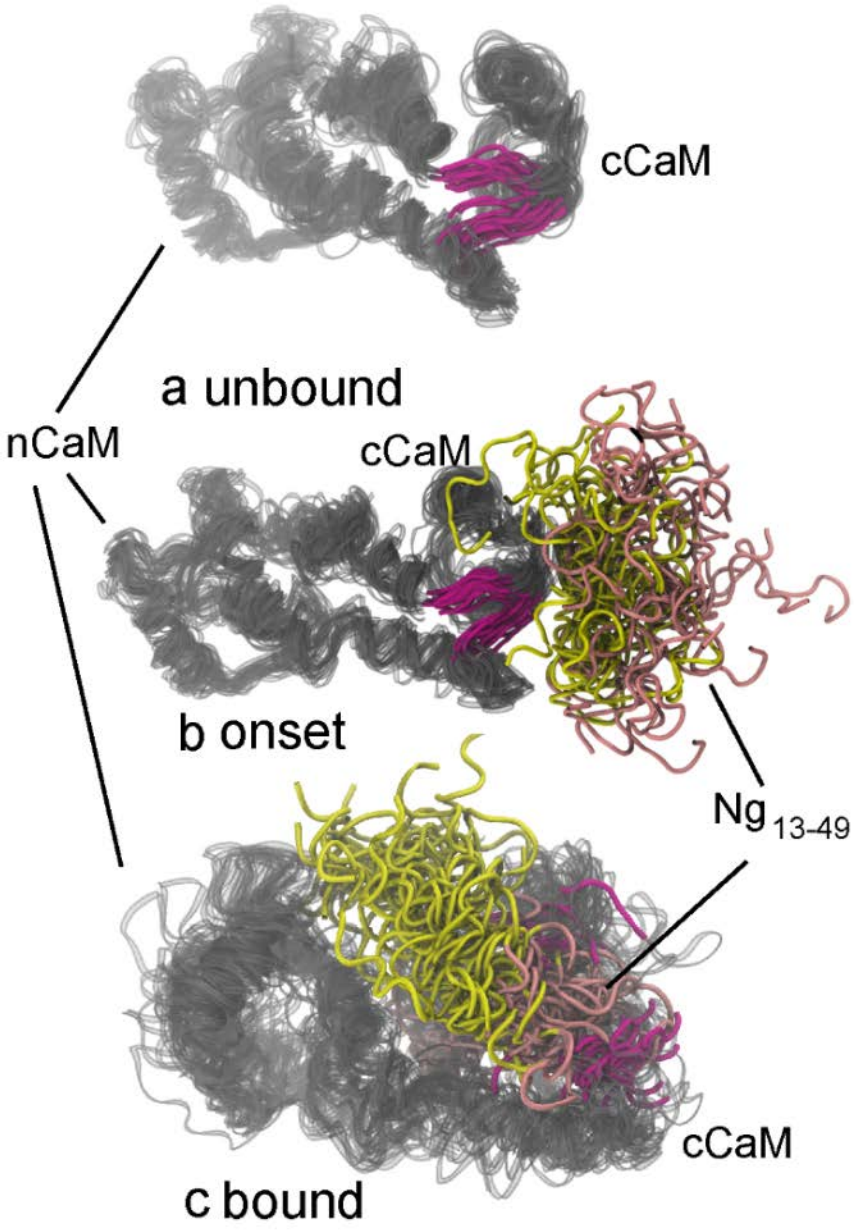

Fig 4.12 Illustration of the structural change in unbound, transition and bound ensemble. The CaM is colored in gray and the residues (residues 99~101 and 135~137) that form EF-hand β-scaffold in cCaM are colored in magenta. The acidic region and IQ motif of Ng $_{13-49}$ are colored in pink and yellow, respectively. (a) unbound state when CaM and Ng $_{13-49}$ are well separated d_{COM} = 20.0 σ (b) IQ motif of Ng $_{13-49}$ interacts with apoCaM and acidic region of Ng $_{13-49}$ has not formed contacts with CaM at d_{COM} = 10.0 σ (c) bound.

In order to better understand the relation between interactions within the two proteins and interfacial contacts, I further investigated the correlation of these pair-wise contacts in Table A11 in the Appendices and found that interactions between acidic region of Ng $_{13\text{-}49}$ and Ca^{2+} binding loops from CaM compete with those within apoCaM (Fig. 4.13). There are complicated correlations between those contacts, which are categorized below.

i. Correlation: as shown in diagonal, the contact formation within CaM, between CaM and Ng $_{13\text{-}49}$, and within Ng $_{13\text{-}49}$ are correlated.

ii. Anti-correlation: Contact formation between nCaM and Ng $_{13\text{-}49}$ anti-correlates contact formation within nCaM; contact formation between cCaM and Ng $_{13\text{-}49}$ anti-correlates contact formation within cCaM. The anti-correlation tells that the interactions within nCaM, cCaM compete with their interaction with Ng $_{13\text{-}49}$. Binding of the target is responsible for the repacking of the nCaM or cCaM either by direct competitive interaction with the Ca^{2+} binding loops or allosteric effects (Sorensen and Shea, 1998).

iii. No correlation: there is no obvious correlation among contacts within nCaM and contacts within cCaM, indicating the two domains of CaM are relative independent on each other.

iv. Mixed correlation: Contact formation between nCaM and cCaM has a complication correlation with contact formation between nCaM and Ng $_{13\text{-}49}$ and between cCaM and Ng $_{13\text{-}49}$.

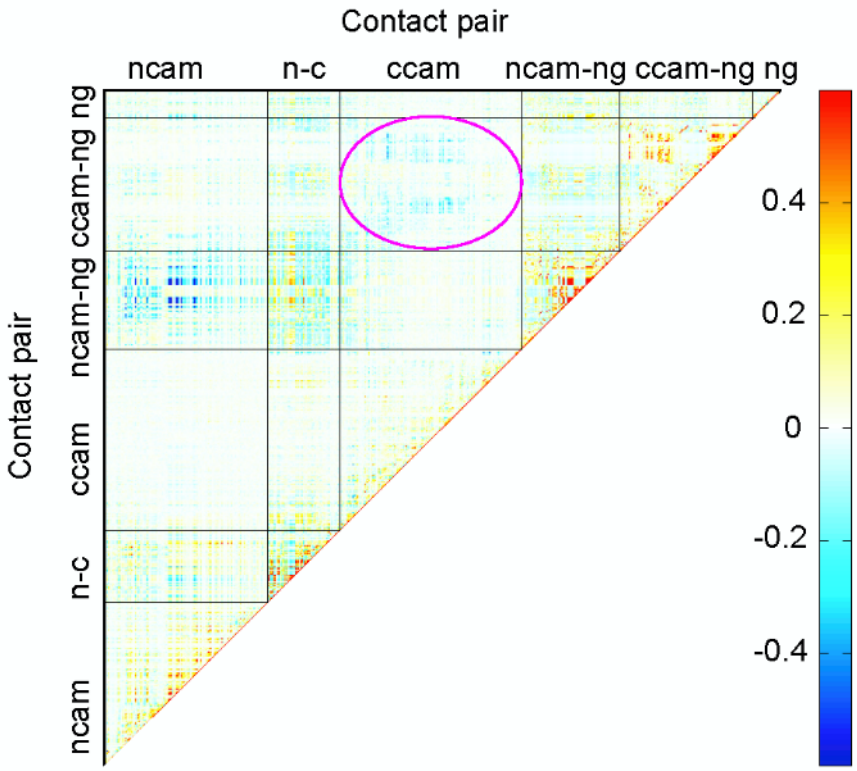

Figure 4.13 Correlation map between contacts formation involving apoCaM and Ng $_{13\text{-}49}$ in bound ensemble (cluster 1). The contact pair list is provided in Table A11 in Appendices. The blue color shows strong anti-correlation between involved contact pairs; the red color shows strong correlation; white means no correlation.

4.3.4 Target binding to CaM modulates the Ca^{2+} binding affinity by fast-growth method

I found that interactions between acidic region of Ng $_{13\text{-}49}$ and Ca^{2+} binding loops from cCaM compete with those within apoCaM. I hypothesized that due to this competition, the β-sheet structure from the Ca^{2+} binding loops (i.e. EF-hand β-scaffold (Grabarek, 2006), where EF hand is a helix-loop-helix motif) becomes less stable, thus promoting CaM's release of Ca^{2+}. In order to validate this hypothesis, I strategically selected apoCaM-Ng $_{13\text{-}49}$ complex structures from cluster 1 of coarse-grained molecular simulations for further

atomistic simulations. Several experimental findings show the characteristics of apoCaM-Ng binding: (i) EF hands in cCaM are open and EF hands in nCaM are closed as suggested from experimental studies (Kumar et al., 2013; Wang et al., 2013); (ii) Ng $_{13\text{-}49}$ has more interactions with cCaM as suggested from experimental studies (Hoffman et al., 2014). This is further shown by the EF-hand angles for CaM in forms of Ca^{2+}-absent, Ca^{2+}-loaded, apoCaM-NgIQ and holoCaM-CaMKII in Table 4.1. Therefore, based on the structural characteristics, I establish the following criteria for selection of apoCaM-Ng $_{13\text{-}49}$ complex structures: (i) Inter-helix angles of EF hands from cCaM must be within (85° to 105°) and those from nCaM must be greater than 125°; (ii) cCaM has more interactions with Ng $_{13\text{-}49}$ than nCaM. As a result, 4 coarse-grained structures were selected from cluster 1 for further reconstruction of all-atomistic protein models (please see ref. (Homouz et al., 2008) for the reconstruction protocol). I computationally placed 4 Ca^{2+} ions in the Ca^{2+} binding sites formed by the EF hands.

By pulling the two Ca^{2+} ions independently from the Ca^{2+} binding sites III and IV from the cCaM, the free energy change between Ca^{2+}-unbound and Ca^{2+}-bound state ($\Delta G = G_{bound} - G_{unbound}$) can be evaluated using Jarzynski's equality (Please see simulations details in method). I computed the free energy changes in the absence and presence of a target ($\Delta\Delta G$). For all of the 4 conformations, the binding of Ng $_{13\text{-}49}$ to CaM destabilizes the Ca^{2+} in site III and site IV, respectively, by showing a positive $\Delta\Delta G$ (Table 4.5). In comparison, I examined the importance of the canonical bound structure for the release of Ca^{2+} from the complex structure of CaM-CaMKII (PDB: 1CDM) where the CaM wraps around and binds antiparallel with well-formed helical structure of CaMKII peptide. Each

site shows a negative ΔΔG, implying that the presence of CaMKII stabilizes the Ca^{2+} in both site III and site IV from CaM.

I found that the distribution of work values does not follow a Gaussian distribution (Fig. 4.14). According to study of Zuckerman's group (Ytreberg and Zuckerman, 2004), the estimation of free energy difference could be very inaccurate using a second order cumulants expansion of the Jarzynski's equality (Eqn. 4.4) or Jarzynski's equality. As seen in Fig 4.15, where the binding free energy ΔG was estimated for Ca^{2+} at binding site III of holoCaM, using a block-average method (subsets of all available work data using Jarzynski's equality). ΔG converged to about -110 kcal/mol, which was far away from experimental values (~ 4 to 5 kcal/mol).

Zuckerman's group developped two extrapolation methods to improve the efficiency of the free energy estimation, linear extrapolation and cumulative integral (CI) extrapolation. Linear extrapolation method estimate the free energy difference by extrapolation to 1/n = 0, where n is total number of work values. CI considers more accurate estimation by using an integral. They have shown the CI extrapolation could reduce the required data by 5-40 fold (Ytreberg and Zuckerman, 2004; Zhang et al., 2006). By using both Jarzynski's equality and CI extrapolation, I summarized the binding free energy in Table 4.5. We found that estimations from Jarzynski's equality were away from experimental values while the CI extrapolation showed fairly good agreement. Later, I would discuss the factors that might contribute to the difference in the binding free energy from experiments.

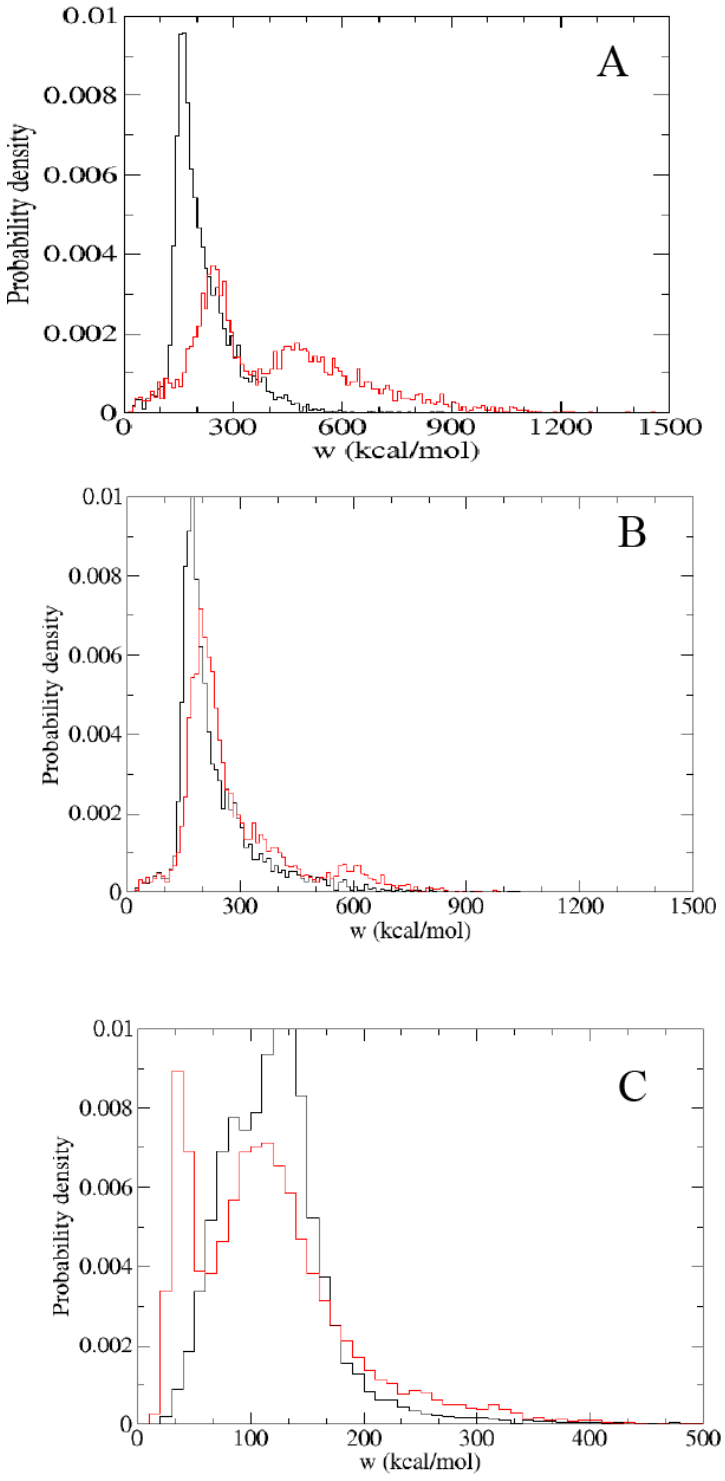

Figure 4.14 Distribution of work for dissociation of Ca^{2+}. (A, B, C) represents the holoCaM, holoCaMKII and holoCaM-Ng $_{13\text{-}49}$, respectively. External work (w) for Ca^{2+} dissociation from site III and site IV are in black and red, respectively.

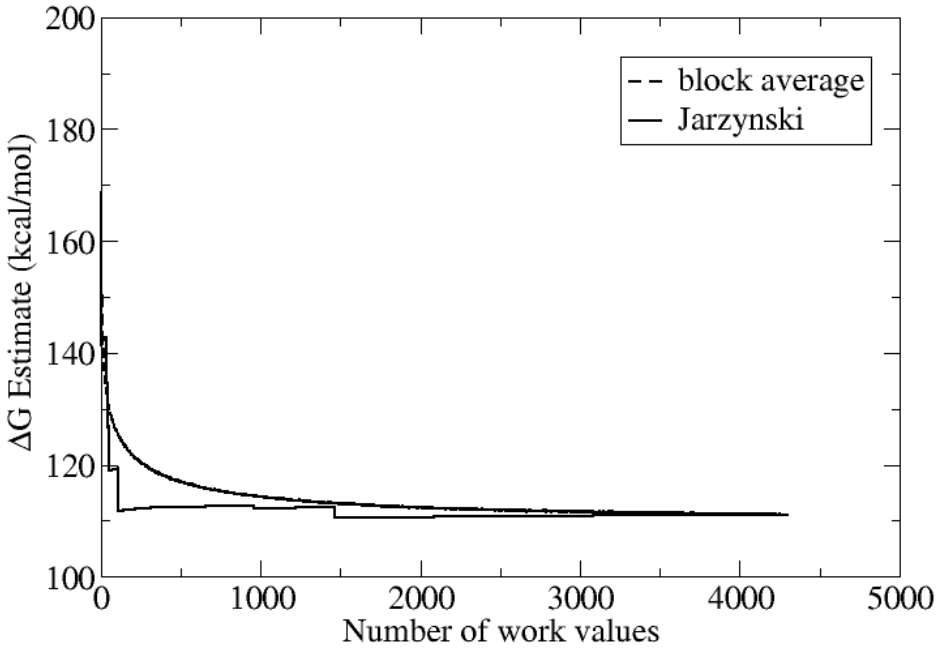

Figure 4.15 The running Jarzynski estimate and subsampled block averaged free energy estimate as a function of the number of trajectories used in the estimate. The method was introduced by Zuckerman's group. The binding free energy of Ca^{2+} at site III of holoCaM is shown as an example.

Table 4.5 Free energy change of Ca^{2+} from simulations using Zuckerman's approach and from experiment.

		holoCaM	CaM-Ng	CaM-CaMKII
ΔG (kcal/mol) Jarzynski's equality	Site III	-111.10	-61.66	-118.28
	Site IV	-143.74	-44.37	-123.89
ΔΔG (kcal/mol) Jarzynski's equality	Site III	-	+49.44	-7.18
	Site IV	-	+99.37	+19.85
ΔG (kcal/mol) CI extrapolation	Site III	-19.24	-19.77	-31.80
	Site IV	-17.59	-3.09	-28.38
ΔΔG (kcal/mol) CI extrapolation	Site III	-	-0.53	-12.56
	Site IV	-	+14.50	-10.77
ΔG (kcal/mol) experiment	Site III/IV	-	+2.5	-3.3

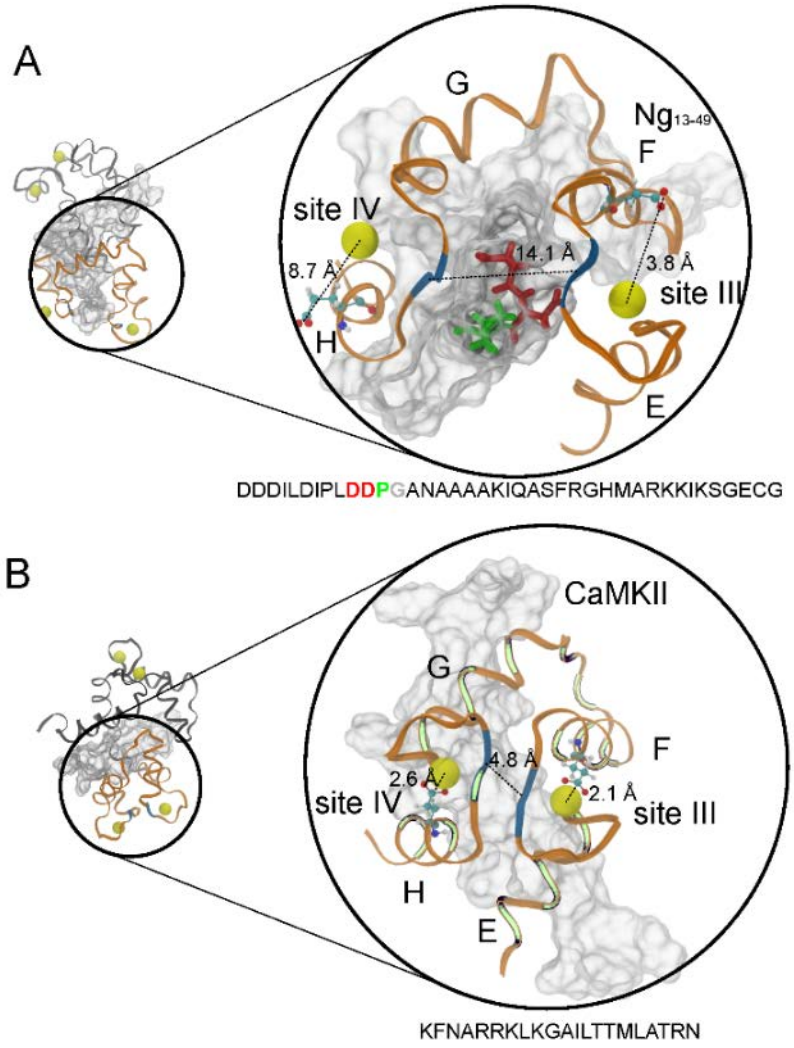

Figure 4.16 The structures of CaM-Ng $_{13-49}$ complex and CaM-CaMKII complex. The structure of CaM-Ng $_{13-49}$ complex is reconstructed from the coarse-grained simulations. This structure shows the largest decrease in Ca^{2+} affinity for site III and site IV. The structure of CaM-CaMKII complex is the crystal structure (PDB ID: 1CDM). Both of the two structures are equilibrated at 300 K after ionization and solvation. The cCaM is in ribbon representation and colored in orange. The residues that form β-sheet in the crystal structure, i.e. residues Y99, I100 from site III and residues Q135, V136 from site IV, are colored in blue. The Ca^{2+} ions are represented as yellow spheres. The Glu12 in each of the two Ca^{2+} binding loops are shown in ball-stick representation. The dotted lines show the distance between Ca^{2+} and the two oxygen donors from Glu12 and the distance between β-strands in cCaM. (A) The Ng $_{13-49}$ is in surface representation (white, transparent). The four residues that stick out from Ng $_{13-49}$ are D22(red) D23(red) P24(green) G25(gray) colored according to their polarity. The sequence of the target is provided below the figure. (B) The CaMKII (292-315) is in surface representation (white, transparent). The residues from CaM which form contacts with CaMKII peptides are colored in light green.

4.3.5 The molecular mechanism of CaM-CaMBT binding for Ca^{2+} release

I examined the reconstructed calcium-bound CaM-Ng $_{13-49}$ models from the coarse-grained structure that leads to a maximum increase of Ca^{2+} binding free energy. In contrast, I also investigated the conformation of the canonical bound complex of CaM-CaMKII.

In Fig. 4.16, Ng $_{13-49}$ bends from the middle at residues P24 and G25. On the one hand, the IQ motif interacts with helix F and G and F/G helix linker from cCaM to stabilize the CaM-Ng $_{13-49}$ complex. On the other hand, the acidic N-term interacts with both nCaM and cCaM. Strikingly, the four residues D22-D23-P24-G25(DDPG) prior to the IQ motif on Ng $_{13-49}$ stick out into the middle of the EF-hand β-scaffold between Ca^{2+} binding loops III and IV from cCaM, disrupting the formation of EF-hand β-scaffold. As a result, the distance between the two β-strands becomes 14.1 Å comparing to about 5.0 Å in intact holoCaM (PDB: 1CLL).

Moreover, the distance between the center of mass of residues DDPG and each of the two β-strands in site III and IV is 9.6 Å and 9.3 Å, respectively (Fig. 4.16A). As a reference, for Ca^{2+} in site III and site IV of the crystal structure of holoCaM without a target (PDB ID: 1CLL) is 2.3 Å and 2.5 Å, respectively. The Glu12 plays a crucial role in stabilizing Ca^{2+} in the binding loops (Grabarek, 2006), however, the presence of DDPG increases the average distances between each of the two oxygen donors from Glu12 and Ca^{2+} in the Ca^{2+} binding loop of CaM-Ng complex (Fig. 4.16A), in comparison to that of CaM-CaMKII complex (Fig. 4.16B), especially for site IV which is slightly closer to center of mass of DDPG than site III. The acidic region of Ng instead of stabilizing the CaM-Ng $_{13-49}$ complex, pries open the EF-hand β-scaffold and promotes the Ca^{2+} release from cCaM.

In comparison, I examined the importance of the canonical bound structure for the release of Ca^{2+} from the complex structure of CaM-CaMKII (PDB: 1CDM Fig. 4.16B). The average distances between the two oxygen donors in the Glu12 remain close to those in the crystal structure of holoCaM without a target (PDB ID: 1CLL). The affinity for calcium ions thus increases drastically.

On the one hand, binding of Ca^{2+} to CaM opens EF hands and exposing of the hydrophobic cores, which facilitates the canonical binding of CaMKII; because of the additional contacts between the helical CaMKII peptide and the Ca^{2+}-CaM that wraps around, binding of CaMKII enhances stability of the Ca^{2+} in bound form. On the other hand, binding of Ng to CaM disrupts the EF-hand β-scaffold in cCaM and destabilizes the Ca^{2+} bound state. This explains the molecular mechanism of "reciprocal relation" (Gaertner et al., 2004; Hoffman et al., 2014) between target binding and affinity for Ca^{2+} for a certain category of CaM binding targets.

4.4 Discussions

4.4.1 Energy landscape of apoCaM-Ng $_{13\text{-}49}$ binding is rugged

In this study, the molecular simulations using a coarse-grained model revealed a rugged free energy landscape of apoCaM-Ng $_{13\text{-}49}$ binding (Fig. 4.6). Conformations of apoCaM-Ng $_{13\text{-}49}$ can thermally fluctuate between the basins separated by 1~2 k_BT. Those transient complexes are commonly observed in protein-protein associations involving IDP(s) experimentally (Tang et al., 2006) (Sugase et al., 2007), as well as in computer simulations (Levy et al., 2004; Alsallaq and Zhou, 2007; Kim et al., 2008; Turjanski et al.,

2008; Higo et al., 2011; Ganguly et al., 2012). One of the advantages of the weakly-bound transient complexes is to enhance the on-rate by long range electrostatic interactions (Shoemaker et al., 2000). There are several factors that lead to the emergence of transient complexes. Generally, the non-specific interactions are thought to be a major contributor. Sugase et al. (Sugase et al., 2007) showed from NMR experiments that the transient complexes of pKID and KIX are stabilized predominantly by nonspecific hydrophobic contacts instead of electrostatic interactions. The importance of non-specific intermolecular interactions was further addressed by computation simulations on pKID and KIX by Turjanski et al. (Turjanski et al., 2008), as well as on NCBD domain of transcription co-activator CBP and the p160 steroid receptor co-activator ACTR by Chen group (Ganguly et al., 2012). In both of the cases, structural/topology based models (Clementi, 2008) were used for the intermolecular interactions. However, apoCaM-Ng $_{13-49}$ complex structure cannot be determined experimentally due to low stability (Ran et al., 2003). In the model, I incorporated statistical nonbonded interactions and electrostatic interactions that do not rely on a specific complex structure. Furthermore, for the Ng that is an intrinsically disordered peptide (IDP), I implemented a sequence-based statistical dihedral angle potential (Karanicolas and Brooks, 2002).

On the other hand, in the apoCaM-Ng $_{13-49}$ complex, Ng $_{13-49}$ remains partially structured as seen in the low probability of contact formation within Ng $_{13-49}$ in Fig. 4.8b. This kind of dynamic disorder of the target in complexes falls to the category of “fuzzy complex” generalized by Fuxreiter et al. (Tompa and Fuxreiter, 2008). In a fuzzy complex, one or both of the proteins could stay dynamic in structure. Recently, Chan’s group integrated both the non-native interactions and conformational entropy in the bound states

of the intrinsically disordered transcriptional activation domain of the Ewing's Sarcoma oncoprotein family (EAD) and its target protein (Song et al., 2013). They implemented a cation-π interaction potential to describe the "fuzzy interactions" for the molecular recognition between the target protein and an IDP. However, the target protein was modeled as a sphere of distributed charges on the surface; thus, no conformational change of the target was considered.

To generalize, the structural flexibility and nonspecific interactions, which correspond to entropic and enthalpic effects respectively, result in the rugged energy landscape that shows multiple transient complexes without a global minimum. In the crystal structure of apoCaM and IQ motif of Ng (NgIQ) (PDB ID: 4E50), the EF hands 3 and 4 reserve similar angles as in holoCaM (open conformation, see Table 4.1) and the EF hands 1 and 2 is similar as in apoCaM. Only about 1.3% of the structures keep similar EF hand angles, i.e. the crystal structure is not the most stable complex structure and may only be stabilized under certain very strict experimental conditions (for example, cCaM and NgIQ were strategically linked by 5-glycine flexible linker to be crystalized by Kumar et.al (Kumar et al., 2013)). Other transient complex structures shown here may be stabilized by certain experimental conditions as well and be used for prediction of CaM-Ng complex structure.

Nevertheless, it was postulated that the residual structure of an IDP is necessary for molecular recognition for CaM-CaMBT (Oldfield et al., 2005). From this study, I find that the residual structures in Ng is crucial for CaM recognition and binding. In the video from Appendix, I showed the contact formation map at each window of the d_{COM} from 20 σ to

0.2 σ (σ = 3.8 Å). I found that the IQ motif started interacting with cCaM at a distance of d_{COM} = 13 σ and retained, indicating its key role in binding with apoCaM. The Creamer group (Dunlap et al., 2013) as well as Clarke group (Rogers et al., 2014), however, indicated that the residual structure of the IDP is not required for binding on-rate, but the binding affinity was weakened upon helix-breaking mutation. It is important to note that in Creamer's study, most of the CaMBTs bind CaM in a canonical form that CaM undergoes large global conformational changes. A conformational and mutually induce fit mechanism (Wang et al., 2013) is needed for the recognition and binding. However, Ng mainly interacts with cCaM and does not requires global conformational changes of CaM. This raises an open question that whether the necessity of residual structure of intrinsically disordered CaMBTs is dependent on their binding mechanisms with CaM.

4.4.2 The affinity of CaM for Ca^{2+} in the presence of target peptides

The change in free energy of Ca^{2+} binding to cCaM computed from non-equilibrium atomistic pulling simulations using Jarzynski's equality reveals that the target peptides Ng $_{13-49}$ and CaMKII have opposite effects on Ca^{2+} binding affinity to cCaM. The binding and dissociation of Ca^{2+} is essentially dependent on the two EF-hand motifs in cCaM, which are connected by F/G helix linker and are known to be rigid in Ca^{2+}-bound CaM (Fig. 1.4). This construct of two-EF-hand motif is often referred as "EF-hand β-scaffold" (Grabarek, 2006). Grabarek et al. (Grabarek, 2006) proposed that the EF-hand β-scaffold immobilizes the Ca^{2+} ion during its initial binding to the Ca^{2+} binding site. At residue level, the competing interaction between the N-terminal acid region of Ng $_{13-49}$ and cCaM is responsible for destabilizing the EF-hand β-scaffold within cCaM (Fig. 4.8c,

circle). As a result, Ng $_{13-49}$ decreases the affinity of Ca^{2+} for cCaM significantly. On the contrary, the CaMKII peptide binds to CaM in a canonical configuration such that the two domains of CaM wraps around the rod-like CaMKII peptide and stabilizes the EF-hand β-scaffold. Therefore, CaMKII peptide strengthens the affinity of Ca^{2+} for cCaM.

From the analysis, the effects on Ca^{2+} binding affinity for cCaM is determined by attribute of the interactions between CaM and the target. It can be applied to explain the change in Ca^{2+} affinity in other complex of calmodulin and Ca^{2+}-dependent/free CaMBTs. According to the structural study of the complex of the Ca^{2+}-activated K^+ channel bound to Ca^{2+} loaded CaM by Adelman group (Schumacher et al., 2001), only nCaM is loaded with Ca^{2+}, while the cCaM EF-hands reserve similar closed structure with those from cCaM. In this complex, a severe disruption in the Ca^{2+} binding loops due to the repacking of the nCaM causes the Glu12 of the Ca^{2+} binding loop to move out of their Ca^{2+} binding positions, releasing the Ca^{2+}. Peptides that bind to CaM canonically were observed for increasing the Ca^{2+} affinity for the cCaM (Olwin et al., 1984), ΔΔG varying from -1.5 kcal/mol for phosphorylase kinase to -3.3 kcal/mol for CaMKII (Peersen et al., 1997) and -4.8 kcal/mol for β-calcineurin (O'Donnell et al., 2011). For Nav 1.1 and PEP-19 that contain an IQ motif, both of them bind mainly to cCaM and an acid region is present prior to the IQ motif in sequence. They lower Ca^{2+} binding affinity to CaM. For instance, binding of the sodium channel Nav1.1 causes an increase in ΔΔG by about +2 kcal/mol [12] and PEP-19 by about +0.2 kcal/mol (Theoharis et al., 2008; Wang et al., 2013). However, for Cav1.1 and Cav 1.2 which contain an IQ motif but without a prior acidic region, they behave like a canonical binding motif and stabilize the Ca^{2+} binding by ΔΔG = -1 kcal/mol and about -

2.6 kcal/mol for Cav 1.1 and Cav 1.2, respectively (Halling et al., 2009). The results support the active role of the acidic region prior to the IQ motif as shown in this study. However, it has not been investigated on the exact residues and the molecular mechanism of how the acidic region regulates Ca^{2+} binding to CaM. I used coarse-grained molecular simulations as well all-atomistic steering simulations and identified key residues DDPG in the acidic region of Ng $_{13\text{-}49}$ that "pries" open β-sheet structure between the two Ca^{2+} binding loops.

Note that although the calculated free energy matches the signs of the experimental values, they differ by an order of magnitude. There are several other computational studies on the mechanisms of Ca^{2+} releasing / binding in CaM using steered molecular dynamics (Lepsik and Field, 2007; Zhang et al., 2008). However, for the calculation of Ca^{2+} binding free energy, the computed values using combined quantum and molecular calculations were about two orders of magnitude away from the experimental values because of large uncertainties of the deformation by allowing backbones to sample wide conformations in the molecular dynamics simulation (Lepsik and Field, 2007). In this study, I fixed only the backbone atoms to maintain the overall shape of the proteins. Major repositions of the side-chains of the Ca^{2+} coordinating residues were observed, especially for the apoCaM-Ng $_{13\text{-}49}$ complex shown by large root mean square deviation (RMSD) of the loops during pulling shown in Fig. 4.17. This is a possible source of the large free energy uncertainties in my calculation.

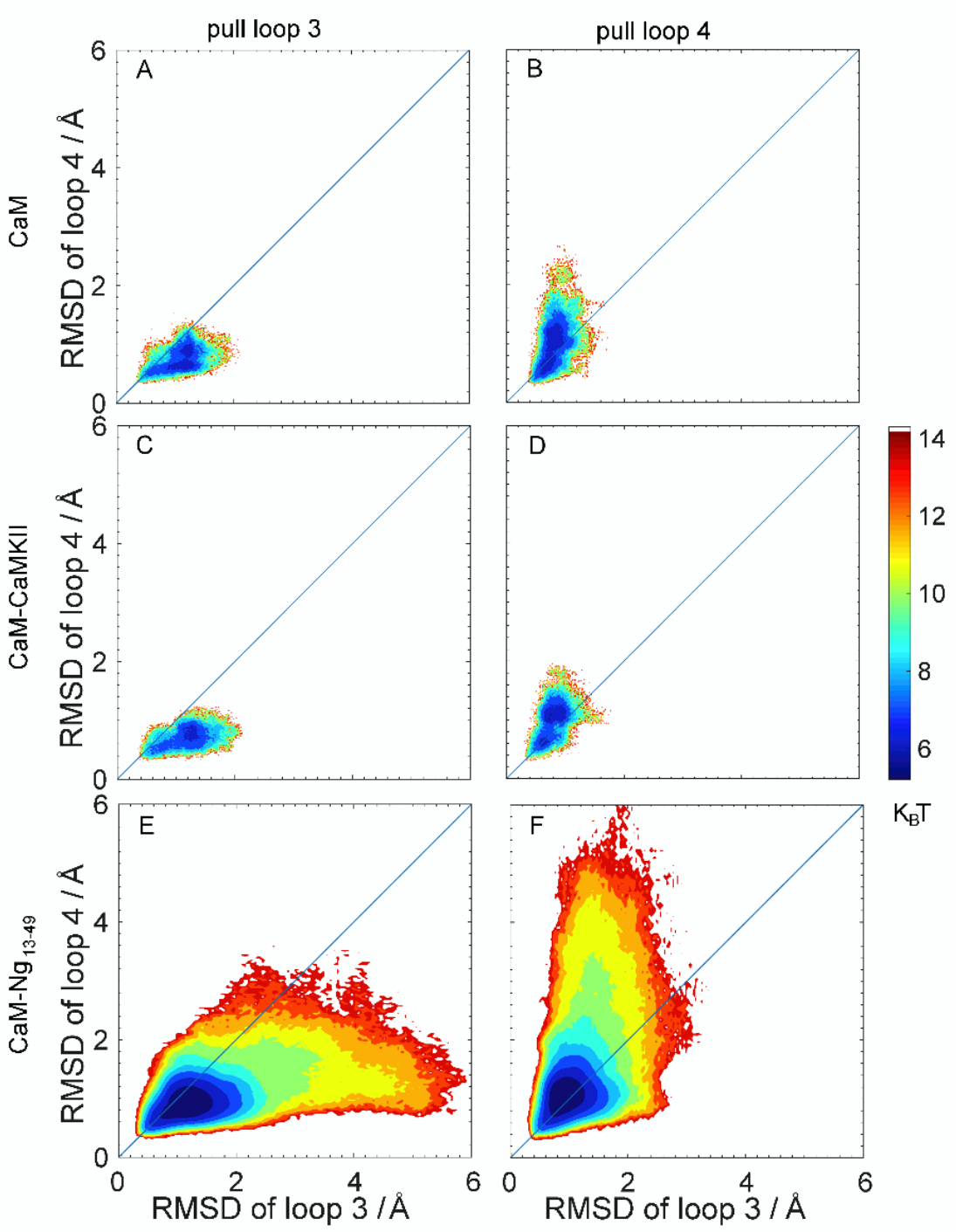

Figure 4.17 Free energy surface of root mean square deviations of the Ca^{2+} loop 3 and loop 4 in the pulling process. (A) and (B) show the free energy surface of pulling Ca^{2+} from holoCaM (PDB: 1CFD) loop 3 and loop 4, respectively. (C) and (D) for pulling Ca^{2+} from CaM-CaMKII complex (PDB: 1CDM). (E) and (F) for pulling Ca^{2+} from CaM-Ng complex.

The main factor that contributes to the over-estimation of the free energy change in our calculation came from the classical force-field that I used in the molecular dynamics simulations. It has been shown that the divalent Ca^{2+} ions do not maintain +2e charge in solution and in complex with proteins due to charge transfer and polarization. In quantum chemistry calculations based on an ensemble of holoCaM structures, Cheung's group showed that the charge for Ca^{2+} was approximately in the range of +1.2e to +1.6e (Wang et al., 2011). Since the bonded interactions between the coordinating oxygen atoms from the Ca^{2+} binding sites and the Ca^{2+} stabilizes the Calcium-bound state, this variation of change would most likely reduce the free energy gap, ΔG.

There are also studies from Jungwirth's group (Kohagen et al., 2014; Kohagen et al., 2014) that tried to incorporate the effect of polarization using a reduced charge of Ca^{2+} (ECCR) and also rescaling the partial charges on the relevant functional groups of the protein (ECCR-P). They found that for binding free energy of Ca^{2+} to CaM, there could be an overestimation of ΔG up to 5 times. By including the polarization effect, the estimation of ΔG could be in the same order of magnitude and hence yielding more accurate estimation of changes in Ca^{2+} affinity for CaM and CaM-CaMBT complexes, which need to be included in further study. Nevertheless, by using classical molecular dynamics simulations, we were able to capture that Ng_{13-49} greatly reduce affinity of Ca^{2+} for cCaM whereas CaMKII peptide increases the affinity.

Chapter 5

Conclusions

From my computational modelling and simulations of the formation of CaM and CaM target binding, I identify key rate-limiting steps that dictate. CaMKI has two hydrophobic motifs (1-5-10 and 1-14) while CaMKII has only one (1-5-10), and our findings have revealed how such seemingly subtle differences may induce a differential response as CaM undergoes structural adjustments to form a canonical functional complex. My contribution emphasizes how CaM's structural flexibility is essential to accommodate the unique conformational changes each of its target undergoes so as to achieve the target recognition and possibly selectivity. My study demonstrates the classical “lock and key” mechanism and the “induced fit” model, as well as “conformation selection”, are insufficient to explain CaM’s target recognition. Additionally, my analysis on CaM target binding provides two detailed mechanistic examples of how the final protein complex is formed through mutually-induced conformational changes between interacting partners.

CaM coordinates the activation of a family of Ca^{2+}-regulated proteins that are crucial for synaptic plasticity at the molecular and cellular level associated with learning and memory in neurons (Xia and Storm, 2005). Ng is a neural-specific postsynaptic IDP, for which Ca^{2+} decreases binding to CaM (Gerendasy et al., 1994). Due to the low binding affinity of apoCaM-Ng $_{13\text{-}49}$ interactions, experimentally it is challenging to determine the bound structures. My coarse-grained molecular simulations of CaM-Ng $_{13\text{-}49}$ binding guided by experimental measurements demonstrated the importance of conformational

flexibility of both the binding partners in various states of the bound complex. My study revealed that the acidic region of Ng $_{13-49}$ sticks out into the middle and "pries open" the EF-hand β-scaffold of cCaM. I speculate that this feature of destabilizing Ca^{2+}-CaM at low Ca^{2+} concentration could possibly explain CaM in apo-form in the resting cell. CaMKII on the other hand plays a pivotal role in learning and memory formation for both long-term potentiation and mechanisms for the modulation of synaptic efficacy (Miller and Kennedy, 1986). My study demonstrated that in an apoCaM-Ng $_{13-49}$ complex, CaM dominantly remains in an extended conformation. This is in contrast to a canonical bound complex where the two domains of CaM wrap around a rod-like CaMKII peptide that essentially stabilizes the Ca^{2+} binding loops of CaM. The embedded CaMKII peptide by two domains increases the binding affinity of Ca^{2+} to CaM. This proposed mechanism leads to a broader understanding of the reciprocal relation between Ca^{2+}/CaM binding and CaM/CaMBT binding (Hoffman et al., 2014). To the best of my knowledge, this is the first computational study providing a detailed description at atomistic level about how binding of CaM with two distinct targets (Ng and CaMKII) influences the release of Ca^{2+} from cCaM as a molecular underpinning of CaM-dependent Ca^{2+} signaling in neurons which has been investigated for decades (Putkey and Waxham, 1996; Waxham et al., 1998; Putkey et al., 2008; Hoffman et al., 2014). The molecular basis for learning and memory formation has aroused attention of the neuroscience community dating back to 1980's (Crick, 1984). I believe this study allows opportunities to connecting the molecular regulations in atomistic detail to the understanding of cellular process cascade of learning and memory formation.

Chapter 6

Appendices

6.1 Side-chain C_α model of CaM and CaMBT system

A side-chain C_α model (SCM) (Cheung et al., 2003) is used to study the association of CaM and CaMBTs at the individual amino acid level. In SCM, each amino acid (except glycine) is represented by two beads: one at the C_α position and the other at the center of mass of the side-chain. A detailed description of the CaM-CaMBT model is given in the published work (Wang et al., 2013).

The potential energy of CaM or CaMBT is given by,

$$E^{intra} = E_{bond} + E_{angle} + E_{dihedral} + E_{chirality} + E_{elec} + E_{HB} + E_{LJ} \qquad \text{Eqn. (A1)}$$

A detailed description of bond stretching term E_{bond}, bond-angle term E_{angle}, side-chain chirality term $E_{chirality}$ can be found in previous work (Homouz et al., 2008). The dihedral angle term $E_{dihedral}$ is described as,

$$E_{dihedral}^{ijkl} = \sum_{n=1,3} k_{\varphi}^{(n)}\left[1 - \cos\left(n \times \left(\varphi_{ijkl} - \varphi_{ijkl}^{0}\right)\right)\right] \qquad \text{Eqn. (A2)}$$

where φ_{ijkl} is the dihedral angle defined over four C_α beads of consecutive residues *i*, *j*, *k* and *l*; φ_{ijkl}^{0} is the corresponding dihedral angle measured in the referenced structure of CaM (PDB ID: 1CLL) and CaMBTs (PDB ID: 2L7L and 1CDM for CaM-CaMKI and CaM-CaMKII, respectively).

For CaM, $k_{\varphi}^{(1)}$ = 40.0 kcal/mol and $k_{\varphi}^{(3)}$ = 20.0 kcal/mol were used for parameterization (Wang et al., 2013) based on the experimental data of unbound CaM (Heidorn et al., 1989; Anthis et al., 2011). CaM retains in an extended, open conformation in solution state. Based on this parameterization the average radius of gyration (R_g) of CaM in an unbound state is ~ 21.1 Å, close to the experimental value 21.3 Å measured from the x-ray scattering experiments (Heidorn et al., 1989). Additionally, the R_g of CaM shows two peaks corresponding to the extended state and compact state at the ratio of ~ 9:1 in agreement with the paramagnetic relaxation enhancement experiment (for details of the parameterization please see the previous work in reference (Wang et al., 2013). For CaMBTs, a set of $k_{\varphi}^{(1)} = 1.2$ kcal/mol and $k_{\varphi}^{(3)} = 0.6$ kcal/mol which are typical values for the strength of dihedral angles in the side-chain C_α model (Cheung et al., 2003), render structural flexibility for these CaMBTs in the unbound form than CaM. The equilibration positions of bond, bond angle and pseudo dihedral angle energies were taken from the crystal structure of CaM (PDB ID: 1CLL). Thus, the Hamiltonian of CaM in the coarse-grained model did not depend on the structure of the bound complex, making it possible to study a variety of CaM-CaMBT interactions.

Debye-Hückel potential (Debye and Hückel, 1923) was used to describe the screening effect of ionic solutions on electrostatic interactions between charged units. The Debye-Hückel potential E_{elec} in Eqn. A1 between beads *i* and *j* follows the equation given by,

$$E_{elec}^{ij} = \frac{q_i q_j}{4\pi\varepsilon_r\varepsilon_0 r_{ij}} e^{-r_{ij}/\sqrt{\varepsilon_r\varepsilon_0 k_B T/2e^2 I}} \qquad \text{Eqn. (A3)}$$

q_i or q_j is the partial charges on bead *i* or *j*, which can be obtained using a combined method of quantum chemistry calculation, statistical physics, and coarse-grained molecular simulations (Wang et al., 2011). In the above equation r_{ij} is the distance between beads *i* and *j*, ε_0 is the permittivity of free space, ε_r is the relative dielectric constant (80 for aqueous solutions), k_B is the Boltzmann constant, *T* is the temperature, *e* is elementary charge and *I* is the ionic strength of the aqueous solution (that was set to 0.1M). The partial charges assigned to CaM and two CaMBTs are given in the previous work (Wang et al., 2013).

In Eqn. A1 the backbone-backbone interactions for the C_α beads are represented by the hydrogen-bonding interactions term E_{HB}, as described in the previous work (Homouz et al., 2008). The Lennard-Jones (LJ) potential E_{LJ} in Eqn. A1 was used to represent the sidechain-sidechain and backbone-sidechain interactions between beads *i* and *j*.

$$E_{LJ}^{ij} = \varepsilon_{ij}\left[\left(\frac{\rho_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\rho_{ij}}{r_{ij}}\right)^{6}\right] \qquad \text{Eqn. (A4)}$$

where r_{ij} is the separation between sidechain beads (or, between backbone and sidechain beads) *i* and *j* ($|i-j| \geq 2$). For sidechain-sidechain interactions the above equation is used, where $\rho_{ij} = f(\rho_i + \rho_j)$ and ρ_i and ρ_j are the van der Waals radii of the side chains *i* and *j*, calculated from the extended structure of CaM (PDB ID: 1CLL); f is the scaling factor to be 0.9 to avoid the clashes between bulky sidechains. The sidechain-sidechain interaction strength ε_{ij} in Eqn. A4 represent the solvent-mediated interaction defined by the Betancourt-Thirumalai (BT) statistical potential $\varepsilon_{ij}^{\mathrm{BT}}$ (Betancourt and Thirumalai, 1999). QW applied the BT statistical potential in the protein model to explore the

conformations beyond the experimentally determined one (Homouz et al., 2009; Wang et al., 2011). For the backbone-sidechain interactions, hard-core repulsion is considered by taking only the $\varepsilon_{ij}\left(\frac{\rho_{ij}}{r_{ij}}\right)^{12}$ term ρ_i for a backbone bead is $0.5\,\sigma$ (σ = 3.8 Å). For the backbone-sidechain interactions $\varepsilon_{ij} = \varepsilon = 0.6$ kcal/mol in Eqn. A4.

For intramolecular interaction strengths the BT potential $\varepsilon_{ij}^{\mathrm{BT}}$ is rescaled to $\varepsilon_{ij}^{\mathrm{intra}}$ due to the explicit inclusion of electrostatic interactions in the Hamiltonian in Eqn. A1, as follows,

$$-\varepsilon_{ij}^{intra} + \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{Q_iQ_j}{r_0} = -\varepsilon_{ij}^{BT} \qquad \text{Eqn. (A5)}$$

Q_i *(*Q_j*)* is the charge summation on the side chain of amino acid *i (j)* from the united atom force field. $r_0 = \rho_i + \rho_j$, where ρ_i and ρ_j are the van der Waals radii of amino acid *i* and *j*, respectively. The values for both $\varepsilon_{ij}^{\mathrm{BT}}$ and $\varepsilon_{ij}^{\mathrm{intra}}$ are given in the previous work (Wang et al., 2013).

The intermolecular interactions between CaM and a CaMBT are given by,

$$E^{inter} = E_{elec} + \lambda(E_{HB} + E_{LJ}) \qquad \text{Eqn. (A6)}$$

E_{elec} is a term for electrostatic interactions (same as Eqn. A3), E_{HB} is a term for backbone hydrogen bond interactions (described in the previous work (Homouz et al., 2008)), and E_{LJ} is a term for side-chain van der Waals interactions (same as Eqn. A4) between CaM and CaMBT. However, the strength λ of the intermolecular hydrogen bond

interactions (E_{HB} in Eqn. A6) and van der Waals interactions (E_{LJ} in Eqn. A6) between CaM and CaMBTs are twice as large as the strength of the intramolecular interactions in order to make the functional CaM-CaMBT complex stable in which the two domains of CaM wrap around the CaMBT. Based on these parameterization the computationally obtained generalized order parameter of the methyl symmetry axis (O^2_{axis}) for the residues in CaM-CaMKI complexes are positively correlated with the measurements published by the Wand group (Marlow et al., 2010), as stated in the supplement of previous work (Wang et al., 2013).

6.2 Structure-based and sequence-based dihedral angle potential

Among the bonded interactions in a protein model, the bond length potential and the bond angle potential are stiff to control the length of covalent bonds and bond geometry. In contrast, the dihedral angle potential has the energy constants that are comparable to the thermal agitation k_BT, so do the non-bonded interactions. Those interactions that can be altered by thermal fluctuations determine the structural transitions of the protein. The dihedral angle is defined as shown in Fig. A1. Dihedral angle potential plays an important role in the local protein structure.

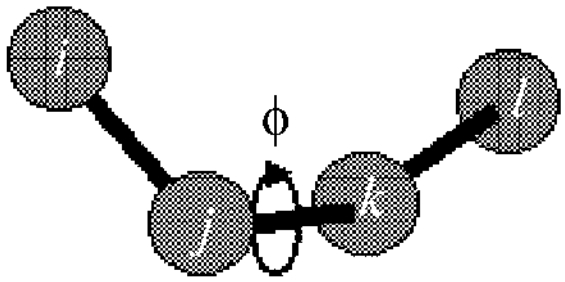

Figure A1 Illustration of a dihedral angle φ. φ is defined between four consecutive beads i, j, k, l. The angle is the planar angle between the plane determined by atoms i, j and k and the plane determined by atoms j, k and l.

In a structure-based or Go-like model, the pseudo-dihedral angle potential is based on a specific structure, which is usually the native structure. The most popular dihedral angle potential used takes the form of addition of two cosine terms, $E_{dihedral}^{ijkl} = \sum_{n=1,3} k_{\varphi}^{(n)}\left[1 - \cos(n \times (\varphi_{ijkl} - \varphi_{ijkl}^{0}))\right]$, where ijkl are four consecutive C_{α} beads. φ_{ijkl} is the dihedral angle, and φ_{ijkl}^{0} means the reference dihedral angle. $k_{\varphi}^{(n)}$ is the barrier height. Combination of two terms results 3 minima that are of different depths (Fig. A2, the global minimum occurs at $\varphi_{ijkl} - \varphi_{ijkl}^{0} = 0$).

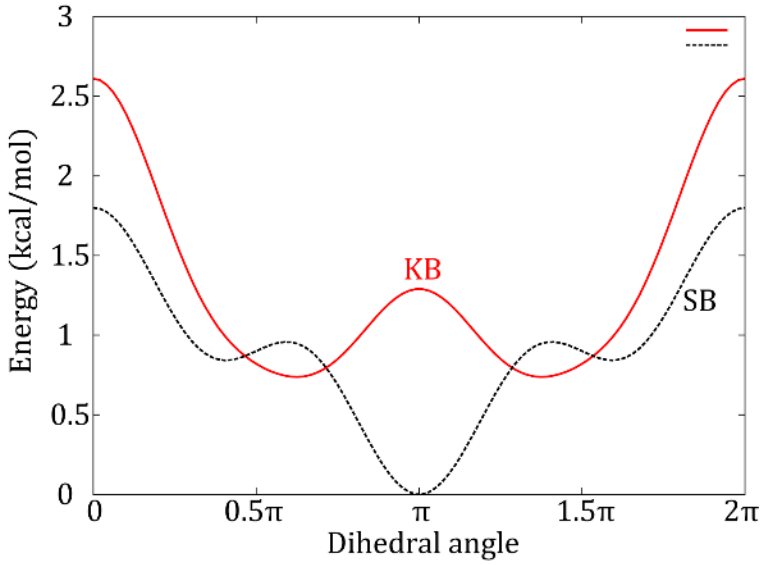

Figure A2 Illustration of a dihedral angle potentials comprised of 2 and 4 cosine terms. The black dotted line shows an example of dihedral angle potential comprised of 2 cosine terms that is widely used in molecular dynamics packages such as GROMACS and AMBER for Structure-Based (SB) models. The red solid line shows an example of dihedral angle potential used by Karanicolas and Brooks' study (KB). It is comprised of 4 cosine terms.

Sequence-dependent dihedral potentials are necessary because in some cases, a specific structure is not available for the protein. The statistical dihedral angle potential that I adopted was developed by Karanicolas and Brooks (Karanicolas and Brooks, 2002) (KB). The dihedral angle between four adjacent α-carbons depends on the backbone dihedral angles of the two middle residues. Brooks' group produced a probability distribution for 400 possible ordered pairs of amino acid residues from a survey of the Protein Data Bank and thus related the probability distribution to potential energy ignoring the entropy contribution. The dihedral angle potential presents two minima corresponding to local α-helical and β-strand geometries. The KB dihedral angle potential was first applied for the purpose of generating non-native intermediate structures of G and L proteins. The statistical potential is modeled as a 4-term cosine series, $E_{dihedral}^{ijkl} =$

$\varepsilon_{KB} \sum_{n=1}^{4} K_{jk,n}\left[1 + cos\left(n\varphi_{ijkl} - \delta_{jk,n}\right)\right]$, where φ_{ijkl} is the dihedral angle formed by four consecutive C_α beads *i, j, k, l* with beads *j* and *k* in the middle, $K_{jk,n}$ and $\delta_{jk,n}$ are statistically determined constants. ε_{KB} is a factor for adjusting the barrier of the dihedral angle potential in relative to other interactions in the current model.

6.3 Mapping coarse-grained simulation time to real time

Natural time unit of the Brownian dynamics simulation is mapped to the real time unit in the over-damped limit (Veitshans et al., 1997) by $\tau = 6\pi\eta a^3/(k_B T)$, where $k_B T$ = 0.66 * 4.184 kJ/mol, viscosity of aqueous solution $\eta = 10^{-3} Pa \cdot s$ and a is the size of the amino acids. In the side-chain C_α model (SCM), the van der Waals (vdW) radius for a C_α beads is 0.45σ and vdW radii of the side-chain beads for CaM vary from 0.39 σ to 1.18 σ. A typical value of the size of a bead is between 0.39 σ and 1.18 σ (σ is 3.8 Å). Therefore, τ is mapped to the real time between 0.01 and 0.37 ns.

6.4 Temperature replica exchange method

The energy landscape of a system could be very rugged and Replica Exchange Method (REM) (Sugita and Okamoto, 1999) is a computational method that enhances sampling (Fig. A3). It executes several replicas (copies) of standard molecular simulations at different temperatures simultaneously and each replica attempts to communicate with another replica every certain time step through a rule based on the Metropolis criterion (Daan and Berend, 2001): the acceptance probability (P_{acc}) for the exchange of two neighboring replica copies, *i* and *j*, is $P_{acc} = \min\{1, \exp[(\beta_i - \beta_j) * (U(r_i) - U(r_j)]\}$, where β=1/k_BT and *U(r)* represented the configurational energy of the system. The exchange ratio

is usually set between 0.2 and 0.3. The integration time step is $10^{-3}\,\tau_L$, where $\tau_L = \sqrt{ma^2/\varepsilon}$ (m is the mass of particle, a is the van der Waals' radius of the bead and ε is the solvent-mediated interaction energy). Thermodynamic properties and errors were calculated using the weighted histogram analysis method (Chodera et al., 2007).

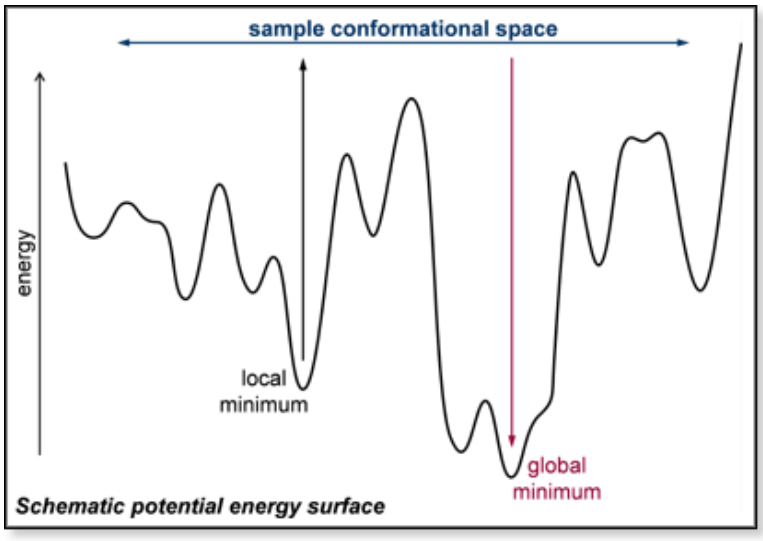

Figure A3 Illustration of how REMD enhances sampling. The schematic energy landscape demonstrates multiple local minima. Therefore, the sampling could be easily trapped in conformations that is not near the native structure. By exchanging with structures at higher temperatures would level the energy of the system and thus the probability of sampling other minima.

6.5 Umbrella sampling method

The umbrella sampling method (US) is a computation technique to enhance the sampling of transient states by introducing a bias potential (Torrie and Valleau, 1977). By introducing the biased potential V(Φ) as a function of the reaction coordinate Φ. The Hamiltonian of the system becomes $H = H_0 + V(\Phi)$, where H_0 is the unbiased Hamiltonian. As shown in Fig. A4, along the reaction coordinate, the bias potential would guarantee sufficient sampling of the intermediates. V(Φ) can be removed using the weighted histogram analysis method.

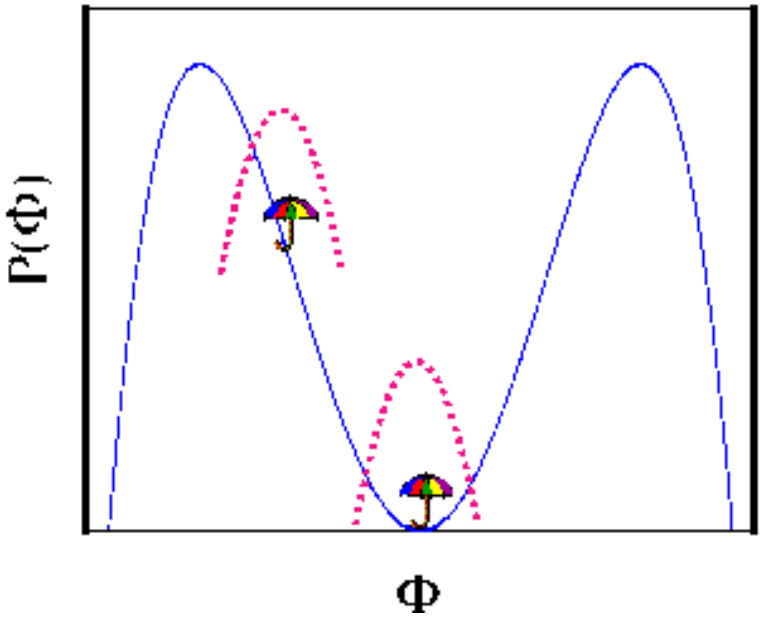

Figure A4 Illustration of setup of Umbrella Sampling. The blue lines are the probability distribution of a given system along reaction coordinate Φ. The red dotted lines represent the biased sampling after applying a harmonic potential at given windows of Φ.

6.6 Calculation of the mean-square fluctuation

The mean square fluctuation (MSF) describes the local flexibility of atoms of the protein over a period of time. For each atom *j*, MSF_j in each trajectory is calculated using the equation

$$MSF_j = \langle (\vec{r}_j - \langle \vec{r}_j \rangle)^2 \rangle \qquad \text{Eqn. (A7)}$$

where $\vec{r}_j$ is the coordinates for the atom *j* and $\langle \cdots \rangle$ denotes time average. The MSF_i for residue *i* is the calculated from C_α atom in the residue. I further averaged MSF_i over all of the successful trajectories to compute $\overline{MSF}$.

6.7 Shape analysis

The shape of configurations can be characterized by two rotationally invariant quantities, the asphericity Δ and the shape parameter S (Dima and Thirumalai, 2004). S and Δ are determined from the inertia tensor, $T_{\alpha\beta}$

$$T_{\alpha\beta} = \frac{1}{2N^2}\sum_{i,j=1}^{N}(r_{i\alpha} - r_{j\alpha})(r_{i\beta} - r_{j\beta}) \qquad \text{Eqn. (A8)}$$

where N is the total number of atoms and α, β represent the X, Y, Z components. The eigenvalues of T are denoted by λ_i (i = 1...N). The $\overline{\lambda}$ is the average of λ_i.

$$\Delta = \frac{3}{2}\frac{\sum_{i=1}^{3}(\lambda_i - \overline{\lambda})^2}{(trT)^2} \qquad \text{Eqn. (A9)}$$

$$S = 27\frac{\prod_{i=1}^{3}(\lambda_i - \overline{\lambda})}{(trT)^3} \qquad \text{Eqn. (A10)}$$

Asphericity (Δ) ranges from 0 to 1, and Δ=0 corresponds to a sphere. Deviation of Δ from 0 indicates the extent of anisotropy. The shape parameter (S) ranges from -0.25 to 2. A positive value of S corresponds to the oblate shape; a negative value of S corresponds to the prolate shape and S = 0 corresponds to a sphere.

6.8 Calculation of φ value

I calculated φ-value from simulation using the free energy perturbation (FEP) method as introduced in this study (Nymeyer et al., 2000). Mimicking the mutation of a specific residue by weakening the interaction between this residue and other residues to a certain percentage (1 - λ) and calculate the free energy change.

$$\Delta G_{mut-wt} = -k_B T \ln < e^{-\beta\lambda|\varepsilon|} >_{wt} \qquad \text{Eqn. (A11)}$$

where $|\varepsilon|$ is the strength of a specific pair, all the average are over the wild type (wt) ensemble. The φ-value for a specific pair is then,

$$\phi = \frac{\Delta G^T_{mut-wt} - \Delta G^U_{mut-wt}}{\Delta G^F_{mut-wt} - \Delta G^U_{mut-wt}} = \frac{\ln(< e^{-\beta\lambda|\varepsilon|} >_{T,wt} / < e^{-\beta\lambda|\varepsilon|} >_{U,wt})}{\ln(< e^{-\beta\lambda|\varepsilon|} >_{F,wt} / < e^{-\beta\lambda|\varepsilon|} >_{U,wt})} \qquad \text{Eqn. (A12)}$$

where T, U and F refer to the Transition state, Unfolded state and Folded state. Assuming λ<< 1, $\Delta G_{mut-wt} = \lambda \left\langle |\varepsilon| \right\rangle$

$$\phi = \frac{\lambda < |\varepsilon| >_{T,wt} - \lambda < |\varepsilon| >_{U,wt}}{\lambda < |\varepsilon| >_{F,wt} - \lambda < |\varepsilon| >_{U,wt}} = \frac{n_T / N_T |\varepsilon| - n_U / N_U |\varepsilon|}{n_F / N_F |\varepsilon| - n_U / N_U |\varepsilon|} = \frac{P_T - P_U}{P_F - P_U} \qquad \text{Eqn. (A13)}$$

where n stands for number of frames in the corresponding ensemble that the specific pair contact is formed, N stands for total number of frames in the ensemble.

If the perturbation is not small, to calculate φ value,

$$\phi = \frac{\ln\left\{\frac{P_T e^{-\beta\lambda|\varepsilon|} + (1 - P_T)}{P_U e^{-\beta\lambda|\varepsilon|} + (1 - P_U)}\right\}}{\ln\left\{\frac{P_F e^{-\beta\lambda|\varepsilon|} + (1 - P_F)}{P_U e^{-\beta\lambda|\varepsilon|} + (1 - P_U)}\right\}} \qquad \text{Eqn. (A14)}$$

Using bond weakening coefficient λ= 0.1, 0.3, 0.6, and 0.9 and assuming each contact has the same energy |ε| = 0.6. The bond φ values are barely dependent on λ.

6.9 Calculation of model free order parameter S^2

A number of N (~ 500) representative coarse-grained structures were selected from the free energy surface of unbound Ng protein using importance sampling method. The model-free NMR order parameter S_i^2 (Lipari and Szabo, 1982) for the backbone N-H bond of residue *i* is calculated on the selected structures by relating to the Root Mean Square Fluctuations (RMSF) of Cα bead through the following relation (Berjanskii and Wishart, 2006),

$$S_i^2 = 1 - 0.5 \ln(1 + \frac{RMSF_i}{23.6\text{Å}} * 10.0) \quad \text{Eqn. (A15)}$$

$$RMSF_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(\boldsymbol{r}_i^j - \frac{1}{N}\sum_{k=1}^{N}\boldsymbol{r}_i^k)^2} \quad \text{Eqn. (A16)}$$

where $\boldsymbol{r}_i^j$ is position vector of Cα bead of residue *i* at frame *j* and *N* is total number of frames after alignment.

6.10 Calculation of "apparent chemical shifts"

In the NMR experimental work by Hoffman et al. (Hoffman et al., 2014), a semi-quantitative comparison of amide chemical shifts for residues in apoCaM that are affected by binding Ng $_{13\text{-}49}$ were presented. A threshold was applied to generate a set of Boolean numbers, with 1 representing significant change in chemical shifts and 0 the other way around. I call the Boolean series "apparent chemical shifts" in this computational study. In the NMR spectroscopy, the chemical shift is the relative resonant frequency in the local induced magnetic field and is diagnostic of the molecular structure. Berjanskii and Wishart

(Berjanskii and Wishart, 2006) showed that the inverse absolute chemical shifts roughly correlates with RMSF, which reflects the flexibility of the structure. Therefore, to compare with change of chemical shifts from unbound to bound apoCaM in the experiment, I approximated the amide chemical shifts by the inverse of RMSF of C_α beads and defined "apparent chemical shifts" as $\Delta = \Theta$ ($RMSF^{unbound}$ / $RMSF^{bound}$ - τ). τ = 50% was used. Δ measures the change of local flexibility of the apoCaM: larger values of δ indicate stabilization of the corresponding residue upon binding with Ng_{13-49}.

6.11 Definition of helicity

The fraction of helical structure, or helicity, of Ng (G25-A42) was calculated using the following formula (Kudlay et al., 2009),

$$f_H = \frac{1}{N-3}\sum_{i=1}^{N-3} < \Theta(\Delta\phi - |\phi_i - \phi_0|) > \qquad \text{Eqn. (A17)}$$

$\Theta(x)$ is the Heaviside step function taking value 1 if $x \geq 0$ and value 0 otherwise. N is the total number of residues, i is the residue index, ϕ_i is the dihedral angle about the residues $i \sim i+3$ from the simulation, $\phi_0 = 49.50°$ is the dihedral angle in a perfect alpha helix and $\Delta\phi = 12.07°$. <•••> denotes ensemble average.

6.12 Definition of Z

I defined an order parameter Z to calculate the total number of intermolecular contacts between apoCaM and Ng_{13-49}. For each residue *i* from apoCaM and residue *j* from Ng_{13-49}, the sidechain-sidechain and backbone-backbone contacts (Z_{ij}^{ss} and Z_{ij}^{bb}, respectively) are determined as

$$Z_{ij}^{ss} = \Theta\left(c - \frac{d_{ij}^{ss}}{\rho_i + \rho_j}\right) \quad \text{Eqn. (A18)}$$

$$Z_{ij}^{bb} = \Theta(c - \frac{d_{ij}^{bb}}{\rho_{HB}}) \quad \text{Eqn. (A19)}$$

d_{ij}^{ss} (d_{ij}^{bb}) is the distance between side-chain (backbone) beads of residue *i* of apoCaM and residue *j* of Ng $_{13\text{-}49}$ in simulation, ρ_i (ρ_j) is van der Waals radius of residue *i(j)*, $\rho_{HB} =$ 4.66 Å is the typical length of a hydrogen bond and cutoff c = 1.3. The total number of contacts Z is the summation of backbone-backbone and sidechain-sidechain contacts over all residues $Z = \sum_{ij}(Z_{ij}^{ss} + Z_{ij}^{bb})$.

6.13 Importance sampling

The structures from US simulations are biased and are not appropriate for structural analyses. I therefore used the Boltzmann reweighting method to sample an ensemble of unbiased structures. The probability of selecting a configuration *x* is

$$w(x) = \begin{cases} 1 & if\ p(x) \geq \rho \\ 0 & else \end{cases} \quad \text{where } p(x) = e^{-G(d_{COM}(x))/k_BT} \quad \text{Eqn. (A20)}$$

$p(x)$ is the probability of the configuration *x* in reweighted ensemble; $G(d_{\mathrm{COM}})$ is the reweighted/unbiased free energy obtained from the WHAM analyses along reaction coordinate d_{COM} (G_{min} is scaled to 0); ρ is a random number generated between 0 and 1; k_B is Boltzmann constant and T is the temperature. The ensemble generated after Boltzmann reweighting thus obeys the canonical distribution and is employed for subsequent analyses. Each structure from the biased ensemble was challenged by acceptance probability $w(x)$. This sample achieves a distribution of P^{sample} (d_{COM}). In order to assess the quality of the sampling, I computed the surprisal value compared with the

original unbiased distribution $P^{\mathrm{ori}}(d_{\mathrm{COM}})$ defined by the following formula (Shoemaker et al., 2000),

$$\text{surprisal} = \sum_r -P^{\mathrm{ori}}(r) \ln\left[P^{\mathrm{sample}}(r)/P^{\mathrm{ori}}(r)\right] \qquad \text{Eqn. (A21)}$$

where the summation is over all the order parameter r (in this case d_{COM}). A low surprisal value ensures that the sampled ensemble can well represent the original distribution.

6.14 Self-organized clustering algorithm

I applied the self-organized neural-net clustering algorithm (Carpenter and Grossberg, 1987) to determine the structures of the CaM-Ng complexes from the umbrella sampling simulations. The detail of the clustering algorithm is described in the study (Guo and Thirumalai, 1997; Grabarek, 2006; Cheung and Thirumalai, 2007; Wang et al., 2013). In this clustering method a vector with *M* elements represents each conformation *j*,

$$x_j = [x_{1j},\ x_{2j},\ \dots,\ x_{Mj}],$$

where $j = 1, 2, \dots N$; and *N* is the number of conformations selected for clustering analysis. The elements x_{ij} $(i = 1, 2, \dots, M)$ are the Euclidean distances between side chains of the polypeptide chain in the conformation *j*. To partition the various conformations into distinct clusters, the clusters are described by the cluster center and the size of the cluster is determined by a radius R_c. A given conformation is assigned to cluster *k* if the distance between the vector x_j and the center of the k^{th} cluster,

$$C_k = \frac{1}{M_k}\sum_{j=1}^{M_k} x_j, \qquad \text{Eqn. (A22)}$$

where $C_k = [C_{1k}, C_{2k}, \ldots, C_{Mk}]$ and M_k is the total number of conformations in C_k. Conformation j belongs to C_k if the Eucledian distance between conformation j and the cluster center k, $d_{jk} = \left|x_j - C_k\right| < R_c$, where R_c is a preassigned value.

6.15 Tables from the coarse-grained simulations

Table A1 Charge distribution on a coarse-grained side-chain C_α model of holoCaM. The unit of charge is $1.6*10^{-19}$ C.

Residue index	Residue name	Charge on C_α	Error on the C_α charge	Charge on side chain	Error on the side chain charge
1	ALA	-0.087	0.001	0.044	0.001
2	ASP	-0.032	0.002	-0.935	0.005
3	GLN	-0.053	0.002	0.044	0.001
4	LEU	-0.061	0.002	0.077	0.001
5	THR	-0.109	0.003	0.087	0.001
6	GLU	-0.063	0.002	-0.872	0.005
7	GLU	-0.075	0.002	-0.856	0.005
8	GLN	-0.052	0.003	0.055	0.001
9	ILE	-0.115	0.003	0.110	0.001
10	ALA	-0.044	0.003	0.049	0.001
11	GLU	-0.061	0.003	-0.833	0.005
12	PHE	-0.056	0.003	0.049	0.001
13	LYS	-0.078	0.003	0.998	0.006
14	GLU	-0.062	0.003	-0.839	0.005
15	ALA	-0.030	0.003	0.038	0.001
16	PHE	-0.064	0.003	0.051	0.001
17	SER	-0.089	0.003	0.074	0.001
18	LEU	-0.053	0.003	0.078	0.001
19	PHE	-0.060	0.003	0.041	0.001
20	ASP	-0.012	0.004	-0.767	0.009
21	LYS	-0.081	0.003	1.000	0.006
22	ASP	-0.028	0.003	-0.740	0.008
23	GLY	-0.184	0.002	0.198	0.001
24	ASP	-0.044	0.002	-0.680	0.007
25	GLY	-0.172	0.002	0.196	0.001
26	THR	-0.063	0.003	0.082	0.002
27	ILE	-0.073	0.003	0.097	0.001
28	THR	-0.117	0.003	0.098	0.001
29	THR	-0.106	0.003	0.091	0.001
30	LYS	-0.074	0.003	1.006	0.006
31	GLU	-0.076	0.003	-0.623	0.006
32	LEU	-0.070	0.004	0.080	0.001

33	GLY	-0.205	0.003	0.193	0.001
34	THR	-0.107	0.003	0.104	0.001
35	VAL	-0.106	0.003	0.114	0.001
36	MET	-0.070	0.003	0.062	0.001
37	ARG	-0.057	0.003	0.991	0.006
38	SER	-0.106	0.003	0.074	0.001
39	LEU	-0.070	0.003	0.078	0.001
40	GLY	-0.204	0.003	0.200	0.002
41	GLN	-0.046	0.002	0.064	0.001
42	ASN	-0.048	0.004	0.007	0.001
43	PRO	-0.266	0.004	0.283	0.002
44	THR	-0.110	0.003	0.092	0.001
45	GLU	-0.063	0.002	-0.824	0.005
46	ALA	-0.056	0.003	0.051	0.001
47	GLU	-0.082	0.003	-0.867	0.005
48	LEU	-0.041	0.003	0.078	0.001
49	GLN	-0.083	0.003	0.052	0.001
50	ASP	-0.045	0.003	-0.933	0.005
51	MET	-0.037	0.003	0.063	0.001
52	ILE	-0.094	0.003	0.111	0.001
53	ASN	-0.060	0.003	-0.004	0.001
54	GLU	-0.061	0.003	-0.826	0.005
55	VAL	-0.089	0.003	0.105	0.001
56	ASP	-0.041	0.004	-0.716	0.008
57	ALA	-0.042	0.002	0.051	0.001
58	ASP	-0.034	0.002	-0.720	0.007
59	GLY	-0.186	0.002	0.199	0.001
60	ASN	-0.033	0.002	0.008	0.002
61	GLY	-0.183	0.003	0.198	0.001
62	THR	-0.090	0.003	0.085	0.002
63	ILE	-0.053	0.003	0.098	0.001
64	ASP	-0.031	0.003	-0.846	0.008
65	PHE	-0.051	0.004	0.042	0.001
66	PRO	-0.279	0.004	0.287	0.002
67	GLU	-0.056	0.003	-0.593	0.006
68	PHE	-0.055	0.004	0.045	0.001
69	LEU	-0.055	0.003	0.079	0.001
70	THR	-0.115	0.003	0.100	0.001

71	MET	-0.073	0.003	0.064	0.001
72	MET	-0.059	0.004	0.071	0.001
73	ALA	-0.059	0.003	0.052	0.001
74	ARG	-0.071	0.003	0.989	0.006
75	LYS	-0.084	0.003	0.995	0.006
76	MET	-0.066	0.003	0.069	0.001
77	LYS	-0.054	0.003	0.973	0.005
78	ASP	-0.036	0.003	-0.879	0.005
79	THR	-0.137	0.003	0.111	0.001
80	ASP	-0.026	0.003	-0.903	0.005
81	SER	-0.105	0.003	0.072	0.001
82	GLU	-0.065	0.003	-0.825	0.005
83	GLU	-0.074	0.002	-0.846	0.005
84	GLU	-0.073	0.003	-0.819	0.005
85	ILE	-0.100	0.003	0.107	0.001
86	ARG	-0.073	0.003	0.933	0.005
87	GLU	-0.068	0.003	-0.840	0.005
88	ALA	-0.033	0.003	0.040	0.001
89	PHE	-0.051	0.003	0.048	0.001
90	ARG	-0.078	0.003	0.986	0.006
91	VAL	-0.093	0.002	0.112	0.001
92	PHE	-0.039	0.003	0.041	0.001
93	ASP	-0.015	0.003	-0.723	0.008
94	LYS	-0.062	0.003	1.006	0.006
95	ASP	-0.031	0.003	-0.677	0.007
96	GLY	-0.172	0.002	0.199	0.001
97	ASN	-0.042	0.002	0.033	0.003
98	GLY	-0.175	0.002	0.197	0.001
99	TYR	-0.005	0.002	0.033	0.001
100	ILE	-0.067	0.003	0.098	0.001
101	SER	-0.110	0.003	0.066	0.002
102	ALA	-0.054	0.003	0.049	0.001
103	ALA	-0.039	0.002	0.053	0.001
104	GLU	-0.056	0.003	-0.572	0.006
105	LEU	-0.059	0.003	0.070	0.001
106	ARG	-0.085	0.003	0.951	0.005
107	HIS	-0.038	0.003	0.057	0.001
108	VAL	-0.093	0.004	0.117	0.001

109	MET	-0.083	0.003	0.066	0.001
110	THR	-0.111	0.003	0.112	0.002
111	ASN	-0.031	0.003	0.007	0.001
112	LEU	-0.074	0.004	0.079	0.001
113	GLY	-0.213	0.003	0.195	0.001
114	GLU	-0.053	0.003	-0.845	0.005
115	LYS	-0.084	0.002	1.022	0.006
116	LEU	-0.060	0.002	0.069	0.001
117	THR	-0.110	0.003	0.093	0.001
118	ASP	-0.011	0.003	-0.885	0.005
119	GLU	-0.079	0.003	-0.879	0.005
120	GLU	-0.077	0.003	-0.847	0.005
121	VAL	-0.096	0.003	0.109	0.001
122	ASP	-0.026	0.003	-0.903	0.005
123	GLU	-0.082	0.003	-0.840	0.005
124	MET	-0.047	0.003	0.059	0.001
125	ILE	-0.093	0.003	0.104	0.001
126	ARG	-0.091	0.003	0.954	0.006
127	GLU	-0.066	0.003	-0.827	0.005
128	ALA	-0.042	0.003	0.040	0.001
129	ASP	-0.028	0.004	-0.734	0.008
130	ILE	-0.097	0.003	0.106	0.001
131	ASP	-0.040	0.003	-0.748	0.009
132	GLY	-0.198	0.002	0.199	0.001
133	ASP	-0.038	0.002	-0.724	0.008
134	GLY	-0.190	0.003	0.197	0.002
135	GLN	-0.022	0.003	0.040	0.001
136	VAL	-0.078	0.003	0.102	0.001
137	ASN	-0.017	0.003	-0.007	0.001
138	TYR	-0.061	0.003	0.042	0.001
139	GLU	-0.078	0.003	-0.875	0.005
140	GLU	-0.058	0.003	-0.633	0.007
141	PHE	-0.065	0.003	0.049	0.004
142	VAL	-0.107	0.003	0.121	0.001
143	GLN	-0.058	0.003	0.056	0.001
144	MET	-0.062	0.003	0.063	0.001
145	MET	-0.063	0.003	0.054	0.004
146	THR	-0.123	0.003	0.092	0.002

147	ALA	-0.013	0.003	0.041	0.001
148	LYS	-0.039	0.002	0.984	0.006
149	*CAL	1.037	0.007	N/A	N/A
150	CAL	1.033	0.007	N/A	N/A
151	CAL	1.049	0.008	N/A	N/A
152	CAL	1.031	0.007	N/A	N/A

* CAL stands for Ca^{2+} ions.

Table A2 Charge distribution on a coarse-grained side-chain C_α model of CaMKI peptide. The unit of charge is $1.6*10^{-19}$ C.

Residue index	Residue name	Charge on C_α	Error on the C_α charge	Charge on side chain	Error on the side chain charge
299	ALA	-0.045	0.002	0.037	0.002
300	LYS	-0.061	0.003	0.959	0.009
301	SER	-0.095	0.003	0.060	0.001
302	LYS	-0.066	0.003	0.959	0.009
303	TRP	-0.057	0.002	0.061	0.001
304	LYS	-0.059	0.003	0.958	0.009
305	GLN	-0.074	0.003	0.059	0.001
306	ALA	-0.028	0.003	0.042	0.001
307	PHE	-0.041	0.003	0.047	0.001
308	ASN	-0.030	0.003	0.002	0.001
309	ALA	-0.030	0.003	0.043	0.001
310	THR	-0.109	0.003	0.089	0.002
311	ALA	-0.038	0.002	0.045	0.001
312	VAL	-0.110	0.003	0.112	0.001
313	VAL	-0.095	0.003	0.107	0.001
314	ARG	-0.076	0.002	0.951	0.009
315	HIS	-0.040	0.003	0.041	0.001
316	MET	-0.051	0.003	0.058	0.001
317	ARG	-0.083	0.002	0.949	0.009
318	LYS	-0.071	0.002	0.972	0.009
319	LEU	-0.076	0.002	0.075	0.001
320	GLN	0.280	0.003	-0.255	0.003

Table A3 Charge distribution on a coarse-grained side chain C_α model of CaMKII peptide. The unit of charge is $1.6*10^{-19}$ C.

Residue Index	Residue Name	Charge on C_α	Error on the C_α charge	Charge on side chain	Error on the side chain charge
292	LYS	0.775	0.005	1.023	0.006
293	PHE	-0.020	0.003	0.053	0.001
294	ASN	-0.025	0.003	0.004	0.001
295	ALA	-0.032	0.003	0.046	0.001
296	ARG	-0.060	0.003	0.962	0.006
297	ARG	-0.072	0.003	0.961	0.006
298	LYS	-0.063	0.003	0.969	0.006
299	LEU	-0.064	0.003	0.070	0.001
300	LYS	-0.083	0.003	0.970	0.006
301	GLY	-0.190	0.003	0.194	0.001
302	ALA	-0.041	0.004	0.046	0.001
303	ILE	-0.097	0.003	0.107	0.001
304	LEU	-0.065	0.003	0.076	0.001
305	THR	-0.120	0.004	0.097	0.002
306	THR	-0.120	0.004	0.091	0.002
307	MET	-0.049	0.003	0.061	0.001
308	LEU	-0.069	0.003	0.074	0.001
309	ALA	-0.043	0.003	0.042	0.001
310	THR	-0.113	0.003	0.087	0.002
311	ARG	-0.077	0.003	0.958	0.006
312	ASN	-0.872	0.005	-0.028	0.001

Table A4 Charge distribution on a coarse-grained side-chain C_α model of full length Ng from rat. The unit of charge is $1.6*10^{-19}$ C.

Residue index	Residue name	Charge on Cα	Error on the Cα charge	Charge on side chain	Error on the side chain charge
1	MET	0.847	0.002	0.091	0.001
2	ASP	-0.028	0.002	-0.898	0.001
3	CYS	-0.037	0.002	0.035	0.001
4	CYS	-0.037	0.002	0.033	0.001
5	THR	-0.098	0.002	0.083	0.001
6	GLU	-0.052	0.002	-0.872	0.002
7	SER	-0.098	0.002	0.066	0.001
8	ALA	-0.022	0.002	0.049	0.001
9	CYS	-0.042	0.002	0.025	0.001
10	SER	-0.077	0.002	0.062	0.001
11	LYS	-0.061	0.003	1.000	0.001
12	PRO	-0.307	0.003	0.316	0.002
13	ASP	-0.010	0.002	-0.926	0.002
14	ASP	-0.037	0.002	-0.927	0.002
15	ASP	-0.040	0.002	-0.930	0.002
16	ILE	-0.070	0.002	0.078	0.001
17	LEU	-0.043	0.002	0.061	0.001
18	ASP	-0.053	0.002	-0.925	0.002
19	ILE	-0.073	0.003	0.080	0.001
20	PRO	-0.301	0.003	0.305	0.002
21	LEU	-0.066	0.002	0.063	0.001
22	ASP	-0.034	0.002	-0.930	0.002
23	ASP	-0.035	0.003	-0.929	0.002
24	PRO	-0.335	0.003	0.320	0.002
25	GLY	-0.182	0.002	0.188	0.001
26	ALA	-0.041	0.002	0.063	0.001
27	ASN	-0.047	0.003	0.011	0.001
28	ALA	-0.043	0.002	0.067	0.001
29	ALA	-0.064	0.002	0.064	0.001
30	ALA	-0.068	0.002	0.068	0.001
31	ALA	-0.057	0.002	0.069	0.001
32	LYS	-0.079	0.002	1.020	0.002
33	ILE	-0.089	0.002	0.103	0.001
34	GLN	-0.077	0.002	0.069	0.001
35	ALA	-0.046	0.002	0.064	0.001
36	SER	-0.091	0.002	0.079	0.001
37	PHE	-0.052	0.002	0.062	0.001
38	ARG	-0.079	0.002	1.027	0.002
39	GLY	-0.181	0.002	0.192	0.001

40	HIS	-0.049	0.002	0.049	0.001
41	MET	-0.069	0.002	0.069	0.001
42	ALA	-0.062	0.002	0.068	0.001
43	ARG	-0.062	0.002	1.020	0.002
44	LYS	-0.074	0.002	1.031	0.002
45	LYS	-0.077	0.002	1.030	0.002
46	ILE	-0.093	0.002	0.110	0.001
47	LYS	-0.092	0.002	1.028	0.002
48	SER	-0.089	0.002	0.069	0.001
49	GLY	-0.176	0.002	0.191	0.001
50	GLU	-0.060	0.002	-0.846	0.002
51	CYS	-0.056	0.002	0.032	0.001
52	GLY	-0.185	0.002	0.194	0.001
53	ARG	-0.066	0.002	1.017	0.002
54	LYS	-0.072	0.002	1.029	0.002
55	GLY	-0.193	0.003	0.192	0.001
56	PRO	-0.341	0.003	0.342	0.002
57	GLY	-0.197	0.002	0.188	0.001
58	PRO	-0.305	0.002	0.322	0.001
59	GLY	-0.192	0.002	0.189	0.000
60	GLY	-0.202	0.002	0.189	0.000
61	PRO	-0.303	0.002	0.325	0.001
62	GLY	-0.203	0.002	0.192	0.001
63	GLY	-0.186	0.002	0.191	0.001
64	ALA	-0.057	0.002	0.064	0.001
65	GLY	-0.197	0.002	0.195	0.001
66	GLY	-0.185	0.002	0.191	0.001
67	ALA	-0.044	0.002	0.061	0.001
68	ARG	-0.093	0.002	1.032	0.002
69	GLY	-0.186	0.002	0.193	0.001
70	GLY	-0.186	0.002	0.193	0.001
71	ALA	-0.058	0.002	0.066	0.001
72	GLY	-0.203	0.002	0.188	0.001
73	GLY	-0.197	0.002	0.191	0.001
74	GLY	-0.202	0.002	0.189	0.001
75	PRO	-0.295	0.002	0.319	0.001
76	SER	-0.092	0.002	0.057	0.001
77	GLY	-0.230	0.002	0.189	0.000
78	ASP	-0.928	0.002	-0.959	0.002

Table A5 Charge distribution on a coarse-grained side-chain C_α model of apoCaM. pH = 7.2, I = 0.15 M. The unit of charge is $1.6*10^{-19}$ C.

Residue index	Residue name	Charge on Cα	Error on the Cα charge	Charge on side chain	Error on the side chain charge
1	ALA	1.564	0.004	0.293	0.002
2	ASP	-0.038	0.003	-0.889	0.002
3	GLN	-0.050	0.003	0.051	0.001
4	LEU	-0.068	0.002	0.081	0.001
5	THR	-0.093	0.002	0.073	0.001
6	GLU	-0.051	0.002	-0.915	0.002
7	GLU	-0.077	0.003	-0.872	0.002
8	GLN	-0.033	0.003	0.052	0.001
9	ILE	-0.104	0.002	0.088	0.001
10	ALA	-0.063	0.003	0.059	0.001
11	GLU	-0.059	0.003	-0.865	0.002
12	PHE	-0.049	0.003	0.049	0.002
13	LYS	-0.087	0.003	1.001	0.001
14	GLU	-0.057	0.002	-0.893	0.002
15	ALA	-0.030	0.002	0.059	0.001
16	PHE	-0.063	0.003	0.050	0.001
17	SER	-0.109	0.003	0.070	0.001
18	LEU	-0.062	0.002	0.078	0.001
19	PHE	-0.060	0.003	0.048	0.001
20	ASP	-0.043	0.003	-0.885	0.002
21	LYS	-0.083	0.002	0.986	0.001
22	ASP	-0.049	0.002	-0.914	0.002
23	GLY	-0.195	0.002	0.190	0.000
24	ASP	-0.062	0.002	-0.948	0.001
25	GLY	-0.152	0.002	0.186	0.001
26	THR	-0.064	0.003	0.064	0.001
27	ILE	-0.060	0.003	0.076	0.001
28	THR	-0.110	0.003	0.077	0.001
29	THR	-0.094	0.003	0.087	0.001
30	LYS	-0.070	0.002	0.998	0.002
31	GLU	-0.059	0.003	-0.866	0.001
32	LEU	-0.081	0.003	0.082	0.001
33	GLY	-0.181	0.003	0.191	0.001
34	THR	-0.108	0.003	0.083	0.001

35	VAL	-0.080	0.003	0.105	0.001
36	MET	-0.068	0.003	0.061	0.001
37	ARG	-0.074	0.003	1.003	0.002
38	SER	-0.090	0.003	0.075	0.001
39	LEU	-0.076	0.003	0.083	0.001
40	GLY	-0.194	0.003	0.188	0.001
41	GLN	-0.034	0.002	0.054	0.001
42	ASN	-0.789	0.003	0.744	0.002
43	PRO	-0.249	0.003	0.267	0.003
44	THR	-0.090	0.002	0.071	0.001
45	GLU	-0.057	0.002	-0.902	0.002
46	ALA	-0.051	0.002	0.047	0.001
47	GLU	-0.050	0.002	-0.880	0.002
48	LEU	-0.066	0.002	0.070	0.001
49	GLN	-0.067	0.002	0.043	0.001
50	ASP	-0.032	0.003	-0.947	0.001
51	MET	-0.036	0.003	0.053	0.001
52	ILE	-0.085	0.003	0.097	0.001
53	ASN	-0.038	0.002	-0.005	0.001
54	GLU	-0.058	0.002	-0.902	0.002
55	VAL	-0.073	0.003	0.094	0.001
56	ASP	-0.066	0.002	-0.913	0.001
57	ALA	-0.063	0.002	0.055	0.001
58	ASP	-0.056	0.002	-0.940	0.002
59	GLY	-0.188	0.002	0.192	0.000
60	ASN	-0.052	0.002	0.005	0.001
61	GLY	-0.160	0.003	0.188	0.001
62	THR	-0.083	0.003	0.074	0.001
63	ILE	-0.080	0.003	0.082	0.001
64	ASP	-0.024	0.002	-0.936	0.006
65	PHE	-0.742	0.004	0.744	0.008
66	PRO	-0.256	0.004	0.256	0.005
67	GLU	-0.051	0.003	-0.878	0.002
68	PHE	-0.037	0.003	0.054	0.008
69	LEU	-0.070	0.004	0.077	0.001
70	THR	-0.103	0.003	0.073	0.001
71	MET	-0.056	0.003	0.069	0.006
72	MET	-0.067	0.003	0.066	0.004

73	ALA	-0.034	0.003	0.055	0.001
74	ARG	-0.079	0.002	0.980	0.002
75	LYS	-0.056	0.003	0.994	0.001
76	MET	-0.051	0.003	0.071	0.006
77	LYS	-0.055	0.002	0.979	0.001
78	ASP	-0.046	0.002	-0.940	0.001
79	THR	-0.086	0.002	0.080	0.001
80	ASP	-0.012	0.003	-0.912	0.002
81	SER	-0.088	0.002	0.065	0.001
82	GLU	-0.064	0.002	-0.862	0.002
83	GLU	-0.068	0.002	-0.869	0.002
84	GLU	-0.061	0.002	-0.875	0.002
85	ILE	-0.093	0.003	0.094	0.001
86	ARG	-0.064	0.003	0.944	0.002
87	GLU	-0.073	0.003	-0.862	0.002
88	ALA	-0.043	0.003	0.061	0.001
89	PHE	-0.054	0.003	0.049	0.001
90	ARG	-0.072	0.002	0.974	0.002
91	VAL	-0.084	0.002	0.097	0.001
92	PHE	-0.075	0.003	0.054	0.001
93	ASP	-0.025	0.003	-0.892	0.001
94	LYS	-0.082	0.002	0.995	0.001
95	ASP	-0.055	0.002	-0.941	0.002
96	GLY	-0.193	0.002	0.195	0.001
97	ASN	-0.041	0.002	0.000	0.001
98	GLY	-0.165	0.002	0.189	0.001
99	TYR	-0.036	0.002	0.031	0.001
100	ILE	-0.074	0.002	0.094	0.001
101	SER	-0.097	0.003	0.060	0.001
102	ALA	-0.061	0.002	0.066	0.001
103	ALA	-0.044	0.002	0.052	0.001
104	GLU	-0.049	0.003	-0.874	0.001
105	LEU	-0.095	0.003	0.076	0.001
106	ARG	-0.056	0.003	0.995	0.002
107	HIS	-0.055	0.003	0.028	0.001
108	VAL	-0.080	0.003	0.094	0.001
109	MET	-0.053	0.003	0.060	0.001
110	THR	-0.088	0.003	0.081	0.001

111	ASN	-0.039	0.003	0.000	0.001
112	LEU	-0.073	0.003	0.081	0.001
113	GLY	-0.194	0.002	0.188	0.001
114	GLU	-0.061	0.002	-0.871	0.002
115	LYS	-0.080	0.002	0.988	0.001
116	LEU	-0.060	0.002	0.072	0.001
117	THR	-0.091	0.003	0.070	0.001
118	ASP	-0.025	0.002	-0.921	0.002
119	GLU	-0.067	0.002	-0.915	0.002
120	GLU	-0.060	0.002	-0.884	0.002
121	VAL	-0.084	0.002	0.086	0.001
122	ASP	-0.030	0.002	-0.903	0.001
123	GLU	-0.058	0.002	-0.888	0.002
124	MET	-0.049	0.002	0.053	0.001
125	ILE	-0.087	0.002	0.087	0.001
126	ARG	-0.060	0.002	0.949	0.002
127	GLU	-0.079	0.002	-0.913	0.002
128	ALA	-0.051	0.002	0.054	0.001
129	ASP	-0.019	0.002	-0.952	0.001
130	ILE	-0.077	0.002	0.077	0.001
131	ASP	-0.038	0.002	-0.950	0.002
132	GLY	-0.210	0.002	0.188	0.001
133	ASP	-0.037	0.002	-0.955	0.001
134	GLY	-0.174	0.002	0.190	0.001
135	GLN	-0.037	0.002	0.029	0.001
136	VAL	-0.078	0.002	0.097	0.001
137	ASN	-0.023	0.002	0.001	0.001
138	TYR	-0.051	0.003	0.033	0.001
139	GLU	-0.071	0.003	-0.893	0.002
140	GLU	-0.052	0.002	-0.893	0.002
141	PHE	-0.061	0.003	0.055	0.003
142	VAL	-0.091	0.003	0.093	0.001
143	GLN	-0.058	0.003	0.044	0.001
144	MET	-0.045	0.002	0.059	0.001
145	MET	-0.053	0.002	0.056	0.003
146	THR	-0.118	0.002	0.076	0.001
147	ALA	-0.050	0.003	0.042	0.001
148	LYS	-2.693	0.005	2.669	0.003

Table A6 Charge distribution on a coarse-grained side-chain C_α model of Ng $^{13-49}$. pH = 7.2, I = 0.15 M. The unit of charge is $1.6*10^{-19}$ C.

Residue index	Residue name	Charge on Cα	Error on the Cα charge	Charge on side chain	Error on the side chain charge
13	ASP	1.574	0.004	-1.659	0.004
14	ASP	-0.020	0.003	-0.921	0.002
15	ASP	-0.034	0.003	-0.911	0.002
16	ILE	-0.071	0.003	0.081	0.001
17	LEU	-0.060	0.002	0.061	0.001
18	ASP	-0.037	0.003	-0.916	0.002
19	ILE	-0.539	0.004	0.540	0.002
20	PRO	-0.265	0.003	0.258	0.003
21	LEU	-0.054	0.003	0.064	0.001
22	ASP	-0.042	0.003	-0.931	0.002
23	ASP	-0.818	0.004	-0.133	0.003
24	PRO	-0.282	0.003	0.261	0.002
25	GLY	-0.181	0.003	0.187	0.001
26	ALA	-0.032	0.002	0.054	0.001
27	ASN	-0.046	0.003	0.005	0.001
28	ALA	-0.040	0.003	0.064	0.001
29	ALA	-0.059	0.003	0.060	0.001
30	ALA	-0.066	0.003	0.062	0.001
31	ALA	-0.055	0.003	0.064	0.001
32	LYS	-0.079	0.003	0.996	0.001
33	ILE	-0.089	0.003	0.100	0.001
34	GLN	-0.072	0.002	0.059	0.001
35	ALA	-0.042	0.003	0.062	0.001
36	SER	-0.092	0.003	0.076	0.001
37	PHE	-0.048	0.003	0.057	0.001
38	ARG	-0.077	0.002	1.002	0.002
39	GLY	-0.175	0.002	0.187	0.001
40	HIS	-0.054	0.002	0.037	0.001
41	MET	-0.068	0.002	0.065	0.001
42	ALA	-0.057	0.003	0.064	0.001
43	ARG	-0.059	0.003	0.984	0.002
44	LYS	-0.061	0.003	0.997	0.002
45	LYS	-0.084	0.002	1.000	0.002

46	ILE	-0.089	0.003	0.106	0.001
47	LYS	-0.075	0.003	1.001	0.002
48	SER	-0.132	0.003	0.066	0.001
49	GLY	-1.086	0.002	0.174	0.001

Table A7 Charge distribution on a coarse-grained side-chain C_α model of apoCaM. pH = 6.3, I = 0.1 M. The unit of charge is $1.6*10^{-19}$ C.

Residue Index	Residue Name	Charge on C_α	Error on the C_α charge	Charge on side chain	Error on the side chain charge
1	ALA	1.565	0.004	0.292	0.001
2	ASP	-0.041	0.003	-0.885	0.002
3	GLN	-0.052	0.002	0.053	0.001
4	LEU	-0.066	0.002	0.080	0.001
5	THR	-0.095	0.002	0.075	0.001
6	GLU	-0.051	0.002	-0.914	0.002
7	GLU	-0.072	0.003	-0.871	0.002
8	GLN	-0.040	0.003	0.053	0.001
9	ILE	-0.098	0.003	0.088	0.001
10	ALA	-0.061	0.003	0.058	0.001
11	GLU	-0.061	0.003	-0.868	0.002
12	PHE	-0.047	0.003	0.047	0.001
13	LYS	-0.090	0.003	0.998	0.001
14	GLU	-0.055	0.003	-0.893	0.002
15	ALA	-0.035	0.002	0.058	0.001
16	PHE	-0.055	0.003	0.053	0.001
17	SER	-0.118	0.003	0.070	0.001
18	LEU	-0.060	0.002	0.076	0.001
19	PHE	-0.054	0.003	0.047	0.001
20	ASP	-0.049	0.003	-0.885	0.002
21	LYS	-0.078	0.002	0.986	0.001
22	ASP	-0.049	0.003	-0.914	0.002
23	GLY	-0.196	0.002	0.191	0.001
24	ASP	-0.064	0.002	-0.945	0.001
25	GLY	-0.150	0.002	0.184	0.001
26	THR	-0.062	0.003	0.065	0.001
27	ILE	-0.066	0.003	0.077	0.001
28	THR	-0.107	0.003	0.079	0.001
29	THR	-0.095	0.003	0.089	0.001
30	LYS	-0.072	0.003	0.995	0.001
31	GLU	-0.058	0.003	-0.866	0.002
32	LEU	-0.079	0.003	0.082	0.001
33	GLY	-0.180	0.003	0.191	0.001
34	THR	-0.108	0.003	0.081	0.001

35	VAL	-0.082	0.003	0.108	0.001
36	MET	-0.066	0.003	0.061	0.001
37	ARG	-0.080	0.002	1.006	0.002
38	SER	-0.088	0.003	0.072	0.001
39	LEU	-0.077	0.003	0.085	0.001
40	GLY	-0.192	0.003	0.188	0.001
41	GLN	-0.033	0.002	0.054	0.001
42	ASN	-0.782	0.004	0.744	0.002
43	PRO	-0.254	0.004	0.265	0.002
44	THR	-0.092	0.002	0.070	0.001
45	GLU	-0.052	0.002	-0.904	0.002
46	ALA	-0.053	0.002	0.049	0.001
47	GLU	-0.052	0.002	-0.883	0.002
48	LEU	-0.064	0.002	0.069	0.001
49	GLN	-0.069	0.002	0.044	0.001
50	ASP	-0.031	0.002	-0.947	0.001
51	MET	-0.037	0.003	0.054	0.001
52	ILE	-0.084	0.003	0.096	0.001
53	ASN	-0.038	0.002	-0.006	0.001
54	GLU	-0.052	0.002	-0.902	0.002
55	VAL	-0.077	0.003	0.093	0.001
56	ASP	-0.062	0.002	-0.915	0.001
57	ALA	-0.063	0.002	0.053	0.001
58	ASP	-0.056	0.002	-0.940	0.001
59	GLY	-0.187	0.002	0.192	0.000
60	ASN	-0.055	0.002	0.004	0.001
61	GLY	-0.162	0.002	0.187	0.001
62	THR	-0.075	0.003	0.073	0.001
63	ILE	-0.083	0.003	0.079	0.001
64	ASP	-0.017	0.003	-0.945	0.001
65	PHE	-0.740	0.004	0.732	0.002
66	PRO	-0.260	0.003	0.251	0.002
67	GLU	-0.055	0.003	-0.880	0.002
68	PHE	-0.045	0.003	0.048	0.007
69	LEU	-0.061	0.002	0.075	0.001
70	THR	-0.103	0.002	0.072	0.001
71	MET	-0.053	0.003	0.061	0.001
72	MET	-0.066	0.003	0.066	0.006

73	ALA	-0.040	0.002	0.057	0.001
74	ARG	-0.073	0.002	0.978	0.002
75	LYS	-0.063	0.002	0.994	0.001
76	MET	-0.051	0.003	0.066	0.001
77	LYS	-0.056	0.003	0.980	0.001
78	ASP	-0.045	0.002	-0.939	0.001
79	THR	-0.083	0.002	0.079	0.001
80	ASP	-0.018	0.003	-0.912	0.001
81	SER	-0.088	0.002	0.067	0.001
82	GLU	-0.062	0.002	-0.865	0.002
83	GLU	-0.065	0.002	-0.869	0.002
84	GLU	-0.066	0.003	-0.875	0.002
85	ILE	-0.093	0.003	0.094	0.001
86	ARG	-0.067	0.002	0.945	0.002
87	GLU	-0.074	0.003	-0.859	0.002
88	ALA	-0.040	0.003	0.061	0.001
89	PHE	-0.055	0.003	0.049	0.001
90	ARG	-0.076	0.003	0.974	0.002
91	VAL	-0.081	0.003	0.097	0.001
92	PHE	-0.074	0.003	0.053	0.003
93	ASP	-0.025	0.003	-0.893	0.002
94	LYS	-0.078	0.002	0.999	0.001
95	ASP	-0.055	0.002	-0.940	0.001
96	GLY	-0.189	0.002	0.194	0.001
97	ASN	-0.044	0.002	0.000	0.001
98	GLY	-0.164	0.002	0.189	0.001
99	TYR	-0.031	0.002	0.032	0.001
100	ILE	-0.075	0.002	0.092	0.001
101	SER	-0.103	0.002	0.063	0.001
102	ALA	-0.062	0.002	0.070	0.001
103	ALA	-0.045	0.002	0.054	0.001
104	GLU	-0.050	0.003	-0.870	0.001
105	LEU	-0.096	0.003	0.080	0.001
106	ARG	-0.053	0.003	0.995	0.002
107	HIS	-0.048	0.003	0.973	0.002
108	VAL	-0.087	0.003	0.098	0.001
109	MET	-0.053	0.003	0.064	0.001
110	THR	-0.087	0.003	0.082	0.001

111	ASN	-0.043	0.003	0.002	0.001
112	LEU	-0.068	0.003	0.085	0.001
113	GLY	-0.193	0.002	0.187	0.001
114	GLU	-0.062	0.002	-0.870	0.002
115	LYS	-0.083	0.002	0.988	0.002
116	LEU	-0.061	0.003	0.074	0.001
117	THR	-0.090	0.003	0.068	0.001
118	ASP	-0.025	0.003	-0.921	0.002
119	GLU	-0.064	0.002	-0.913	0.002
120	GLU	-0.061	0.002	-0.883	0.002
121	VAL	-0.086	0.002	0.085	0.001
122	ASP	-0.029	0.002	-0.901	0.001
123	GLU	-0.061	0.002	-0.888	0.002
124	MET	-0.045	0.002	0.052	0.001
125	ILE	-0.084	0.003	0.086	0.001
126	ARG	-0.065	0.002	0.950	0.002
127	GLU	-0.075	0.002	-0.912	0.002
128	ALA	-0.051	0.002	0.053	0.001
129	ASP	-0.023	0.003	-0.955	0.001
130	ILE	-0.072	0.003	0.075	0.001
131	ASP	-0.040	0.002	-0.951	0.002
132	GLY	-0.211	0.002	0.190	0.001
133	ASP	-0.038	0.002	-0.955	0.001
134	GLY	-0.172	0.002	0.190	0.001
135	GLN	-0.042	0.002	0.030	0.001
136	VAL	-0.077	0.002	0.099	0.001
137	ASN	-0.025	0.002	0.003	0.001
138	TYR	-0.053	0.002	0.033	0.001
139	GLU	-0.068	0.002	-0.891	0.002
140	GLU	-0.050	0.002	-0.896	0.001
141	PHE	-0.059	0.003	0.048	0.001
142	VAL	-0.093	0.003	0.093	0.001
143	GLN	-0.056	0.002	0.045	0.001
144	MET	-0.047	0.002	0.060	0.001
145	MET	-0.052	0.002	0.058	0.001
146	THR	-0.118	0.002	0.077	0.001
147	ALA	-0.055	0.002	0.041	0.001
148	LYS	-2.680	0.005	2.666	0.003

Table A8 Charge distribution on a coarse-grained side-chain C_α model of Ng $_{13\text{-}49}$. pH = 6.3, I = 0.1 M. The unit of charge is $1.6*10^{-19}$ C.

Residue Index	Residue Name	Charge on C_α	Error on the C_α charge	Charge on side chain	Error on the side chain charge
13	ASP	1.570	0.004	-1.657	0.003
14	ASP	-0.017	0.003	-0.918	0.002
15	ASP	-0.035	0.003	-0.909	0.002
16	ILE	-0.075	0.003	0.083	0.001
17	LEU	-0.059	0.003	0.063	0.001
18	ASP	-0.033	0.003	-0.918	0.002
19	ILE	-0.539	0.004	0.541	0.002
20	PRO	-0.269	0.003	0.257	0.002
21	LEU	-0.055	0.003	0.065	0.001
22	ASP	-0.039	0.002	-0.932	0.002
23	ASP	-0.813	0.004	-0.137	0.003
24	PRO	-0.286	0.003	0.265	0.002
25	GLY	-0.179	0.003	0.186	0.001
26	ALA	-0.031	0.002	0.056	0.001
27	ASN	-0.049	0.003	0.007	0.001
28	ALA	-0.045	0.003	0.063	0.001
29	ALA	-0.051	0.003	0.059	0.001
30	ALA	-0.070	0.003	0.063	0.001
31	ALA	-0.048	0.003	0.062	0.001
32	LYS	-0.084	0.003	0.998	0.001
33	ILE	-0.085	0.003	0.100	0.001
34	GLN	-0.074	0.002	0.060	0.001
35	ALA	-0.041	0.002	0.063	0.001
36	SER	-0.090	0.003	0.075	0.001
37	PHE	-0.050	0.003	0.057	0.001
38	ARG	-0.080	0.002	1.001	0.002
39	GLY	-0.174	0.003	0.189	0.001
40	HIS	-0.056	0.002	0.039	0.001
41	MET	-0.069	0.002	0.065	0.001
42	ALA	-0.062	0.002	0.067	0.001
43	ARG	-0.057	0.003	0.987	0.002
44	LYS	-0.064	0.003	0.996	0.002
45	LYS	-0.081	0.002	1.000	0.002
46	ILE	-0.091	0.003	0.105	0.001

47	LYS	-0.073	0.003	0.999	0.002
48	SER	-0.131	0.003	0.065	0.001
49	GLY	-1.086	0.002	0.176	0.001

Table A9 Solvent mediated interaction between the pairs of 20 amino acids given by the Betancourt-Thirumalai statistical potential ε^{BT} in kcal/mol.

ε^{BT}	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	0.72																			
R	0.44	0.52																		
N	0.46	0.59	0.62																	
D	0.42	1.03	0.67	0.44																
C	0.76	0.41	0.43	0.37	1.4															
Q	0.47	0.67	0.63	0.53	0.58	0.52														
E	0.34	1.05	0.61	0.36	0.32	0.54	0.33													
G	0.62	0.52	0.54	0.5	0.65	0.48	0.31	0.72												
H	0.47	0.58	0.54	0.73	0.71	0.47	0.67	0.46	0.8											
I	0.81	0.49	0.27	0.28	0.89	0.52	0.37	0.47	0.49	0.96										
L	0.82	0.55	0.38	0.23	0.9	0.55	0.38	0.52	0.54	1.07	1.09									
K	0.48	0.3	0.68	1.01	0.39	0.72	1.12	0.53	0.44	0.47	0.5	0.37								
M	0.74	0.5	0.41	0.23	0.89	0.61	0.46	0.55	0.7	0.96	1.01	0.47	0.94							
F	0.8	0.55	0.43	0.31	0.92	0.62	0.4	0.53	0.71	0.99	1.07	0.53	1.13	1.09						
P	0.56	0.61	0.52	0.45	0.71	0.63	0.44	0.61	0.63	0.57	0.65	0.53	0.7	0.71	0.64					
S	0.51	0.53	0.52	0.59	0.55	0.45	0.54	0.54	0.51	0.39	0.44	0.54	0.41	0.54	0.5	0.52				
T	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6			
W	0.84	0.85	0.65	0.56	1.04	0.67	0.69	0.74	0.88	0.99	1.02	0.77	1.16	1.07	1.04	0.56	0.6	1.04		
Y	0.69	0.82	0.59	0.64	0.7	0.71	0.7	0.62	0.73	0.8	0.86	0.84	0.91	0.89	0.84	0.56	0.6	0.93	0.76	
V	0.83	0.5	0.37	0.2	0.91	0.5	0.35	0.58	0.49	1.01	1.08	0.5	0.88	1	0.65	0.45	0.6	0.97	0.76	1.03

Table A10 Solvent mediated interaction between the pairs of 20 amino acids given by the rescaled intramolecular potential ε^{intra} in kcal/mol.

ε^{intra}	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	0.72																			
R	0.44	1.07																		
N	0.46	0.58	0.62																	
D	0.42	0.46	0.68	0.39																
C	0.76	0.39	0.43	0.50	1.40															
Q	0.47	0.70	0.63	0.87	0.58	0.52														
E	0.34	0.57	0.61	0.21	0.34	0.52	0.76													
G	0.62	0.73	0.54	0.72	0.65	0.49	0.10	0.72												
H	0.47	0.59	0.54	0.21	0.71	0.47	0.66	0.47	0.80											
I	0.81	0.55	0.27	0.20	0.89	0.52	0.32	0.51	0.49	0.97										
L	0.82	0.58	0.38	0.43	0.90	0.55	0.35	0.53	0.54	1.08	1.09									
K	0.48	0.86	0.68	0.20	0.37	0.74	0.63	0.76	0.45	0.53	0.53	0.94								
M	0.74	0.53	0.41	0.30	0.89	0.61	0.43	0.57	0.70	0.96	1.01	0.50	0.94							
F	0.80	0.57	0.43	0.28	0.92	0.63	0.38	0.54	0.71	0.99	1.07	0.55	1.14	1.09						
P	0.56	0.76	0.52	0.58	0.70	0.64	0.31	0.70	0.63	0.59	0.66	0.68	0.70	0.72	0.69					
S	0.51	0.55	0.52	0.53	0.55	0.45	0.53	0.55	0.51	0.39	0.45	0.56	0.41	0.54	0.50	0.52				
T	0.60	0.66	0.60	0.56	0.60	0.60	0.54	0.64	0.60	0.61	0.60	0.67	0.60	0.60	0.62	0.60	0.61			
W	0.84	0.85	0.65	0.64	1.04	0.67	0.69	0.75	0.88	0.99	1.02	0.77	1.16	1.07	1.04	0.56	0.60	1.04		
Y	0.69	0.82	0.59	0.22	0.70	0.71	0.69	0.63	0.73	0.80	0.86	0.84	0.91	0.89	0.84	0.56	0.60	0.93	0.76	
V	0.83	0.48	0.37	0.20	0.91	0.50	0.37	0.57	0.49	1.01	1.08	0.49	0.88	1.00	0.64	0.45	0.60	0.97	0.76	1.03

Table A11 Contact pair list used for correlation analysis. The non-specific contact pairs are determined from the difference contact map analysis. A contact pair is selected if the magnitude of the change in the probability of the contact pair in cluster 1 (Fig 4) is greater than 0.2 from unbound state to the bound state. The contact pairs are listed in a sequence of categories: within nCaM, between nCaM and cCaM, within cCaM, between nCaM and Ng $_{13\text{-}49}$, between cCaM and Ng $_{13\text{-}49}$ and within Ng $_{13\text{-}49}$. The sequence of apoCaM is from NMR structure (PDB: 1CFD) and sequence of Ng $_{13\text{-}49}$ is provided in the *Materials and Methods* in the main text.

Index	Residue Index	Residue Index	Interaction Category	Index	Residue Index	Residue Index	Interaction Category
1	4	12	nCaM	2	4	65	nCaM
3	4	69	nCaM	4	4	8	nCaM
5	4	9	nCaM	6	5	9	nCaM
7	6	10	nCaM	8	7	11	nCaM
9	8	12	nCaM	10	9	13	nCaM
11	9	65	nCaM	12	9	69	nCaM
13	10	14	nCaM	14	11	15	nCaM
15	11	39	nCaM	16	12	16	nCaM
17	12	39	nCaM	18	12	41	nCaM
19	12	65	nCaM	20	12	68	nCaM
21	12	69	nCaM	22	12	72	nCaM
23	13	17	nCaM	24	13	20	nCaM
25	13	65	nCaM	26	14	18	nCaM
27	15	38	nCaM	28	15	39	nCaM
29	16	20	nCaM	30	16	27	nCaM
31	16	35	nCaM	32	16	38	nCaM
33	16	39	nCaM	34	16	65	nCaM
35	16	68	nCaM	36	18	38	nCaM
37	19	27	nCaM	38	19	31	nCaM
39	19	34	nCaM	40	19	35	nCaM
41	19	37	nCaM	42	19	38	nCaM
43	20	26	nCaM	44	20	27	nCaM
45	21	31	nCaM	46	22	26	nCaM
47	22	28	nCaM	48	22	31	nCaM
49	22	62	nCaM	50	24	60	nCaM
51	24	62	nCaM	52	26	62	nCaM
53	26	63	nCaM	54	26	64	nCaM
55	27	31	nCaM	56	27	32	nCaM
57	27	35	nCaM	58	27	62	nCaM
59	27	63	nCaM	60	27	68	nCaM

61	28	32	nCaM	62	28	62	nCaM
63	29	49	nCaM	64	29	52	nCaM
65	29	56	nCaM	66	29	62	nCaM
67	30	45	nCaM	68	30	49	nCaM
69	31	35	nCaM	70	32	36	nCaM
71	32	48	nCaM	72	32	52	nCaM
73	32	63	nCaM	74	32	68	nCaM
75	32	71	nCaM	76	34	38	nCaM
77	35	39	nCaM	78	35	68	nCaM
79	36	41	nCaM	80	36	42	nCaM
81	36	43	nCaM	82	36	48	nCaM
83	36	68	nCaM	84	36	71	nCaM
85	36	72	nCaM	86	36	75	nCaM
87	37	41	nCaM	88	37	42	nCaM
89	39	68	nCaM	90	39	72	nCaM
91	41	72	nCaM	92	41	75	nCaM
93	43	47	nCaM	94	43	48	nCaM
95	43	75	nCaM	96	44	48	nCaM
97	45	49	nCaM	98	46	50	nCaM
99	47	51	nCaM	100	47	75	nCaM
101	48	52	nCaM	102	48	71	nCaM
103	48	75	nCaM	104	49	53	nCaM
105	49	56	nCaM	106	50	54	nCaM
107	51	55	nCaM	108	51	71	nCaM
109	51	74	nCaM	110	51	75	nCaM
111	52	56	nCaM	112	52	63	nCaM
113	52	71	nCaM	114	53	57	nCaM
115	54	71	nCaM	116	54	74	nCaM
117	55	63	nCaM	118	55	67	nCaM
119	55	71	nCaM	120	56	60	nCaM
121	58	67	nCaM	122	63	67	nCaM
123	63	68	nCaM	124	63	71	nCaM
125	64	68	nCaM	126	65	69	nCaM
127	67	71	nCaM	128	67	74	nCaM
129	68	72	nCaM	130	69	73	nCaM
131	70	74	nCaM	132	71	75	nCaM
133	1	114	nCaM:cCaM	134	1	143	nCaM:cCaM
135	1	147	nCaM:cCaM	136	1	148	nCaM:cCaM

137	1	81	nCaM:cCaM	138	1	85	nCaM:cCaM
139	2	124	nCaM:cCaM	140	2	146	nCaM:cCaM
141	2	147	nCaM:cCaM	142	2	148	nCaM:cCaM
143	2	76	nCaM:cCaM	144	2	77	nCaM:cCaM
145	2	80	nCaM:cCaM	146	2	81	nCaM:cCaM
147	2	82	nCaM:cCaM	148	2	84	nCaM:cCaM
149	3	109	nCaM:cCaM	150	3	114	nCaM:cCaM
151	3	115	nCaM:cCaM	152	3	116	nCaM:cCaM
153	3	124	nCaM:cCaM	154	3	145	nCaM:cCaM
155	3	147	nCaM:cCaM	156	3	148	nCaM:cCaM
157	4	147	nCaM:cCaM	158	4	76	nCaM:cCaM
159	5	147	nCaM:cCaM	160	5	148	nCaM:cCaM
161	6	115	nCaM:cCaM	162	7	114	nCaM:cCaM
163	7	115	nCaM:cCaM	164	7	116	nCaM:cCaM
165	7	120	nCaM:cCaM	166	7	124	nCaM:cCaM
167	7	148	nCaM:cCaM	168	8	115	nCaM:cCaM
169	8	120	nCaM:cCaM	170	8	147	nCaM:cCaM
171	8	76	nCaM:cCaM	172	8	81	nCaM:cCaM
173	8	84	nCaM:cCaM	174	11	115	nCaM:cCaM
175	11	120	nCaM:cCaM	176	11	76	nCaM:cCaM
177	12	76	nCaM:cCaM	178	14	115	nCaM:cCaM
179	39	76	nCaM:cCaM	180	41	76	nCaM:cCaM
181	41	79	nCaM:cCaM	182	41	84	nCaM:cCaM
183	66	143	nCaM:cCaM	184	67	139	nCaM:cCaM
185	67	143	nCaM:cCaM	186	69	76	nCaM:cCaM
187	72	76	nCaM:cCaM	188	73	77	nCaM:cCaM
189	73	80	nCaM:cCaM	190	74	80	nCaM:cCaM
191	75	79	nCaM:cCaM	192	76	80	cCaM
193	76	81	cCaM	194	76	84	cCaM
195	77	138	cCaM	196	77	82	cCaM
197	78	83	cCaM	198	79	83	cCaM
199	79	84	cCaM	200	80	146	cCaM
201	81	145	cCaM	202	81	146	cCaM
203	81	85	cCaM	204	82	138	cCaM
205	82	139	cCaM	206	82	142	cCaM
207	82	143	cCaM	208	82	146	cCaM
209	82	86	cCaM	210	82	89	cCaM
211	83	87	cCaM	212	84	88	cCaM

213	85	112	cCaM	214	85	114	cCaM
215	85	142	cCaM	216	85	145	cCaM
217	85	146	cCaM	218	85	89	cCaM
219	86	138	cCaM	220	86	139	cCaM
221	86	142	cCaM	222	86	143	cCaM
223	86	90	cCaM	224	87	91	cCaM
225	88	112	cCaM	226	88	92	cCaM
227	89	105	cCaM	228	89	108	cCaM
229	89	109	cCaM	230	89	112	cCaM
231	89	114	cCaM	232	89	138	cCaM
233	89	141	cCaM	234	89	142	cCaM
235	89	145	cCaM	236	89	93	cCaM
237	89	100	cCaM	238	90	138	cCaM
239	90	94	cCaM	240	92	104	cCaM
241	92	107	cCaM	242	92	108	cCaM
243	92	111	cCaM	244	92	112	cCaM
245	92	100	cCaM	246	93	104	cCaM
247	93	138	cCaM	248	93	97	cCaM
249	93	99	cCaM	250	93	100	cCaM
251	94	104	cCaM	252	97	135	cCaM
253	99	135	cCaM	254	99	136	cCaM
255	99	137	cCaM	256	99	138	cCaM
257	100	104	cCaM	258	100	105	cCaM
259	100	108	cCaM	260	100	135	cCaM
261	100	136	cCaM	262	100	138	cCaM
263	101	105	cCaM	264	101	135	cCaM
265	102	106	cCaM	266	102	130	cCaM
267	102	131	cCaM	268	102	135	cCaM
269	102	136	cCaM	270	103	107	cCaM
271	104	108	cCaM	272	104	111	cCaM
273	105	109	cCaM	274	105	130	cCaM
275	105	136	cCaM	276	105	138	cCaM
277	105	141	cCaM	278	106	110	cCaM
279	106	118	cCaM	280	106	121	cCaM
281	106	122	cCaM	282	106	125	cCaM
283	106	126	cCaM	284	106	130	cCaM
285	106	131	cCaM	286	107	111	cCaM
287	108	112	cCaM	288	109	114	cCaM

289	109	116	cCaM	290	109	121	cCaM
291	109	124	cCaM	292	109	125	cCaM
293	109	130	cCaM	294	109	141	cCaM
295	109	145	cCaM	296	110	114	cCaM
297	110	118	cCaM	298	114	145	cCaM
299	116	120	cCaM	300	116	121	cCaM
301	116	124	cCaM	302	116	145	cCaM
303	117	121	cCaM	304	118	122	cCaM
305	120	124	cCaM	306	121	125	cCaM
307	122	126	cCaM	308	123	127	cCaM
309	123	144	cCaM	310	123	148	cCaM
311	124	128	cCaM	312	124	141	cCaM
313	124	144	cCaM	314	124	145	cCaM
315	124	148	cCaM	316	125	129	cCaM
317	125	130	cCaM	318	125	141	cCaM
319	126	131	cCaM	320	127	140	cCaM
321	127	144	cCaM	322	127	148	cCaM
323	128	140	cCaM	324	128	141	cCaM
325	128	144	cCaM	326	129	140	cCaM
327	130	135	cCaM	328	130	136	cCaM
329	130	141	cCaM	330	136	141	cCaM
331	137	141	cCaM	332	138	142	cCaM
333	139	143	cCaM	334	140	144	cCaM
335	141	145	cCaM	336	142	146	cCaM
337	143	147	cCaM	338	143	148	cCaM
339	144	148	cCaM	340	1	16	nCaM:Ng
341	1	18	nCaM:Ng	342	2	13	nCaM:Ng
343	2	18	nCaM:Ng	344	2	22	nCaM:Ng
345	2	40	nCaM:Ng	346	3	15	nCaM:Ng
347	3	16	nCaM:Ng	348	3	17	nCaM:Ng
349	3	18	nCaM:Ng	350	3	20	nCaM:Ng
351	3	41	nCaM:Ng	352	3	43	nCaM:Ng
353	3	45	nCaM:Ng	354	4	16	nCaM:Ng
355	4	21	nCaM:Ng	356	4	41	nCaM:Ng
357	6	15	nCaM:Ng	358	7	44	nCaM:Ng
359	8	32	nCaM:Ng	360	8	45	nCaM:Ng
361	9	17	nCaM:Ng	362	9	41	nCaM:Ng
363	11	18	nCaM:Ng	364	11	32	nCaM:Ng

365	11	44	nCaM:Ng	366	12	17	nCaM:Ng
367	12	18	nCaM:Ng	368	12	19	nCaM:Ng
369	12	21	nCaM:Ng	370	12	41	nCaM:Ng
371	13	17	nCaM:Ng	372	15	16	nCaM:Ng
373	15	17	nCaM:Ng	374	16	17	nCaM:Ng
375	16	19	nCaM:Ng	376	18	16	nCaM:Ng
377	19	16	nCaM:Ng	378	26	19	nCaM:Ng
379	26	20	nCaM:Ng	380	27	21	nCaM:Ng
381	29	22	nCaM:Ng	382	32	21	nCaM:Ng
383	35	17	nCaM:Ng	384	38	15	nCaM:Ng
385	38	16	nCaM:Ng	386	38	17	nCaM:Ng
387	39	15	nCaM:Ng	388	39	16	nCaM:Ng
389	39	17	nCaM:Ng	390	39	18	nCaM:Ng
391	39	19	nCaM:Ng	392	41	34	nCaM:Ng
393	62	20	nCaM:Ng	394	62	21	nCaM:Ng
395	62	22	nCaM:Ng	396	63	20	nCaM:Ng
397	63	21	nCaM:Ng	398	65	17	nCaM:Ng
399	65	19	nCaM:Ng	400	65	21	nCaM:Ng
401	65	41	nCaM:Ng	402	65	46	nCaM:Ng
403	67	38	nCaM:Ng	404	67	44	nCaM:Ng
405	67	47	nCaM:Ng	406	68	17	nCaM:Ng
407	68	19	nCaM:Ng	408	68	41	nCaM:Ng
409	69	19	nCaM:Ng	410	69	20	nCaM:Ng
411	69	21	nCaM:Ng	412	69	41	nCaM:Ng
413	70	38	nCaM:Ng	414	72	19	nCaM:Ng
415	72	41	nCaM:Ng	416	74	37	nCaM:Ng
417	74	38	nCaM:Ng	418	74	43	nCaM:Ng
419	74	47	nCaM:Ng	420	76	20	cCaM:Ng
421	76	41	cCaM:Ng	422	76	43	cCaM:Ng
423	77	43	cCaM:Ng	424	77	44	cCaM:Ng
425	80	43	cCaM:Ng	426	81	27	cCaM:Ng
427	81	43	cCaM:Ng	428	82	18	cCaM:Ng
429	82	32	cCaM:Ng	430	82	37	cCaM:Ng
431	82	38	cCaM:Ng	432	82	40	cCaM:Ng
433	82	41	cCaM:Ng	434	82	43	cCaM:Ng
435	82	44	cCaM:Ng	436	84	33	cCaM:Ng
437	84	34	cCaM:Ng	438	84	38	cCaM:Ng
439	85	17	cCaM:Ng	440	85	18	cCaM:Ng

441	85	33	cCaM:Ng	442	86	23	cCaM:Ng
443	86	38	cCaM:Ng	444	88	33	cCaM:Ng
445	89	17	cCaM:Ng	446	89	18	cCaM:Ng
447	89	19	cCaM:Ng	448	89	21	cCaM:Ng
449	89	23	cCaM:Ng	450	89	37	cCaM:Ng
451	99	21	cCaM:Ng	452	99	22	cCaM:Ng
453	99	23	cCaM:Ng	454	99	24	cCaM:Ng
455	99	34	cCaM:Ng	456	99	38	cCaM:Ng
457	100	21	cCaM:Ng	458	105	17	cCaM:Ng
459	105	19	cCaM:Ng	460	105	21	cCaM:Ng
461	106	21	cCaM:Ng	462	109	17	cCaM:Ng
463	109	19	cCaM:Ng	464	109	21	cCaM:Ng
465	112	17	cCaM:Ng	466	112	33	cCaM:Ng
467	112	37	cCaM:Ng	468	114	15	cCaM:Ng
469	114	17	cCaM:Ng	470	114	32	cCaM:Ng
471	114	33	cCaM:Ng	472	114	36	cCaM:Ng
473	114	41	cCaM:Ng	474	114	43	cCaM:Ng
475	114	44	cCaM:Ng	476	115	18	cCaM:Ng
477	115	23	cCaM:Ng	478	115	32	cCaM:Ng
479	115	36	cCaM:Ng	480	115	40	cCaM:Ng
481	115	42	cCaM:Ng	482	115	43	cCaM:Ng
483	115	44	cCaM:Ng	484	115	45	cCaM:Ng
485	116	32	cCaM:Ng	486	120	32	cCaM:Ng
487	120	43	cCaM:Ng	488	120	44	cCaM:Ng
489	120	45	cCaM:Ng	490	124	32	cCaM:Ng
491	125	21	cCaM:Ng	492	130	21	cCaM:Ng
493	135	20	cCaM:Ng	494	135	22	cCaM:Ng
495	135	36	cCaM:Ng	496	136	20	cCaM:Ng
497	137	20	cCaM:Ng	498	138	18	cCaM:Ng
499	138	19	cCaM:Ng	500	138	20	cCaM:Ng
501	138	22	cCaM:Ng	502	138	23	cCaM:Ng
503	138	27	cCaM:Ng	504	138	32	cCaM:Ng
505	138	33	cCaM:Ng	506	138	36	cCaM:Ng
507	138	40	cCaM:Ng	508	138	43	cCaM:Ng
509	138	44	cCaM:Ng	510	141	17	cCaM:Ng
511	141	19	cCaM:Ng	512	141	21	cCaM:Ng
513	142	16	cCaM:Ng	514	142	17	cCaM:Ng
515	142	18	cCaM:Ng	516	145	16	cCaM:Ng

517	145	17	cCaM:Ng	518	145	32	cCaM:Ng
519	145	33	cCaM:Ng	520	145	41	cCaM:Ng
521	145	43	cCaM:Ng	522	145	44	cCaM:Ng
523	146	16	cCaM:Ng	524	146	33	cCaM:Ng
525	146	35	cCaM:Ng	526	146	36	cCaM:Ng
527	146	42	cCaM:Ng	528	147	47	cCaM:Ng
529	18	32	Ng	530	20	26	Ng
531	20	27	Ng	532	20	28	Ng
533	22	27	Ng	534	23	27	Ng
535	27	31	Ng	536	27	32	Ng
537	28	32	Ng	538	29	33	Ng
539	30	34	Ng	540	30	36	Ng
541	32	36	Ng	542	32	37	Ng
543	33	37	Ng	544	34	38	Ng
545	36	40	Ng	546	37	41	Ng
547	38	43	Ng	548	40	44	Ng
549	41	45	Ng	550	41	46	Ng
551	43	47	Ng				

Bibliography

Alsallaq, R. and H. X. Zhou (2007). "Energy landscape and transition state of protein-protein association." Biophysical journal **92**(5): 1486-1502.

Alsallaq, R. and H. X. Zhou (2007). "Prediction of protein-protein association rates from a transition-state theory." Structure **15**(2): 215-224.

Anthis, N. J., M. Doucleff, et al. (2011). "Transient, Sparsely Populated Compact States of Apo and Calcium-Loaded Calmodulin Probed by Paramagnetic Relaxation Enhancement: Interplay of Conformational Selection and Induced Fit." Journal of the American Chemical Society **133**(46): 18966-18974.

Arai, M., K. Sugase, et al. (2015). "Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding." Proceedings of the National Academy of Sciences of the United States of America **112**(31): 9614-9619.

Bahler, M. and A. Rhoads (2002). "Calmodulin signaling via the IQ motif." FEBS letters **513**(1): 107-113.

Barth, A., S. R. Martin, et al. (1998). "Specificity and symmetry in the interaction of calmodulin domains with the skeletal muscle mysoin light chain kinase target sequence." J. Biol. Chem. **273**: 2174-2183.

Berchtold, M. W. and A. Villalobo (2014). "The many faces of calmodulin in cell proliferation, programmed cell death, autophagy, and cancer." Biochim Biophys Acta **1843**(2): 398-435.

Berjanskii, M. and D. S. Wishart (2006). "NMR: prediction of protein flexibility." Nature protocols **1**(2): 683-688.

Best, R. B., Y. G. Chen, et al. (2005). "Slow protein conformational dynamics from multiple experimental structures: The helix/sheet transition of arc repressor." Structure **13**(12): 1755-1763.

Betancourt, M. R. and D. Thirumalai (1999). "Pair potentials for protein folding: Choice of reference states and sensitivity of predicted native states to variations in the interaction schemes." Protein Science **8**(2): 361-369.

Bhattacharya, S., C. G. Bunick, et al. (2004). "Target selectivity in EF-hand calcium binding proteins." Biochimica et Biophysica Acta **1742**: 69-79.

Bjelkmar, P., P. Larsson, et al. (2010). "Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models." Journal of Chemical Theory and Computation **6**(2): 459-466.

Briggs, G. E. and J. B. Haldane (1925). "A Note on the Kinetics of Enzyme Action." Biochemical Journal **19**(2): 338-339.

Bryngelson, J. D. and P. G. Wolynes (1987). "Spin glasses and the statistical mechanics of protein folding." Proc Natl Acad Sci U S A **84**(21): 7524-7528.

Carpenter, G. A. and S. Grossberg (1987). "ART 2: self-organization of stable category recognition codes for analog input patterns." Applied optics **26**(23): 4919-4930.

Case, D., T. Darden, et al. (2008). AMBER 10. University of California, San Francisco.

Cavalli, A., M. Vendruscolo, et al. (2005). "Comparison of sequence-based and structure-based energy functions for the reversible folding of a peptide." Biophysical journal **88**(5): 3158-3166.

Chance, B. (1940). "The accelerated flow method for rapid reactions." Journal of the Franklin Institute **229**(4): 455-476.
Chang, J. W., E. Schumacher, et al. (1997). "Dendritic translocation of RC3/neurogranin mRNA in normal aging, Alzheimer disease and fronto-temporal dementia." Journal of neuropathology and experimental neurology **56**(10): 1105-1118.
Cheung, M. S., J. M. Finke, et al. (2003). "Exploring the interplay between topology and secondary structural formation in the protein folding problem." Journal of Physical Chemistry B **107**(40): 11193-11200.
Cheung, M. S. and D. Thirumalai (2007). "Effects of crowding and confinement on the structures of the transition state ensemble in proteins." Journal of Physical Chemistry B **111**: 8250-8257.
Chin, D. and A. R. Means (2000). "Calmodulin: a prototypical calcium sensor." Trends Cell Biol **10**(8): 322-328.
Chodera, J. D., W. C. Swope, et al. (2007). "Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations." Journal of Chemical Theory and Computation **3**(1): 26-41.
Choi, J. and M. Husain (2006). "Calmodulin-mediated cell cycle regulation." Cell Cycle **5**: 2183-2186.
Clementi, C. (2008). "Coarse-grained models of protein folding: toy models or predictive tools?" Current Opinion in Structural Biology **18**(1): 10-15.
Colbran, R. J. and A. M. Brown (2004). "Calcium/calmodulin-dependent protein kinase II and synaptic plasticity." Current opinion in neurobiology **14**(3): 318-327.
Colomer, J. M. and A. R. Means (2000). "Chronic elevation of calmodulin in the ventricles of transgenic mice increases the autonomous activity of calmodulin-dependent protein kinase II, which regulates atrial natriuretic factor gene expression." Mol Endocrinol **14**(8): 1125-1136.
Crick, F. (1984). "Neurobiology - Memory and Molecular Turnover." Nature **312**(5990): 101-101.
Crivici, A. and M. Ikura (1995). "Molecular and structural basis of target recognition by calmodulin." Annual Review of Biophysics and Biomolecular Structure **24**: 85-116.
Daan, F. and S. Berend (2001). Understanding Molecular Simulation, Academic Press, Inc.
Darden, T., D. York, et al. (1993). "Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems." Journal of Chemical Physics **98**(12): 10089-10092.
Davidsson, P. and K. Blennow (1998). "Neurochemical dissection of synaptic pathology in Alzheimer's disease." International psychogeriatrics / IPA **10**(1): 11-23.
Debye, P. and E. Hückel (1923). "The theory of electrolytes. I. Lowering of freezing point and related phenomena." Physikalische Zeitschrift **24**: 185-206.
Deutch, J. M. and B. U. Felderho (1973). "Hydrodynamic Effect in Diffusion-Controlled Reaction." Journal of Chemical Physics **59**(4): 1669-1679.
Diez-Guerra, F. J. (2010). "Neurogranin, a link between calcium/calmodulin and protein kinase C signaling in synaptic plasticity." IUBMB life **62**(8): 597-606.
Dima, R. I. and D. Thirumalai (2004). "Asymmetry in the shapes of folded and denatured states of proteins." Journal of Physical Chemistry B **108**(21): 6564-6570.
Dunker, A. K., M. S. Cortese, et al. (2005). "Flexible nets. The roles of intrinsic disorder in protein interaction networks." The FEBS journal **272**(20): 5129-5148.

Dunker, A. K., J. D. Lawson, et al. (2001). "Intrinsically disordered protein." Journal of Molecular Graphics and Modelling **19**(1): 26-59.
Dunker, A. K. and Z. Obradovic (2001). "The protein trinity--linking function and disorder." Nat Biotechnol **19**(9): 805-806.
Dunker, A. K., Z. Obradovic, et al. (2000). "Intrinsic protein disorder in complete genomes." Genome Inform Ser Workshop Genome Inform **11**: 161-171.
Dunlap, T. B., J. M. Kirk, et al. (2013). "Thermodynamics of binding by calmodulin correlates with target peptide alpha-helical propensity." Proteins-Structure Function and Bioinformatics **81**(4): 607-612.
Elcock, A. H., R. R. Gabdoulline, et al. (1999). "Computer simulation of protein-protein association kinetics: Acetylcholinesterase-fasciculin." Journal of molecular biology **291**(1): 149-162.
Ermak, D. L. and J. A. Mccammon (1978). "Brownian Dynamics with Hydrodynamic Interactions." Journal of Chemical Physics **69**(4): 1352-1360.
Fischer, E. (1894). "Einfluss der Configuration auf die Wirkung der Enzyme." Ber Dtsch Chem Ges. **27**: 2984-2993.
Forest, A., M. T. Swulius, et al. (2008). "Role of the N- and C-lobes of calmodulin in the activation of Ca(2+)/calmodulin-dependent protein kinase II." Biochemistry **47**(40): 10587-10599.
Fuxreiter, M., I. Simon, et al. (2004). "Preformed structural elements feature in partner recognition by intrinsically unstructured proteins." Journal of molecular biology **338**(5): 1015-1026.
Gabdoulline, R. R. and R. C. Wade (1997). "Simulation of the diffusional association of Barnase and Barstar." Biophysical journal **72**(5): 1917-1929.
Gabdoulline, R. R. and R. C. Wade (2001). "Protein-protein association: Investigation of factors influencing association rates by Brownian dynamics simulations." Journal of molecular biology **306**(5): 1139-1155.
Gaertner, T. R., J. A. Putkey, et al. (2004). "RC3/Neurogranin and Ca2+/calmodulin-dependent protein kinase II produce opposing effects on the affinity of calmodulin for calcium." The Journal of biological chemistry **279**(38): 39374-39382.
Ganguly, D. and J. Chen (2011). "Topology-based modeling of intrinsically disordered proteins: Balancing intrinsic folding and intermolecular interactions." Proteins: Structure, Function, and Bioinformatics **79**(4): 1251-1266.
Ganguly, D., S. Otieno, et al. (2012). "Electrostatically accelerated coupled binding and folding of intrinsically disordered proteins." Journal of molecular biology **422**(5): 674-684.
Ganguly, D., W. Zhang, et al. (2012). "Synergistic folding of two intrinsically disordered proteins: searching for conformational selection." Mol Biosyst **8**(1): 198-209.
Gerendasy, D. D., S. R. Herron, et al. (1994). "Mutational and biophysical studies suggest RC3/neurogranin regulates calmodulin availability." The Journal of biological chemistry **269**(35): 22420-22426.
Gifford, J. L., H. Ishida, et al. (2011). "Fast methionine-based solution structure determination of calcium-calmodulin complexes." Journal of Biomolecular NMR **50**: 71-81.

Gifford, J. L., M. P. Walsh, et al. (2007). "Structures and metal-ion-binding properties of the Ca2+-binding helix-loop-helix EF-hand motifs." Biochemical Journal **405**: 199-221.

Grabarek, Z. (2006). "Structural basis for diversity of the EF-hand calcium-binding proteins." Journal of molecular biology **359**(3): 509-525.

Grossfield, A. (2013). "WHAM: an implementation of the weighted histogram analysis method." http://membrane.urmc.rochester.edu/content/wham **Version 2.0.4**.

Guo, Z. and D. Thirumalai (1997). "The nucleation-collapse mechanism in protein folding: evidence for the non-uniqueness of the folding nucleus." Folding and Design **2**(6): 377-391.

Halling, D. B., D. K. Georgiou, et al. (2009). "Determinants in CaV1 channels that regulate the Ca2+ sensitivity of bound calmodulin." The Journal of biological chemistry **284**(30): 20041-20051.

Harpaz, Y., N. Elmasry, et al. (1994). "Direct observation of better hydration at the N terminus of an alpha-helix with glycine rather than alanine as the N-cap residue." Proc Natl Acad Sci U S A **91**(1): 311-315.

Hartridge, H. and F. J. W. Roughton (1923). "A Method of Measuring the Velocity of Very Rapid Chemical Reactions." Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **104**(726): 376-394.

Heidorn, D. B., P. A. Seeger, et al. (1989). "Changes in the Structure of Calmodulin Induced by a Peptide Based on the Calmodulin-Binding Domain of Myosin Light Chain Kinase." Biochemistry **28**(16): 6757-6764.

Heppel, L. A. and R. J. Hilmoe (1951). "Purification of yeast inorganic pyrophosphatase." The Journal of biological chemistry **192**(1): 87-94.

Heppel, L. A. and R. J. Hilmore (1951). "Purification and properties of 5-nucleotidase." The Journal of biological chemistry **188**(2): 665-676.

Hess, B., H. Bekker, et al. (1997). "LINCS: A linear constraint solver for molecular simulations." Journal of Computational Chemistry **18**(12): 1463-1472.

Higo, J., Y. Nishimura, et al. (2011). "A Free-Energy Landscape for Coupled Folding and Binding of an Intrinsically Disordered Protein in Explicit Solvent from Detailed All-Atom Computations." Journal of the American Chemical Society **133**(27): 10448-10458.

Hoeflich, K. P. and M. Ikura (2002). "Calmodulin in action: diversity in target recognition and activation mechanisms." Cell **108**: 739-742.

Hoffman, L., A. Chandrasekar, et al. (2014). "Neurogranin alters the structure and calcium binding properties of calmodulin." The Journal of biological chemistry **289**(21): 14644-14655.

Homouz, D., M. Perham, et al. (2008). "Crowded, cell-like environment induces shape changes in aspherical protein." Proceedings of the National Academy of Sciences of the United States of America **105**(33): 11754-11759.

Homouz, D., H. Sanabria, et al. (2009). "Modulation of calmodulin plasticity by the effect of macromolecular crowding." Journal of molecular biology **391**: 933-943.

Huang, Y. and Z. Liu (2009). "Kinetic Advantage of Intrinsically Disordered Proteins in Coupled Folding–Binding Process: A Critical Assessment of the “Fly-Casting” Mechanism." Journal of molecular biology **393**(5): 1143-1159.

Huber, G. A. and J. A. McCammon (2010). "Browndye: A software package for Brownian dynamics." Computer Physics Communications **181**(11): 1896-1905.
Humphrey, W., A. Dalke, et al. (1996). "VMD: visual molecular dynamics." Journal of molecular graphics **14**(1): 33-38, 27-38.
Iakoucheva, L. M., C. J. Brown, et al. (2002). "Intrinsic disorder in cell-signaling and cancer-associated proteins." Journal of molecular biology **323**(3): 573-584.
Jackson, S. E. and A. R. Fersht (1991). "Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition." Biochemistry **30**(43): 10428-10435.
Jarzynski, C. (1997). "Nonequilibrium equality for free energy differences." Physical Review Letters **78**(14): 2690-2693.
Jorgensen, W. L., J. Chandrasekhar, et al. (1983). "Comparison of Simple Potential Functions for Simulating Liquid Water." Journal of Chemical Physics **79**(2): 926-935.
Jurado, L. A., P. S. Chockalingam, et al. (1999). "Apocalmodulin." Physiological reviews **79**(3): 661-682.
Karanicolas, J. and C. L. Brooks, 3rd (2002). "The origins of asymmetry in the folding transition states of protein L and protein G." Protein Sci **11**(10): 2351-2361.
Kastner, J. (2011). "Umbrella sampling." Wiley Interdisciplinary Reviews-Computational Molecular Science **1**(6): 932-942.
Kendrew, J. C., G. Bodo, et al. (1958). "A three-dimensional model of the myoglobin molecule obtained by x-ray analysis." Nature **181**(4610): 662-666.
Kim, Y. C., A. Bhattacharya, et al. (2014). "Macromolecular crowding effects on coupled folding and binding." The journal of physical chemistry. B **118**(44): 12621-12629.
Kim, Y. C., C. Tang, et al. (2008). "Replica exchange simulations of transient encounter complexes in protein-protein association." Proceedings of the National Academy of Sciences of the United States of America **105**(35): 12855-12860.
Kohagen, M., M. Lepsik, et al. (2014). "Calcium Binding to Calmodulin by Molecular Dynamics with Effective Polarization." The journal of physical chemistry letters **5**(22): 3964-3969.
Kohagen, M., P. E. Mason, et al. (2014). "Accurate description of calcium solvation in concentrated aqueous solutions." The journal of physical chemistry. B **118**(28): 7902-7909.
Koshland, D. E. J. (1958). "Application of a theory of enzyme specificity of protein synthesis. ." Proc Nat Acad Sci USA. **44**: 98-104.
Kretsinger, R. H. (1996). "EF-hands reach out." Nature structural biology **3**(1): 12-15.
Kubelka, J., J. Hofrichter, et al. (2004). "The protein folding 'speed limit'." Curr Opin Struct Biol **14**(1): 76-88.
Kuboniwa, H., N. Tjandra, et al. (1995). "Solution structure of calcium-free calmodulin." Nature structural biology **2**(9): 768-776.
Kubota, Y., J. A. Putkey, et al. (2008). "IQ-motif proteins influence intracellular free Ca2+ in hippocampal neurons through their interactions with calmodulin." Journal of neurophysiology **99**(1): 264-276.
Kudlay, A., M. S. Cheung, et al. (2009). "Crowding effects on the structural transitions in a flexible helical homopolymer." Physical Review Letters **102**(11): 118101.

Kumar, V., V. P. Chichili, et al. (2013). "Structural basis for the interaction of unstructured neuron specific substrates neuromodulin and neurogranin with Calmodulin." Scientific reports **3**: 1392.
Lakey, J. H. and E. M. Raggett (1998). "Measuring protein-protein interactions." Curr Opin Struct Biol **8**(1): 119-123.
Lee, K. H. and J. Chen (2016). "Multiscale enhanced sampling of intrinsically disordered protein conformations." Journal of Computational Chemistry **37**(6): 550-557.
Lepsik, M. and M. J. Field (2007). "Binding of calcium and other metal ions to the EF-hand loops of calmodulin studied by quantum chemical calculations and molecular dynamics simulations." The journal of physical chemistry. B **111**(33): 10012-10022.
Levinthal, C. (1968). "Are There Pathways for Protein Folding." Journal De Chimie Physique Et De Physico-Chimie Biologique **65**(1): 44-45.
Levy, Y., P. G. Wolynes, et al. (2004). "Protein topology determines binding mechanism." Proc Nat Acad Sci USA. **101**: 511-516.
Liedberg, B., C. Nylander, et al. (1983). "Surface-Plasmon Resonance for Gas-Detection and Biosensing." Sensors and Actuators **4**(2): 299-304.
Lindorff-Larsen, K., S. Piana, et al. (2010). "Improved side-chain torsion potentials for the Amber ff99SB protein force field." Proteins-Structure Function and Bioinformatics **78**(8): 1950-1958.
Lipari, G. and A. Szabo (1982). "Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity." Journal of the American Chemical Society **104**(17): 4546-4559.
Lisman, J., H. Schulman, et al. (2002). "The molecular basis of CaMKII function in synaptic and behavioural memory." Nature reviews. Neuroscience **3**(3): 175-190.
Llinas, R., J. A. Gruner, et al. (1991). "Regulation by synapsin I and Ca(2+)-calmodulin-dependent protein kinase II of the transmitter release in squid giant synapse." The Journal of physiology **436**: 257-282.
Lu, Q. and J. Wang (2008). "Single molecule conformational dynamics of adenylate kinase: energy landscape, structural correlations, and transition state ensembles." Journal of the American Chemical Society **130**: 4772-4783.
Ma, B., S. Kumar, et al. (1999). "Folding funnels and binding mechanisms." Protein Eng **12**(9): 713-720.
Marlow, M. S., J. Dogan, et al. (2010). "The role of conformational entropy in molecular recognition by calmodulin." Nature Chemical Biology **6**(5): 352-358.
Meador, W. E., A. R. Means, et al. (1993). "Modulation of calmodulin plasticity in molecular recognition on the basis of x-ray structures." Science **262**(5140): 1718-1721.
Miller, S. G. and M. B. Kennedy (1986). "Regulation of brain type II Ca2+/calmodulin-dependent protein kinase by autophosphorylation: a Ca2+-triggered molecular switch." Cell **44**(6): 861-870.
Mohan, R. R., G. A. Huber, et al. (2016). "Electrostatic Steering Accelerates C3d:CR2 Association." The journal of physical chemistry. B.
Monod, J., J. Wyman, et al. (1965). "On the nature of allosteric transitions: a plausible model." J. Mol. Biol. **12**: 88-118.

Moriguchi, S., F. Han, et al. (2006). "Decreased calcium/calmodulin-dependent protein kinase II and protein kinase C activities mediate impairment of hippocampal long-term potentiation in the olfactory bulbectomized mice." Journal of neurochemistry **97**(1): 22-29.

Northrup, S. H., S. A. Allison, et al. (1984). "Brownian Dynamics Simulation of Diffusion-Influenced Bimolecular Reactions." Journal of Chemical Physics **80**(4): 1517-1526.

Northrup, S. H., J. O. Boles, et al. (1988). "Brownian Dynamics of Cytochrome-C and Cytochrome-C Peroxidase Association." Science **241**(4861): 67-70.

Northrup, S. H. and H. P. Erickson (1992). "Kinetics of Protein Protein Association Explained by Brownian Dynamics Computer-Simulation." Proceedings of the National Academy of Sciences of the United States of America **89**(8): 3338-3342.

Northrup, S. H. and J. T. Hynes (1979). "Short-Range Caging Effects for Reactions in Solution .1. Reaction-Rate Constants and Short-Range Caging Picture." Journal of Chemical Physics **71**(2): 871-883.

Northrup, S. H. and J. T. Hynes (1979). "Short-Range Caging Effects for Reactions in Solution .2. Escape Probability and Time-Dependent Reactivity." Journal of Chemical Physics **71**(2): 884-893.

Nymeyer, H., N. D. Socci, et al. (2000). "Landscape approaches for determining the ensemble of folding transition states: success and failure hinge on the degree of frustration." Proc Natl Acad Sci U S A **97**(2): 634-639.

O'Donnell, S. E., L. Yu, et al. (2011). "Recognition of beta-calcineurin by the domains of calmodulin: thermodynamic and structural evidence for distinct roles." Proteins-Structure Function and Bioinformatics **79**(3): 765-786.

Okazaki, K., N. Koga, et al. (2006). "Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations." Proceedings of the National Academy of Sciences of the United States of America **103**(32): 11844-11849.

Okazaki, K. I. and S. Takada (2008). "Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms." Proceedings of the National Academy of Sciences of the United States of America **105**(32): 11182-11187.

Oldfield, C. J., Y. Cheng, et al. (2005). "Coupled folding and binding with alpha-helix-forming molecular recognition elements." Biochemistry **44**(37): 12454-12470.

Olwin, B. B., A. M. Edelman, et al. (1984). "Quantitation of energy coupling between Ca2+, calmodulin, skeletal muscle myosin light chain kinase, and kinase substrates." The Journal of biological chemistry **259**(17): 10949-10955.

Onuchic, J. N., Z. Luthey-Schulten, et al. (1997). "Theory of protein folding: the energy landscape perspective." Annual review of physical chemistry **48**: 545-600.

Onuchic, J. N. and P. G. Wolynes (2004). "Theory of protein folding." Curr Opin Struct Biol **14**(1): 70-75.

Pak, J. H., F. L. Huang, et al. (2000). "Involvement of neurogranin in the modulation of calcium/calmodulin-dependent protein kinase II, synaptic plasticity, and spatial learning: a study with knockout mice." Proceedings of the National Academy of Sciences of the United States of America **97**(21): 11232-11237.

Pang, Z. P., P. Cao, et al. (2010). "Calmodulin controls synaptic strength via presynaptic activation of calmodulin kinase II." The Journal of neuroscience : the official journal of the Society for Neuroscience **30**(11): 4132-4142.

Parrinello, M. and A. Rahman (1981). "Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method." Journal of Applied Physics **52**(12): 7182-7190.

Peersen, O. B., T. S. Madsen, et al. (1997). "Intermolecular tuning of calmodulin by target peptides and proteins: differential effects on Ca2+ binding and implications for kinase activation." Protein Sci **6**(4): 794-807.

Plaxco, K. W., K. T. Simons, et al. (1998). "Contact order, transition state placement and the refolding rates of single domain proteins." Journal of molecular biology **277**(4): 985-994.

Price, E. S., M. S. DeVore, et al. (2010). "Detecting intramolecular dynamics and multiple Forster Resonance Energy Transfer states by Fluorescence Correlation Spectroscopy." J. Phys. Chem. B. **114**: 5895-5902.

Putkey, J. A. and M. N. Waxham (1996). "A peptide model for calmodulin trapping by calcium/calmodulin-dependent protein kinase II." The Journal of biological chemistry **271**(47): 29619-29623.

Putkey, J. A., M. N. Waxham, et al. (2008). "Acidic/IQ motif regulator of calmodulin." The Journal of biological chemistry **283**(3): 1401-1410.

Ran, X., H. H. Miao, et al. (2003). "Structural and dynamic characterization of a neuron-specific protein kinase C substrate, neurogranin." Biochemistry **42**(17): 5143-5150.

Ratcliff, G. C. and D. A. Erie (2001). "A Novel Single-Molecule Study To Determine Protein−Protein Association Constants." Journal of the American Chemical Society **123**(24): 5632-5635.

Ravikumar, K. M., W. Huang, et al. (2012). "Coarse-grained simulations of protein-protein association: an energy landscape perspective." Biophysical journal **103**(4): 837-845.

Rhoads, A. R. and F. Friedberg (1997). "Sequence motifs for calmodulin recognition." FASEB J **11**(5): 331-340.

Rogers, J. M., C. T. Wong, et al. (2014). "Coupled folding and binding of the disordered protein PUMA does not require particular residual structure." Journal of the American Chemical Society **136**(14): 5197-5200.

Roux, B. (1995). "The Calculation of the Potential of Mean Force Using Computer-Simulations." Computer Physics Communications **91**(1-3): 275-282.

Samiotakis, A., D. Homouz, et al. (2010). "Multiscale investigation of chemical interference in proteins." The Journal of chemical physics **132**(17): 175101.

Schreiber, G., G. Haran, et al. (2009). "Fundamental aspects of protein-protein association kinetics." Chemical Review **109**: 839-860.

Schrodinger, LLC (2015). The PyMOL Molecular Graphics System, Version 1.8.

Schumacher, M. A., A. F. Rivard, et al. (2001). "Structure of the gating domain of a Ca2+-activated K+ channel complexed with Ca2+/calmodulin." Nature **410**(6832): 1120-1124.

Shoemaker, B. A., J. J. Portman, et al. (2000). "Speeding molecular recognition by using the folding funnel: The fly-casting mechanism." Proceedings of the National Academy of Sciences of the United States of America **97**(16): 8868-8873.

Slaughter, B. D., M. W. Allen, et al. (2004). "Single-molecule resonance energy transfer and fluorescence correlation spectroscopy of calmodulin in solution." Journal of Physical Chemistry B **108**(29): 10388-10397.

Slaughter, B. D., R. J. Bieber-Urbauer, et al. (2005). "Single-molecule tracking of sub-milisecond domain motion in calmodulin." J. Phys. Chem. B. **109**: 12658-12662.

Smock, R. G. and L. M. Gierasch (2009). "Sending signals dynamically." Science **324**(5924): 198-203.

Smoluchowski, M. V. (1916). "Drei Vortrage uber Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen." Physik. Zeit. **17**: 557-585.

Soderling, T. R. (1999). "The Ca-calmodulin-dependent protein kinase cascade." Trends in Biochemical Sciences **24**(6): 232-236.

Song, J., S. C. Ng, et al. (2013). "Polycation-pi interactions are a driving force for molecular recognition by an intrinsically disordered oncoprotein family." PloS Computational Biology **9**(9): e1003239.

Sorensen, B. R. and M. A. Shea (1998). "Interactions between domains of apo calmodulin alter calcium binding and stability." Biochemistry **37**(12): 4244-4253.

Sparrow, M. P., U. Mrwa, et al. (1981). "Calmodulin is essential for smooth muscle contraction." FEBS letters **125**(2): 141-145.

Stewart, J. J. (2007). "Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements." J Mol Model **13**(12): 1173-1213.

Stigler, J., F. Ziegler, et al. (2011). "The complex folding network of single calmodulin molecules." Science **334**(6055): 512-516.

Sugase, K., H. J. Dyson, et al. (2007). "Mechanism of coupled folding and binding of an intrinsically disordered protein." Nature **447**(7147): 1021-1025.

Sugita, Y. and Y. Okamoto (1999). "Replica-exchange molecular dynamics method for protein folding." Chemical Physics Letters **314**(1-2): 141-151.

Swulius, M. T. and M. N. Waxham (2008). "Ca(2+)/calmodulin-dependent protein kinases." Cellular and molecular life sciences : CMLS **65**(17): 2637-2657.

Szollosi, D., T. Horvath, et al. (2014). "Discrete molecular dynamics can predict helical prestructured motifs in disordered proteins." PloS one **9**(4): e95795.

Tang, C., J. Iwahara, et al. (2006). "Visualization of transient encounter complexes in protein-protein association." Nature **444**(7117): 383-386.

Theoharis, N. T., B. R. Sorensen, et al. (2008). "The neuronal voltage-dependent sodium channel type II IQ motif lowers the calcium affinity of the C-domain of calmodulin." Biochemistry **47**(1): 112-123.

Tidow, H. and P. Nissen (2013). "Structural diversity of calmodulin binding to its target sites." The FEBS journal **280**(21): 5551-5565.

Tompa, P. and M. Fuxreiter (2008). "Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions." Trends in Biochemical Sciences **33**(1): 2-8.

Torrie, G. M. and J. P. Valleau (1977). "Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling." Journal of Computational Physics **23**(2): 187-199.

Tripathi, S. and J. J. Portman (2009). "Inherent flexibility determines the transition mechanisms of the EF-hands of calmodulin." Proceedings of the National Academy of Sciences of the United States of America **106**(7): 2104-2109.

Tripathi, S., Q. Wang, et al. (2015). "Conformational frustration in calmodulin-target recognition." J Mol Recognit **28**(2): 74-86.

Tsai, C. J., S. Kumar, et al. (1999). "Folding funnels, binding funnels, and protein function." Protein Science **8**: 1181-1190.

Turjanski, A. G., J. S. Gutkind, et al. (2008). "Binding-induced folding of a natively unstructured transcription factor." PloS Computational Biology **4**(4): e1000060.

Uversky, V. N. (2002). "Natively unfolded proteins: A point where biology waits for physics." Protein Science **11**(4): 739-756.

Uversky, V. N., J. R. Gillespie, et al. (2000). "Why are "natively unfolded" proteins unstructured under physiologic conditions?" Proteins-Structure Function and Bioinformatics **41**(3): 415-427.

Uversky, V. N., C. J. Oldfield, et al. (2005). "Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling." J Mol Recognit **18**(5): 343-384.

Veitshans, T., D. Klimov, et al. (1997). "Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties." Folding & design **2**(1): 1-22.

Walsh, M. P. (1994). "Calmodulin and the regulation of smooth muscle contraction." Molecular and cellular biochemistry **135**(1): 21-41.

Wang, J., Y. Wang, et al. (2011). "Multi-scaled explorations of binding-induced folding of intrinsically disordered protein inhibitor IA3 to its target Enzyme." PloS Computational Biology **7**: e1001118.

Wang, Q., K. C. Liang, et al. (2011). "The Effect of Macromolecular Crowding, Ionic Strength and Calcium Binding on Calmodulin Dynamics." PloS Computational Biology **7**(7): 16.

Wang, Q., P. Zhang, et al. (2013). "Protein recognition and selection through conformational and mutually induced fit." Proc Natl Acad Sci U S A **110**(51): 20545-20550.

Wang, X., L. W. Xiong, et al. (2013). "The calmodulin regulator protein, PEP-19, sensitizes ATP-induced Ca2+ release." The Journal of biological chemistry **288**(3): 2040-2048.

Ward, J. J., J. S. Sodhi, et al. (2004). "Prediction and functional analysis of native disorder in proteins from the three kingdoms of life." Journal of molecular biology **337**(3): 635-645.

Waxham, M. N., A. L. Tsai, et al. (1998). "A mechanism for calmodulin (CaM) trapping by CaM-kinase II defined by a family of CaM-binding peptides." Journal of Biological Chemistry **273**: 17579-17584.

Weinkam, P., J. Pons, et al. (2012). "Structure-based model of allostery predicts coupling between distant sites." Proceedings of the National Academy of Sciences of the United States of America **109**(13): 4875-4880.

Whitford, P. C., O. Miyashita, et al. (2007). "Conformational transitions of adenylate kinase: Switching by cracking." Journal of molecular biology **366**(5): 1661-1671.

Whitlock, J. R., A. J. Heynen, et al. (2006). "Learning induces long-term potentiation in the hippocampus." Science **313**(5790): 1093-1097.
Williamson, J. R. (2000). "Induced fit in RNA-protein recognition." Nature structural biology **7**: 834-837.
Wolynes, P. G. and J. M. Deutch (1976). "Slip Boundary-Conditions and Hydrodynamic Effect on Diffusion Controlled Reactions." Journal of Chemical Physics **65**(1): 450-454.
Wright, P. E. and H. J. Dyson (1999). "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm." Journal of molecular biology **293**(2): 321-331.
Xia, Z. and D. R. Storm (2005). "The role of calmodulin as a signal integrator for synaptic plasticity." Nature reviews. Neuroscience **6**(4): 267-276.
Yamauchi, T. (2005). "Neuronal Ca2+/calmodulin-dependent protein kinase II--discovery, progress in a quarter of a century, and perspective: implication for learning and memory." Biological & pharmaceutical bulletin **28**(8): 1342-1354.
Yamniuk, A. P. and H. J. Vogel (2004). "Calmodulin's flexibility allows for promiscuity in its interactions with target proteins and peptides." Molecular Biotechnology **27**: 33-57.
Yang, Y., E. Faraggi, et al. (2011). "Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates." Bioinformatics **27**(15): 2076-2082.
Ytreberg, F. M. and D. M. Zuckerman (2004). "Efficient use of nonequilibrium measurement to estimate free energy differences for molecular systems." Journal of Computational Chemistry **25**(14): 1749-1759.
Zhang, D., J. Gullingsrud, et al. (2006). "Potentials of mean force for acetylcholine unbinding from the alpha7 nicotinic acetylcholine receptor ligand-binding domain." Journal of the American Chemical Society **128**(9): 3019-3026.
Zhang, Y., H. Tan, et al. (2008). "Ca(2+) dissociation from the C-terminal EF-hand pair in calmodulin: a steered molecular dynamics study." FEBS letters **582**(9): 1355-1361.
Zhong, L. and N. Z. Gerges (2010). "Neurogranin and synaptic plasticity balance." Communicative & integrative biology **3**(4): 340-342.
Zientara, G. P., J. A. Nagy, et al. (1982). "Dynamics of Protein Domain Coalescence .2." Journal of Physical Chemistry **86**(5): 824-832.
Zuckerman, D. M. (2004). "Simulation of an ensemble of conformational transitions in a united-residue model of calmodulin." Journal of Physical Chemistry B **108**(16): 5127-5137.
Zwanzig, R., A. Szabo, et al. (1992). "Levinthal's paradox." Proc Natl Acad Sci U S A **89**(1): 20-22.