Assessing the Role of CaM's Linker Flexibility in Target Binding

Bin Sun[1] and Peter M. Kekenes-Huskey[*,1]

[1] Department of Cell and Molecular Physiology, Loyola University Chicago, Maywood, IL, USA 60153
* Corresponding author. E-mail address: pkekeneshuskey@luc.edu (P. Kekenes-Huskey)
March 15, 2021

1 Abstract

Calmodulin (CaM) is a universal Ca^{2+} binding protein known to bind at least 300 targets. The selectivity and specificity towards these targets are partially attributed to the protein's flexible alpha-helical linker that connects its N- and C- domains. However, how this flexible linker mediates the driving forces guiding CaM's binding to regulatory targets is not well-established. Therefore, we utilized coarse-grained (CG) Martini molecular dynamics simulations to probe interrelationships between CaM/target assembly and the role of its linker region. As a model system, we simulated the binding of CaM to the CaM binding region (CaMBR) of calcineurin (CaN). The simulations were conducted assuming a 'wild-type' calmodulin with normal flexibility of its linker between the N- and C-terminal domains, as well as a labile, highly flexible linker variant. For the wild-type model, 98% of the 600 simulations across three ionic strengths adopted a tightly-bound complex within 2 μs of simulation time; of these, 1.7% sampled the fully-bound state observed in experimentally-determined crystallographic structure. By calculating the mean-first-passage-time for these simulations, we estimated the association rate to be $k_a = 6.3 \times 10^8\,\mathrm{M}^{-1}\,\mathrm{s}^{-1}$, which is similar to the experimentally-determined rate of $2.2 \times 10^8\,\mathrm{M}^{-1}\,\mathrm{s}^{-1}$ [1]. Further, our simulations recapitulated the inverse relationship between the association rate and solution ionic strength reported in the literature. In contrast, although over 97% of the labile linker simulations formed tightly-bound complexes, only 0.3% achieved the fully-bound configuration and its ionic strength dependence is attenuated. These effects appear to stem from a difference in the ensembles of extended and collapsed states controlled by the linker properties. Specifically, the labile linker variant samples fewer extended states compatible with target peptide binding. Therefore, our simulations suggest that variations in the CaM linker's propensity for alpha-helical secondary structure can modulate the kinetics of target binding. This finding is important, given that some CaM variants found in the human population and post-translational modifications sites fall within this linker region, which may alter the protein's normal regulatory

functions.

2 Introduction

Calmodulin (CaM) is a highly-expressed, 16.7 kDa globular protein comprising two domains connected by a linker (Fig. 1). CaM regulates hundreds of protein targets [2] in a Ca^{2+}-dependent manner. How CaM maintains selectivity towards these targets with varying affinities (K_d from nM [3] to μM [4]) has been studied for decades [5–16]. These studies have generated valuable insights into factors contributing to its binding selectivity, which includes hydrophobic anchoring residues from targets that interact with CaM, CaM's conformationl heterogeneity at the binding surface and its Ca^{2+}-binding sensitivity. Of these, the flexibility of its linker is believed to play a prominent role in shaping the conformational ensemble it adopts [17, 18] as well as its regulation of target enzymes [19].

A key basis for CaM's selectivity is attributed to the variety of binding modes it adopts when bound to its mostly peptide-based targets (reviewed in [20]). These include an extended conformation, a collapsed conformation in which its N- and C- domains wrap around the target, or intermediate configurations that permit CaM/target stoichiometries of 1:1, 2:2, or 2:1. Therefore, it is apparent that CaM's ability to bind various targets in part stems from the linker's ability to assume different conformations [21, 22]. Further, this flexible linker is believed to exert an entropic role in tuning target affinity, as target-binding induced mobility changes of linker residues correlate with the conformational entropy measured for the entire protein [22]. As an example, Katyal et al [23] demonstrated that tethering the N- and C- CaM termini via disulfide bond increases the entropic favorability of binding by reducing the ensemble of thermodynamically-accessible states available to the linker.

Nonetheless, while many thermodynamic details of CaM/target binding are increasingly understood, the protein's binding kinetics remain enigmatic. Toward addressing this gap, molecular dynamics (MD) simulations have commonly

been used to probe the CaM/target binding process. One of the prominent challenges in these simulations is the fact that most CaM targets are intrinsically disordered peptides (IDP) [24]. Such IDPs adopt stable secondary structures upon binding through "coupled binding and folding" mechanisms that are case-dependent [25] and generally a combination of conformational-selection and induced-fit [26, 27] binding strategies. IDP binding mechanisms can unravel over micro-seconds and longer time scales that are generally inaccessible to conventional all-atom protein descriptions [28]. Efforts to bridge this limitation include adaptive sampling techniques such as goal-oriented sampling, weighted ensemble (WE) and coarse-grained techniques, which have been used to study the process of IDP/DNA binding [29], globular proteins binding [30] and and protein/ligand binding [31]. Of these approaches, coarse-grained (CG) molecular dynamics simulations are attractive due to their sampling efficiency and ability to preserve important molecular details of IDP-binding [32–36]. Such simulations could provide important insights into the time-dependent nature of CaM/target assembly, but have generally been limited to study their mechanisms instead of rate [12, 33]. Resolving calmodulin (CaM) binding kinetics via coarse-grained modeling could shed insight into fundamental properties that determine CaM's ability to regulate targets. This strategy could be important in resolving how known missense mutations [37] and post-translation modifications [38] in CaM's linker domain may influence its function.

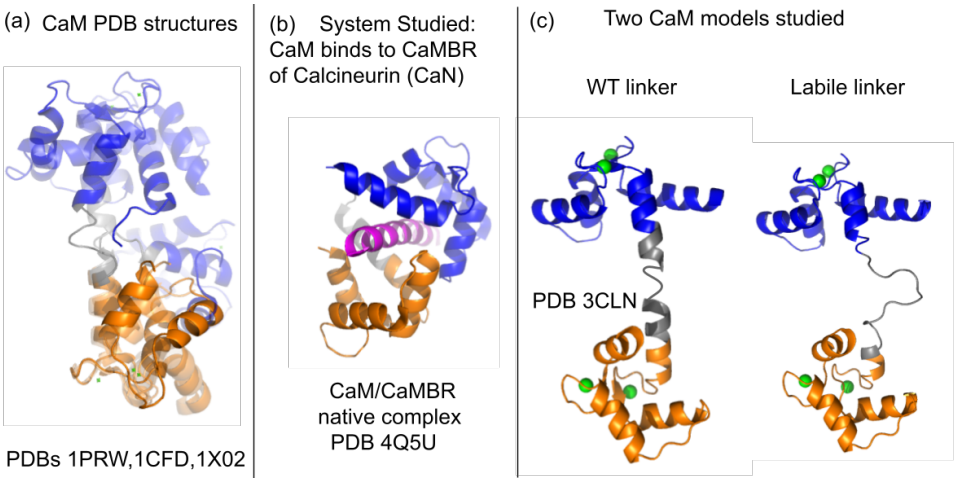

Figure 1: a) Three CaM crystal structures aligned on C-domain to show the native flexibility of the linker [2]. The N- and C- domains are colored blue and orange, respectively. The central linker (residues 73 to 87) is colored gray. The PDBs of these structures are 1PRW, 1CFD and 1X02. (b) The process of CaM binding to the CaM binding region (CaMBR) of calcineurin (CaN) is studied in this work. (c) The two CaM models that has a WT linker and a labile, highly-flexibilble linker, respectively, were studied in present work..

In this study, we simulated the binding process of a model system, CaM and the CaM binding region (CaMBR) of calcineurin, using a Martini CG model with explicit solvents and ions. This permitted us to sample native-like CaM/CaMBR complex structures from unbiased sampling over a range of ionic strengths and under different linker flexibilities. Using these simulations, we have identified a mechanistic basis for how CaM linker flexibility shapes the kinetics of target binding and its dependence on the solvent ionic strength.

3 Materials and methods

3.1 Martini CG simulations

We elected to use Martini CG given its advantages over the 'structure based' model (SBM, also called G$\bar{o}$-like model [39]). 1) Martini CG has higher molecular resolution than G$\bar{o}$-based CG model. The Martini CG mapping ratio is on average four atoms to one CG bead [40] while the commonly used G$\bar{o}$-based model are about ten non-hydrogen atoms to one bead [41]. This higher resolution permits detailed description of protein side-chain interactions with waters, which can impact protein mobility [42]. 2) No reference-structure based potential is needed, therefore the Martini CG potential have better transferability to arbitrary proteins without refitting and the simulation time scale can be directly interpreted. Therefore, Martini CG is well-suited for modeling the kinetics of CaM/target binding.

To this end, we established a Martini CG model based on the extended CaM structure PDB 3CLN [43]. The CaMBR of CaN is extracted from the CaM/-CaMBR complex PDB 4Q5U [44]. Both CaM and CaMBR structures were used to construct the CG model with the Martini protein force field V2.2 [45, 46]. Martini CG's time scale is four times faster than all-atom simulations [47]. The mapping ratio we used assumed four heavy atoms to one bead on average. Martini CG does not typically sample protein secondary structure thus it relies on user-provided secondary structure information to assign proper backbone

parameters of bonds, angle, and dihedral terms [45]. Therefore, we calculated the secondary structure of the CaM and CaMBR via the DSSP program [48, 49]. We constructed two CaM models, one has the WT linker (residues 73-87) flexibility and one has a labile, highly flexible linker (Fig. 2). The WT linker is based on PDB 3CLN and has a 4-residue hinge (residues 78-81) with DSSP predicted secondary structure as turns ("TTTT"). The labile linker in which the entire linker was subject to a short annealing MD in vacuum by Amber [50] to destroy the alpha-helix structure with DSSP predicted secondary structure as coils and bend ("CCCSCCSSSCCCSSS" where "C" and "S" refers to coil and bend, respectively). After defining the system, the CaMBR was randomly placed around CaM via the GMX INSERT-MOLECUES command with a minimum distance between CaMBR and CaM of 40 Å to minimize bias.

For each system, 200 trials were run with ionic strengths of 0, 0.15, and 0.5M NaCl, respectively. Elastic constraints within CaM's N-/C- domains were introduced to maintain an open domain conformation [51], which was justified by the observations that the N-/C- domain conformations are unchanged after binding CaMBR ($<$ 2 Å RMSD). The Martini V2.0 ion force field was used to describe the NaCl ions added into the system and appropriate numbers were added to set the aforementioned ionic strengths. The system was solvated into a cubic water box with dimension of 150 Å per side using the GMX SOLVATE command with pre-equilibrated waters from the Martini website. The standard Martini water model was used, in which four all-atom water molecules are combined into a single coarse-grained bead [52]. The system was first subject to 2000 steps of energy minimization using the steepest descent algorithm with constraints on the protein. The minimized system was equilibrated after being heated to 300 K over 6 ns. Constraints were imposed on the proteins during the heating stage. A 2 μs production run was initiated from the equilibrated system in the NPT ensemble using a 30 fs time step. The Berendsen temperature and pressure couplings were used to maintain 300 K and standard pressure. All simulations were performed using GROMACS version 2020.3 [53]. The back mapping from CG model to the all-atom model was done according to the procedure proposed

in [54].

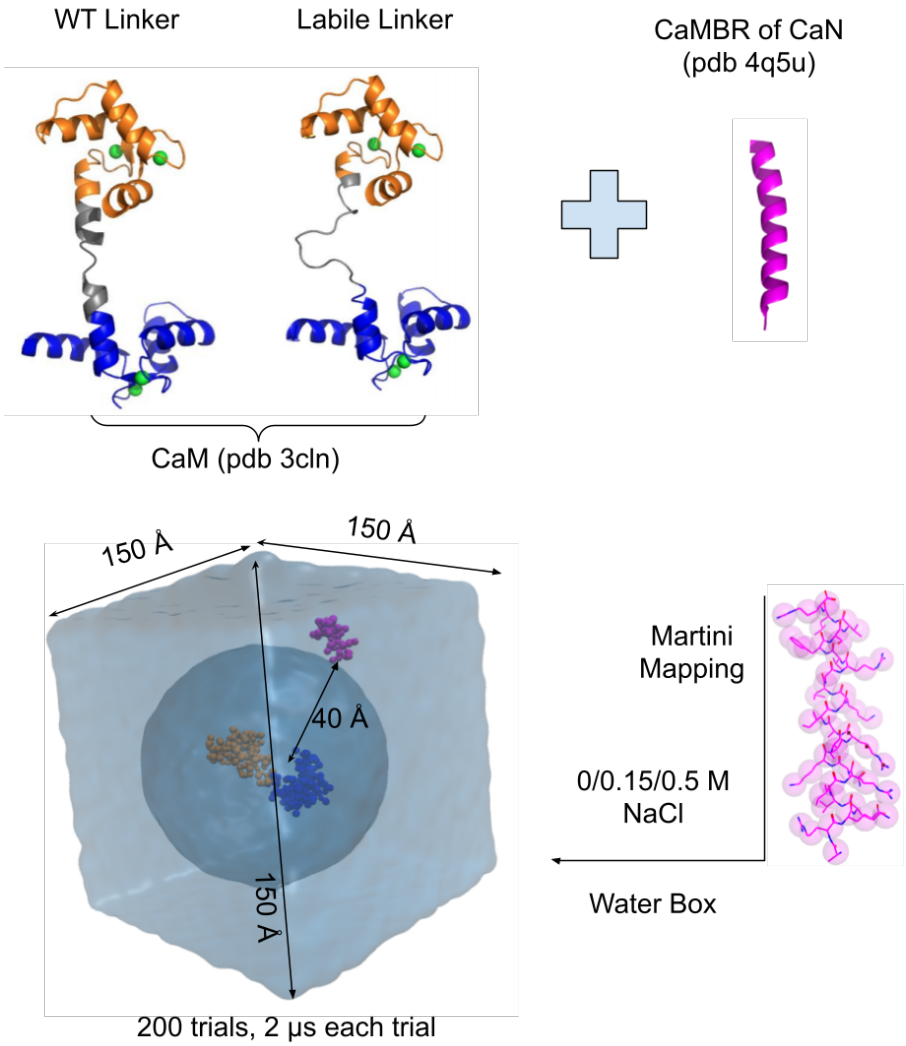

Figure 2: Martini CG setup. The all-atom CaM and CaMBR of CaN structures are from PDB 3CLN and PDB 4Q5U, respectively. Two CaM models with WT linker and labile linker were considered. After mapping the all-atom structures into Martini CG structures, the CaMBR was randomly placed around CaM with minimum distance >40 Å. The binding process was simulated by molecular dynamics without any further constraints imposed on the proteins. .

3.2 Analyses

The AUTOIMAGE command from the CPPTRJ [55] program that centers and images the trajectory was used for periodic boundary condition treatment. The trajectory was then converted to PDB format with bonds added using the GMX TRJCONV command of Gromacs. The PDB format trajectory was used for all analyses and visualization. The contacts between CaMBR and CaM were calculated assuming that one contact represents as any pair of beads that is within 5.5 Å. The CaMBR/CaM center of mass distance, RMSD to the native-like complex and CaM's radius of gyration (R_G) were calculated via the CPPTRAJ program. The CG structure of PDB 4Q5U served as a reference structure for the RMSD calculation. The trajectories were projected onto a plane according to the CaM/CaMBR center of mass distance and RMSD relative to the reference structure. The projection densities were estimated by a Gaussian kernels using the GAUSSIAN_KDE from the scipy python library. The potential was estimated via Boltzmann inversion $-k_b T \ln(\rho/\rho_{min})$ where k_b is Boltzmann's constant, T is temperature, ρ is the point density after projection and ρ_{min} is the minimum density.

3.3 Association rate calculations based on first passage time to bound state

CaM and CaMBR were deemed bound when the structures assumed 50 or more contacts. We later justify that CaMBR/CaM complexes by this definition are thermodynamically favourable and located near the fully-bound state (Fig. 5a). The first passage time (T_{fp}) of reaching the bound state can then be used to estimate the association rate (k_a) [56]. To illustrate the concept, we use the

following expression:

$$\text{CaMBR} + \text{CaM} \xrightarrow{\text{k}_a} \text{CaMBR} \cdot \text{CaM} \tag{1}$$

$$\frac{d[\text{CaMBR} \cdot \text{CaM}]}{dt} = k_a[\text{CaMBR}][\text{CaM}] \tag{2}$$

$$[\text{CaMBR}] = [\text{CaM}] = [c] \tag{3}$$

$[c] = 1/(N_A V_{box})$ is the equivalent concentration of one molecule in the simulation box with volume $V_{box} = 3.375 \times 10^{-21}$ L and N_A is Avogadro's constant. In unit time ($dt = 1$), there are $1/T_{fp}$ association events, for which each event leads to the formation of CaMBR.CaM. Thus $d[\text{CaMBR} \cdot \text{CaM}] = (1/T_{fp})[c] = k_a[c][c]$ which leads to:

$$k_a = \frac{1}{T_{fp}} \frac{V_{box}}{1/N_A} \tag{4}$$

Hence, k_a is essentially the inverse of T_{fp} [56]. We used bootstrapping with replacement [57] to calculate the T_{fp} from our simulations. For this, we randomly generated 1000 distributions of T_{fp} values using the original 200 values generated by the simulations. Each of those 1000 distributions were fitted to an exponential distribution $A \exp(-\tau x)$, which is analogous to a Poisson process undergoing rare (binding) events. The expected value for T_{fp}, $\mathbf{E} = 1/\tau$, is reported. Scripts for these analyses are available in https://github.com/huskeypm/pkh-lab-analyses (2021-CaMBRmartini)

4 Results and Discussion

4.1 Linker flexibility impacts CaM/CaMBR assembly

To investigate how the CaM's linker flexibility affects the binding process between CaM and the CaMBR of CaN, we performed extensive binding simulations using the Martini CG models with two CaM models that have WT and labile linker flexibility, respectively. The control of linker flexibility is achieved

by setting the secondary structural properties of the linker. Hence a linker with a higher helical character is more rigid than a coil. For the target-free CaM crystal structure (PDB 3CLN), the linker (residues 73-87) was anticipated to have moderate flexibility. This was based on the DSSP result [48, 49] suggesting that a four-residue 'hinge' of consisting of residues 78-81 assumed 'turn' (T) secondary structure properties, whereas the remainder were ascribed rigid alpha-helical character. For the labile linker we assumed the entire linker consisted of labile coils and turns. To study the contributions of long-range electrostatic interactions in driving diffusion[58, 59], we performed 200 simulation trials at 0, 0.15, and 0.5 M ionic strength, respectively. The trajectories were projected along two axis: a) the CaMBR/CaM center of mass distance b) the RMSD to native CaM/CaMBR complex crystal structure PDB 4Q5U (Fig. 3) to visualize the sampled conformational space. The population densities in the projected space was used to infer the potential of mean force via inverting the Boltzmann equation, e.g., $E_{PMF} = -k_b T \ln(\rho/\rho_{min})$ where k_b is Boltzmann's constant, T is temperature, ρ is the point density after projection and ρ_{min} is the minimum density.

Our simulations indicate that each system configuration exhibited significant sampling of a loosely-bound state, which is evidenced by the region of more negative E_{PMF} values found where the RMSD<20 Å and CaMBR/CaM distance< 35 Å. We note that the loosely-bound states most prominently sample a region modestly displaced from the reference crystal structure. We also observed some binding events that lead to the native-like bound complex. The resulting structures have an average RMSD of 6.3 Å to the crystal structure of the CaMBR/CaM complex; although this value is perceived as relatively large for atomistic-resolution structures, by visual inspection (Fig. S2) it is apparent that the trial poses strongly resemble the native structure. Additionally, the ∼6 kcal/mol potential difference of this region relative to the unbound state indicates that the binding process is thermodynamically favourable. Interestingly, a bottle-neck is also apparent when the RMSD is within $25 - 40$Å and the CaMBR/CaM distance is between $60 - 100$ Å for non-zero ionic strengths

(see the dashed circles in Fig. 3a). We attribute this bottle-neck region to two factors: 1) Electrostatic screening by ions and 2) impaired CaM/CN alignment. We show in Fig. S3 that the electrostatic potentials about CaM and CaMBR are prominent, complementary and nonuniform. The largely negative electrostatic potential presented by CaM can be expected to facilitate the binding of the positively-charged CaMBR peptide, as is well-established in other protein/protein complexes [58]. However, the negatively-charged potential can also drive off-target associations of the two proteins, that compete with binding to the native bound configuration. With increasing ionic strength, it appears that the binding to the native bound configuration is disfavored, which manifests in a greater population of CaMBR/CaM binding poses that are off-target.

The nonuniformity of the CaM electrostatic potential also appears to impose constraints on the CaMBR's 'angle of approach' (Fig. 3c). This is shown by the asymmetry in the distribution of CaMBR configurations about CaM. In other words, the nonuniform electrostatic potential of CaM and its highly labile conformational ensemble necessitates proper alignment of CaM with the CaMBR to facilitate productive binding. This mechanism of constraining the angle of approach has previously been observed in lysozyme/α-lactalbumin assembly [60]. In that study, it was shown the binding process for two charged protein substrates a strong preference for a narrow set of approach angles [60, 61]. Further, the tendency for the two substrates to align decreased with increasing ionic strength. Therefore, we anticipate that the interplay of electrostatics and protein/protein alignment for CaM/CaMBR gives rise to a 'funneled' landscape that is consistent with the Lund et al study [61].

CaM's linker flexibility appears to impact the degree to which the bottle-neck region is favored as the ionic strength is altered. In the WT linker CaM model, as the ionic strength is increased from 0.15 to 0.5 M, we see a strong redistribution of states toward the bottle-neck region. For the labile linker, this change is much less pronounced. We believe this is because the labile linker imposes less significant constraints on the CaM/CaMBR angle of approach. This is evident based on the distribution of CaMBR configurations about CaM (Fig. 3c) that

fall within the bottle neck region. Notably, the CaMBR configurations are more diffusely and uniformly distributed about CaM, which evidences the broader angles of approach afforded for a labile linker relative to thw WT. We will further discuss how this distribution shapes association rates between CaM and CaMBR in Sect. 4.3.

We also report for WT and labile linker CaM models the number of binding events that lead to fully-bound configuration consistent with the reference crystal structure. We identify this bound state as the region near (0,0) of the native RMSD and CaMBR/CaM distance projection plane. For the WT linker CaM model, we observe ten events relative to only two for the labile linker model, out of ∼590 trials culminating in loosely-associated assemblies. This shows that an abnormally highly flexible linker significantly reduces the probability of achieving the native-like bound state. It is worth speculating that this reduction in probability may have consequences in CaM's ability to regulate its targets. This interpretation is supported by observations that CaM can bind to its targets in different structural states but only a subset of structural states can activate the target enzyme [62].

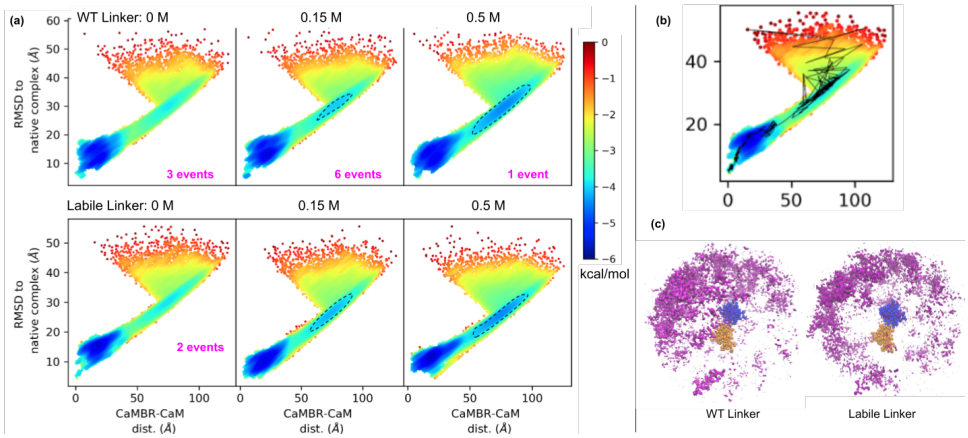

Figure 3: (a) Simulation trajectories projected onto 2D plane with axis: RMSD to native complex PDB 4Q5U and CaMBR/CaM center of mass distance. The dashed lines highlight the difference caused by increasing ionic strength. Energy is calculated via Boltzmann weights. The numbers of binding events that lead to native-like bound state are indicated in each panel. (b) An example trajectory of a binding event that leads to a native-like bound state. (c) Volumetric occupancy density isosurface (isovalue = 0.00015) of CaMBR beads around the CaM calculated using structures at the bottle-neck region indicated by dashed circles in panel a (RMSD$\sim$ 30 Å and CaMBR/CaM distance $\sim$75 Å) .

4.2 Linker flexibility determines the CaM conformation ensemble

To determine the basis for how the labile linker reduces the probability of achieving native bound states, we calculated the radius of gyration (R_G) of CaM just prior to forming the loosely-bound assembly. In the WT linker CaM model, the R_G distributions can be fitted to a two-peak Gaussian distribution with maxima at 15.2 and 18.0 Å for all three ionic strengths. These two peaks correspond to a highly compact structure and an extended structure, respectively. This distribution is consistent with the extended [43] and compact CaM [63] structures that have been experimentally-observed in the absence of target. Transition path calculations have suggested that the extended CaM formation is slightly more favorable than the compact one with ∼3-4 kcal/mol theromodynamic advantage and these two CaM states are separated by a ∼10 kcal/mol barrier [64]. Our R_G data of the WT linker CaM model qualitatively agree with the transition path results as the integrated extended CaM probability density is greater than that of the compact states and moreover, the probability density have limit overlap.

We contrast these data with those of the labile linker CaM model. For this configuration, a third peak emerges between the compact and extended distributions. The corresponding maxima are at 15.2, 17.0 and 18.0 Å, respectively. This additional probability density represents an intermediate state between the typical compact/extended conformation; the higher amplitude of which relative to the collapsed and extended states suggests that the intermediate state competes with two extreme CaM configurations. In other words, the alpha-helical linker of the WT model constrains the CaM conformational ensemble toward states the support productive binding. We illustrate this by labeling in green the configurations the amenable to CaMBR binding (green shaded areas in Fig. 4). It is clear from this representation that the WT linker exhibits a higher percentage of states facilitating CaMBR binding relative to the labile linker.

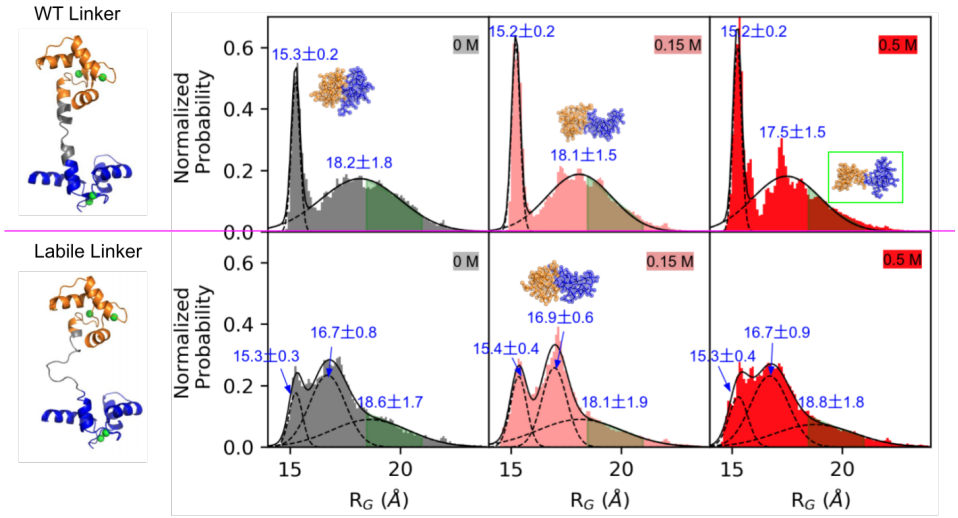

Figure 4: Radius of Gyration (R_G) of CaM at different ionic strengths before forming the loosely-bound state with CaMBR. The probability is normalized such that the area under the curve is 1. The green shaded areas show the R_Gs of all 12 binding events that lead to native-like CaMBR/CaM complex with $R_G = 19.7 \pm 1.3$ Å. The three structures in the top row have R_G values of 15.2, 18.0 and 19.7 Å, respectively. The one structure in the bottom row has R_G value of 17.0 Å..

4.3 Higher linker flexibility attenuates the associate rate sensitivities to ionic strength change at elevated ionic strength

We next relate the markedly different conformation ensembles adopted by the WT and labile linkers to the kinetics of CaM/CaMBR assembly. For this, we compute a CaM/CaMBR association rate by assuming that binding is a one-step process from unbound to loosely-bound state. We base our assumption on observations that a fluorescence signal monitoring binding yields a monophasic distribution in time [1]. Specifically, Cook et al used an acrylodan probe that reports changes in the proteins' hydrophobicity and polarity as they assemble. Changes in the probes fluorescence signify CaM/CaMBR association, although they do not necessarily reflect formation of the fully-bound, native-like state. Hence, we defined the CaM and CaMBR loosely-bound state by the set of configurations that adopt 50 inter-protein contacts. In Fig. 5a we verify that the loosely-bound state complexes (in gray) are thermodynamically favourable and located near the fully-bound state.

To estimate the association rate, k_a, we report the first passage time T_{fp} of reaching the loosely-bound state in Fig. 5b,c. Our calculated k_a via Eq. 4 of the WT linker CaM model at 0.5 M ionic strength is $5.9 \times 10^8\,\mathrm{M}^{-1}\,\mathrm{s}^{-1}$ (Table 1). This compares favorably to the experimental value of $2.2 \times 10^8\,\mathrm{M}^{-1}\,\mathrm{s}^{-1}$ [1] and suggests the accuracy of our strategy for computing k_a. We demonstrate in Fig. 5c that the increased concentration of ions reduces the k_a values for both configurations. This is consistent with experimental measurements [1] and our Brownian dynamic simulations using isolated CaM domains and CaMBR (Fig. S1) [59]. However, we notes some differences in these trends for the WT and labile linkers. For the WT linker, increasing the ionic strength monotonically decreases the k_a. However, this dependency is weaker for the labile linker case. We speculate that this weaker dependency is consistent with our observations that fewer labile linker configurations are confined to the bottle neck region at high ionic strengths relative to the WT (Fig. 3). This is

expected as the thermodynamic stability suggested in the bottle-neck region likely contributes to the dwelling time of CaM/CaM structures in this state, which is ultimately reflected in the T_{fp}. In total, our simulations indicate that the rates of forming loosely assembled CaM/CaMBR configurations are comparable for the WT and labile configurations and are strongly driven by electrostatic interactions. Importantly, though, the WT configuration strongly favors configurations that lead to fully-bound assemblies relative to the labile CaM.

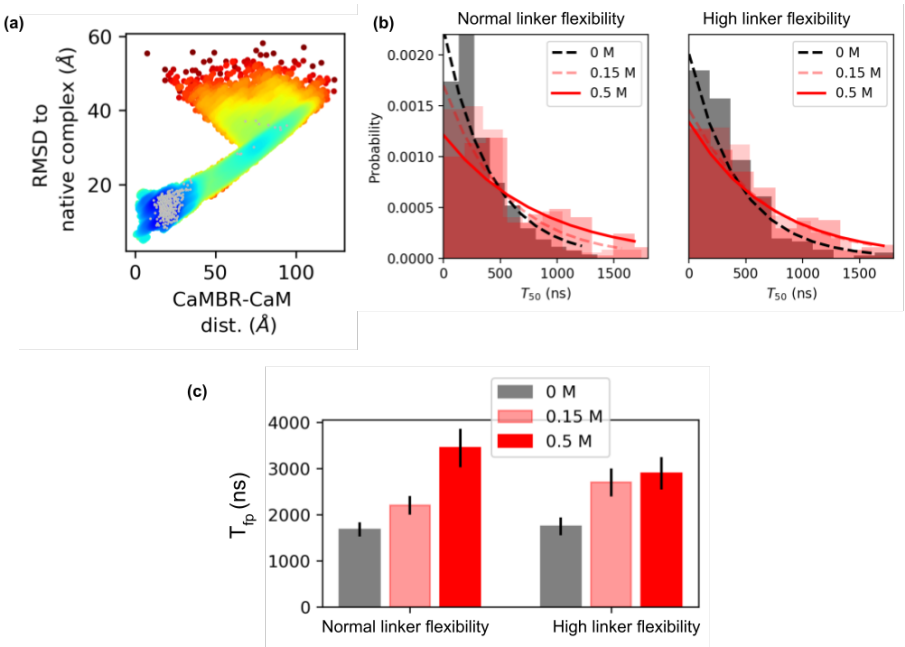

Figure 5: a) Locations of loosely-bound state complexes defined as 50 inter-contacts between CaMBR and CaM. b) Distribution of T_{fp}s for a sample consisting of 200 simulation trials (with replacement). The distribution was fitted to $A\exp(-\tau x)$ and the expected T_{fp} is $1/\tau$ ns. c) Average T_{fp} after bootstrapping using 1000 samples. .

Table 1: The mean first passage time (T_{fp}) to form the loosely-bound state (Fig. 5) and the corresponding k_a calculated via Eq. 4

	Normal linker flexibility			High linker flexibility		
	0 M	0.15 M	0.5 M	0 M	0.15 M	0.5 M
Encounter Time (ns)	1675±156	2199±203	3442±418	1745±190	2696±301	2894±352
k_a ($10^8 M^{-1}s^{-1}$)	12.13±1.12	9.24±0.85	5.91±0.72	11.65±1.27	7.54±0.84	7.02±0.85
Expt. k_a ($10^8 M^{-1}s^{-1}$)			2.2 ± 0.44			

5 Limitations

We note a few limitations of our study that could be refined in subsequent works. In our Martini CG model setup, two assumptions were made: 1) the N-/C- domain of CaM were treated as rigid and 2) The secondary structure of CaMBR of CaN is pre-defined as an alpha-helix. The first assumption is reasonable as our comparisons of the crystal structures for CaM and CaMBR/CaM complex indicate that the structures do not significantly change (<2 Å RMSD). The latter assumption could mask any conformational rearrangement that the CaMBR may reflect as part of an "induced fit" binding mechanism. However, we anticipate that this limitation should only have marginal impact on our results reported here. This is motivated by the increasingly accepted notion that IDP binding mechanisms are characterized by a "couple binding and folding" paradigm comprising "conformational selection" and "induced fit" [27]. The ratio of these two mechanisms is determined by IDP's propensity for folding in isolated state [65]. Our previous studies [59] show that when the CaMBR is simulated in the absence of CaM, it prefers a partially folded α helix structures (68.7% of the ensemble falls within 7-9 Å RMSD of a perfect alpha-helix). We believe this implies that "conformational selection" constitutes the predominant binding mechanism for CaMBR to CaM.

In our association rate calculations, we assumed a one-step binding process based on experimental measurements. However, a two-step binding process consisting of 1) an unbound to loosely-bound (encounter complex) assembly and 2) a loosely-bound to native bound assembly through structural reorganization, is also feasible. This mechanism has been observed in other IDP binding processes

[66–68]. To investigate this potential mechanism, experimental strategies that can monitor both the encounter complex and bound states are necessary. Such a strategy has been applied to melittin binding to CaM [69], for which a dansyl probe used as a fluorescence resonance energy transfer (FRET) acceptor from a melittin tryptophan residue. The approach indicated that the fluorescence signal is biphasic in time, therefore suggesting that there exists a fast rate of encounter-complex assembly, followed by a slower structural reorganization into the native complex. Analogous experimental studies for CaN binding would be helpful in elucidating this potential mechanism.

6 Conclusions

CaM's flexible linker plays an important role in target binding. In this study, we utilized the Martini coarse grained (CG) model to simulate the binding process between CaM and a representative protein target, the CaMBR of CaN. Our simulations examined the fully-bound native-like CaMBR/CaM complexes under normal and high linker flexibility conditions. The latter exhibited few events that lead to fully-bound configurations, suggesting that a higher linker flexibility reducing the probability of achieving productive binding. Our data indicate that the enhanced linker flexibility disturbs CaM's conformational ensemble in a manner that limits the sampling of states compatible with fully-bound assemblies. This limited sampling likely impacts CaM's ability to regulate targets [62], such as the CaN phosphatase considered in our study.

Our results further highlight the importance of electrostatics interactions in driving CaMBR and CaM assembly. Notably, the nonuniform electrostatic potential along the CaM solvent exposed surface imposes constraints on angles of approach leading to protein/protein association. We demonstrate that this stereospecific constraint is more significant for the WT linker relative to the highly labile configuration, suggesting that linker lability shapes the assembly mechanism. Further, these properties influence kinetics of CaM/CaMBR assembly and their sensitivity to changes in ionic strength. In sum, our study

emphasizes the important role of CaM's linker properties in shaping the thermodyanmics and kinetics of CaM/target assembly. Our findings could therefore shed light into how CaM target regulation could be impacted by modulation of CaM's linker properties, as might be expected for the linker-localized CaM missense mutations (M77I and S82R) [37] and post-translational modification (T80 and S82) [38].

7 Acknowledgements

We thank Peter Varughese for critical review of the manuscript. Research reported in this publication, release was supported by the Maximizing Investigators' Research Award (MIRA) (R35) from the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under grant number R35GM124977. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) [70], which is supported by the National Science Foundation under grant ACI-1548562.

S1 Supplementary Information (SI): Assessing the Role of CaM's Linker Flexibility in Target Binding

Bin Sun[1] and PM Kekenes-Huskey[1]*

[1] Department of Cell and Molecular Physiology, Loyola University Chicago, Maywood, IL, USA 60153

* Corresponding author. E-mail address: pkekeneshuskey@luc.edu (P. Kekenes-Huskey)

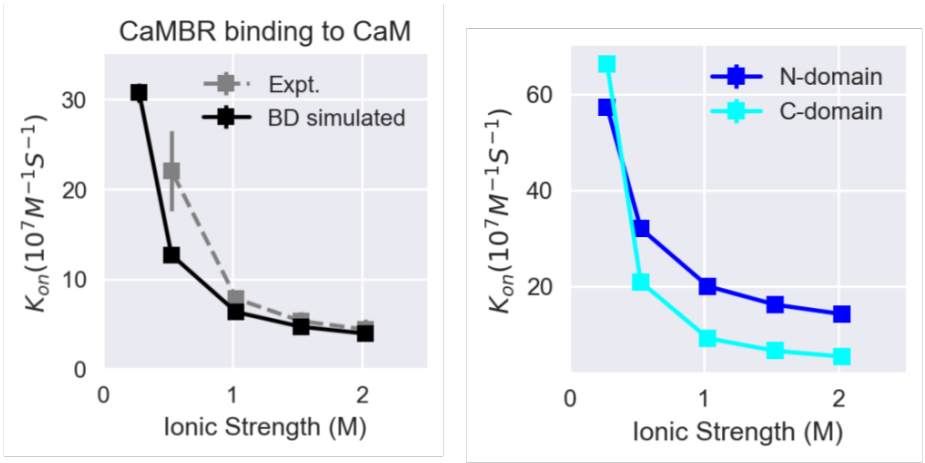

Figure S1: Association rate between CaM and CaMBR from Brownian dynamics (BD) simulations using the BrowDye package. In the BD simulations, the association between CaMBR and isolated CaM N-/C- domains was simulated with association rate k_n and k_c, respectively. The final associate rate k_a is achieved via $1/k_a = 1/k_n + 1/k_c$ [59].

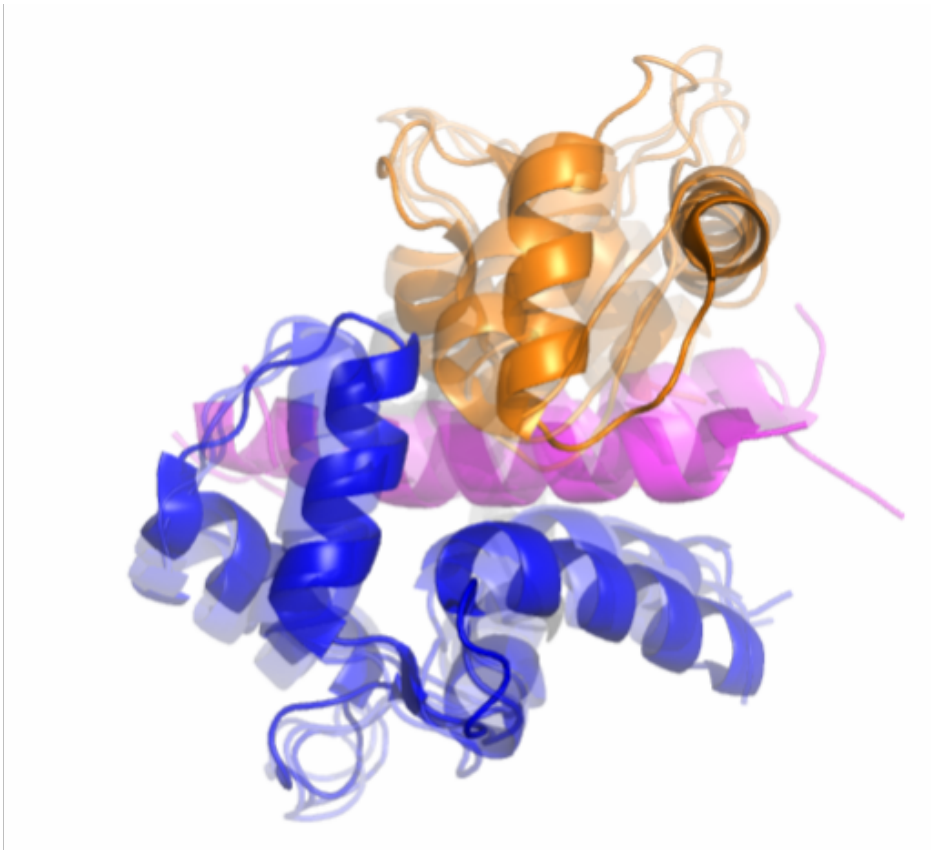

Figure S2: Superimposition of simulated fully-bound native-like CaM/-CaMBR complex structure with the crystal complex structure PDB 4Q5U. The simulated structures are shown as transparent cartoons. .

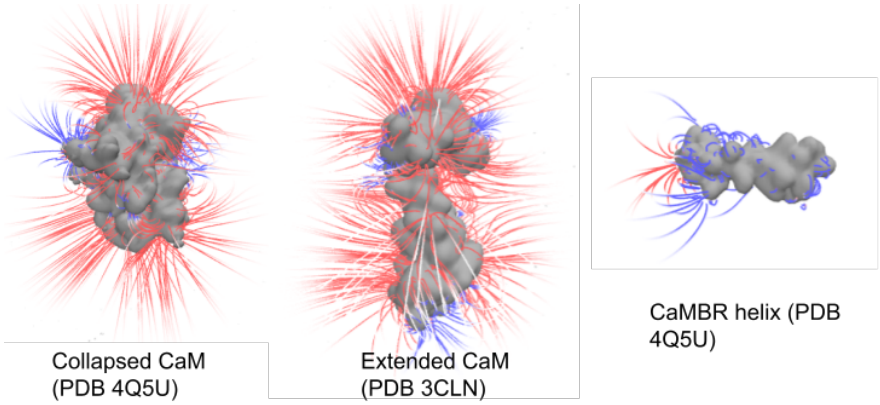

Figure S3: Electrostatic potentials (-0.2 to 0.2) of collapsed and extend CaM structures and the CaMBR helix. The electrostatic potentials were calculated via the APBS program at 0.15 monovalent ionic strength. The pqr files of each structure were generated using the pdb2pqr program with amber force field used.

References

[1] E. C. Cook and T. P. Creamer. "Influence of electrostatic forces on the association kinetics and conformational ensemble of an intrinsically disordered protein". In: *Proteins: Structure, Function and Bioinformatics* (Aug. 2020), prot.25979. ISSN: 10970134.

[2] P. Kursula. "The many structural faces of calmodulin: A multitasking molecular jackknife". In: *Amino Acids* 46.10 (Oct. 2014), pp. 2295–2304. ISSN: 14382199.

[3] J. M. Shifman and S. L. Mayo. "Exploring the origins of binding specificity through the computational redesign of calmodulin". In: *Proceedings of the National Academy of Sciences of the United States of America* 100.23 (Nov. 2003), pp. 13274–13279. ISSN: 00278424.

[4] P. M. Bayley, W. A. Findlay, and S. R. Martin. "Target recognition by calmodulin: Dissecting the kinetics and affinity of interaction using short peptide sequences". In: *Protein Science* 5.7 (July 1996), pp. 1215–1228. ISSN: 09618368.

[5] J. Gsponer et al. "A Coupled Equilibrium Shift Mechanism in Calmodulin-Mediated Signal Transduction". In: *Structure* 16.5 (May 2008), pp. 736–746. ISSN: 09692126.

[6] S. E. O'Donnell et al. "Recognition of β-calcineurin by the domains of calmodulin: Thermodynamic and structural evidence for distinct roles". In: *Proteins: Structure, Function and Bioinformatics* 79.3 (Mar. 2011), pp. 765–786. ISSN: 08873585.

[7] M. Zhang et al. "Structural basis for calmodulin as a dynamic calcium sensor". In: *Structure* 20.5 (May 2012), pp. 911–923. ISSN: 09692126.

[8] F. Van Petegem. "Slaying a giant: Structures of calmodulin and protein kinase a bound to the cardiac ryanodine receptor". In: *Cell Calcium* 83 (2019), p. 102079. ISSN: 15321991.

[9] I. De Diego et al. "Molecular basis of the death-associated protein kinase-calcium/calmodulin regulator complex". In: *Science Signaling* 3.106 (Jan. 2010), ra6 LP –ra6. ISSN: 19450877.

[10] M. S. Marlow et al. "The role of conformational entropy in molecular recognition by calmodulin". In: *Nature Chemical Biology* 6.5 (Apr. 2010), pp. 352–358. ISSN: 15524469.

[11] D. Shukla, A. Peck, and V. S. Pande. "Conformational heterogeneity of the calmodulin binding interface". In: *Nature Communications* 7.1 (Apr. 2016), pp. 1–11. ISSN: 20411723.

[12] F. Liu et al. "Molecular mechanism of multispecific recognition of Calmodulin through conformational changes". In: *Proceedings of the National Academy of Sciences of the United States of America* 114.20 (May 2017), E3927–E3934. ISSN: 10916490.

[13] G. Fiorin et al. "Using metadynamics to understand the mechanism of calmodulin/target recognition at atomic detail". In: *Biophysical Journal* 91.8 (Oct. 2006), pp. 2768–2777. ISSN: 00063495.

[14] C. Yang, G. S. Jas, and K. Kuczera. "Structure, dynamics and interaction with kinase targets: Computer simulations of calmodulin". In: *Biochimica et Biophysica Acta - Proteins and Proteomics* 1697.1-2 (Mar. 2004), pp. 289–300. ISSN: 15709639.

[15] A. M. Westerlund and L. Delemotte. "Effect of Ca2+on the promiscuous target-protein binding of calmodulin". In: *PLoS Computational Biology* 14.4 (Apr. 2018), e1006072. ISSN: 15537358.

[16] R. Mahling et al. "Ca2+-Saturated calmodulin binds tightly to the N-terminal domain of A-type fibroblast growth factor homologous factors". In: *Journal of Biological Chemistry* (2021), p. 100458. ISSN: 0021-9258.

[17] N. J. Anthis and G. M. Clore. "The length of the calmodulin linker determines the extent of transient interdomain association and target affin-

ity". In: *Journal of the American Chemical Society* 135.26 (July 2013), pp. 9648–9651. ISSN: 00027863.

[18] J. Wang et al. "Mapping central α-helix linker mediated conformational transition pathway of calmodulin via simple computational approach". In: *Journal of Physical Chemistry B* 118.32 (2014), pp. 9677–9685. ISSN: 15205207.

[19] M. F. VanBerkum, S. E. George, and A. R. Means. "Calmodulin activation of target enzymes. Consequences of deletions in the central helix". In: *Journal of Biological Chemistry* 265.7 (1990), pp. 3750–3756. ISSN: 00219258.

[20] H. Tidow and P. Nissen. "Structural diversity of calmodulin binding to its target sites". In: *FEBS Journal*. Vol. 280. 21. John Wiley & Sons, Ltd, Nov. 2013, pp. 5551–5565.

[21] A. P. Yamniuk and H. J. Vogel. "Calmodulin's flexibility allows for promiscuity in its interactions with target proteins and peptides". In: *Applied Biochemistry and Biotechnology - Part B Molecular Biotechnology* 27.1 (2004), pp. 33–57. ISSN: 10736085.

[22] D. M. Smith, T. P. Straatsma, and T. C. Squier. "Retention of conformational entropy upon calmodulin binding to target peptides is driven by transient salt bridges". In: *Biophysical Journal* 103.7 (Oct. 2012), pp. 1576–1584. ISSN: 00063495.

[23] P. Katyal et al. "Binding and backbone dynamics of protein under topological constraint: calmodulin as a model system". In: *Chemical Communications* 54.64 (2018), pp. 8917–8920. ISSN: 1364548X.

[24] T. B. Dunlap et al. "Thermodynamics of binding by calmodulin correlates with target peptide α-helical propensity". In: *Proteins: Structure, Function and Bioinformatics* 81.4 (Apr. 2013), pp. 607–612. ISSN: 08873585.

[25] J. Yang et al. "Features of molecular recognition of intrinsically disordered proteins via coupled folding and binding". In: *Protein Science* 28.11 (2019), pp. 1952–1965. ISSN: 1469896X.

[26] J. Dogan, S. Gianni, and P. Jemth. "The binding mechanisms of intrinsically disordered proteins". In: *Physical Chemistry Chemical Physics* 16.14 (2014), pp. 6323–6331. ISSN: 14639076.

[27] L. Mollica et al. "Binding mechanisms of intrinsically disordered proteins: Theory, simulation, and experiment". In: *Frontiers in Molecular Biosciences* 3.SEP (2016), p. 52. ISSN: 2296889X.

[28] P. Das, S. Matysiak, and J. Mittal. "Looking at the Disordered Proteins through the Computational Microscope". In: *ACS Central Science* 4.5 (May 2018), pp. 534–542. ISSN: 23747951.

[29] A. P. Collins and P. C. Anderson. "Complete Coupled Binding-Folding Pathway of the Intrinsically Disordered Transcription Factor Protein Brinker Revealed by Molecular Dynamics Simulations and Markov State Modeling". In: *Biochemistry* 57.30 (July 2018), pp. 4404–4420. ISSN: 15204995.

[30] A. S. Saglam and L. T. Chong. "Protein-protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations". In: *Chemical Science* 10.8 (Feb. 2019), pp. 2360–2372. ISSN: 20416539.

[31] P. C. Souza et al. "Protein–ligand binding with the coarse-grained Martini model". In: *Nature Communications* 11.1 (2020), p. 3714. ISSN: 20411723.

[32] Y. Levy, P. G. Wolynes, and J. N. Onuchic. "Protein topology determines binding mechanism". In: *Proceedings of the National Academy of Sciences of the United States of America* 101.2 (2004), pp. 511–516. ISSN: 00278424.

[33] Q. Wang et al. "Protein recognition and selection through conformational and mutually induced fit". In: *Proceedings of the National Academy of Sciences of the United States of America* 110.51 (2013), pp. 20545–20550. ISSN: 00278424.

[34] S. Kmiecik et al. *Coarse-Grained Protein Models and Their Applications*. July 2016.

[35] Z. R. Xie, J. Chen, and Y. Wu. "Predicting Protein-protein Association Rates using Coarse-grained Simulation and Machine Learning". In: *Scientific Reports* 7.1 (Apr. 2017), pp. 1–17. ISSN: 20452322.

[36] X. Chu and J. Wang. "Position-, disorder-, and salt-dependent diffusion in binding-coupled-folding of intrinsically disordered proteins". In: *Physical Chemistry Chemical Physics* 21.10 (2019), pp. 5634–5645. ISSN: 14639076.

[37] H. H. Jensen et al. "Human calmodulin mutations". In: *Frontiers in Molecular Neuroscience* 11 (2018), p. 396. ISSN: 16625099.

[38] G. Benaim and A. Villalobo. "Phosphorylation of calmodulin: Functional implications". In: *European Journal of Biochemistry* 269.15 (2002), pp. 3619–3631. ISSN: 00142956.

[39] Y. Ueda, H. Taketomi, and N. Gō. "Studies on protein folding, unfolding, and fluctuations by computer simulation. II. A. Three-dimensional lattice model of lysozyme". In: *Biopolymers* 17.6 (1978), pp. 1531–1548. ISSN: 10970282.

[40] S. J. Marrink and D. P. Tieleman. "Perspective on the martini model". In: *Chemical Society Reviews* 42.16 (July 2013), pp. 6801–6822. ISSN: 14604744.

[41] S. Takada et al. "Modeling Structural Dynamics of Biomolecular Complexes by Coarse-Grained Molecular Simulations". In: *Accounts of Chemical Research* 48.12 (Nov. 2015), pp. 3026–3035. ISSN: 15204898.

[42] F. Sterpone, P. Derreumaux, and S. Melchionna. "Protein simulations in fluids: Coupling the OPEP coarse-grained force field with hydrodynamics". In: *Journal of Chemical Theory and Computation* 11.4 (2015), pp. 1843–1853. ISSN: 15499626.

[43] Y. S. Babu, C. E. Bugg, and W. J. Cook. "Structure of calmodulin refined at 2.2 Å resolution". In: *Journal of Molecular Biology* 204.1 (1988), pp. 191–204. ISSN: 00222836.

[44] T. B. Dunlap et al. "Stoichiometry of the Calcineurin Regulatory Domain-Calmodulin Complex". In: *Biochemistry* 53.36 (2014), pp. 5779–5790. ISSN: 15204995.

[45] L. Monticelli et al. "The MARTINI coarse-grained force field: Extension to proteins". In: *Journal of Chemical Theory and Computation* 4.5 (May 2008), pp. 819–834. ISSN: 15499618.

[46] D. H. De Jong et al. "Improved parameters for the martini coarse-grained protein force field". In: *Journal of Chemical Theory and Computation* 9.1 (Jan. 2013), pp. 687–697. ISSN: 15499618.

[47] S. J. Marrink, A. H. De Vries, and A. E. Mark. "Coarse Grained Model for Semiquantitative Lipid Simulations". In: *Journal of Physical Chemistry B* 108.2 (Jan. 2004), pp. 750–760. ISSN: 15206106.

[48] W. Kabsch and C. Sander. "Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features". In: *Biopolymers* 22.12 (Dec. 1983), pp. 2577–2637.

[49] W. G. Touw et al. "A series of PDB-related databanks for everyday needs". eng. In: *Nucleic Acids Research* 43.D1 (Jan. 2015), pp. D364–D368.

[50] R. Salomon-Ferrer, D. A. Case, and R. C. Walker. "An overview of the Amber biomolecular simulation package". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 3.2 (2013), pp. 198–210. ISSN: 17590876.

[51] X. Periole et al. "Combining an elastic network with a coarse-grained molecular force field: Structure, dynamics, and intermolecular recognition". In: *Journal of Chemical Theory and Computation* 5.9 (Sept. 2009), pp. 2531–2543. ISSN: 15499618.

[52] S. O. Yesylevskyy et al. "Polarizable water model for the coarse-grained MARTINI force field". In: *PLoS Computational Biology* 6.6 (June 2010). Ed. by M. Levitt, pp. 1–17. ISSN: 1553734X.

[53] S. Páll et al. "Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS". In: *Journal of Chemical Physics* 153.13 (2020), p. 134110.

[54] T. A. Wassenaar et al. "Going backward: A flexible geometric approach to reverse transformation from coarse grained to atomistic models". In: *Journal of Chemical Theory and Computation* 10.2 (Feb. 2014), pp. 676–690. ISSN: 15499618.

[55] D. R. Roe and T. E. Cheatham. "PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data". In: *Journal of Chemical Theory and Computation* 9.7 (2013), pp. 3084–3095.

[56] H. X. Zhou. "Rate theories for biologists". eng. In: *Quarterly Reviews of Biophysics* 43.2 (May 2010), pp. 219–293. ISSN: 00335835.

[57] E. Bradley and T. R.J. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, 1994.

[58] G. Schreiber, G. Haran, and H. X. Zhou. "Fundamental aspects of protein - Protein association kinetics". In: *Chemical Reviews* 109.3 (2009), pp. 839–860. ISSN: 00092665.

[59] B. Sun et al. "Electrostatic control of calcineurin's intrinsically-disordered regulatory domain binding to calmodulin". In: *Biochimica et Biophysica Acta - General Subjects* 1862.12 (2018), pp. 2651–2659. ISSN: 18728006.

[60] A. Kurut et al. "Anisotropic interactions in protein mixtures: Self assembly and phase behavior in aqueous solution". In: *Journal of Physical Chemistry Letters* 3.6 (Mar. 2012), pp. 731–734. ISSN: 19487185.

[61] M. Lund. "Anisotropic protein-protein interactions due to ion binding". In: *Colloids and Surfaces B: Biointerfaces* 137 (Jan. 2016), pp. 17–21. ISSN: 18734367.

[62] S. D. Walton et al. "Divergent soybean calmodulins respond similarly to calcium transients: Insight into differential target regulation". In: *Frontiers in Plant Science* 8 (Feb. 2017). ISSN: 1664462X.

[63] J. L. Fallon and F. A. Quiocho. "A closed compact structure of native Ca2+-calmodulin". In: *Structure* 11.10 (2003), pp. 1303–1307. ISSN: 09692126.

[64] F. Delfino et al. "Structural Transition States Explored With Minimalist Coarse Grained Models: Applications to Calmodulin". In: *Frontiers in Molecular Biosciences* 6 (2019), p. 104. ISSN: 2296889X.

[65] M. Arai et al. "Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding". In: *Proceedings of the National Academy of Sciences of the United States of America* 112.31 (2015), pp. 9614–9619. ISSN: 10916490.

[66] W. T. Chu et al. "Role of non-native electrostatic interactions in the coupled folding and binding of PUMA with Mcl-1". In: *PLoS Computational Biology* 13.4 (2017), pp. 1–20. ISSN: 15537358.

[67] G. J. Gerlach et al. "A disordered encounter complex is central to the yeast Abp1p SH3 domain binding pathway". In: *PLoS Computational Biology* 16.9 (2020), pp. 1–32. ISSN: 15537358.

[68] W. T. Chu, S. L. Shammas, and J. Wang. "Charge Interactions Modulate the Encounter Complex Ensemble of Two Differently Charged Disordered Protein Partners of KIX". In: *Journal of Chemical Theory and Computation* 16.6 (June 2020), pp. 3856–3868. ISSN: 15499626.

[69] E. Kovacs et al. "Dissociation of calmodulin-target peptide complexes by the lipid mediator sphingosylphosphorylcholine: Implications in calcium signaling". In: *Journal of Biological Chemistry* 285.3 (Jan. 2010), pp. 1799–1808. ISSN: 00219258.

[70] J. Towns et al. "XSEDE: Accelerating Scientific Discovery". In: *Comput. Sci. Eng.* 16.5 (Sept. 2014), pp. 62–74. ISSN: 1521-9615.