Protein recognition and selection through conformational and mutually induced fit

Qian Wang[a,1], Pengzhi Zhang[a,1], Laurel Hoffman[b], Swarnendu Tripathi[a], Dirar Homouz[c], Yin Liu[b], M. Neal Waxham[b], and Margaret S. Cheung[a,2]

[a]Department of Physics, University of Houston, Houston, TX 77204; [b]Department of Neurobiology and Anatomy, University of Texas Medical School, Health Science Center, Houston, TX 77030; and [c]Department of Applied Math and Sciences, Khalifa University, Abu Dhabi, United Arab Emirates 127788

Edited* by Peter G Wolynes, Rice University, Houston, TX, and approved November 8, 2013 (received for review July 15, 2013)

Protein–protein interactions drive most every biological process, but in many instances the domains mediating recognition are disordered. How specificity in binding is attained in the absence of defined structure contrasts with well-established experimental and theoretical work describing ligand binding to protein. The signaling protein calmodulin presents a unique opportunity to investigate mechanisms for target recognition given that it interacts with several hundred different targets. By advancing coarse-grained computer simulations and experimental techniques, mechanistic insights were gained in defining the pathways leading to recognition and in how target selectivity can be achieved at the molecular level. A model requiring mutually induced conformational changes in both calmodulin and target proteins was necessary and broadly informs how proteins can achieve both high affinity and high specificity.

coarse-grained molecular simulations | stopped-flow fluorescence techniques | conformational flexibility | hydrophobic motif | calmodulin binding target

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 (1, 2). 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 (3–6) 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 (7) initially proposed for rigid ligand binding to proteins, and require adopting more dynamic models, such as induced fit (8) or conformational selection (9). 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 (10), mutually induced fit (11), fly-casting (12), or folding and binding (13, 14)—have begun to integrate molecule dynamics to describe folding and binding involving flexible molecules in protein–protein interactions, especially for the intrinsically disordered proteins (15–17) 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, we show that 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.

To establish experimental constraints for the coarse-grained simulations, we measured the association rate 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 (18). These peptides represent two distinct motif classes of CaM-binding molecules, and high-resolution NMR and crystal structures of the bound complexes are available (19, 20). 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 (Fig. 1*A*). 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, we used a coarse-grained model and performed Brownian dynamics simulation of the association between CaM and each of the two CaMBTs. It is important to note 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.

Initial computational studies on the association rates of proteins were typically based on the diffusion of rigid objects in

Significance

Protein–protein interactions drive most every biological process, but in many instances the domains mediating recognition are disordered. How specificity in binding is attained in the absence of defined structure is a fascinating problem but not well understood. Calmodulin presents a unique opportunity to investigate mechanisms for selectivity given that it interacts with several hundred different potential targets. In our work, a combined experimental and theoretical approach is taken to define how target selectivity occurs at the molecular level. Our study shows that the binding interactions require mutually induced conformational changes in both calmodulin and the target protein, and broadly informs how intrinsically disordered proteins can achieve both high affinity and high specificity.

Author contributions: Q.W., L.H., M.N.W., and M.S.C. designed research; Q.W., P.Z., L.H., D.H., M.N.W., and M.S.C. performed research; L.H., S.T., D.H., Y.L., M.N.W., and M.S.C. contributed new reagents/analytic tools; Q.W., P.Z., L.H., S.T., M.N.W., and M.S.C. analyzed data; and Q.W., P.Z., L.H., S.T., M.N.W., and M.S.C. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

[1]Q.W. and P.Z. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: mscheung@uh.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1312788110/-/DCSupplemental.

BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Downloaded at Palestinian Territory, occupied on December 9, 2021

www.manaraa.com

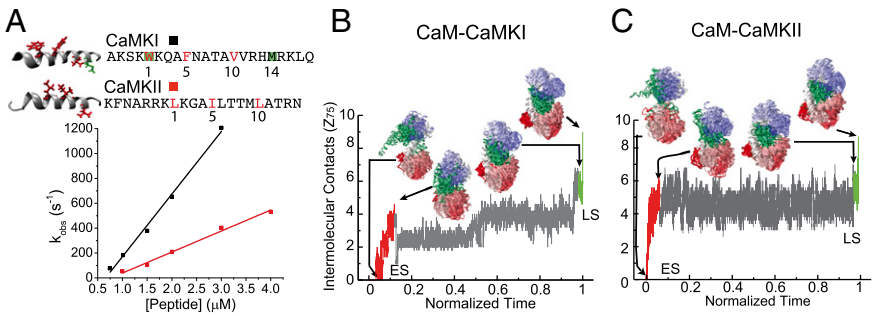

Fig. 1. Association of CaM and CaMBT. (*A*) Hydrophobic motifs are highlighted in CaMKI and CaMKII peptide structures and sequences. Rates for association, k_{obs}, of CaMKI (black) and CaMKII (red) peptides with CaM were measured and plotted as a function of peptide concentration. A linear function was used to fit the data where the resulting slope represents the on rate, k_{on}. On rates were determined to be 37.88×10^7 and 15.40×10^7 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 intermolecular contacts between the 75th residue on CaM and the target (Z_{75}) is plotted as a function of the normalized time. The superimposed structures of the CaM–CaMBT complexes are shown for the first transition (ES, in red). Similarly, the complex structures are shown for the second transition (LS, in green). CaM from *N* to C termini is colored from red to blue. CaMBT is green.

which the structural flexibility of a protein was ignored (21–23). 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 (13, 24–26). Others extended this structure-based model approach by introducing a set of experimentally determined structures in the Hamiltonian (27–32) (such as multiple basin structure-based models) or by including additional nonspecific interactions (33) to investigate the transitions between two or more distinct conformations of a protein (e.g., calcium-bound and unbound conformations of CaM; refs. 34, 35). 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, we 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, we computed the association rates of CaM and CaMBTs using an experimentally constrained coarse-grained molecular simulation and investigated the molecular mechanisms in postcollisional events. We were able to provide a mechanism of mutually induced fit for the understanding on how intrinsically disordered proteins achieve protein recognition and selection.

Results and Discussion

Association Rates of CaM–CaMBT. Time-dependent changes in fluorescence were monitored as CaM–Acr was rapidly mixed with excess peptide under pseudo–first-order binding conditions, and k_{on} was determined by the slope of k_{obs} values as a function of peptide concentration (Fig. 1*A*). On rates were 37.88×10^7 and 15.40×10^7 M^{-1} s^{-1} for CaMKI and CaMKII peptides, respectively. This difference in on rate cannot be accounted for from differences in diffusion rates, as the peptides are not significantly different in net charges (see ref. 18 for the effect of net charges on k_{on}) and in size (2.5 and 2.3 kDa, respectively), and diffuse more quickly than the larger CaM molecule (16.8 kDa).

We used the experimentally measured association rates of CaMKI and CaMKII with CaM to guide the determination of the computed association rates for the two targets from our simulations (Table 1 and its error analysis in *SI Appendix*, Table S1). Association rates were determined using stopped-flow techniques, where an environmentally sensitive fluorescent dye conjugated to the residue Cys75 of CaM reports association with CaMBTs. The observed changes in fluorescence do not distinguish between various orientations of a CaMBT that causes an environmental change. Correspondingly, our computational approach was designed to accommodate all types of binding; we did not dictate a particular orientation of the CaMBT to be a criterion to define association. Instead we defined the number of intermolecular contacts (Z_{75}) between residue 75 of CaM and the CaMBT to define successful trajectories and computationally assess the calculation of the association rates.

The definition of a successful association lies in the threshold of Z_{75} (see *SI Appendix* for additional discussion on Z_{75}). 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 that is statistically significant (*SI Appendix*, Table S1), 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}). We therefore used the threshold of $Z_{75} = 9$ as the criterion to define the end of the association process in the simulations.

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. 1 *B* and *C* and *SI Appendix*, Figs. S4 and S5): the early-stage transition (ES) and the late-stage transition (LS). ES denotes the first sharp transition shown in red (Fig. 1 *B* and *C* for CaMKI and CaMKII, respectively). After 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 LS transition encompassing the last 2% of a trajectory, which is shown in green in Fig. 1 *B* and *C*. In this late stage, the *N* domain of CaM (nCaM) collapses toward, and wraps around both CaMKI and CaMKII.

To further explore the CaM–CaMBT association process, we analyzed the association at the different stages ES (ES1 and ES2) and LS (LS1 and LS2) by several parameters listed in Table 2. ES1

Table 1. Calculated association rates obtained by simulations

	CaM–CaMKI		CaM–CaMKII	
Threshold Z_{75}	β	k_a (10^7 M^{-1} s^{-1})	β	k_a (10^7 M^{-1} s^{-1})
5	0.512	572.68 ± 0.71	0.517	577.63 ± 0.71
6	0.359	415.64 ± 0.64	0.402	460.88 ± 0.66
7	0.237	282.29 ± 0.55	0.230	274.41 ± 0.54
8	0.148	180.07 ± 0.45	0.114	139.85 ± 0.40
9	0.045	56.14 ± 0.26	0.018	22.61 ± 0.17
10	0.002	2.52 ± 0.06	0.001	1.26 ± 0.04

The association rates k_a were computed from Eq. **1**, and the standard error was estimated assuming number of successful trajectories follows Poisson distribution; β is the probability of successful association.

Downloaded at Palestinian Territory, occupied on December 9, 2021
www.manaraa.com

Table 2. Structural assessment of CaM and the target at the early and late stages of association

Stages of association	Δ* (CaM)	S† (CaM)	$D_{75\text{-target}}$‡ (σ)	CO¶ (CaMBT)
CaM–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				
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

*Asphericity Δ of CaM (35) is defined between 0 and 1. Here, 0 refers to the shape of a sphere and 1 refers to rod.
†Shape parameter S of CaM (35) ranging between −0.25 and 2 is defined as following: $S < 0$ is oblate, $S = 0$ is spherical, and $S > 0$ is prolate.
‡$D_{75\text{-target}}$ 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.
¶CO is the relative contact order of the target. The average sequence distance between all pairs of contacting residues normalized by the total sequence length shown in *SI Appendix*, Eq. **S7**.

is the time point when Z_{75} becomes nonzero and ES2 is the time point immediately before Z_{75} reaches plateau. LS1 is the first time point of LS, and LS2 is the last time point. At the beginning of the early stage (ES1), the distance between the target and CaM is still far apart ($D_{75\text{-target}} > 6\sigma$; σ = 3.8 Å). Initially both CaMKI and CaMKII bind to cCaM, as shown in the probability of intermolecular contact map at ES1 between CaM and CaMBT (*SI Appendix*, Fig. S6 *A* and *E*), and the shape of CaM is rather extended ($S \sim 0.6$ in Table 2). For both the CaMBTs, the relative contact order (CO) is the same (CO = 0.25 in Table 2) at ES1. At the end of the early stage (ES2), $D_{75\text{-target}}$ decreases significantly to ∼3σ accompanied by a significant decrease in the shape parameter (S) of CaM at $S \sim 0.3$. 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 *SI Appendix*, Fig. *S6 B* and *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).

Between the two major transitions, from ES2 to the beginning of LS1, we observed that $D_{75\text{-target}}$ significantly decreases to ∼1.8σ for both CaM–CaMKI and CaM–CaMKII (Table 2). 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 for CaMKII CO increases to 0.32. Between ES2 and LS1 a significant number of contacts to the cCaM are lost (especially from helix F and G of CaM), and those to the nCaM increase (*SI Appendix*, Fig. S6 *C* and *G*). At the end of LS (LS2), we observed a further decrease in $D_{75\text{-target}}$ to ∼1.5σ for both systems without further changes in the overall shape of CaM, as well as the probability of contact formation (*SI Appendix*, Fig. S6 *D* and *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 supports the inference 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). Our 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) (5) of CaMKI and CaMKII throughout the association, however, is not significantly different (H_{CaMKI} and its SE ranges from 0.41 ± 0.04–0.46 ± 0.04 and H_{CaMKII} and its SE ranges from 0.34 ± 0.04–0.42 ± 0.06). We propose that this is because H only captures the contact formation of the helical pitches, but 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, we suggest that 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.

Structural Changes in CaM and CaMBTs During the Course of Association. To investigate the above hypothesis, we next analyzed the structural variation of the targets upon association measured by the root mean square fluctuations of CaMKI and CaMKII at the different stages of association (*SI Appendix*, Fig. S7 *A* and *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.

We then computationally assessed the roles of hydrophobic motifs of the CaMBTs (*SI Appendix*, Table S2) during association by plotting the 3D 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. The native intermolecular contacts are defined only for 1–14 and 1–5-10 motifs of CaMKI and 1–5-10 motif of CaMKII (see *SI Appendix* for the definition of Q_I). 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 (Fig. 2 *A* and *B*, respectively) than CaMKII, which has only the 1–5-10 motif (Fig. 2*C*). These unique features are circled at $Q_I = 6$ between ES ($Z_{75} \sim 3$) and LS ($Z_{75} \sim 7$) in Fig. 2. Also evident in these 3D 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. From these results, we infer that 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

Downloaded at Palestinian Territory, occupied on December 9, 2021

www.manaraa.com

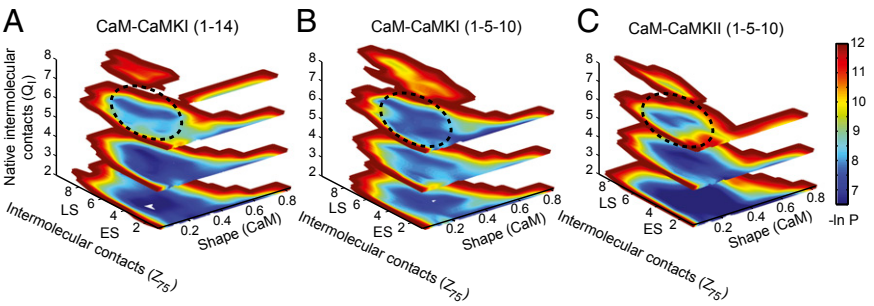

Fig. 2. Interplay between the structural changes in CaM and the CaM–CaMBT interactions. The 3D distributions are plotted in log scale -ln *P*, where *P* is the probability as a function of the number of selected native intermolecular contacts between CaM and the target (Q_I), the number of intermolecular contact formation between the 75th residue on CaM and the target (Z_{75}) (see *SI Appendix* for the definitions of Q_I and 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; and (*C*) 1–5-10 motif of CaMKII. The ES and 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 color bar, red and blue indicate low and high probability, respectively.

structural unwinding of the central helix of CaM revealed through NMR and crystallography (19, 20).

Next, we analyzed the orientation of the target against the major axis of CaM and their secondary structures after forming functional complexes. We used a phase diagram of cos(θ) and ΔZ to analyze the orientation of bound targets relative to CaM in *SI Appendix*, Fig. S7 *C* and *D* for CaM–CaMKI and CaM–CaMKII, respectively. There are several dominant binding orientations including normal, inverse, and others shown in *SI Appendix*, Table S3. Nearly all CaMKI peptides bind in either normal (49%) or inverse orientations (49%), with equal probability of each. CaMKII can bind in normal (30%), inverse (40%), and other orientations (30%). Specifically, when we 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. In contrast, CaMKII binding trajectories showed that nearly 40% of the bound complex adopts bent helical structures (*SI Appendix*, Fig. S7*C*) with a large CO value (*SI Appendix*, Fig. S7*D* and Table 2).

Principal Component Analysis (PCA) and the Cross-Correlation 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 toward bound complexes. To quantify these differences, we performed 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. 3, *SI Appendix*, Figs. *S8* and *S9*). We observed that PC2 of CaM–CaMKI distinguished the trajectories of association leading to a normal or inverse binding orientation (Fig. 3 *A* and *C*). When CaMKI starts in an inverse mode of binding (Fig. 3*A*), there are strong and broad anticorrelations among the contacts across interdomain and intradomain CaM motions (*SI Appendix*, Fig. S8*B*). At the late stage of the trajectory when CaMKI flips to a normal orientation, the strained interaction among interdomain motion diminishes (*SI Appendix*, Fig. S8*C*). Fig. 3*C* 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 movements of CaM are strongly anticorrelated (*SI Appendix*, Fig. S8*E*) even after they form a complex.

A similar pattern from the PCA is observed for the binding of CaM to CaMKII. Fig. 3*B* 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 and PC2 (*SI Appendix*, Fig. S9 *B* and *C*, respectively), we note that before the formation of a functional complex, the intradomain and interdomain anticorrelated 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. 3*D*), the patterns differ at PC2 where the anticorrelation among the interdomain contacts of CaM grows instead of shrinks (*SI Appendix*, Fig. S9*F*, lower triangle). We 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.

Protein Recognition Through Conformational and Mutually Induced Fit. We observed that at the onset of an association between CaM and CaMBTs, the formation of a transient complex is initiated,

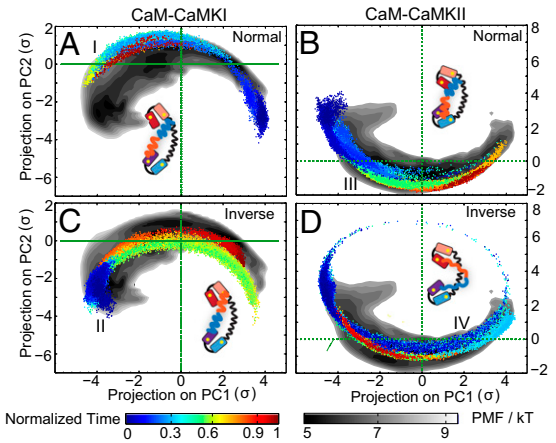

Fig. 3. PCA on the trajectories between the early stage and the late stage of the association. PC1 characterizes the interdomain movement of calmodulin and PC2 denotes the combined intradomain and interdomain movements of calmodulin. The PMF (in gray scale) is obtained by projecting structures from all of 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; σ is 3.8 Å.

Downloaded at Palestinian Territory, occupied on December 9, 2021
www.manaraa.com

and then a newly formed energy landscape emerges that accounts for conformational adjustments required to form the partially associated complex (Fig. 4*A*). The distinction between CaM–CaMKI and CaM–CaMKII occurs at target recognition (Fig. 4*B*), where differences in the structural fluctuation of the targets induce different responses in the structure of CaM. These fluctuations involve both interdomain and intradomain movements as the target-binding pockets in CaM adjust to the unique structurally changing targets, and CaM collapses from a normally more extended conformation. The target recognition happens during the process of conformational and mutually induced fit because both CaM and the target must undergo significant structural changes to access new conformations that rarely exist before binding.

To further understand the nature of intermolecular interactions (hydrophobic and electrostatic) in the mechanism of the conformational and mutually induced fit, we calculated the average number of contacts for each residue of CaMKI and CaMKII with CaM at the different stages of association. Our results clearly indicate that most of the charged residues of CaMBTs (labeled with an asterisk in *SI Appendix*, Fig. S10) have significant interactions with CaM at ES1. Nonetheless, at the later stages of association, especially at ES2 and LS1, we 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 our study, we 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) and 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.

Our contribution here 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. Our 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. In addition, our 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.

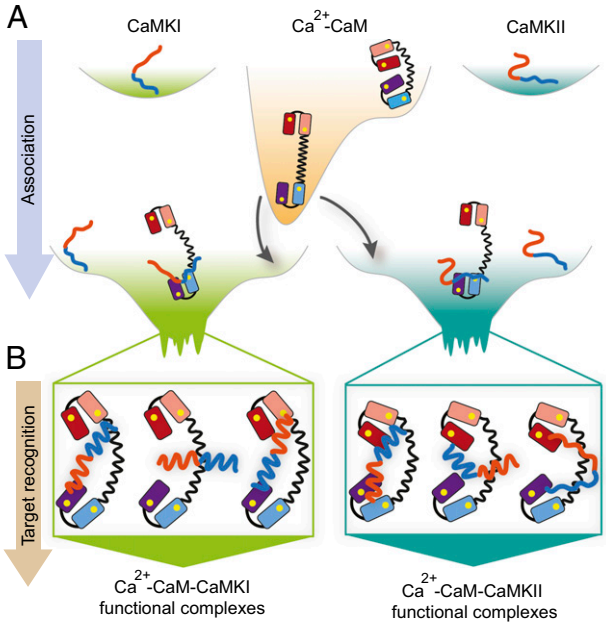

Fig. 4. Schematic energy landscape of calmodulin-target recognition through mutually induced fit. (*A*) Before association, CaMKI and CaMKII targets each remain unstructured in the unbound state, and Ca^{2+}–CaM samples both the extended and the collapsed conformations that are different from the conformation 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 before the formation of the functional complexes of Ca^{2+}–CaM–CaMKI and Ca^{2+}–CaM–CaMKII.

Materials and Methods

Measurement of the Association Rates by Stopped-Flow Fluorescence Spectroscopy. On rates were measured using an Applied Photophysics Ltd model SX20MV sequential stopped flow spectrofluorimeter with an instrument dead time of 1.7 ms. Reactions were carried out by rapidly mixing solutions from two syringes in equal volumes at 4 °C. One syringe contained 200 nM acrylodan labeled CaM in a standard buffer (1 mM $CaCl_2$, 25 mM Mops, 150 mM KCl, 0.1 mg/mL BSA, 0.1 mM EGTA, pH 7.2) and the other contained the indicated amount of peptide in the standard buffer (see *SI Appendix* for the sample purification and preparation). Data were collected for six binding reactions and the average was fit to a single exponential equation to determine k_{obs} (*SI Appendix*, Fig. S1). On rates were calculated for each condition by plotting k_{obs} as a function of peptide concentration and determining the slope of the linear fit (Fig. 1).

Coarse-Grained Protein Model. A side-chain C_α model (36) that includes two beads per amino acid (except glycine) was adopted to represent CaM and each of the two CaMBTs (CaMKI and CaMKII). The total potential energy E of the system is given by $E = E_{CaM} + E_{target} + E_{CaM-target}$. E_{CaM} and E_{target} represent the potential energy of the CaM and a target peptide, respectively. Each consists of the structural term, nonbonded van der Waals interaction and electrostatic interaction (see the *SI Appendix* for detailed description of each of these energy terms). The electrostatic interaction between the two beads with partial charges (5) in solution is described by the Debye–Hückel (37) potential to include the screening effect of ions. The nonbonded van der Waals interaction is represented by a Lennard–Jones (LJ) potential (*SI Appendix*, Eq. **S5**). The solvent-mediated interaction between beads i and j each with a distinct type of amino acid ε_{ij} is scaled to avoid overly compensating the electrostatic interactions due to the introduction of a Debye–Hückel potential. Details are provided in the *SI Appendix* and the scaled strength of LJ interactions within CaM or a target peptide ε_{ij}^{intra} is provided in *SI Appendix*, Table S7. The energy between CaM and a target peptide $E_{CaM\text{-}target}$ consists of both van der Waals interaction (LJ potential) and electrostatic interaction. The strength of the LJ interaction between CaM and a target peptide ε_{ij}^{inter} was set to 2 ε_{ij}^{intra} to enhance the stability of the functional complex when the two domains of CaM wrap around the target.

Parameterization of CaM and Targets. For CaM, aimed at developing a transferrable potential for CaM to interact with CaMBTs of distinct types, we used the crystal structure of an extended form of unbound CaM (PDB ID: 1CLL) (38) as a reference. In specific, the dihedral potential of CaM was parameterized to reflect the experimentally obtained measurements. The average radius of gyration of CaM after parameterization is ~21.2Å, close to the value 21.3 Å measured by the X-ray scattering experiments (39). The distribution of the radius of gyration of CaM (*SI Appendix*, Fig. S11) shows two peaks corresponding to the extended state and the collapsed state at the ratio of 9.25:1, close to 9:1 from the paramagnetic relaxation enhancement measurement (6). For the targets, the dihedral strength was parameterized to render a highly disordered ensemble of structures in the unbound state and allows the sampling of helical conformations in the bound state. Expression and parameters of then Hamiltonian for the two systems can be found in the *SI Appendix*.

Association Simulations. Coarse-grained molecular simulations were performed using an in-house version of AMBER10 (40) to investigate the association process through the Brownian equations of motion. The simulation temperature is $k_BT/\varepsilon = 1.1$, where $\varepsilon = 0.6$ kcal/mol and k_B is the Boltzmann constant. The viscosity of the system was set to 0.001 Pa · S in aqueous

Downloaded at Palestinian Territory, occupied on December 9, 2021

www.manaraa.com

solution. We started with an algorithm developed by McCammon and coworkers (41). The association rate k_a is calculated by,

$$k_a = 4\pi Db\left[\frac{\beta}{1-(1-\beta)\Omega}\right], \quad [1]$$

where D is the diffusion coefficient of the system, which is the sum of the diffusion coefficients of the target and CaM. In our calculations, we used the experimental measurement on D (42) for CaM (7.8×10^{-7} cm^2/s). For the targets, we approximated the diffusion coefficient to be 2 times that of CaM based on the ratio of their sizes in terms of their radius of gyrations, because experimental values were not available; β is the probability of successful association events. CaM was randomly distributed in a spherical surface that is $b = 15\sigma$ (σ = 3.8 Å) away from the center of mass of a binding target. The interactions between CaM and the target are isotropic beyond this distance. Here, Ω is the returning rate, which was set to 0.2 in the simulation (43, 44). In other words, the simulation stopped if the distance between CaM and a binding target exceeded 75σ or reached the maximum time of 240 million steps. A total of 1,000 pairs of CaM and CaMBT coordinates were selected for the kinetic simulation. Details about preparation of initial structures and Brownian dynamics simulation are provided in the *SI Appendix*.

Analysis of Data from Simulations. The $D_{75\text{-}target}$ defines the distance between residue 75 of CaM and the center of mass of the target. S (CaM) measures the shape of CaM during association from an unbound extended conformation to the bound state. The CO is the measure of the locality of the contacts between residues along a sequence normalized by the total length of the sequence (*SI Appendix*, Eq. **S7**). Here, θ is defined as the angle between the vector pointing from the center of mass of cCaM to that of nCaM and the vector pointing from the geometrical center of the last four to that of the first four amino acids of a CaMBT; ΔZ is defined as the difference in the number of intermolecular contacts of a CaMBT at its two ends to the CaM domains [see *SI Appendix*, Figs. S2 and S3 for a detailed description of $\cos(\theta)$ and ΔZ, respectively].

ACKNOWLEDGMENTS. M.S.C. thanks Drs. P. G. Wolynes and J. N. Onuchic for the helpful discussion. We thank the National Institutes of Health (1R01GM097553) for the support, and Rice University (Blue BioU) and Extreme Science and Engineering Discovery Environment (XSEDE) for the computing resources.

1. Tidow H, Nissen P (2013) Structural diversity of calmodulin binding to its target sites. *FEBS J* 280(21):5551–5565, 10.1111/febs.12296.
2. Yamniuk AP, Vogel HJ (2004) Calmodulin's flexibility allows for promiscuity in its interactions with target proteins and peptides. *Mol Biotechnol* 27(1):33–57.
3. Slaughter BD, Bieber-Urbauer RJ, Johnson CK (2005) Single-molecule tracking of submillisecond domain motion in calmodulin. *J Phys Chem B* 109(26):12658–12662.
4. Homouz D, Sanabria H, Waxham MN, Cheung MS (2009) Modulation of calmodulin plasticity by the effect of macromolecular crowding. *J Mol Biol* 391(5):933–943.
5. Wang Q, Liang KC, Czader A, Waxham MN, Cheung MS (2011) The effect of macromolecular crowding, ionic strength and calcium binding on calmodulin dynamics. *PLOS Comput Biol* 7(7):e1002114.
6. Anthis NJ, Doucleff M, Clore GM (2011) Transient, sparsely populated compact states of apo and calcium-loaded calmodulin probed by paramagnetic relaxation enhancement: Interplay of conformational selection and induced fit. *J Am Chem Soc* 133(46):18966–18974.
7. Fischer E (1894) Einfluss der Configuration auf die Wirkung der Enzyme [Configuration of the influence on the action of enzymes]. *Ber Dtsch Chem Ges* 27:2984–2993. German.
8. Koshland DEJ (1958) Application of a theory of enzyme specificity of protein synthesis. *Proc Natl Acad Sci USA* 44(2):98–104.
9. Monod J, Wyman J, Changeux JP (1965) On the nature of allosteric transitions: A plausible model. *J Mol Biol* 12:88–118.
10. Tsai CJ, Kumar S, Ma BY, Nussinov R (1999) Folding funnels, binding funnels, and protein function. *Protein Sci* 8(6):1181–1190.
11. Williamson JR (2000) Induced fit in RNA-protein recognition. *Nat Struct Biol* 7(10):834–837.
12. Shoemaker BA, Portman JJ, Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proc Natl Acad Sci USA* 97(16):8868–8873.
13. Levy Y, Wolynes PG, Onuchic JN (2004) Protein topology determines binding mechanism. *Proc Natl Acad Sci USA* 101(2):511–516.
14. Chu X, Gan L, Wang E, Wang J (2013) Quantifying the topography of the intrinsic energy landscape of flexible biomolecular recognition. *Proc Natl Acad Sci USA* 110(26):E2342–E2351.
15. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J Mol Biol* 293(2):321–331.
16. Dunker AK, et al. (2001) Intrinsically disordered protein. *J Mol Graph Model* 19(1):26–59.
17. Uversky VN (2002) Natively unfolded proteins: A point where biology waits for physics. *Protein Sci* 11(4):739–756.
18. Waxham MN, Tsai AL, Putkey JA (1998) A mechanism for calmodulin (CaM) trapping by CaM-kinase II defined by a family of CaM-binding peptides. *J Biol Chem* 273(28):17579–17584.
19. Gifford JL, Ishida H, Vogel HJ (2011) Fast methionine-based solution structure determination of calcium-calmodulin complexes. *J Biomol NMR* 50(1):71–81.
20. Meador WE, Means AR, Quiocho FA (1993) Modulation of calmodulin plasticity in molecular recognition on the basis of x-ray structures. *Science* 262(5140):1718–1721.
21. Gabdoulline RR, Wade RC (2001) Protein-protein association: Investigation of factors influencing association rates by Brownian dynamics simulations. *J Mol Biol* 306(5):1139–1155.
22. Alsallaq R, Zhou HX (2007) Prediction of protein-protein association rates from a transition-state theory. *Structure* 15(2):215–224.
23. Elcock AH, Gabdoulline RR, Wade RC, McCammon JA (1999) Computer simulation of protein-protein association kinetics: Acetylcholinesterase-fasciculin. *J Mol Biol* 291(1):149–162.
24. Turjanski AG, Gutkind JS, Best RB, Hummer G (2008) Binding-induced folding of a natively unstructured transcription factor. *PLOS Comput Biol* 4(4):e1000060.
25. Wang J, et al. (2011) Multi-scaled explorations of binding-induced folding of intrinsically disordered protein inhibitor IA3 to its target enzyme. *PLOS Comput Biol* 7(4):e1001118.
26. Chu X, et al. (2012) Importance of electrostatic interactions in the association of intrinsically disordered histone chaperone Chz1 and histone H2A.Z-H2B. *PLOS Comput Biol* 8(7):e1002608.
27. Best RB, Chen YG, Hummer G (2005) Slow protein conformational dynamics from multiple experimental structures: The helix/sheet transition of arc repressor. *Structure* 13(12):1755–1763.
28. Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG (2006) Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc Natl Acad Sci USA* 103(32):11844–11849.
29. Okazaki KI, Takada S (2008) Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms. *Proc Natl Acad Sci USA* 105(32):11182–11187.
30. Weinkam P, Pons J, Sali A (2012) Structure-based model of allostery predicts coupling between distant sites. *Proc Natl Acad Sci USA* 109(13):4875–4880.
31. Whitford PC, Miyashita O, Levy Y, Onuchic JN (2007) Conformational transitions of adenylate kinase: Switching by cracking. *J Mol Biol* 366(5):1661–1671.
32. Lu Q, Wang J (2008) Single molecule conformational dynamics of adenylate kinase: Energy landscape, structural correlations, and transition state ensembles. *J Am Chem Soc* 130(14):4772–4783.
33. Ravikumar KM, Huang W, Yang S (2012) Coarse-grained simulations of protein-protein association: An energy landscape perspective. *Biophys J* 103(4):837–845.
34. Tripathi S, Portman JJ (2009) Inherent flexibility determines the transition mechanisms of the EF-hands of calmodulin. *Proc Natl Acad Sci USA* 106(7):2104–2109.
35. Zuckerman DM (2004) Simulation of an ensemble of conformational transitions in a united-residue model of calmodulin. *J Phys Chem B* 108(16):5127–5137.
36. Cheung MS, Finke JM, Callahan B, Onuchic JN (2003) Exploring the interplay between topology and secondary structural formation in the protein folding problem. *J Phys Chem B* 107(40):11193–11200.
37. Debye P, Hückel E (1923) The theory of electrolytes. I. Lowering of freezing point and related phenomena. *Phys Z* 24:185–206.
38. Chattopadhyaya R, Meador WE, Means AR, Quiocho FA (1992) Calmodulin structure refined at 1.7 A resolution. *J Mol Biol* 228(4):1177–1192.
39. Heidorn DB, 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.
40. Case D, et al. (2008) AMBER 10 (University of California, San Francisco).
41. Northrup SH, Allison SA, McCammon JA (1984) Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J Chem Phys* 80(4):1517–1526.
42. Slaughter BD, Allen MW, Unruh JR, Urbauer RJB, Johnson CK (2004) Single-molecule resonance energy transfer and fluorescence correlation spectroscopy of calmodulin in solution. *J Phys Chem B* 108(29):10388–10397.
43. Northrup SH, Boles JO, Reynolds JCL (1988) Brownian dynamics of cytochrome c and cytochrome c peroxidase association. *Science* 241(4861):67–70.
44. Gabdoulline RR, Wade RC (1997) Simulation of the diffusional association of barnase and barstar. *Biophys J* 72(5):1917–1929.

Downloaded at Palestinian Territory, occupied on December 9, 2021

www.manaraa.com