Biophysical Journal
Article

Biophysical Society

Opposing Intermolecular Tuning of Ca^{2+} Affinity for Calmodulin by Neurogranin and CaMKII Peptides

Pengzhi Zhang,[1] Swarnendu Tripathi,[1] Hoa Trinh,[1] and Margaret S. Cheung[1,2,*]
[1]Department of Physics, University of Houston, Houston, Texas; and [2]Center for Theoretical Biological Physics, Rice University, Houston, Texas

ABSTRACT We investigated the impact of bound calmodulin (CaM)-target compound structure on the affinity of calcium (Ca^{2+}) by integrating coarse-grained models and all-atomistic simulations with nonequilibrium physics. We focused on binding between CaM and two specific targets, Ca^{2+}/CaM-dependent protein kinase II (CaMKII) and neurogranin (Ng), as they both regulate CaM-dependent Ca^{2+} signaling pathways in neurons. It was shown experimentally that Ca^{2+}/CaM (holoCaM) binds to the CaMKII peptide with overwhelmingly higher affinity than Ca^{2+}-free CaM (apoCaM); the binding of CaMKII peptide to CaM in return increases the Ca^{2+} affinity for CaM. However, this reciprocal relation was not observed in the Ng peptide ($Ng_{13–49}$), which binds to apoCaM or holoCaM with binding affinities of the same order of magnitude. Unlike the holoCaM-CaMKII peptide, whose structure can be determined by crystallography, the structural description of the apoCaM-$Ng_{13–49}$ is unknown due to low binding affinity, therefore we computationally generated an ensemble of apoCaM-$Ng_{13–49}$ structures by matching the changes in the chemical shifts of CaM upon $Ng_{13–49}$ binding from nuclear magnetic resonance experiments. Next, we computed the changes in Ca^{2+} affinity for CaM with and without binding targets in atomistic models using Jarzynski's equality. We discovered the molecular underpinnings of lowered affinity of Ca^{2+} for CaM in the presence of $Ng_{13–49}$ by showing that the N-terminal acidic region of Ng peptide pries open the β-sheet structure between the Ca^{2+} binding loops particularly at C-domain of CaM, enabling Ca^{2+} release. In contrast, CaMKII peptide increases Ca^{2+} affinity for the C-domain of CaM by stabilizing the two Ca^{2+} binding loops. We speculate that the distinctive structural difference in the bound complexes of apoCaM-$Ng_{13–49}$ and holoCaM-CaMKII delineates the importance of CaM's progressive mechanism of target binding on its Ca^{2+} binding affinities.

INTRODUCTION

Calcium (Ca^{2+}) is exquisitely used by a cell for transducing external stimuli through its gradient of extracellular (~1000 μM) and intracellular (~0.1 μM) concentration (1). A broad spectrum of Ca^{2+} signals are encoded by protein calmodulin (CaM) (2) (the structure is shown in Fig. 1 *a*) through specific binding with various targets regulating CaM-dependent Ca^{2+} signaling pathways in neurons (3). The targets enhance CaM's affinity for Ca^{2+} by up to 30 times or accelerate dissociation of Ca^{2+} from CaM (4). Therefore, this mechanism tunes CaM's capacity of encoding global Ca^{2+} signals by its differential binding affinity for Ca^{2+}. Among those targets, Ca^{2+}/CaM-dependent protein kinase II (CaMKII) (5–7) (the structure and sequence of the CaM-binding domain are shown in Fig. 1 *b*) and neurogranin (Ng) (8) (the sequence of the $Ng_{13–49}$ peptide is shown in Fig. 1 *c*) play an essential role in synaptic plasticity (9), which is critical to learning and memory formation shown by early behavior experiments on Ng-mutated or Ng-knocked-out mice (10,11). We summarized several experimental observations (3,7) in Fig. S1 on how CaM sensitizes downstream CaM-binding enzymes to Ca^{2+} regulating synaptic plasticity in memory formation: CaM activates CaMKII at high Ca^{2+} levels (6) while the availability of CaM is controlled by a group of postsynaptic proteins including Ng at a lower Ca^{2+} level (8,12).

Waxham et al. (13,14) used fluorescence experiments to show that the CaMKII kinase increases the affinity of Ca^{2+} for CaM while Ng significantly decreases the affinity of Ca^{2+} for Ca^{2+}-binding loop III and IV (Fig. 1 *a*) in C-domain of CaM (cCaM) by up to 60-fold. These experiments raised questions about the dependence of CaM's affinity for Ca^{2+} ions on structural models of CaM (13,15). Recent studies by Le Novère et al. (16,17) demonstrated the alteration of Ca^{2+} affinity for CaM by plotting the binding affinity curves with elaborated allosteric models and fitted rates. They investigated the effect of competing targets on CaM by showing the enhancement of binding affinity in

Submitted August 3, 2016, and accepted for publication January 23, 2017.
*Correspondence: mscheung@uh.edu
Editor: Michael Feig.
http://dx.doi.org/10.1016/j.bpj.2017.01.020

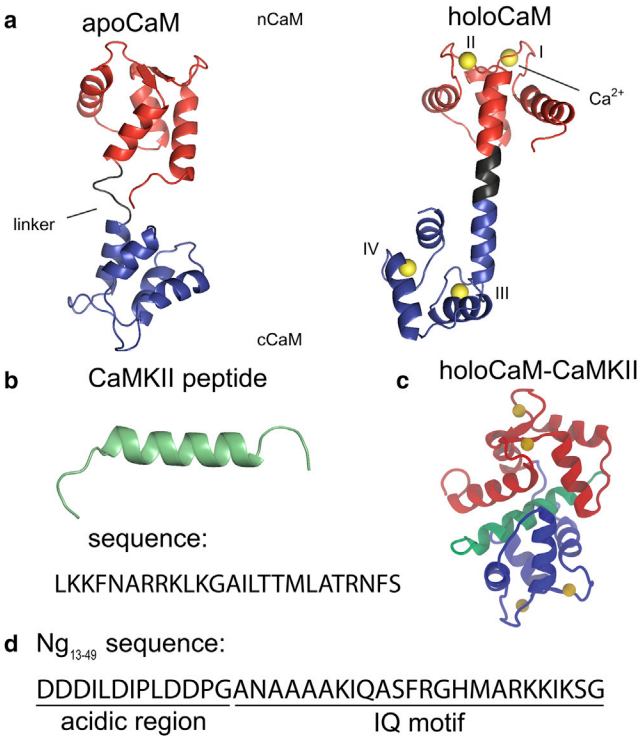

FIGURE 1 Illustration of structures of CaM and sequences of the target peptides. (*a*) Solution structure of Ca^{2+}-free CaM (apoCaM) (PDB: 1CFD) and crystal structure of Ca^{2+}/CaM (holoCaM) (PDB: 1CLL). CaM is colored as follows, red → nCaM (residue 1–76), gray → central linker (residue 77–82), blue → cCaM (residue 83–148), and the Ca^{2+} ions are colored in green; Ca^{2+} binding sites are labeled from I to IV. (*b*) Structure of Ca^{2+}/CaM-dependent kinase II (CaMKII) peptide from the holoCaM-CaMKII crystal structure (PDB: 1CDM) and the sequence of the peptide. (*c*) X-ray structure of holoCaM-CaMKII (PDB: 1CDM). CaM and the Ca^{2+} ions are colored as in (*a*) and the CaMKII peptide is colored in green. (*d*) Sequence of the neurogranin peptide ($Ng_{13–49}$). The two minimum compositions of the $Ng_{13–49}$ peptide are marked as "acidic region" and "IQ motif." To see this figure in color, go online.

the presence of some targets (such as CaMKII), as well as the diminishment in the presence of other targets (such as Ng). Dieckmann et al. (18) used steady-state fluorescence and nuclear magnetic resonance (NMR) experiments and found that the binding of endothelial nitric oxide synthase (eNOS) to CaM enhances CaM's affinity for Ca^{2+} while phosphorylated eNOS does not due to electrostatic repulsion. They speculated that the phosphorylation on eNOS diminished its helical propensity, possibly affecting CaM's affinity for Ca^{2+}. On the other hand, the structural basis of the regulation of Ca^{2+} affinity for CaM by Ng and CaMKII remains elusive.

CaM has two domains separated by a flexible central linker. Each domain consists of two helix-linker-helix motifs (marked as I–IV in Fig. 1 *a*), which can bind a Ca^{2+} ion. The flexible central linker allows adjustable CaM conformations, thus accommodates the binding of distinct CaM binding targets (CaMBTs) (19–23). In a canonical complex of Ca^{2+}/CaM (holoCaM) and CaMKII peptide, the two domains of holoCaM wrap around the CaMKII peptide, which forms a helical structure (24). In our previous work, by using a combined approach of coarse-grained molecular simulations and experiments, we have established the binding mechanism of holoCaM and CaMKII peptide (25,26) that follows "conformational and mutually induced fit" (27). Both holoCaM and CaMKII undergoes large conformational changes in the two-step binding mechanism.

Unlike the CaMKII peptide in complex with holoCaM, which was solved by X-ray crystallography decades ago (24), the weak binding between Ng and Ca^{2+}-free CaM (apoCaM) makes the determination of its three-dimensional complex structure challenging (28). Ng belongs to the IQ motif family. In this family, the CaMBTs share a common motif (IQXXXRGXXXR), where X stands for any amino acid. Recently, Kumar et al. (29) determined the crystal structure of apoCaM and IQ motif of Ng (NgIQ), however, NgIQ was artificially covalently linked with cCaM in the bound structure. Furthermore, a previous experimental study (30) revealed that the IQ motif alone cannot represent the biophysical characteristics of the intact Ng protein. Inclusion of the acidic region in the N-terminal before the IQ motif (or $Ng_{13–49}$), which are largely composed of Asp or Glu residues, reproduced the Ng-mediated affinity of Ca^{2+} to apoCaM, as well as the intermolecular interactions between the intact Ng and apoCaM. However, the main obstacle to investigate apoCaM-$Ng_{13–49}$ binding is a lack of experimentally solved structures because of their weak binding affinity (approximately micromolar) (30). To determine the apoCaM-$Ng_{13–49}$ complex structures, we modeled the interaction between intrinsically disordered $Ng_{13–49}$ and apoCaM by introducing statistical dihedral angle potentials of $Ng_{13–49}$ (31). After obtaining the appropriate coarse-grained models of apoCaM-$Ng_{13–49}$, we reconstructed them into all-atomistic structures (32). We noted that in comparison to the canonical complex of CaM and CaMBT that includes the two domains of CaM wrapping around a helical CaMBT, the conformation of CaM in the complexes of apoCaM-$Ng_{13–49}$ remains extended instead. The conformations of Ng in the bound complexes are not entirely helical.

The reconstructed atomistic models of the holoCaM-$Ng_{13–49}$ complex and the crystal structure of holoCaM-CaMKII complex (Protein Data Bank (PDB): 1CDM) allow us to develop a hypothesis that the conformation of the Ca^{2+} binding loops from cCaM in a holoCaM-CaMBT complex dictates cCaM's affinity for Ca^{2+}. To validate this hypothesis, we computationally applied all-atomistic steered molecular dynamics (MD) simulations of holoCaM-CaMBT (33) and used Jarzynski's equality (JE) (34) to calculate the change in Ca^{2+} binding free energy. As a reference, the same simulations were performed on a conformation of holoCaM without a CaMBT (PDB: 1CLL).

One of the challenges is the modeling of the bivalent Ca^{2+} ions. It has been shown that the charge on Ca^{2+} is

$< +2e$ due to charge transfer and polarization effects in the solution and especially in the Ca^{2+}-bound state (35–37). Moreover, Jungwirth et al. (36,38) suggested that both the size and the charge require adjustment to account for the polarization effects. Following our previous work (39), we computed the charges of calcium ions according to the protonation states of CaM by employing a semiempirical quantum chemistry program MOPAC (40). Thereby we revealed a mechanism that the presence of Ng peptide destabilizes the Ca^{2+} binding to cCaM, whereas the CaMKII peptide stabilizes the Ca^{2+} binding to cCaM. We speculate that CaM's progressive mechanism of target binding, which yields distinctive complexes of apoCaM-$Ng_{13–49}$ and holoCaM-CaMKII, modulates Ca^{2+} affinities for CaM.

MATERIALS AND METHODS

Coarse-grained protein or peptide models

We performed coarse-grained molecular simulations of the intact Ng protein as well as the binding between apoCaM and $Ng_{13–49}$ peptide. Sample preparation of the intact Ng or the $Ng_{13–49}$ peptide is provided in Section I of the Supporting Material. The Hamiltonian of the coarse-grained protein models including Ng protein and the apoCaM-$Ng_{13–49}$ complex is described in Section II of the Supporting Material. Ca^{2+} ions are absent from the coarse-grained models. The details about the coarse-grained molecular simulations are included in Section III of the Supporting Material.

Free energy calculation using JE

Preparation of the structures

The preparation of the coordinates as initial conditions for the JE calculation is highly nontrivial for the computation of Ca^{2+} binding free energy ΔG. For both systems, we determined the accurate protonation states by using H++ server (41) according to the protein conformation, pH, and the ionic strength of the solution. We then assigned the partial charges based on the geometry of the proteins by using a semiempirical quantum chemistry program MOPAC (40). We applied those protonated protein structures and partial charges for minimization and further procedures to carry out all-atomistic calculations of the Ca^{2+} binding free energy (find the details in Section IV of the Supporting Material). For holoCaM and holoCaM-CaMKII, which have experimentally determined structures, we fixed the heavy atoms during the preparation; for holoCaM-Ng, we constrained the atoms with harmonic potentials for the preparation of the structures (holoCaM: calcium-bound calmodulin).

holoCaM-CaMKII and holoCaM. We fixed the positions of Ca^{2+} atoms and backbone heavy atoms from holoCaM (PDB: 1CLL) or holoCaM-CaMKII (PDB: 1CDM) during the minimization, solvation, ionization, and the equilibration procedures (see Section IV of the Supporting Material for details).

holoCaM-$Ng_{13–49}$. Since there is no experimentally determined structure, we reconstructed the all-atomistic structures from the selected coarse-grained side-chain-C_α models of apoCaM-$Ng_{13–49}$ using SCAAL method (32) (see Section IV.1 of the Supporting Material for the criteria of selecting apoCaM-$Ng_{13–49}$ models). We estimated the position of each Ca^{2+} ion as the center of mass of the side chains of the corresponding Ca^{2+}-coordinating residues from the four Ca^{2+} binding loops. We inserted the Ca^{2+} ions into these positions as an initial condition for the next minimization process with the all-atomistic AMBER99SB-ILDN force field to define their final positions. During the minimization, only the backbone heavy atoms (excluding Ca^{2+}) of the holoCaM-$Ng_{13–49}$ were constrained. It was then followed by solvation of TIP3P water molecules and ionization of Na^+/Cl^- ions, which required another round of energy minimization and equilibration. The backbone heavy atoms and the Ca^{2+} ions were constrained (but not fixed) for the holoCaM-$Ng_{13–49}$ complexes. After preparation, the positions of the Ca^{2+} ions were robust and did not drift away because 1) there were ~5–7 oxygen atoms from the side chains in the loop within 4 Å of each Ca^{2+} ion; 2) our calculation showed that the average stability of the Ca^{2+} ions at binding site III and IV is $\Delta G = -16.0$ and -12.8 kcal/mol, respectively.

Pulling simulations and computation of the free energy differences

We performed all-atomistic steered molecular dynamics (MD) simulations (see Sections IV.3 and IV.4 in the Supporting Material for the procedures and details of the steered MD simulations) to calculate the Ca^{2+} binding free energy (ΔG) using JE (34). JE relates ΔG and the irreversible work along an ensemble of non-equilibrium trajectories pulling the Ca^{2+} from the bound state to the unbound state (Eq.uation S10). The nonequilibrium trajectories (number M = 100~150) were generated by steered MD simulations with all-atomistic models. The difference in the Ca^{2+} binding free energy ΔG between holoCaM-CaMBT (CaMBT: calmodulin binding targets) and CaM is defined as $\Delta\Delta G = \Delta G^{\text{holoCaM-CaMBT}} - \Delta G^{\text{holoCaM}}$. $\Delta\Delta G$ allows us to evaluate the influence of different CaMBTs on Ca^{2+} binding to CaM, where $\Delta G^{\text{holoCaM-CaMBT}}$ and $\Delta G^{\text{holoCaM}}$ were estimated from independent simulations of pulling Ca^{2+} from the holoCaM-CaMBT and the holoCaM, respectively.

Analyses: calculation of "apparent chemical shifts"

In the NMR experimental work by Hoffman et al. (30), a semiquantitative comparison of amide chemical shifts were presented for residues in apoCaM (calmodulin without calcium ions) that are affected by binding of $Ng_{13–49}$. In the NMR spectroscopy, the chemical shift is the relative resonant frequency in the local induced magnetic field and is diagnostic of the molecular structure. A threshold was applied to generate a set of Boolean numbers, with "1" representing a significant change in the chemical shifts and "0" indicating no change. We refer to the Boolean series as "apparent chemical shifts" throughout this study. Berjanskii and Wishart (42) showed that the inverse absolute chemical shifts roughly correlates with root mean square fluctuations (RMSF), which reflects the flexibility of the structure. Therefore, to compare with the change of chemical shifts from unbound to bound apoCaM in the experiment, we approximated the amide chemical shifts by the inverse of RMSF of C_α beads and defined "apparent chemical shifts" as $\Delta = \Theta\,(\text{RMSF}^{\text{unbound}}/\text{RMSF}^{\text{bound}} - \tau)$. $\tau = 50\%$ was used. Δ measures the change of local flexibility of the apoCaM: larger values of Δ indicate stabilization of the corresponding residue upon binding with $Ng_{13–49}$.

RESULTS

Building weakly bound apoCaM-$Ng_{13–49}$ complexes in coarse-grained models

Because of the lack of experimentally determined apoCaM-$Ng_{13–49}$ complex structures, we used low-resolution information on the bound complex of apoCaM-$Ng_{13–49}$ from NMR measurements (30) for building plausible models from coarse-grained molecular simulations. The NMR experiments showed the change in the backbone chemical shifts of apoCaM upon $Ng_{13–49}$ binding (Fig. 2 *a*). It revealed the residual information on the structural changes

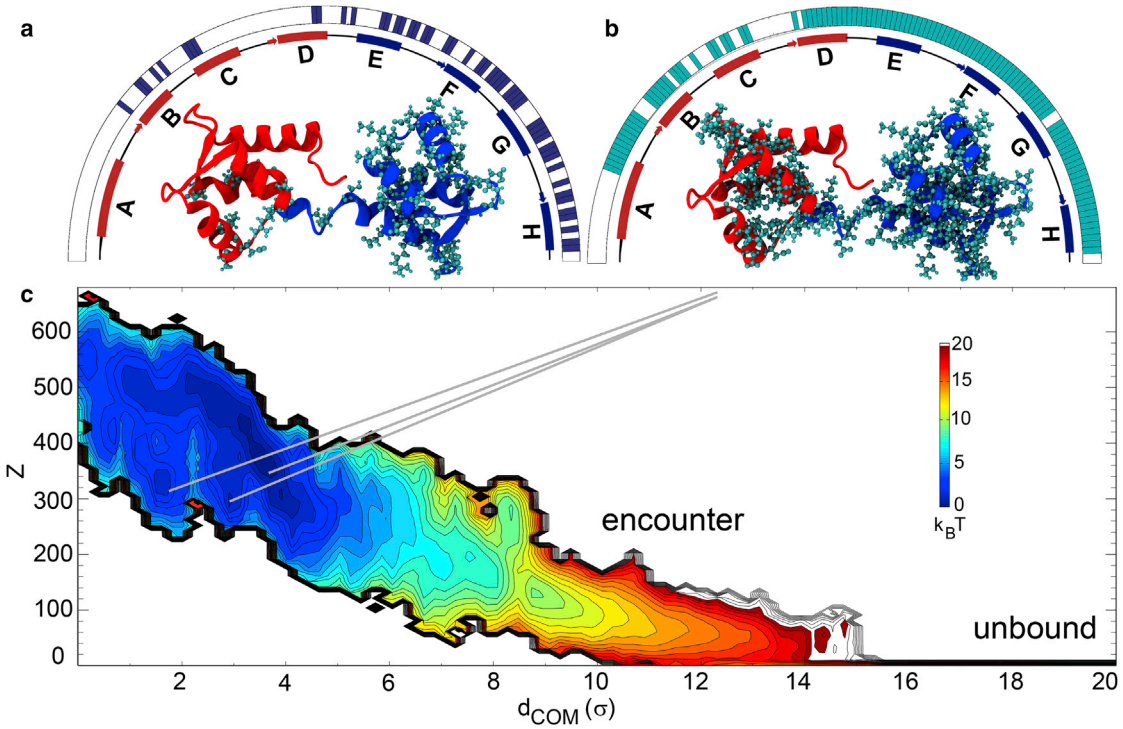

FIGURE 2 Reweighted PMF of apoCaM-$Ng_{13–49}$ binding and changes in chemical shifts for apoCaM upon $Ng_{13–49}$ binding. (*a* and *b*) The residues that experienced significant changes in chemical shifts are projected on the apoCaM structure (PDB: 1CFD) as a ball-and-stick representation from experimental measurement in (*a*) (30) and from our calculations according to the coarse-grained molecular simulations in (*b*). (*c*) The PMF from the coarse-grained molecular simulations is plotted against the center of mass distance between apoCaM and $Ng_{13–49}$ (d_{COM}) and the number of intermolecular contacts (Z). The coarse-grained molecular simulations were performed at pH = 6.3 and ionic strength = 0.1 M according to the conditions of the nuclear magnetic resonance experiments (30). We used a reduced unit of length $\sigma = 3.8$Å. The color is scaled in k_BT, $T = 1.1\ \varepsilon/k_B$, where ε is reduced unit of energy and k^B is Boltzmann constant. For visual guidance, we provided bars above the schematic representation of secondary structures in red and blue segments of apoCaM on the half circles in (*a* and *b*). To see this figure in color, go online.

of apoCaM upon binding with $Ng_{13–49}$ (30), but did not guarantee a direct contact between apoCaM and $Ng_{13–49}$. The experiment showed that the major structural changes occur in cCaM as well as Helices B and C in nCaM, as in Fig. 2 *a*.

One of the prominent reasons that we used coarse-grained simulations is to achieve efficient sampling of the conformational space (see Sections I–III in the Supporting Material for details about the coarse-grained simulations). We implemented the statistical potentials into the structural Hamiltonian of the $Ng_{13–49}$, as well as the intermolecular interaction between apoCaM and $Ng_{13–49}$. For the modeling of apoCaM, we kept limited structural information that favors an extended conformation and matches the structural description of the experimental measurements (25,43). We found that apoCaM-$Ng_{13–49}$ complex samples an ensemble of varying conformations through nonspecific intermolecular interactions, indicated by the broad distribution of the bound complex in the potential of mean force (PMF) (see Fig. 2 *c*).

Once we generated the ensemble of bound complexes, we strategically selected the most probable complex structures of apoCaM-$Ng_{13–49}$ with the previously mentioned NMR measurement as a guide (Fig. 2 *a*) by following the procedures: 1) we reweighed the ensemble of apoCaM-$Ng_{13–49}$ structures from the umbrella sampling simulations (44) using Weighted Histogram Analysis Method (45,46), and employed an importance sampling method (32) (see Section V.4 in the Supporting Material) to select a subset of 23,722 structures from five million independently sampled conformations. The surprisal value between the sample distribution $P^{\text{sample}}(d_{\text{COM}})$ and the original unbiased distribution $P^{\text{ori}}(d_{\text{COM}}) = 0.14$ to show that the sample distribution represents the original distribution well. 2) We performed a clustering analysis on the sampled structures (see Section V.5 in the Supporting Material for the clustering method); 3) we computed the "apparent chemical shifts" (see Analyses: calculation of "apparent chemical shifts" for the definition, Fig. 2 *b*) from the major clusters. The population of the major four clusters (comprising 86%, Table S2) and the correlation coefficients between computationally obtained "apparent chemical shifts" from the four clusters and the experimental NMR data are shown in Table S2. We identified the cluster (the dominant cluster) of complex structures, which best matched the experiment (30), with a correlation coefficient of 0.36. From the dominant cluster (i.e., cluster 1 in Table S2), cCaM is mostly stabilized by binding with $Ng_{13–49}$ and helices B and C are partially stabilized (Fig. 2 *c*).

We next analyzed the structures of the weakly bound apoCaM-$Ng_{13–49}$ by comparing the probability of contact formation from the unbound state (defined by $d_{COM} \approx 20$ σ in Fig. 2 *c*) in Fig. S3. In the selected cluster (cluster 1) of the apoCaM-$Ng_{13–49}$ complexes, contacts within apoCaM (triangles in solid lines in Fig. S3 *c*) became less probable upon $Ng_{13–49}$ binding. In addition, the probability of interdomain contacts between Helix A from nCaM and inter-domain linker, helices G and H from cCaM (the solid rectangles in Fig. S3 *c*) decreased to allow interaction with $Ng_{13–49}$ (the solid rectangles in Fig. S3 *a*). In contrast, after binding with apoCaM, $Ng_{13–49}$ diminished interactions between IQ motif and the acidic region (Fig. S3 *b*, the rectangle in dotted lines); subsequently, probability of forming α-helix structure in the IQ motif increases (Fig. S3 *b*, the rectangle in solid line). Provided that there was no bias to the formation of α-helix structure in $Ng_{13–49}$, our model predicted formation of secondary structures in the IQ motif after binding with apoCaM.

For the intermolecular interaction of the weakly bound complex from the main cluster, IQ motif of $Ng_{13–49}$ interacted with Helix A, B/C Helix linker, and Helix D from nCaM, and interdomain linker, Helix E and Helix H, and the F/G Helix linker from cCaM (Fig. S3 *a*, ellipse in dotted lines). These regions from CaM (either apoCaM or holoCaM) also participate in canonical binding of holoCaM with CaMKI or CaMKII (25). Although the IQ motif of $Ng_{13–49}$ formed a helical structure and interacts with cCaM of high probability, the acidic region of $Ng_{13–49}$ remains unstructured (Fig. S3 *b*, no change in contact formation along diagonal for the acidic region). It is highly probable for the acidic region to interact with both nCaM and cCaM, especially with F/G helix linker and the central linker (Fig. S3 *a*, ellipse in solid lines). We further calculated the correlation among these pairwise contacts in Section V.6 of the Supporting Material, and found that the interaction between acidic region of $Ng_{13–49}$ and Ca^{2+} binding loops from apoCaM anticorrelated with those within cCaM (shown in magenta ellipse in Fig. S4).

Molecular mechanism of apoCaM-$Ng_{13–49}$ binding

We next investigated the conformational changes of apoCaM-$Ng_{13–49}$ during the association process along their center-of-mass separation (d_{COM}) for better understanding of the molecular mechanism of binding. The unbound structure was taken from $d_{COM} = 20.0$ σ, where the interaction between the two proteins is negligible; the encounter state structure (25) was taken from $d_{COM} = 10.0$ σ, where the two proteins began to make contact; the bound state structure was take from $d_{COM} = 2.8$ σ, where the PMF(d_{COM}) reached the minimum (see the magenta curve Fig. S2 for the PMF). Superposed structures of apoCaM-$Ng_{13–49}$ complex at $d_{COM} = 20.0$, 10.0, and 2.8 σ are shown, representing unbound, encounter, and bound states, respectively (Fig. 3).

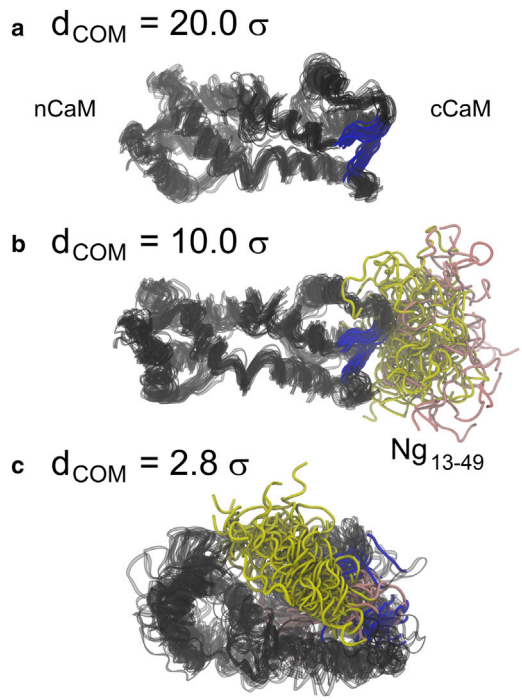

FIGURE 3 Illustration of the structural changes in unbound, encounter, and bound ensemble of CaM-$Ng_{13–49}$ complexes. Structures of CaM in (*a*) was taken from the unbound state when CaM and $Ng_{13–49}$ are well separated at $d_{COM} = 20.0$ σ. $\sigma = 3.8$ Å. (*b*) Structures from the encounter of binding when apoCaM and Ng is separated at $d_{COM} = 10.0$ σ. (*c*) Structures from the bound state at $d_{COM} = 2.8$ σ. For visual guidance, we superposed 20 sets of structures in each panel. The CaM is colored in black; the residues (residues 99~101 in Ca^{2+} binding loop III and residues 135~137 in Ca^{2+} binding loop IV), which form EF-hand β-scaffold in cCaM are colored in blue; the acidic region and IQ motif of $Ng_{13–49}$ are colored in pink and yellow, respectively. To see this figure in color, go online.

There was no apparent structural change in apoCaM from the unbound state (Fig. 3 *a*) to the encounter complex (Fig. 3 *b*) that the conformation of apoCaM remains extended. The onset of binding in Fig. 3 *b* shows that the IQ motif of the $Ng_{13–49}$ mainly interacted with cCaM, which was not involved in a global conformational change of apoCaM. From the encounter complex (Fig. 3 *b*) to the bound state (Fig. 3 *c*), $Ng_{13–49}$ remained partially structured and mainly gained interactions with nCaM and local conformational change of apoCaM was observed. In particular, the EF-hand β-scaffold (47), which stabilizes the two helix-loop-helix motifs (EF-hand) in cCaM (see Fig. S5 for illustration of the EF-hand β-scaffold), is broken by insertion of the acidic region of $Ng_{13–49}$.

To understand their binding at the residual level, we plotted the probability of contact formation at $d_{COM} =$ 20.0, 10.0, and 2.8 σ in Fig. S6. At $d_{COM} = 20.0$ σ, or the unbound state, there was no contact formed between apoCaM and $Ng_{13–49}$ (in Fig. S6 *a*). The contacts within $Ng_{13–49}$ were ubiquitous and of low probability, corresponding to the transient nature of conformations of an intrinsically disordered protein/peptide (IDP). Upon initial

binding with apoCaM at $d_{COM} = 10.0\ \sigma$, only IQ motif from the $Ng_{13–49}$ gained contacts with Helix F and Ca^{2+} binding loop IV from cCaM. This indicates a crucial role of the IQ motif in the recognition by apoCaM. Those contacts were kept when d_{COM} became 2.8 σ, which suggests that the IQ motif also facilitated the stabilization of the apoCaM-$Ng_{13–49}$ complex. At this stage, the contacts between the acidic region of $Ng_{13–49}$ and both of nCaM and cCaM formed broadly but with low probability. For apoCaM, from unbound to bound states, most of the contact formation within apoCaM remained the same (of high probability), except when $d_{COM} = 2.8\ \sigma$, the side chain-side chain contacts between Helix E and Helix H, between Helix F and Ca^{2+} binding site IV, and the β-sheet contacts between Ca^{2+} binding sites III and IV (encircled in Fig. S6). In particular, the probability of the contacts that form the EF-hand β-scaffold (encircled) became less when the acidic region of $Ng_{13–49}$ formed contacts with cCaM. Ng became partially helical in the bound form. We will relate this finding to its importance of specific functions under Discussion.

Binding of the CaMBTs to CaM modulates the Ca^{2+} binding affinity validated by atomistic pulling simulations

We found that interactions between the acidic region of $Ng_{13–49}$ and the Ca^{2+} binding loops from cCaM-apoCaM competed with those within apoCaM. We hypothesized that, due to this competition, the EF-hand β-scaffold, which is shown to control the opening and closing of the EF-hands (47,48), become less stable, thus facilitating the release of Ca^{2+} from CaM. To test this hypothesis, we strategically selected apoCaM-$Ng_{13–49}$ complex structures from the most dominant cluster (i.e., cluster 1 in Table S2) in the coarse-grained molecular simulations for the calculation of Ca^{2+} binding free energy. The selection was based on the characteristics of apoCaM binding with either Ng protein or Ng peptides according to the results from several experimental studies: 1) EF-hands in cCaM are open and EF-hands in nCaM are closed from X-ray crystallography and NMR experiments (29,49); 2) $Ng_{13–49}$ has more interactions with cCaM as suggested from NMR studies, $\Delta Z = Z_n - Z_c < 0$ (30). Z_n (Z_c) is the number of intermolecular contacts between Ng and nCaM (cCaM). (See Section V in the Supporting Material for the definition of Z.) This was further shown by the EF-hand angles in CaM in forms of Ca^{2+}-absent, Ca^{2+}-loaded, apoCaM-NgIQ, and holo-CaM-CaMKII in Table S3. As a result, four coarse-grained structures were selected from the major cluster (cluster 1) for further reconstruction of all-atomistic protein models with four Ca^{2+} ions included (see (50) for the protocol of reconstructing the all-atomistic model from a side-chain-C_α model).

We evaluated the free energy difference between Ca^{2+}-unbound and Ca^{2+}-bound state ($\Delta G = G_B - G_U$, see Fig. 4 for illustration) using JE (34,51) by pulling the two Ca^{2+} ions independently from the Ca^{2+} binding sites III and IV of the cCaM (see simulations details in IV.5 from the Supporting Material). We further computed the free energy changes in the absence or the presence of a target ($\Delta\Delta G = \Delta G^{holoCaM\text{-}CaMBT} - \Delta G^{holoCaM}$, see Fig. 4 for illustration). $\Delta\Delta G$ allows us to evaluate the influence of CaMBT on Ca^{2+} binding to CaM. If $\Delta\Delta G > 0$, it means that the CaMBT destabilizes the bound state and thus decreases the Ca^{2+} affinity, as illustrated in Fig. 4. If $\Delta\Delta G < 0$, CaMBT enhances Ca^{2+} affinity for CaM.

We found that the distribution of work values does not follow Gaussian distribution (Fig. S14). The estimation of free energy difference ΔG could be very inaccurate by directly applying JE or its second-order cumulants approximation (51) (we reported the direct estimation from JE in Table S6). As shown in Fig. S13, the binding free energy ΔG was estimated for Ca^{2+} at binding site III or site IV of holoCaM and holoCaM-CaMKII using a running JE estimate or a block-average method (BA, i.e., using subsets of all available work data using Jarzynski's equality). ΔG converged to values, which deviate from experimental values (~ −5 to −4 kcal/mol). To improve the efficiency of the free energy estimation, we implemented the cumulative integral (CI) extrapolation method developed by Ytreberg and Zuckerman (51). CI uses an integral for more accurate estimations than the linear extrapolation method by extrapolating to $1/n \rightarrow 0$, where n is total number of work values. It was shown that CI extrapolation could reduce the required data by 5- to 40-fold (51,52). Therefore, we adopted the CI extrapolation method to estimate the free energy difference.

We summarized the binding free energies of Ca^{2+} estimated by the CI extrapolation in Table 1 for a total of six

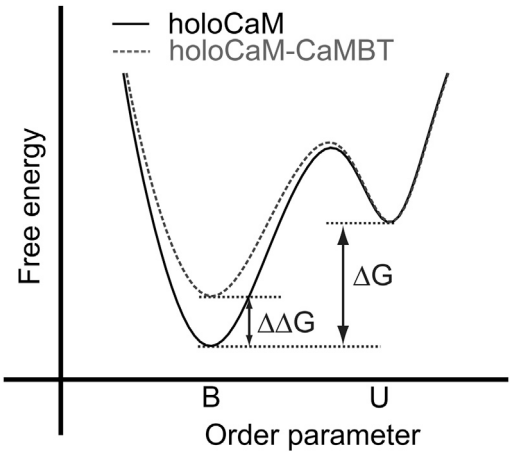

FIGURE 4 Illustration of the definitions of the binding free energy ΔG and $\Delta\Delta G$. $\Delta G = G_B - G_U$. $\Delta\Delta G = \Delta G^{holoCaM\text{-}CaMBT} - \Delta G^{holoCaM}$. B and U stand for bound and unbound states of the Ca^{2+}, respectively. $\Delta\Delta G > 0$ means that the CaMBT destabilizes the bound state and thus decreases the Ca^{2+} affinity.

TABLE 1 Difference in Binding Free Energy of Ca^{2+} Calculated from Nonequilibrium Molecular Simulations and from the Experiments at pH = 7.4.

		holoCaM-Ng_{13-49}	holoCaM-CaMKII
$\Delta\Delta G^{CI}$ (kcal/mol)	Site III	9.2 ± 2.2	−2.5
	Site IV	22.4 ± 0.9	−1.7
$\Delta\Delta G^{exp}$ (kcal/mol)	Site III/IV	2.5	−3.3

The CI extrapolation was used in the calculation of binding free energy of Ca^{2+} $\Delta\Delta G^{CI}$ from the simulations. The averaged value of $\Delta\Delta G$ statistical errors from the four conformations of the holoCaM-Ng_{13-49} complex are provided. The experimental binding free energy $\Delta\Delta G^{exp}$ for holoCaM-Ng_{13-49} was from literature (30) and the experimental values of the holo-CaM-CaMKII were from (14).

conformations: four holoCaM-Ng_{13-49} (showing averaged values), one holoCaM and one holoCaM-CaMKII. For all of the four conformations of holoCaM-Ng_{13-49}, the binding of Ng_{13-49} to CaM destabilized the Ca^{2+} in site III and site IV, respectively, by showing an averaged positive $\Delta\Delta G$ (Table 1). In comparison, we examined the importance of the canonical bound structure for the release of Ca^{2+} from the complex structure of holoCaM-CaMKII (PDB: 1CDM) where the CaM wraps around and binds with a well-formed helical structure of CaMKII peptide at an antiparallel position. From CI extrapolation, each site presented a negative $\Delta\Delta G$, implying that the presence of CaMKII stabilizes the Ca^{2+} in both site III and site IV from CaM.

Conformation of bound CaM-CaMBT complex dictates Ca^{2+} release

HoloCaM-CaMKII forms a canonical bound complex, while apoCaM-Ng_{13-49} does not. In the following, we unveiled the molecular mechanism that governs varying Ca^{2+} binding affinity for cCaM in complexes of CaM and distinct CaMBTs. We examined one holoCaM-Ng_{13-49} model reconstructed from the coarse-grained structure that led to a maximal increase of Ca^{2+} binding free energy, as well as the crystal structure of holoCaM-CaMKII (PDB: 1CLL).

In Fig. 5 *a*, Ng_{13-49} is bent from the middle between residues P24 and G25 from the acidic regions (in *green* and in *gray* on the amino acid sequence). On the one hand, the IQ motif interacts with Helix F, Helix G, and F/G helix linker from cCaM to stabilize the holoCaM-Ng_{13-49} complex. On the other hand, the acidic N-term interacts with both nCaM and cCaM. Surprisingly, the four residues D22-D23-P24-G25 (DDPG) (in ball-and-stick representation) before the IQ motif on Ng_{13-49} stick out and insert into the middle of the EF-hand β-scaffold (colored in *blue* segments) between Ca^{2+} binding loops III and IV from cCaM, disrupting the stability of EF-hand β-scaffold. As a result, the distance between the two β-strands becomes 7.9 ± 1.2 Å (average over the four conformations)

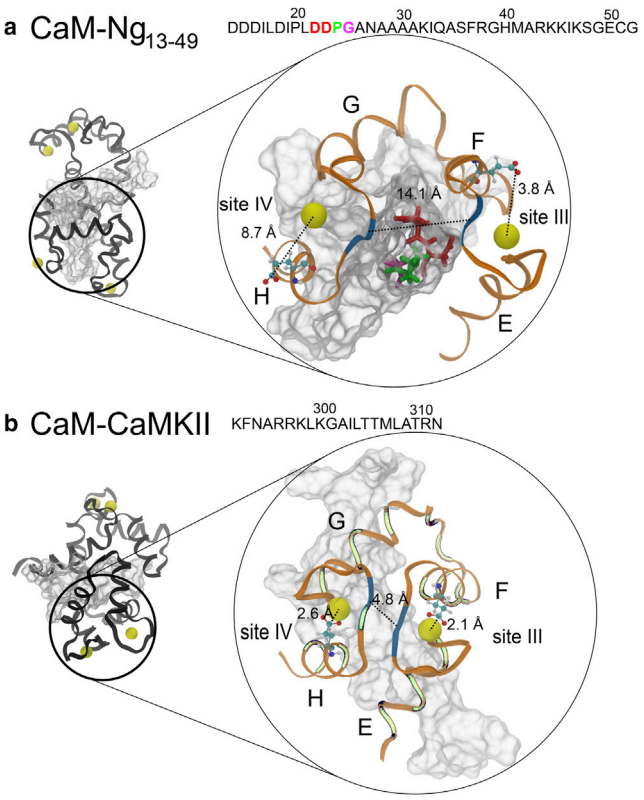

FIGURE 5 Differential intermolecular interactions in holoCaM-Ng_{13-49} and holoCaM-CaMKII complex structure tune Ca^{2+} binding affinity for cCaM. (*a*) The atomistic structure of holoCaM-Ng_{13-49} complex was reconstructed from the coarse-grained simulations that led to the largest decrease in Ca^{2+} affinity for site III and site IV. (*b*) The structure of holoCaM-CaMKII complex, showing a wrap-around binding pattern, is from the X-ray crystallography (PDB: 1CDM). The complete structures of the complexes are on the left, CaM is in black, Ca^{2+} ions are represented by yellow beads, and the CaMBTs are in white. The sequences of the CaMBTs are provided above the figures. For visual guidance, the cCaM and the CaMBTs are enlarged and recolored on the right. The cCaM is in ribbon representation and colored in orange. The residues forming EF-hand β-scaffold in the crystal structure, i.e., residues Y99, I100 from site III, and residues Q135, V136 from site IV, are colored in blue. The invariant Glu residue in the 12th position of each Ca^{2+}-binding loop contributes two oxygen atoms to the coordination of the Ca^{2+} ion and is shown in ball-and-stick representation. The dotted lines show the distance between Ca^{2+} and the two oxygen atoms from Glu12 and the distance between the two β-strands in cCaM. The four residues that stick out from Ng_{13-49} are D22 (*red*) D23 (*red*) P24 (*green*) G25 (*magenta*) in (*a*).The residues from CaM, which form contacts with CaMKII peptides, are colored in light green in (*b*). To see this figure in color, go online.

comparing to ~5.0 Å in the intact holoCaM (PDB: 1CLL). In contrast, the distance decreases slightly to 4.8 Å in holoCaM-CaMKII (PDB: 1CDM).

Moreover, the average distance between the Ca^{2+} and the Ca^{2+} coordinating oxygen atoms from the 12th residue (Glu-12), or $\overline{d}^{Glu\text{-}Ca}$, in site III and IV from cCaM is 4.5 ± 0.7 and 3.7 ± 0.5 Å, respectively (Fig. 5 *a*). As a reference, for Ca^{2+} in site III and site IV of the crystal structure of holoCaM without a target (PDB: 1CLL), $\overline{d}^{Glu\text{-}Ca}$ = 2.3 and

2.5 Å, respectively. The 12th residue in the Ca^{2+} binding loop, or Glu-12, plays a crucial role in stabilizing Ca^{2+} in the binding loop (1,47). However, the presence of the acidic region from Ng_{13-49} increases $\overline{d}^{\text{Glu-Ca}}$ in the holoCaM-Ng_{13-49} complex (Fig. 5 *a*), especially for site IV, which is slightly closer to the center of mass of the DDPG motif than site III. The motif DDPG from the acidic region of Ng_{13-49} pries open the EF-hand β-scaffold and facilitates the release of the Ca^{2+} from cCaM. In comparison, we examined the importance of the canonical bound structure for the release of Ca^{2+} from the crystal structure of holoCaM-CaMKII (PDB: 1CDM) in Fig. 5 *b*. In contrast, $\overline{d}^{\text{Glu-Ca}}$ for sites III and IV (Fig. 5 *b*) are similar to those in the crystal structure of holoCaM without a target (PDB: 1CLL).

To understand the evolutionary role of the acidic region of Ng_{13-49} and its biological significance in modulating Ca^{2+} binding affinity of CaM, we performed sequence alignment of Ng from different species available in the protein knowledgebase (UniProtKB, http://www.uniprot.org/uniprot/) (Table S14). Sequence alignment results of Ng revealed that the amino acids in the acidic region are either conserved across the species or found as amino acids with similar physicochemical properties (see Fig. S15). In particular, we noted that the cluster of residues "DIPLDDP" from the acidic region of Ng that show anti-correlation with the binding loops III and IV of CaM based on the pairwise contacts (see Fig. S4; Table S8) are conserved across all the species (see Fig. S15). Moreover, the "DDPG" motif from the acidic region of Ng is conserved (see Fig. S15) across the higher vertebrate species (from rodent to human; see Table S14). Interestingly, in the lower vertebrate species the Gly (G) residue of this motif is found as Asp (D) (in reptiles and birds) or Glu (E) (in fish) or Ala (A) (in zebrafish), whereas the rest of the residues (DDP) are conserved across all the vertebrates (see Fig. S15). We also found that the DDPG motif, which maintains close proximity to the middle part of both the Ca^{2+} binding loops III and IV in the holoCaM-Ng_{13-49} complex (see Fig. S16) plays an important role in disrupting the EF-hand β-scaffold in CaM as we mentioned before. We speculate that the acidic region of Ng and especially the DDPG motif is evolutionarily important for their biological role in suppressing and/or fine-tuning the Ca^{2+} binding affinity of CaM.

DISCUSSION

The energy landscape of apoCaM-Ng_{13-49} binding is rugged

The weakly bound complexes are commonly observed in protein-protein associations involving intrinsic disordered proteins (IDPs) in both experiments (53,54) and computer simulations (55–60). Long-range electrostatic interactions enhance the rate of association of several weakly bound complexes (61). Generally, the nonspecific interactions, including hydrophobic and electrostatic interactions, are known to be a major contributor to the weakly bound complexes involving IDPs. Sugase et al. (54) showed from NMR experiments that the weakly bound transient complexes of the phosphorylated kinase inducible domain (pKID) of transcription factor CREB and the kinase-induced domain of the CREB-binding protein were predominantly stabilized by nonspecific hydrophobic contacts instead of electrostatic interactions. The importance of nonspecific intermolecular interactions was further addressed by computer simulations on phosphorylated kinase inducible domain and kinase-induced domain by Turjanski et al. (55), as well as on nuclear-receptor coactivator binding domain of CREB-binding protein and the p160 steroid receptor coactivator ACTR by Chen et al. (57). In both cases structure- or topology- based models (62) were used for the intermolecular interactions that define a specific bound structure.

Determination of the structure of the apoCaM-Ng_{13-49} complex experimentally is challenging due to its low stability (28). To overcome this problem, we incorporated statistical nonbonded interactions and electrostatic interactions, which do not rely on a specific complex structure, in our protein model. In addition, to model the disordered Ng_{13-49}, we implemented a sequence-based statistical dihedral angle potential (31) in the Hamiltonian. Our study reveals a rugged free energy landscape of apoCaM-Ng_{13-49} binding (Fig. 2 *c*). The conformations of apoCaM-Ng_{13-49} thermally fluctuate between shallow basins separated by 1~2 k_BT ($T = 1.1\ \epsilon/k_B$, where ϵ is reduced unit of energy and k^B is Boltzmann constant.).

The weakly bound complexes comprises "fuzzy" structures, in which either one of the binding partners or both stay dynamic (63). According to our simulations of the apoCaM-Ng_{13-49} complex, Ng_{13-49} remains partially structured as shown in the low probability of contact formation within Ng_{13-49} (Fig. S3 *b* and S6 *c*). Our model of the CaM model (25) samples a wide range of states by incorporating non-native interactions between residues. This idea of modeling beyond a simple structure-based model meets the experimental exploration of CaM with single-molecule experiments that its energy landscape is rather rugged and complex with plenty of intermediates (19). Our approach differs from the approaches adopted by those studies that the conformational changes of the target were ignored. For example, in the investigation of the interactions between the transcriptional activation domain of an oncoprotein and its target by Chan et al. (64), the target protein was modeled as a sphere of distributed charges on the surface.

To generalize, the structural flexibility and nonspecific interactions, corresponding to entropic and enthalpic effects, respectively, result in the rugged energy landscape of apoCaM-Ng_{13-49} that presents multiple weakly bound complexes without a global minimum.

IQ motif of $Ng_{13–49}$ is crucial for the initial binding of apoCaM

It was postulated that the residual structure of an IDP is necessary for molecular recognition in CaM-CaMBT complex formation (65). We found that the residual structure in Ng is crucial for recognition and binding with CaM. We found that the IQ motif started to interact with cCaM at a distance of $d_{COM} = 10\ \sigma$ and remained during the binding process, indicating its key role in binding with apoCaM. Creamer et al. (66) using CaM-CaMBT systems as well as Clarke et al. (67) using the BH3 motif of the largely disordered protein p53 upregulated modulator of apoptosis (PUMA) and a folded protein induced myeloid leukemia cell differentiation protein (MCL-1), however, postulated that the residual structure of the IDP did not affect binding kinetics, but weakened the binding affinity upon helix-breaking mutation. It is important to note that in Creamer's study, most of the CaMBTs bind CaM in a canonical form where CaM undergoes large global conformational changes. In such scenarios, the "conformational and mutually induced fit" mechanism (25) is needed for the recognition and binding (see The binding mechanism of CaM-CaMBT might be key to regulating Ca^{2+} binding affinity on the binding mechanism of CaM-CaMBT). However, Ng predominantly interacts with cCaM and does not form a full helical structure, therefore the binding does not require global conformational changes of CaM.

The acidic region is key to $Ng_{13–49}$ for tuning CaM's affinity for Ca^{2+} comparing to other IQ-motif peptides

Our computation of the change in free energy of Ca^{2+} binding to cCaM from nonequilibrium atomistic pulling simulations reveals that the target peptides $Ng_{13–49}$ and CaMKII have opposing effects on Ca^{2+} binding affinity to cCaM. It is important to note that the CaMKII peptide binds to CaM in a canonical form in which the two domains of CaM wrap around a helical CaMKII peptide (24), whereas $Ng_{13–49}$ forms weakly bound complexes with a rather extended CaM. Such a difference results in distinct conformations of the EF-hand motifs and the Ca^{2+} binding loops in CaM.

The binding and dissociation of Ca^{2+} to/from cCaM essentially depend on the two EF-hand motifs, which are connected by the F/G helix linker and are rigid in holoCaM (Fig. S5). This construct of EF-hand motif pair is often referred as the "EF-hand β-scaffold" (47) proposed by Grabarek (47). The EF-hand β-scaffold immobilizes the Ca^{2+} ion during its initial binding to the Ca^{2+} binding site. Our findings with regard to target binding that modulate CaM's calcium affinity is supported by experimental observations (1). An enhanced structural stability of an EF-hand β-scaffold through target binding in a canonical CaM-CaMBT form increases CaM's calcium affinity, which is discussed in the detailed review by Gifford et al. (1). In those experiments, a severe disruption in the Ca^{2+} binding loops due to the repacking of cCaM causes the Ca^{2+} coordinating oxygen atoms from the Glu-12 of the Ca^{2+} binding loop to move away from their Ca^{2+} binding positions and the release of the Ca^{2+} (68). In addition, several experiments showed that when CaMBT peptides bind to CaM canonically, the Ca^{2+} affinity for cCaM increases (69): $\Delta\Delta G$ varies from -1.5 kcal/mol for phosphorylase kinase, -3.3 kcal/mol for CaMKII (70) to -4.8 kcal/mol for β-calcineurin (71).

We found from this work that during the binding process, on one hand, IQ motif of $Ng_{13–49}$ carries the function of recognition and stabilization. On the other hand, the role of the acidic region of $Ng_{13–49}$ might relate to the biological functions of the protein because of its influence on the stability of EF-hand β-scaffold (Fig. S6 *c*) regulating Ca^{2+} binding to CaM. Ng belongs to the IQ motif CaMBTs family. IQ-motif CaMBTs such as the sodium channel ($Na_V1.2$), the calcium channel ($Ca_V1.1$ and $Ca_V1.2$), PEP-19, and Ng present distinct effects on CaM's affinity for Ca^{2+}.

The presence of the acidic region rather than the IQ motif determines whether the interactions between CaM and the IQ-motif CaMBT are disruptive or constructive. For $Na_V1.2$ and PEP-19 (Purkinje cell protein 4; PCP 4) that contain an IQ motif and a previous acid region, both of them bind to cCaM. Binding of them lowers the Ca^{2+} binding affinity for cCaM: binding of the $Na_V1.2$ causes an increase in $\Delta\Delta G$ of about $+2$ kcal/mol (72) and PEP-19 an increase of about $+0.2$ kcal/mol (49). However, for $Ca_V1.1$ and $Ca_V1.2$, which contain an IQ motif but no such previous acidic region, they behave like a canonical binding motif, where the CaMBT forms a helical structure wrapped by the compact CaM, and stabilize the Ca^{2+} binding by $\Delta\Delta G = -1$ and -2.6 kcal/mol, respectively (73). The results support the active role of the acidic region, but do not imply the molecular mechanism of how it regulates Ca^{2+} binding to CaM. By carrying out multiscale molecular simulations, we identified the key residues DDPG in the acidic region of $Ng_{13–49}$ that "pry" open the β-sheet structure between the two Ca^{2+} binding loops.

The binding mechanism of CaM-CaMBT might be key to regulating Ca^{2+} binding affinity

A number of recent advancements have been made in experimental and computational approaches to examine protein-protein interactions involving IDPs (74–78). A key emerging concept is the idea of "specificity on demand" where protein interactions are dictated by concerted interactions between both binding partners. From this view, the process of binding cannot be inferred even if the structures of the initial reactants and final complex are known. The importance of structural flexibility motivated the development of theories for binding mechanisms (79–83), such as

fly casting (61,84), folding upon binding (58,85), "dock-and-coalesce" (74,86), and conformational selection (87,88). These multiple steps mechanisms underlie possible kinetic bottlenecks due to subsequent conformational changes in a post-collision event (74,86). Despite differences in these theories, past efforts that focus on IDPs binding to known structures have not dealt with the complexities of the conformational variability inherent in the binding process that can ultimately lead to different final products involving potentially large changes in both partners. This view is highlighted by our efforts that revealed that CaM binding to its targets is mediated by a conformational and mutually induced fit model (25).

In our previous study (25), we discovered that the binding between CaM and CaMBTs that forms a canonical bound complex, such as the CaMKII peptide, require a "conformational and mutually induced fit" mechanism. In this mechanism, the binding of the CaMKII peptide and CaM presents a two-stage process: at the onset of binding (as in an early stage), the binding is driven by diffusion and electrostatic steering effect that both CaM and CaMKII do not undergo large conformational changes. At the late stage, both experience large conformational changes that the two domains of CaM wrap around the emerging rod-like CaMKII peptide. The latter event of binding dictates a small, but significant, difference in the association rates of CaM with various CaMBTs. However, this is not the case for apoCaM-$Ng_{13–49}$ binding. After the onset of binding where, the IQ motif of $Ng_{13–49}$ binds to cCaM, there is no subsequent large conformational change of CaM.

The difference in the bound complexes of CaMBT-CaM along its progressive binding mechanism delineates their effects on Ca^{2+} binding affinity. The Ca^{2+} affinity is increased by binding of CaMKII peptide. The EF-hand motifs of CaM wrap around the rod-like CaMKII peptide; hence, they enhances the stability of the "EF-hand β-scaffold" that retains Ca^{2+} in Ca^{2+} binding loops. In contrast, the Ca^{2+} affinity is decreased by binding of Ng peptide because of lack of such wrap-and-enhance pattern of CaM. In addition, insertion of the DDPG motif into the "EF-hand β-scaffold" further destabilizes the Ca^{2+} binding loops. Therefore, we speculate that the benefit of a progressive binding mechanism of CaM that can be modulated by CaMBTs underlies their distinct effects on CaM's Ca^{2+} binding affinities.

Force fields affect the magnitude of free energy estimation

From Table 1, the signs of ΔΔG from our computed values agreed with the experimental measurements. For holoCaM-CaMKII, they were in the same order of magnitude, while for the holoCaM-$Ng_{13–49}$ complexes, the computed values were about one order of magnitude larger than the experimental measured ones. Majorly the force field contributed to the accuracy in the calculation of ΔΔG. We break the reasoning down into three categories relating to force fields:

Initial structures might matter

We found that the current all-atomistic force fields, including AMBER99SB-ILDN and CHARMM27, alter the initial structures after minimization and equilibration if the positions of the heavy atoms were not strictly fixed in space. This is true for coordinates from the crystal structures of holoCaM and holoCaM-CaMKII, as well as the all-atomistic structures of holoCaM-$Ng_{13–49}$ reconstructed from coarse-grained models. The shifts in the positions are especially prominent for the Ca^{2+} ions. According to the study by Shukla et al. (89) on the same structure of holoCaM (PDB: 1CLL) using the AMBER99SB-ILDN force field, the root mean-square deviation of the equilibrated structure from the crystal structure was around 1 Å. This is in agreement with our simulations using the same force field that after minimization and equilibration without fixing the heavy atoms, the displacement between the Ca^{2+} atom and the center of mass of the corresponding Ca^{2+} binding loop (d^{Ca}) changed by δd^{Ca} ~0.5 Å for holoCaM, δd^{Ca} ~1.0 Å for holoCaM-CaMKII, and δd^{Ca} ~1.0–2.1 Å for holoCaM-$Ng_{13–49}$. This shift in d^{Ca} affected the free energy of any Ca^{2+} bound state. For a crude estimation of the free energy differences due to a shift in d^{Ca}, suppose the average force is 500 kJ/mol/nm, an underestimation of $\Delta G^{holoCaM}$ would be by 6 kcal/mol if $\delta d^{Ca} = 0.5$ Å. Likewise, an underestimate of $\Delta G^{holoCaM\text{-}CaMKII}$ would be by 12 kcal/mol if $\delta d^{Ca} = 1$ Å. Such deficiency would cause a net underestimation of $\Delta\Delta G^{holoCaM\text{-}CaMKII}$ by 6 kcal/mol.

Therefore, we adopted the following strategy in our estimation of ΔΔG: for a system that has experimentally determined structure, such as holoCaM or holoCaM-CaMKII, we fixed the positions of the Ca^{2+} atoms and the backbone heavy atoms during the energy minimization and equilibration simulations; otherwise, for holoCaM-$Ng_{13–49}$ which has no experimentally determined structure, we reconstructed the all-atomistic coordinates from coarse-grained models of apoCaM-$Ng_{13–49}$ and constrained the positions of the Ca^{2+} ions and the backbone heavy atoms.

We determined the initial positions of the Ca^{2+} ions carefully. We estimated the position of each Ca^{2+} ion as the center of mass of the side chains of the corresponding Ca^{2+} coordinating residues and minimized the potential energy with the all-atomistic AMBER99SB-ILDN force field for optimization. During the energy minimization, only the backbone heavy atoms (excluding Ca^{2+}) of holoCaM-$Ng_{13–49}$ were constrained. Afterward, 1) there were ~5–7 oxygen atoms in the loop within 4 Å of each Ca^{2+} ion; 2) later calculation showed that the Ca^{2+} ions at binding site III and IV had average $\Delta G = -16.0$ and -12.8 kcal/mol, respectively. In summary, the final positions of the Ca^{2+} ions of holoCaM-$Ng_{13–49}$ in the initial structures were

robustly determined. Our results with respect to the sign of the differences in the free energy calculation whether targets are present or not were not affected.

The protonation states of the proteins might matter

ApoCaM is a highly charged protein with a net charge of $-24e$, where e is the elementary charge. The IDP target peptide Ng_{13-49} contains 12 charged residues. Therefore, accurate determination of the partial charges according to the protonation states particularly at the Ca^{2+}-binding loops impacts the estimation of the binding free energy of Ca^{2+} for CaM. The standard procedure of assigning protonation states according to a neural pH and its pKa eludes the correct prediction of the actual protonation states for Glu and Asp amino-acids in the Ca^{2+}-binding loops of CaM. Therefore, it requires a careful treatment on the protonation states. We employed the H++ program (41) to predict the protonation states based on pH and electrostatic energies that vary with the protein conformation and ionic strength of the solution.

After determining the protonation states of the proteins (complexes) at pH = 7.4 and ionic strength of = 0.15 M using H++ server (41), we found that there was a significant change in the net charges in comparison with the outcome from the standard protonation protocol of the MD package GROMACS (90). In the standard protocol, the Asp and the Glu amino acids in the Ca^{2+}-binding are deprotonated at neutral pH and they carry negative charges. However, from the predictions by H++, they are protonated instead. For holoCaM, the newly protonated residues by H++ included D20, D56, E67, D93, and D129 from the Ca^{2+} binding loops I–IV. For holoCaM-CaMKII, the newly protonated residues by H++ included D20, D56, E67, D93, D133, and E140. For holoCaM-Ng_{13-49}, the results depended on the conformations: most of the protonated residues by H++ included D22, E67, and D93. Because the acetyl groups (-COO^-) from the Ca^{2+} binding loops stabilize the Ca^{2+}-bound state, the numerical estimate of free energies was way off from the experimental values. Once the side chains of these residues in the Ca^{2+} binding loop were adequately protonated, the order of magnitude in Ca^{2+} binding free energy from computer simulations became comparable to the experimentally measured ones (shown in Table 1).

As a comparative study, we explored the robustness of our results by changing the pH from 7.4 to 6.8. According to H++, the protonation states of holoCaM remained the same at the range of pH $\in$ [6.9–7.8]. However, at pH $\in$ [6.5–6.8], E104 at Ca^{2+} binding loop III was additionally protonated. This changed the net charge of the loop III as well as the charge of the corresponding Ca^{2+} ion. We found that the Ca^{2+} binding free energy of loop III (ΔG) at pH = 6.8 increased by 3.15 kcal/mol comparing to that at pH = 7.4. ΔG for holoCaM-CaMKII remains the same at pH = 6.8 as the protonation states did not change. Therefore, $\Delta\Delta G$ for holoCaM-CaMKII still kept the same negative sign at pH = 6.8 as at pH = 7.4. For holoCaM-Ng_{13-49}, depending on different structures, the gain of additional protonated residues at pH = 6.8 varied: residues D58, D11 (from Ng) for the first structure, H107 for the second structure, E84 for the third structure, and D95, H107 for the fourth structure. We found that $\Delta\Delta G$ for holoCaM-Ng_{13-49} remained the same positive sign at pH = 6.8 as at pH = 7.4. Our results remain robust and can be further tested by experiments.

We provided the calculated $\Delta\Delta G$ values of Ca^{2+} binding site III for both holoCaM-CaMKII and holoCaM-Ng_{13-49} in Table S7. When the pH was lowered to 6.8, the $\Delta\Delta G$ values decreased for holoCaM-CaMKII (favoring retaining of the Ca^{2+}) and increased for holoCaM-Ng_{13-49}, (favoring release of Ca^{2+}) as illustrated in Fig. S17. The range of pH we explored here is typical for CaM to carry its specific function in a cell. We concluded that our finding is consistently robust with regard to how CaM encodes calcium incidents by binding to distinctive targets.

The charges of the Ca^{2+} ions depend on the environment

It has been shown that the charge on Ca^{2+} can be $< +2e$ due to charge transfer and polarization in the solution and especially in the Ca^{2+}-bound state (35–37). We computed the charges of calcium ions according to the protonation states of CaM by employing a semiempirical quantum chemistry program MOPAC (40). Similar to our previous work (39), we found that the charges on Ca^{2+} ions in holoCaM were approximately in the range of $+1.2e$ to $+1.6e$. However, we did not rescale the size of the calcium ions in this study, which can be critical to the development of proper effective polarizable force fields for Ca^{2+} according to Jungwirth et al. (36,38).

Despite those issues, our calculated free energy successfully matched the signs of the experimental values, whereas they differed by an order of magnitude for the holoCaM-Ng_{13-49} complexes. By using classical MD simulations, we were able to capture that Ng_{13-49} greatly reduces affinity of Ca^{2+} for cCaM whereas CaMKII peptide increases the affinity. More importantly, we provided a molecular mechanism of how the CaMBTs tune CaM's capacity of sensing Ca^{2+}.

The systematic error from the parameters used in steered MD simulations was minimized for the estimation of free energy changes

Computationally, free energy difference can be estimated from equilibrium ensemble using methods including umbrella sampling (44), free energy perturbation (91), and thermodynamic integration (92), as well as from nonequilibrium ensemble by using JE based on nonequilibrium work (51). We employed the nonequilibrium work method by carrying out many steered MD simulations for its simplicity and a particular reaction coordinate is not required. The drawback of this method is that the accuracy is highly dependent on

sufficient sampling of the small work (34). Although theoretically JE does not depend on the pathways of pulling when removing Ca^{2+} from its binding site, in practice, pulling along some relevant pathways leads to small work and small variance of the work distribution; while pulling along irrelevant pathways usually yields larger work and larger variance (51). In the latter case, significantly more trajectories are required for ample sampling. Therefore, in several studies, the pulling was only along one or more predetermined pathways repeatedly (52,93–95). However, for complex systems, it becomes even more difficult to find the relevant pathways. To solve this problem, we first searched the pathways leading to small work by pulling the Ca^{2+} from each binding site for ~100 times along randomly generated directions. Second, we increased the number of simulations by pulling along those specific pathways that generate smallest work until the differences in the free energy converges between the direct JE estimation and the CI extrapolation estimation. We minimized the systematic error from sampling by running the simulations until the extrapolation from CI as well as JE converges. The signs of $\Delta\Delta G$ for both systems we studied agreed with the experimental measured ones.

There are other computational studies on the mechanisms of Ca^{2+} releasing/binding in CaM (96,97) arguing that the deformation of protein structures attributes the discrepancies in the free energy calculation by allowing backbones to sample wide conformations in the pulling simulations (97). Even though the protein conformation can deform during pulling, we argue that this part of the work associated with protein deformation from the course of pulling Ca^{2+} should be part of the free energy calculation according to the concept of the JE: only the initial state and the final state determine the free energy difference.

CONCLUSIONS

CaM coordinates the activation of a family of Ca^{2+}-regulated proteins that are crucial for synaptic plasticity at the molecular and cellular level associated with learning and memory in neurons (3). Ng is a neural-specific postsynaptic IDP, for which Ca^{2+} decreases binding to CaM (12). Due to the low binding affinity of apoCaM-Ng_{13-49} interactions, experimentally it is challenging to determine the bound structures. Our coarse-grained molecular simulations of apoCaM-Ng_{13-49} binding guided by experimental measurements demonstrated the importance of conformational flexibility of both the binding partners in various states of the bound complex. Our study revealed that the acidic region of Ng_{13-49} (the DDPG residues right before the IQ motif; refer to Fig. 1 for the sequence of Ng_{13-49}) sticks out into the middle and "pries open" the EF-hand β-scaffold of cCaM. We speculate that this feature of destabilizing Ca^{2+}/CaM at low Ca^{2+} concentration could possibly explain CaM in apo-form in the resting cell. CaMKII on the other hand plays a pivotal role in learning and memory formation for both long-term potentiation and mechanisms for the modulation of synaptic efficacy (98). Our study demonstrated that in an apoCaM-Ng_{13-49} complex, CaM dominantly remains in an extended conformation. This is in contrast to a canonical bound complex where the two domains of CaM wrap around a rod-like CaMKII peptide that essentially stabilizes the Ca^{2+} binding loops of CaM. The embedded CaMKII peptide by two domains increases the binding affinity of Ca^{2+} to CaM. This proposed mechanism leads to a broader understanding of the reciprocal relation between Ca^{2+}/CaM binding and CaM/CaMBT binding (30) that involves major conformational changes in both partners. To the best of our knowledge, this is the first computational study providing a detailed description at atomistic level about how binding of CaM with two distinct targets (Ng and CaMKII) influences the release of Ca^{2+} from cCaM as a molecular underpinning of CaM-dependent Ca^{2+} signaling in neurons, which has been investigated for decades (27,30,99,100). Molecular basis for learning and memory formation has aroused attention of the neuroscience community dating back to 1980s (101). We believe this study allows opportunities to connecting the molecular regulations in atomistic detail to the understanding of cellular process cascade of learning and memory formation.

SUPPORTING MATERIAL

Supporting Material, seventeen figures, and fourteen tables are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(17)30116-9.

AUTHOR CONTRIBUTIONS

M.S.C. and P.Z. designed the research; P.Z., S.T., and H.T. performed the research; P.Z., S.T., and H.T. analyzed data; M.S.C., P.Z., and S.T. wrote the article.

ACKNOWLEDGMENTS

The authors thank Dr. Neal Waxham and Dr. Dirar Homouz for stimulating discussion and Dr. John Putkey for sharing the chemical shifts data of apoCaM. P.Z. thanks Dr. Edward P. O'Brien for sharing the table of KB dihedral angle potential. All authors acknowledge the Center for Advanced Computing and Data System and the Research Computing Center at the University of Houston for providing high performance computing resources and support.

This research was funded by the National Institute of Health (1R01GM097553), the National Science Foundation (ACI: 1531814), the Center for Advanced Computing and Data Systems at UH, and the Extreme Science and Engineering Discovery Environment (XSEDE, MCB070068). S.T. was supported by Cancer Prevention and Research Institute of Texas (RP140113). M.S.C. acknowledges support from National Science Foundation (PHY-1427654).

SUPPORTING CITATIONS

References (102–124) appear in the Supporting Material.

REFERENCES

1. Gifford, J. L., M. P. Walsh, and H. J. Vogel. 2007. Structures and metal-ion-binding properties of the Ca^{2+}-binding helix-loop-helix EF-hand motifs. *Biochem. J.* 405:199–221.
2. Yamniuk, A. P., and H. J. Vogel. 2004. Calmodulin's flexibility allows for promiscuity in its interactions with target proteins and peptides. *Mol. Biotechnol.* 27:33–57.
3. Xia, Z., and D. R. Storm. 2005. The role of calmodulin as a signal integrator for synaptic plasticity. *Nat. Rev. Neurosci.* 6:267–276.
4. Saucerman, J. J., and D. M. Bers. 2012. Calmodulin binding proteins provide domains of local Ca^{2+} signaling in cardiac myocytes. *J. Mol. Cell. Cardiol.* 52:312–316.
5. Yamauchi, T. 2005. Neuronal Ca^{2+}/calmodulin-dependent protein kinase II–discovery, progress in a quarter of a century, and perspective: implication for learning and memory. *Biol. Pharm. Bull.* 28:1342–1354.
6. Lisman, J., H. Schulman, and H. Cline. 2002. The molecular basis of CaMKII function in synaptic and behavioural memory. *Nat. Rev. Neurosci.* 3:175–190.
7. Colbran, R. J., and A. M. Brown. 2004. Calcium/calmodulin-dependent protein kinase II and synaptic plasticity. *Curr. Opin. Neurobiol.* 14:318–327.
8. Díez-Guerra, F. J. 2010. Neurogranin, a link between calcium/calmodulin and protein kinase C signaling in synaptic plasticity. *IUBMB Life.* 62:597–606.
9. Segal, M. 2005. Dendritic spines and long-term plasticity. *Nat. Rev. Neurosci.* 6:277–284.
10. Pak, J. H., F. L. Huang, …, K. P. Huang. 2000. Involvement of neurogranin in the modulation of calcium/calmodulin-dependent protein kinase II, synaptic plasticity, and spatial learning: a study with knockout mice. *Proc. Natl. Acad. Sci. USA.* 97:11232–11237.
11. Silva, A. J., R. Paylor, …, S. Tonegawa. 1992. Impaired spatial learning in alpha-calcium-calmodulin kinase II mutant mice. *Science.* 257:206–211.
12. Gerendasy, D. D., S. R. Herron, …, J. G. Sutcliffe. 1994. Mutational and biophysical studies suggest RC3/neurogranin regulates calmodulin availability. *J. Biol. Chem.* 269:22420–22426.
13. Kubota, Y., J. A. Putkey, …, M. N. Waxham. 2008. IQ-motif proteins influence intracellular free Ca^{2+} in hippocampal neurons through their interactions with calmodulin. *J. Neurophysiol.* 99:264–276.
14. Gaertner, T. R., J. A. Putkey, and M. N. Waxham. 2004. RC3/Neurogranin and Ca^{2+}/calmodulin-dependent protein kinase II produce opposing effects on the affinity of calmodulin for calcium. *J. Biol. Chem.* 279:39374–39382.
15. Kubota, Y., J. A. Putkey, and M. N. Waxham. 2007. Neurogranin controls the spatiotemporal pattern of postsynaptic Ca^{2+}/CaM signaling. *Biophys. J.* 93:3848–3859.
16. Stefan, M. I., S. J. Edelstein, and N. Le Novère. 2008. An allosteric model of calmodulin explains differential activation of PP2B and CaMKII. *Proc. Natl. Acad. Sci. USA.* 105:10768–10773.
17. Lai, M., D. Brun, …, N. Le Novère. 2015. Modulation of calmodulin lobes by different targets: an allosteric model with hemiconcerted conformational transitions. *PLoS Comput. Biol.* 11:e1004063.
18. Piazza, M., V. Taiakina, …, T. Dieckmann. 2014. Solution structure of calmodulin bound to the target peptide of endothelial nitric oxide synthase phosphorylated at Thr495. *Biochemistry.* 53:1241–1249.
19. Stigler, J., F. Ziegler, …, M. Rief. 2011. The complex folding network of single calmodulin molecules. *Science.* 334:512–516.
20. Tidow, H., and P. Nissen. 2013. Structural diversity of calmodulin binding to its target sites. *FEBS J.* 280:5551–5565.
21. Slaughter, B. D., R. J. Bieber-Urbauer, and C. K. Johnson. 2005. Single-molecule tracking of sub-millisecond domain motion in calmodulin. *J. Phys. Chem. B.* 109:12658–12662.
22. Zhang, M., T. Tanaka, and M. Ikura. 1995. Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nat. Struct. Biol.* 2:758–767.
23. Yang, C., G. S. Jas, and K. Kuczera. 2004. Structure, dynamics and interaction with kinase targets: computer simulations of calmodulin. *Biochim. Biophys. Acta.* 1697:289–300.
24. Meador, W. E., A. R. Means, and F. A. Quiocho. 1993. Modulation of calmodulin plasticity in molecular recognition on the basis of x-ray structures. *Science.* 262:1718–1721.
25. Wang, Q., P. Zhang, …, M. S. Cheung. 2013. Protein recognition and selection through conformational and mutually induced fit. *Proc. Natl. Acad. Sci. USA.* 110:20545–20550.
26. Tripathi, S., Q. Wang, …, M. S. Cheung. 2015. Conformational frustration in calmodulin-target recognition. *J. Mol. Recognit.* 28:74–86.
27. Waxham, M. N., A. L. Tsai, and J. A. Putkey. 1998. A mechanism for calmodulin (CaM) trapping by CaM-kinase II defined by a family of CaM-binding peptides. *J. Biol. Chem.* 273:17579–17584.
28. Ran, X., H. H. Miao, …, D. Yang. 2003. Structural and dynamic characterization of a neuron-specific protein kinase C substrate, neurogranin. *Biochemistry.* 42:5143–5150.
29. Kumar, V., V. P. Chichili, …, J. Sivaraman. 2013. Structural basis for the interaction of unstructured neuron specific substrates neuromodulin and neurogranin with Calmodulin. *Sci. Rep.* 3:1392.
30. Hoffman, L., A. Chandrasekar, …, M. N. Waxham. 2014. Neurogranin alters the structure and calcium binding properties of calmodulin. *J. Biol. Chem.* 289:14644–14655.
31. Karanicolas, J., and C. L. Brooks, 3rd. 2002. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* 11:2351–2361.
32. Samiotakis, A., D. Homouz, and M. S. Cheung. 2010. Multiscale investigation of chemical interference in proteins. *J. Chem. Phys.* 132:175101.
33. Hendrix, D. A., and C. Jarzynski. 2001. A "fast growth" method of computing free energy differences. *J. Chem. Phys.* 114:5974–5981.
34. Jarzynski, C. 1997. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* 78:2690–2693.
35. Dudev, T., and C. Lim. 2003. Principles governing Mg, Ca, and Zn binding and selectivity in proteins. *Chem. Rev.* 103:773–788.
36. Kohagen, M., P. E. Mason, and P. Jungwirth. 2014. Accurate description of calcium solvation in concentrated aqueous solutions. *J. Phys. Chem. B.* 118:7902–7909.
37. Mamatkulov, S., M. Fyta, and R. R. Netz. 2013. Force fields for divalent cations based on single-ion and ion-pair properties. *J. Chem. Phys.* 138:024505.
38. Kohagen, M., M. Lepšík, and P. Jungwirth. 2014. Calcium binding to calmodulin by molecular dynamics with effective polarization. *J. Phys. Chem. Lett.* 5:3964–3969.
39. Wang, Q., K. C. Liang, …, M. S. Cheung. 2011. The effect of macromolecular crowding, ionic strength and calcium binding on calmodulin dynamics. *PLoS Comput. Biol.* 7:e1002114.
40. Stewart, J. J. 2007. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* 13:1173–1213.
41. Anandakrishnan, R., B. Aguilar, and A. V. Onufriev. 2012. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* 40:W537–W541.
42. Berjanskii, M., and D. S. Wishart. 2006. NMR: prediction of protein flexibility. *Nat. Protoc.* 1:683–688.
43. Anthis, N. J., M. Doucleff, and G. M. Clore. 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.* 133:18966–18974.

44. Roux, B. 1995. The calculation of the potential of mean force using computer-simulations. *Comput. Phys. Commun.* 91:275–282.

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

46. Kumar, S., D. Bouzida, …, J. M. Rosenberg. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J. Comput. Chem.* 13:1011–1021.

47. Grabarek, Z. 2006. Structural basis for diversity of the EF-hand calcium-binding proteins. *J. Mol. Biol.* 359:509–525.

48. Dupuis, L., and N. Mousseau. 2012. Understanding the EF-hand closing pathway using non-biased interatomic potentials. *J. Chem. Phys.* 136:035101.

49. Wang, X., L. W. Xiong, …, J. A. Putkey. 2013. The calmodulin regulator protein, PEP-19, sensitizes ATP-induced Ca^{2+} release. *J. Biol. Chem.* 288:2040–2048.

50. Homouz, D., M. Perham, …, P. Wittung-Stafshede. 2008. Crowded, cell-like environment induces shape changes in aspherical protein. *Proc. Natl. Acad. Sci. USA.* 105:11754–11759.

51. Ytreberg, F. M., and D. M. Zuckerman. 2004. Efficient use of nonequilibrium measurement to estimate free energy differences for molecular systems. *J. Comput. Chem.* 25:1749–1759.

52. Zhang, D., J. Gullingsrud, and J. A. McCammon. 2006. Potentials of mean force for acetylcholine unbinding from the alpha7 nicotinic acetylcholine receptor ligand-binding domain. *J. Am. Chem. Soc.* 128:3019–3026.

53. Tang, C., J. Iwahara, and G. M. Clore. 2006. Visualization of transient encounter complexes in protein-protein association. *Nature.* 444:383–386.

54. Sugase, K., H. J. Dyson, and P. E. Wright. 2007. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature.* 447:1021–1025.

55. Turjanski, A. G., J. S. Gutkind, …, G. Hummer. 2008. Binding-induced folding of a natively unstructured transcription factor. *PLoS Comput. Biol.* 4:e1000060.

56. Higo, J., Y. Nishimura, and H. Nakamura. 2011. A free-energy landscape for coupled folding and binding of an intrinsically disordered protein in explicit solvent from detailed all-atom computations. *J. Am. Chem. Soc.* 133:10448–10458.

57. Ganguly, D., W. Zhang, and J. Chen. 2012. Synergistic folding of two intrinsically disordered proteins: searching for conformational selection. *Mol. Biosyst.* 8:198–209.

58. Levy, Y., P. G. Wolynes, and J. N. Onuchic. 2004. Protein topology determines binding mechanism. *Proc. Natl. Acad. Sci. USA.* 101:511–516.

59. Alsallaq, R., and H. X. Zhou. 2007. Energy landscape and transition state of protein-protein association. *Biophys. J.* 92:1486–1502.

60. Kim, Y. C., C. Tang, …, G. Hummer. 2008. Replica exchange simulations of transient encounter complexes in protein-protein association. *Proc. Natl. Acad. Sci. USA.* 105:12855–12860.

61. Shoemaker, B. A., J. J. Portman, and P. G. Wolynes. 2000. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. USA.* 97:8868–8873.

62. Clementi, C. 2008. Coarse-grained models of protein folding: toy models or predictive tools? *Curr. Opin. Struct. Biol.* 18:10–15.

63. Tompa, P., and M. Fuxreiter. 2008. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* 33:2–8.

64. Song, J., S. C. Ng, …, H. S. Chan. 2013. Polycation-π interactions are a driving force for molecular recognition by an intrinsically disordered oncoprotein family. *PLoS Comput. Biol.* 9:e1003239.

65. Oldfield, C. J., Y. Cheng, …, A. K. Dunker. 2005. Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry.* 44:12454–12470.

66. Dunlap, T. B., J. M. Kirk, …, T. P. Creamer. 2013. Thermodynamics of binding by calmodulin correlates with target peptide α-helical propensity. *Proteins.* 81:607–612.

67. Rogers, J. M., C. T. Wong, and J. Clarke. 2014. Coupled folding and binding of the disordered protein PUMA does not require particular residual structure. *J. Am. Chem. Soc.* 136:5197–5200.

68. Schumacher, M. A., A. F. Rivard, …, J. P. Adelman. 2001. Structure of the gating domain of a Ca^{2+}-activated K^+ channel complexed with Ca^{2+}/calmodulin. *Nature.* 410:1120–1124.

69. Olwin, B. B., A. M. Edelman, …, D. R. Storm. 1984. Quantitation of energy coupling between Ca^{2+}, calmodulin, skeletal muscle myosin light chain kinase, and kinase substrates. *J. Biol. Chem.* 259:10949–10955.

70. Peersen, O. B., T. S. Madsen, and J. J. Falke. 1997. Intermolecular tuning of calmodulin by target peptides and proteins: differential effects on Ca^{2+} binding and implications for kinase activation. *Protein Sci.* 6:794–807.

71. O'Donnell, S. E., L. Yu, …, M. A. Shea. 2011. Recognition of β-calcineurin by the domains of calmodulin: thermodynamic and structural evidence for distinct roles. *Proteins.* 79:765–786.

72. Theoharis, N. T., B. R. Sorensen, …, M. A. Shea. 2008. The neuronal voltage-dependent sodium channel type II IQ motif lowers the calcium affinity of the C-domain of calmodulin. *Biochemistry.* 47:112–123.

73. Halling, D. B., D. K. Georgiou, …, S. L. Hamilton. 2009. Determinants in Ca_V1 channels that regulate the Ca^{2+} sensitivity of bound calmodulin. *J. Biol. Chem.* 284:20041–20051.

74. Qin, S., X. Pang, and H.-X. Zhou. 2011. Automated prediction of protein association rate constants. *Structure.* 19:1744–1751.

75. Zheng, W., N. P. Schafer, …, P. G. Wolynes. 2012. Predictive energy landscapes for protein-protein association. *Proc. Natl. Acad. Sci. USA.* 109:19244–19249.

76. Khait, R., and G. Schreiber. 2012. FRETex: a FRET-based, high-throughput technique to analyze protein-protein interactions. *Protein Eng. Des. Sel.* 25:681–687.

77. Rogers, J. M., A. Steward, and J. Clarke. 2013. Folding and binding of an intrinsically disordered protein: fast, but not 'diffusion-limited'. *J. Am. Chem. Soc.* 135:1415–1422.

78. Schreiber, G., G. Haran, and H. X. Zhou. 2009. Fundamental aspects of protein-protein association kinetics. *Chem. Rev.* 109:839–860.

79. Uversky, V. N. 2002. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 11:739–756.

80. Wright, P. E., and H. J. Dyson. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293:321–331.

81. Papoian, G. A., and P. G. Wolynes. 2003. The physics and bioinformatics of binding and folding-an energy landscape perspective. *Biopolymers.* 68:333–349.

82. Sickmeier, M., J. A. Hamilton, …, A. K. Dunker. 2007. DisProt: the database of disordered proteins. *Nucleic Acids Res.* 35:D786–D793.

83. Chu, X., L. Gan, …, J. Wang. 2013. Quantifying the topography of the intrinsic energy landscape of flexible biomolecular recognition. *Proc. Natl. Acad. Sci. USA.* 110:E2342–E2351.

84. Chen, J. 2009. Intrinsically disordered p53 extreme C-terminus binds to S100Bββ through "fly-casting". *J. Am. Chem. Soc.* 131:2088–2089.

85. Ganguly, D., and J. Chen. 2009. Atomistic details of the disordered states of KID and pKID. Implications in coupled binding and folding. *J. Am. Chem. Soc.* 131:5214–5223.

86. Zhou, H. X., X. Pang, and C. Lu. 2012. Rate constants and mechanisms of intrinsically disordered proteins binding to structured targets. *Phys. Chem. Chem. Phys.* 14:10466–10476.

87. Tsai, C. J., S. Kumar, …, R. Nussinov. 1999. Folding funnels, binding funnels, and protein function. *Protein Sci.* 8:1181–1190.

88. Boehr, D. D., R. Nussinov, and P. E. Wright. 2009. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* 5:789–796.

89. Shukla, D., A. Peck, and V. S. Pande. 2016. Conformational heterogeneity of the calmodulin binding interface. *Nat. Commun.* 7:10910.

90. Pall, S., M. J. Abraham, …, E. Lindahl. 2015. Tackling exascale software challenges in molecular dynamics simulations with GROMACS. *Lect. Notes Comput. Sci.* 8759:3–27.

91. Zwanzig, R. W. 1954. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* 22:1420–1426.

92. Kirkwood, J. G. 1935. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* 3:300–313.

93. Park, S., and K. Schulten. 2004. Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.* 120:5946–5961.

94. Park, S., F. Khalili-Araghi, …, K. Schulten. 2003. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J. Chem. Phys.* 119:3559–3566.

95. Chen, P. C., and S. Kuyucak. 2009. Mechanism and energetics of charybdotoxin unbinding from a potassium channel from molecular dynamics simulations. *Biophys. J.* 96:2577–2588.

96. Zhang, Y., H. Tan, …, G. Chen. 2008. Ca^{2+} dissociation from the C-terminal EF-hand pair in calmodulin: a steered molecular dynamics study. *FEBS Lett.* 582:1355–1361.

97. Lepsík, M., and M. J. Field. 2007. Binding of calcium and other metal ions to the EF-hand loops of calmodulin studied by quantum chemical calculations and molecular dynamics simulations. *J. Phys. Chem. B.* 111:10012–10022.

98. Miller, S. G., and M. B. Kennedy. 1986. Regulation of brain type II Ca^{2+}/calmodulin-dependent protein kinase by autophosphorylation: a Ca^{2+}-triggered molecular switch. *Cell.* 44:861–870.

99. Putkey, J. A., and M. N. Waxham. 1996. A peptide model for calmodulin trapping by calcium/calmodulin-dependent protein kinase II. *J. Biol. Chem.* 271:29619–29623.

100. Putkey, J. A., M. N. Waxham, …, Q. K. Kleerekoper. 2008. Acidic/IQ motif regulator of calmodulin. *J. Biol. Chem.* 283:1401–1410.

101. Crick, F. 1984. Memory and molecular turnover. *Nature*. 312:101.

102. Jurado, L. A., P. S. Chockalingam, and H. W. Jarrett. 1999. Apocalmodulin. *Physiol. Rev.* 79:661–682.

103. Waterhouse, A. M., J. B. Procter, …, G. J. Barton. 2009. Jalview version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 25:1189–1191.

104. Bähler, M., and A. Rhoads. 2002. Calmodulin signaling via the IQ motif. *FEBS Lett.* 513:107–113.

105. Sugita, Y., and Y. Okamoto. 1999. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314:141–151.

106. Sanbonmatsu, K. Y., and A. E. García. 2002. Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins.* 46:225–234.

107. Cheung, M. S., J. M. Finke, …, J. N. Onuchic. 2003. Exploring the interplay between topology and secondary structural formation in the protein folding problem. *J. Phys. Chem. B.* 107:11193–11200.

108. Yang, Y., E. Faraggi, …, Y. Zhou. 2011. Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. *Bioinformatics.* 27:2076–2082.

109. Debye, P., and E. Hückel. 1923. The theory of electrolytes. I. Lowering of freezing point and related phenomena. *Phys. Z.* 24:185–206.

110. Betancourt, M. R., and D. Thirumalai. 1999. Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Sci.* 8:361–369.

111. Andreasen, T. J., C. W. Luetje, …, D. R. Storm. 1983. Purification of a novel calmodulin binding protein from bovine cerebral cortex membranes. *Biochemistry.* 22:4615–4618.

112. Dima, R. I., and D. Thirumalai. 2004. Asymmetry in the shapes of folded and denatured states of proteins. *J. Phys. Chem. B.* 108:6564–6570.

113. Camacho, C. J., and D. Thirumalai. 1993. Kinetics and thermodynamics of folding in model proteins. *Proc. Natl. Acad. Sci. USA.* 90:6369–6372.

114. Lindorff-Larsen, K., S. Piana, …, D. E. Shaw. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins.* 78:1950–1958.

115. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

116. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald - an N.Log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

117. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single-crystals - a new molecular-dynamics method. *J. Appl. Phys.* 52:7182–7190.

118. Hess, B., H. Bekker, …, J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.

119. Lipari, G., and A. Szabo. 1982. Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* 104:4546–4559.

120. Kudlay, A., M. S. Cheung, and D. Thirumalai. 2009. Crowding effects on the structural transitions in a flexible helical homopolymer. *Phys. Rev. Lett.* 102:118101.

121. Weinkam, P., E. V. Pletneva, …, P. G. Wolynes. 2009. Electrostatic effects on funneled landscapes and structural diversity in denatured protein ensembles. *Proc. Natl. Acad. Sci. USA.* 106:1796–1801.

122. Carpenter, G. A., and S. Grossberg. 1987. ART 2: self-organization of stable category recognition codes for analog input patterns. *Appl. Opt.* 26:4919–4930.

123. Cheung, M. S., and D. Thirumalai. 2007. Effects of crowding and confinement on the structures of the transition state ensemble in proteins. *J. Phys. Chem. B.* 111:8250–8257.

124. Guo, Z., and D. Thirumalai. 1997. The nucleation-collapse mechanism in protein folding: evidence for the non-uniqueness of the folding nucleus. *Fold. Des.* 2:377–391.

Biophysical Journal, Volume *112*

Supplemental Information

Opposing Intermolecular Tuning of Ca^{2+} Affinity for Calmodulin by Neurogranin and CaMKII Peptides

Pengzhi Zhang, Swarnendu Tripathi, Hoa Trinh, and Margaret S. Cheung

Supporting Text

I. Sample preparation

I.1 Intact Ng

Ng is a 78-residue intrinsically disordered protein with IQ motif (1, 2) composed of hydrophobic and basic amino acids. Available experimental measurement for residual structure of Ng is for the full-length protein Ng from mouse (3), therefore, the sequence is used in the modeling and parameterization of Ng, as shown below,

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

Following the work by Hoffman et al. (4), we divided the Ng protein into several units, including the acidic N terminal (residues 13-25), the IQ motif (residues 26-49) and the poly-Glycine C terminal (residues 50-78). The underscored sequence stands for the IQ motif which partially retains residual structure (residues 25-42) in the unbound state. To reproduce the residual structure, replica exchange method (REM) (5, 6) was employed for calculation of average helicity of the residual structure and the backbone model-free order parameter S^2.

I.2 $Ng_{13\text{-}49}$ peptide

Hoffman et al. showed from calorimetry and NMR experiments that the IQ motif alone does not reproduce similar Ng-mediated affinity of Ca^{2+} for Ca^{2+}-free CaM (apoCaM), or the pattern of intermolecular interaction between Ng and apoCaM (4). A combination of the acidic region in the N-terminal and the IQ motif yields the minimum composition of Ng for its function.

Therefore the Ng_{13-49} peptide is used to study the interaction with apoCaM (PDB ID: 1CFD) that allows the comparison of the results from computer simulations with several experimental measurements, including dissociation constant of apoCaM and Ng_{13-49}, the affinity of Ca^{2+} to CaM, and the changes in the chemical shifts of apoCaM upon Ng_{13-49} binding (4). The sequence of the Ng_{13-49} peptide is provided below,

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

II. Coarse-grained protein or peptide models

II. 1 Hamiltonian and parametrization of the coarse-grained protein models

As described above, Ng is partly ordered in solution. The fragment G25-A42 has a fraction of 22% and 28% residual structure calculated from the C_α and the H_α chemical shifts in the nuclear magnetic resonance (NMR) experiment (3), respectively. The rest part of the peptide remains unstructured. In order to compare with results from several experiments (3, 4), peptide Ng_{13-49} comprising the IQ domain with adjacent acidic region as well as the full sequence Ng were used in the simulations (please find the sequences in session I from the *Supporting Information*.).

We adopted a side-chain-C_α coarse-grained model to represent protein/peptides developed by Cheung et al (7). In this model, each amino acid (except glycine, which is represented by a C_α bead) is represented by two beads: the C_α bead is located at the C_α position to represent the backbone atoms; the side-chain bead is located at the center of mass of the side-chain atoms to represent the side-chain atoms. The use of side-chain-C_α model enables us to capture large structural fluctuations of the proteins at a large timescale that warrants a wide search of the phase space.

The total potential energy E of Ng or $Ng_{13\text{-}49}$ is given by,

$$E^{Ng} = E_{bond} + E_{angle} + E_{dihedral} + E_{chirality} + E_{elec} + E_{HB} + E_{LJ} \qquad \text{Eq. (S1)}$$

The bonded interactions include the bond energy (E_{bond}), the bond-angle energy (E_{angle}), and the dihedral-angle energy ($E_{dihedral}$) the L-isomer restraint of chirality ($E_{chirality}$). E_{bond}, E_{angle}, and $E_{chirality}$ constrain the bond length, bond angle and side-chain chirality through harmonic potentials. The formula can be found in our previous work (8). The equilibrium bond length, bond angle and side-chain chirality parameters were taken from the crystal structure of apoCaM and IQ motif of Ng (PDB ID: 405E), and those of the missing segment were obtained from the structures predicted by the Sparks-X protein structure prediction server (9). Since these terms are mainly constrained by chemical rules, they do not vary significantly by the conformations.

The electrostatic energy (E_{elec}) between each two beads (C_α or side-chain bead) i and j with partial charges is described by Debye-Hückel potential (10) to include the screening effect of the electrolyte solution.

$$E_{elec}^{ij} = \frac{q_i q_j}{4\pi\epsilon_r\epsilon_0 r_{ij}} \exp\left(-r_{ij} \Big/ \sqrt{\frac{\epsilon_r\epsilon_0 k_B T}{2e^2 I}}\right) \qquad \text{Eq. (S2)}$$

where q_i (q_j) is the partial charges on bead i (j); e is the elementary charge (see the method of generating the partial charges in the next section II.2 from the *Supporting Information*); ϵ_r is the relative dielectric constant and is set to 80 for aqueous solution; ϵ_0 is the permittivity of the vacuum; r_{ij} is the distance between beads i and j; k_B is Boltzmann constant; T = 1.1 ε / k_B is the temperature; I = 0.15 M is the ionic strength.

For the backbone hydrogen bonding interactions, we adopted an angular-dependent function that captures directional properties of backbone hydrogen bonds (7),

$$E_{HB}^{ij} = \sum_{|j-i|>3} A(\rho)\varepsilon\left[\left(\frac{\rho_{ij}^0}{r_{ij}}\right)^6 - 2\left(\frac{\rho_{ij}^0}{r_{ij}}\right)^{12}\right] \quad \text{Eq. (S3)}$$

$$A(\rho) = \frac{1}{\left[1+(1-\cos^2\rho)(1-\frac{\cos\rho}{\cos\rho_\alpha})\right]^2} \quad \text{Eq. (S4)}$$

where the strength ε = 0.6 kcal/mol; r_{ij} is the distance between backbone beads *i* and *j*; and ρ_{ij}^0 = 4.6 Å is the typical length of a hydrogen bond; $A(\rho)$ measures the structural alignment of two interacting strands; ρ is the pseudo dihedral angle between backbone beads of the two interacting strands (7); ρ_α is the pseudo dihedral angle of a canonical helical turn, 0.466 rad. $A(\rho)$ preserves the tendency to form β-strands (ρ = 0 or π) or α-helices ($\rho = \rho_\alpha$) as it is maximized to 1 in either of the two cases.

For the non-bonded interactions (E_{LJ}), between C_α bead *i* from the backbone (bb) and side-chain (sc) bead *j*, a pure repulsive interaction was considered,

$$E_{LJ}^{sc-bb} = \varepsilon\left(\frac{\rho_{ij}^0}{r_{ij}}\right)^{12} \quad \text{Eq. (S5)}$$

where the strength ε = 0.6 kcal/mol; i $\rho_{ij}^0 = 0.9(\rho_i^0 + \rho_j^0)$, ρ_i^0 is the size of backbone bead *i*, 0.5 σ (σ = 3.8 Å) and ρ_j^0 is the size of side-chain bead *j*, which is the van der Waals radius of the side-chain. 0.9 is a scaling factor to remove clashes between bulky side-chains. For the non-bonded interactions between side-chain bead *i* and side-chain bead *j*, we applied the Lennard-Jones potential,

$$E_{LJ}^{sc-sc} = \varepsilon_{ij}\left[\left(\frac{\rho_{ij}^0}{r_{ij}}\right)^{12} - 2\left(\frac{\rho_{ij}^0}{r_{ij}}\right)^6\right] \quad \text{Eq. (S6)}$$

where ε_{ij} is solvent mediated interactions between the involved amino acids obtained from Betancourt-Thirumalai's study (11). $\rho_{ij}^0 = 0.9(\rho_i^0 + \rho_j^0)$, $\rho_{i(j)}^0$ is the van der Waals radius of side-chain bead *i* (*j*). r_{ij} is the distance between beads *i* and *j* in Eqs. S3, S5 and S6. Since the electrostatic interactions were explicitly included in the Hamiltonian, the Lennard-Jones potential coefficient ε_{ij} was rescaled as described in our previous work (8).

To best describe the intrinsic disorder nature of Ng or $Ng_{13\text{-}49}$, we employed a backbone dihedral angle potential that is sequence-specific but independent on protein topology. For this we used the model introduced by Karanicolas and Brooks (12), or the KB model. The dihedral angle between four adjacent α-carbons depends on the backbone dihedral angles of the two middle residues. Brooks' group produced a probability distribution for 400 possible ordered pairs of amino acid residues from a survey of the Protein Data Bank and thus related the probability distribution to potential energy ignoring the entropy contribution. The dihedral angle potential presents two minima corresponding to local α-helical and β-strand geometries. The statistical potential is modeled as a 4-term cosine series,

$$E_{dihedral}^{ijkl} = \varepsilon_{KB} \sum_{n=1}^{4} K_{ij,n} \left[1 + \cos(n\varphi_{ijkl} - \delta_{jk,n}) \right] \quad \text{(S7)}$$

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

We tested a total of three types of dihedral potentials for a comparative study. The first dihedral angle potential is a sequence-based model as described above. The second dihedral angle potential is a structure-based potential, or SB model, where the structure of a specific

segment of Ng obtained from the crystal structure (PDB code: 4E50 for Ng) was used as the reference (please see Table S4 for the residues in this segment). The SB model is composed of two-term cosine-series,

$$E_{dihedral}^{ijkl} = \varepsilon_{SB} \sum_{n=1,3} k_{\varphi}^{n} \left[1 - \cos\left(n \times \left(\varphi_{ijkl} - \varphi_{ijkl}^{0}\right)\right)\right] \quad \text{Eq. (S8)}$$

where *i, j, k* and *l* are four consecutive C_α beads. φ_{ijkl} is the dihedral angle formed by those four beads. The equilibration position of the dihedral angle φ_{ijkl}^{0} for this specific segment G25-A42 was taken from the crystal structure (PDB code: 4E50 for Ng). $k_{\varphi}^{1} = 2k_{\varphi}^{3} = 2\varepsilon$. ε_{SB} is used to adjust the barrier of the dihedral angle potential. For the unstructured segment $\varepsilon_{SB} = 0$.

The third dihedral potential is the hybrid of the two models, or Hybrid model, as shown in Table S4. We replaced the dihedral angle potential of the unstructured segment (beyond G25-A42) from the SB model with a KB model.

Different barrier values of the structure-based dihedral potential ε_{SB} and the statistical dihedral potential ε_{KB} were tested. Fig. S7 shows the helicity of the fragment G25-A42 decreases with temperature in all the cases. At the temperature $T = 1.1\ \varepsilon/k_B$, for the SB model, the helicity fell within the experimental range when $\varepsilon_{SB} = 0.3$ (Fig. S7A); for the Hybrid model, the helical fraction fell within the experimental range with when $\varepsilon_{SB} = 0.3$, and ε_{KB} did not make any influence in the range from 1.8 to 2.2 (Fig. S7B); for the KB model, the helical fraction was best matched when ε_{KB} is in the range from 1.5 to 2.2 (Fig. S7C).

To determine which best represents the properties of Ng among all the three models, we further investigated them at the residue level. Therefore, for those models with parameters that matched the overall helicity (Fig. S7), we computed the nuclear magnetic resonance (NMR)

model free order parameter S^2 for the backbone beads from our coarse-grained simulations (please find the details of calculation in V.1 from the *Supporting Information*), and compared with data from the NMR experiments (3). The computed S^2 (Table S5) positively correlated with the experimental values in all cases. Interestingly, for the KB model, the correlation was overall higher than the SB and Hybrid model, especially for ε_{KB} = 1.8. Because Ng protein is intrinsically disordered (13), The KB dihedral angle potential of no bias to any specific structure enables sampling of a broad spectrum of Ng conformations, whereas the SB model and the Hybrid model of full or partial bias to a specific structure limit the flexibility to explore more conformations. Therefore, for modeling intrinsically disordered protein Ng, we adopted the sequence-based KB dihedral potential model. ε_{KB} = 1.8 was used in the following study of apoCaM-$Ng_{13\text{-}49}$ binding simulations.

II.2 Determination of partial charges for the coarse-grained models

II.2.a Neurogranin protein

We adopted a multi-scale method developed by Cheung group (14) to assign partial charges to the intact Ng protein. Firstly, we ran REMD simulations without electro-static interactions (please find the details in III.1 in the *Supporting Information*) and obtained the free energy surface $F(\Delta, \chi)$ as a function of asphericity Δ and overlap function χ. Δ measures the shape of the protein: it is like a rod or a sphere when Δ = 1 or 0, respectively (15). χ measures the similarity to the reference structure (16). Using the free energy, we thus selected about 400-600 frames from the REMD simulations of Ng protein through importance sampling (17). We reconstructed the coarse-grained structures into atomistic protein models (17), used H++ server (18) to predict the protonation states and computed the partial charges using the semi-empirical

quantum chemistry program MOPAC (19). We collected the partial charges on the backbone (side-chain) atoms as the charge for the C_α (sidechain) bead. We then obtained the averaged partial charges on the C_α and side-chain beads over all the structures. We repeated the same process for all the three models shown in Table S4 and the three sets of charges we obtained were highly similar to each other (the correlation coefficients were ~1.00). We therefore used only one set of charges for all the three models. The charges on Ng protein are provided in Table S9.

II.2.b Apo-calmodulin and neurogranin peptide system

We followed the same procedures as for the Ng protein. The free energy surface $F(d_{COM}, Z_{inter})$ was obtained from US simulations of apoCaM and Ng_{13-49} without electrostatic interactions by using WHAM (20, 21) (please see details about the US simulations in III.2 in the *Supporting Information*). d_{COM} is the distance between center of mass of apoCaM and center of mass of Ng_{13-49}, and Z_{inter} is the total number of intermolecular contacts (please see the definition of a contact in V.3 from the *Supporting Information*). Two groups of partial charges were generated for apoCaM and Ng_{13-49} according to the experimental conditions: the fluorescence experiments for measurement of the dissociation constant of apoCaM and Ng_{13-49} were conducted at pH = 7.2 and at ionic strength I = 0.15 M; the NMR experiments for determining the change in chemical shifts of apoCaM upon binding Ng_{13-49} were conducted at pH = 6.3 and at ionic strength I = 0.10 M. The protonation states of histidine residues were determined by the H++ server (18) before using MOPAC (19). The calculated partial charges of apoCaM and Ng_{13-49} in the above two conditions were provided in Table S10-S13.

III. Coarse-grained Molecular Simulations

III.1 Replica Exchange Molecular Dynamics simulations

To study dynamics of unbound Ng protein, which has a rugged energy landscape because of its intrinsic disorder, we applied replica exchange molecular dynamics (REMD) (5, 6) to enhance the sampling. For the three dihedral angle models, a range of barriers of dihedral angel potential were investigated: for the SB (structure-based) model, ε_{SB} = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6 and 0.7; for the Hybrid model, ε_{SB} = 0.3, 0.4 for the part with residual structure (G25-A42), in combination with ε_{KB} = 1.8, 2.0 and 2.2 for the rest; for the KB model, ε_{KB} = 1.0, 1.5, 1.8, 2.0, 2.2, 2.5 and 3.0. For each of the models, 24 replicas were distributed at temperatures T = 0.80, 0.82, 0.83, 0.85, 0.87, 0.88, 0.90, 0.93, 0.97, 1.00, 1.03, 1.07, 1.10, 1.13, 1.17, 1.20, 1.23, 1.27, 1.30, 1.33, 1.40, 1.47, 1.57, 1.67 ε / k_B to produce ample exchanges between the replicas (ε is the reduced energy unit = 0.6 kcal/mol, and k_B is the Boltzmann constant). The average acceptance ratio of exchanging among the neighboring replicas ranges from 0.57 to 0.77. An exchange between neighboring replicas was attempted every 100,000 time steps. In order to explore more conformation space, in each REMD simulation, a total of 6 sets of distinct initial configurations were used, making up a total of 1200 exchange attempts for each model.

III.2 Umbrella sampling simulations

Umbrella sampling (US) method (22) was used to explore the thermodynamic properties of the apoCaM-Ng system. The distance between the center of mass of apoCaM and the center of mass of Ng_{13-49} d_{COM} was used as the reaction coordinate. d_{COM} was restrained by a harmonic force with spring constant 66.7 ε/σ^2 (σ = 3.8 Å is reduced length unit). The equilibrium positions of the harmonic force range from 0.2 σ to 20.0 σ with a bin size of 0.2 σ, making up a total of 100

windows. For each window, 10,000,000 time-steps of constrained molecular dynamic simulation were carried out.

To generate the initial structures at each window of d_{COM} for the US simulations, molecular dynamics simulations were carried out for the apoCaM-Ng_{13-49} complex from the bound form (PDB code: 4E50, in the coarse-grained model) at a high temperature $T = 1.33\ \varepsilon / k_B$ for dissociation (ε = 0.6 kcal/mol is reduced energy unit, and k_B is Boltzmann constant). A total of 5 sets of initial configurations for the following umbrella sampling simulations were generated from 5 dissociation trajectories.

US simulations were then performed without electrostatic interactions for generating ensemble of structures to generate ensemble of structures for partial charge determination (see 2.3 for the details). After we obtained the partial charges for apoCaM and Ng_{13-49}, US simulations were carried out with charge-charge interactions to study how the strength of non-electrostatic intermolecular interactions influences the binding affinity and to determine the optimal strength by comparing with the experiments (4). Using this strength of non-electrostatic intermolecular interactions, US simulations were carried up at appropriate experimental conditions to study the binding thermodynamic properties of apoCaM-Ng_{13-49} system. Each set of US simulations were performed from 5 different initial structures at the temperature $T = 1.1\ \varepsilon/k_B$. We analyzed the data using Weighted Histogram Analysis Method (WHAM) (20, 21).

IV. Calculation of difference in binding free energy using Jarzynski's equality from steered molecular dynamics simulations

IV.1 Selection of initial atomistic structures

holoCaM-Ng$_{13-49}$ complex: We employed coarse-grained molecular simulations to efficiently sample a broad ensemble of complex structures. We used the experimental measurements as a guide to strategically select several structures from the major cluster of coarse-grained complex structures. According to the crystal structure of apoCaM and IQ motif of Ng (NgIQ) (PDB ID: 4E50), the two EF-hand motifs from cCaM are open (EF-hand angles ≈ 101°) and the two EF-hand motifs from nCaM are closed (EF-hand angles ≈ 129°) (Table S3). The EF-hand angles were computed as the angle between the vectors that define the orientation of the two helices in the EF-hand motif; the vectors were defined from the center of mass of the C_α beads of the first 4 residues to the center of mass of the C_α beads of the last 4 residues in a helix of the EF-hand motif. Moreover, NgIQ binds to cCaM in the crystal structure and the NMR experiments (4) showed that Ng_{13-49} mainly interacts with cCaM. Therefore, we select the 4 structures from the simulations by following the criteria: (i) EF-hand motifs from cCaM must be open (EF-hand angles range from 85° to 105°) and EF-hand motifs from nCaM must be closed (EF-hand angles are greater than 125°); (ii) cCaM has more interactions with Ng_{13-49} than nCaM ($Z_c - Z_n > 0$). Using SCAAL method (17), we reconstructed all-atomistic structures of apoCaM-Ng_{13-49} from the selected sidechain-C_α configurations, as shown in Fig. S8. Four Ca^{2+} ions were added and their positions were estimated as the center of mass of sidechain beads of the corresponding Ca^{2+} co-ordination residues. The Ca^{2+} positions were further optimized by energy minimization on the all-atomistic structures of Ca^{2+}-CaM-Ng_{13-49} (holoCaM-Ng_{13-49}) using a gradient descent algorithm provided by the molecular dynamics software package GROMACS (version 5.0.6) (23) with AMBER99SB-ILDN force-field (24). The tolerance for the maximum force was set to 500 kJ / mol / nm to remove clashes between atoms.

holoCaM-CaMKII complex: We used the crystal structure of calcium-bound CaM-CaMKII complex (PDB ID: 1CDM) for Ca^{2+} binding free energy calculations.

holoCaM: We used the crystal structure of holoCaM (PDB ID: 1CLL) for assessing the Ca^{2+} binding free energy in the absence of the CaMBT as a reference.

IV.2 Protonation of the initial structures for pulling simulations

The charge distribution impacts accurate estimation of the binding free energy of the Ca^{2+} ions. We first took into consideration of the pH and ionic strength by using H++ server (18) to predict the protonation states of the all-atomistic structure. We input the following parameters to H++ as used in the experiment: pH = 7.4, ionic strength I = 0.15 M, the external dielectric constant is 78.4 and the internal dielectric constant is 4. After we obtained the protonated states of all the residues, we performed an energy minimization using steepest descent algorithm and the tolerance for the maximum force was set to 500 kJ / mol / nm. Then we assigned partial charges according to the geometry of the proteins (input is the protonated structure in PDB format) by using a semi-empirical quantum chemistry program MOPAC (19). We applied those protonated protein structures and partial charges for further all-atomistic calculations.

IV.3 Steered molecular dynamics simulations

Classical molecular dynamics simulations were carried out using GROMACS molecular dynamics package (version 5.0.6) (23). We used the AMBER99SB-ILDN force field (24) except the charge assignment. The partial charges were generated from MOPAC as explained in the

previous session. The rigid three-site TIP3P model (25) was used for water molecules. We neutralized the system with Na^+ and Cl^- ions and maintained an ionic strength of 150 mM.

The size of the box is about 10 nm x 10 nm x 10 nm. The proteins are placed at least 1 nm away from the edges of the cubic box. Periodic boundary conditions were employed to mimic the macroscopic settings for electrolytes. Electrostatic interactions between periodic images were treated using the particle mesh Ewald approach (26), with a grid size of 0.16 nm, fourth-order cubic interpolation and a tolerance of 10^{-5}. A cutoff of 10 Å was used for van der Waals interactions and real space Coulomb interactions as well as for updating neighbor lists.

We adopted a gradient descent algorithm for energy minimization. Then we gradually heated the system temperature by 50 K per 0.1 ns to 298.15 K using NVT ensemble. We carried out 1 ns NPT equilibration with heavy atoms of the proteins (including the four Ca^{2+} ions) fixed. The proteins as well as the Ca^{2+} ions were afterwards released and were further equilibrated for another 1 ns. All NPT simulations maintained a constant pressure of 1 bar and temperature of 298.15 K using the Parrinello-Rahman barostat (27). The bond lengths in proteins were constrained using the LINCS algorithm of Hess (28). The equation of motion was integrated using a 2-fs time steps. As discussed in the main text, the positions of the Ca^{2+} ions in the bound state change during the minimization and equilibration stage, which yields inaccurate estimation of the binding free energy, therefore, we froze the positions of the Ca^{2+} ions as well as the backbone heavy atoms of the protein (or protein complex) during these preparation stage. They were free to move afterwards in the pulling simulations.

For each of the initial structures including 1 structure for each of holoCaM and holoCaM-CaMKII and four selected structure of holoCaM-Ng_{13-49}, I pulled the Ca^{2+} from site III and IV independently to 2 nm away where the interaction between the Ca^{2+} and the corresponding Ca^{2+}

binding loop is negligible. During the pulling simulations, the positions of the proteins may shift and this would cause inaccurate estimation of the distance between Ca^{2+} and the corresponding Ca^{2+} binding loop. Therefore, we fixed the C_α atom of the 100th or 136th residue in CaM, which showed smallest RMSD in a separate equilibration simulation, while pulling Ca^{2+} from the binding site III or IV, respectively. The force constant *k* and speed of pulling *v* of the reference position are described in the next section. The direction of the pulling force was randomly assigned and pointed away from the center of mass of CaM to avoid clashes between CaM and the Ca^{2+}. The pulling direction was selected if the angle Ω between the pulling vector and the vector connecting center of mass of CaM and center of mass of the corresponding Ca^{2+} binding loop was within 90 degrees since Ω > 90 leads to a large work (as shown in Fig. S12). The displacement of the Ca^{2+} ion and the pulling forces were output every 20 fs for calculation of the work as shown in Eq. S9. The coordinates and velocities of the system were saved every 1 ps.

The setup of the pulling simulations is illustrated in Fig. S9. The Ca^{2+} is constrained to a reference position that is moving along $\vec{x}$ direction at a speed of $|\vec{v}|$ = 0.001 nm/ps. The force constant of the spring *k* is set to 10000 kJ / mol / nm^2 to guarantee that the Ca^{2+} strictly follows the reference position. Therefore, the pulling force is calculated as,

$$\vec{f}(t) = -k(\vec{x}(t) - \vec{v}t - \vec{x}_0) \qquad \text{Eq. (S9)}$$

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

IV.4 Justification of parameters used in steered molecular dynamics simulations

We set the pulling speed *v* and the spring constant of the external force *k* to effectively estimate Ca^{2+} binding free energy from steered molecular dynamics. We explored different

combinations of spring constant and pulling speed to rationalize the parameter setting. For example, to pull the Ca^{2+} from site III of holoCaM, we used k = 100, 1000, 10000, 100000 kJ/mol/nm^2 and v = 0.001, 0.01, 0.02 nm/ps. The simulation time was 4 ns, 0.4 ns and 0.2 ns to ensure that Ca^{2+} was pulled the same distance. In Fig. S10A, with fairly slow pulling speed, using k = 1000, 10000 and 10000 kJ/mol/nm^2, the Ca^{2+} follows the motion of the dummy bead. For k = 100 kJ/mol/nm^2, the pulling force is so weak to pull the Ca^{2+} out during the entire simulation. For k = 1000 kJ/mol/nm^2, there exists a lag in the beginning of the simulation, indicating a favored interaction between CaM and Ca^{2+} at site III. However, the tail that deviates from the straight line indicates that the thermal fluctuations in the unbound state of Ca^{2+} dominate its motion. By increasing v to 0.01 or 0.02 nm/ps, k = 1000, 10000 or 100000 kJ/mol/nm^2 meet the stiff spring approximation, shown in Figs. S9B and S9C. Comparing to v = 0.001 nm/ps, the displacement curves are much smoother.

During the exploration of factors that influence work values, we found that the pulling speed v is more sensitive. Jarzynski's group (29) showed that JE does not depend on the pulling speed. In their study, they showed for the same amount of simulation time, the estimation of the chemical potential of a simple Lennard-Jones fluid did not vary for different switching time (comparable to pulling speed v). However, in practice, the free energy calculation is more efficient with fewer slow pulling trajectories than more fast pulling trajectories given the same amount of simulation time (39). For a complex system of pulling Ca^{2+} form a holoCaM-CaMBT compound, the efficiency of convergence of distribution of work values depend on the pulling speed v. Moreover, it has been shown that slower pulling speed v reduces the perturbation from the pulling force compared to the level of the thermal fluctuations of the binding site (30). Therefore, in this study, we set a relatively slow pulling speed v = 0.001 nm/ps with a stiff spring

constant k = 10,000 kJ/mol/nm^2 to guarantee that the Ca^{2+} follows the moving reference bead that accounts for small work. To improve the efficiency of the free energy estimation, furthermore, we also employed cumulative integral (CI) extrapolation method developed by Zuckerman's group (31), which is discussed in the next session.

In Fig. S11, the typical force profiles lead to the same conclusion. At v = 0.001 nm/ps, k = 10,000 and 100,000 kJ/mol/nm^2 demonstrated the same trend in the force profiles except more thermal fluctuations in the latter, indicating a converged behavior of the dissociation of Ca^{2+} in this parameter range. Therefore, k = 10,000 kJ/mol/nm^2 were used for the pulling simulations.

IV.5 Estimation of binding free energy using Jarzynski's equality and Cumulative Integral extrapolation

After obtaining the work (w) distribution from the pulling simulations as explained in IV.3 in the *Supporting Information*, we estimated the binding free energy $\Delta G^{\text{holoCaM-CaMBT}}$ between a one-Ca^{2+}-missing complex of holoCaM-CaMBT (G_U) and a Ca^{2+}-loaded complex of holoCaM-CaMBT (G_B) according to JE (32),

$$\begin{aligned} \Delta G^{\text{holoCaM-CaMBT}} &= G_{\text{B}} - G_{\text{U}} \\ &= k_B T ln \langle e^{-\frac{W}{k_B T}} \rangle \\ &\approx k_B T ln \sum_{n=1}^{M} e^{-w_n/k_B T}/M \end{aligned} \qquad \text{Eq. (S10)}$$

where k_B is Boltzmann constant, T is the temperature, M is number of pulling simulations (M = 100 ~ 150), w_n is the work in nth pulling calculated as,

$$w_n(\text{B} \rightarrow \text{U}) = \int_0^t \overrightarrow{f_n}(\tau) \cdot \vec{v}_n d\tau \qquad \text{Eq. (S11)}$$

B and U stands for bound and unbound states of the Ca^{2+} ion, respectively.

However, accurate estimation of the binding free energy ΔG relies on the sufficient sampling of small work (32). Zuckerman's group (31) showed that using a cumulative integral (CI) to extrapolate the free energy estimation to $n \rightarrow \infty$ can help reduce the required simulation data by 5 to 40 fold (30, 31), where n is the number of work values. In Table S6, we showed the estimation of change in Ca^{2+} binding free energy (ΔΔG) by using JE or CI estimation. We found that JE or CI produced the same signs of the ΔΔG, however, the magnitudes were not the same. The discrepancy in the free energy calculation lies in the requirement of sufficient sampling of small work for direct JE calculations, which can be overcome by CI extrapolation (28).

V. Analyses

V.1 Calculation of model free order parameter S^2

A number of N (~ 500) representative coarse-grained structures were selected from the free energy surface of unbound Ng protein using importance sampling method (17). The NMR model-free order parameter S_i^2 (33) for the backbone N-H bond of residue *i* is calculated on the selected structures by relating to the Root Mean Square Fluctuations (RMSF) of the C_α bead through the following empirical relation (34),

$$S_i^2 = 1 - 0.5 * \ln(1 + \frac{RMSF_i}{23.6\text{Å}} * 10.0) \qquad \text{Eq. (S12)}$$

$$RMSF_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(\boldsymbol{r}_i^j - \frac{1}{N}\sum_{k=1}^{N}\boldsymbol{r}_i^k)^2} \qquad \text{Eq. (S13)}$$

where $\boldsymbol{r}_i^j$ is the position of the C_α bead of residue *i* at frame *j* and *N* is the total number of frames. $RMSF_i$ was computed after structural alignment.

V.2 Definition of helicity

The fraction of helical structure, or helicity, of Ng (G25-A42) was calculated using the following formula (35),

$$f_H = \frac{1}{N-3}\sum_{i=1}^{N-3} < \Theta(\Delta\phi - |\phi_i - \phi_0|) > \quad \text{Eq. (S14)}$$

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

V.3 Definition of Z

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

$$Z_{ij}^{ss} = \Theta\left(c - \frac{d_{ij}^{ss}}{\rho_i + \rho_j}\right) \quad \text{Eq. (S15)}$$

$$Z_{ij}^{bb} = \Theta(c - \frac{d_{ij}^{bb}}{\rho_{HB}}) \quad \text{Eq. (S16)}$$

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

V.4 Sampling protein configuration for structure analyses

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

$$w(x) = \begin{cases} 1 & if\ p(x) \geq \rho \\ 0 & else \end{cases} \quad \text{where } p(x) = e^{-G(d_{COM}(x))/k_B T} \qquad \text{Eq. (S17)}$$

$p(x)$ is the probability of the configuration x in reweighted ensemble; $G(d_{\mathrm{COM}})$ is the reweighted/unbiased free energy obtained from the WHAM analyses along reaction coordinate d_{COM} (G_{min} is scaled to 0); ρ is a random number generated between 0 and 1; k_B is Boltzmann constant and T is the temperature. The ensemble generated after Boltzmann reweighting thus obeys the canonical distribution and is employed for subsequent analyses. Each structure from the biased ensemble was challenged by acceptance probability $w(x)$ and a total of 23,722 structures were sampled. This sample achieves a distribution of $P^{\mathrm{sample}}(d_{\mathrm{COM}})$. In order to assess the quality of the sampling, we computed the surprisal value compared with the original unbiased distribution $P^{\mathrm{ori}}(d_{\mathrm{COM}})$ defined by the following formula (36),

$$\text{surprisal} = \sum_r -P^{\mathrm{ori}}(r) \ln\left[P^{\mathrm{sample}}(r)/P^{\mathrm{ori}}(r)\right] \qquad \text{Eq. (S18)}$$

where the summation is over all the order parameter r (in this case d_{COM}). A surprisal value of 0.14 ensures that the sampled ensemble can well represent the original distribution.

V.5 Self-organized clustering algorithm

We applied the self-organized neural-net clustering algorithm (37-39) to determine the structures of the apoCaM-$\mathrm{Ng}_{13\text{-}49}$ complexes from the umbrella sampling simulations. In this clustering method a vector with M elements represents each conformation j, $x_j = [x_{1j},\ x_{2j},\ \dots,\ x_{Mj}]$, where $j = 1, 2, \dots N$ and N is the number of conformations selected for clustering analysis. The element x_{ij} ($i = 1, 2, \dots, M$) stands for the Euclidean distance between side-chains of the

polypeptide chain in the conformation j. To partition the various conformations into distinct clusters, the clusters are described by the cluster center and the size of the cluster is determined by a radius R_c. A given conformation is assigned to cluster k if the distance between the vector x_j and the center of the k^{th} cluster,

$$C_k = \frac{1}{M_k}\sum_{j=1}^{M_k} x_j, \qquad \text{Eq. (S19)}$$

where $C_k = [C_{1k}, C_{2k}, \ldots, C_{Mk}]$ and $M_k$is the total number of conformations in C_k.

Conformation j belongs to C_k if the Euclidean distance between conformation j and the cluster center k, $d_{jk} = \left|x_j - C_k\right| < R_c$, where R_c is a preassigned value. In the current study we used the native contact pairs from the unbound structure of apoCaM and $Ng_{13\text{-}49}$ as the M elements and a following cutoff distance R_c = 25 σ (σ = 3.8 Å) to categorizes the structures into seven distinct clusters.

V.6 Correlation between intra- and inter- molecular interactions in apoCaM-$Ng_{13\text{-}49}$ binding

In order to better understand the relation between interactions within the two proteins and intermolecular contacts, we built a correlation map between contacts (Fig. S4). The correlation between two contact pairs m and n is computed as in the following equation,

$$corr_{mn} = \frac{< q_m q_n > - < q_m >< q_n >}{\sqrt{< q_m^2 > - < q_m >^2}\sqrt{< q_n^2 > - < q_n >^2}} \qquad \text{Eq. (S20)}$$

where $q_{m(n)}$ is the contact state of the contact pair m (n): 1 if m (n) is a contact, 0 if not. The list of contacts are provided in Table S8. The correlation between those contacts falls into several categories.

a) Correlation: as shown in diagonal, the contact formation within apoCaM, between apoCaM and Ng_{13-49}, and within Ng_{13-49} are correlated.

b) Anti-correlation: contact formation between nCaM and Ng_{13-49} anti-correlates with contact formation within nCaM; contact formation between cCaM and Ng_{13-49} anti-correlates contact formation within cCaM. The anti-correlation tells that the contact formation within nCaM or that within cCaM competes with its interaction with Ng_{13-49}. We infer that binding of the target is responsible for the repacking of the nCaM or cCaM by direct competitive interaction with the Ca^{2+} binding loops.

c) No correlation: there is no apparent correlation among contact formation within nCaM and contact formation within cCaM, indicating the two domains of CaM are relative independent.

d) Mixed correlation: contact formation between nCaM and cCaM has a mixed correlation with contact formation between nCaM and Ng_{13-49} as well as contact formation between cCaM and Ng_{13-49}.

Supporting Figures

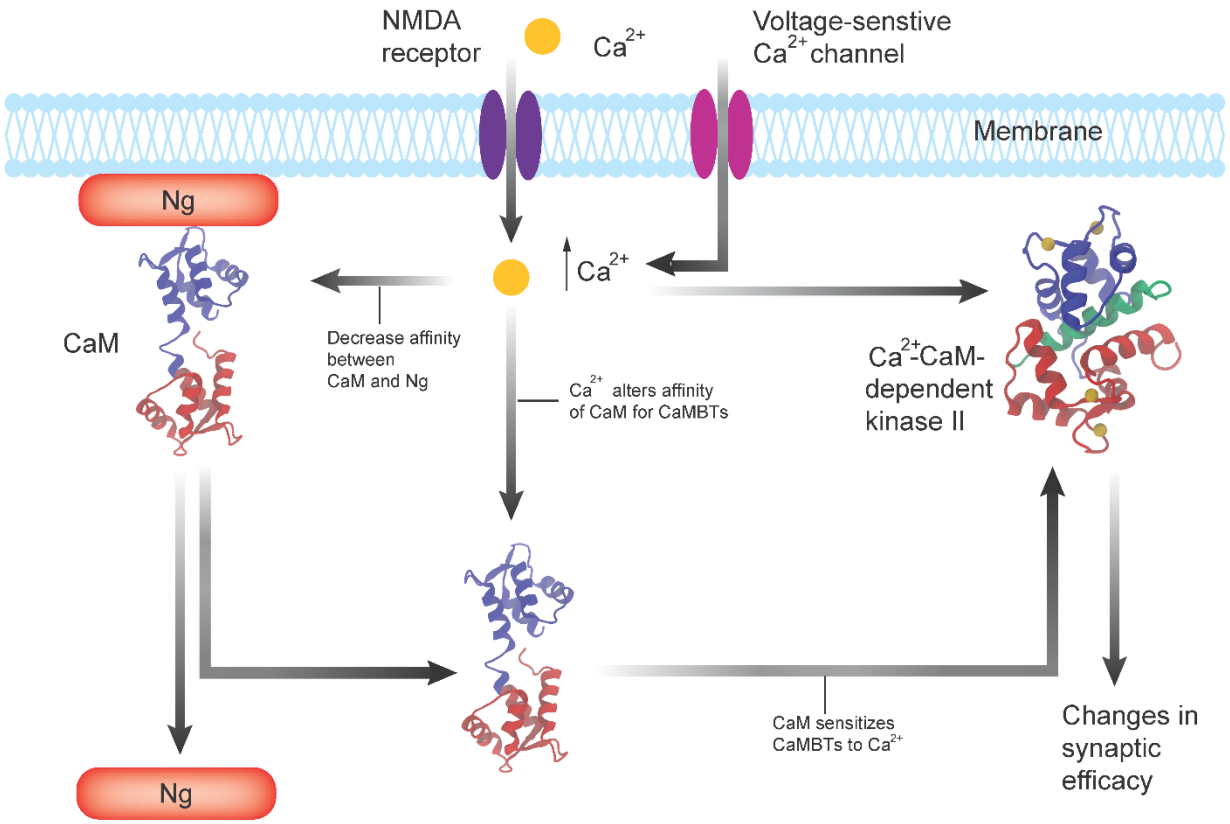

Fig. S1. Overview of the effects of CaM-dependent Ca^{2+} signaling and effects of CaM binding targets (CaMBTs) on changes in synaptic plasticity. Many of the effects of intracellular Ca^{2+} on synaptic plasticity are mediated through CaM-regulated proteins. Increase in intracellular Ca^{2+}, generated through the activity of NMDA (N-methyl-D-aspartate) receptors or voltage-sensitive Ca^{2+} channels, results in the release of CaM that is bound to Ng. CaM mediates the Ca^{2+} stimulation of CaMKII which is required for changes in synaptic plasticity. The structures of Ca^{2+}-free CaM (apoCaM, PDBID: 1CFD) and Ca^{2+}-CaM-CaMKII peptide (PDB ID: 1CDM) are provided. CaM is colored as follows, red → nCaM (residue 1 to 76), gray → central linker (residue 77 to 82), blue → cCaM (residue 83 to 148) and the CaMKII peptide is colored in green.

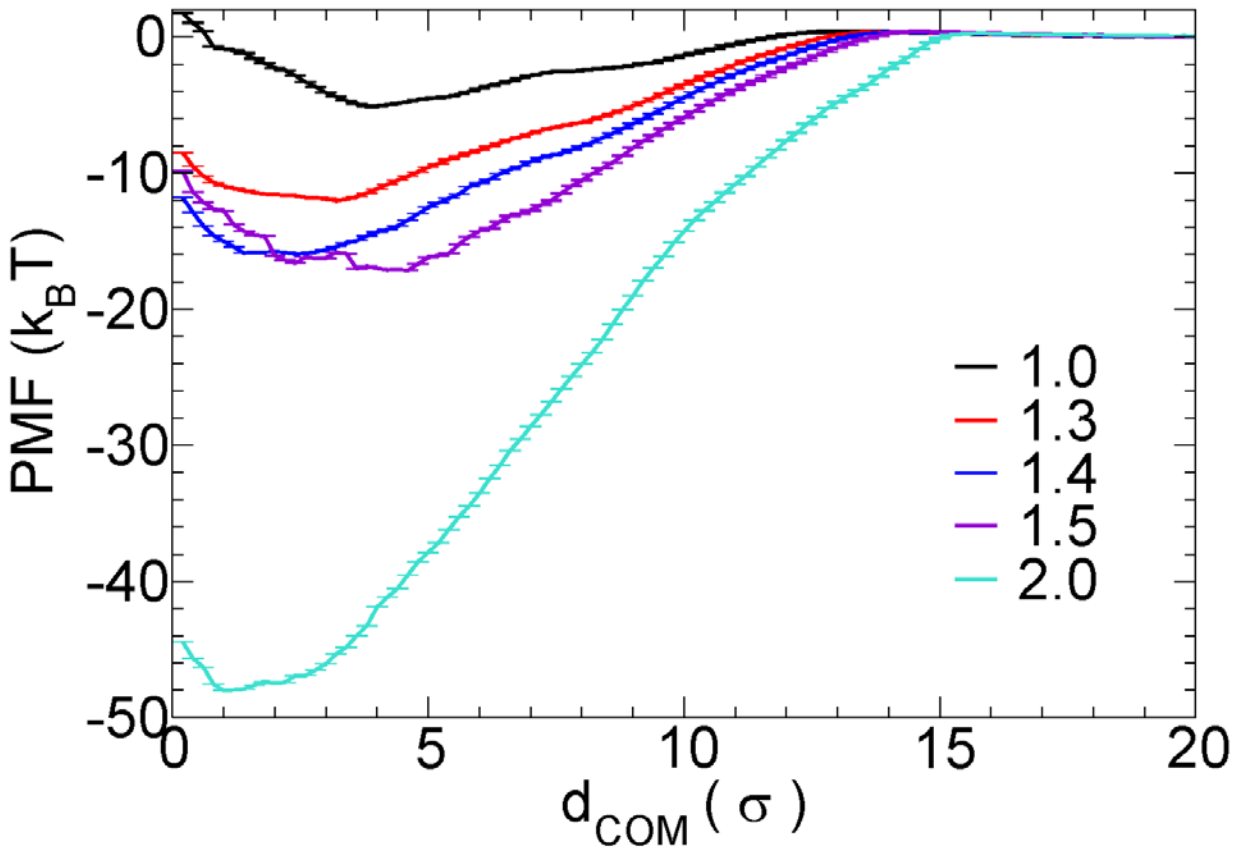

Fig. S2 Reweighted potential of mean force of apoCaM and Ng_{13-49}. The PMF was reweighted from umbrella sampling simulations using WHAM at varying scaling factors of the inter-molecular nonbonded interaction λ (excluding electrostatic interactions). KB statistical dihedral angle potential was employed. d_{COM} is the distance between center of mass of apoCaM and center of mass of Ng_{13-49}. σ equals 3.8 Å. T = 1.1 ε / k_B. ε = 0.6 kcal / mol.

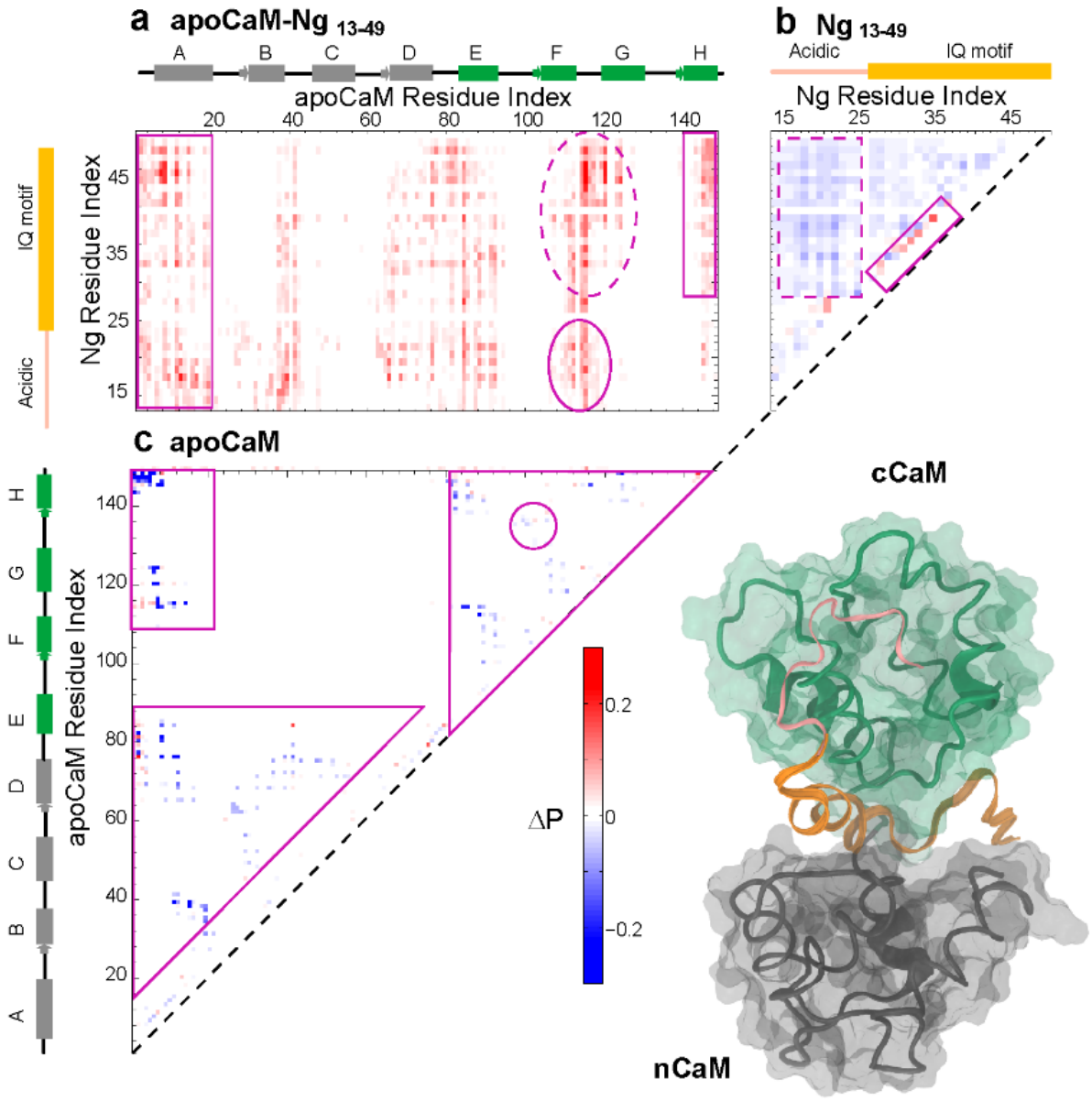

Fig. S3 Difference in probability of contact formation between the bound ensemble of the most dominant cluster (cluster 1) and the unbound ensemble. The difference maps for apoCaM intramolecular (a), Ng_{13-49} intramolecular (b) and apoCaM-Ng_{13-49} intermolecular (c) contacts are provided. d_{COM} = 20.0 s is used for the unbound state. The schematic representation of the helices of apoCaM and Ng_{13-49} are provided along the axes. The representative structure, which has minimal root mean square deviation from the averaged structure, is in ribbon representation and colored according to the secondary structures: nCaM is in gray, cCaM is in green, acidic region of Ng_{13-49} is in pink and IQ motif of Ng_{13-49} is in orange. The regions of particular interest are encircled by ellipses, rectangles, circles and triangles. The color bar represents the difference in probability of contact formation ΔP between the bound and the

unbound conformations where positive ΔP indicates an increase of contact probabilities in the bound ensemble.

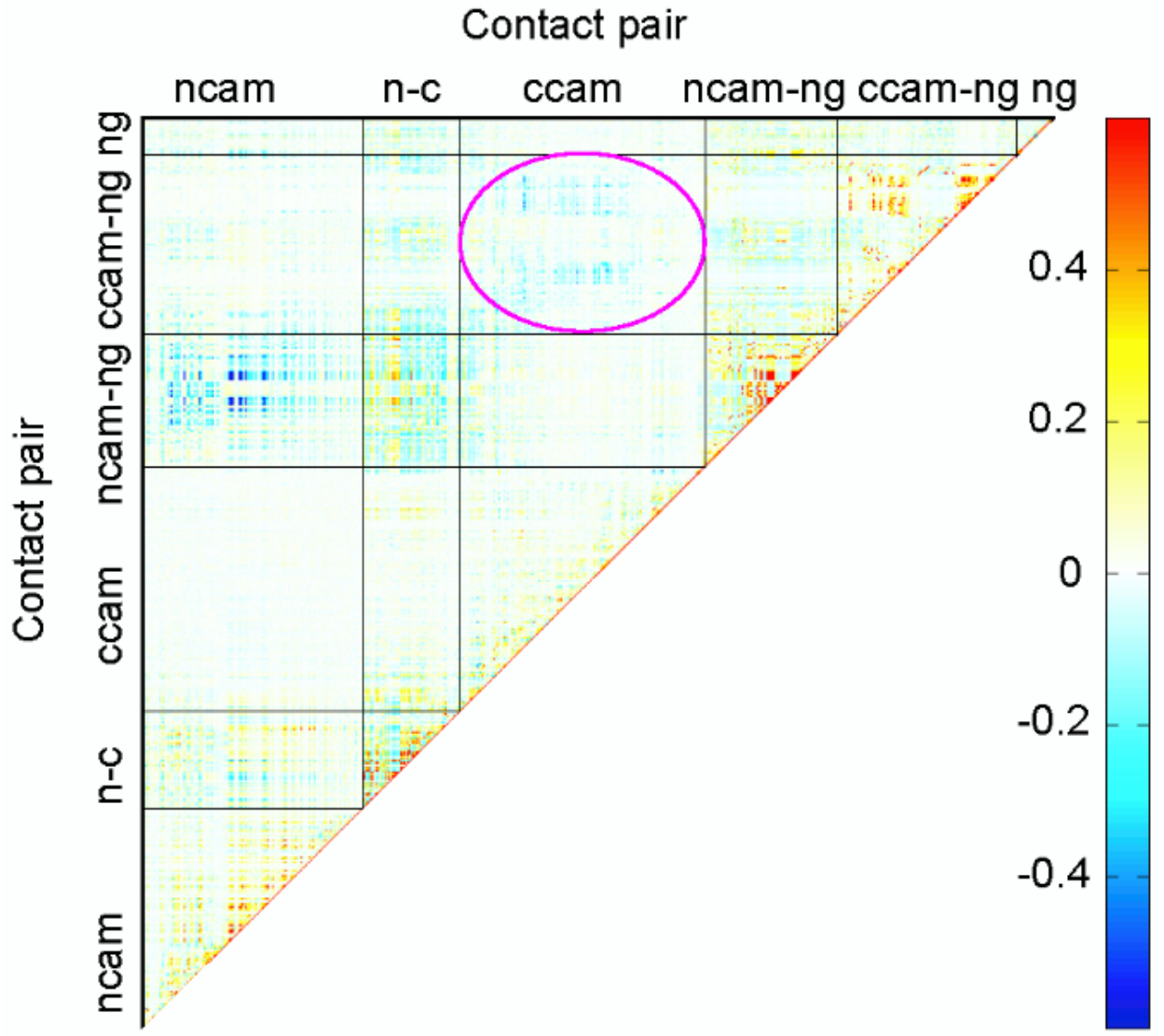

Fig. S4 Correlation map between contacts formation involving apoCaM and Ng_{13-49} in bound ensemble (cluster 1). The contact pair list is provided in Table S8. The blue color shows strong anti-correlation between involved contact pairs; the red color shows strong correlation; white means no correlation.

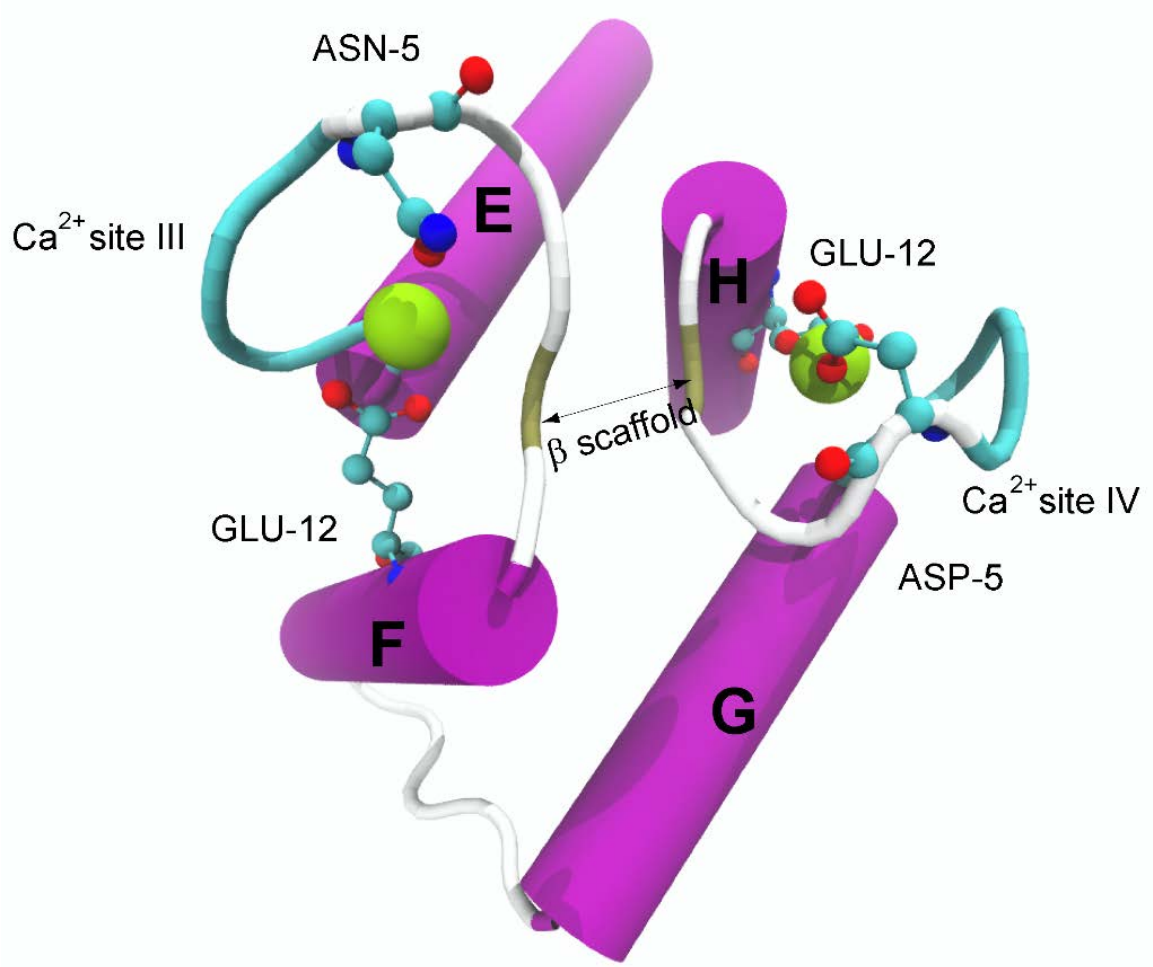

Fig. S5 Illustration of the EF-hand β-scaffold. The structure is from C-terminus of the crystal structure of holoCaM (PDB: 1CLL). The calcium ions are shown in yellow beads. The 5th (+Z coordinating ligand) and 12th (-Z coordinating ligand) residues in the Ca^{2+} binding loops are shown in ball-and-stick representation.

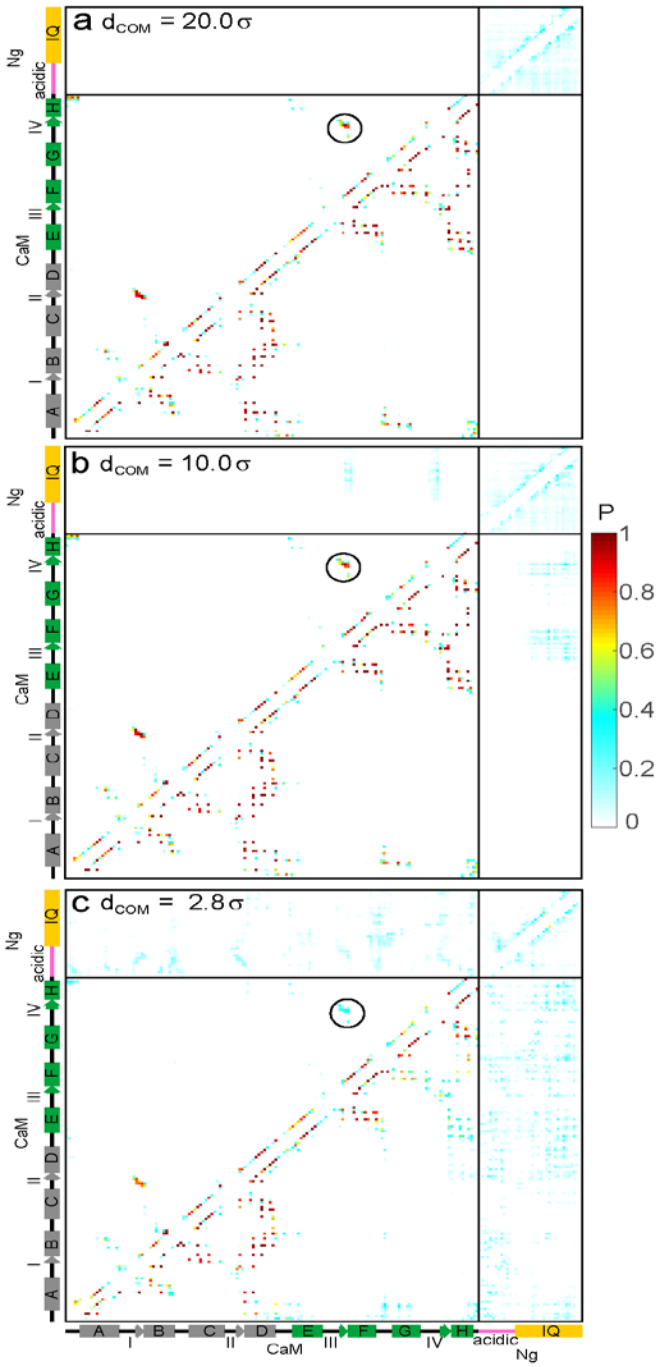

Fig. S6 Probability of contact formation in unbound, encounter and bound ensemble of apoCaM-Ng$_{13-49}$. In panel (a), the probability map of contact formation for the unbound state was calculated using the ensemble when apoCaM and Ng$_{13-49}$ are well separated at d_{COM} = 20.0 σ. σ = 3.8 Å. (b) the probability map of contact formation for the encounter complex at d_{COM} = 10.0 σ and (c) in the bound state (d_{COM} = 2.8 σ) are provided. The upper and lower triangles show contacts maps for backbone-backbone conatcts and sidechain-sidechain contacts, respectively. The schematic representation of the helices of apoCaM and Ng$_{13-49}$ are provided along the axes. The region encircled by a circle marks the interaction between the two β-strands in the EF-hand β-scaffold. The color bar represents the probability of contact formation P.

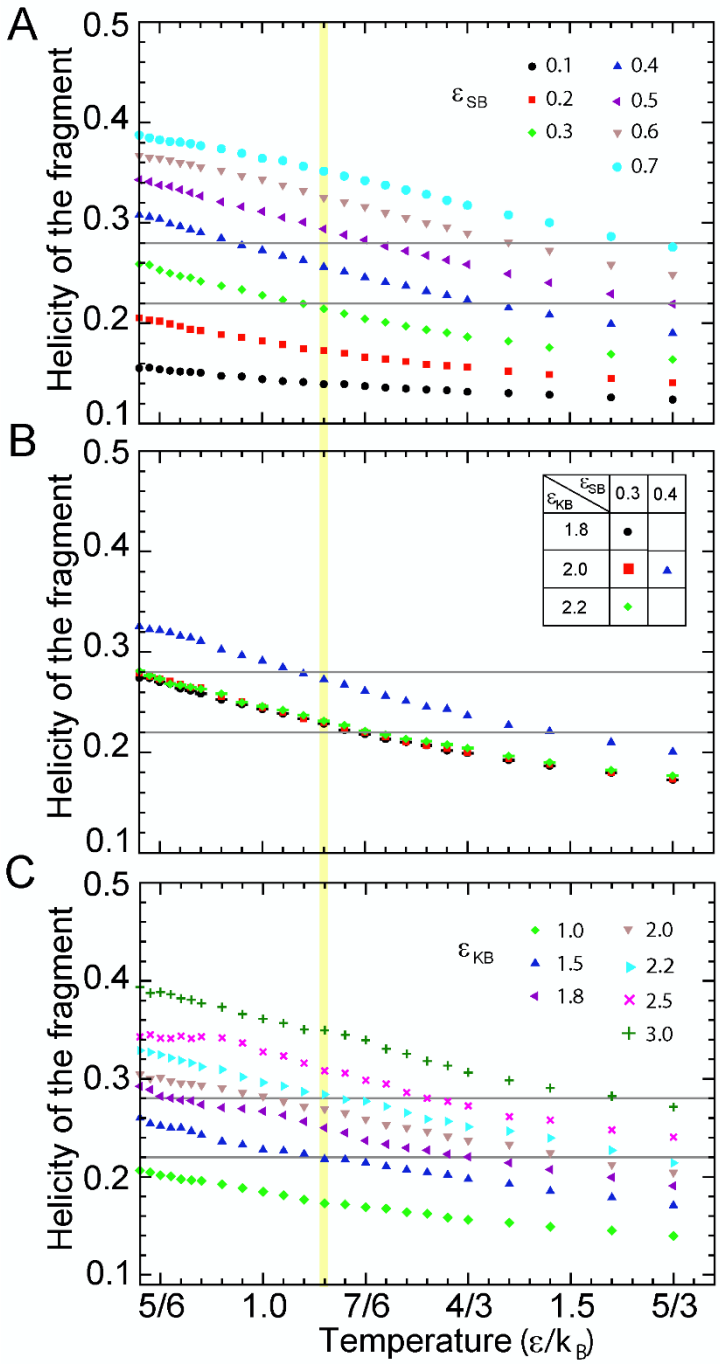

Fig. S7 Calculated helicity of the Ng protein fragment with residual structure matches with experiment using three dihedral angle potentials. The helicity of the fragment G25-A42 was computed from REMD simulations of the intact Ng protein with the coarse-grained side-chain C_α model. Three models of dihedral angle potential were used: (A) Structure-Based (SB) potential; (B) Hybrid potential; (C) Karanicolas-Brooks (KB) statistical potential. The two gray lines mark the upper and lower limit of the helicity of the segment from the experiment (3); the yellow shade marks the corresponding reduced temperature 1.1 ε / k_B. Helicity was estimated based on the dihedral angles between four consecutive C_α beads (see definition in V.2 in *Supporting Information*).

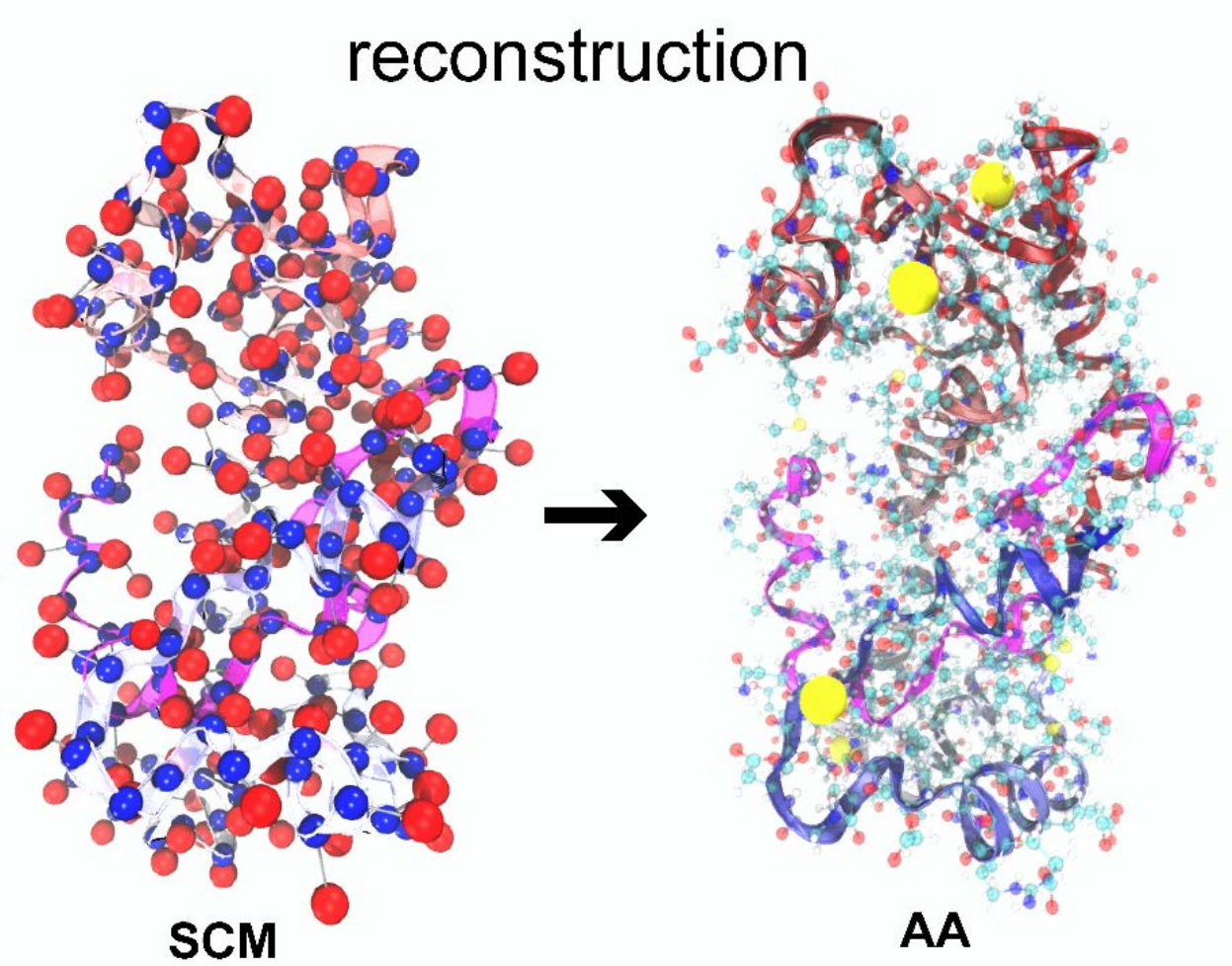

Fig. S8 Illustration of protein reconstruction from coarse-grained configuration to atomistic configuration for the holoCaM-Ng$_{13-49}$ complex. The backbone of CaM and Ng$_{13-49}$ is shown in ribbon representation. CaM is colored from red (N-Domain) to blue (C-Domain); Ng$_{13-49}$ is shown in magenta. (left) The Sidechain-C$_\alpha$ model (SCM), where the backbone atoms are represented by the C$_\alpha$ beads (blue balls) and sidechain atoms are represented by the sidechain beads (red balls). (right) All-atomistic (AA) configuration of the holoCaM-Ng$_{13-49}$ complex with Ca^{2+} (yellow balls) added to the Ca^{2+} binding loops. The other atoms are colored according to the atom names (oxygen atoms are in red, carbon in cyan, hydrogen in white, nitrogen in blue, etc.).

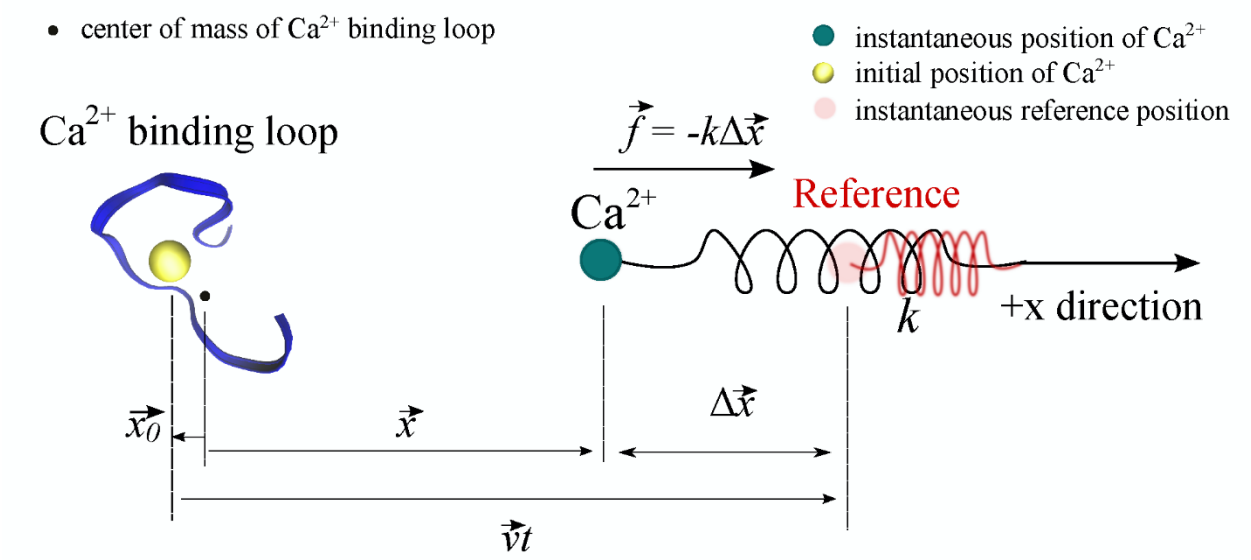

Fig. S9 Schematic illustration of pulling Ca^{2+} from one of the calcium-binding loop of CaM. The rest of the system including CaM and the solvent molecules is not shown in this illustration for better visualization.

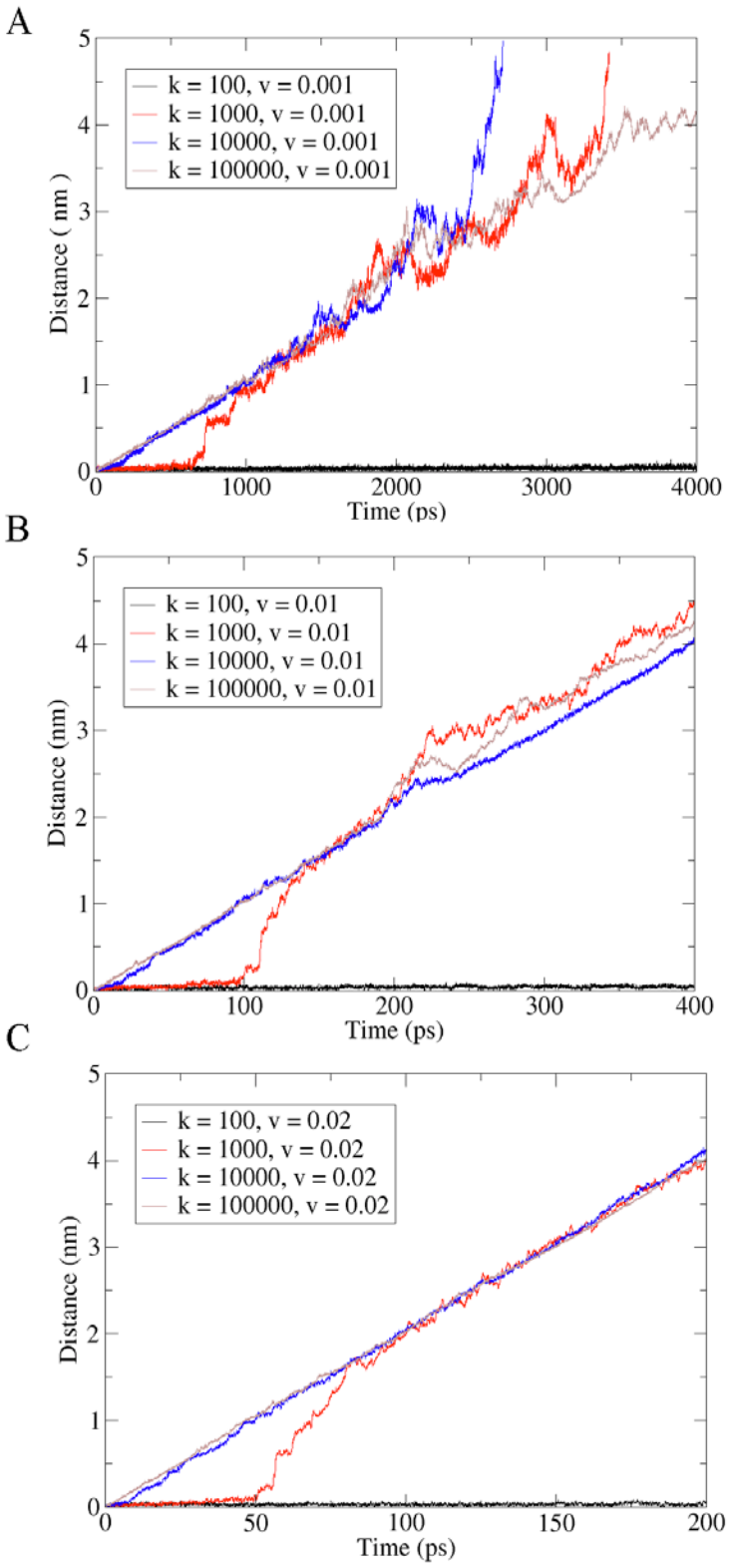

Fig. S10 The displacement of the Ca^{2+} from the Ca^{2+} binding loop during the pulling simulations of Ca^{2+} from the Ca^{2+} binding site III of holoCaM. The unit of k is kJ/mol/nm^2 and the unit of v is nm/ps. (A) (B) (C) show the case pulling speed v = 0.001, v = 0.01 and v = 0.02 nm/ps.

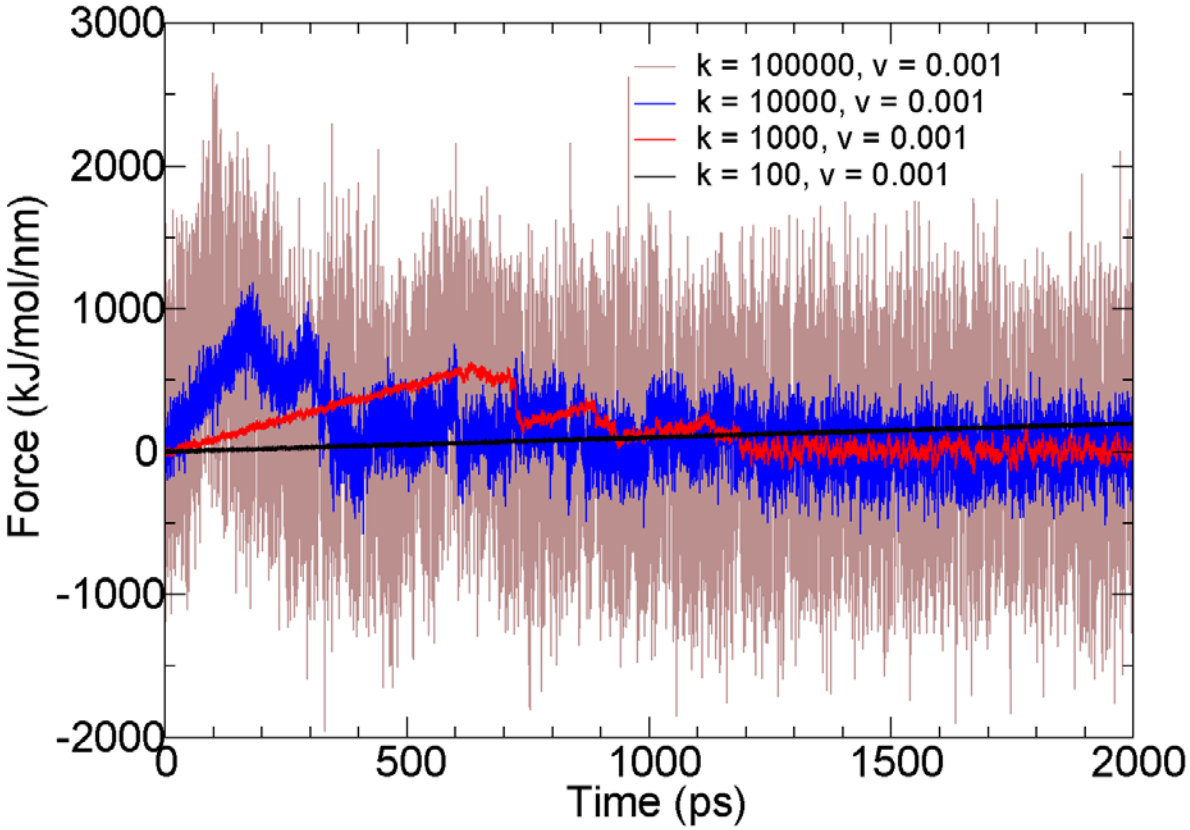

Fig. S11 The profile of external force in pulling simulations of Ca^{2+} from the Ca^{2+} binding site III of holoCaM. The pulling speed v = 0.001 nm / ps. The unit of spring constant k is kJ/mol/nm^2 and the unit of pulling speed v is nm/ps.

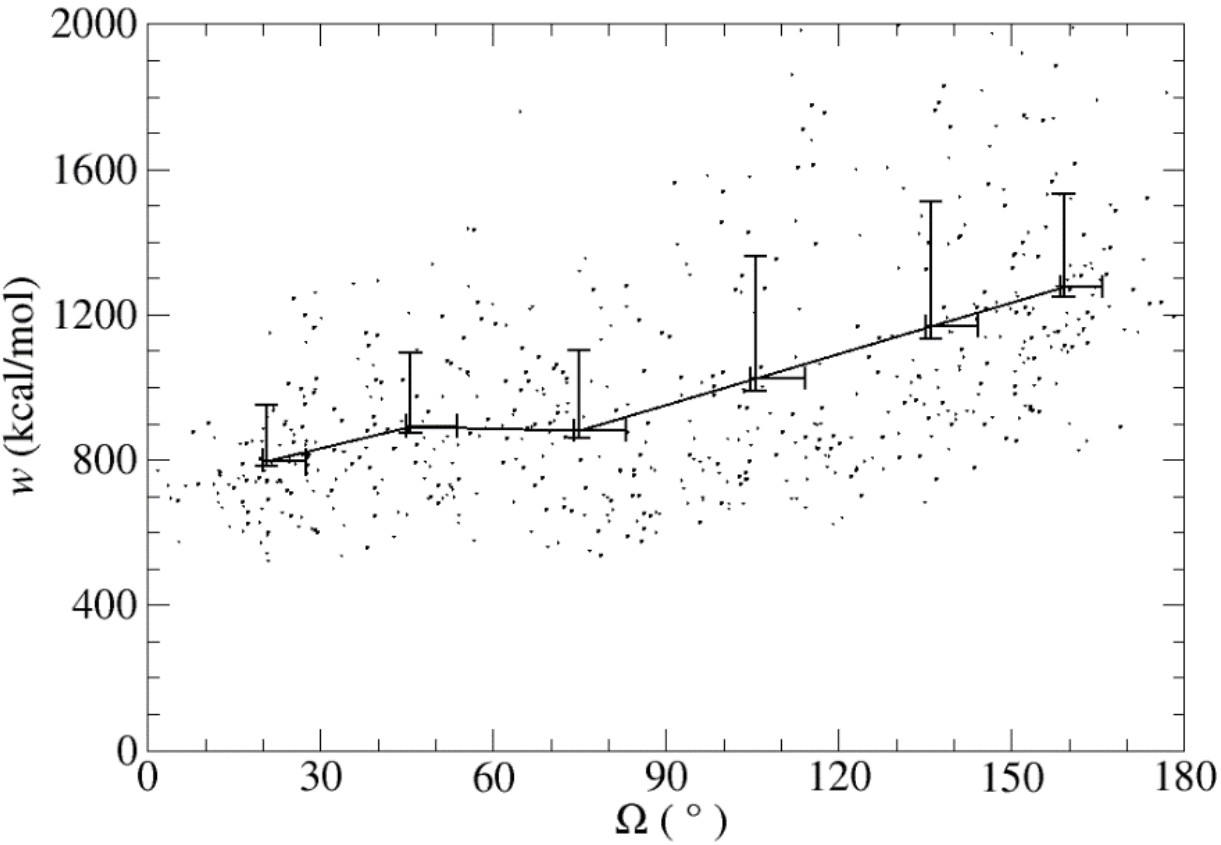

Fig. S12 Work done by external pulling force of pulling Ca^{2+} at different pulling angles. At each pulling angle zone Ω in [0, 30], [30, 60], [60, 90], [90, 120], [120, 150], [150, 180], the Ca^{2+} were pulled from site IV of holoCaM for 100 times. The average work, deviation (upper bar) as well as standard error (lower bar) are provided. The pulling speed v = 10 nm / ns and the spring constant of the pulling force k = 1000 kJ / mol / nm^2.

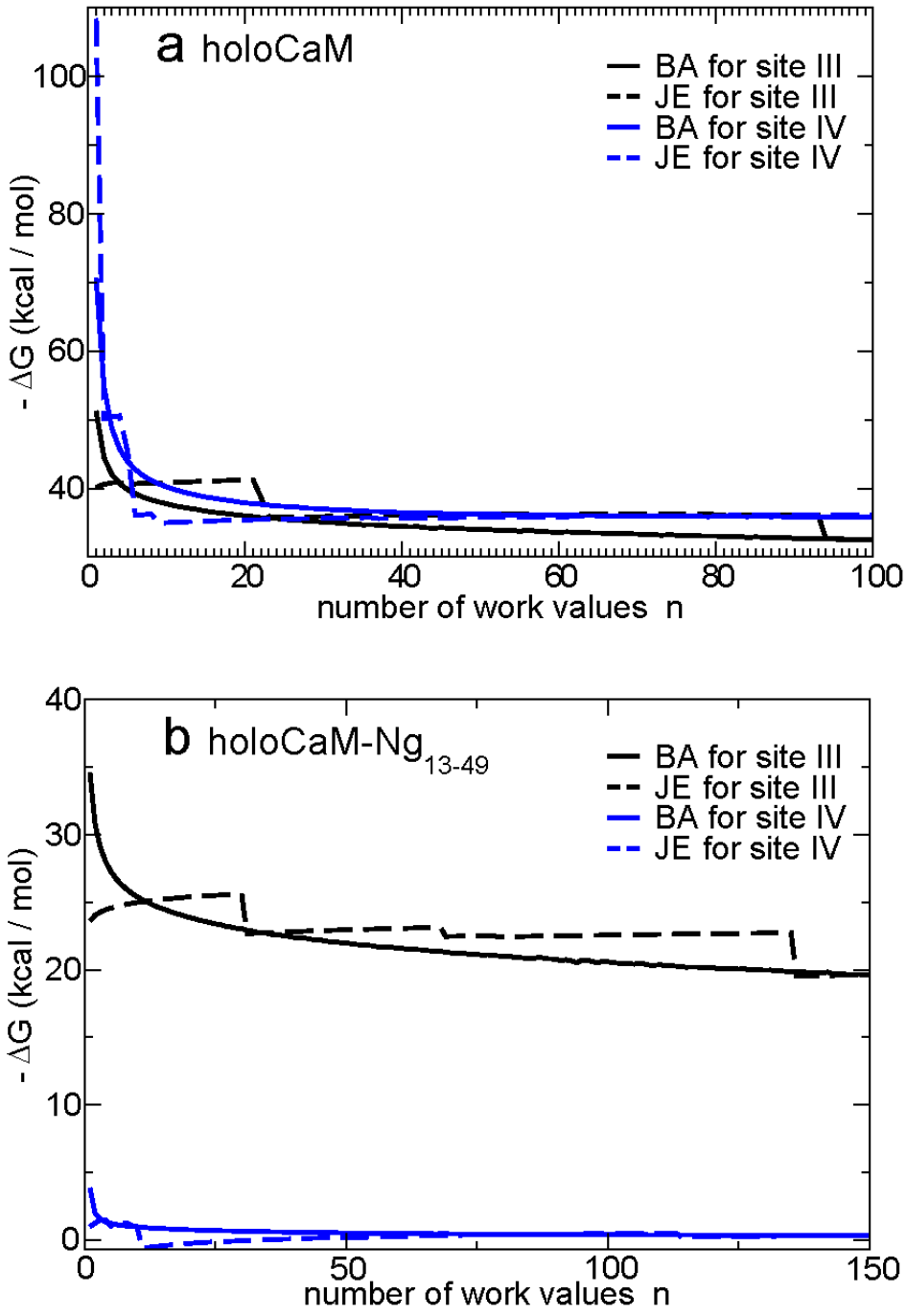

Fig. S13 The running Jarzynski equality (JE) estimate and subsampled block-averaged (BA) estimate are plotted as a function of the number of trajectories used in the estimate. The binding free energies of Ca^{2+} at site III and IV of holoCaM using the running JE method and the BA method are shown in panel (a) and the binding free energies of Ca^{2+} at site III and IV of holoCaM-Ng_{13-49} (using the conformation shown in Fig. 5) are shown in panel (b).

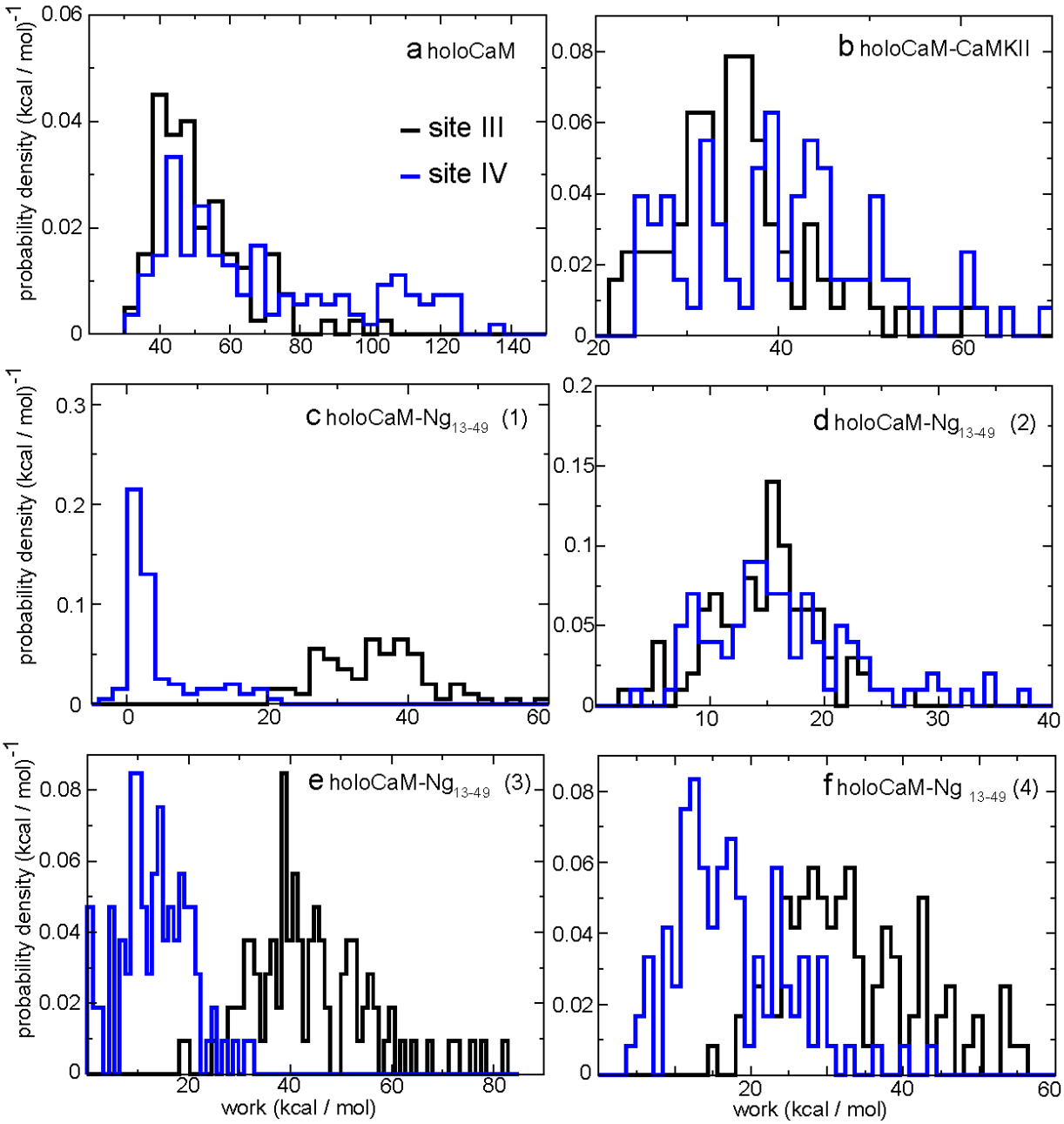

Figure S14 Distribution of work for dissociation of Ca^{2+}. (a) and (b) represent the holoCaM and holoCaM-Ng_{13-49}, respectively. (c-f) represent the four conformations of holoCaM-Ng_{13-49}.

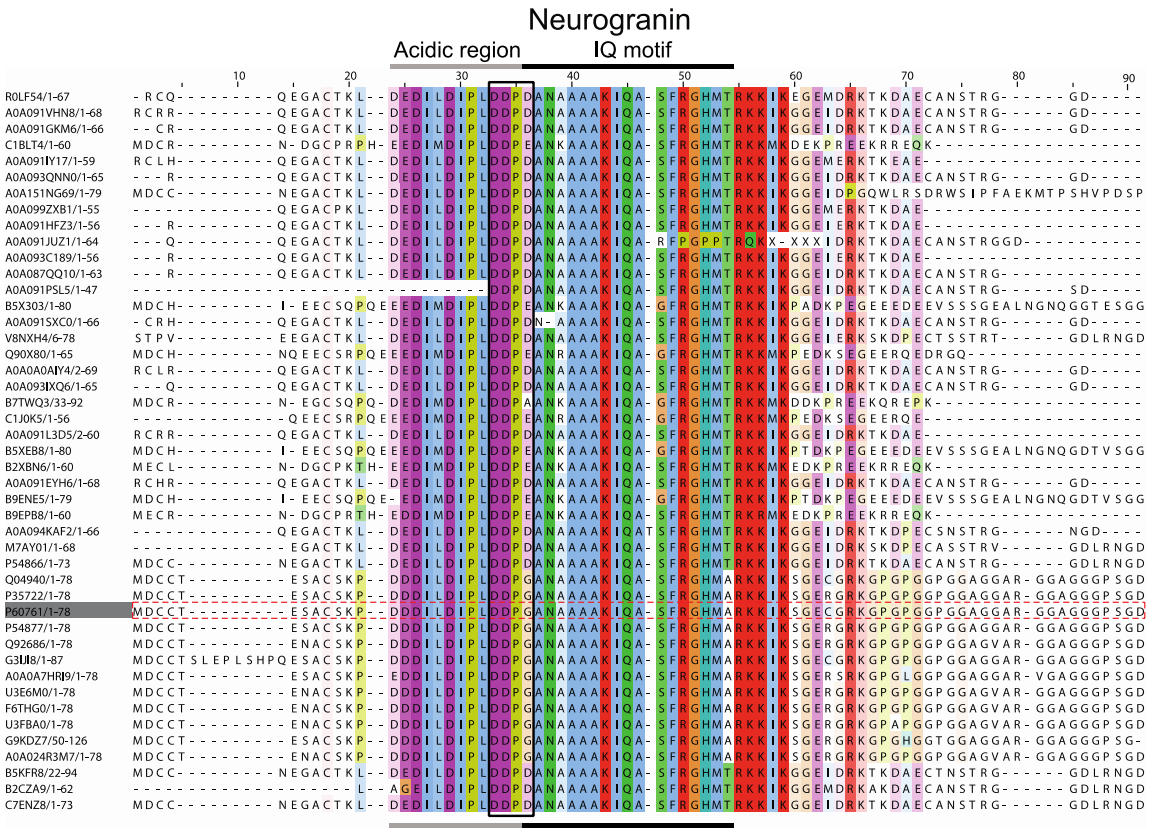

Fig. S15 Sequence alignment of neurogranin (Ng). 45 sequences of Ng (see Table S14 for the details of species associated with each sequence) were selected from UniProtKB (www.uniprot.org) by searching the keyword “neurogranin”. The sequences were aligned using the “Align” tool in UniProt and visualized in Jalview (40). The ClustalX color scheme in Jalview was used to highlight the residues and shading intensity of the color is based on the conservation over all the sequences in the alignment. The “acidic region” and the “IQ motif” of Ng are indicated by the grey and black line, respectively. The “DDPG (D/E/A)” motif is indicated inside the black area. The sequence of Ng from mouse (UniProt id P60761) that was used in this study is shown inside the red dotted area.

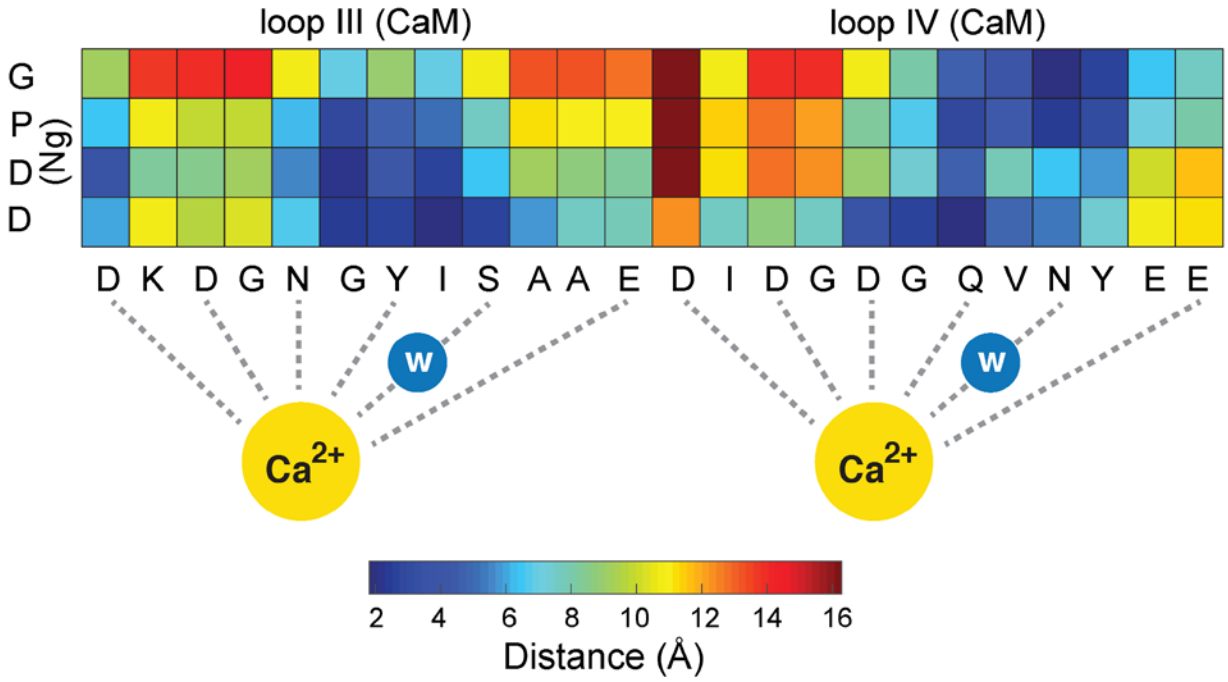

Fig. S16 Distance map between the Ca^{2+} binding loops (III and IV) of CaM and the "DDPG" motif of Ng. The distance here refers to the closest distance between the corresponding residues. The Ca^{2+} ions (shown as yellow spheres) coordinate (indicated by grey dotted lines) with the first, third, fifth, seventh, ninth (through water molecule (w), shown as a blue sphere) and twelfth residue of the loops. The distance to distance map is based on the holoCaM-Ng_{13-49} complex structure in Fig. 5a in the main text.

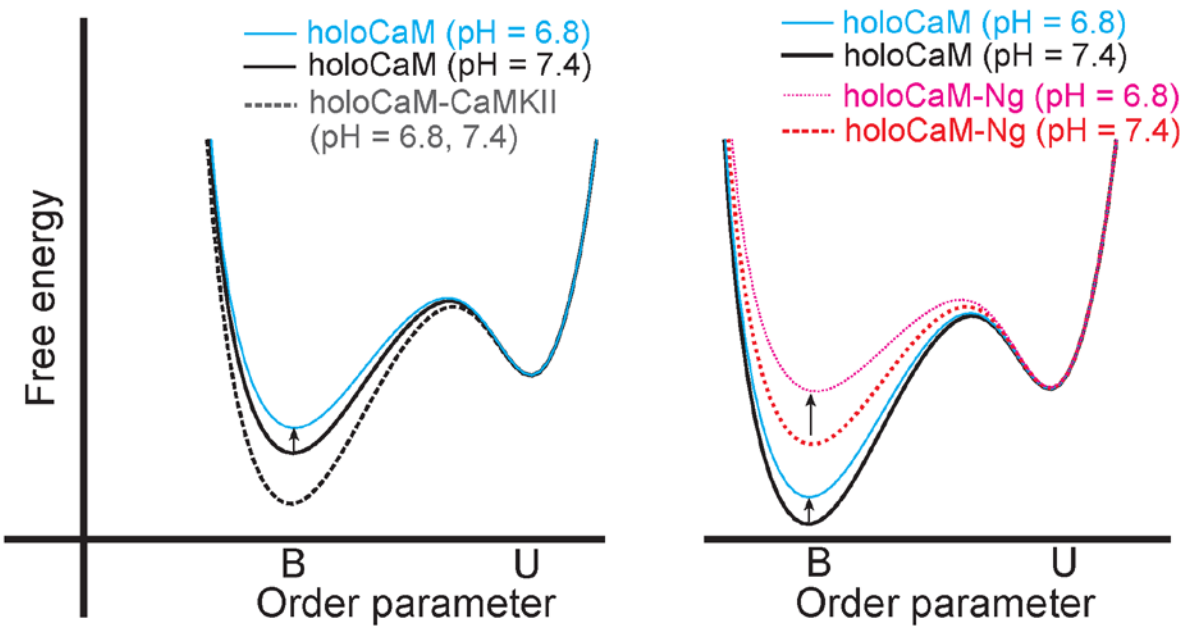

Fig. S17 Illustration of the changes in the binding free energy ΔG and ΔΔG from pH = 7.4 to pH = 6.8. $\Delta G = G_B - G_U$. $\Delta\Delta G = \Delta G^{holoCaM\text{-}CaMBT} - \Delta G^{holoCaM}$. B and U stand for bound and unbound states of the Ca^{2+}, respectively. The arrows show increase of ΔG for Ca^{2+} in all the three systems.

Supporting Tables

Table S1 Dissociation constant K_d of apoCaM and Ng_{13-49} with several values of strength of non-electrostatic intermolecular hydrogen bonding and van der Waals interactions λ. λ is shown in Eq 4. Experimentally measured K_d = 680 nM (4).

Strength λ	ΔPMF (k_BT)	K_d (nM)
1.0	5.12	6.0×10^{6}
1.3	12.04	5.9×10^{3}
1.4	15.94	119.0
1.5	17.16	35.4
2.0	48.01	1.4×10^{-11}

Table S2 The population of major clusters from computer simulations and the correlation coefficient of their computational "apparent chemical shifts" with the data from NMR experiments (4).

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

Table S3 EF-hand angles in several forms of CaM. Definition of EF hand angles is provided in IV.1 from the *Supporting Information*.

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

Table S4 The composition of the three dihedral potentials for modeling Ng or Ng_{13-49}. SB stands structure-based and KB stands for Karanicolas-Brooks statistical potential (12).

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

Table S5 Correlation between computed and experimentally-derived NMR model free order parameter S^2.

Models	SB		Hybrid	KB	
	ε_{SB} = 0.3	ε_{SB} = 0.4		ε_{KB} = 1.8	ε_{KB} = 2.0
Correlation Coefficient	0.40	0.73	0.29	0.78	0.64

Table S6 Difference in binding free energy of Ca^{2+} ($\Delta\Delta G$) calculated from non-equilibrium molecular simulations and from the experiments at pH = 7.4. Direct use of Jarzynski's equality and a cumulants integral extrapolation was used in the calculation of binding free energy of Ca^{2+} from the simulations.

System		holoCaM-Ng_{13-49} (average)	holoCaM-CaMKII
$\Delta\Delta G$ (kcal/mol) Jarzynski's equality	Site III	8.5±2.7	-0.1
	Site IV	23.9±1.1	-3.4
$\Delta\Delta G$ (kcal/mol) CI extrapolation	Site III	9.2±2.2	-0.5
	Site IV	22.4±0.9	-7.0
$\Delta\Delta G$ (kcal/mol) experiment	Site III/IV	+2.5	-3.3

Table S7 Difference in binding free energy of Ca^{2+} ($\Delta\Delta G$) for Ca^{2+} at site III calculated from non-equilibrium molecular simulations at pH = 6.8. Direct use of Jarzynski's equality and a cumulants integral extrapolation was used in the calculation of binding free energy of Ca^{2+} from the simulations.

System	holoCaM-Ng_{13-49} (average)	holoCaM-CaMKII
$\Delta\Delta G$ (kcal/mol) Jarzynski's equality	14.6±2.8	+0.7
$\Delta\Delta G$ (kcal/mol) CI extrapolation	10.5±2.1	-5.6

Table S8 Contact pair list used for correlation analysis. The non-specific contact pairs are determined from the difference contact map analysis. A contact pair is selected if the magnitude of the change in the probability of the contact pair in cluster 1 (Fig. S3) is greater than 0.2 from unbound state to the bound state. The contact pairs are listed in a sequence of categories: within nCaM, between nCaM and cCaM, within cCaM, between nCaM and Ng_{13-49}, between cCaM and Ng_{13-49} and within Ng_{13-49}. The sequence of apoCaM is from NMR structure (PDB: 1CFD) and sequence of Ng_{13-49} is provided in the *Materials and Methods* in the main text.

Index	Residue Index	Residue Index	Interaction Category	Index	Residue Index	Residue Index	Interaction Category
1	4	12	nCaM	2	4	65	nCaM
3	4	69	nCaM	4	4	8	nCaM
5	4	9	nCaM	6	5	9	nCaM
7	6	10	nCaM	8	7	11	nCaM

9	8	12	nCaM	10	9	13	nCaM
11	9	65	nCaM	12	9	69	nCaM
13	10	14	nCaM	14	11	15	nCaM
15	11	39	nCaM	16	12	16	nCaM
17	12	39	nCaM	18	12	41	nCaM
19	12	65	nCaM	20	12	68	nCaM
21	12	69	nCaM	22	12	72	nCaM
23	13	17	nCaM	24	13	20	nCaM
25	13	65	nCaM	26	14	18	nCaM
27	15	38	nCaM	28	15	39	nCaM
29	16	20	nCaM	30	16	27	nCaM
31	16	35	nCaM	32	16	38	nCaM
33	16	39	nCaM	34	16	65	nCaM
35	16	68	nCaM	36	18	38	nCaM
37	19	27	nCaM	38	19	31	nCaM
39	19	34	nCaM	40	19	35	nCaM
41	19	37	nCaM	42	19	38	nCaM
43	20	26	nCaM	44	20	27	nCaM
45	21	31	nCaM	46	22	26	nCaM
47	22	28	nCaM	48	22	31	nCaM
49	22	62	nCaM	50	24	60	nCaM
51	24	62	nCaM	52	26	62	nCaM
53	26	63	nCaM	54	26	64	nCaM
55	27	31	nCaM	56	27	32	nCaM
57	27	35	nCaM	58	27	62	nCaM
59	27	63	nCaM	60	27	68	nCaM
61	28	32	nCaM	62	28	62	nCaM
63	29	49	nCaM	64	29	52	nCaM
65	29	56	nCaM	66	29	62	nCaM
67	30	45	nCaM	68	30	49	nCaM
69	31	35	nCaM	70	32	36	nCaM
71	32	48	nCaM	72	32	52	nCaM
73	32	63	nCaM	74	32	68	nCaM
75	32	71	nCaM	76	34	38	nCaM
77	35	39	nCaM	78	35	68	nCaM
79	36	41	nCaM	80	36	42	nCaM
81	36	43	nCaM	82	36	48	nCaM
83	36	68	nCaM	84	36	71	nCaM
85	36	72	nCaM	86	36	75	nCaM

87	37	41	nCaM	88	37	42	nCaM
89	39	68	nCaM	90	39	72	nCaM
91	41	72	nCaM	92	41	75	nCaM
93	43	47	nCaM	94	43	48	nCaM
95	43	75	nCaM	96	44	48	nCaM
97	45	49	nCaM	98	46	50	nCaM
99	47	51	nCaM	100	47	75	nCaM
101	48	52	nCaM	102	48	71	nCaM
103	48	75	nCaM	104	49	53	nCaM
105	49	56	nCaM	106	50	54	nCaM
107	51	55	nCaM	108	51	71	nCaM
109	51	74	nCaM	110	51	75	nCaM
111	52	56	nCaM	112	52	63	nCaM
113	52	71	nCaM	114	53	57	nCaM
115	54	71	nCaM	116	54	74	nCaM
117	55	63	nCaM	118	55	67	nCaM
119	55	71	nCaM	120	56	60	nCaM
121	58	67	nCaM	122	63	67	nCaM
123	63	68	nCaM	124	63	71	nCaM
125	64	68	nCaM	126	65	69	nCaM
127	67	71	nCaM	128	67	74	nCaM
129	68	72	nCaM	130	69	73	nCaM
131	70	74	nCaM	132	71	75	nCaM
133	1	114	nCaM:cCaM	134	1	143	nCaM:cCaM
135	1	147	nCaM:cCaM	136	1	148	nCaM:cCaM
137	1	81	nCaM:cCaM	138	1	85	nCaM:cCaM
139	2	124	nCaM:cCaM	140	2	146	nCaM:cCaM
141	2	147	nCaM:cCaM	142	2	148	nCaM:cCaM
143	2	76	nCaM:cCaM	144	2	77	nCaM:cCaM
145	2	80	nCaM:cCaM	146	2	81	nCaM:cCaM
147	2	82	nCaM:cCaM	148	2	84	nCaM:cCaM
149	3	109	nCaM:cCaM	150	3	114	nCaM:cCaM
151	3	115	nCaM:cCaM	152	3	116	nCaM:cCaM
153	3	124	nCaM:cCaM	154	3	145	nCaM:cCaM
155	3	147	nCaM:cCaM	156	3	148	nCaM:cCaM
157	4	147	nCaM:cCaM	158	4	76	nCaM:cCaM
159	5	147	nCaM:cCaM	160	5	148	nCaM:cCaM
161	6	115	nCaM:cCaM	162	7	114	nCaM:cCaM
163	7	115	nCaM:cCaM	164	7	116	nCaM:cCaM

165	7	120	nCaM:cCaM	166	7	124	nCaM:cCaM
167	7	148	nCaM:cCaM	168	8	115	nCaM:cCaM
169	8	120	nCaM:cCaM	170	8	147	nCaM:cCaM
171	8	76	nCaM:cCaM	172	8	81	nCaM:cCaM
173	8	84	nCaM:cCaM	174	11	115	nCaM:cCaM
175	11	120	nCaM:cCaM	176	11	76	nCaM:cCaM
177	12	76	nCaM:cCaM	178	14	115	nCaM:cCaM
179	39	76	nCaM:cCaM	180	41	76	nCaM:cCaM
181	41	79	nCaM:cCaM	182	41	84	nCaM:cCaM
183	66	143	nCaM:cCaM	184	67	139	nCaM:cCaM
185	67	143	nCaM:cCaM	186	69	76	nCaM:cCaM
187	72	76	nCaM:cCaM	188	73	77	nCaM:cCaM
189	73	80	nCaM:cCaM	190	74	80	nCaM:cCaM
191	75	79	nCaM:cCaM	192	76	80	cCaM
193	76	81	cCaM	194	76	84	cCaM
195	77	138	cCaM	196	77	82	cCaM
197	78	83	cCaM	198	79	83	cCaM
199	79	84	cCaM	200	80	146	cCaM
201	81	145	cCaM	202	81	146	cCaM
203	81	85	cCaM	204	82	138	cCaM
205	82	139	cCaM	206	82	142	cCaM
207	82	143	cCaM	208	82	146	cCaM
209	82	86	cCaM	210	82	89	cCaM
211	83	87	cCaM	212	84	88	cCaM
213	85	112	cCaM	214	85	114	cCaM
215	85	142	cCaM	216	85	145	cCaM
217	85	146	cCaM	218	85	89	cCaM
219	86	138	cCaM	220	86	139	cCaM
221	86	142	cCaM	222	86	143	cCaM
223	86	90	cCaM	224	87	91	cCaM
225	88	112	cCaM	226	88	92	cCaM
227	89	105	cCaM	228	89	108	cCaM
229	89	109	cCaM	230	89	112	cCaM
231	89	114	cCaM	232	89	138	cCaM
233	89	141	cCaM	234	89	142	cCaM
235	89	145	cCaM	236	89	93	cCaM
237	89	100	cCaM	238	90	138	cCaM
239	90	94	cCaM	240	92	104	cCaM
241	92	107	cCaM	242	92	108	cCaM

243	92	111	cCaM	244	92	112	cCaM
245	92	100	cCaM	246	93	104	cCaM
247	93	138	cCaM	248	93	97	cCaM
249	93	99	cCaM	250	93	100	cCaM
251	94	104	cCaM	252	97	135	cCaM
253	99	135	cCaM	254	99	136	cCaM
255	99	137	cCaM	256	99	138	cCaM
257	100	104	cCaM	258	100	105	cCaM
259	100	108	cCaM	260	100	135	cCaM
261	100	136	cCaM	262	100	138	cCaM
263	101	105	cCaM	264	101	135	cCaM
265	102	106	cCaM	266	102	130	cCaM
267	102	131	cCaM	268	102	135	cCaM
269	102	136	cCaM	270	103	107	cCaM
271	104	108	cCaM	272	104	111	cCaM
273	105	109	cCaM	274	105	130	cCaM
275	105	136	cCaM	276	105	138	cCaM
277	105	141	cCaM	278	106	110	cCaM
279	106	118	cCaM	280	106	121	cCaM
281	106	122	cCaM	282	106	125	cCaM
283	106	126	cCaM	284	106	130	cCaM
285	106	131	cCaM	286	107	111	cCaM
287	108	112	cCaM	288	109	114	cCaM
289	109	116	cCaM	290	109	121	cCaM
291	109	124	cCaM	292	109	125	cCaM
293	109	130	cCaM	294	109	141	cCaM
295	109	145	cCaM	296	110	114	cCaM
297	110	118	cCaM	298	114	145	cCaM
299	116	120	cCaM	300	116	121	cCaM
301	116	124	cCaM	302	116	145	cCaM
303	117	121	cCaM	304	118	122	cCaM
305	120	124	cCaM	306	121	125	cCaM
307	122	126	cCaM	308	123	127	cCaM
309	123	144	cCaM	310	123	148	cCaM
311	124	128	cCaM	312	124	141	cCaM
313	124	144	cCaM	314	124	145	cCaM
315	124	148	cCaM	316	125	129	cCaM
317	125	130	cCaM	318	125	141	cCaM
319	126	131	cCaM	320	127	140	cCaM

321	127	144	cCaM	322	127	148	cCaM
323	128	140	cCaM	324	128	141	cCaM
325	128	144	cCaM	326	129	140	cCaM
327	130	135	cCaM	328	130	136	cCaM
329	130	141	cCaM	330	136	141	cCaM
331	137	141	cCaM	332	138	142	cCaM
333	139	143	cCaM	334	140	144	cCaM
335	141	145	cCaM	336	142	146	cCaM
337	143	147	cCaM	338	143	148	cCaM
339	144	148	cCaM	340	1	16	nCaM:Ng
341	1	18	nCaM:Ng	342	2	13	nCaM:Ng
343	2	18	nCaM:Ng	344	2	22	nCaM:Ng
345	2	40	nCaM:Ng	346	3	15	nCaM:Ng
347	3	16	nCaM:Ng	348	3	17	nCaM:Ng
349	3	18	nCaM:Ng	350	3	20	nCaM:Ng
351	3	41	nCaM:Ng	352	3	43	nCaM:Ng
353	3	45	nCaM:Ng	354	4	16	nCaM:Ng
355	4	21	nCaM:Ng	356	4	41	nCaM:Ng
357	6	15	nCaM:Ng	358	7	44	nCaM:Ng
359	8	32	nCaM:Ng	360	8	45	nCaM:Ng
361	9	17	nCaM:Ng	362	9	41	nCaM:Ng
363	11	18	nCaM:Ng	364	11	32	nCaM:Ng
365	11	44	nCaM:Ng	366	12	17	nCaM:Ng
367	12	18	nCaM:Ng	368	12	19	nCaM:Ng
369	12	21	nCaM:Ng	370	12	41	nCaM:Ng
371	13	17	nCaM:Ng	372	15	16	nCaM:Ng
373	15	17	nCaM:Ng	374	16	17	nCaM:Ng
375	16	19	nCaM:Ng	376	18	16	nCaM:Ng
377	19	16	nCaM:Ng	378	26	19	nCaM:Ng
379	26	20	nCaM:Ng	380	27	21	nCaM:Ng
381	29	22	nCaM:Ng	382	32	21	nCaM:Ng
383	35	17	nCaM:Ng	384	38	15	nCaM:Ng
385	38	16	nCaM:Ng	386	38	17	nCaM:Ng
387	39	15	nCaM:Ng	388	39	16	nCaM:Ng
389	39	17	nCaM:Ng	390	39	18	nCaM:Ng
391	39	19	nCaM:Ng	392	41	34	nCaM:Ng
393	62	20	nCaM:Ng	394	62	21	nCaM:Ng
395	62	22	nCaM:Ng	396	63	20	nCaM:Ng
397	63	21	nCaM:Ng	398	65	17	nCaM:Ng

399	65	19	nCaM:Ng	400	65	21	nCaM:Ng
401	65	41	nCaM:Ng	402	65	46	nCaM:Ng
403	67	38	nCaM:Ng	404	67	44	nCaM:Ng
405	67	47	nCaM:Ng	406	68	17	nCaM:Ng
407	68	19	nCaM:Ng	408	68	41	nCaM:Ng
409	69	19	nCaM:Ng	410	69	20	nCaM:Ng
411	69	21	nCaM:Ng	412	69	41	nCaM:Ng
413	70	38	nCaM:Ng	414	72	19	nCaM:Ng
415	72	41	nCaM:Ng	416	74	37	nCaM:Ng
417	74	38	nCaM:Ng	418	74	43	nCaM:Ng
419	74	47	nCaM:Ng	420	76	20	cCaM:Ng
421	76	41	cCaM:Ng	422	76	43	cCaM:Ng
423	77	43	cCaM:Ng	424	77	44	cCaM:Ng
425	80	43	cCaM:Ng	426	81	27	cCaM:Ng
427	81	43	cCaM:Ng	428	82	18	cCaM:Ng
429	82	32	cCaM:Ng	430	82	37	cCaM:Ng
431	82	38	cCaM:Ng	432	82	40	cCaM:Ng
433	82	41	cCaM:Ng	434	82	43	cCaM:Ng
435	82	44	cCaM:Ng	436	84	33	cCaM:Ng
437	84	34	cCaM:Ng	438	84	38	cCaM:Ng
439	85	17	cCaM:Ng	440	85	18	cCaM:Ng
441	85	33	cCaM:Ng	442	86	23	cCaM:Ng
443	86	38	cCaM:Ng	444	88	33	cCaM:Ng
445	89	17	cCaM:Ng	446	89	18	cCaM:Ng
447	89	19	cCaM:Ng	448	89	21	cCaM:Ng
449	89	23	cCaM:Ng	450	89	37	cCaM:Ng
451	99	21	cCaM:Ng	452	99	22	cCaM:Ng
453	99	23	cCaM:Ng	454	99	24	cCaM:Ng
455	99	34	cCaM:Ng	456	99	38	cCaM:Ng
457	100	21	cCaM:Ng	458	105	17	cCaM:Ng
459	105	19	cCaM:Ng	460	105	21	cCaM:Ng
461	106	21	cCaM:Ng	462	109	17	cCaM:Ng
463	109	19	cCaM:Ng	464	109	21	cCaM:Ng
465	112	17	cCaM:Ng	466	112	33	cCaM:Ng
467	112	37	cCaM:Ng	468	114	15	cCaM:Ng
469	114	17	cCaM:Ng	470	114	32	cCaM:Ng
471	114	33	cCaM:Ng	472	114	36	cCaM:Ng
473	114	41	cCaM:Ng	474	114	43	cCaM:Ng
475	114	44	cCaM:Ng	476	115	18	cCaM:Ng

477	115	23	cCaM:Ng	478	115	32	cCaM:Ng
479	115	36	cCaM:Ng	480	115	40	cCaM:Ng
481	115	42	cCaM:Ng	482	115	43	cCaM:Ng
483	115	44	cCaM:Ng	484	115	45	cCaM:Ng
485	116	32	cCaM:Ng	486	120	32	cCaM:Ng
487	120	43	cCaM:Ng	488	120	44	cCaM:Ng
489	120	45	cCaM:Ng	490	124	32	cCaM:Ng
491	125	21	cCaM:Ng	492	130	21	cCaM:Ng
493	135	20	cCaM:Ng	494	135	22	cCaM:Ng
495	135	36	cCaM:Ng	496	136	20	cCaM:Ng
497	137	20	cCaM:Ng	498	138	18	cCaM:Ng
499	138	19	cCaM:Ng	500	138	20	cCaM:Ng
501	138	22	cCaM:Ng	502	138	23	cCaM:Ng
503	138	27	cCaM:Ng	504	138	32	cCaM:Ng
505	138	33	cCaM:Ng	506	138	36	cCaM:Ng
507	138	40	cCaM:Ng	508	138	43	cCaM:Ng
509	138	44	cCaM:Ng	510	141	17	cCaM:Ng
511	141	19	cCaM:Ng	512	141	21	cCaM:Ng
513	142	16	cCaM:Ng	514	142	17	cCaM:Ng
515	142	18	cCaM:Ng	516	145	16	cCaM:Ng
517	145	17	cCaM:Ng	518	145	32	cCaM:Ng
519	145	33	cCaM:Ng	520	145	41	cCaM:Ng
521	145	43	cCaM:Ng	522	145	44	cCaM:Ng
523	146	16	cCaM:Ng	524	146	33	cCaM:Ng
525	146	35	cCaM:Ng	526	146	36	cCaM:Ng
527	146	42	cCaM:Ng	528	147	47	cCaM:Ng
529	18	32	Ng	530	20	26	Ng
531	20	27	Ng	532	20	28	Ng
533	22	27	Ng	534	23	27	Ng
535	27	31	Ng	536	27	32	Ng
537	28	32	Ng	538	29	33	Ng
539	30	34	Ng	540	30	36	Ng
541	32	36	Ng	542	32	37	Ng
543	33	37	Ng	544	34	38	Ng
545	36	40	Ng	546	37	41	Ng
547	38	43	Ng	548	40	44	Ng
549	41	45	Ng	550	41	46	Ng
551	43	47	Ng				

Table S9 Charge distribution on a coarse-grained side-chain C_α model of full length Ng.

Residue Index	Residue Name	Charge on C_α	Error of the C_α charge	Charge on side-chain	Error of the side-chain charge
1	MET	0.847	0.002	0.091	0.001
2	ASP	-0.028	0.002	-0.898	0.001
3	CYS	-0.037	0.002	0.035	0.001
4	CYS	-0.037	0.002	0.033	0.001
5	THR	-0.098	0.002	0.083	0.001
6	GLU	-0.052	0.002	-0.872	0.002
7	SER	-0.098	0.002	0.066	0.001
8	ALA	-0.022	0.002	0.049	0.001
9	CYS	-0.042	0.002	0.025	0.001
10	SER	-0.077	0.002	0.062	0.001
11	LYS	-0.061	0.003	1.000	0.001
12	PRO	-0.307	0.003	0.316	0.002
13	ASP	-0.010	0.002	-0.926	0.002
14	ASP	-0.037	0.002	-0.927	0.002
15	ASP	-0.040	0.002	-0.930	0.002
16	ILE	-0.070	0.002	0.078	0.001
17	LEU	-0.043	0.002	0.061	0.001
18	ASP	-0.053	0.002	-0.925	0.002
19	ILE	-0.073	0.003	0.080	0.001
20	PRO	-0.301	0.003	0.305	0.002
21	LEU	-0.066	0.002	0.063	0.001
22	ASP	-0.034	0.002	-0.930	0.002
23	ASP	-0.035	0.003	-0.929	0.002
24	PRO	-0.335	0.003	0.320	0.002
25	GLY	-0.182	0.002	0.188	0.001
26	ALA	-0.041	0.002	0.063	0.001
27	ASN	-0.047	0.003	0.011	0.001
28	ALA	-0.043	0.002	0.067	0.001
29	ALA	-0.064	0.002	0.064	0.001
30	ALA	-0.068	0.002	0.068	0.001
31	ALA	-0.057	0.002	0.069	0.001
32	LYS	-0.079	0.002	1.020	0.002
33	ILE	-0.089	0.002	0.103	0.001
34	GLN	-0.077	0.002	0.069	0.001
35	ALA	-0.046	0.002	0.064	0.001
36	SER	-0.091	0.002	0.079	0.001
37	PHE	-0.052	0.002	0.062	0.001
38	ARG	-0.079	0.002	1.027	0.002
39	GLY	-0.181	0.002	0.192	0.001
40	HIS	-0.049	0.002	0.049	0.001
41	MET	-0.069	0.002	0.069	0.001
42	ALA	-0.062	0.002	0.068	0.001
43	ARG	-0.062	0.002	1.020	0.002
44	LYS	-0.074	0.002	1.031	0.002

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

The unit of charge is $1.6*10^{-19}$ C.

Table S10 Charge distribution on a coarse-grained side-chain C_α model of apoCaM at pH = 7.2, I = 0.15 M.

Residue Index	Residue Name	Charge on C_α	Error of the C_α charge	Charge on side-chain	Error of the side-chain charge
1	ALA	1.564	0.004	0.293	0.002
2	ASP	-0.038	0.003	-0.889	0.002
3	GLN	-0.050	0.003	0.051	0.001
4	LEU	-0.068	0.002	0.081	0.001
5	THR	-0.093	0.002	0.073	0.001
6	GLU	-0.051	0.002	-0.915	0.002

7	GLU	-0.077	0.003	-0.872	0.002
8	GLN	-0.033	0.003	0.052	0.001
9	ILE	-0.104	0.002	0.088	0.001
10	ALA	-0.063	0.003	0.059	0.001
11	GLU	-0.059	0.003	-0.865	0.002
12	PHE	-0.049	0.003	0.049	0.002
13	LYS	-0.087	0.003	1.001	0.001
14	GLU	-0.057	0.002	-0.893	0.002
15	ALA	-0.030	0.002	0.059	0.001
16	PHE	-0.063	0.003	0.050	0.001
17	SER	-0.109	0.003	0.070	0.001
18	LEU	-0.062	0.002	0.078	0.001
19	PHE	-0.060	0.003	0.048	0.001
20	ASP	-0.043	0.003	-0.885	0.002
21	LYS	-0.083	0.002	0.986	0.001
22	ASP	-0.049	0.002	-0.914	0.002
23	GLY	-0.195	0.002	0.190	0.000
24	ASP	-0.062	0.002	-0.948	0.001
25	GLY	-0.152	0.002	0.186	0.001
26	THR	-0.064	0.003	0.064	0.001
27	ILE	-0.060	0.003	0.076	0.001
28	THR	-0.110	0.003	0.077	0.001
29	THR	-0.094	0.003	0.087	0.001
30	LYS	-0.070	0.002	0.998	0.002
31	GLU	-0.059	0.003	-0.866	0.001
32	LEU	-0.081	0.003	0.082	0.001
33	GLY	-0.181	0.003	0.191	0.001
34	THR	-0.108	0.003	0.083	0.001
35	VAL	-0.080	0.003	0.105	0.001
36	MET	-0.068	0.003	0.061	0.001
37	ARG	-0.074	0.003	1.003	0.002
38	SER	-0.090	0.003	0.075	0.001
39	LEU	-0.076	0.003	0.083	0.001
40	GLY	-0.194	0.003	0.188	0.001
41	GLN	-0.034	0.002	0.054	0.001
42	ASN	-0.789	0.003	0.744	0.002
43	PRO	-0.249	0.003	0.267	0.003
44	THR	-0.090	0.002	0.071	0.001
45	GLU	-0.057	0.002	-0.902	0.002
46	ALA	-0.051	0.002	0.047	0.001
47	GLU	-0.050	0.002	-0.880	0.002
48	LEU	-0.066	0.002	0.070	0.001
49	GLN	-0.067	0.002	0.043	0.001

50	ASP	-0.032	0.003	-0.947	0.001
51	MET	-0.036	0.003	0.053	0.001
52	ILE	-0.085	0.003	0.097	0.001
53	ASN	-0.038	0.002	-0.005	0.001
54	GLU	-0.058	0.002	-0.902	0.002
55	VAL	-0.073	0.003	0.094	0.001
56	ASP	-0.066	0.002	-0.913	0.001
57	ALA	-0.063	0.002	0.055	0.001
58	ASP	-0.056	0.002	-0.940	0.002
59	GLY	-0.188	0.002	0.192	0.000
60	ASN	-0.052	0.002	0.005	0.001
61	GLY	-0.160	0.003	0.188	0.001
62	THR	-0.083	0.003	0.074	0.001
63	ILE	-0.080	0.003	0.082	0.001
64	ASP	-0.024	0.002	-0.936	0.006
65	PHE	-0.742	0.004	0.744	0.008
66	PRO	-0.256	0.004	0.256	0.005
67	GLU	-0.051	0.003	-0.878	0.002
68	PHE	-0.037	0.003	0.054	0.008
69	LEU	-0.070	0.004	0.077	0.001
70	THR	-0.103	0.003	0.073	0.001
71	MET	-0.056	0.003	0.069	0.006
72	MET	-0.067	0.003	0.066	0.004
73	ALA	-0.034	0.003	0.055	0.001
74	ARG	-0.079	0.002	0.980	0.002
75	LYS	-0.056	0.003	0.994	0.001
76	MET	-0.051	0.003	0.071	0.006
77	LYS	-0.055	0.002	0.979	0.001
78	ASP	-0.046	0.002	-0.940	0.001
79	THR	-0.086	0.002	0.080	0.001
80	ASP	-0.012	0.003	-0.912	0.002
81	SER	-0.088	0.002	0.065	0.001
82	GLU	-0.064	0.002	-0.862	0.002
83	GLU	-0.068	0.002	-0.869	0.002
84	GLU	-0.061	0.002	-0.875	0.002
85	ILE	-0.093	0.003	0.094	0.001
86	ARG	-0.064	0.003	0.944	0.002
87	GLU	-0.073	0.003	-0.862	0.002
88	ALA	-0.043	0.003	0.061	0.001
89	PHE	-0.054	0.003	0.049	0.001
90	ARG	-0.072	0.002	0.974	0.002
91	VAL	-0.084	0.002	0.097	0.001
92	PHE	-0.075	0.003	0.054	0.001

93	ASP	-0.025	0.003	-0.892	0.001
94	LYS	-0.082	0.002	0.995	0.001
95	ASP	-0.055	0.002	-0.941	0.002
96	GLY	-0.193	0.002	0.195	0.001
97	ASN	-0.041	0.002	0.000	0.001
98	GLY	-0.165	0.002	0.189	0.001
99	TYR	-0.036	0.002	0.031	0.001
100	ILE	-0.074	0.002	0.094	0.001
101	SER	-0.097	0.003	0.060	0.001
102	ALA	-0.061	0.002	0.066	0.001
103	ALA	-0.044	0.002	0.052	0.001
104	GLU	-0.049	0.003	-0.874	0.001
105	LEU	-0.095	0.003	0.076	0.001
106	ARG	-0.056	0.003	0.995	0.002
107	HIS	-0.055	0.003	0.028	0.001
108	VAL	-0.080	0.003	0.094	0.001
109	MET	-0.053	0.003	0.060	0.001
110	THR	-0.088	0.003	0.081	0.001
111	ASN	-0.039	0.003	0.000	0.001
112	LEU	-0.073	0.003	0.081	0.001
113	GLY	-0.194	0.002	0.188	0.001
114	GLU	-0.061	0.002	-0.871	0.002
115	LYS	-0.080	0.002	0.988	0.001
116	LEU	-0.060	0.002	0.072	0.001
117	THR	-0.091	0.003	0.070	0.001
118	ASP	-0.025	0.002	-0.921	0.002
119	GLU	-0.067	0.002	-0.915	0.002
120	GLU	-0.060	0.002	-0.884	0.002
121	VAL	-0.084	0.002	0.086	0.001
122	ASP	-0.030	0.002	-0.903	0.001
123	GLU	-0.058	0.002	-0.888	0.002
124	MET	-0.049	0.002	0.053	0.001
125	ILE	-0.087	0.002	0.087	0.001
126	ARG	-0.060	0.002	0.949	0.002
127	GLU	-0.079	0.002	-0.913	0.002
128	ALA	-0.051	0.002	0.054	0.001
129	ASP	-0.019	0.002	-0.952	0.001
130	ILE	-0.077	0.002	0.077	0.001
131	ASP	-0.038	0.002	-0.950	0.002
132	GLY	-0.210	0.002	0.188	0.001
133	ASP	-0.037	0.002	-0.955	0.001
134	GLY	-0.174	0.002	0.190	0.001
135	GLN	-0.037	0.002	0.029	0.001

136	VAL	-0.078	0.002	0.097	0.001
137	ASN	-0.023	0.002	0.001	0.001
138	TYR	-0.051	0.003	0.033	0.001
139	GLU	-0.071	0.003	-0.893	0.002
140	GLU	-0.052	0.002	-0.893	0.002
141	PHE	-0.061	0.003	0.055	0.003
142	VAL	-0.091	0.003	0.093	0.001
143	GLN	-0.058	0.003	0.044	0.001
144	MET	-0.045	0.002	0.059	0.001
145	MET	-0.053	0.002	0.056	0.003
146	THR	-0.118	0.002	0.076	0.001
147	ALA	-0.050	0.003	0.042	0.001
148	LYS	-2.693	0.005	2.669	0.003

The unit of charge is $1.6*10^{-19}$ C.

Table S11 Charge distribution on a coarse-grained side chain C_α model of Ng_{13-49} at pH = 7.2, I = 0.15 M.

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

40	HIS	-0.054	0.002	0.037	0.001
41	MET	-0.068	0.002	0.065	0.001
42	ALA	-0.057	0.003	0.064	0.001
43	ARG	-0.059	0.003	0.984	0.002
44	LYS	-0.061	0.003	0.997	0.002
45	LYS	-0.084	0.002	1.000	0.002
46	ILE	-0.089	0.003	0.106	0.001
47	LYS	-0.075	0.003	1.001	0.002
48	SER	-0.132	0.003	0.066	0.001
49	GLY	-1.086	0.002	0.174	0.001

The unit of charge is $1.6*10^{-19}$ C.

Table S12 Charge distribution on a coarse-grained side chain C_α model of apoCaM at pH = 6.3, I = 0.1 M.

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

29	THR	-0.095	0.003	0.089	0.001
30	LYS	-0.072	0.003	0.995	0.001
31	GLU	-0.058	0.003	-0.866	0.002
32	LEU	-0.079	0.003	0.082	0.001
33	GLY	-0.180	0.003	0.191	0.001
34	THR	-0.108	0.003	0.081	0.001
35	VAL	-0.082	0.003	0.108	0.001
36	MET	-0.066	0.003	0.061	0.001
37	ARG	-0.080	0.002	1.006	0.002
38	SER	-0.088	0.003	0.072	0.001
39	LEU	-0.077	0.003	0.085	0.001
40	GLY	-0.192	0.003	0.188	0.001
41	GLN	-0.033	0.002	0.054	0.001
42	ASN	-0.782	0.004	0.744	0.002
43	PRO	-0.254	0.004	0.265	0.002
44	THR	-0.092	0.002	0.070	0.001
45	GLU	-0.052	0.002	-0.904	0.002
46	ALA	-0.053	0.002	0.049	0.001
47	GLU	-0.052	0.002	-0.883	0.002
48	LEU	-0.064	0.002	0.069	0.001
49	GLN	-0.069	0.002	0.044	0.001
50	ASP	-0.031	0.002	-0.947	0.001
51	MET	-0.037	0.003	0.054	0.001
52	ILE	-0.084	0.003	0.096	0.001
53	ASN	-0.038	0.002	-0.006	0.001
54	GLU	-0.052	0.002	-0.902	0.002
55	VAL	-0.077	0.003	0.093	0.001
56	ASP	-0.062	0.002	-0.915	0.001
57	ALA	-0.063	0.002	0.053	0.001
58	ASP	-0.056	0.002	-0.940	0.001
59	GLY	-0.187	0.002	0.192	0.000
60	ASN	-0.055	0.002	0.004	0.001
61	GLY	-0.162	0.002	0.187	0.001
62	THR	-0.075	0.003	0.073	0.001
63	ILE	-0.083	0.003	0.079	0.001
64	ASP	-0.017	0.003	-0.945	0.001
65	PHE	-0.740	0.004	0.732	0.002
66	PRO	-0.260	0.003	0.251	0.002
67	GLU	-0.055	0.003	-0.880	0.002
68	PHE	-0.045	0.003	0.048	0.007
69	LEU	-0.061	0.002	0.075	0.001
70	THR	-0.103	0.002	0.072	0.001
71	MET	-0.053	0.003	0.061	0.001

72	MET	-0.066	0.003	0.066	0.006
73	ALA	-0.040	0.002	0.057	0.001
74	ARG	-0.073	0.002	0.978	0.002
75	LYS	-0.063	0.002	0.994	0.001
76	MET	-0.051	0.003	0.066	0.001
77	LYS	-0.056	0.003	0.980	0.001
78	ASP	-0.045	0.002	-0.939	0.001
79	THR	-0.083	0.002	0.079	0.001
80	ASP	-0.018	0.003	-0.912	0.001
81	SER	-0.088	0.002	0.067	0.001
82	GLU	-0.062	0.002	-0.865	0.002
83	GLU	-0.065	0.002	-0.869	0.002
84	GLU	-0.066	0.003	-0.875	0.002
85	ILE	-0.093	0.003	0.094	0.001
86	ARG	-0.067	0.002	0.945	0.002
87	GLU	-0.074	0.003	-0.859	0.002
88	ALA	-0.040	0.003	0.061	0.001
89	PHE	-0.055	0.003	0.049	0.001
90	ARG	-0.076	0.003	0.974	0.002
91	VAL	-0.081	0.003	0.097	0.001
92	PHE	-0.074	0.003	0.053	0.003
93	ASP	-0.025	0.003	-0.893	0.002
94	LYS	-0.078	0.002	0.999	0.001
95	ASP	-0.055	0.002	-0.940	0.001
96	GLY	-0.189	0.002	0.194	0.001
97	ASN	-0.044	0.002	0.000	0.001
98	GLY	-0.164	0.002	0.189	0.001
99	TYR	-0.031	0.002	0.032	0.001
100	ILE	-0.075	0.002	0.092	0.001
101	SER	-0.103	0.002	0.063	0.001
102	ALA	-0.062	0.002	0.070	0.001
103	ALA	-0.045	0.002	0.054	0.001
104	GLU	-0.050	0.003	-0.870	0.001
105	LEU	-0.096	0.003	0.080	0.001
106	ARG	-0.053	0.003	0.995	0.002
107	HIS	-0.048	0.003	0.973	0.002
108	VAL	-0.087	0.003	0.098	0.001
109	MET	-0.053	0.003	0.064	0.001
110	THR	-0.087	0.003	0.082	0.001
111	ASN	-0.043	0.003	0.002	0.001
112	LEU	-0.068	0.003	0.085	0.001
113	GLY	-0.193	0.002	0.187	0.001
114	GLU	-0.062	0.002	-0.870	0.002

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

The unit of charge is $1.6*10^{-19}$ C.

Table S13 Charge distribution on a coarse-grained side chain C_α model of Ng_{13-49}. pH = 6.3, I = 0.1 M.

Residue Index	Residue Name	Charge on C_α	Error of the C_α charge	Charge on the side-chain	Error of the side-chain charge
13	ASP	1.570	0.004	-1.657	0.003
14	ASP	-0.017	0.003	-0.918	0.002
15	ASP	-0.035	0.003	-0.909	0.002

16	ILE	-0.075	0.003	0.083	0.001
17	LEU	-0.059	0.003	0.063	0.001
18	ASP	-0.033	0.003	-0.918	0.002
19	ILE	-0.539	0.004	0.541	0.002
20	PRO	-0.269	0.003	0.257	0.002
21	LEU	-0.055	0.003	0.065	0.001
22	ASP	-0.039	0.002	-0.932	0.002
23	ASP	-0.813	0.004	-0.137	0.003
24	PRO	-0.286	0.003	0.265	0.002
25	GLY	-0.179	0.003	0.186	0.001
26	ALA	-0.031	0.002	0.056	0.001
27	ASN	-0.049	0.003	0.007	0.001
28	ALA	-0.045	0.003	0.063	0.001
29	ALA	-0.051	0.003	0.059	0.001
30	ALA	-0.070	0.003	0.063	0.001
31	ALA	-0.048	0.003	0.062	0.001
32	LYS	-0.084	0.003	0.998	0.001
33	ILE	-0.085	0.003	0.100	0.001
34	GLN	-0.074	0.002	0.060	0.001
35	ALA	-0.041	0.002	0.063	0.001
36	SER	-0.090	0.003	0.075	0.001
37	PHE	-0.050	0.003	0.057	0.001
38	ARG	-0.080	0.002	1.001	0.002
39	GLY	-0.174	0.003	0.189	0.001
40	HIS	-0.056	0.002	0.039	0.001
41	MET	-0.069	0.002	0.065	0.001
42	ALA	-0.062	0.002	0.067	0.001
43	ARG	-0.057	0.003	0.987	0.002
44	LYS	-0.064	0.003	0.996	0.002
45	LYS	-0.081	0.002	1.000	0.002
46	ILE	-0.091	0.003	0.105	0.001
47	LYS	-0.073	0.003	0.999	0.002
48	SER	-0.131	0.003	0.065	0.001
49	GLY	-1.086	0.002	0.176	0.001

The unit of charge is $1.6*10^{-19}$ C.

Table S14. Detail of the neurogranin (Ng) sequences obtained from UniprotKB (www.uniprot.org). 45 sequences of Ng with unique Uniprot ID were used for the sequence alignment (see Fig. S15).

Table S14(a)

Entry	Entry name	Status	Protein names
R0LF54	R0LF54_ANAPL	unreviewed	Neurogranin (Fragment)
A0A091VHN8	A0A091VHN8_NIPNI	unreviewed	Neurogranin (Fragment)
A0A091GKM6	A0A091GKM6_9AVES	unreviewed	Neurogranin (Fragment)
C1BLT4	C1BLT4_OSMMO	unreviewed	Neurogranin
A0A091IY17	A0A091IY17_CALAN	unreviewed	Neurogranin (Fragment)
A0A093QNN0	A0A093QNN0_9PASS	unreviewed	Neurogranin (Fragment)
A0A151NG69	A0A151NG69_ALLMI	unreviewed	Neurogranin
A0A099ZXB1	A0A099ZXB1_TINGU	unreviewed	Neurogranin (Fragment)
A0A091HFZ3	A0A091HFZ3_BUCRH	unreviewed	Neurogranin (Fragment)
A0A091JUZ1	A0A091JUZ1_COLST	unreviewed	Neurogranin (Fragment)
A0A093C189	A0A093C189_TAUER	unreviewed	Neurogranin (Fragment)
A0A087QQ10	A0A087QQ10_APTFO	unreviewed	Neurogranin (Fragment)
A0A091PSL5	A0A091PSL5_HALAL	unreviewed	Neurogranin (Fragment)
B5X303	B5X303_SALSA	unreviewed	Neurogranin
A0A091SXC0	A0A091SXC0_NESNO	unreviewed	Neurogranin (Fragment)
V8NXH4	V8NXH4_OPHHA	unreviewed	Neurogranin (Fragment)
Q90X80	Q90X80_GILMI	unreviewed	Neurogranin
A0A0A0AIY4	A0A0A0AIY4_CHAVO	unreviewed	Neurogranin (Fragment)
A0A093IXQ6	A0A093IXQ6_PICPB	unreviewed	Neurogranin (Fragment)
B7TWQ3	B7TWQ3_DANRE	unreviewed	Neurogranin (Fragment)
C1J0K5	C1J0K5_GILSE	unreviewed	Neurogranin (Fragment)
A0A091L3D5	A0A091L3D5_CATAU	unreviewed	Neurogranin (Fragment)
B5XEB8	B5XEB8_SALSA	unreviewed	Neurogranin
B2XBN6	B2XBN6_SALSA	unreviewed	Neurogranin
A0A091EYH6	A0A091EYH6_CORBR	unreviewed	Neurogranin (Fragment)
B9ENE5	B9ENE5_SALSA	unreviewed	Neurogranin
B9EPB8	B9EPB8_SALSA	unreviewed	Neurogranin
A0A094KAF2	A0A094KAF2_ANTCR	unreviewed	Neurogranin (Fragment)
M7AY01	M7AY01_CHEMY	unreviewed	Neurogranin (Fragment)
P54866	NEUG_SERCA	reviewed	Neurogranin (NG) (Canarigranin)

Q04940	NEUG_RAT	reviewed	Neurogranin (Ng) (Protein kinase C substrate 7.5 kDa protein) (RC3) [Cleaved into: NEUG(55-78)]
P35722	NEUG_BOVIN	reviewed	Neurogranin (NG) (B-50 immunoreactive C-kinase substrate) (BICKS) (p17) [Cleaved into: NEUG(55-78)]
P60761	NEUG_MOUSE	reviewed	Neurogranin (Ng) (RC3) [Cleaved into: NEUG(55-78)]
P54877	NEUG_CAPHI	reviewed	Neurogranin (NG) (Protein kinase C substrate 7.5 kDa protein) (RC3) [Cleaved into: NEUG(55-78)]
Q92686	NEUG_HUMAN	reviewed	Neurogranin (Ng) (RC3) [Cleaved into: NEUG(55-78)]
G3IJI8	G3IJI8_CRIGR	unreviewed	Neurogranin
A0A0A7HRI9	A0A0A7HRI9_HETGA	unreviewed	Neurogranin
U3E6M0	U3E6M0_CALJA	unreviewed	Neurogranin
F6THG0	F6THG0_MACMU	unreviewed	Neurogranin (Uncharacterized protein)
U3FBA0	U3FBA0_CALJA	unreviewed	Neurogranin
G9KDZ7	G9KDZ7_MUSPF	unreviewed	Neurogranin (Fragment)
A0A024R3M7	A0A024R3M7_HUMAN	unreviewed	Neurogranin (Protein kinase C substrate, RC3), isoform CRA_a
B5KFR8	B5KFR8_TAEGU	unreviewed	Putative neurogranin
B2CZA9	B2CZA9_COTJA	unreviewed	RC3/neurogranin (Fragment)
C7ENZ8	C7ENZ8_LARAR	unreviewed	RC3/neurogranin

Table S14(b)

Entry	Gene names	Organism	Length
R0LF54	Anapl_11925	Anas platyrhynchos (Mallard) (Anas boschas)	67
A0A091VHN8	Y956_07381	Nipponia nippon (Crested ibis) (Ibis nippon)	68
A0A091GKM6	N303_10815	Cuculus canorus (common cuckoo)	66
C1BLT4	NEUG	Osmerus mordax (Rainbow smelt) (Atherina mordax)	60
A0A091IY17	N300_07147	Calypte anna (Anna's hummingbird) (Archilochus anna)	59
A0A093QNN0	N305_02066	Manacus vitellinus (golden-collared manakin)	65
A0A151NG69	NRGN Y1Q_0010222	Alligator mississippiensis (American alligator)	109
A0A099ZXB1	N309_12965	Tinamus guttatus (White-throated tinamou)	55
A0A091HFZ3	N320_07902	Buceros rhinoceros silvestris	56
A0A091JUZ1	N325_03244	Colius striatus (Speckled mousebird)	64
A0A093C189	N340_06990	Tauraco erythrolophus (Red-crested turaco)	56
A0A087QQ10	AS27_01314	Aptenodytes forsteri (Emperor penguin)	63
A0A091PSL5	N329_06668	Haliaeetus albicilla (White-tailed sea-eagle)	47
B5X303	NEUG	Salmo salar (Atlantic salmon)	94
A0A091SXC0	N333_05461	Nestor notabilis (Kea)	66

V8NXH4	NRGN L345_07452	Ophiophagus hannah (King cobra) (Naja hannah)	78
Q90X80		Gillichthys mirabilis (Long-jawed mudsucker)	65
A0A0A0AIY4	N301_09391	Charadrius vociferus (Killdeer) (Aegialitis vocifera)	69
A0A093IXQ6	N307_15420	Picoides pubescens (Downy woodpecker) (Dryobates pubescens)	65
B7TWQ3	nrgna nrgn	Danio rerio (Zebrafish) (Brachydanio rerio)	92
C1J0K5		Gillichthys seta (Shortjaw mudsucker)	56
A0A091L3D5	N323_11621	Cathartes aura (Turkey vulture) (Vultur aura)	60
B5XEB8	NEUG	Salmo salar (Atlantic salmon)	94
B2XBN6		Salmo salar (Atlantic salmon)	60
A0A091EYH6	N302_11071	Corvus brachyrhynchos (American crow)	68
B9ENE5	NEUG	Salmo salar (Atlantic salmon)	93
B9EPB8	NEUG	Salmo salar (Atlantic salmon)	60
A0A094KAF2	N321_02863	Antrostomus carolinensis (Chuck-will's-widow) (Caprimulgus carolinensis)	66
M7AY01	UY3_13287	Chelonia mydas (Green sea-turtle) (Chelonia agassizi)	68
P54866	NRGN HAT14	Serinus canaria (Island canary) (Fringilla canaria)	73
Q04940	Nrgn	Rattus norvegicus (Rat)	78
P35722	NRGN	Bos taurus (Bovine)	78
P60761	Nrgn	Mus musculus (Mouse)	78
P54877	NRGN	Capra hircus (Goat)	78
Q92686	NRGN	Homo sapiens (Human)	78
G3IJI8	I79_024023	Cricetulus griseus (Chinese hamster) (Cricetulus barabensis griseus)	87
A0A0A7HRI9		Heterocephalus glaber (Naked mole rat)	78
U3E6M0	NRGN	Callithrix jacchus (White-tufted-ear marmoset)	78
F6THG0	NRGN	Macaca mulatta (Rhesus macaque)	78
U3FBA0	NRGN	Callithrix jacchus (White-tufted-ear marmoset)	78
G9KDZ7		Mustela putorius furo (European domestic ferret) (Mustela furo)	126
A0A024R3M7	NRGN hCG_1732008	Homo sapiens (Human)	78
B5KFR8		Taeniopygia guttata (Zebra finch) (Poephila guttata)	94
B2CZA9		Coturnix coturnix japonica (Japanese quail) (Coturnix japonica)	62
C7ENZ8		Larus argentatus (Herring gull)	73

References

1. Jurado, L. A., P. S. Chockalingam, and H. W. Jarrett. 1999. Apocalmodulin. Physiological reviews 79:661-682.
2. Bahler, M., and A. Rhoads. 2002. Calmodulin signaling via the IQ motif. FEBS Lett 513:107-113.
3. Ran, X., H. H. Miao, F. S. Sheu, and D. Yang. 2003. Structural and dynamic characterization of a neuron-specific protein kinase C substrate, neurogranin. Biochemistry 42:5143-5150.
4. Hoffman, L., A. Chandrasekar, X. Wang, J. A. Putkey, and M. N. Waxham. 2014. Neurogranin alters the structure and calcium binding properties of calmodulin. J Biol Chem 289:14644-14655.
5. Sugita, Y., and Y. Okamoto. 1999. Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 314:141-151.
6. Sanbonmatsu, K. Y., and A. E. Garcia. 2002. Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. Proteins 46:225-234.
7. Cheung, M. S., J. M. Finke, B. Callahan, and J. N. Onuchic. 2003. Exploring the interplay between topology and secondary structural formation in the protein folding problem. J. Phys. Chem. B 107:11193-11200.
8. Wang, Q., P. Zhang, L. Hoffman, S. Tripathi, D. Homouz, Y. Liu, M. N. Waxham, and M. S. Cheung. 2013. Protein recognition and selection through conformational and mutually induced fit. Proceedings of the National Academy of Sciences of the United States of America 110:20545-20550.
9. Yang, Y., E. Faraggi, H. Zhao, and Y. Zhou. 2011. Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. Bioinformatics 27:2076-2082.
10. Debye, P., and E. Hückel. 1923. The theory of electrolytes. I. Lowering of freezing point and related phenomena. Physikalische Zeitschrift 24:185-206.
11. Betancourt, M. R., and D. Thirumalai. 1999. Pair potentials for protein folding: Choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. Protein Sci. 8:361-369.
12. Karanicolas, J., and C. L. Brooks, 3rd. 2002. The origins of asymmetry in the folding transition states of protein L and protein G. Protein Sci 11:2351-2361.
13. Andreasen, T. J., C. W. Luetje, W. Heideman, and D. R. Storm. 1983. Purification of a novel calmodulin binding protein from bovine cerebral cortex membranes. Biochemistry 22:4615-4618.
14. Wang, Q., K. C. Liang, A. Czader, M. N. Waxham, and M. S. Cheung. 2011. The Effect of Macromolecular Crowding, Ionic Strength and Calcium Binding on Calmodulin Dynamics. PLoS Comput. Biol. 7:16.
15. Dima, R. I., and D. Thirumalai. 2004. Asymmetry in the shapes of folded and denatured states of proteins. J. Phys. Chem. B 108:6564-6570.
16. Camacho, C. J., and D. Thirumalai. 1993. Kinetics and Thermodynamics of Folding in Model Proteins. Proceedings of the National Academy of Sciences of the United States of America 90:6369-6372.

17. Samiotakis, A., D. Homouz, and M. S. Cheung. 2010. Multiscale investigation of chemical interference in proteins. J Chem Phys 132:175101.
18. Anandakrishnan, R., B. Aguilar, and A. V. Onufriev. 2012. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic Acids Res 40:W537-541.
19. Stewart, J. J. 2007. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. J Mol Model 13:1173-1213.
20. Kumar, S., D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. 1992. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. J Comput Chem 13:1011-1021.
21. Grossfield, A. 2013. WHAM: an implementation of the weighted histogram analysis method. http://membrane.urmc.rochester.edu/content/wham Version 2.0.4.
22. Roux, B. 1995. The Calculation of the Potential of Mean Force Using Computer-Simulations. Comput Phys Commun 91:275-282.
23. Pall, S., M. J. Abraham, C. Kutzner, B. Hess, and E. Lindahl. 2015. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. Lect Notes Comput Sc 8759:3-27.
24. Lindorff-Larsen, K., S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins-Structure Function and Bioinformatics 78:1950-1958.
25. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of Simple Potential Functions for Simulating Liquid Water. Journal of Chemical Physics 79:926-935.
26. Darden, T., D. York, and L. Pedersen. 1993. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. Journal of Chemical Physics 98:10089-10092.
27. Parrinello, M., and A. Rahman. 1981. Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. J Appl Phys 52:7182-7190.
28. Hess, B., H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. 1997. LINCS: A linear constraint solver for molecular simulations. J Comput Chem 18:1463-1472.
29. Hendrix, D. A., and C. Jarzynski. 2001. A "fast growth" method of computing free energy differences. Journal of Chemical Physics 114:5974-5981.
30. Zhang, D., J. Gullingsrud, and J. A. McCammon. 2006. Potentials of mean force for acetylcholine unbinding from the alpha7 nicotinic acetylcholine receptor ligand-binding domain. J Am Chem Soc 128:3019-3026.
31. Ytreberg, F. M., and D. M. Zuckerman. 2004. Efficient use of nonequilibrium measurement to estimate free energy differences for molecular systems. J Comput Chem 25:1749-1759.
32. Jarzynski, C. 1997. Nonequilibrium equality for free energy differences. Phys Rev Lett 78:2690-2693.
33. Lipari, G., and A. Szabo. 1982. Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity. J Am Chem Soc 104:4546-4559.
34. Berjanskii, M., and D. S. Wishart. 2006. NMR: prediction of protein flexibility. Nat Protoc 1:683-688.

35. Kudlay, A., M. S. Cheung, and D. Thirumalai. 2009. Crowding effects on the structural transitions in a flexible helical homopolymer. Phys Rev Lett 102:118101.
36. Weinkam, P., E. V. Pletneva, H. B. Gray, J. R. Winkler, and P. G. Wolynes. 2009. Electrostatic effects on funneled landscapes and structural diversity in denatured protein ensembles. Proc Natl Acad Sci U S A 106:1796-1801.
37. Carpenter, G. A., and S. Grossberg. 1987. ART 2: self-organization of stable category recognition codes for analog input patterns. Appl Opt 26:4919-4930.
38. Cheung, M. S., and D. Thirumalai. 2007. Effects of crowding and confinement on the structures of the transition state ensemble in proteins. J. Phys. Chem. B 111:8250-8257.
39. Guo, Z., and D. Thirumalai. 1997. The nucleation-collapse mechanism in protein folding: evidence for the non-uniqueness of the folding nucleus. Folding and Design 2:377-391.
40. Waterhouse, A. M., J. B. Procter, D. M. Martin, M. Clamp, and G. J. Barton. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189-1191.