Supplementary Information

Conformational Frustration in Calmodulin-Target Recognition

Swarnendu Tripathi[1], Qian Wang[1], Pengzhi Zhang[1], Laurel Hoffman[2], M. Neal Waxham[2], and Margaret S. Cheung[1,3]

[1] Department of Physics, University of Houston, Houston, TX 77204

[2] Department of Neurology and Anatomy, University of Texas Health Science Center, Houston, TX 77030

[3] Center for Theoretical Biological Physics, Rice University, Houston, TX, 77005

Method:

Algorithm to calculate the association rate

We adopted the classic algorithm developed by Northrup, Allison and McCammon (NAM) [1] to compute the association rate k_a as described below,

$$k_a = k(b)\left[\frac{\beta}{1-(1-\beta)\Omega}\right]. \qquad \text{Eqn (S1)}$$

k(b) is the diffusion-controlled rate constant at a relative separation greater than *b* (Fig. S1), where the forces between CaM and the CaMBT (CaMKI or CaMKII) are centrosymmetric. Then *k(b)* is solved from the well known Smoluchowski's equation [2],

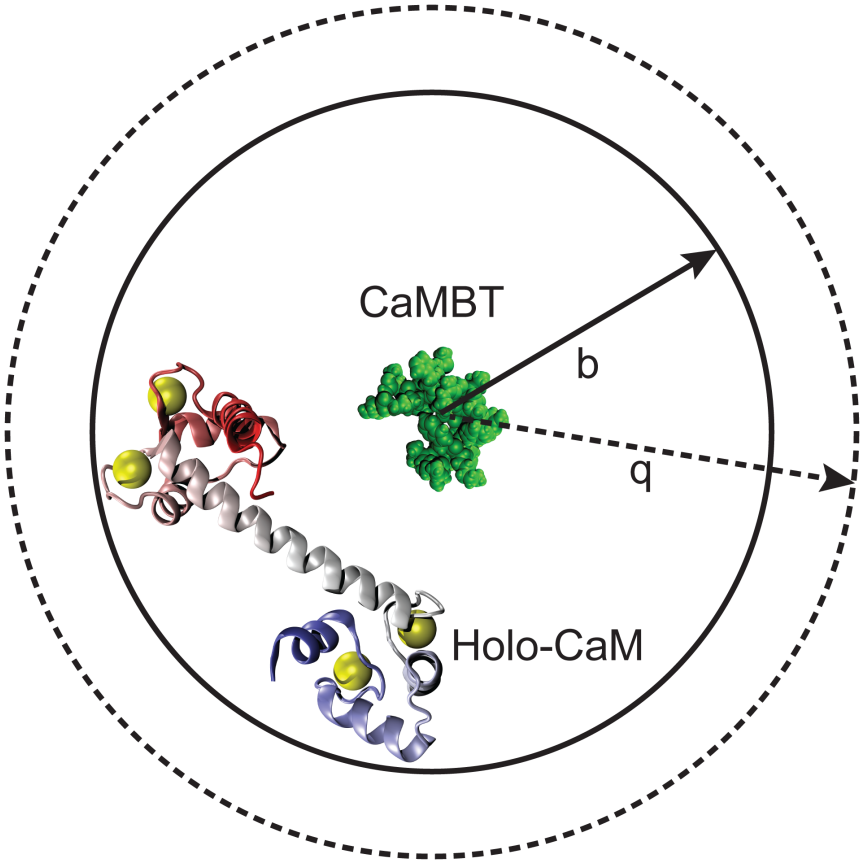
$$k(b) = 4\pi D b, \qquad \text{Eqn (S2)}$$

where D is the translational diffusion coefficient for the relative motion and β is the probability of successful association. $\Omega = \frac{b}{q}$ is the returning rate from the outer sphere q to the inner sphere b (shown in Fig. S1). In our Brownian Dynamics (BD) simulation of protein-protein association the effect of hydrodynamic interactions (HI) is not included.

Northrup *et al.* [3] showed that at the separation greater than the distance b (Fig. S1), D is defined as the sum of bulk self-diffusion coefficients of the two objects. For our CaM-CaMBT systems, D is D_{CaM}+D_{CaMBT}. The focus of our experimentally constrained molecular simulation is the ratio of k_a between CaM-CaMKI and CaM-CaMKII. This ratio is determined by β in Eqn (S1), but not the value of D in Eqn (S2). Therefore, we used the experimentally measured [4] D_{CaM} for the calculation of the association rates. For the CaMBTs, without available experimental values, we approximated D_{CaMBT} to be two times of D_{CaM} based on the ratio in sizes measured by the radius of gyration ($R_g^{\mathrm{CaM}} \sim$ 21 Å and $R_g^{\mathrm{CaMKI/CaMKII}} \sim 9$ Å).

Natural time unit of the Brownian dynamics simulation (described in the *Materials and Method* section of the *Main text*) is mapped to the real time unit in the overdamped limit by [5] $\tau = \frac{6\pi\eta a^3}{k_B T}$, where $k_B T = 0.66 \times 4.184$ kJ/mol, viscosity $\eta = 10^{-3} Pa \cdot s$ and a is the size of the amino acids. In our side-chain C_α model (SCM), the van der Waals (vdW) radius for a C_α beads is 0.45σ and vdW radii of the side-chain beads for CaM vary from 0.39σ to 1.18σ. A typical value of the size of a bead is between 0.39σ and 1.18σ (σ is 3.8 Å). Therefore, τ is mapped to the real time between 0.01-0.37 ns.

Figure S1. A schematic diagram for the simulations of the association of CaM–CaMBT (CaMKI/CaMKII) performed in the current study based on the NAM algorithm. The distances *b* and *q* represent the inner and outer radius of the spherical shells at 15σ and 75σ (σ is 3.8 Å), respectively.



Details of the Molecular Dynamics Simulations

The atomistic initial structures were reconstructed from sidechain-C_α configurations of the final complexes in the Brownian dynamics trajectory simulation [6], using SCAAL method [7]. All molecular dynamics simulations were performed using version 4.6.5 of the GROMACS molecular dynamics package. The proteins were modeled with two force fields, AMBER-99SB-ILDN force field [8] and CHARMM27 force field with CMAP corrections for protein and ions [9]. The rigid three-site TIP3P model [10] was used to simulate water molecules. The systems were neutralized by Na^+ and Cl^- ions, maintaining an ionic strength of 100 mM. The systems were minimized using steepest descent method before a short 0.1 ns NVT equilibration at 300K and a following 1 ns NPT equilibration with proteins fixed. The proteins were afterwards released and were further equilibrated for 5 ns. Another 10 ns NPT simulation was carried out for data collection. All NPT simulations maintained a constant pressure of 1 bar and temperature of 300 K using the Parrinello-Rahman barostat [11]. The bond lengths in proteins are constrained using the LINCS algorithm of Hess [12]. The equations of motion were integrated using a 2-fs time steps. Snapshots were saved for analysis once every 1 ps. Periodic boundary conditions were employed to mimic the macroscopic setting for electrolytes. Long-range electrostatic interactions between periodic images were treated using the particle mesh Ewald approach [13], with a grid size of 0.16 nm, fourth-order cubic interpolation and a tolerance of 10^{-5}. Neighbor lists were updated every 5 time steps. A cutoff of 10 Å was used for van der Waals interactions and real space Coulomb interactions as well as for updating neighbor lists.

Calculation of the Mean-square Fluctuation (MSF)

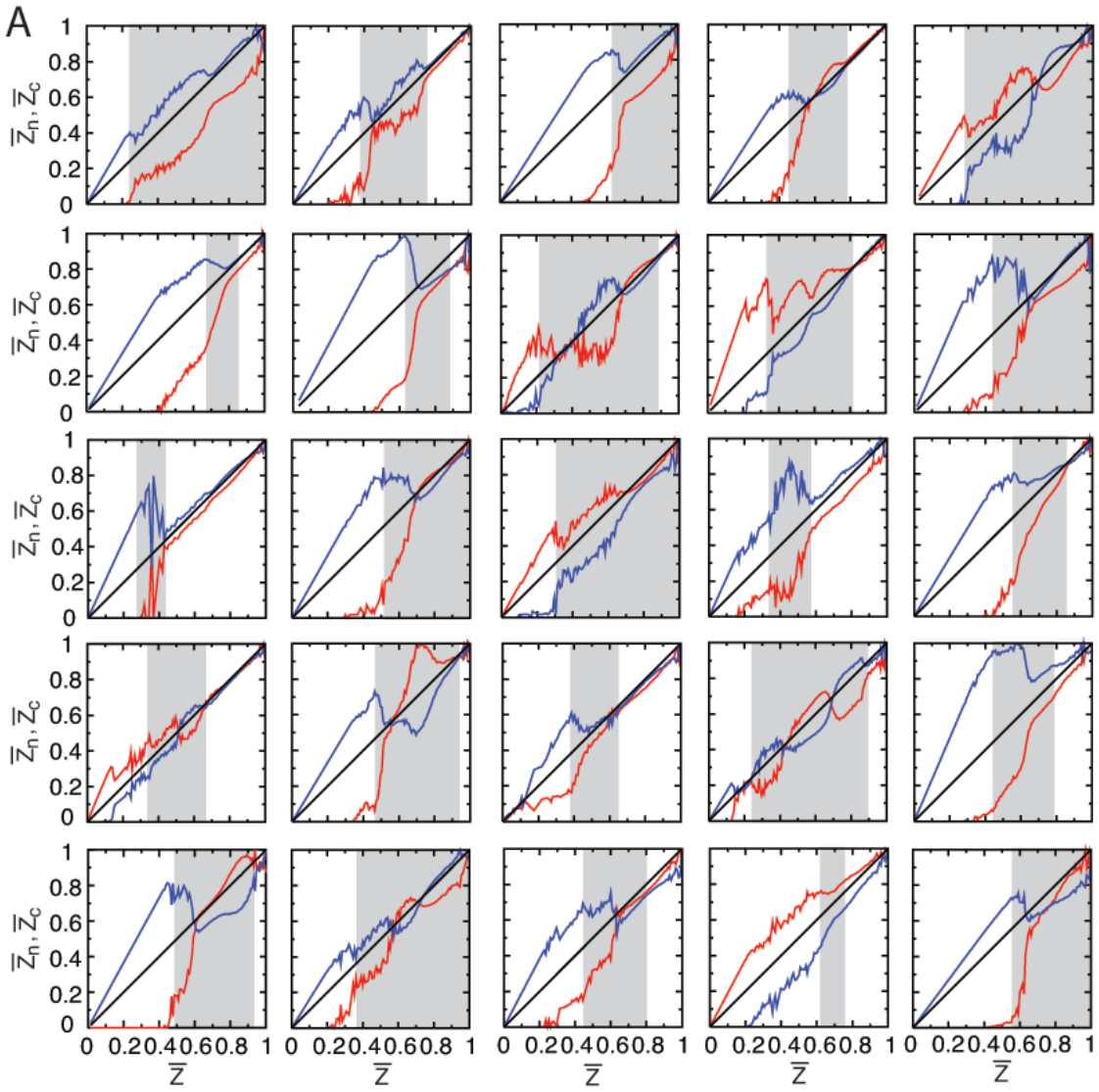
The MSF_j for atom *j* in each trajectory is calculated using the equation

$$MSF_j = \langle \left(\vec{r_j} - \langle \vec{r_j} \rangle\right)^2 \rangle \qquad \text{Eqn (S3)}$$

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

Results and Discussion

Fig. S2 The binding route analysis on each trajectory from the successful associations between CaM and CaMKI. (A) and (B) illustrate the formation of the normalized intermolecular contacts $\overline{Z_n}$ (in red) and $\overline{Z_c}$ (in blue) from the N- and C-terminal domain of CaM compared to the average $\bar{Z}$ (normalized) and the target CaMKI. $\bar{Z}$ is defined as the total number of intermolecular contacts between CaM and CaMKI, such that $\bar{Z} = \overline{Z_n + Z_c}$. See the *Models and Method* section for a detailed discussion on $\bar{Z}$, $\overline{Z_n}$ and $\overline{Z_c}$ in the *Main text*. In each plot, the diagonal line is drawn as a reference for the intermolecular contacts $\overline{Z_n}$ or $\overline{Z_c}$ that ideally follow the average $\bar{Z}$. In each plot the shaded region indicates backtracking, as discussed in the *Main text*.


A
$\overline{z}_n, \overline{z}_c$
0
0.2
0.4
0.6
0.8
1
$\overline{z}$

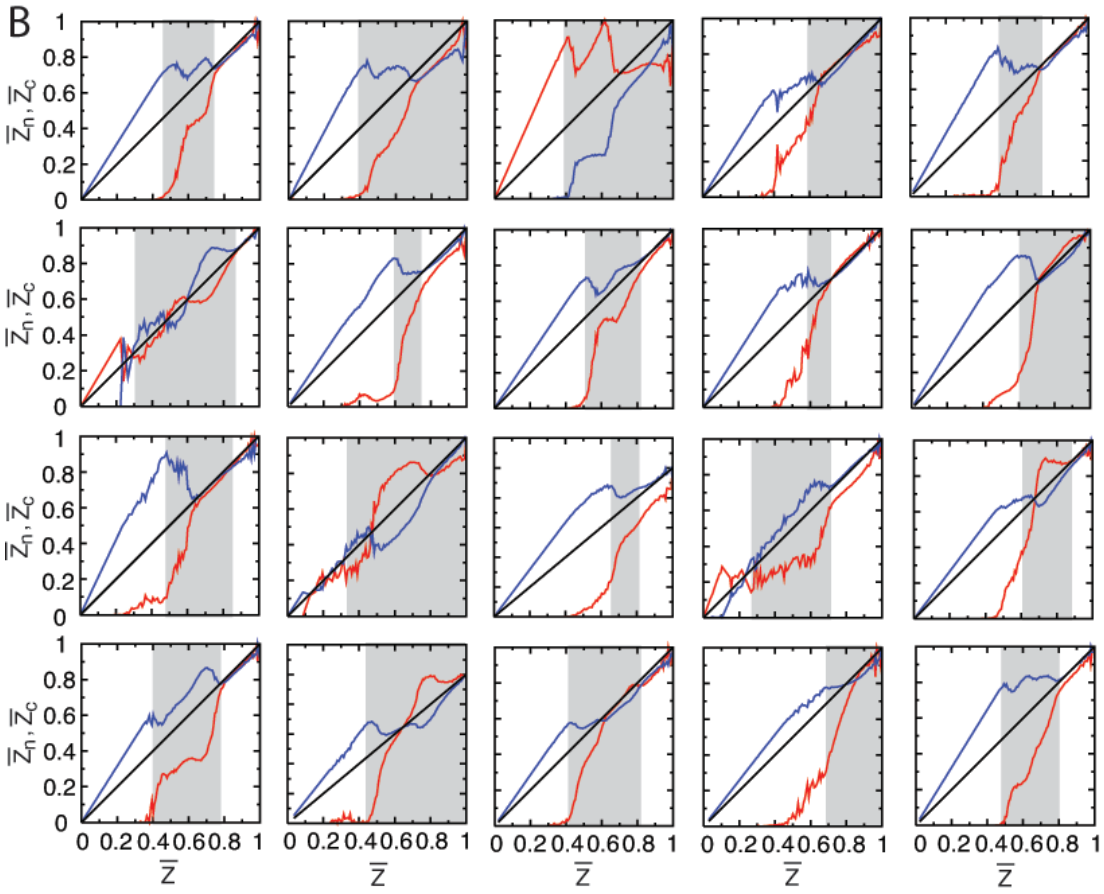
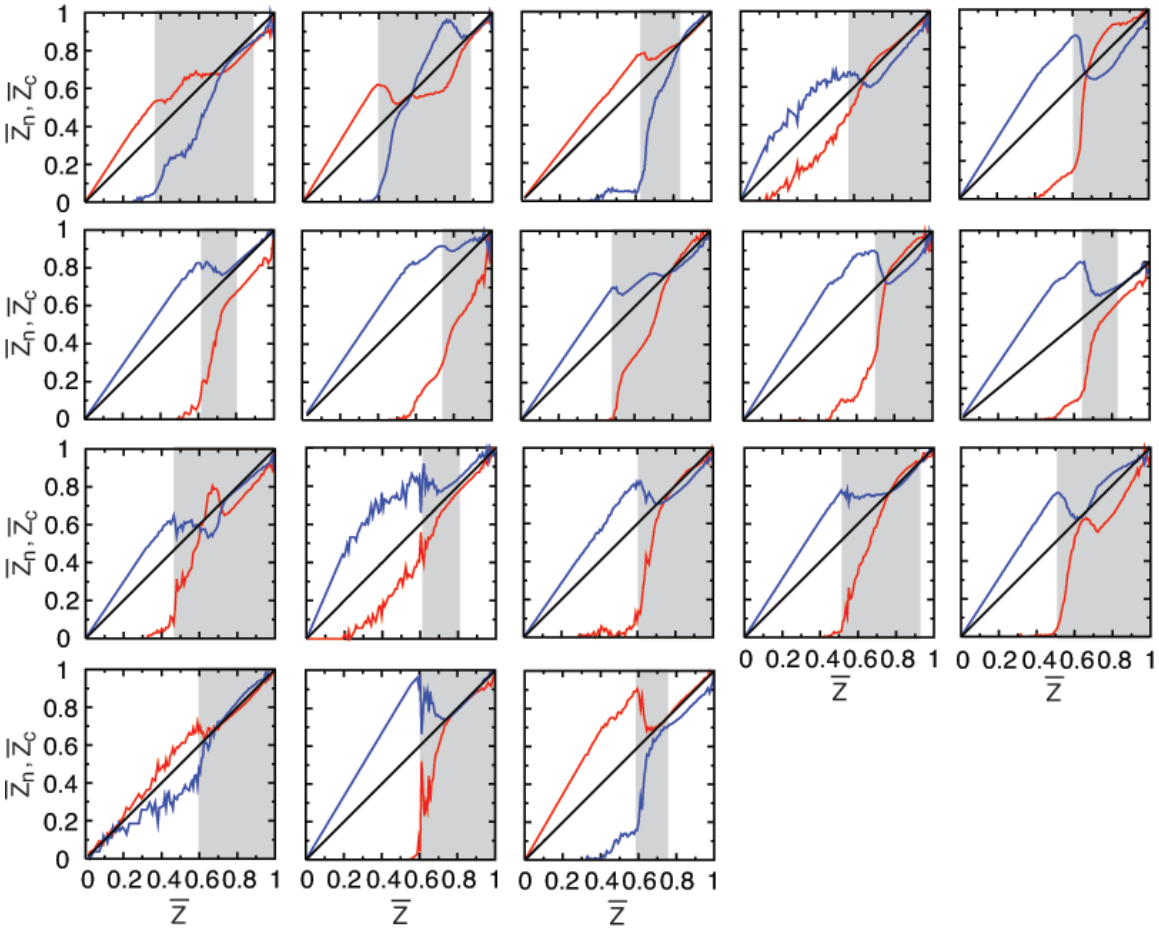

B
$\overline{z}_n, \overline{z}_c$
0
0.2
0.4
0.6
0.8
1
$\overline{z}$

Fig. S3 The binding route analysis on each trajectory from the successful associations between CaM and CaMKII. This plot illustrates the formation of the normalized intermolecular contacts $\overline{Z_n}$ (in red) and $\overline{Z_c}$ (in blue) from the N- and C-terminal domain of CaM compared to the average $\bar{Z}$ (normalized) and the target CaMKII. Definitions of the parameters $\overline{Z_n}$, $\overline{Z_c}$ and $\bar{Z}$ are similar as described in Fig. S2. In each plot the shaded region indicates backtracking, as discussed in the *Main text*.



Interactions of the $Ca2^{+}$- binding loops and the helix-linkers (of CaM domains) with the CaMBTs.

Based on our contact maps in Fig. 4 (for CaM-CaMKI) and Fig. 5 (for CaM-CaMKII) in the *Main text* along the binding route, we extended our analysis specifically for the Ca^{2+}-binding loops and helix-linkers of the CaM domains to further understand their role in binding the CaMBTs.

The CaM-CaMKI Complex

Fig. S4(A) shows that at $\bar{Z}$~0.3, all parts in the nCaM are far away from CaMKI. This is also evident from the contact map in Fig. 4(A), which shows that nCaM barely makes any contacts with CaMKI. In Fig. S4(B), only the FG helix-linker and the Ca^{2+} from loop IV (Ca4) of cCaM form contacts with CaMKI as their distributions show a peak at ~2σ.

From Fig. S4(C) at $\bar{Z}$~0.6, only the BC helix-linker (of nCaM) interacts with CaMKI as shown by a distribution that peaks at ~3σ. It is also supported by the contact map analysis on this region in Fig. 4(B). From Fig. S4(D), we found that all parts of cCaM are far from CaMKI. This shows that in order to form contacts between nCaM and CaMKI, some contacts between cCaM (specifically from FG-linker) and CaMKI needed to be broken [see Fig. 4(B)].

At higher $\bar{Z}$~0.8 from Fig. S4(E), the peak of the distribution between the BC-linker of nCaM and CaMKI moves away to a distance greater than 4.5σ. That means less contact in this region as shown in Fig. 4(C). In the case of FG-linker of the cCaM we

found that it moved closer to CaMKI in Fig. S4(F), making more intermolecular contacts with CaMKI [see Fig. 4(C)].

The CaM-CaMKII Complex

For the CaM-CaMKII system we noted that at $\bar{Z}$~0.2 all parts of nCaM remained far away from CaMKII by a distance greater than 12σ, as shown in Fig. S5(A). In Fig. S5(B) we found that only the FG helix-linker is close to the CaMKII at a distance of ~2σ; There are significant number of contacts formed in this region with CaMKII, as shown in Fig. 5(A).

Interestingly, at $\bar{Z}$~0.3 the Ca^{2+} binding loop I (Ca1) is particularly close to CaMKII at a distance less than 4σ [see Fig. S5(C)]. There are interactions of CaMKII especially with Ca1 in Fig. 5(B). In Fig. S5(D) there is a bimodal distribution of the distance between CaMKII and the FG-linker, in which half of the population is close to CaMKII and the other half is far away. Consequently, the probability of contact between CaMKII and the FG-linker is weakened [see Fig. 5(B)]. From Fig. S5(E) and (F), we noted that both the helix-linker BC of nCaM and the FG-linker of cCaM showed two distinct peaks. One of the two peaks remained much closer to CaMKII at a distance less than 4σ, which contributed to significant interactions on the contact maps [see Fig. 5(C)] in these regions. On the other hand, we found that all the Ca^{2+}-binding loops of CaM remained at a distance greater than 5σ from CaMKII providing few interactions.

Figure S4. Distributions of the distances between CaMKI and the Ca^{2+} ions, and the distances between CaMKI and the helix-linkers of the CaM domains during association. (A), (C) and (E) represent the distributions of the distances between CaMKI and the Ca^{2+} ions (Ca1 from binding loop I shown in red line and Ca2 from binding loop II shown in blue line) and the distance between CaMKI and the BC helix-linker (shown in green line) from the N-terminal domain of CaM, at $\bar{Z}$~0.3, 0.6 and 0.8, respectively (see Fig. 3(A) in the *Main text*). Similarly, (B), (D) and (F) represent the distributions of the distances between CaMKI and the Ca^{2+} ions (Ca3 from binding loop III shown in red line and Ca4 from binding loop IV shown in blue line) and the distance between CaMKI and the FG helix-linker (shown in green line) from the C-terminal domain of CaM, at $\bar{Z}$~0.3, 0.6 and 0.8, respectively. See the *Model and Methods* section in the *Main text* for the definition of $\bar{Z}$. All the distances were calculated from the center of mass of the CaMBTs. For the helix linkers the center of mass was used as the reference.

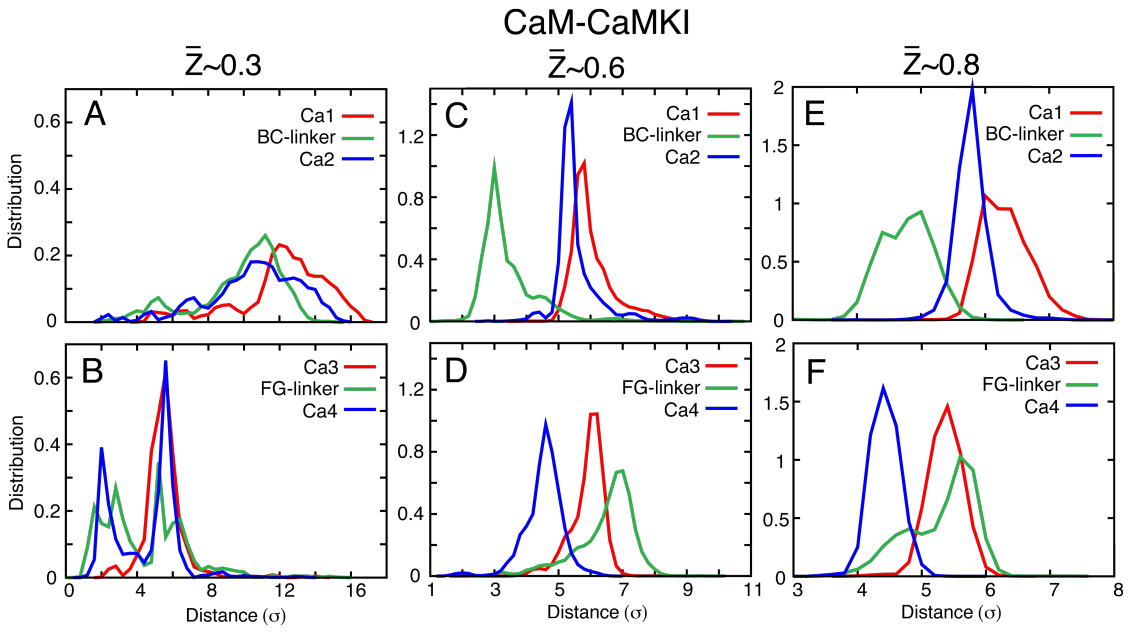


Figure S5. Distributions of the distances between CaMKII and the Ca^{2+} ions and the distances between CaMKII and the helix-linkers of the CaM domains during association. (A), (C) and (E) represent the distributions of the distances between CaMKII and the Ca^{2+} ions (Ca1 from binding loop I shown in red line and Ca2 from binding loop II shown in blue line) and the distance between CaMKII and the BC helix-linker (shown in green line) from the N-terminal domain of CaM, at $\bar{Z}$~0.2, 0.3 and 0.6, respectively (see Fig. 3(B) in the *Main text*). Similarly, (B), (D) and (F) represent the distributions of the distances between CaMKII and the Ca^{2+} ions (Ca3 from binding loop III shown in red line and Ca4 from binding loop IV shown in blue line) and the distance between CaMKII and the FG helix-linker (shown in green line) from the C-terminal domain Ca^{2+}-CaM, at $\bar{Z}$~0.2, 0.3 and 0.6, respectively. Distances were calculated using the same procedure as described in Fig. S4.

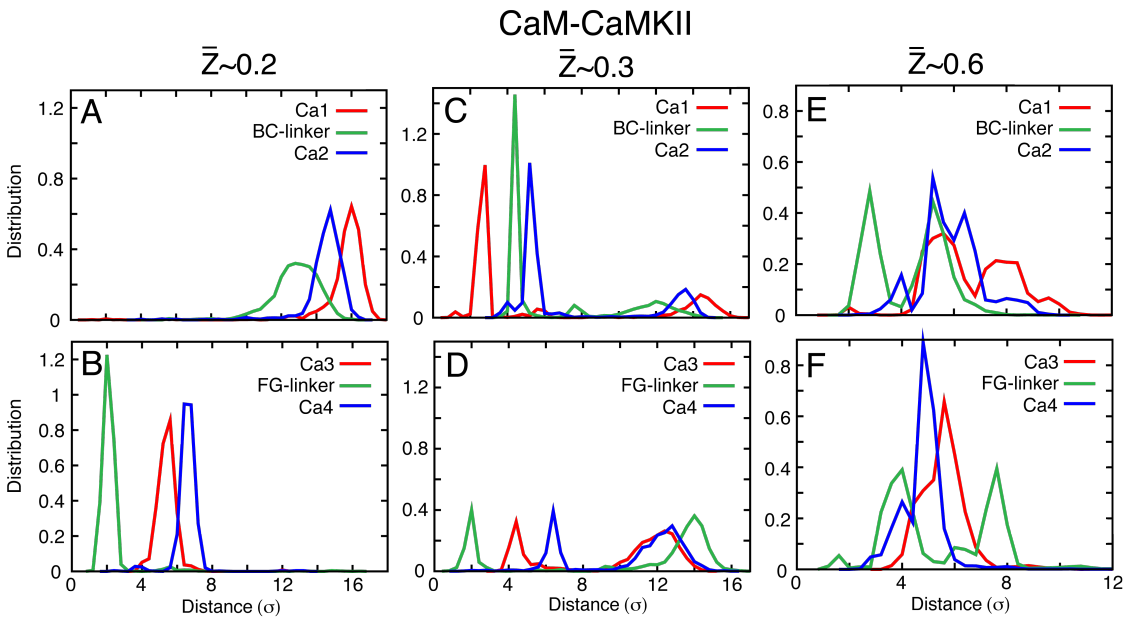


Figure S6. Probability of contact formation within CaM during binding with CaMBTs. (A), (B) and (C) represent the contact maps calculated between the amino acids from the side-chain beads within CaM (from the CaM-CaMKI complex), at $\bar{Z}$~0.3, 0.6 and 0.8, respectively (see Fig. 3(A) in the *Main text*). Similarly, (D), (E) and (F) represent the contact maps calculated between the amino acids from the side-chain beads within CaM (from the CaM-CaMKII complexes), at $\bar{Z}$~0.2, 0.3 and 0.6, respectively (see Fig. 3(B) in the *Main text*).

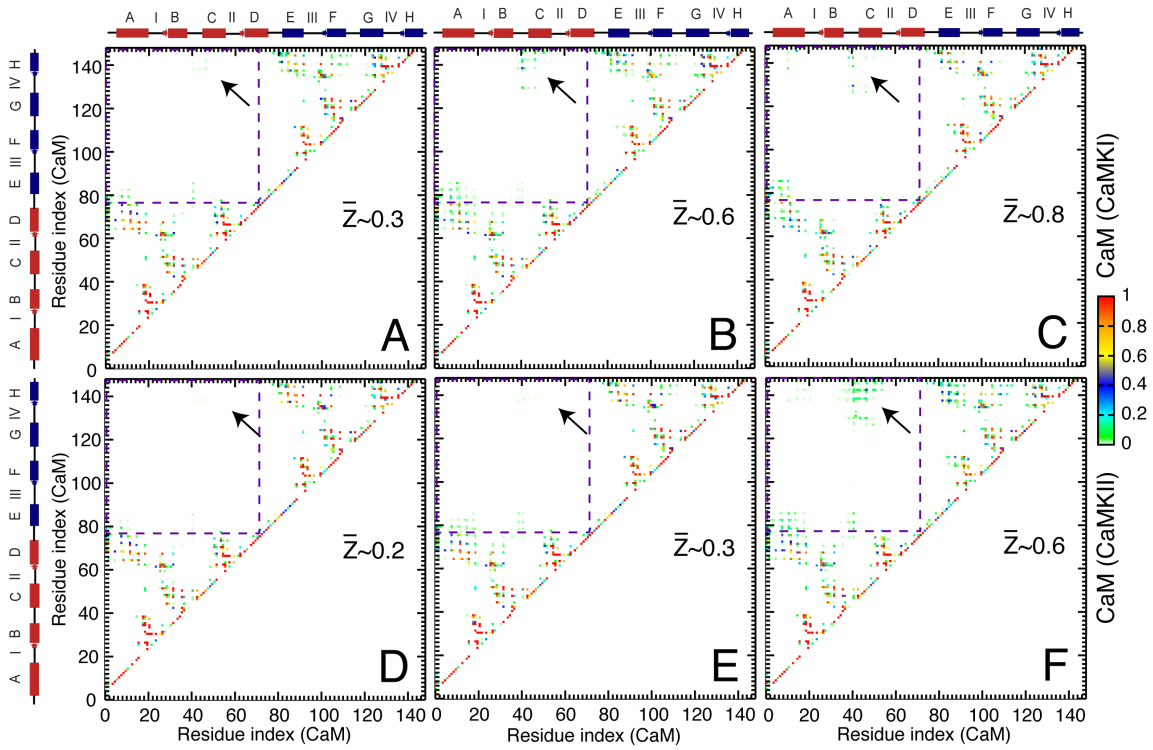


Figure S7. Superimposed structures of CaMBTs from the bound CaM-CaMBT complexes. We performed a clustering analysis [14] on the CaMBT from the bound CaM-CaMBT complexes for the trajectories that show successful association. For both CaMKI and CaMKII the first three most dominant clusters are shown. In all the structures CaMBTs are colored from blue (N-terminal end) to red (C-terminal end). For each cluster, the average value of the relative contact order (CO) [15] is indicated. In each cluster only ten structures of CaMBTs are shown for clarity.

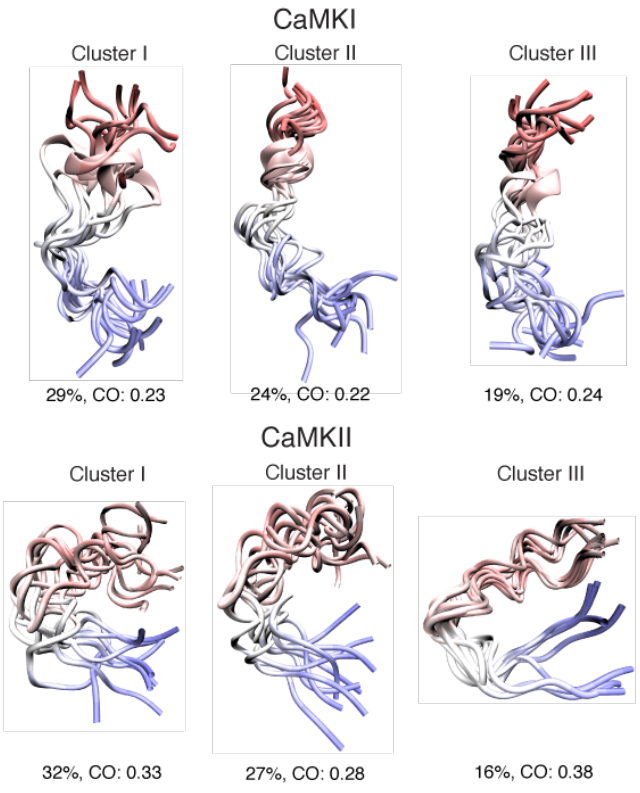
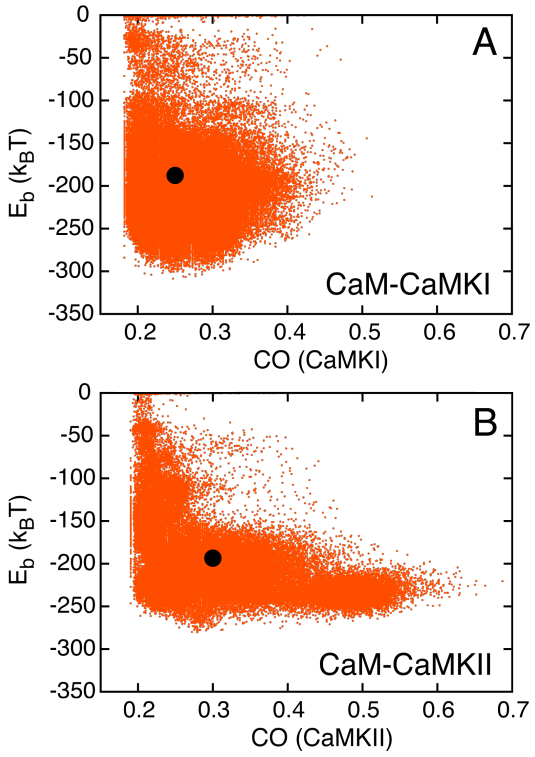


Figure S8 Binding energy (E_b) of CaM-CaMKI and CaM-CaMKII. (A) and (B) represent the two dimensional scattered plots of binding energy (E_b, in unit of the multiplication of the Boltzmann's constant and temperature, k_BT) as a function of the relative contact order (CO) of the CaMBTs, for CaM-CaMKI and CaM-CaMKII, respectively. The plots were made from the data of the trajectories of successful association for both systems. The black filled circle in each plot indicates the ensemble-averaged value of E_b and CO. E_b between CaM and CaMBT is calculated as the sum of intermolecular van der Waals and electrostatic energies.



Comparison of MSF (from all-atom MD simulation) and B-factor of CaM and CaMBTs

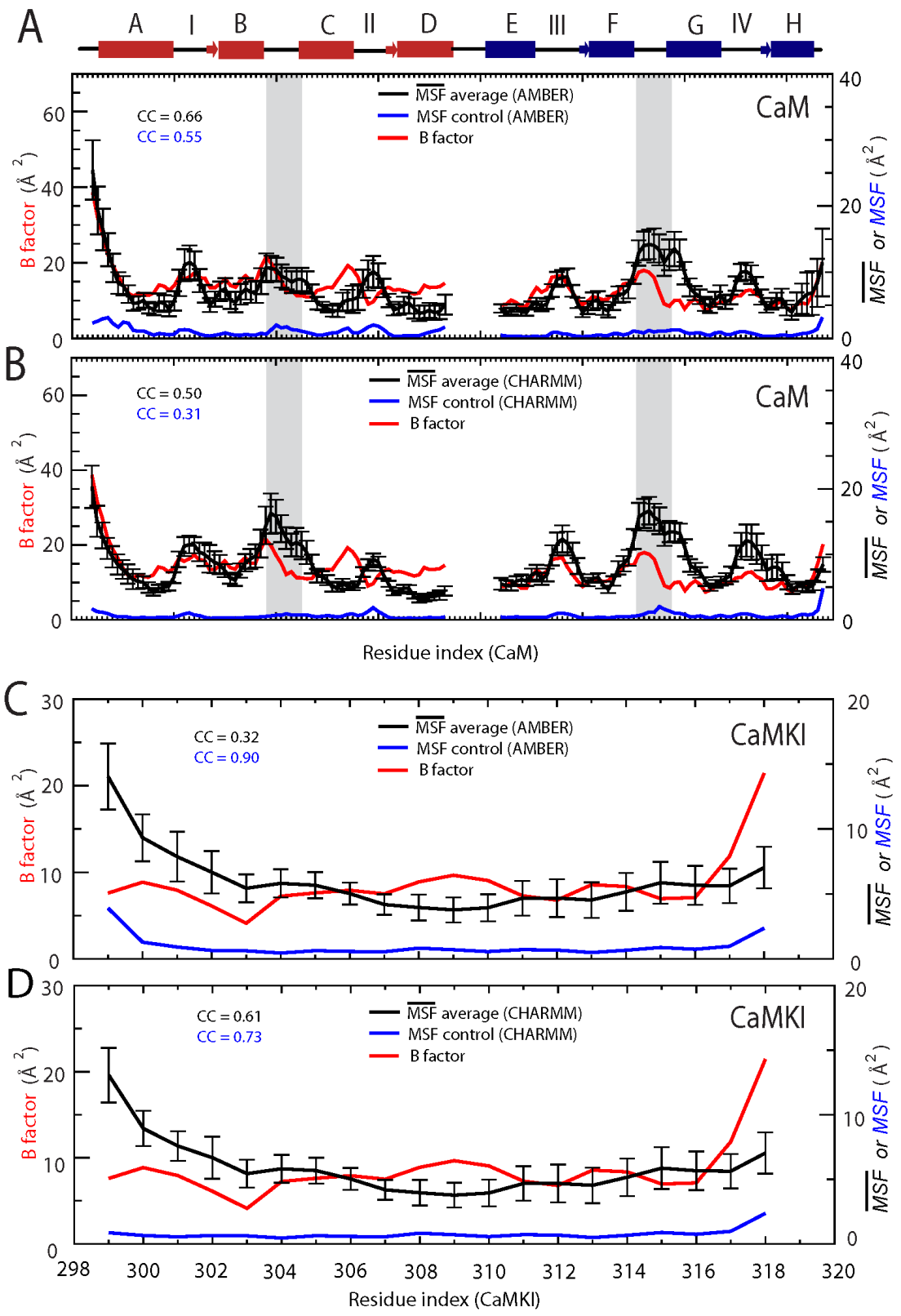
CaM-CaMKI

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

$\overline{MSF}$ was calculated for C_α atoms from atomistic simulations using the initial configurations reconstructed from 45 coarse-grained structures. A control simulation was carried out using the native complex structure (2L7L) and MSF was calculated accordingly for C_α atoms in Figure S9.

Figure S9. **Mean square fluctuations (MSF) from all-atomistic simulations and the B-factor experiments for CaM and CaMKI from the X-ray structure (PDB ID: 1MXE)**. (A) $\overline{MSF}$ (MSF) of CaM from molecular dynamics simulation using the AMBER99SB-ILDN force field (black from the reconstructed protein representations and blue from the control) and the B-factor (red curve). (B) $\overline{MSF}$ (MSF) of CaM from molecular dynamics simulations using the CHARMM27 force field and the B-factor (red

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


A
A I B C II D E III F G IV H
CaM
CC = 0.66
CC = 0.55
MSF average (AMBER)
MSF control (AMBER)
B factor
B
CaM
CC = 0.50
CC = 0.31
MSF average (CHARMM)
MSF control (CHARMM)
B factor
B factor (Å 2)
MSF or MSF (Å 2)
Residue index (CaM)
C
CaMKI
CC = 0.32
CC = 0.90
MSF average (AMBER)
MSF control (AMBER)
B factor
D
CaMKI
CC = 0.61
CC = 0.73
MSF average (CHARMM)
MSF control (CHARMM)
B factor
Residue index (CaMKI)

CaM-CaMKII

B factor for CaM-CaMKII is available in native complex (PDB code: 1CDM). The B factor values for C_α atoms were collected. $\overline{MSF}$ was calculated for C_α atoms from atomistic simulations using the initial configurations reconstructed from 18 coarse-grained structures. A control simulation was carried out using the native complex structure (1CDM) and MSF was calculated accordingly for C_α atoms in Figure S10.

Figure S10. Mean square fluctuations (MSF) from all-atomistic simulations and the B-factor experiments for CaM and CaMKII from the X-ray structure (PDB ID: 1CDM). (A) $\overline{MSF}$ (MSF) of CaM from molecular dynamics simulation using the AMBER99SB-ILDN force field (black from the reconstructed protein representations and blue from the control) and the B-factor (red curve). (B) $\overline{MSF}$(MSF) of CaM from molecular dynamics simulations using the CHARMM27 force field and the B-factor (red curve). (C) $\overline{MSF}$ (MSF) of CaMKII from molecular dynamics simulation using the AMBER99SB-ILDN force field (black from the reconstructed protein representations and blue from the control) and the B-factor (red curve). (D) $\overline{MSF}$(MSF) of CaMKII from molecular dynamics simulation using the CHARMM27 force field (black from the reconstructed protein representations and blue from the control) and B-factor (red curve). The correlation coefficient (CC) of average $\overline{MSF}$ and control MSF with respect to the B-factor from experiments is shown in each of the plot for both CaM and CaMKII. The $\overline{MSF}$ (or MSF) of CaM (or CaMKII) was calculated for the functional complexes of CaM-CaMKII from all the successful trajectories. The control MSF of CaM (or CaMKII) was calculated from the NMR structure of CaM-CaMKII complex (PDB ID: 1CDM).

The $\overline{MSF}$ (or *MSF*) of residue 74-83 from the central-linker of CaM are not shown in (A) and (B) and were not considered for the calculation of the CC. The shaded regions in (A) and (B) indicate the BC and FG helix-linker of CaM.

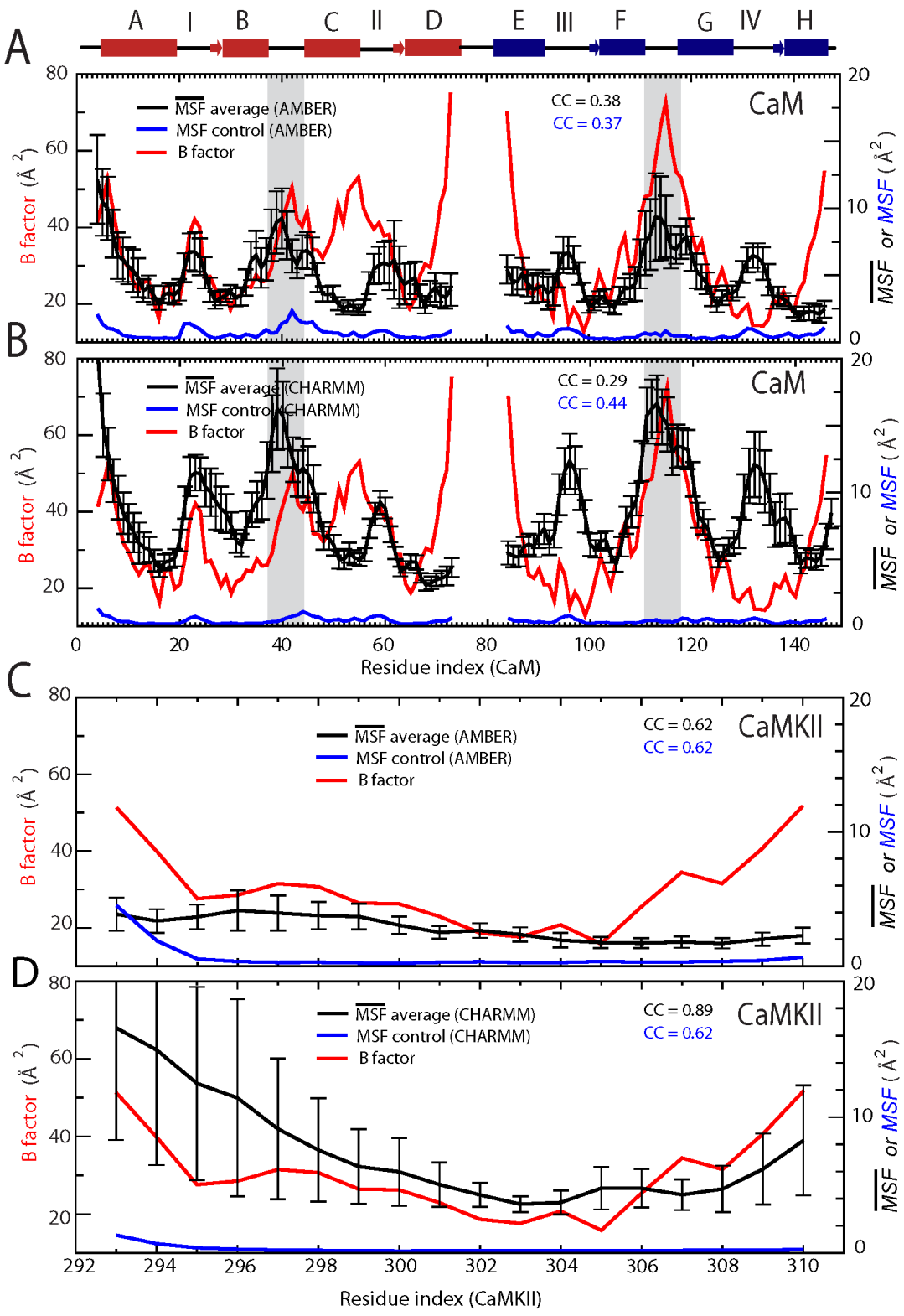

A
A I B C II D E III F G IV H
B
C
D
MSF average (AMBER)
MSF control (AMBER)
B factor
CC = 0.38
CC = 0.37
CaM
MSF average (CHARMM)
MSF control (CHARMM)
CC = 0.29
CC = 0.44
CC = 0.62
CC = 0.62
CaMKII
CC = 0.89
CC = 0.62
B factor (Å 2)
MSF or MSF (Å 2)
Residue index (CaM)
Residue index (CaMKII)

Table S1 Intermolecular interactions (Z) between CaM and CaMBT (from binding route analysis) at the early stage (ES) and the late stage (LS) of association. Intermolecular interactions are defined by $\bar{Z}$ (total normalized intermolecular contacts between CaM and CaMBT) from the binding route analysis (Fig. 3 in the *Main text*). The early stage and the late stage of the association of CaM and CaMBT are defined by Z_{75} (number of intermolecular contacts between the residue Lys75 of CaM and any amino acids from the CaMBT) [7]. The early stage (ES) and late stage (LS) are defined by the two major sharp transitions in Z_{75} during the association of CaM-CaMBT [7]. The parameters $\bar{Z}$ and Z_{75} are defined in the *Materials and Methods* section of the *Main text*.

Normalized $\bar{Z}$ (Binding route analysis)	Z_{75}
CaM-CaMKI	
$\bar{Z}$~0.3	0.25 (±0.01) (Before the ES)
$\bar{Z}$~0.6	5.22 (±0.01) (Transition from the ES to the LS)
$\bar{Z}$~0.8	5.20 (±0.01) (Transition from the ES to the LS)
CaM-CaMKII	
$\bar{Z}$~0.2	0.00 (±0.00) (Before the ES)
$\bar{Z}$~0.3	0.0 (±0.00) (Before the ES)
$\bar{Z}$~0.6	3.08 (±0.02) (Transition from the ES to the LS)

References

1. Northrup, S.H., S.A. Allison, and J.A. McCammon, *Brownian Dynamics Simulation of Diffusion-Influenced Bimolecular Reactions.* Journal of Chemical Physics, 1984. 80(4): p. 1517-1526.
2. Smoluchowski, M.V., *Drei Vortrage uber Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen.* Physik. Zeit., 1916. 17: p. 557-585.
3. Northrup, S.H. and J.T. Hynes, *Short-Range Caging Effects for Reactions in Solution .1. Reaction-Rate Constants and Short-Range Caging Picture.* Journal of Chemical Physics, 1979. 71(2): p. 871-883.
4. Slaughter, B.D., et al., *Single-molecule resonance energy transfer and fluorescence correlation spectroscopy of calmodulin in solution.* Journal of Physical Chemistry B, 2004. 108(29): p. 10388-10397.
5. Veitshans, T., D. Klimov, and D. Thirumalai, *Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties.* Folding & design, 1997. 2(1): p. 1-22.
6. Wang, Q., et al., *Protein recognition and selection through conformational and mutually induced fit.* Proc Natl Acad Sci U S A, 2013. 110(51): p. 20545-50.
7. Samiotakis, A., D. Homouz, and M.S. Cheung, *Multiscale investigation of chemical interference in proteins.* Journal of Chemical Physics, 2010. 132(17).
8. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field.* Proteins-Structure Function and Bioinformatics, 2010. 78(8): p. 1950-1958.
9. Bjelkmar, P., et al., *Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models.* Journal of Chemical Theory and Computation, 2010. 6(2): p. 459-466.
10. Jorgensen, W.L., et al., *Comparison of Simple Potential Functions for Simulating Liquid Water.* Journal of Chemical Physics, 1983. 79(2): p. 926-935.
11. Parrinello, M. and A. Rahman, *Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method.* Journal of Applied Physics, 1981. 52(12): p. 7182-7190.
12. Hess, B., et al., *LINCS: A linear constraint solver for molecular simulations.* Journal of Computational Chemistry, 1997. 18(12): p. 1463-1472.
13. Darden, T., D. York, and L. Pedersen, *Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems.* Journal of Chemical Physics, 1993. 98(12): p. 10089-10092.
14. Cheung, M.S. and D. Thirumalai, *Effects of crowding and confinement on the structures of the transition state ensemble in proteins.* Journal of Physical Chemistry B, 2007. 111(28): p. 8250-8257.
15. Plaxco, K.W., K.T. Simons, and D. Baker, *Contact order, transition state placement and the refolding rates of single domain proteins.* J Mol Biol, 1998. 277(4): p. 985-94.