SI Appendix

Fei Liu[1,2], Xiakun Chu[1,2], H. Peter Lu[3], Jin Wang[1,2,4,*]

1. College of Physics, Jilin University, Changchun, Jilin, P.R. China 130012
2. State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, Jilin, P.R. China 130022
3. Center for Photochemical Sciences, Department of Chemistry, Bowling Green State University, Bowling Green, OH 43403
4. Department of Chemistry and Physics, State University of New York at Stony Brook, Stony Brook, New York 11794-3400, United States

* jin.wang.1@stonybrook.edu

1 Materials and Methods

1.1 Reference PDB structure

We developed a two-basin structure-based model to explore the process that the Calmodulin bound to its target skeletal myosin light chain kinase (skMLCK). For the target free state, we use 1CLL(PDB) as reference structure, the Ca^{2+}-CaM adopts a dumbbell conformation before the binding of skMLCK(Fig. S1A). For the CaM-skMLCK complexes, we use 2BBM(PDB) as reference structure[6]. In this structure, CaM adopts the compact conformation with both N and C terminal domains binding skMLCK(Fig. S1B). The binding skMLCK peptide has 26 residues and are defined 1-4-8-14 by the spacing of hydrophobic anchor residues(Fig. S1B). The green, blue, and red regions are respectively the N-domain(residue 5-63), linker(residue 64-92) and C-domain(residue 93-146) of CaM(Fig. S1).

1.2 Coarse graining

In our model, each amino acid is represented by two beads (expect for glycine). The residue in backbone is represented by the CA bead, which is located on the C_α atom. The CB bead represents the side chain atom, which is located at center of mass of the side-chain atoms. There are four kinds of charged residues (include Asp(-), Glu(-), Lys(+), Arg(+)) and eight kinds of hydrophobic residue(Ala,Val,Leu,Ile,Met,Trp,Phe,Tyr) in our model.

1.3 Construction of two-basin potential and state-specific native contacts

The key to construct a model with two basins is to build a mixed contact map which integrates two structural informations together[10,16,17]. The contact map gives all possible interactions between pairs of residues in a given structure. In our work, the contacts are defined by contact map obtained with Contacts of Structural Units(CSU) algorithm[1].

We integrate a mixed contact map into our model by $U_{Attraction}$ in Energy Hamiltonian[10,16,17]. The $U_{Attraction}$ is give by the expression:

$$U_{Attracition} = U_{LJ}(Core) + U_{LJ}(open) + U_{LJ}(closed) + U_{LJ}(Ligand) + U_{LJ}(Intra-Ligand)$$

$$= \epsilon_{Core}U_{LJ} + \epsilon_{Specific}(\epsilon_{open}U_{LJ} + \epsilon_{closed}U_{LJ}) + \epsilon_{Ligand}U_{LJ} + \epsilon_{Intra-Ligand}U_{LJ}$$

$$U_{LJ} = \epsilon_{LJ}[5(\frac{\sigma_{ij}}{r_0})^{12} - 6(\frac{\sigma_{ij}}{r_0})^{10}]$$

The details of building the mixed contact map are shown as following:

1.3.1 Reference distances

We calculate the distances r_{ij}^{open} and r_{ij}^{closed} between the two beads i and j in the library, r_{ij}^{open} and r_{ij}^{closed} represent the distance of bead in open state(PDB ID :1CLL) and closed state(PDB ID :2BBM), respectively.

1.3.2 Core sublibrary γ^{Core}

We define $R_{ij}(X/Y) = \frac{r_{ij}^X}{r_{ij}^Y}$. For a bead pair between i and j, if both $R_{ij}(open/closed)$ and $R_{ij}(closed/open)$ are less than R_{cut}, this pair will be considered as a member of core contact library (γ^{Core}). We used $R_{cut} = 1.25$ in present work. A total of 252 contacts were obtained in the sublibrary γ^{Core}.

1.3.3 State-open sublibrary γ^{open}

For a pair with $R_{ij}(closed/open) > R_{cut}$, it will be considered as a member of state-open contact library γ^{open}. We used $R_{cut} = 1.4$ here. There were 16 contacts in γ^{open}(Fig. S7A).

1.3.4 State-closed sublibrary γ^{closed}

For a pair with $R_{ij}(open/closed) > R_{cut}$, it will be considered as a member of specific contacts ligand bound state. The number of specific contacts in bound state are 11(Fig. S7E).

1.3.5 Ligand sublibrary $\gamma^{binding}$

We used CSU[1] algorithm getting 98 ligand(skMLCK)-acceptor(Ca^{2+}-CaM) contacts from closed structure(2BBM). It includes 27, 31, 40 contacts between the skMLCK with the N-domain(Fig. S7B), the linker(Fig. S7C) and the C-domain(Fig. S7D) of Ca^{2+}-CaM, respectively.

1.3.6 Intra-Ligand sublibrary $\gamma^{intra-Ligand}$

We used CSU[1] algorithm getting 15 intra-skMLCK contacts with closed structure(2BBM)(Fig. 7F).

1.4 Repulsive interactions

The σ_{NC} is excluded distance between non-native pairs to provide excluded volume repulsion. The σ_{NC} is 4.0 $\mathring{A}$, the $\epsilon_{NC} = 1KJ/mol$. All pairs in γ^{all} are not considered in this term[10,16,17].

$$U_{Repulsive} = \epsilon_{NC}(\frac{\sigma_{NC}}{r_{ij}})^{12}$$

1.5 Electrostatic potential

The electrostatic potential is represented by the Debye-Huckel model[15]. The k^{-1} is Debye screening length which is directly affected by salt concentration. The $C_{salt} = 0.15M$ is the salt concentration. $k = 0.32\sqrt{C_{salt}}nm^{-1}$. The ϵ is dielectric constant and was set to 80. The q_i is the point charge of the ith bead, r_{ij} is the distance between two charged beads, $K_{coulomb} = 4\pi\epsilon_0 = 332kcalmol^{-1} = 138kJmol^{-1}nm^{-1}$. Residue charges are correspond to pH 7, such that q_i =+e for Lys and Arg, -e for Asp and Glu, where e is the elementary charge. The B(K)is the salt-dependent coefficient. The exact physical meaning B(K), k, $K_{coulomb}$, q_i, q_j can be found here[15].

$$U_{Electrostatic} = K_{coulomb}B(K)\frac{q_iq_jexp(-kr_{ij})}{\epsilon r_{ij}}$$

1.6 Hydrophobic potential

The hydrophobic interactions occur among hydrophobic residues. Eight kinds of hydrophobic residue(Ala, Val, Leu, Ile, Met, Trp, Phe, Tyr) in our model. The CB-CB hydrophobic potential is represented by 6-12 LJ potential, comparing to the 10-12 potential for non-hydrophobic and hydrophobic CA residues. The σ_{ij} for CB-CB hydrophobic residues is 5.0 $\mathring{A}$.

$$U_{hydrophobic(CB-CB)} = 4\epsilon_{Hydrophobic}[(\frac{\sigma_{ij}}{r_0})^{12} - (\frac{\sigma_{ij}}{r_0})^6]$$

1.7 Local Potential and the flexibility of Hinge Regions

The local potential is divided into bond stretching, angle bending and torsion energy[10,16,17].

$U_{Local} = U_{Bonds} + U_{Angles} + U_{Dihedrals}$

$U_{Bonds} = \sum_{bonds} K_b(r - r_0)^2$

$U_{Angles} + U_{Dihedral} = \epsilon_{angle}(U_{Non-hinge} + \epsilon_{Hinge}U_{Hinge})$

$U_{Non-hinge} = K_\theta(\theta - \theta_0)^2 + K_\phi[(1 - \cos(\phi - \phi_0)) + 0.5(1 - \cos(\phi - \phi_0))]$

$U_{Hinge} = K_\theta(\theta - \theta_0)^2 + K_\phi[(1 - \cos(\phi - \phi_0)) + 0.5(1 - \cos(\phi - \phi_0))]$

The bond energies U_{Bonds} is summed over the energy of all covalent bonds. The r is the distance of pseudo bond between adjacent atoms and with the subscript 0, it corresponds to r in open-form or closed-form. The $K_r = 10000KJ/(mol\ nm^2)$ is the bond force constant.

The expression of pseudo angle and pseudo dihedral are divided into non-hinge and hinge regions. The θ is the angle between two adjacent bonds, and the θ_0 is the reference angle in the native structure. The angle ϕ is the pseudo dihedral between the two planes formed by four connected atoms and ϕ_0 is the reference angle in the native structure. The angle constant $K_\theta = 20kJ/mol$. There are three different dihedral angles in our model. To form a dihedral angle need 4 atoms. For CB-CA-CA-CB dihedral angles and CA-CA-CA-CB dihedral angles, $K_\phi = 0.2kJ/mol$. If all four atoms are CA,$K_\phi = 0.8kJ/mol$. Improper torsion angles (or improper dihedral angles) are implemented using the same equation. Previous theoretical investigations reported that large-amplitude conformational change may cause structural strain in special portions of the protein (such as the hinge region)[2]. It is assumed that the strong local strain energies could be released by breaking fragile local interactions. The local interactions were weakened by decreasing the site-specific constants ϵ_{Hinge} as in the previous two-basin model. The region of the hinge was determined by the angle differences of pseudo angle and pseudo dihedral between open state and close state. Hinges were located wherever pseudo angle or pseudo dihedral angle differences were greater than threshold values. Because of the hinge region has large amplitude conformational change. It may cause structural strain which will led to the local energies. In our model, the threshold values of pseudo angle and pseudo dihedral are 14° and 50°, respectively. The energy is released by breaking the fragile local interactions[2,10,16,17]. We adopt the similar strategy just like the previous research[2,10,16,17]. The site-specific constant is rescaled by min(1,C_θ/E_{θ_i}), where the threshold $C_\theta = 1.0, E_{\theta_i} = K_\theta * |\theta_O - \theta_C|^2$. Similarly, we rescale site-specific constants for the dihedral angle, by min(1,C_ϕ/E_{ϕ_i}), in which the thresholds are set as $C_\phi = 1.0, E_{\phi i} = K_\phi * [1 - cos(|\phi_O - \phi_C|) + 0.5(1 - cos(3|\phi_O - \phi_C|))]$.

1.8 Model the Ca^{2+} explicitly

Each Ca^{2+} ion represents as a bead is located at the active site of EF-hand by attaching to the nearest negatively charged residues. In Figure S8, we represented one example of Ca^{2+} tethering to ASP20, ASP22, ASP24 and GLU31. The localization restraint is realized by spring-like potential in the distance and angle in native structure($U_{Bonds} = K_b(r - r_0)^2$, $U_{angle} = K_\theta(\theta - \theta_0)^2$). The other interactions involving Ca^{2+} are only electrostatic. For a calcium ion, +2 charge units was assigned. However, the effective charge of the calcium in a protein environment will be less than +2 in our work because so many negative charge residues are around it(Fig. S8). The angle constant $K_\theta = 20kJ/mol$. The $K_r = 10000KJ/(mol\ nm^2)$.

1.9 Harmonic constraint

In our simulation, the CaM and its target were placed in a cubic box and a harmonic potential was added if the two chains were farther than 9.0 nm. Many studies have shown that this constraint is effective and will not affect the binding mechanism between the CaM and its target[10,13,14,17].

1.10 Simulation parameters

To achieve a better sampling, we used Replica exchange Molecular Dynamics (REMD)[9] to get the thermodynamics landscape(Fig. 1A) of calmodulin binding with the skMLCK. We used 25 replicas and the neighbor replica attempted to exchange with each other at every 2000 MD steps. For all replicas, the total simulation time is 1.25 μs. For the thermodynamics landscape(Fig. 1A), the parameters of structure based conformational dynamic model are: $Temperature = 1.5\epsilon_{LJ}$, $\epsilon_{Core} = 2.5$, $\epsilon_{Specific} = 1.0$, $\epsilon_{open} = 1.0$, $\epsilon_{closed} = 1.2$, $\epsilon_{Ligand} = 1.2$, $\epsilon_{Intra-Ligand} = 1.2$, $\epsilon_{Hinge} = 10.0$, $salt\ concentration(C_{salt}) = 0.15mol/L$, $\epsilon_{hydrophobic} = 1.0$. The landscapes in different temperatures are shown in Fig. S16. Additionally, in order to study how the electrostatic interactions, the hydrophobic interaction and the flexibility influence the thermodynamic and dynamics properties of the system(Fig. 2, Fig. 3), we perform a group of constant temperature simulations and the total simulations are running 20 μs. Except ϵ_{Hinge}, $salt\ concentration(C_{salt})$, $\epsilon_{hydrophobic}$, which need to be respectively changed in different analyses(Fig. 2, 3), the other parameters for temperature simulations are same with REMD simulation.

1.11 Fraction of formation of certain state-specific contacts

The traditional folding Q has been an exact reaction coordinate in measuring the degree of folding in the studies of protein folding[3]. However, for functional dynamics, most contacts are shared by the native functional states, which will not be broken in the process of functional dynamics. So we should only measure the formation of state-specific native contacts(See **Construction of two-basin potential and state-special native contacts**) instead of all native contacts to describe the functional landscape[10,16,17]. The Q of state-specific contacts has same definition with the Q of all contacts. Their difference is that they will describe the folding degree of different regions.They can be expressed by:

$$Q_{open,close,binding,intra-ligand} = \sum_{\gamma^{open,close,binding,intra-ligand}} \frac{1-((R_{ij}-D_0)/R_{ij}^{nat})^n}{1-((R_{ij}-D_0)/R_{ij}^{nat})^m}$$

where, $m = 20, n = 10, R_{ij}^{nat}$ is native distance of atom i and atom j. $D_0 = 0.3 * R_{ij}^{nat}$. The Q_{open} is the fraction of intra-native contacts for Ca^{2+}-CaM in the open structure("O", PDB:1CLL) and the Q_{close} is the fraction of intra-native contacts for Ca^{2+}-CaM in the closed structure("C", PDB:2BBM). The $Q_{binding}$ is the fraction of native contacts between Ca^{2+}-CaM and skMLCK in the closed structure. The $Q_{intra-ligand}$ is the fraction of native contacts for intra-skMLCK contacts from closed structure.

1.12 Cut-off Algorithm to Count non-native Contacts

We used a cut-off algorithm to describe the interactions between the two chains, which can describe non-native contacts formed in the trajectory. If the distance between two atoms is shorter than 5 $\mathring{A}$, we define the contact value as 1. If the distance is between 5 $\mathring{A}$ and 8 $\mathring{A}$ the contact value is 0.5[10,18]. The non-native contacts for each states are normalized by the total numbers of each states in the trajectory.

2 Comparing with the FRET experiment

We compare our simulation with a FRET experiment[4] qualitatively. The FRET experiment shows the fluctuations of the N-terminal domain of CaM between the conformations of bounded and loosely bounded to the target peptide, while the C-terminal domain of CaM is overall still associated with the target peptide[4]. The FRET experiment also shows the distribution of FRET efficiency which was analyzed to reveal the distance fluctuations between the N-terminal domains of CaM and its target(Fig. S12). The broad distribution of FRET efficiency revealed that there were intermediate states in the interaction between CaM and its target[4]. From structural analysis based our modeling results, CaM in I1 region, though not very stable due to the flexibility of the central linker, provides strong evidence of the intermediate states that the targets only bind to the C-terminal domain of CaM and the conformation of the central linker does not change. In addition, an 3/4 closeness intermediate "L-B" is found in our work, whose C-terminal domain is also tightly bound to the target peptide. These results in our simulation also shown the dynamic trend that the N-terminal domain of CaM fluctuations bound and loosely bound to the target peptide while the C-terminal domain of CaM is overall still associated with the target peptide. Our work is accordance with the FRET experiment.

3 Structural characteristic for each state

We used the fraction of state specific native contact probability map to explore the structural features of each state. The Q_{Open} is the fraction of intra-native contacts for Ca^{2+}-CaM in the open structure("O", PDB:1CLL) and the Q_{Closed} is the fraction of special intra-native contacts for Ca^{2+}-CaM in the closed

structure("C", PDB:2BBM). The $Q_{binding}$ is the fraction of special native contacts between Ca^{2+}-CaM and skMLCK in the closed structure. The Fig. S2, Fig. S3 and Fig. S4 are respectively the native contacts probability map for Q_{Ooepn}, Q_{Closed}, $Q_{binding}$ in different states and regions. We have found the "LB" is a partial binding and partial closeness state comparing to the free open state("O"). Compare the difference of the contact map for "O" and "LB"(Fig. S2,S3,S4), with only minor changes in the two domains of CaM, we found the long helical structure of the central linker is broken and bending of the linker contributes(black square in Fig S2) to the closeness of Ca^{2+}-CaM in the "LB" state. The N-terminal of skMLCK mainly binds to the C-domain as well as the linker close to C-domain of CaM(red square in Fig S4), which make the "LB" stable. From structural analysis (Fig. S2,S3,S4), CaM in I1 region, though it is not very stable due to the flexibility of the central linker[8], rather than LB state, resembles the intermediate states proposed by experimental study. In addition, if only have the breaking in central linker(I3), the conformation of CaM is also not very stable. The "LB" is more stable than "I1" and "I3", which caused by both breaking of central linker and the binding of skMLCK.

Compare the difference of the contact map for "LB" and "C"(Fig. S2,S3,S4), the contact probability map for Q_{Open} (Fig. S2)shows the native contacts involve in the helix structure of the central linker are still broken(Fig. S2). The C-terminal of skMLCK will bind to the N-domain of CaM(Fig. S4). The binding between the N-domain of CaM and the C-terminal of skMLCK drives the two domains close to each other. The contact probability map for Q_{Closed} shows that native contact between the N-domain and C-domain of CaM formed from "LB" to "C"(Fig. S3, green square). The major role of the contacts between the N-domain and C-domain of CaM is to help the stable conformation, because its number is far less than the contacts between the N-domain of CaM and the C-terminal of skMLCK. These behaviors make the two domains wrap around the skMLCK peptide to form a compact globular complexes("C") from "LB".

4 The non-native interaction drive conformational change of the target

The skMLCK changes its conformation only when it binds to CaM. We found both the non-native electrostatic and non-native hydrophobic interaction play the role in driving the conformational change of the skMLCK(Fig. S10A and Fig. S10B). Increase in non-native hydrophobic interaction can drive all low helicities of skMLCK to form the higher helicity conformation(Fig. S10B). Increase in the non-native electrostatic interaction can also drive the low helicity of skMLCK to form the higher helicity conformation(Fig. 10A). However, this influence are slight and limited compared with the non-native hydrophobic interaction.

5 The simulations for CaM and smMLCK

To learn the binding mechanism between CaM with different target, we also develop a two basins structure-based model to calculate the landscape of the process that the calmodulin bound to smooth muscle myosin light chain kinase(smMLCK) binding peptide[7]. The smMLCK is termed 1-8-14 by it conserved position of hydrophobic residues[11](Fig. S13). For the target free state, we use 1CLL(PDB) as reference structure, the Ca^{2+}-CaM adopts a dumbbell conformation before the binding of skMLCK(Fig. S1A). For the CaM-smMLCK complexes, we use 1CDL(PDB)[11] as reference structure(Fig. S13). The method of developing two basins structure-based model for smMLCK and CaM is same with the method for CaM and skMLCK, the all simulation parameters for smMLCK are same with the parameters for the simulation of skMLCK.

References

[1] Sobolev V, Sorokine A, Prilusky J (1999) Automated analysis of interatomic contacts in proteins Bioinformatics15, 327−32

[2] Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG. (2006) Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations Proc Natl Acad Sci USA 103, 11844−9.

[3] Cho SS, Levy Y, Wolynes PG (2006) P versus q: structural reaction coordinates capture protein folding on smooth landscapes.Proc Natl Acad Sci USA 103, 586C591.

[4] Liu RC, Hu DH, Tan X, Lu HP (2006) Revealing two-state protein-protein interactions of calmodulin by single-molecule spectroscopy. Journal of the American Chemical Society 128,10034−10042.

[5] Sun H, Squier TC (2000) Ordered and cooperative binding of opposing globular domains of calmodulin to the plasma membrane Ca-ATPase. J. Biol. Chem. 2000, 275 (3), 1731−1738.

[6] Ikura M (1992) Solution Structure of a Calmodulin-Target Peptide Complex by Multidimensional Nmr. Science 256, 632−638.

[7] Meador WE, Means AR, Quiocho FA (1992) Target Enzyme Recognition by Calmodulin−2.4−Angstrom Structure of a Calmodulin-Peptide Complex. Science 257, 1251−255.

[8] Meador WE, Means AR, Quiocho FA (1993) Modulation of Calmodulin Plasticity in Molecular Recognition on the Basis of X-Ray Structures. Science 262, 1718−1721 (1993).

[9] Liu RC, Hu DH, Tan X, Lu HP (2006) Okamoto Y, Sugita Y (1999) Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 314: 141−151.

[10] Chu X, Liu F, Maxwell BA, Wang Y, Suo Z (2014) Dynamic Conformational Change Regulates the Protein-DNA Recognition: An Investigation on Binding of a Y-Family Polymerase to Its Target DNA. Plos Comput Biol 10(9).

[11] Yamniuk AP, Vogel HJ (2004) Calmodulin's flexibility allows for promiscuity in its interactions with target proteins and peptides. Molecular biotechnology 27, 33-57.

[12] Tripathi S, Portman JJ (2009) Inherent flexibility determines the transition mechanisms of the EF-hands of calmodulin. P Natl Acad Sci USA 106(7):2104-2109.

[13] Wang Y, Chu X, Longhi S, Roche P, Han W, et al. (2013) Multiscaled exploration of coupled folding and binding of an intrinsically disordered molecular recognition element in measles virus nucleoprotein. Proc Natl Acad Sci USA 110: E3743CE3752.

[14] Levy Y, Cho SS, Onuchic JN, Wolynes PG (2005) A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. J Mol Biol 346: 1121C1145

[15] Azia A, Levy Y (2009) Nonnative electrostatic interactions can modulate protein folding: Molecular dynamics with a grain of salt. J Mol Biol 393: 527C542.

[16] Wang Y, Tang C, Wang EK, Wang J (2012) Exploration of Multi-State Conformational Dynamics and Underlying Global Functional Landscape of Maltose Binding Protein. Plos Comput Biol 8(4).

[17] Wang Y, Gan L, Wang E, Wang J (2012) Exploring the dynamic functional landscape of adenylate kinase modulated by substrates. J Chem Theory Comput 9: 84−95

[18] Wang J, et al. (2011) Multi-scaled explorations of binding-induced folding of intrinsically disordered protein inhibitor IA3 to its target enzyme. PLoS Comput Biol7(4):e1001118.

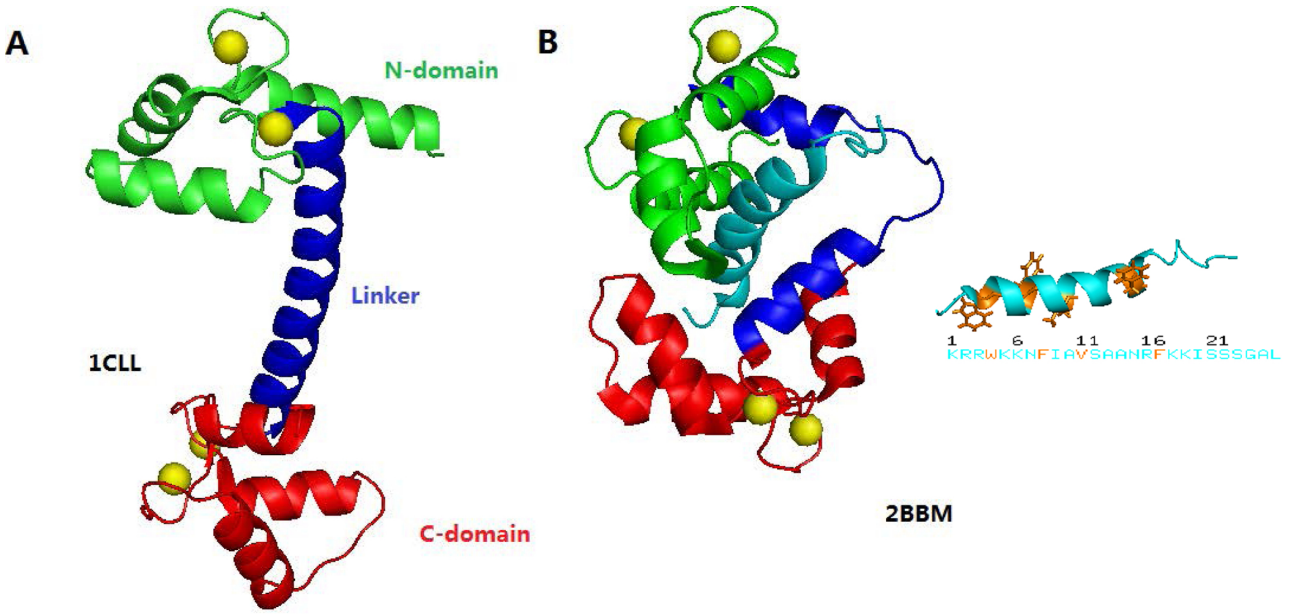

Fig.S 1: The PDB structures for model. For the target free state, we use 1CLL(PDB) as reference structure, the Ca^{2+}-CaM adopts a dumbbell conformation before the binding of skMLCK. For the CaM-skMLCK complexes, we use 2BBM(PDB) as reference structure. In this structure, CaM adopts the compact conformation with both N- and C-domains binding skMLCK. The binding skMLCK peptide has 26 residues and are defined 1-4-8-14 by the spacing of hydrophobic anchor residues. The green, blue, red regions are respectively the N-domain(residue 5-63), linker(residue 64-92) and C-domain(residue 93-146) of CaM.

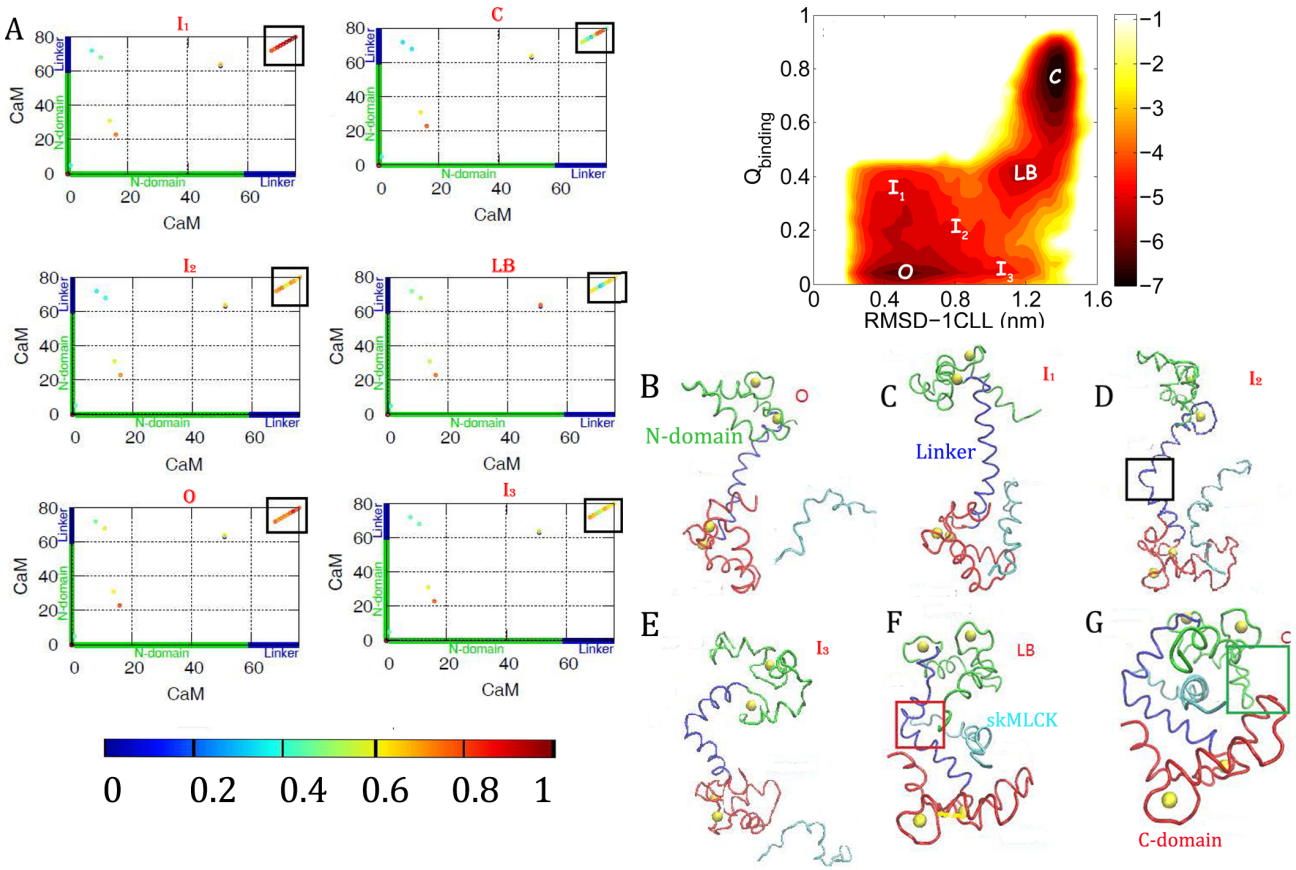

Fig.S 2: (A) The native contacts probability maps for Q_{Open} in different states and regions are shown. The contacts probability map Q_{Open} shows the helix structure of the central of the linker will be broken and the bend of the linker led to the Ca^{2+}-CaM closed in the "LB". For the pathway "LB" to "C", the contacts probability map for Q_O shows the native contact involve in in the helix structure of the central of the linker is still broken. (B) The target free state, it is the open state(O) of Ca^{2+}-Calmodulin before the binding of skMLCK. (C) The conformation extracted from I1 region is shown , the skMLCK only binds to the C-domain of CaM and the conformation of CaM did not change comparing to the open state "O" in this region. (D) The conformation extracted from I2 region is shown, the skMLCK only binds to the C-domain of CaM but the degree of both binding and closing are less than intermediate "LB". (E) The conformation extracted from I3 region is shown, the skMLCK did not bind to the CaM and the degree of the closing of CaM allow the skMLCK bind to it. (F) The intermediate "LB" which is partly binding and partly closed comparing to the completely free open state "O". (G) The Ca^{2+}-Calmodulin-skMLCK complexes are shown, calmodulin adopts the compact conformation of both N- and C-terminal domains binding to skMLCK.

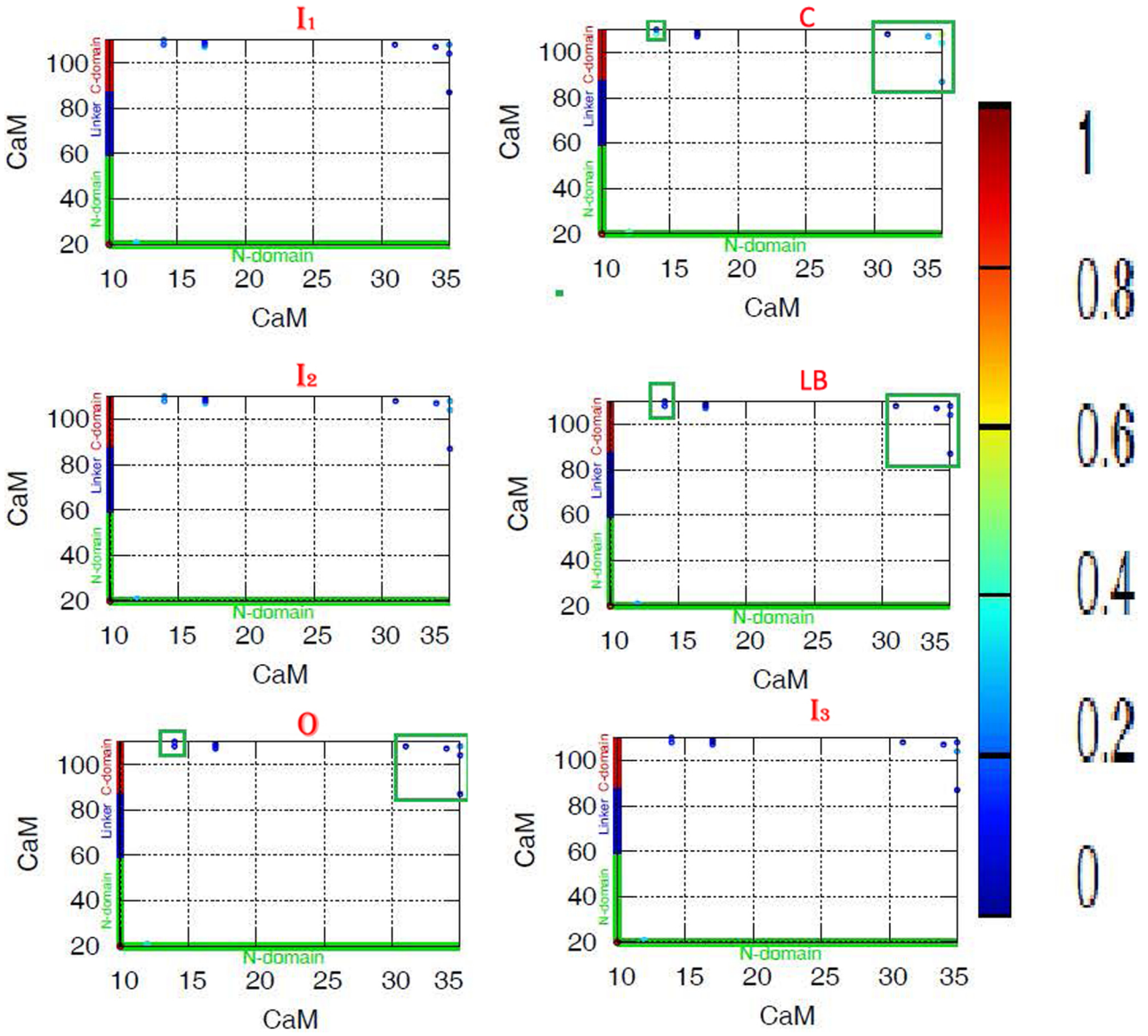

Fig.S 3: The native contacts probability map Q_{Closed} in different states and regions are shown. The Q_{Closed} did not form in both"LB" and "O" state. The contacts probability map for Q_{Closed} show that native contact between the N-domain and C-domain of Ca^{2+}-CaM formed from "LB" to "C", which make the two domains wrap around the -helical peptide to form a compact globular complexes("C").

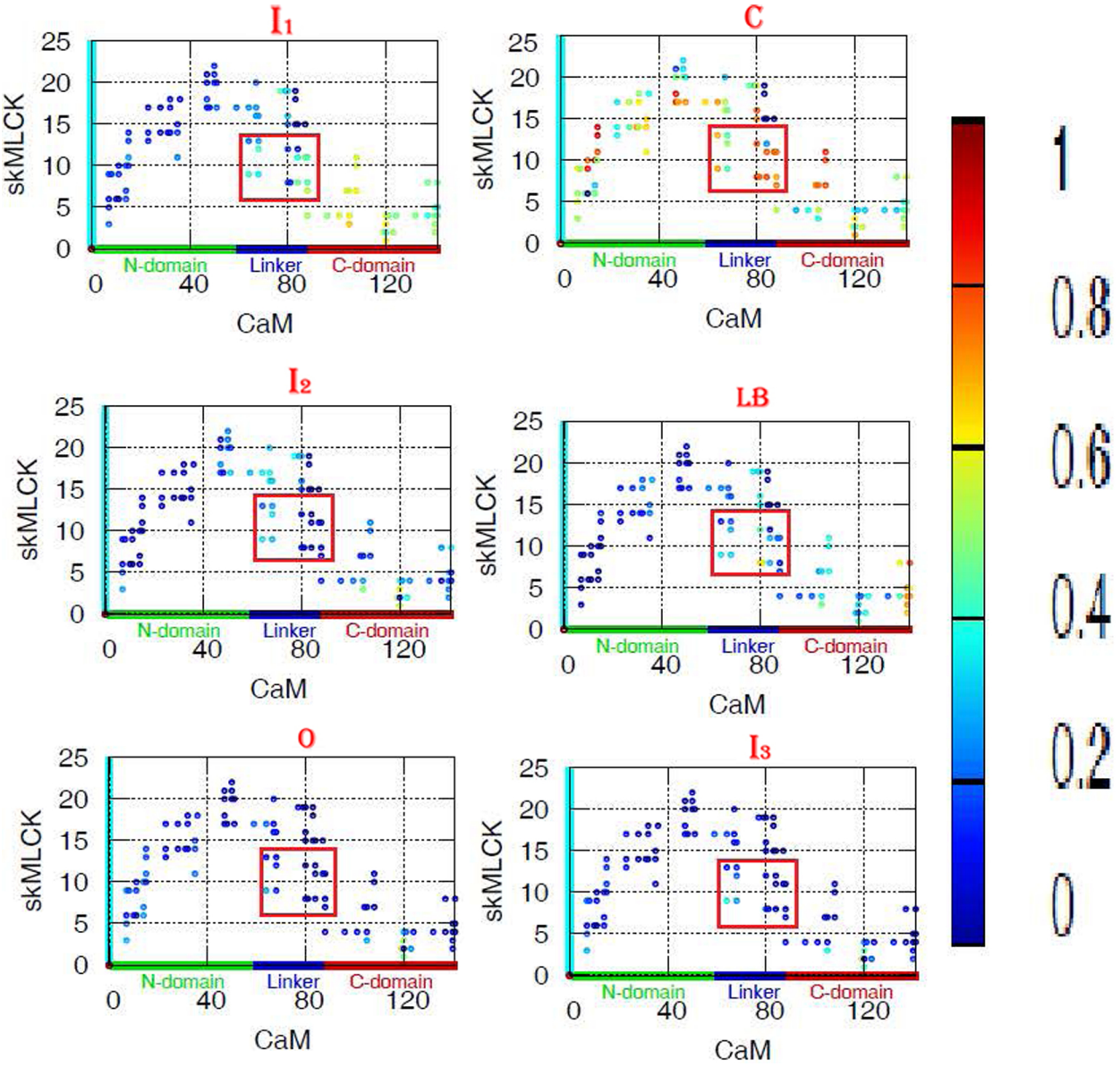

Fig.S 4: The native contacts probability map for $Q_{binding}$ in different states and regions are shown. We have found the "LB" is a partly binding and partly closed state comparing to the free open state. From "LB" to "C", the C-terminal of skMLCK will bind to the N-domain of Ca^{2+}-CaM

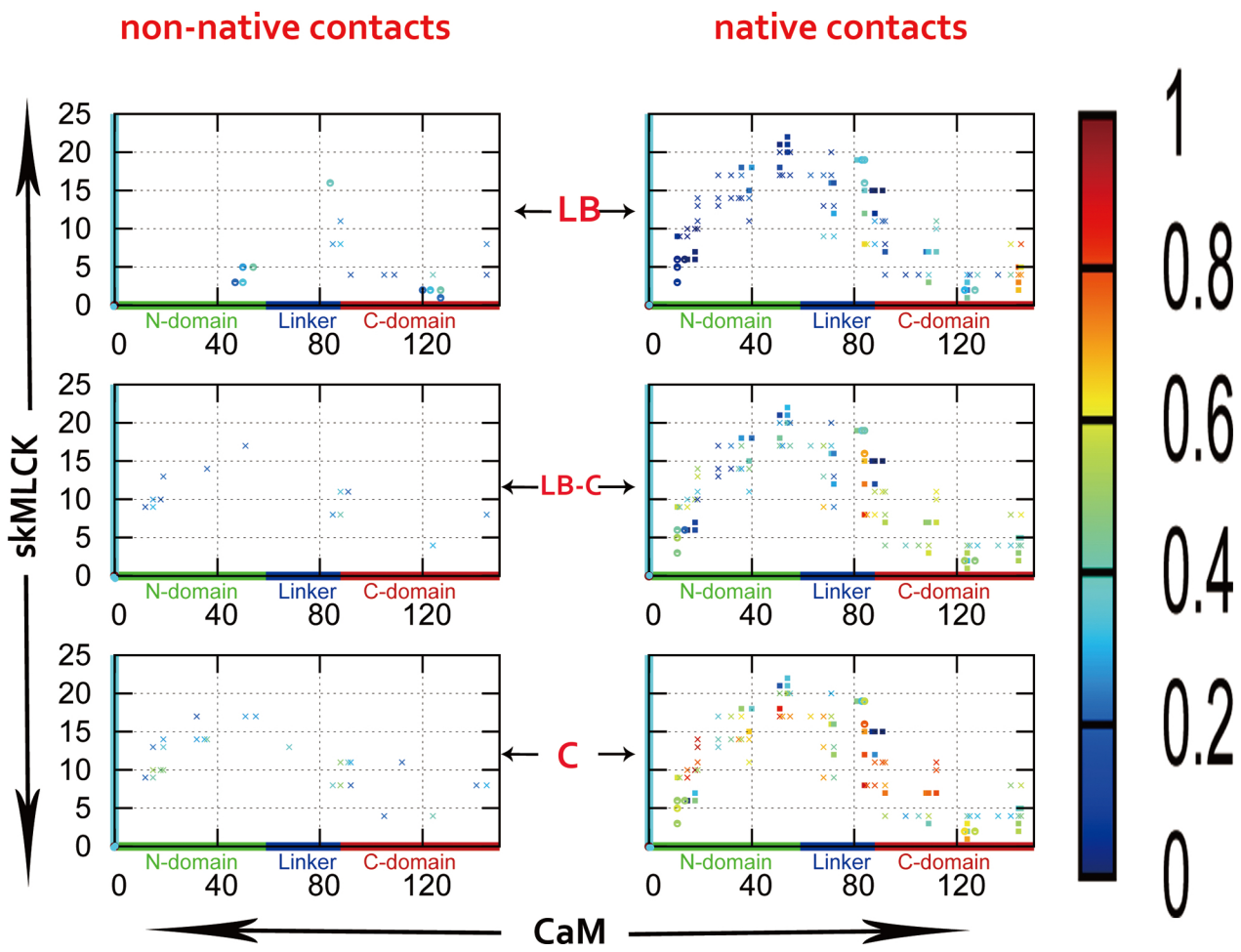

Fig.S 5: The non-native and native contact map for contact each state. The cross stands for hydrophobic contact, dots stands for electric contact and the square stands for plain interaction. The X-axis and Y-axis stand for the residue of CaM and skMLCK, respectively. The "LB-C" stands for the intermediate which is located on the middle of the pathway from LB to C.

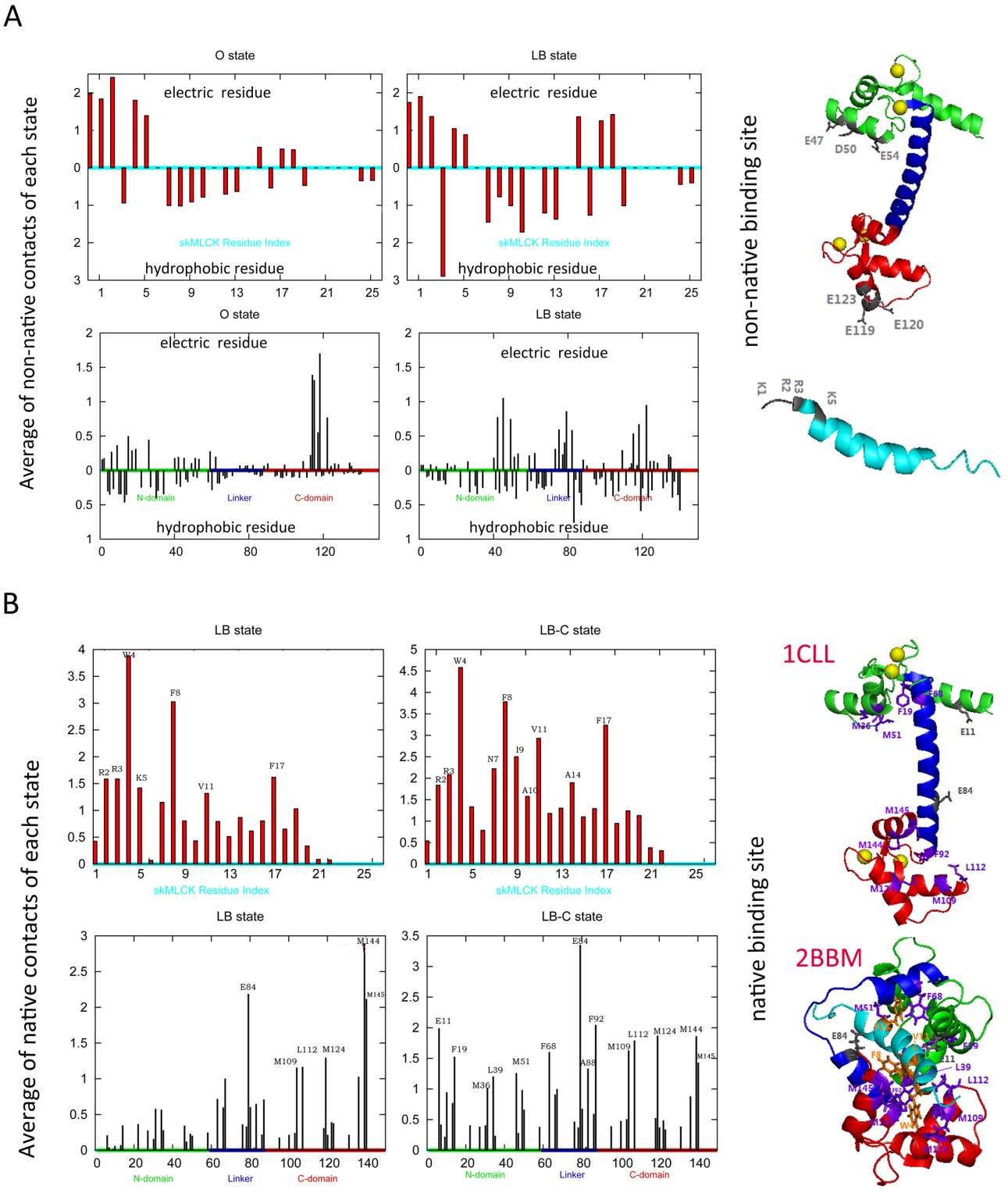

Fig.S 6: (A)The average of non-native contact map for each state. The X-axis stand for the residue of CaM or skMLCK. In each subfigure, the positive and negative Y-axis stand for the average of non-native electric and hydrophobic contact, respectively. Additionally, the key non-native binding site in the O and LB state are shown in PDB structure. (B) Average of native contacts of each residue in skMLCK and CaM for LB and LB-C state are shown. In addition, the residues which have high value of average of native contacts in the intermediate LB and LB-C are shown in PDB structure. The 1CLL and 2BBM are reference structure for the O and C state, respectively.

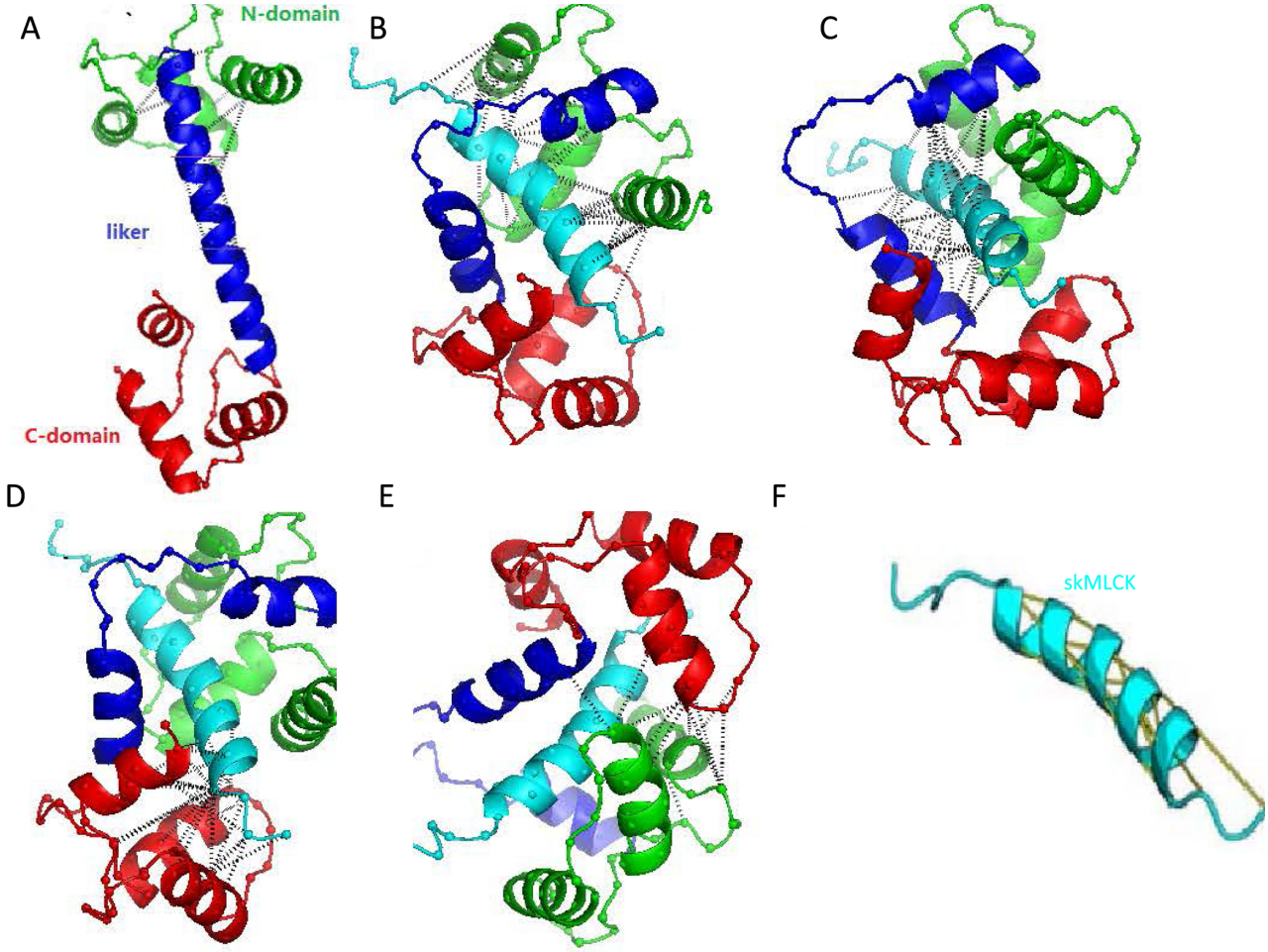

Fig.S 7: The special contacts for the mixed contact map. The contact map gives all possible interactions between pairs of residues in a given structure. In our work, the contacts are defined by contact map obtained with Contacts of Structural Units(CSU) algorithm[1].

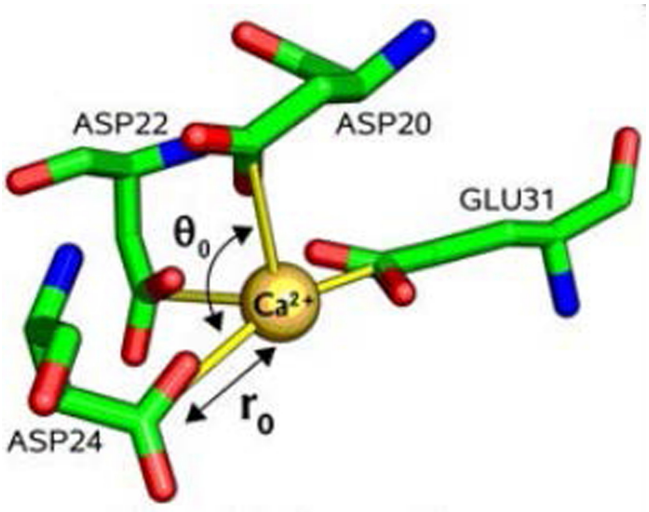

Fig.S 8: Bond interaction between Ca^{2+} and CaM. Each Ca^{2+} ion represents as a bead is located at the active site of EF-hand by attaching to the nearest negatively charged residues. We represented one example of Ca^{2+} tethering to ASP20, ASP22, ASP24 and GLU31. The localization restraint is realized by spring-like potential in the distance and angle in native structure. The other interactions involving Ca^{2+} are only electrostatic.

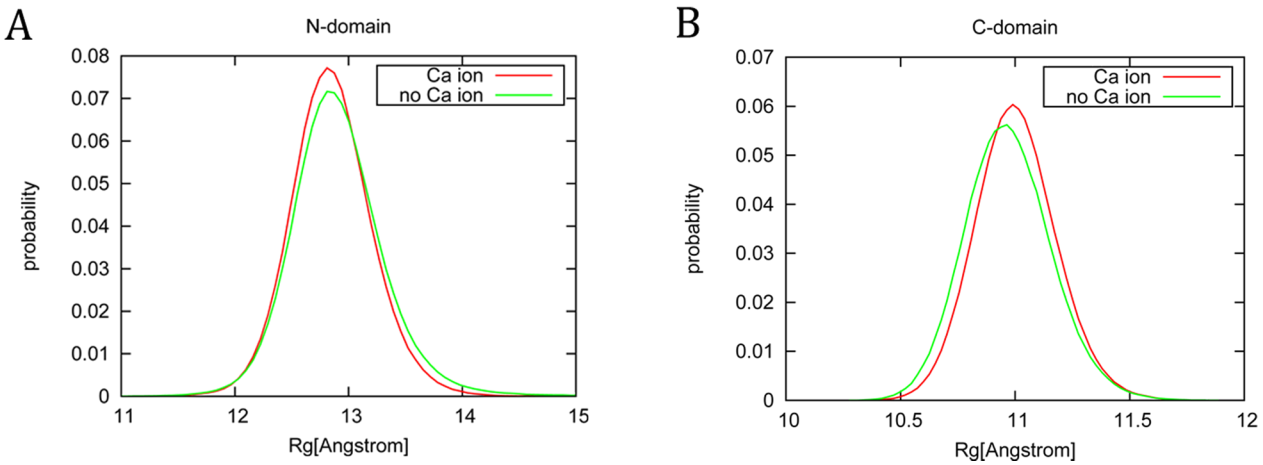

Fig.S 9: (A) The distributions of the radius of gyration(Rg)of the N-domain of CaM under the simulations with and without Ca^{2+}, respectively. (B)The distributions of the radius of gyration(Rg)of the N-domain of CaM. The peak of Rg distribution is smaller in the C-domain than in the N-domain. This is due to the C-domain of CaM is inherently less flexible than its N-domain[12]. The Rg distribution of N-domain indicates that the N-domain is more compact upon the binding of Ca^{2+}. The contrary is the case of C-domain. This is may due to the "cracking" or local partial unfolding of C-domain during the binding of Ca^{2+}[12].

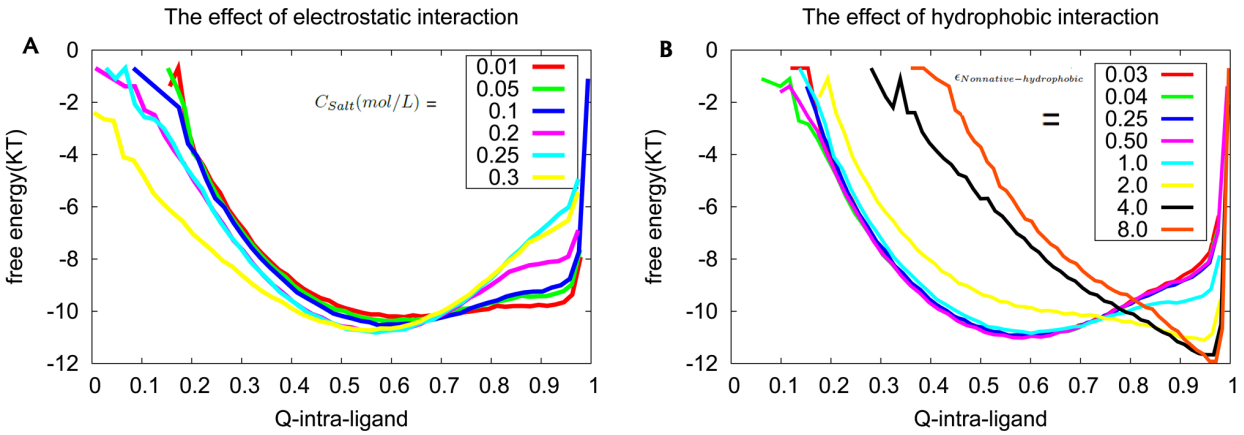

Fig.S 10: (A) and (B) are the one dimensional free energy landscape of "Q-intra-ligand" in different C_{Salt} and $\epsilon_{Nonnative-hydrophobic}$, respectively. The fraction of intra-native contacts of skMLCK(Q-intra-ligand) is to monitor the helicity(conformational change) of skMLCK. The $\epsilon_{Nonnative-hydrophobic}$ is the parameter that determines the strength of L-J potential of the hydrophobic nonnative contacts in the Hamiltonian Energy(SI), altering the strength of the nonnative hydrophobic interaction between the CaM and the skMLCK. The C_{Salt} is the salt concentration in Debye-Huckel model, which determines the strength of the electrostatic interaction.

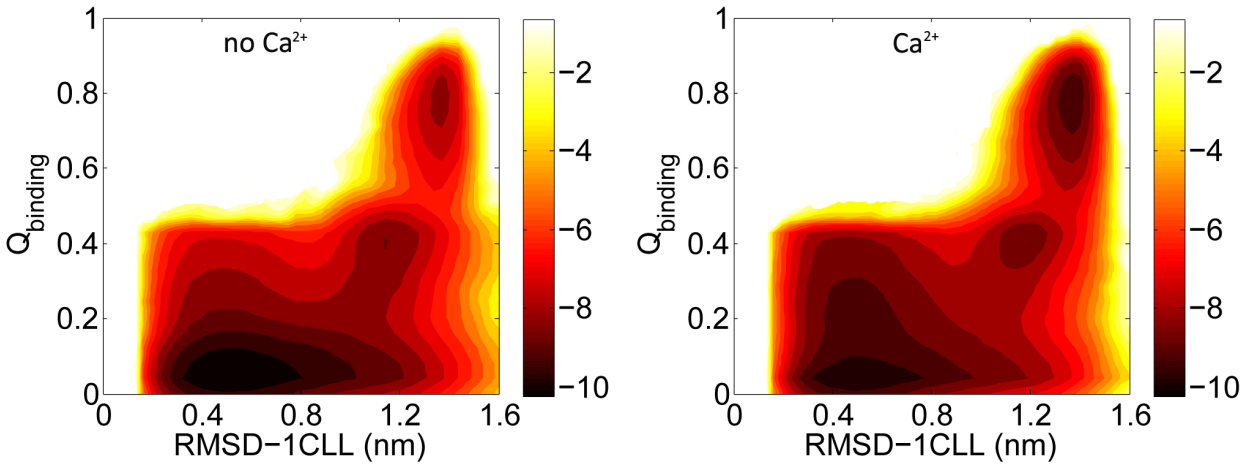

Fig.S 11: The effect of the Ca^{2+} on the landscape. Comparing the landscapes of the Ca^{2+}-CaM binding to skMLCK with and without Ca^{2+}, we found the "C" and "LB" are more stable in the presence of the Ca^{2+}.

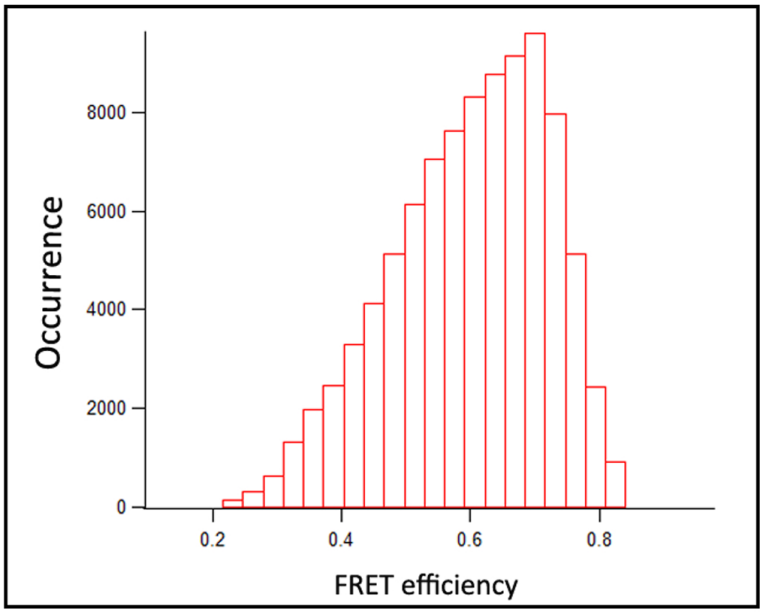

Fig.S 12: The figure showed the distribution of FRET efficiency which was analyzed to reveal the distance of fluctuations between the N-terminal domains of CaM[4]. The broad distribution of FRET efficiency revealed that there were intermediate states in the interaction between CaM and its target.

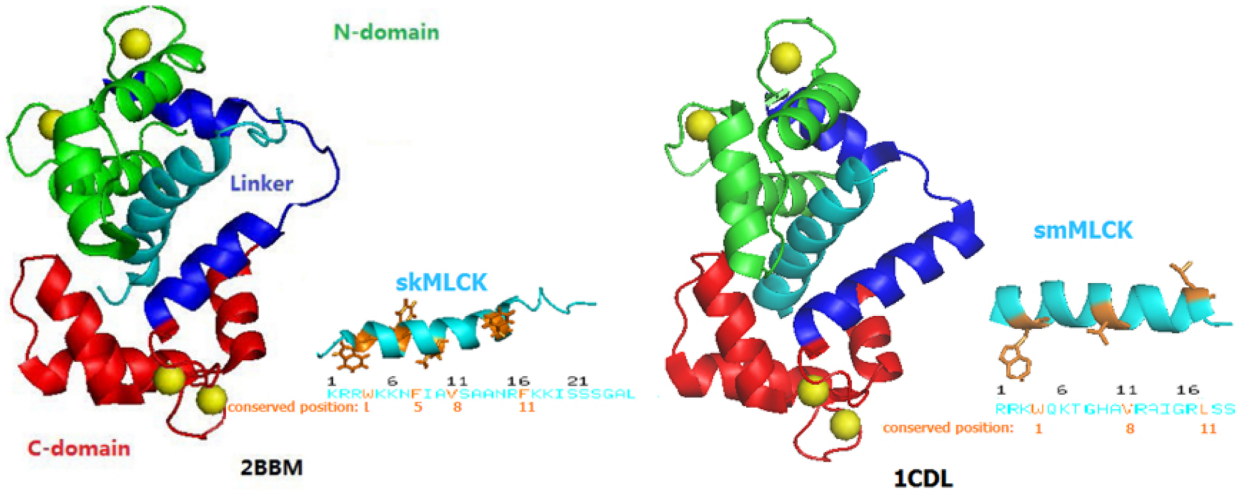

Fig.S 13: Comparing PDB structures of 2BBM and 1CDL. The reason we select smMLCK as the object comparing with the skMLCK is that the smMLCK is termed 1-8-14 by it conserved position of hydrophobic residues. The smMLCK lack the conserved position of hydrophobic "5" relative to skMLCK, which is proved to be important to form the "LB" state in the simulation of skMLCK.

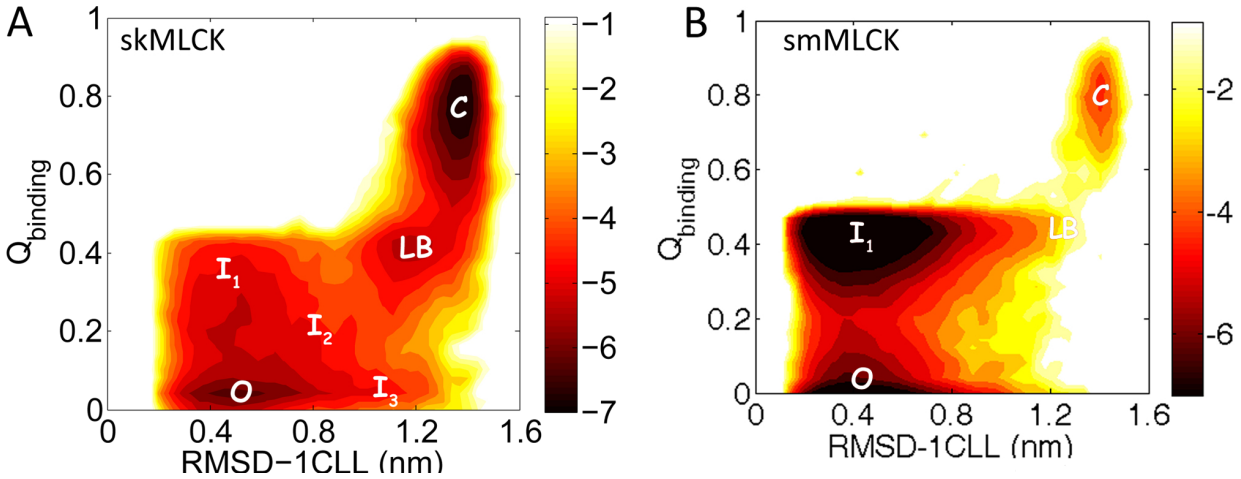

Fig.S 14: (A) The two dimensional contour map free energy of landscape in "RMSD-1CLL" and "$Q_{binding}$" for skMLCK, we adopt the root mean square deviation(RMSD-1CLL) relative to structure 1CLL to monitor the structural change of Ca^{2+}-Calmodulin and use the fraction of native contacts between Ca^{2+}-CaM and skMLCK($Q_{binding}$) to monitor their degree of freedom. We identified three stable states and three unstable regions in the landscape. There are three stable states in the landscape. Except the "O" and "C" states, which respectively indicated the open and close states, there is a stable intermediate "LB" state in the landscape. We found that all three pathways go through the intermediate "LB" state from the transition of open to close states. Each pathway respectively passes through the unstable regions "I1","I2","I3" before getting to the "LB". (B) The landscape for the binding between the CaM and smMLCK, all simulation parameters and situation for sampling this landscape is same with the parameters for the simulation of skMLCK. We found the very little state which is similar with "LB" is captured and the state which is similar with "I1" is more stable in the landscape of the binding between the CaM and smMLCK. The smMLCK more tends to select a pathway similar with the atypical "induced-fit' relative to the mixture mechanism for skMLCK.

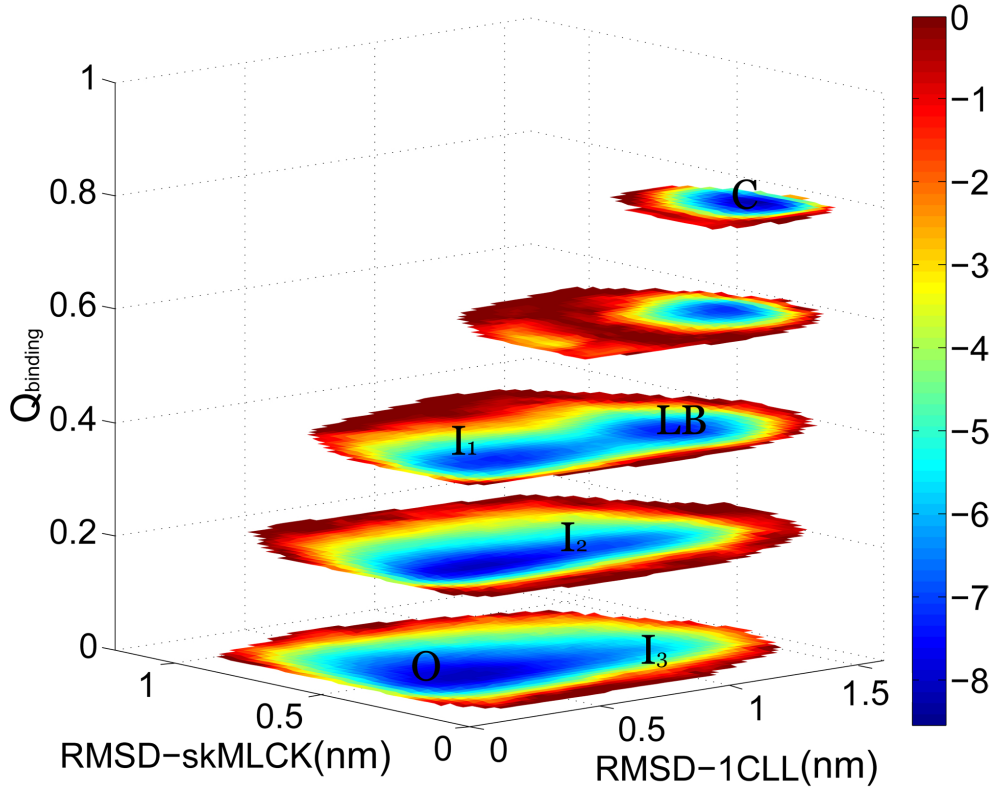

Fig.S 15: The free energy landscape along "RMSD-1CLL","$Q_{binding}$" and RMSD-skMLCK. We identified three stable states and three unstable regions in the landscape. Besides the "O" and "C" states, corresponding to "open" and "closed" states respectively, there is a stable intermediate "LB" state . We found that all three pathways go through the intermediate "LB" state from the transition of open to close states. Each pathway respectively passes through the unstable regions "I1","I2","I3" before getting to the "LB". The distribution of "RMSD-skMLCK" for "C" state is narrower and smaller than the "O" state, "I3" state, and the distributions for "I1" state and "I2" state are in the middle. This indicates that the skMLCK binding peptide gradually forms α-helical structures from random coil along the binding to CaM.

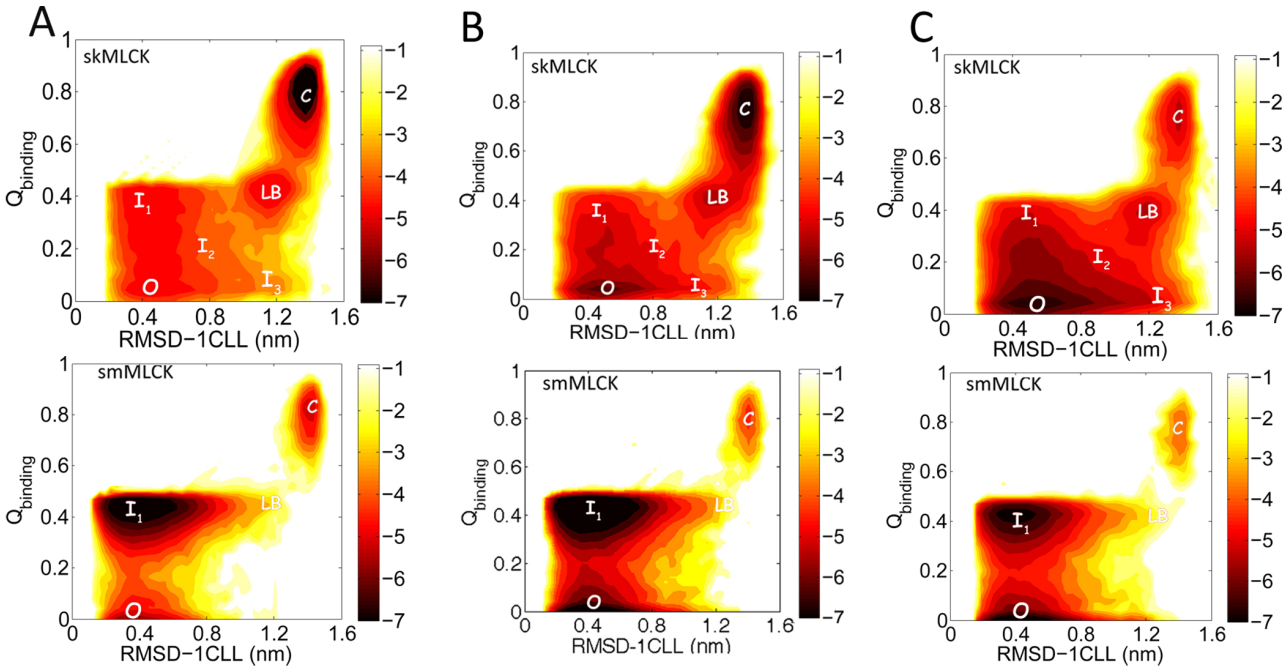

Fig.S 16: The binding free energy landscapes in different temperatures are shown for skMLCK and smMLCK. The temperatures are $1.45\epsilon_{LJ}$, $1.5\epsilon_{LJ}$ and $1.55\epsilon_{LJ}$ for subfigure A, B and C, respectively. We find the high temperature favors the unbinding state and the low temperature favors the binding state.