

The non-canonical interaction between calmodulin and calcineurin contributes to the differential regulation of plant-derived calmodulins on calcineurin.

Bin Sun¹, Xuan Fang¹, Christopher N. Johnson^{2,3}, Garrett Hauck²,
Jonathan P. Davis² and Peter M. Kekeneshuskey^{*,1}

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

² Department of Physiology and Cell Biology, The Ohio State University,
Columbus, OH, USA 43210

³ Department of Chemistry, Mississippi State University, Starkville MS, 39759

* Corresponding author. E-mail address: pkeneshuskey@luc.edu (P.
Kekeneshuskey)
June 18, 2021

Abstract

Calmodulin (CaM) is an important Ca^{2+} signaling hub that regulates many protein signaling pathways. In recent years, several CaM homologs expressed in plants have been shown to regulate mammalian targets and they are attractive for gene therapy. However, the molecular basis of how the CaM homologs mutations impact target activation is unclear, which limits efforts to engineer their functional properties. To understand these mechanisms, we examined two CaM isoforms found in soybean plants that differentially regulate a mammalian target, calcineurin (CaN). These CaM isoforms, sCaM-1 and sCaM-4 share $> 90\%$ and $\sim 78\%$ identity with the mammalian CaM (mCaM), respectively, activate CaN with comparable or reduced activity relative to mCaM. We used molecular simulations and experimental assays to probe whether calcium and protein-protein binding interactions are altered in plant CaMs relative to mCaM as a basis for differential CaN regulations. We found that the two sCaMs' Ca^{2+} -binding properties such as coordination and affinity are comparable to mCaM. Further, the binding of CaM to the CaM binding region (CaMBR) in CaN is also comparable among the three CaMs, as evidenced by calculated binding free energies and experimental measured EC_{50} [CaM]. However, mCaM and sCaM-1 exhibited stronger binding with a secondary region of CaN's regulatory domain that is weakened for sCaM-4. This secondary interaction is likely to affect the turnover rate (k_{cat}) of CaN based on our modeling of enzyme activity and is consistent with our experimental data. Together, our data show how plant-derived CaM variants can alter target activation through interactions beyond Ca^{2+} -binding and canonical CaMBR binding, which may extend beyond the mammalian CaN target.

2 Introduction

Calmodulin (CaM) is a 16.7 kDa Ca^{2+} sensor ubiquitously expressed in eukaryotic cells [1] with an invariant sequence. CaM regulates at least 300 targets in cellular processes including muscle contraction, proliferation and immune response [2, 3]. CaM is also essential to plant physiology in response to environment stimuli [4]. In contrast to vertebrates, plants have different CaM isoforms [5, 6]. Two soybean CaM isoforms, sCaM-1 and sCaM-4, share $\sim 90\%$ and $\sim 78\%$ identity with mCaM, respectively, and have been extensively studied [7–9]. Interestingly, these two plant CaM variants show differential or even reciprocal regulation abilities on human protein targets [10, 11]. As an example, for calcineurin (CaN), a phosphatase, the sCaM-1 exhibits similar ability to activate the protein target as human CaM while sCaM-4 has significantly reduced ability [10]. CaN is a ubiquitously expressed serine/threonine phosphatase in all human tissues [12] and is regulated by CaM in a Ca^{2+} -dependent manner. Disregulation of CaN is related to diseases such as autoimmunity [13], ventricular hypertrophy [14] and Alzheimer’s disease [15]. The positive correlation between sequence similarity and activation of CaN of these soybean CaMs with mCaM makes them good model systems to study the effects of CaM mutation on target activation.

CaM mutations found in human have been reported to disturb Ca^{2+} -handling in cardiac cells [16]. To date the underlying molecular basis remains unclear. Several mechanisms may explain altered functions in CaM variants. CaM’s Ca^{2+} binding property undoubtedly plays an important role in its target activation. It has been widely accepted that Ca^{2+} binding promotes CaM’s N/C domains to expose the hydrophobic patch for target binding [17–19]. Although apo-CaM can bind targets [20], presence of Ca^{2+} facilitates the formation of native complex [21], or even reverses CaM’s activating role [22]. Besides Ca^{2+} -binding, the association between CaM and the CaM binding region (CaMBR) of targets is also important. A single mutation in the CaMBR region could abolish CaM binding [23]. Lastly, interactions between CaM and target region beyond the CaMBR can affect target regulation as well. For example, the ryanodine receptor (RYR) isoforms are differentially regulated by CaM, albeit bearing the same CaMBR region [24], suggesting that isoform-dependent region beyond CaMBR contributes to regulation. Another example is the activation of calcineurin (CaN) by CaM, in which a ‘distal helix’ region C-terminal to CaMBR interacts with CaM to fully remove the autoinhibitory domain of CaN [25]. Namely, studies [25–29] have provided a model that explains the molecular interactions between CaN and CaM that are vital to CaN’s complete activation (Fig. 1) [30]. In the model, Ca^{2+} -saturated CaM binds the CaM binding region (CaMBR) (A391-R414) on the CaN regulatory domain, followed by a secondary interaction between CaM and a ‘distal helix’ (DH, K441-K466) that ultimately removes the autoinhibitory domain (AID) from the CaN catalytic domain. These mechanisms could be utilized by CaM variants to alter target regulations.

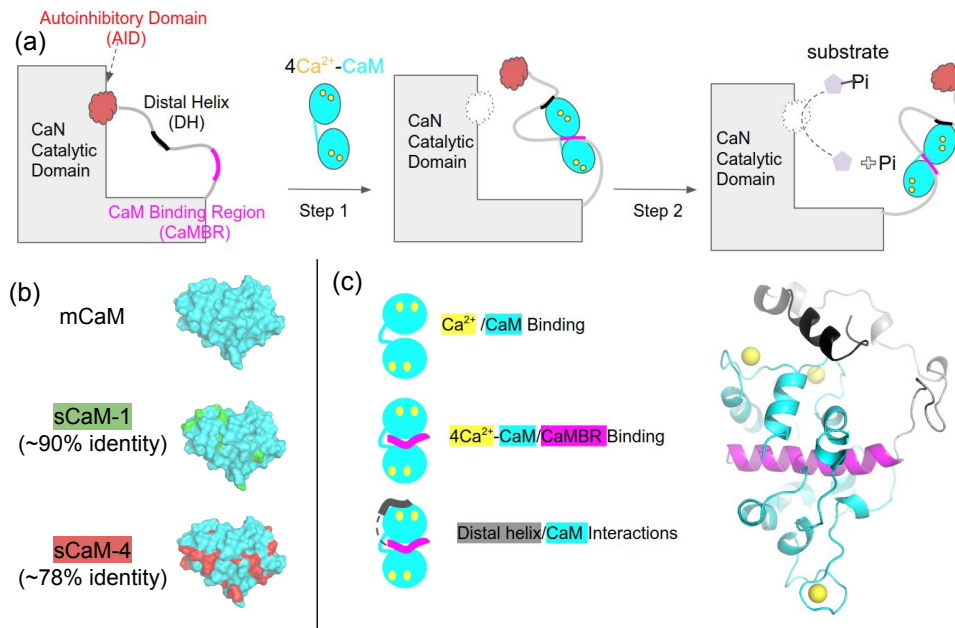


Figure 1: (a) Proposed CaN activation model by CaM [30]. Step 1, Ca²⁺-saturated CaM binds to the CaM binding region (CaMBR) and the distal helix (DH) motif of CaN, which displaces the autoinhibitory domain (AID) from its catalytic site. Step 2, Entry of substrate to CaN's catalytic site enables dephosphorylation of substrate. (b) Two soybean CaM isoforms sCaM-1 and sCaM-4 share ~90% and ~78% sequence identity with mammalian CaM (mCaM). The unique residues are colored green and red for sCaM-1 and sCaM-4, respectively. These two sCaMs have differential regulations on human CaN. Detailed sequence comparison is in Fig. S1. (c) The three molecular interactions we examined in this study to explore the origins of the two sCaMs' different activating abilities on human CaN.

In the present study, we therefore used CaN/CaM as a model system to explore the molecular basis of CaM target activation with different variants. We explored the origination of the two sCaM's differentiated regulation abilities on human CaN. We hypothesised that the soybean CaM variants use one of these mechanisms to alter mammalian CaN activity: 1) Ca²⁺ binding properties to CaM, 2) CaM binding to CaMBR of CaN, and 3) region of the CaN regulatory domain distal to the CaMBR, the 'distal helix' motif. We found that the two sCaMs maintain comparable Ca²⁺ and CaMBR binding abilities to mCaM. However, the distal helix/CaM interaction patterns are different. This implicates the distal helix interaction as the cause for the different activation capabilities. We further pinpointed a hot-spot region on CaM that contributes most to the distal helix interaction difference between the two sCaMs and mCaM, This region lies in between residues N60-E87 and can be used in future CaM reengineering studies. Lastly, using a revised activation kinetic model, we found that CaM activates CaN by converting it to an activated form that exhibits altered kinetic properties. Altogether, our computational and experimental results showed that DH interactions are central to CaN activation by CaM.

3 Materials and methods

3.1 Calcineurin phosphatase assay using MUF-P substrate

The mCaM and sCaM-4 were overexpressed in *E. coli* and purified as described in [31] using phenyl sepharose. CaN was overexpressed from a pETagHisCN plasmid using *E. coli* BL21 (DE3) CodonPlus RIL cells. Cells were lysed by sonication, and CaN was purified using Ni-NTA and CaM-sepharose. Protein purity (>95%) was determined by Coomassie staining and SDS denaturing gel electrophoresis. Purified CaN fractions were pooled and dialyzed into 10 mM MOPS,

150 mM KNL, 1 mM TCEP at pH7.0. CaN and CaM concentrations were determined by UV-VIS spectroscopy. Experimental sample conditions were 2 mL of 150 CaN in assay buffer comprised of 200 mM MOPS, 150 mM KCl, 3 mM MgCl₂, 2 mM EGTA at pH 7.0. The MUP concentrations was set to 0, 100, 200, 400 and 600 μ M for substrate titrations of CaN and held at 100 μ M for EC₅₀ experiments. Assay buffer EGTA concentration was calibrated based on a dose response of CaM tyrosine fluorescence (corresponding to CaM-C domain) against Ca²⁺. PerkinElmer spectrometer parameters were as follows: excitation 365 nm, emission 445 nm, data sampling interval 0.2 second, temperature = 20 °C. CaM titrations ranged from 5 nM to 3 μ M. Ca²⁺ titrations ranged pCa9 to pCa4. A minimum of 3 min. of raw fluorescence data were collected for each titration point and fit to a linear line using Kelidograph. The slope values were normalized and plotted against the titrant (CaM or Ca²⁺). Data was fit to a 4 parameter dose response curve using Prism. The resulting EC₅₀ values were plotted in prism. Statistical significance was determined (95% confidence, p < 0.05) using an Anova test.

3.2 Kinetic Analyses

To investigate the activation effect of CaM on CaN, we first fit the Michaelis-Menten model to the kinetic data. It should be noted that due to the inner filter effect of MUF at high concentrations (data not shown), the kinetic assays were done at low concentrations. To estimate the kinetic parameters, the double reciprocal form of the model was fit to the data:

$$\frac{1}{v} = \frac{K_m}{V_{max}} [S] + \frac{1}{V_{max}} \quad (1)$$

This Michaelis-Menten model is based on the assumption that activator binds to both the enzyme and the enzyme-substrate complex. However, in the activation of CaN by CaM, multiple molecular interactions are involved and thus a more sophisticated model is necessary.

To gain further insight into the activation mechanism of CaM and compare the kinetic properties of different CaM isoforms, we proposed a revised activation model for the data (Fig. 2). In this revised model, we assume Michaelis-Menten mechanism for isolated CaN. However, upon binding to CaM, CaN is then converted into the activated form that has different kinetic properties as specified by the α and β factors. The fitting of the revised model was done in two steps. For the CaM free data set, the K_m and V_{max} was determined by fitting the double reciprocal form of the Michaelis-Menten model to this data set. We then assumed that mCaM and the soybean isoforms have the same K_a ($0.038 \pm 0.0048 \mu$ M), which we validated in this study via experimental assays. Together with the K_m and V_{max} determined from the previous step, the reciprocal form of the revised model (Eq. 2) was fit to the CaM and sCaM data sets with K_m , V_{max} , and K_a as shared and fixed parameters (K_m and V_{max} are intrinsic properties of CaN) while α and β were used as free parameters.

$$\frac{1}{v} = \frac{1}{V_{max} S \left(\frac{1}{K_m} + \frac{\alpha A}{\beta K_a K_m} \right) \left(1 + \frac{S}{K_m} + \frac{A}{K_a} + \frac{A S}{\beta K_a K_m} \right)} \quad (2)$$

where A is CaM, K_a is the binding constant of CaM to CaN, α is scales of V_{max} and β is scales of K_m .

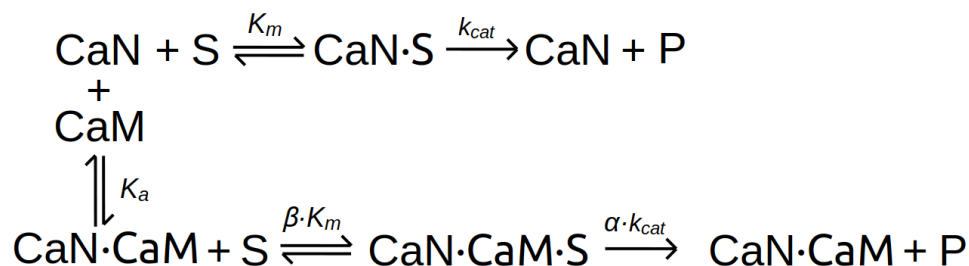


Figure 2: A kinetic model that considers multiple molecular interactions involved in CaN’s enzymatic activity after being activated by CaM.

3.3 Rosetta Comparative Modeling of sCaM-1 and sCaM-4 structures

The available structural information for sCaM-1 and sCaM-4 deposited in PDB databank are for isolated N- or C- domain structures [11] (PDB ID: 2RO8/2RO9 for sCaM-1 and 2ROA/2ROB for sCaM-4, respectively). Although there is one complete sCaM-4 structure in complex with vacuolar calcium ATPase BCA1 peptide deposited with PDB ID 2L1W [7], the binding mode of target peptide is significantly different relative to the mammalian CaM/CaMBR structure. For the sCaM structure, the N- and C-domain binds to the two ends of peptide, as opposed to the conventional ‘wrap round’ manner in which N-/C- domain collapse around the middle part of target peptide. Therefore, we built complete sCaM-1 and sCaM-4 structures that bind CaMBR in the same manner as mammalian CaM. The complete structures of sCaM-1 and sCaM-4 were built using the Rosetta comparative modeling approach [32]. For each model, the mammalian CaM/CaMBR complex crystal structure (PDB ID 4Q5U) as well as the N-/C- domain structures were used as templates (the weights of each template are 1.0, 0.8 and 0.64, which are the default values in the RosettaCM hybridize mover). The HYBRIDIZE function of Rosetta was used and the specific protocol and flags are given in Sect. S2.1. For each soybean CaM, the highest ranked structure generated by Rosetta comparative modeling was subject to $3 \times 2 \mu\text{s}$ all-atom MD simulation refinement.

3.4 All-atom MD simulations to refine the modeled sCaM structures

All-atom MD simulation was performed to refine the Rosetta modelled soybean CaM structures. The ff14SB [33] force field was used for protein atoms. The system was solvated in TIP3P [34] waterbox with the distance between protein to waterbox wall being 14 Å. K^+ and Cl^- ions were added to maintain 0.15 M salt concentration. The system was subjected to an energy minimization, for which all atoms except hydrogens, water and KCl ions were constrained by the *ibelly* functionality. The cutoff value for non-bond interactions was set to 10 Å. A 2 fs timestep was chosen, as SHAKE [35] constraints were applied on bonds involving hydrogen atoms. Two heating procedures were performed to heat the system from 0 to 300 K using the Amber18 SANDER.MPI engine [36]. In the first heating stage, the *ibelly* function was used to keep all system except the water box and KCl ions fixed. The water box was heated to 300 K over a 100 ps interval under the NVT ensemble. For the second heating stage, the entire system was heated from 0 to 300 K over 500 ps under the NPT ensemble, for which the backbone atoms were constrained by a harmonic potential (force constants of $3 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$). Thereafter, an additional 1 ns equilibrium stage was conducted at 300 K under the same constraints, but with a reduced force constant of $1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. The Langevin dynamics was used during the simulation. These equilibrium simulations were followed by triplicate $2 \mu\text{s}$ production-level molecular dynamics (MD) simulations. The simulation length is sufficient to refine the Rosetta modelled structure and give reliable sCaM structure. This is based on the reports that the timescale of conformational fluctuation in the MD refinement of modeled structure is approximate $1 \mu\text{s}$ for systems with mean residue number around 120 [37], which is approximate the length of CaM.

3.5 Docking distal helix structure to CaMs

3.5.1 Replica exchange molecular dynamics (REMD) simulations of isolated distal helix region

REMD was performed on the distal helix fragment (residues K441-K466 of CaN, amino acid sequence is shown in Fig. S6) to sample representative conformations of the isolated DH. 12 replicas of the system were constructed and the temperature range of the REMD was 270 to 454.72 K, which was calculated based on the protocol proposed in [38]. The temperature range ensures that the exchange probability between neighbouring replicas is approximately 0.4 as proposed in [39]. The system was parameterized with ff99SBildn force field [40] coupled with an generalized Born implicit solvent model. The detailed REMD simulation procedure has been described in [30]. The production run of the REMD was 100 ns and clustering analysis were performed on the trajectory from 297 K. Representative structure of the most populated cluster was selected for docking to CaM structures. The selected DH conformation is shown in Fig. S6.

3.5.2 Docking DH to CaM surface and extensive MD refinement

The representative structure of the DH generated by REMD simulations was docked to surface of sCaM-1, sCaM-4 and mCaM via the ZDOCK webserver [41]. The CaM structures that were used for the docking are from the MD refinements of Rosetta modeled structures. Specifically, clustering analysis were first performed on the MD trajectories and the representative structure of the most populated cluster was selected. For mCaM, one representative structure was chosen and for the two sCaMs, two representative structures were chosen for each case. The procedure of clustering analysis and identifying the two representative structures is detailed in Sect. S2.3. After preparing the DH and CaM structures, the DH was docked to CaM following the docking procedure in [30]. Four putative binding sites were pre-defined on CaMs that locate closely to the grooves formed by helix bundles, as these places are more likely to accommodate helix-helix interactions [42]. The selected four putative sites covered most of the CaM's solvent-accessible surface. The residues in Table S2 from each site on CaM were specified as contact residues during docking to ensure the initially generated distal helix poses are near the putative binding sites. The detailed docking procedure was reported in [30].

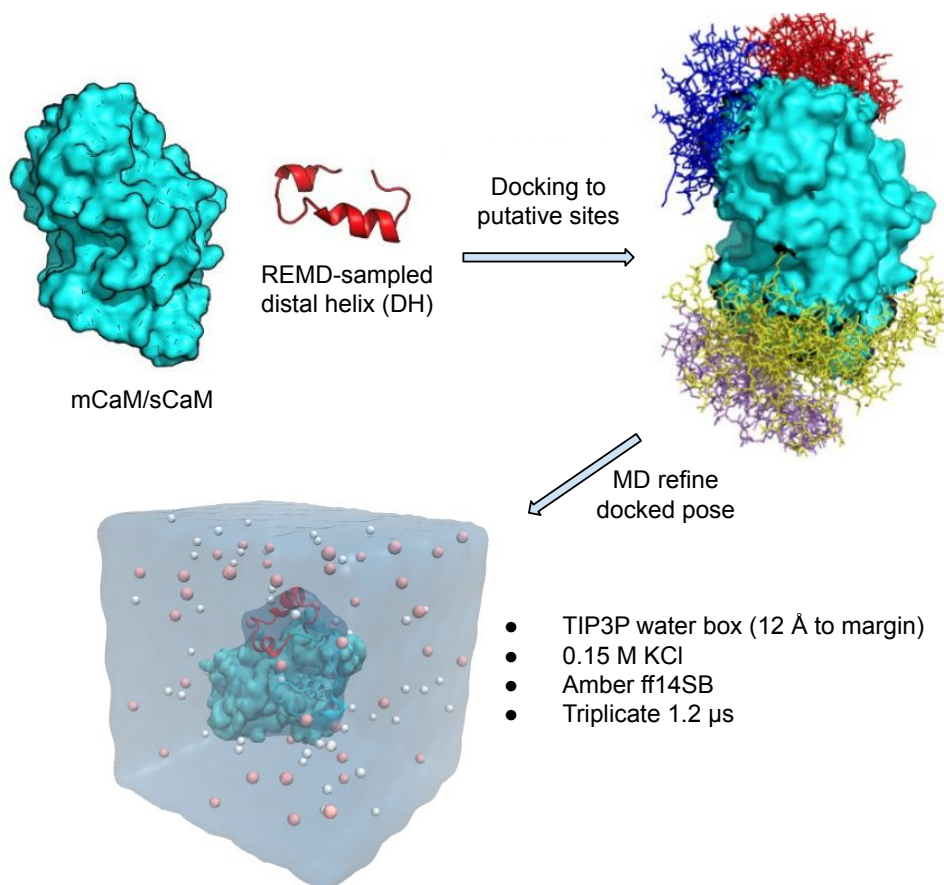


Figure 3: Scheme of docking the REMD-sampled distal helix (DH) conformation to CaM's surface via the ZDOCK webserver [41]. For each CaM case, four putative binding sites on CaM that are localized to the grooves between helices were pre-defined as contact site during the docking. The putative site residues of CaMs are listed in Table S2. For each site, the highest-scored pose was further subjected to all-atom MD refinement.

The highest-scored CaM/DH interaction pose at each site was selected for further MD refinement with triplicate 1.2 μs runs. The detailed procedure of docking and MD refinement has been reported in [30]. In summary, 72 μs refinement MD simulations were performed with 28.8 μs for each sCaM-1 and sCaM-4 and 14.4 μs for mammalian CaM. All molecular dynamics simulations were performed via Amber16 package [43] and are summarized in Table S1.

3.6 Analyses Methods

Clustering analysis, root mean squared deviations (RMSD)/root mean squared fluctuations (RMSF) calculations, hydrogen bonds, and secondary structure analysis were performed via CPPTRAJ [44]. The reference structure used for these analyses was the CaM/CaMBR crystal structure (PDB ID: 4Q5U [27]). Secondary structure for each residue was calculated using CPPTRAJ with the Define Secondary Structure of Proteins (DSSP) algorithm [45].

3.6.1 Projection of MD sampled conformations onto 2D plane using the sketch-map dimensionality reduction method

Sketch-map (<https://github.com/cosmo-epfl/sketchmap>) is a non-linear dimensionality reduction method that we used to visualize CaM/DH complex conformation space sampled in the MD simulations [46, 47]. Each frame from the MD trajectories was represented by a point

in the high-dimensional space $\vec{X}_D = (D_1, D_2, \dots, D_N)$ and D_1, D_2, \dots, D_N are the minimums distance between C_α atoms of CaMBR and distal helix. In the present study, N was set as 32: D_1 to D_{14} represents the minimum distance of the 14 residues in CaMBR (E394-A407) to the distal helix. Namely, for residue i , we first calculated its distance to each residue in the distal helix and assign the smallest value to D_i . And D_{15} to D_{32} are the reverse minimum distances, namely, the minimum distance of each residue in the distal helix (S438-D455) to CaMBR residues. This definition of high-dimensional points ensures that the relative conformation information between distal helix to CaM/CaMBR is entirely encoded in the coordinates of high-dimensional points, as illustrated in [48]. These high-dimensional points were projected onto a 2D plane as described in [48] to generate the sketch-map representation of MD sampling.

4 Results and Discussion

4.1 Experimental assays revealed that sCaM-1 and mCaM exhibited stronger activation of CaN than sCaM-4.

We anticipated that sCaM-1 would more strongly activate human CaN relative to sCaM-4, based on the form's higher sequence similarity to mCaM. Indeed, a study by Cho *et al.* demonstrated that sCaM-1 exhibited a stronger activation of CaN relative to sCaM-4, based on their measurements of CaN activity as a function of CaM with 100 μ M MUF-P as a substrate [10]. To investigate the kinetic mechanisms by which the CaM variants differentially activate CaN, we performed pMUF substrate titrations of CaN alone as well as CaN with different CaM isoforms. As shown in Fig. 4, the presence of CaM (regardless of the isoform relative to isolated CaN) significantly increased the activity of CaN. Consistent with Cho *et al.* observations, sCaM-1 enhanced CaN's activity to a similar extent as mCaM while sCaM-4 did so to a much lesser extent.

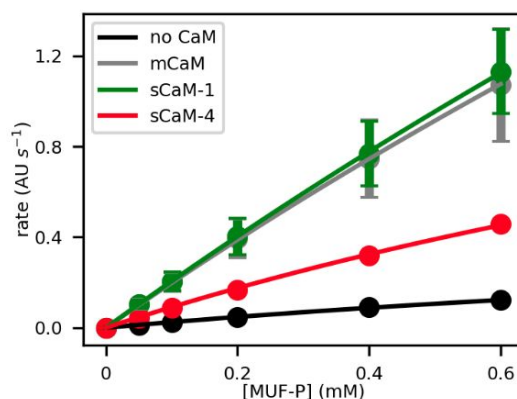


Figure 4: **Experimental CaN activity.** Substrate titrations of CaN alone as well as CaN with different CaM isoforms.

4.2 The different activation effects between sCaM-4 and mCaM/sCaM-1 are not caused by their Ca²⁺ binding properties and their CaMBR binding properties

We next sought to determine the molecular mechanisms behind these experimental trends *via* computational modelings, which included homology modeling, protein-protein docking and molecular dynamics simulations. As shown in Fig. 1a, the previous model for CaM/CaN activation includes 1) Ca²⁺ binding to isolated CaM. 2) binding of Ca²⁺-saturated CaM to the CaM binding region (CaMBR) of CaN and 3) the interaction of the distal helix (DH) motif from from CaN to CaM. We examined these interactions for the mCaM and sCaM isoforms in the activation of CaN.

Because the complete sCaM-1/4 structures complexed with CaMBR of CaN are not yet available, we performed homology modeling followed by extensive all-atom MD refinement to obtain structural models. We show in Sect. S2.2 that the modelled sCaM/CaMBR complex structures were reasonable as supported by converged RMSD values over the μ s MD time-scale, low fluctuations of backbone atoms ($< 2 \text{ \AA}$ RMSF for most part) and a more compact structure of sCaM-4 that was consistent with experimental observations (Fig. S3). Based on the modelled structures, we examined key molecular interactions that may contribute to the sCaM isoforms' differential activation of mammalian CaN.

4.2.1 Ca²⁺ binding property

Ca²⁺ plays a vital role in CaM's target regulation [49]. In this section we examined if the Ca²⁺-binding properties were similar among mCaM and the sCaM isoforms. CaM has four EF hands, two in each N-/C- domain. MD simulations indicated that the EF-hands in the modelled sCaM structures were stable as evidenced by the $\sim 3 \text{ \AA}$ RMSF values (Fig. S3). Not surprisingly, mCaM had the smallest RMSF of $\sim 1 \text{ \AA}$ at EF-hands. sCaM-1 had the largest RMSF among the three CaMs because the anti-parallel β -sheet between EF-hands was disrupted. We anticipated this was due to a structural organization to maintain Ca²⁺ affinity, similar to β -sheet reductions observed in parvalbumins upon Ca²⁺ binding [50].

Nonetheless, sCaM-1 maintained the same Ca²⁺ coordination number as the other two CaMs. Fig. S4(a,b) reports the integrated radial distribution functions (IRDF) of amino acid oxygens and water oxygens around Ca²⁺. In previous studies, we have shown intrinsic Ca²⁺ affinities in EF-hands tend to correlate with oxygen RDFs [50, 51]. In all CaM isoforms, the IRDF curves reached a plateau around 2.3 \AA , which indicates the average distance between Ca²⁺ and the coordinating oxygens RDFs were indistinguishable across all cases, which suggests similar Ca²⁺ binding

affinities. Interestingly, as shown in Fig. 5a, the number of amino acid oxygens and water oxygens varied somewhat: sCaM-1 had the lowest amino acid oxygen number around 5.5 for Ca^{2+} at the first EF-hand of N-domain and second EF-hand of C-domain but was compensated for by the highest water oxygen numbers around 2. In addition, the Ca^{2+} s bound at the C-domain of the two sCaMs had more water oxygens compared with mCaM. Nevertheless, the total Ca^{2+} coordination numbers were comparable, all were in the range of seven to eight oxygens (Fig. 5a), in agreement with the Ca^{2+} coordination number measured from other Ca^{2+} -binding proteins [30, 51–54] and in aqueous solution [55]. The one oxygen variation should have negligible impact on Ca^{2+} binding free energy in the EF-hand site as we demonstrated previously difference of ± 1 oxygen do not appreciably impact Ca^{2+} binding free energy in the EF-hand sites [50]. Since changes in Ca^{2+} affinity can enlist conformational changes outside of the EF-hands [50, 56], we experimentally measured the EC_{50} [Ca^{2+}] of mCaM and sCaM-4 in activating CaN (Fig. 6(b-c)). This was obtained by measuring CaN activity at varying Ca^{2+} concentrations with saturated CaM. The mean values of the EC_{50} for mCaM and sCaM-4 were 0.68 and 0.49 μM , respectively. Converting this EC_{50} to energy via $-RT \ln \text{EC}_{50}$, which implies 0.19 kcal/mol energy difference between mCaM and sCaM-4 at room temperature, and thus Ca^{2+} binding properties were comparable. These data together showed that the three CaM isoforms maintained comparable Ca^{2+} binding properties, which suggested the differential CaN activation ability was not caused by Ca^{2+} binding.

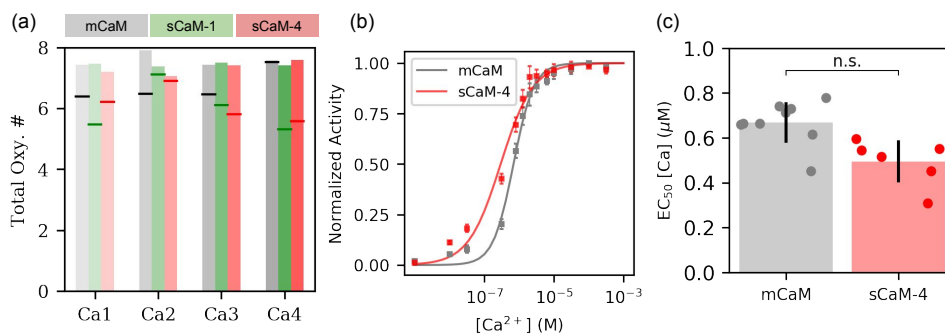


Figure 5: **Ca²⁺ binding properties.** (a) The total number of coordinating oxygens (amino acid oxygen plus water oxygen) around Ca²⁺. The two Ca²⁺s at the N-domain EF-hands were indexed as "Ca1" and "Ca2", and the Ca²⁺s at C-domain were indexed as "Ca3" and "Ca4". The horizontal lines within each bar indicate the number of amino acid oxygens. (b) Dependence of CaN activity on Ca²⁺ under saturating CaM condition. (c) Experimental EC₅₀ [Ca²⁺] for mCaM and sCaM-4. n.s. $p > 0.05$.

4.2.2 CaMBR binding property

We next examined if the binding between the CaMBR and CaM differed for the three CaM isoforms, as a basis for their different activities. We reported in Fig. 6a the contacts between CaMBR and CaM. Overall, the CaMBR/CaM interaction pattern shown in mCaM were largely maintained by the two sCaMs, especially the contacts from CaM helices formed by residues Q8-S17, T43-N53, K77-R90. In the simulations of all CaMs, the CaMBR exhibited small fluctuations with RMSF $< 2 \text{ \AA}$ (Fig. S3b), which suggested that the CaMBR was highly stable when bound to three CaMs. The largest difference was observed in the C-domain: both sCaMs exhibited contacts with the CaMBR via two helices through residues S101-M109 and D118-I125, while mCaM had negligible contacts with CaMBR at these two helices. However, the MMGBSA-calculated binding free energies between the CaMBR and CaM (Fig. 6b) were comparable across the three CaMs and were highly favorable: -128.8 ± 3.3 , -127.1 ± 2.0 , -125.4 ± 1.9 kcal/mol for mCaM, sCaM-1 and sCaM-4, respectively. To verify, we measured the EC₅₀ [CaM] for mCaM and sCaM-4 in activating CaN. This was obtained by measuring CaN activity with varying CaM concentrations under saturating Ca²⁺. As shown in Fig. 6c, the dependence curves of CaN activity on mCaM and sCaM-4 were indistinguishable and the calculated EC₅₀ [CaM] for mCaM and sCaM-4 were equal ($\sim 0.038 \mu\text{M}$, Fig. 6d). Although EC₅₀ is not merely the CaM binding affinity towards CaN, CaM and CaMBR binding affinity contributes large part to the measured EC₅₀. This is because CaMBR region is universal in CaM-targets and this region is the primary CaM recognition site when activating targets [2]. Thus, the almost identical EC₅₀ values for mCaM and sCaM-4 suggest that these two CaMs are highly likely to have the same CaMBR binding affinities. Additionally, it has been shown that the two sCaM's bind to the target with comparable affinities and are also comparable to other mCaM/targets binding affinities [11, 20], which support our interpretation of EC₅₀ [CaM] data. The two sCaM did not show significant difference in binding CaMBR when compared with mCaM, based on 1) Most of the contacts shown in CaMBR/mCaM binding are maintained in the two sCaMs. 2) The MMGBSA-calculated CaMBR/CaM binding free energies are comparable. Therefore the different CaN activation ability of sCaMs is not likely due to CaMBR binding.

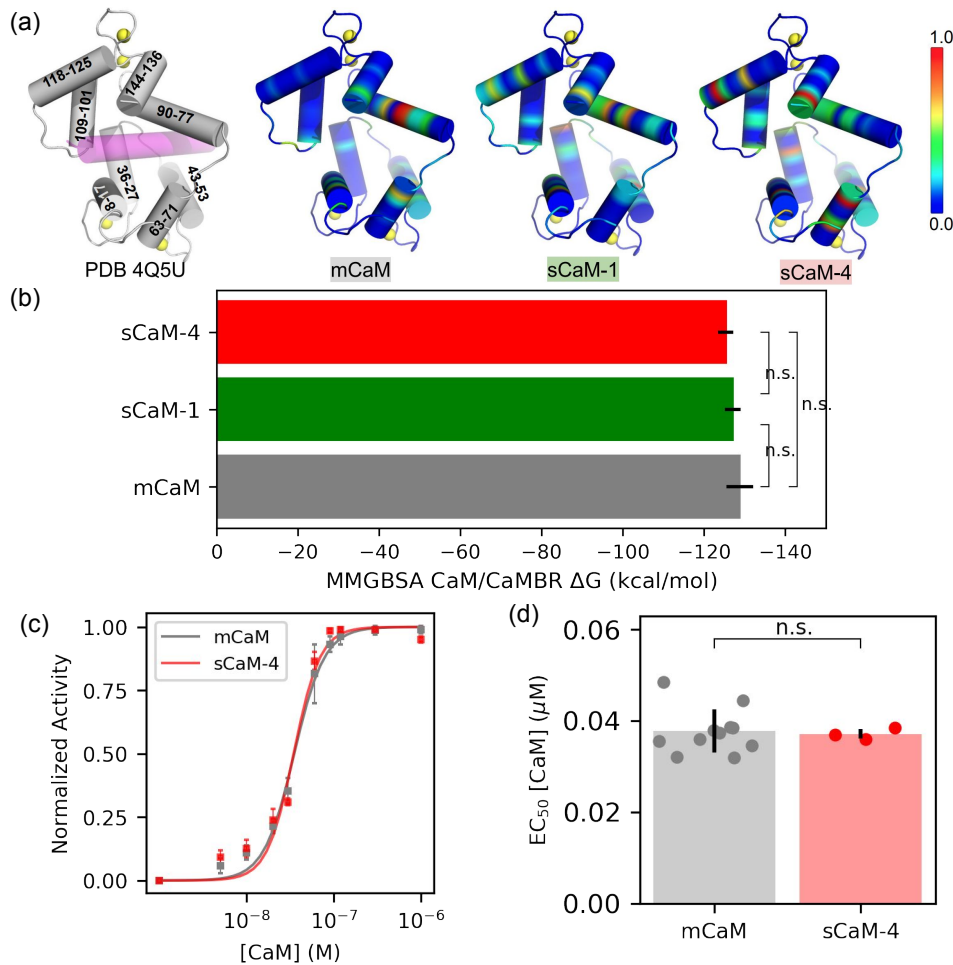


Figure 6: **CaMBR binding properties.** (a) Normalized contact probability (with respect to the residue pair that has the most accumulated contacts from the MD sampling) of CaMBR with CaM projected on mCaM/CaMBR complex structure PDB 4Q5U. The residues numbers comprising the helices are labeled. (b) MM-GBSA calculated binding free energy between CaMBR and CaM using MD trajectories. The difference between the three CaMs were non-significant as determined by student's t-test. (c) Dependence of CaN activity on CaM under saturating Ca^{2+} condition. (d) Experimental EC_{50} [CaM] for mCaM and sCaM-4.

4.3 mCaM/sCaM-1 exhibited stronger binding with distal helix than sCaM-4

We next examined if CaM/distal helix binding differed among the three CaMs as the final potential mechanism for differential CaN activation (Fig. 1c). To analyse the CaM/distal helix binding models, we first performed 100 ns REMD simulation of the isolated distal helix to sample the representative distal helix conformation as initially described in [30] (see Fig. S6 for REMD results). The representative distal helix was then docked to CaM isoforms via protein-protein docking, followed by extensive all-atom MD simulations. To visualize the CaM/distal helix conformational space, the MD trajectories were projected onto a 2D plane via the Sketch-map method [46], using the distal helix to CaMBR distance as the coordinate basis (Fig. 7a).

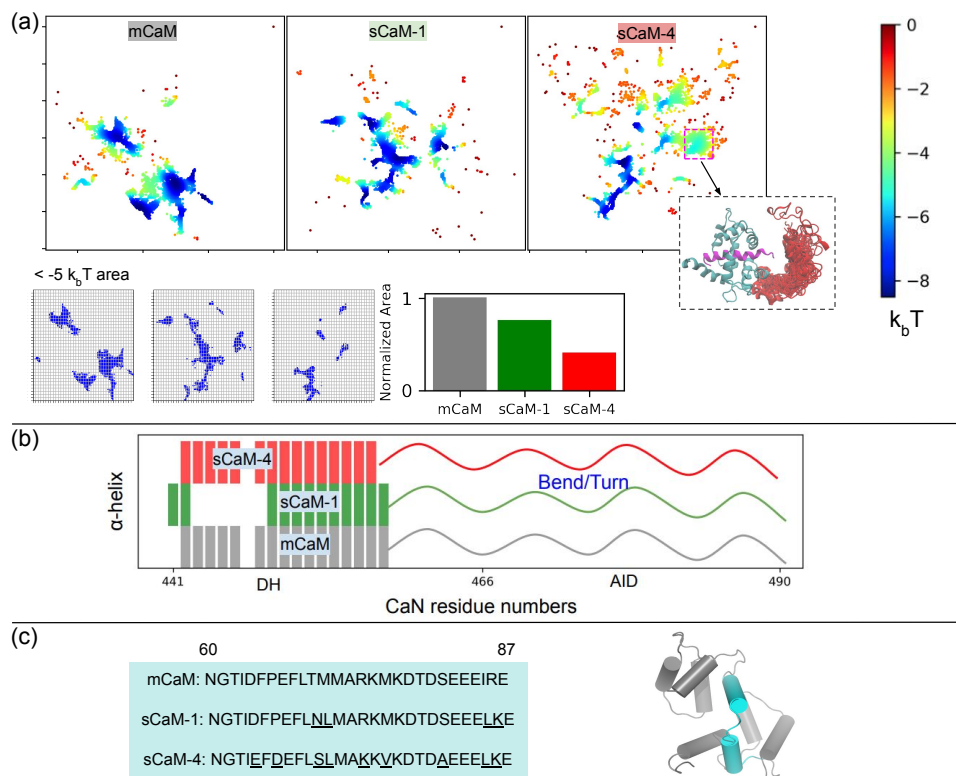


Figure 7: (a) Poses of the distal helix (DH) interacting with CaM. The poses are colored based on their energies using $-k_b T \ln(\rho/\rho_0)$ where k_b is Boltzmann constant, T is temperature and ρ is the point density after projection using Sketch-Map dimensionality reduction method [46]. ρ_0 is density of the peripheral points in sCaM-4 (dark red points). Region that are below $-5 k_b T$ are used as a reference for stable binding. In sCaM-4, representative structures from unstable region ($> -5 k_b T$) are shown with DH colored red. (b) Overall α -helicity of residues comprising the DH and AID of CaN from MD simulations. A bar depicts a residue that has α -helix as dominant secondary structure, which was calculated via the CPPTRJ program [44]. (c) Sequence comparison of residues N60-E87 of two sCaMs with mCaM. The region contributes most to the DH/CaM contact map difference between mCaM/sCaM-1 and sCaM-4 shown in Fig. S7. Non-conserved residues are underlined. The location of this sequence region is highlighted as cyan in the mCaM crystal structure PDB ID: 4Q5U.

The densities (ρ) of the projected points reflected the distribution of DH poses on CaM's surface. For easier interpretation, we converted the densities to free energies using $-k_b T \ln(\rho/\rho_0)$ where ρ_0 is the density corresponding to the peripheral points in sCaM-4. As shown in Fig. 7a, the free energy minimum of DH/CaM interaction were all about 8.5 kcal/mol for all three cases, which suggested that there were thermodynamic-favourable binding sites for the distal helix on all CaM isoforms. This is consistent with our experimental data demonstrating that all CaM isoforms are capable of activating CaN. Using $< -5 k_b T$ as an arbitrary criterion to define thermodynamic favourable area, mCaM had the largest area while sCaM-1 and sCaM-4 had 75% and 40% of mCaM, respectively (Fig. 7b). Visual inspection showed that high energy conformations of DH from sCaM-4 were more spread instead of being concentrated on CaM. The larger interacting area provided by sCaM-1 for the distal helix interaction is consistent with sCaM-1's complete activation of CaN versus partial activation by sCaM-4 (see Fig. 4).

We next performed contact map analyses to identify the key CaM residues that contributed to CaM/DH interactions among these three CaMs. We found that mCaM had the strongest contact in terms of number of contacting residues as well as contact persistence measured by the accumulated contacts (Fig. S7). sCaM-1 had comparable contact persistence with sCaM-4 but with more pairs of contacts. The notable difference was that mCaM residues (T62, I63 and D64) and

no CaM			
K_a (μM)	0.038 \pm 0.0048 @		
V_{max} (RFU/s)	0.44 \pm 0.16		
K_m (mM)	1.73 \pm 0.88		
	mCaM	sCaM-1	sCaM-4
α	19.43 \pm 6.39	39.69 \pm 13.25 *	7.95 \pm 1.24 *
β	2.61 \pm 0.85	4.70 \pm 1.23 *	2.20 \pm 0.42
α/β	8.47 \pm 1.84	8.41 \pm 1.60	3.64 \pm 0.15 *

Table 1: Kinetic parameters obtained by fitting model in Fig. 2 to experimental data. Student’s t test was used to determine the significance of difference between mCaM and the soybean isoforms. * $p < 0.05$. @ obtained from $EC_{50}[\text{CaM}]$ data in Fig. 6

sCaM-1 residues (D64, F65, P66 and F68) had more contacts with DH than sCaM-4 (Fig. S7). These residues locate closely to the linker of CaM that connects the N- and C- domains. In this region, sCaM-1 has six different residues relative to mCaM while sCaM-4 has ten (Fig. 7c). The larger sequence difference of sCaM-4 relative to mCaM/sCaM-1 is likely to weaken the distal helix interaction. sCaM-4 exhibited weaker CaM/DH interaction than mCaM and sCaM-1, which likely contributes to sCaM-4’s reduced CaN activation ability.

Higher helical content of DH may facilitate the AID displacement via shortening the distance between DH and AID. We calculated this propensity and determined that the overall α -helical residues in DH were comparable for these three CaM (Fig. 7b). The distal helix in all cases adopted α -helix structure at the C-termini. When interacting with CaM, the number of α -helical residues were comparable for these three CaM, with 16, 12, and 15 residues for mCaM, sCaM-4 and sCaM-4, respectively. Hence, the CaMs don’t appear to impact distal helix upon activating CaN. This suggests distal helix helicity may be a precursor to activity showed by these CaM isoforms.

4.4 Kinetic fitting suggested that stronger CaM/DH led to larger apparent k_{cat} of CaN.

To determine how CaM/CaN interactions impact CaN’s kinetics, beyond standard molecular modeling, we fit the kinetic data shown in Fig. 4 to the kinetic model in Fig. 2. This model includes additional CaM/CaN interactions studied via simulations. As shown in Fig. 1a, multiple interactions such as CaM/CaMBR interaction, CaM/DH interaction and activated CaN/substrate interactions are involved in the dephosphorylation process. There are three assumptions in the proposed model: 1) CaM binds only to the enzyme but not the enzyme-substrate complex, in accordance with the mechanism of CaN activation by CaM (Fig. 2). 2) The three CaM isoforms share the same K_a as indicated by the $EC_{50}[\text{CaM}]$ data where mCaM and sCaM-4 showed nearly identical $EC_{50}[\text{CaM}]$ (Fig. 6). 3) Once bound by CaM and activated, CaN exhibits modified kinetic properties described by α and β factors which rescale k_{cat} and K_m , respectively.

Firstly, fitting yielded an α value much greater than 1 for all CaM cases, indicating that CaM significantly enhanced the catalysis of CaN (k_{cat}). All CaM isoforms also exhibited $\beta > 1$, indicating a modified K_m . However, it should be noted that interpreting K_m as the substrate affinity is not appropriate here for the following reasons, which are: Suppose the binding of S to CaN can be described by k_{on} and k_{off} , the on and off rates of S to CaN. $K_m = \frac{k_{off} + k_{cat}}{k_{on}}$ therefore approximates $K_{d,S} = \frac{k_{off}}{k_{on}}$ if $k_{cat} \ll k_{on}, k_{off}$. In this case, a change in α is independent of a change in β . However, as our kinetic data revealed, both α and β increased, implying a possible correlation between these two parameters. Consequently, interpretation of K_m as the affinity for substrate is nuanced.

To examine the impact of different CaMs on CaN kinetics, we computed α/β as a measure of catalytic efficiency. Both mCaM and sCaM-1 showed nearly identical α/β and are about twice that of sCaM-4. This indicates that upon CaM binding, CaN is converted into an more activated form (*i.e.* AID-removed CaN), in contrast to its autoinhibited form. And more importantly, this

activated form possesses enhanced catalytic capability. In the previous sections, both experimental and computational data demonstrated that the Ca^{2+} and CaMBR binding did not contribute much to the different activation effects of CaM isoforms. Meanwhile, the three isoforms exhibited differences in CaM/DH interaction. Hence the difference in catalytic efficiency could be attributed to the difference in DH binding. Kinetic analyses suggested that CaM increased the turn over rate. CaN activated by mCaM and sCaM-1 had significantly larger turn over rate (k_{cat}) than that of sCaM-4.

To test this hypothesis, using a polymer model, we showed in Sect. S2.5 that the CaM/DH interaction is positively correlated with the effective AID concentration ($[\text{AID}]_{\text{eff}}$), a quantitative measure of the extent of CaN being autoinhibited. In the case of mCaM and sCaM-1, the CaM/DH interaction led to lower $[\text{AID}]_{\text{eff}}$ than that of sCaM-4. Hence, it is possible that the DH/CaM binding in mCaM and sCaM-1 pulled the CaN's AID away from its catalytic site to a greater extent through the interactions with DH, resulting in a CaN form that showed higher catalytic efficiency (α/β). Further studies will be needed to delineate other potential mechanisms, including interactions of CaM with the well folded catalytic domain.

5 Conclusions

CaM is an important Ca^{2+} sensor that translates Ca^{2+} signaling to various protein-protein interactions. It was long believed that CaM has an invariant amino acid sequence in vertebrates [57] and has only one amino acid difference to several insect CaM sequences [58]. However, recent studies have revealed that CaM mutations in human populations have so far been related to cardiac diseases [31, 59], likely through disturbing Ca^{2+} -handling in cardiac cells [16]. The molecular mechanisms underlying calmodulinopathy are still emerging [60]. Understanding how plant-derived CaM isoforms affect the activation of mammalian targets helps to explain the altered functions of CaM variants found in human. In the present study, we examined the molecular basis of two soybean CaM isoforms that differentially activate human CaN. Two soybean CaMs, sCaM-4 and sCaM-1, share moderate to high sequence identity to human CaM (mCaM), but exhibit differential regulation of human CaN: sCaM-1 was comparable with mCaM and sCaM-4 was significantly reduced. Based on a recent CaM/CaN interaction model [30], we investigated three molecular interactions that may affect CaM's activation on CaN. 1) Ca^{2+} binding to CaM. 2) CaM binding to the CaM binding region (CaMBR) of CaN and 3) CaM binding to the distal helix (DH) motif of CaN. By using molecular dynamics simulations and experimental measurements, we excluded Ca^{2+} binding and CaMBR binding as factors that differentiates the sCaMs. Instead, it is the CaM/DH interaction that strongly differentiates sCaM-4 versus mCaM/sCaM-1. Namely, mCaM and sCaM-1 both exhibited stronger interactions with distal helix than sCaM-4. Kinetic modeling revealed that the prime effect of the two sCaMs was to form CaN/CaM complex with different turnover (k_{cat}) rates, which is likely resulted from altered CaM/DH interactions. Using a polymer model, we found that the CaM/DH interaction correlates with the effective [AID] concentration around the catalytic site of CaN, which could impact the turnover rate. Our study shed insights into how CaM mutation may affect target activation through interaction beyond the canonical CaMBR binding and Ca^{2+} binding, which may extend to other CaM regulated targets such as CaM protein kinase II (CaMKII [61]), CaM kinase kinase (CaMKK [62]), myosin light chain kinase (MLCK [63] and death-associated protein kinase (DAPK) [64]) that utilize autoinhibitory domains.

6 Acknowledgements

Research reported in this publication was supported by the Maximizing Investigators' Research Award (MIRA) (R35) from the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under grant number R35GM124977 (PKH). This work was also supported by American Heart Association under grant number 20CDA35310757 (CNJ) and NIH under grant 5R01HL138579-04 (JPD). This work used the Extreme Science and Engineering Discovery Environment (XSEDE) [65], which is supported by National Science Foundation under grant number ACI-1548562.

References

- [1] F. C. Stevens. *Calmodulin: an introduction*. 1983.
- [2] H. Tidow and P. Nissen. “Structural diversity of calmodulin binding to its target sites”. In: *FEBS Journal*. Vol. 280. 21. John Wiley & Sons, Ltd, Nov. 2013, pp. 5551–5565.
- [3] L. Racioppi and A. R. Means. “Calcium/calmodulin-dependent kinase IV in immune and inflammatory responses: novel routes for an ancient traveller”. In: *Trends in Immunology* 29.12 (2008), pp. 600–607. ISSN: 14714906.
- [4] W. A. Snedden and H. Fromm. “Calmodulin, calmodulin-related proteins and plant responses to the environment”. In: *Trends in Plant Science* 3.8 (1998), pp. 299–304. ISSN: 13601385.
- [5] Sang Hyoung Lee et al. “Identification of a novel divergent calmodulin isoform from soybean which has differential ability to activate calmodulin-dependent enzymes”. In: *Journal of Biological Chemistry* 270.37 (Sept. 1995), pp. 21806–21812. ISSN: 00219258.
- [6] Y. Zhao et al. “Genome-wide identification and functional analyses of calmodulin genes in Solanaceous species”. In: *BMC Plant Biology* 13.1 (2013), p. 70. ISSN: 14712229.
- [7] H. Ishida and H. J. Vogel. “The solution structure of a plant calmodulin and the CaM-binding domain of the vacuolar calcium-ATPase BCA1 reveals a new binding and activation mechanism”. In: *Journal of Biological Chemistry* 285.49 (Dec. 2010), pp. 38502–38510. ISSN: 00219258.
- [8] J. L. Gifford et al. “Comparing the calcium binding abilities of two soybean calmodulins: Towards understanding the divergent nature of plant calmodulins”. In: *Plant Cell* 25.11 (Nov. 2013), pp. 4512–4524.
- [9] S. D. Walton et al. “Divergent soybean calmodulins respond similarly to calcium transients: Insight into differential target regulation”. In: *Frontiers in Plant Science* 8 (Feb. 2017). ISSN: 1664462X.
- [10] M. J. Cho et al. “Reciprocal regulation of mammalian nitric oxide synthase and calcineurin by plant calmodulin isoforms”. In: *Biochemistry* 37.45 (1998), pp. 15593–15597. ISSN: 00062960.
- [11] H. Ishida et al. “The solution structures of two soybean calmodulin isoforms provide a structural basis for their selective target activation properties”. In: *Journal of Biological Chemistry* 283.21 (May 2008), pp. 14619–14628. ISSN: 00219258.
- [12] J. Roy and M. S. Cyert. “Identifying New Substrates and Functions for an Old Enzyme: Calcineurin”. In: *Cold Spring Harbor Perspectives in Biology* 12.3 (2020).
- [13] Y. J. Park et al. “The Role of Calcium-Calcineurin-NFAT Signaling Pathway in Health and Autoimmune Diseases”. In: *Frontiers in Immunology* 11 (2020), p. 195. ISSN: 16643224.
- [14] O. F. Bueno et al. “Calcineurin and hypertrophic heart disease: novel insights and remaining questions”. In: *Cardiovascular Research* 53.4 (Mar. 2002), pp. 806–821. ISSN: 0008-6363.
- [15] L. C. Reese and G. Tagliatela. “A Role for Calcineurin in Alzheimer’s Disease”. In: *Current Neuropharmacology* 9.4 (2011), pp. 685–692. ISSN: 1570-159X.
- [16] H. H. Jensen et al. “Human calmodulin mutations”. In: *Frontiers in Molecular Neuroscience* 11 (2018), p. 396. ISSN: 16625099.
- [17] H. Kuboniwa et al. “Solution structure of calcium-free calmodulin”. In: *Nature Structural Biology* 2.9 (1995), pp. 768–776. ISSN: 10728368.
- [18] Y. S. Babu et al. “Three-dimensional structure of calmodulin”. In: *Nature* 315.6014 (1985), pp. 37–40. ISSN: 00280836.
- [19] D. Shukla, A. Peck, and V. S. Pande. “Conformational heterogeneity of the calmodulin binding interface”. In: *Nature Communications* 7.1 (Apr. 2016), pp. 1–11. ISSN: 20411723.
- [20] A. P. Yamniuk and H. J. Vogel. “Calmodulin’s flexibility allows for promiscuity in its interactions with target proteins and peptides”. In: *Applied Biochemistry and Biotechnology - Part B Molecular Biotechnology* 27.1 (2004), pp. 33–57. ISSN: 10736085.

- [21] C. Ghosh and B. Jana. "Role of Calcium in Modulating the Conformational Landscape and Peptide Binding Induced Closing of Calmodulin". In: *Journal of Physical Chemistry B* 125.9 (Mar. 2021), pp. 2317–2327. ISSN: 15205207.
- [22] G. G. Rodney et al. "Calcium binding to calmodulin leads to an N-terminal shift in its binding site on the ryanodine receptor". In: *Journal of Biological Chemistry* 276.3 (2001), pp. 2069–2074. ISSN: 00219258.
- [23] X. ZHANG and S. K. JOSEPH. "Effect of mutation of a calmodulin binding site on Ca²⁺ regulation of inositol trisphosphate receptors". In: *Biochemical Journal* 360.2 (Nov. 2001), pp. 395–400. ISSN: 0264-6021. eprint: <https://portlandpress.com/biochemj/article-pdf/360/2/395/707933/bj3600395.pdf>.
- [24] N. Yamaguchi et al. "Calmodulin regulation and identification of calmodulin binding region of type-3 ryanodine receptor calcium release channel". In: *Biochemistry* 44.45 (2005), pp. 15074–15081. ISSN: 00062960.
- [25] T. B. Dunlap et al. "Thermodynamics of binding by calmodulin correlates with target peptide α -helical propensity". In: *Proteins: Structure, Function and Bioinformatics* 81.4 (Apr. 2013), pp. 607–612. ISSN: 08873585.
- [26] T. B. Dunlap et al. "The distal helix in the regulatory domain of calcineurin is important for domain stability and enzyme function". In: *Biochemistry* 52.48 (2013), pp. 8643–8651. ISSN: 00062960.
- [27] T. B. Dunlap et al. "Stoichiometry of the Calcineurin Regulatory Domain-Calmodulin Complex". In: *Biochemistry* 53.36 (2014), pp. 5779–5790.
- [28] B. Sun et al. "Electrostatic control of calcineurin's intrinsically-disordered regulatory domain binding to calmodulin". In: *Biochimica et Biophysica Acta - General Subjects* 1862.12 (2018), pp. 2651–2659. ISSN: 18728006.
- [29] B. Sun and Keken-Huskey. "Assessing the Role of Calmodulin's Linker Flexibility in Target Binding". In: *International Journal of Molecular Science* 22.9 (2021). ISSN: 1422-0067.
- [30] B. Sun et al. "Calmodulin-Calcineurin Interaction beyond the Calmodulin-Binding Region Contributes to Calcineurin Activation". In: *Biochemistry* 58.39 (2019), pp. 4070–4085. ISSN: 15204995.
- [31] L. Crotti et al. "Calmodulin Mutations Associated With Recurrent Cardiac Arrest in Infants". In: *Circulation* 127.9 (2013), pp. 1009–1017.
- [32] Y. Song et al. "High-resolution comparative modeling with RosettaCM". In: *Structure* 21.10 (2013), pp. 1735–1742. ISSN: 09692126.
- [33] J. A. Maier et al. "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB". In: *Journal of Chemical Theory and Computation* 11.8 (Aug. 2015), pp. 3696–3713. ISSN: 15499626. arXiv: [15334406](https://arxiv.org/abs/15334406).
- [34] W. L. Jorgensen et al. "Comparison of simple potential functions for simulating liquid water". In: *The Journal of Chemical Physics* 79.2 (1983), pp. 926–935.
- [35] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen. "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes". In: *Journal of Computational Physics* 23.3 (1977), pp. 327–341. ISSN: 0021-9991.
- [36] D. Case et al. "AMBER 2018". In: ().
- [37] A. Raval et al. "Refinement of protein structure homology models via long, all-atom molecular dynamics simulations". In: *Proteins: Structure, Function and Bioinformatics* 80.8 (Aug. 2012), pp. 2071–2079. ISSN: 08873585.
- [38] A. Patriksson and D. van der Spoel. "A temperature predictor for parallel tempering simulations". In: *Phys. Chem. Chem. Phys.* 10 (15 2008), pp. 2073–2077.
- [39] S. Patel, E. Vierling, and F. Tama. "Replica exchange molecular dynamics simulations provide insight into substrate recognition by small heat shock proteins". In: *Biophysical Journal* 106.12 (June 2014), pp. 2644–2655. ISSN: 15420086.

- [40] K. Lindorff-Larsen et al. "Improved side-chain torsion potentials for the Amber ff99SB protein force field". In: *Proteins: Structure, Function and Bioinformatics* 78.8 (2010), pp. 1950–1958. ISSN: 08873585.
- [41] B. G. Pierce et al. "ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers". eng. In: *Bioinformatics* 30.12 (June 2014), pp. 1771–1773. ISSN: 14602059.
- [42] B. N. Bullock, A. L. Jochim, and P. S. Arora. "Assessing helical protein interfaces for inhibitor design". In: *Journal of the American Chemical Society* 133.36 (2011), pp. 14220–14223. ISSN: 00027863. arXiv: [NIHMS150003](#).
- [43] D. Case et al. "AMBER 2016". In: ().
- [44] D. R. Roe and T. E. Cheatham. "PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data". In: *Journal of Chemical Theory and Computation* 9.7 (2013), pp. 3084–3095.
- [45] W. Kabsch and C. Sander. "Dictionary of protein secondary structure: Pattern recognition of hydrogen bonded and geometrical features". In: *Biopolymers* 22.12 (Dec. 1983), pp. 2577–2637.
- [46] M. Ceriotti, G. A. Tribello, and M. Parrinello. "Simplifying the representation of complex free-energy landscapes using sketch-map". In: *Proceedings of the National Academy of Sciences* 108.32 (Aug. 2011), pp. 13023–13028. ISSN: 0027-8424.
- [47] G. A. Tribello, M. Ceriotti, and M. Parrinello. "Using sketch-map coordinates to analyze and bias molecular dynamics simulations". In: *Proceedings of the National Academy of Sciences* 109.14 (Apr. 2012), pp. 5196–5201. ISSN: 0027-8424.
- [48] A. Berg and C. Peter. "Simulating and analysing configurational landscapes of protein-protein contact formation". In: *Interface Focus* 9.3 (June 2019), p. 20180062. ISSN: 20428901.
- [49] A. M. Westerlund and L. Delemotte. "Effect of Ca²⁺ on the promiscuous target-protein binding of calmodulin". In: *PLoS Computational Biology* 14.4 (Apr. 2018), e1006072. ISSN: 15537358.
- [50] K. Immadisetty, B. Sun, and P. M. Kekenus-Huskey. "Structural Changes beyond the EF-Hand Contribute to Apparent Calcium Binding Affinities: Insights from Parvalbumins". In: *The Journal of Physical Chemistry B* 0.0 (2021), null. ISSN: 1520-6106.
- [51] A. N. Kucharski et al. "Understanding Ion Binding Affinity and Selectivity in β -Parvalbumin Using Molecular Dynamics and Mean Spherical Approximation Theory". In: *Journal of Physical Chemistry B* 120.33 (2016), pp. 8617–8630. ISSN: 15205207.
- [52] T. Dudev and C. Lim. *Competition among metal ions for protein binding sites: Determinants of metal ion selectivity in proteins*. Jan. 2014.
- [53] B. Sun et al. "Thermodynamics of Cation Binding to the Sarcoendoplasmic Reticulum Calcium ATPase Pump and Impacts on Enzyme Function". In: *Journal of Chemical Theory and Computation* 15.4 (Apr. 2019), pp. 2692–2705. ISSN: 15499626.
- [54] B. Sun and P. M. Kekenus-Huskey. "Molecular Basis of S100A1 Activation and Target Regulation Within Physiological Cytosolic Ca²⁺ Levels". eng. In: *Frontiers in Molecular Biosciences* 7 (June 2020), p. 77. ISSN: 2296889X.
- [55] F. Jalilehvand et al. "Hydration of the calcium ion. An EXAFS, large-angle X-ray scattering, and molecular dynamics simulation study". In: *Journal of the American Chemical Society* 123.3 (2001), pp. 431–441. ISSN: 00027863.
- [56] Z. Grabarek. "Structural Basis for Diversity of the EF-hand Calcium-binding Proteins". In: *Journal of Molecular Biology* 359.3 (2006), pp. 509–525. ISSN: 00222836.
- [57] D. Chin and A. R. Means. "Calmodulin: A prototypical calcium sensor". In: *Trends in Cell Biology* 10.8 (2000), pp. 322–328. ISSN: 09628924.
- [58] W. Wang et al. "Double-stranded RNA targeting calmodulin reveals a potential target for pest management of *Nilaparvata lugens*". In: *Pest Management Science* 74.7 (2018), pp. 1711–1719. ISSN: 15264998.

- [59] M. Nyegaard et al. "Mutations in Calmodulin Cause Ventricular Tachycardia and Sudden Cardiac Death". In: *The American Journal of Human Genetics* 91.4 (2012), pp. 703–712. ISSN: 0002-9297.
- [60] W. J. Chazin and C. N. Johnson. "Calmodulin Mutations Associated with Heart Arrhythmia: A Status Report". In: *International Journal of Molecular Sciences* 21.4 (2020). ISSN: 1422-0067.
- [61] W. E. Meador, A. R. Means, and F. A. Quijcho. "Modulation of calmodulin plasticity in molecular recognition on the basis of X-ray structures". In: *Science* 262.5140 (1993), pp. 1718–1721. ISSN: 00368075.
- [62] H. Kurokawa et al. "Target-induced conformational adaptation of calmodulin revealed by the crystal structure of a complex with nematode Ca²⁺/calmodulin-dependent kinase kinase peptide". In: *Journal of Molecular Biology* 312.1 (Sept. 2001), pp. 59–68. ISSN: 00222836.
- [63] J. Gsponer et al. "A Coupled Equilibrium Shift Mechanism in Calmodulin-Mediated Signal Transduction". In: *Structure* 16.5 (May 2008), pp. 736–746. ISSN: 09692126.
- [64] I. De Diego et al. "Molecular basis of the death-associated protein kinase-calcium/calmodulin regulator complex". In: *Science Signaling* 3.106 (Jan. 2010), ra6 LP –ra6. ISSN: 19450877.
- [65] J. Towns et al. "XSEDE: Accelerating Scientific Discovery". In: *Computing in Science & Engineering* 16.5 (Sept. 2014), pp. 62–74. ISSN: 1521-9615.
- [66] H. Fan. "Refinement of homology-based protein structures by molecular dynamics simulation techniques". In: *Protein Science* 13.1 (Jan. 2003), pp. 211–220. ISSN: 0961-8368.
- [67] Z. Xiang. "Advances in Homology Protein Structure Modeling". In: *Current Protein and Peptide Science* 7.3 (2006), pp. 217–227. ISSN: 1389-2037.
- [68] D. Van Valen, M. Haataja, and R. Phillips. "Biochemistry on a leash: The roles of tether length and geometry in signal integration proteins". In: *Biophysical Journal* 96.4 (Feb. 2009), pp. 1275–1292. ISSN: 15420086.
- [69] W. Borchers et al. "Optimal Affinity Enhancement by a Conserved Flexible Linker Controls p53 Mimicry in MdmX". In: *Biophysical Journal* 112.10 (May 2017), pp. 2038–2042. ISSN: 15420086.

S1 Supplementary Information (SI)

Bin Sun¹, Xuan Fang¹, Christopher N. Johnson^{2,3}, Garrett Hauck²,
Jonathan P. Davis² and Peter M. Kekeneshuskey^{*,1}

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

² Department of Physiology and Cell Biology, The Ohio State University, Columbus, OH, USA
43210

³ Department of Chemistry, Mississippi State University Starkville MS, 39759

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

Table S1: All molecular dynamics simulations performed in present study

Systems	Starting structure	Simulation length
sCaM-1	Rosetta modeled structure	3 x 2 μ s = 6 μ s
sCaM-4	Rosetta modeled structure	3 x 2 μ s = 6 μ s
mCaM	PDB ID 4Q5U	3 x 2 μ s = 6 μ s
Isolated Distal helix (DH)	Replica-Exchange MD from sequence	100 ns
sCaM-1 + DH	2 MD refined sCaM-1 structures + DH docked at 4 sites	2 x 4 x 3 x 1.2 μ s = 28.8 μ s
sCaM-4 + DH	2 MD refined sCaM-4 structures + DH docked at 4 sites	2 x 4 x 3 x 1.2 μ s = 28.8 μ s
mCaM + DH	Representative structure of MD simulations of 4Q5U + DH docked at 4 sites	1 x 4 x 3 x 1.2 μ s = 14.4 μ s
sCaM-1 + DH + AID	Representative sCaM-1 + DH at each site with AID fused to DH C-termini	8 x 1 μ s = 8 μ s
sCaM-4 + DH + AID	Representative sCaM-4 + DH at each site with AID fused to DH C-termini	8 x 1 μ s = 8 μ s
mCaM + DH + AID	Representative mCaM + DH at each site with AID fused to DH C-termini	4 x 1 μ s = 4 μ s
		total: ~110 μ s

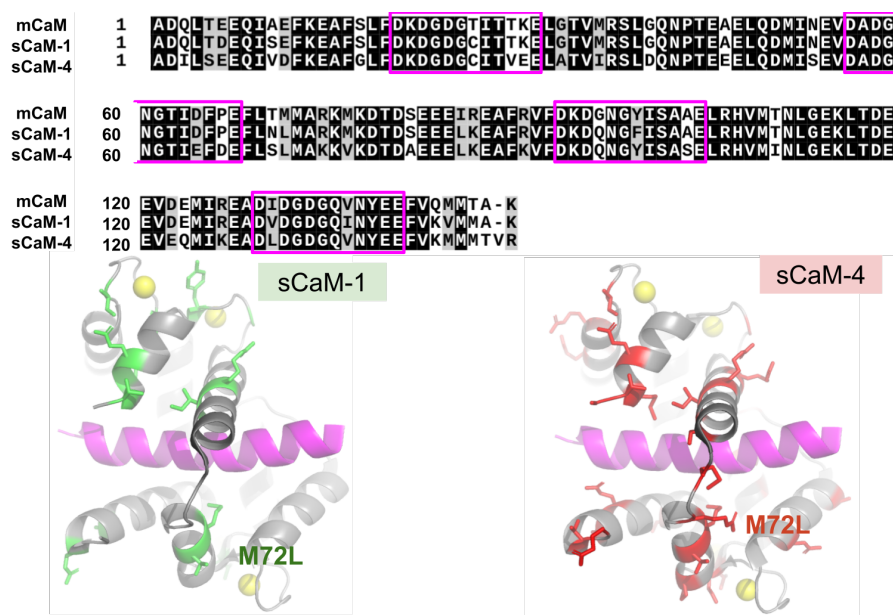


Figure S1: Sequence comparison between mammalian CaM (mCaM) and the sCaM-1/sCaM-4 soybean isoforms. The EF-hand residues are highlighted with magenta boxes. The lower panel highlights residues in soybean CaMs that are conserved in the human CaM/CaMBR complex structure (PDB 4Q5U [27]). Unique residues are shown in sticks and colored green and red for sCaM-1 and sCaM-4, respectively.

S2 Supplement Method

S2.1 Protocol and flags of Rosetta Comparative Modeling

Listing 1: HybridizeMover Protocol

```

1 <ROSETTASCRIPTS>
2 <SCOREFXNS>
3   <stage1 weights=score3 symmetric=0>
4     <Reweight scoretype=atom_pair_constraint weight=0.5/>
5   </stage1>
6   <stage2 weights=score4.smooth_cart symmetric=0>
7     <Reweight scoretype=atom_pair_constraint weight=0.5/>
8   </stage2>
  
```

```
9 <fullatom weights=talaris2013_cart symmetric=0>
10 <Reweight scoretype=atom_pair_constraint weight=0.5/>
11 </fullatom>
12 </SCOREFXNS>
13 <MOVERS>
14 <Hybridize name=hybridize stage1_scorefxn=stage1 stage2_scorefxn=stage2
    fa_scorefxn=fullatom batch=1 stage1_increase_cycles=1.0
    stage2_increase_cycles=1.0>
15 <Template pdb="4q5uA.pdb" cst_file="AUTO" weight= 1.000 />
16 <Template pdb="sCaMN.pdb" cst_file="AUTO" weight= 0.800 />
17 <Template pdb="sCaMC.pdb" cst_file="AUTO" weight= 0.800 />
18 </Hybridize>
19 <FastRelax name="relax" scorefxn=talaris2013 />
20 </MOVERS>
21 <PROTOCOLS>
22 <Add mover=hybridize/>
23 <Add mover=relax/>
24 </PROTOCOLS>
25 <OUTPUT scorefxn=talaris2013 />
26 </ROSETTASCRIPTS>
```

Flags:

Listing 2: Flags

```
1 # i/o
2 -in:file:fasta soyCaM.fasta
3 -nstruct 10
4 -parser:protocol HybridizeMover.xml
5 -out:file:silent_struct.type binary
6
7 # relax options
8 -relax:minimize_bond_angles
9 -relax:minimize_bond_lengths
10 -relax:jump_move true
11 -default_max_cycles 200
12 -relax:min_type lbfgs_armijo_nonmonotone
13 -relax:jump_move true
14 -score:weights stage3_rlx.wts
15 -use_bicubic_interpolation
16 -hybridize:stage1_probability 1.0
17 -sog_upper_bound 15
18
19 # reduce memory footprint
20 -chemical:exclude_patches LowerDNA UpperDNA Cterm_amidation SpecialRotamer
    VirtualBB ShoveBB VirtualDNAPhosphate VirtualNTerm CTermConnect sc_orbitals
    pro_hydroxylated_case1 pro_hydroxylated_case2 ser_phosphorylated
    thr_phosphorylated tyr_phosphorylated tyr_sulfated lys_dimethylated
    lys_monomethylated lys_trimethylated lys_acetylated glu_carboxylated
    cys_acetylated tyr_diiodinated N_acetylated C_methylamidated
    MethylatedProteinCterm
```

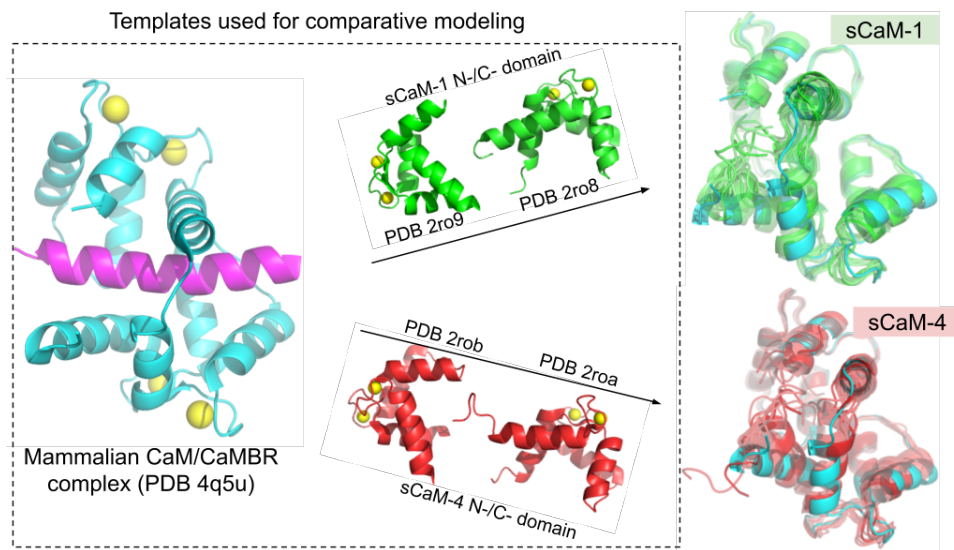



Figure S2: Rosetta comparative modeling of sCaM/CaMBR complex structures using human CaM/CaMBR crystal structure (PDB ID: 4q5u) and isolated sCaM domain structures as templates. The 10 modelled structures are superimposed on mammalian CaM template and are color green and red for sCaM-1 and sCaM-4, respectively.

S2.2 Homology modeling combined with MD refinement gives reliable sCaM/CaMBR complex structures

Although the separate N-/C- domains structures of sCaM-1 and sCaM-4 with Ca^{2+} bound in the EF-hands have been resolved [11], complete sCaM-1/4 structures complexed with CaMBR of CaN are not yet unavailable. We thus utilized the Rosetta comparative modeling tools [32] to build the complete structures of sCaM isoforms to provide a structural basis for exploring CaM/-CaN interactions. Starting from the highest-scored Rosetta modeled structures, we performed 2 μs MD in triplicates to refine the modeled structures as well as to investigate the structural and dynamic features of the sCaM isoforms. In addition, we also performed MD for the human CaM/CaMBR complex starting from the crystal structure PDB 4Q5U as a comparison.

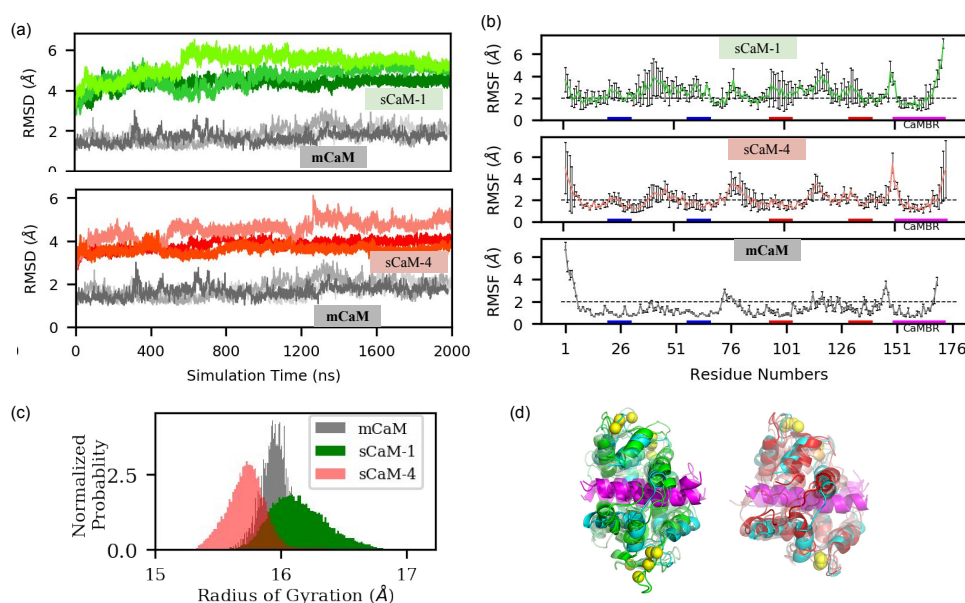


Figure S3: (a) root mean squared deviations (RMSD) of backbone atom of CaM/CaMBR. The first frame of the trajectory was used as reference of RMSD calculation. (b) root mean squared fluctuations (RMSF) of non-hydrogen atom in the triplicate MD simulations. The blue and red bars at the x axis represent the EF hands in the N and C domain, respectively. All frames were fitted to the first frame before RMSF calculation. The error bars are standard deviations calculated from the triplicate runs. (c) Radius of gyration distribution (d) Representative structures of sCaM-1 (green) and sCaM-4 (red) from triplicate MD simulations. The human CaM/CaMBR structure (PDB 4Q5U [27]) is colored cyan with CaMBR colored magenta. Ca²⁺ ions are represented as yellow balls.

As shown in Fig. S3a, the backbone atom root mean squared deviations (RMSD) values of sCaM-1 and sCaM-4 converges around 5 and 4 Å respectively. Meanwhile, the mCaM case started from the crystal structure maintains low RMSD values around 2 Å. The larger RMSD values for sCaM-1 and sCaM-4 are expected and it reflects the structural relaxation of the comparative modeled structures relative to its initial structure. Despite larger RMSD values for the two sCaM isoforms, these MD refinements suggest that modeled structures are stable over the 2 μ s time-scale simulations. More importantly, electrophoresis experiments of purified sCaMs using SDS-polyacrylamide gel in the presence of 5 mM CaCl₂ [5] has shown that sCaM-4 has greater mobility than sCaM-1, implying that Ca²⁺ bound sCaM-4 is more compact. Our MD simulations support this experimental observation, as shown in Fig. S3d, sCaM-1 has significantly larger radius of gyration than sCaM-4. Experiments also show that Ca²⁺ binding causes more hydrophobic patch exposure in sCaM-1 than sCaM-4 as measured by 1-anilino-8-naphthalenesulfonate (ANS) fluorescence spectroscopy [11]. The less-packed hydrophobic core of sCaM-1 might be the reason why it is more extended than sCaM-4. By integration of comparative modeling with appropriately selected templates and extensive atomistic MD simulation refinement, we were able to model reliable complete sCaM structures bound with CaMBR of CaN, as evident by 1) RMSD values that are converged over the μ s MD time-scale. Compact structure of sCaM-4 that is consistent with experimental observations. This integrated protocol has been widely used to build protein structures that have eluded experimental structural determination [66, 67].

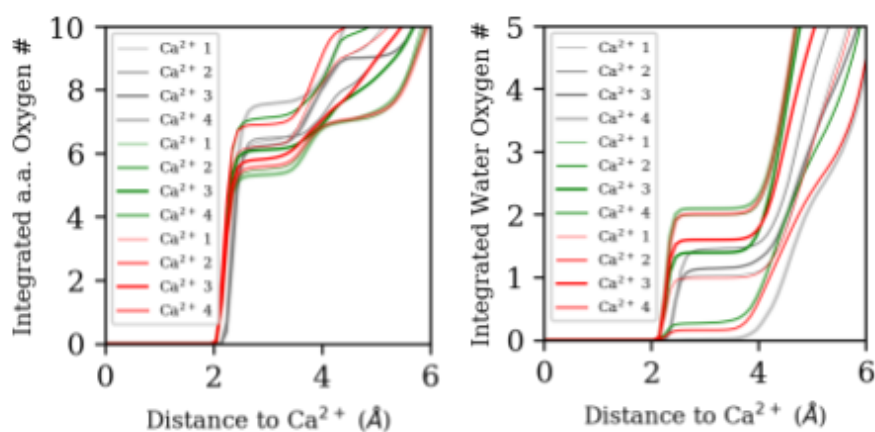


Figure S4: Integrated amino acid (a.a.) and water oxygen numbers around Ca²⁺ ions.

S2.3 Clustering analysis on sCaM MD trajectory to identify representative structures that will be used to dock with DH

To perform the clustering analysis, we first divided the triplicate MD runs for each sCaM into two groups, based on the RMSD values (Fig. S5). Thus for each sCaM, we have two groups A and B. For each group, a clustering analysis was performed using the cpptraj module of Amber with the RMS of backbone atoms as distance metric. The hierarchical agglomerative (bottom-up) approach was used and the MD frames were divided into ten clusters. The representative structure of the most populated cluster was picked out (the bottom panel of Fig. S5)

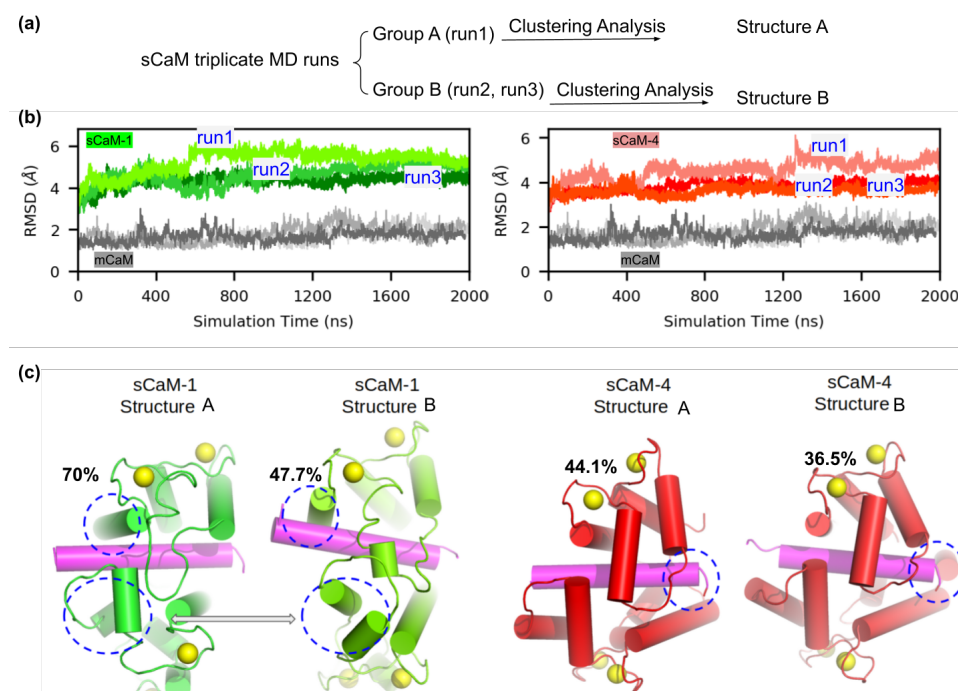


Figure S5: (a) Strategy of clustering analysis on the MD refinement trajectories. (b) Backbone atom RMSD of triplicate MD refinement simulations. The first frame was used as reference. (c) Two representative structures for each sCaM were selected to dock distal helix. The blue dashed circles highlight the structural difference of the two structures of the same sCaM.

S2.4 REMD of distal helix and putative docking sites on CaM surface

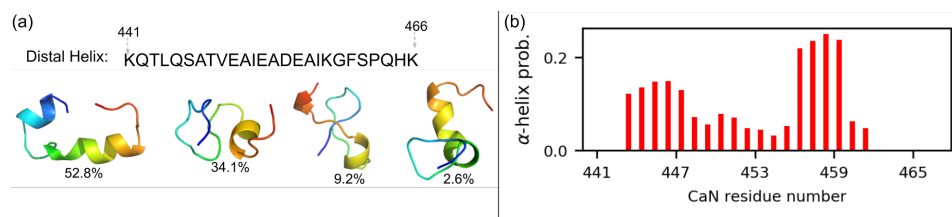


Figure S6: Structural property of distal helix sampled via 100 ns REMD simulation. (a) Sequence and representative structures of the first four most populated clusters. (b) α -helix probability of each residue.

Table S2: Residues at the each tentative binding site on sCaM-1/sCaM-4 used in ZDOCK to predict distal helix interaction at each site. The slightly different residues at each site between hCaM and sCaM-1/sCaM-4 was to make sure these selected residues are exposed on surface.

Tentative Site	hCaM	sCaM-1 Residues	sCaM-4 Residues
Site A	R86, F89, V142, Y138	K86, R90, Y138, V142	K86, K90, Y138, K143
Site B	R106, I125, D118, D122	R106, D118, D122, R126	R106, D118, E122, K126
Site C	I9, F12, L69, F65	I9, K13, F65, N70	V10, K13, F65, S70
Site D	K30, T34, R37, Q49	K30, V35, S38, Q49	E30, T34, R37, Q49

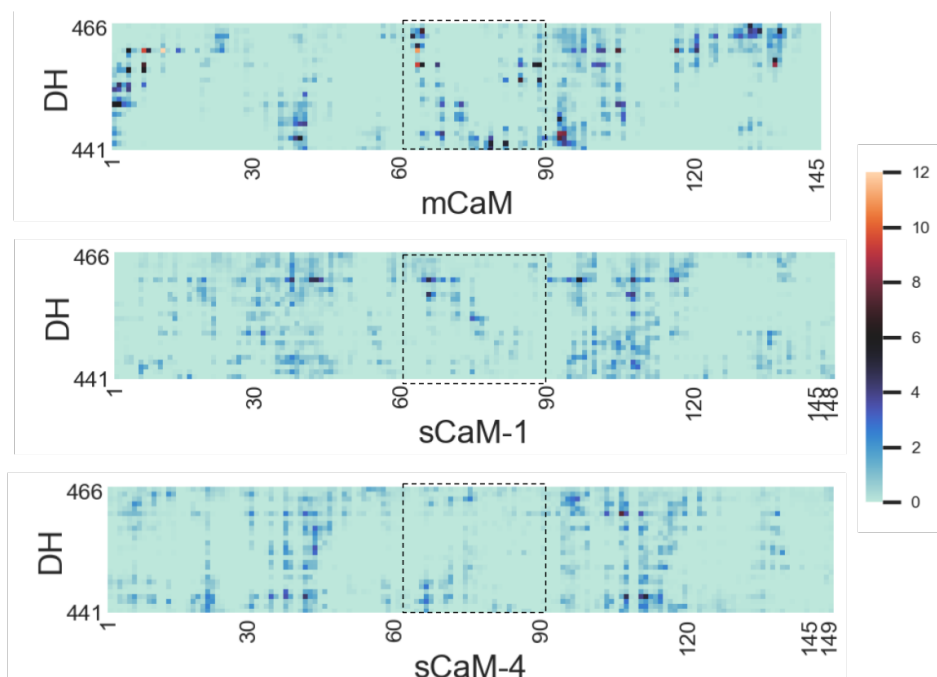


Figure S7: Contact map analysis between DH and CaM. The dashed boxes highlight the regions where sCaM-1 and mCaM have similar contact that differs from sCaM-4. The contact is defined by atoms that are within 7 Å. All contacts from the atoms comprising the residue are summed up to serve as residue-residue contact data shown here. Thus the color bar represents the total percentages of all atoms in one residue that have contacts with other residues

S2.5 The functional consequence of the distal helix/CaM interaction is quantitatively explained by a tether model that gives $[AID]_{\text{eff}}$ concentration as the indication of CaN activation.

Previously we have successfully used K_m to validate our simulations under the hypothesis that the binding of CaM to CaN RD alters the AID distribution around the active site, which can be shown in the apparent K_m measured by experiments [30]. To relate the distal helix/CaM interaction to actual CaN activation, we utilized a tethering model our lab developed in [30] which gives the effective AID concentration ($[AID]_{\text{eff}}$) near the catalytic site of CaN. Our hypothesis is that larger $[AID]_{\text{eff}}$ leads to larger K_D of substrate. Here we briefly summarize the key idea of the tethering model: in the absence of CaM, the regulatory domain of CaN is treated as a random coil and the AID is tethered by this coil (Fig. 1). The $[AID]_{\text{eff}}$ around the catalytic site is determined by the length of the coil: when the length shortens, the $[AID]_{\text{eff}}$ is reduced near the catalytic site, which reduces competition for substrate binding. Interacting with CaM will shorten the length of regulatory domain as the CaMBR and distal helix regions become rigid and locked by CaM, leading to much smaller $[AID]_{\text{eff}}$. Quantitatively, the $[AID]_{\text{eff}}$ is given by Van Valen et al [68],

$$[AID]_{\text{eff}} = \left(\frac{3}{4\pi\zeta L} \right)^{3/2} \exp \left(-\frac{3D^2}{4\zeta L} \right) \quad (\text{S1})$$

where D depicts the distance between N-termini of regulatory domain and catalytic site, which we assume is the distance between M387 and E481 in PDB 4OR9 [30]. ζ is the persistence length determined from literature [69]. L is linker length that we sampled from MD simulations. To obtain L values, we started from MD refined CaM/DH complex structures and added the AID fragment to the system and performed extensive all-atom MD to sample the relative position of

AID around CaM as originally described in [30]. We measured the distance between the center of mass (COM) of AID to the COM of CaM (Fig. S8a,b) and used the values as L in Eq. S1. The distance distribution shows that the three CaMs have L values ranging from 18 to 40 Å.

Putting the values in Fig. S8b as L into Eq. S1, we obtained the $[AID]_{\text{eff}}$ distribution for the three CaMs. The $[AID]_{\text{eff}}$ distribution is much like the CaM-AID distance distribution. For interpretation, we assume the $[AID]_{\text{eff}}$ value of the highest peak of mCaM is the critical value and the $[AID]_{\text{eff}}$ s that are smaller and larger than this critical value will exert activating and non-activating effect on CaN, respectively (the yellow line in Fig. S8c). We calculated the areas of the activating and non-activating peaks in Fig. S8c to assess the overall effect of these CaMs on CaN. By this measurement, we show that mCaM and sCaM-1 both have ~65 % activating effect on CaN while sCaM-4 only has 46 % (Fig. S8d). These trends suggest that sCaM-4's weaker CaN activating power is due to its more non-activating effect on CaN than sCaM-1 and mCaM. In other words, all CaMs are capable of activating CaN, but sCaM-4 to a lesser extent.

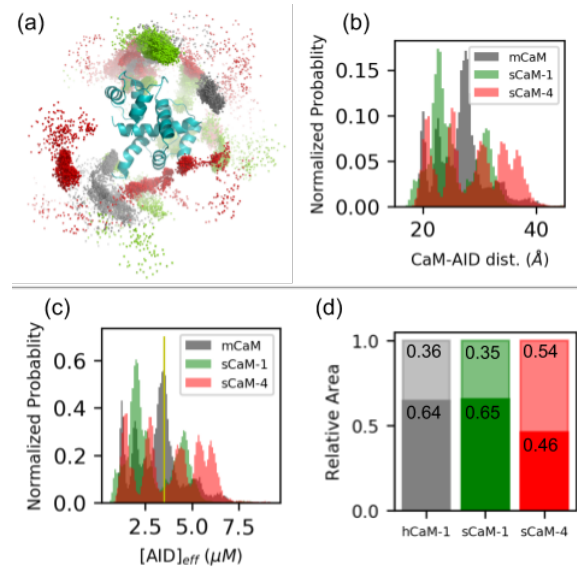


Figure S8: (a) Distribution of AID center of mass (COM) around CaM. (b) Distance between the COM of CaM and COM of AID. (c) The estimated effective AID concentration ($[AID]_{eff}$) calculated via the tethering model with the CaM-AID distance as input. (d) Ratio of activating area versus inhibitory area calculated from panel c. Using $[AID]_{eff} = 3.5 \mu M$ as thresholding value.