Markov State Modelling Reveals Heterogeneous Drug-Inhibition Mechanism of Calmodulin

Annie M. Westerlund[1†], Akshay Sridhar[1†], Leo Dahl[1], Alma Andersson[1,2], Anna-Yaroslava Bodnar[1], Lucie Delemotte[1*]

[1] Department of Applied Physics, Science for Life Laboratory, KTH Royal Institute of Technology, 17121, Solna, Sweden
[2] Division of Gene Technology, Science for Life Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden

* Corresponding Author: lucied@kth.se
† Contributed equally to the work

Submission to PLoS Computational Biology

Abstract

Calmodulin (CaM) is a calcium sensor which binds and regulates a wide range of target-proteins. This implicitly enables the concentration of calcium to influence many downstream physiological responses, including muscle contraction, learning and depression. The antipsychotic drug trifluoperazine (TFP) is a known CaM inhibitor. By binding to various sites, TFP prevents CaM from associating to target-proteins. However, the molecular and state-dependent mechanisms behind CaM inhibition by drugs such as TFP are largely unknown. Here, we build a Markov state model (MSM) from adaptively sampled molecular dynamics simulations and reveal the structural and dynamical features behind the inhibitory mechanism of TFP-binding to the C-terminal domain of CaM. We specifically identify three major TFP binding-modes from the MSM macrostates, and distinguish their effect on CaM conformation by using a systematic analysis protocol based on biophysical descriptors and tools from machine learning. The results show that depending on the binding orientation, TFP effectively stabilizes features of the calcium-unbound CaM, either affecting the CaM hydrophobic binding pocket, the calcium binding sites or the secondary structure content in the bound domain. The conclusions drawn from this work may in the future serve to formulate a complete model of pharmacological modulation of CaM, which furthers our understanding of how these drugs affect signaling pathways as well as associated diseases.

Author summary

Calmodulin (CaM) is a calcium-sensing protein which makes other proteins dependent on the surrounding calcium concentration by binding to these proteins. Such protein-protein interactions with CaM are vital for calcium to control many physiological pathways within the cell. The antipsychotic drug trifluoperazine (TFP) inhibits CaM's ability to bind and regulate other proteins. Here, we use molecular dynamics simulations together with Markov state modeling and machine learning to understand the structural and dynamical features by which TFP bound to the one domain of CaM prevents association to other proteins. We find that TFP encourages CaM to adopt a conformation that is like the one stabilized in absence of calcium: depending on the binding orientation of TFP, the drug indeed either affects the CaM hydrophobic binding pocket, the calcium binding sites or the secondary structure content in the domain. Understanding TFP binding is a first step towards designing better drugs targeting CaM.

Introduction

Cellular Ca^{2+}-signaling pathways affect a multitude of physiological processes, including learning and memory, muscle contraction, metabolism, and long-term depression [1,2]. Calmodulin (CaM) is a small and highly-conserved Ca^{2+}-sensing protein which contributes to the ubiquity of Ca^{2+}-signaling by binding and regulating a wide range of target-proteins. Indeed, CaM has been found to regulate voltage-gated ion channels, G-Protein coupled receptors, NMDA receptors and even proteins with opposite cellular function like kinases and phosphatases [3,4].

CaM consists of N-terminal and C-terminal lobes (N-CaM and C-CaM), as well as a linker connecting the two lobes (Fig. 1A). The binding promiscuity is partly due to the flexible linker, which permits a wrapping of the lobes around target-proteins, but also due to the conformational plasticity of the two lobes [3,5–7]. Each of the N-CaM and C-CaM lobes contain two (helix-loop-helix) EF-hand motifs (Fig. 1B) which each form a Ca^{2+} binding site [8–10]. The helices within each lobe enclose a hydrophobic pocket containing bulky aromatic residues and methionines that make up the interaction surface for target-proteins [11,12]. In line with the known binding promiscuity of CaM, various NMR, small-angle X-ray scattering (SAXS) and molecular dynamics (MD) simulation studies have demonstrated the protein's conformational flexibility [5,13–16]. Moreover, previous studies have demonstrated that Ca^{2+}-ion binding in the two lobes affects properties of the conformational landscape by inducing a change in the interhelical angles within the EF motifs [5,8,9,17]. Specifically, the Ca^{2+}-free (apo) CaM tends to adopt a closed conformation, shielding the hydrophobic residues from the surrounding solvent. Ca^{2+}-bound (holo) CaM, on the other hand, may adopt more open conformations where the target-protein interacting residues are exposed to solvent [5,13].

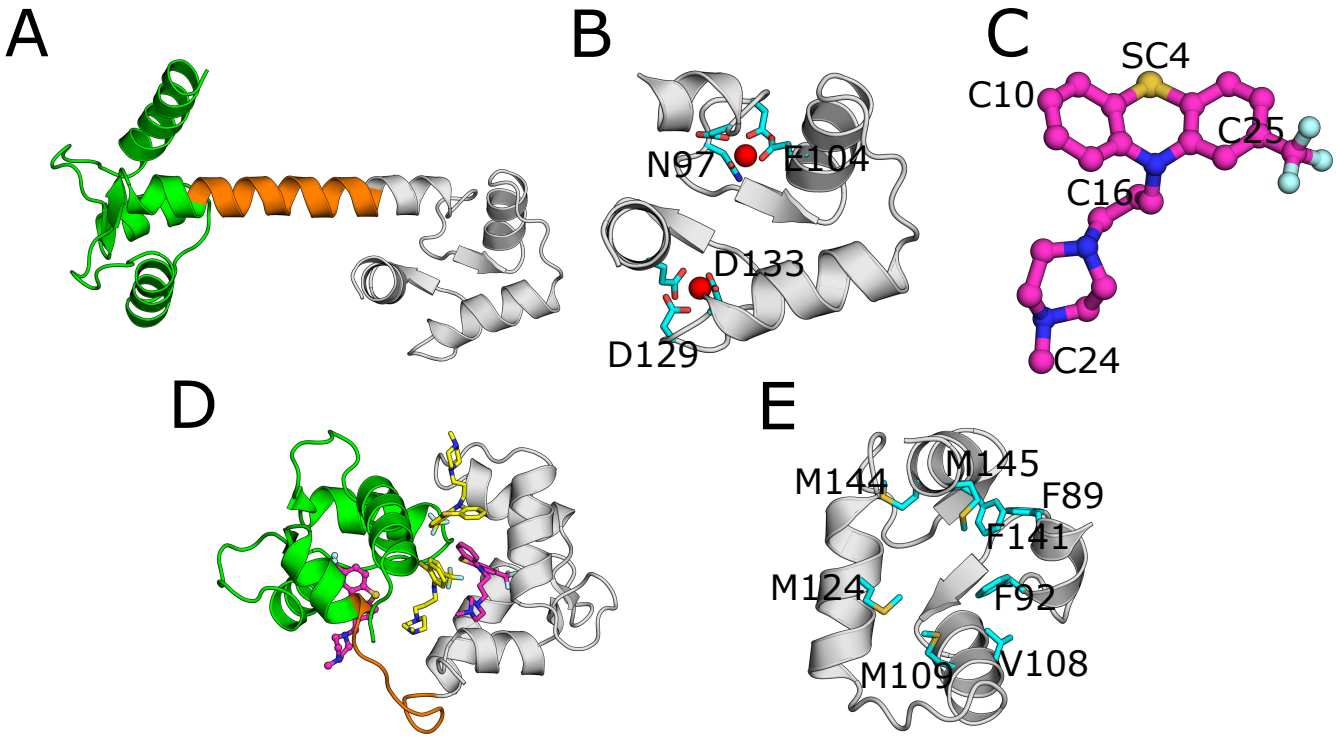

Figure 1

(A) The molecular structure of Calmodulin (PDB: 1CLL [8]). The N-CaM, linker and C-CaM are shown in green, orange and grey respectively. **(B)** The two EF-hand motifs making up the structure of C-CaM. The

bound Ca^{2+} ions are shown as red spheres together with their coordinating residues. **(C)** The molecular structure of Trifluoperazine. The five atoms distributed across the ligand used to featurize the trajectory are labelled. **(D)** Trifluoperazine binding configurations within calmodulin (PDB: 1LIN [23]). Ligands at the inter-lobe sites are colored yellow and those at the N-CaM/C-CaM intra-lobe sites are colored purple. **(E)** The hydrophobic methionine and aromatic residues making up the intra-lobe binding site within C-CaM.

Antipsychotic phenothiazines make up a class of drugs which are known CaM modulators [18–21]. By binding to CaM, the drugs effectively inhibit target-protein association. Interestingly, the hydrophobicity of these drugs correlates with the ability to inhibit CaM function, suggesting important interactions with the hydrophobic binding pockets of the two lobes [19]. Trifluoperazine (TFP, Fig. 1C) is a potent member of this family of drugs whose interaction with CaM is well-studied [22–25]. Apart from antipsychotic purposes, TFP has also been investigated as treatment for various cancer forms [20]. Crystallographic structures of TFP bound to CaM revealed that the drug is capable of binding to CaM with varying stoichiometries at both inter- and intra-lobe binding sites (Fig. 1D).

TFP binding at the lower-affinity interlobe site is suggested to inhibit target-protein association by causing a compact CaM structure [26–28]. In contrast, the effect on CaM's ability to bind target-proteins due to within-lobe drug binding is still poorly understood. In experimentally resolved CaM-TFP complex structures, the TFP within C-CaM occupies two different poses where the drug interacts with the hydrophobic methionine and aromatic residues (Fig. 1E) [25]. However, the mechanism and kinetics of transitions between the binding poses, as well as possible inhibitory effects on target-protein binding, remains unclear.

Molecular Dynamics (MD) simulations provide atomistic resolution of protein dynamic behavior and may be used to reveal the impact of drug-binding on protein function. However, conformational changes of CaM lobes, as well as phenothiazine binding, occur on the microsecond (μs) timescale; beyond the range of classical unbiased MD simulations [13,29]. Markov state modeling (MSM) is a computational technique that allows the stitching together of multiple short independent simulations and thus circumvent the sampling issue [30,31].

Here, we analyze the ensemble of TFP binding-modes in C-CaM, the most favorable binding site for TFP [32].This is done by constructing an MSM from ~21 μs adaptively sampled atomistic simulations of CaM in presence of a bound TFP molecule. The obtained free energy landscape and kinetic network then validate the sampled drug-binding modes and illustrate their coarse structural and dynamical attributes. Through further systematic and comparative analysis of the MSM macrostates with previously sampled trajectories of TFP-unbound CaM, we uncover the state-dependent molecular determinants through which TFP prevents CaM from binding target-proteins.

Materials and methods

Calmodulin-TFP system preparation and equilibration

To enable fast sampling of TFP bound to the multitude of C-CaM conformational states, the crystallographically resolved TFP binding poses (PDB ID: 4RJD) [25] were placed into three of the prominent C-CaM states identified in our prior work (state 2, 5 and 6) [5]. These states were extracted from MD simulations of the 3CLN structure [33]. Each CaM-TFP complex together with the bound Ca^{2+} ions were placed in a box with a solution of 22598 water molecules together with Na^{+} counter-ions. The system interaction energies were minimized until the maximum force on an atom was < 1000 kJ/mol and equilibrated by a 20 ps NVT ensemble with harmonic restraints of 1000 kJ/mol/nm^2 applied to protein and TFP atoms. Pressure was then relaxed, and the box scaled using a Berendsen barostat [34] for 20 ps with similar restraints on the protein and TFP atoms.

Replica exchange solute tempering and adaptive sampling simulations

Replica exchange solute tempering (REST) simulations [35,36] were initiated from each of the three bound states with 5 replicas run for 500 ns each. Exchanges between adjacent replicas were attempted every 4 ps and the 'temperature' span (303.15 K to 330 K) was determined using the temperature predictor of Patriksson and van der Spoel [37] by only considering the CaM-TFP atoms. Protein-ligand contacts of the lowest-temperature replica were then projected onto the two slowest degrees of freedom (called 'tICs' - see 'Selecting distance-based features and projecting the data onto slow degrees of freedom'). The two tICs were obtained from the continuous REST trajectories. To initiate the first round of 100 ns-long unbiased MD simulations of each C-CaM/TFP binding configuration, we sampled 10 representative conformations per configuration uniformly over the grid spanning the first two tICs.

To enhance sampling of unbiased MD simulations, we used a counts-based method to sample new initial configurations. This was done by first building an MSM (see 'Construction of a Markov state model') using the existing unbiased MD simulations and randomly selecting microstates to sample from. Specifically, the probability of sampling an initial configuration from a microstate was inversely proportional to its stationary probability [38]. Thus, microstates with a low stationary probability were more likely selected for seeding new simulations. This simple counts-based method has been demonstrated to be efficient for exploring new states [38,39]. Through eight rounds of such adaptive sampling, we accumulated a combined simulation time of 20.8 µs across the initial C-CaM/TFP binding configuration.

Molecular dynamics simulation parameters

The MD simulations were performed using Gromacs 2018 [40] with a timestep of 2 fs. For the REST simulations, we used Plumed 2.3.5 [41,42]. Long-range electrostatic interactions were calculated using the particle mesh Ewald method [43] and hydrogen-bond lengths were constrained using LINCS [44]. Pressure and temperature were maintained through the use of the

Parrinello–Rahman barostat [45] (1 bar) and Nose-Hoover (300 K) thermostat [46] respectively. The protein was described using the Charmm36 force field [47], water using the TIP3P model, Ca^{2+}-ions using the Charmm27 parameters of Liao et al. [48] and TFP parameters were generated using STaGE [49].

Selecting distance-based features and projecting the data onto slow degrees of freedom

To build the MSM, we first determined distance features to describe TFP binding-modes within C-CaM. The minimum distances between all C-CaM residues and five TFP atoms distributed across the ligand's structure (C25, C24, C10, S, C16) were calculated. We then constructed different distance-based feature types defined by various contact cutoffs (ranging from 5 to 8 Å). Moreover, we evaluated regular and inverse distances at these different contact cutoffs.

To maximize the kinetic variance within features, we evaluated the different feature types using the variational approach for Markov processes (VAMP2) score [50,51]. Each feature type was scored at 5 lag-times (2 ns, 5 ns, 10 ns, 15 ns, 20 ns) using the top ten eigenvalues and 5-fold cross-validation. The feature type maximizing the VAMP2 score across lag-times was then selected for building the MSM. SI Fig. 1 shows the mean scores for different lag-times. The scores suggest that these feature types are broadly similar. However, the feature type given by a 6 Å cutoff and inverse distances yielded the highest mean score across lag-times. This feature type was therefore selected for MSM construction.

The data was projected with time-lagged independent component analysis (tICA) [52] using time-lag correlation matrices (tICA time-lag of 20 ns). A tICA projection provides a low-dimensional representation of the data along the slowest degrees of freedom. Overall, we used the python package MDTraj [53] to compute distance-based features, while pyEMMA [54] was used for scoring the distance-based features, projecting the data, building the MSM and extracting the final macrostates.

Construction of a Markov state model

The 10 slowest time-lagged independent components (tICs) were obtained from the full set of simulations and used as input to k-means clustering to obtain 200 microstates, where the number of microstates was determined by increasing it in intervals of 50 until the results described herein were stable. To select a Markovian, or memoryless, lag-time for the MSM, we first estimated a set of transition probability matrices at varying lag-times. Each such matrix describes the probability of transitioning between microstates at a specific lag-time. By plotting the implied timescales [55,56] of the obtained transition matrices, we identified a Markovian lag-time where the timescales converge (15 ns - SI Fig. 2). The final MSM was constructed using this lag-time, and used to assign probability weights (π_i) to each trajectory frame i. Equilibrium averages of any system characteristic *O* could then be obtained with a weighted sum,

$$\langle O \rangle = \frac{\sum \pi_i O_i}{\sum \pi_i}.$$

The number of MSM metastable states was selected using eigenvalue spectral analysis of the MSM transition matrix (SI Fig. 3). PCCA++ spectral clustering [57,58] was then used to estimate the probability of a trajectory frame belonging to each macrostate. Finally, we identified core-states by assigning a trajectory frame to one of the macrostates if the state probability was >80%. The rest of the trajectory frames were left as transition points. The core states allow us to better distinguish the states due to reduced noise in state-definition.

Identification of important residues using machine learning

To pinpoint C-CaM residues participating in important residue-interactions of each identified TFP binding state, we used the *demystifying* toolkit [59]. This approach is based on explainable artificial intelligence (AI). In short, a machine learning model can be trained to recognize the three major states based on each frame's internal coordinates and metastable state-assignment. We may then ask the model which input features contributed the most to making a classification (state assignment) decision. Here, the input features were given by the inverse distances between all C-CaM residue pairs which were less than 6.5 Å in at least one frame and more than 6.5 Å in another. We trained random forest (RF) models using scikit-learn [60] and default parameters as determined in Fleetwood et al. [59] to recognize each metastable state given the input features and calculated the importance of residue-interactions for distinguishing between states by their Gini-importance (Figure 4A-C) [61]. Moreover, we calculated the Kullback-Leibler divergence of inverse residue-distance distributions for comparison. For both models, we trained one model per state in a one-versus-the-rest fashion. Finally, the per-residue importance was calculated as a sum over all the residues' interactions. The obtained residue importance was normalized between 0 and 1.

Results and discussion

Validation of Markov state model and identification of three major metastable states

We projected two-dimensional MSM free-energy landscapes along the three slowest time-dependent independent components (tICs). The best state separation is observed when the data is projected along tIC1-tIC3 (Figure 2A, SI Fig. 4). This free-energy landscape suggests the presence of three well-defined major local minima. The projection of the TFP-bound C-CaM structures onto the tIC space reveals how the free energy basins obtained using MD simulations compare with experimentally resolved structures: the 4RJD (Chain - or CaM molecule - B) and 1CTR structures indeed differ from 1A29 and 1LIN along tIC3. Interestingly, Chain A of the 4RJD structure is displaced from the other structures along tIC1, located close to a separate free-energy minimum. To further validate these free energy minima, we analyze the paths taken by the various sampled trajectories and confirm that we observe multiple transitions between the free-energy basins. Hence, the coverage of the conformational space by the combined set of MD simulations is extensive (SI Fig. 5).

Based on the MSM spectral analysis, we selected and extracted seven metastable states (SI Fig. 3). Figure 2B shows the resulting clustering, where each dot corresponds to a trajectory frame, colored by its macrostate assignment. The seven states (S_1-S_7), which are validated by Chapman-Kolmogorov tests (SI Fig. 6), are well separated in the tIC space (SI Fig. 4). We note, however, that the three major macrostates (S_5, S_6, S_7) account for ~98% of the stationary probability distribution (Fig. 2B). We thus proceeded to characterize these three major TFP binding-modes and their effect on C-CaM conformations.

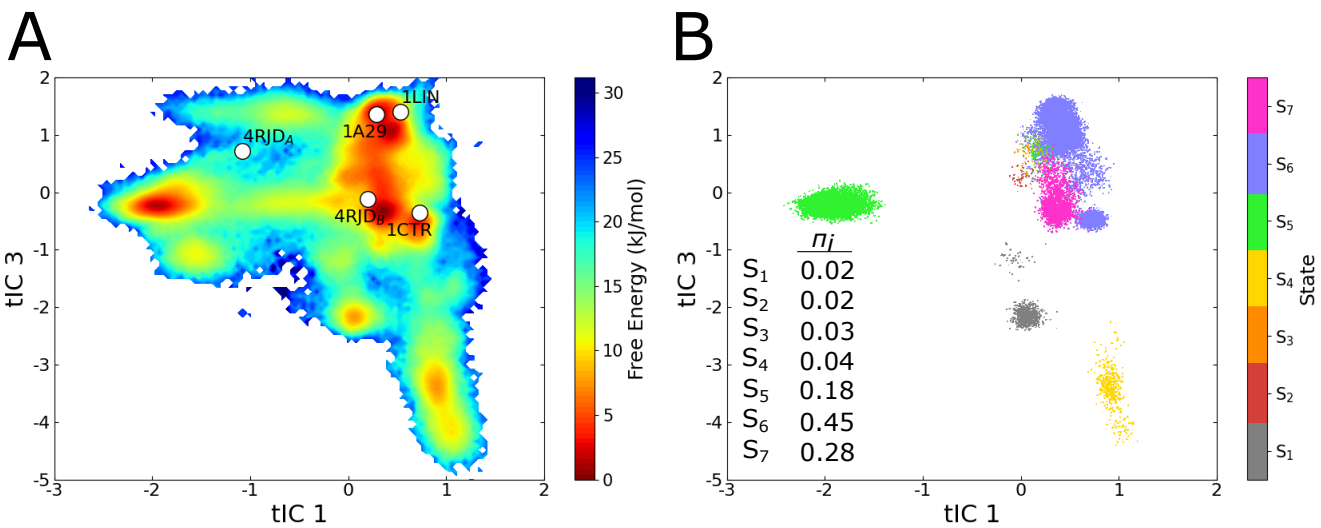

Figure 2

(A) Projection of the Markov state model free-energy surface along the slowest time-dependent independent components (tIC1 and tIC3). The projection of the 4RJD (Chains A & B) [25], 1LIN [23], 1A29 [24] and 1CTR [22] crystallographic structures onto the surface are shown as white dots. **(B)** S_1-S_7 macrostate assignments along the tICs obtained from PCCA++ clustering. Each trajectory frame is represented as a dot within the scatter plot. The stationary probability distributions of the seven macrostates are shown in the inset.

The metastable states represent distinct TFP binding modes

We first identified the coarse molecular features of each metastable state by calculating the ligand root mean square displacement (RMSD) across frames in the state after alignment of the C-CaM Cα atoms. The trajectory frame with the lowest mean RMSD was then selected as a representative structure. A visual inspection of the resulting representative binding modes provides an initial explanation of their separation in tIC-space (Fig. 3A). The trifluoromethyl (CF_3) moiety of TFP within S_5/S_6 and S_7, for example, occupy reversed orientations, either pointing towards or away from the hydrophobic pocket, respectively. Consistently, the experimentally resolved structures projected onto the free energy basins display similar 'flipped' configurations: the CF_3 group points towards the hydrophobic pocket in 4RJD (Chain A) [25], 1A29 [24] and 1LIN [23], and away from the pocket in 4RJD (Chain B) [25]and 1CTR [22].

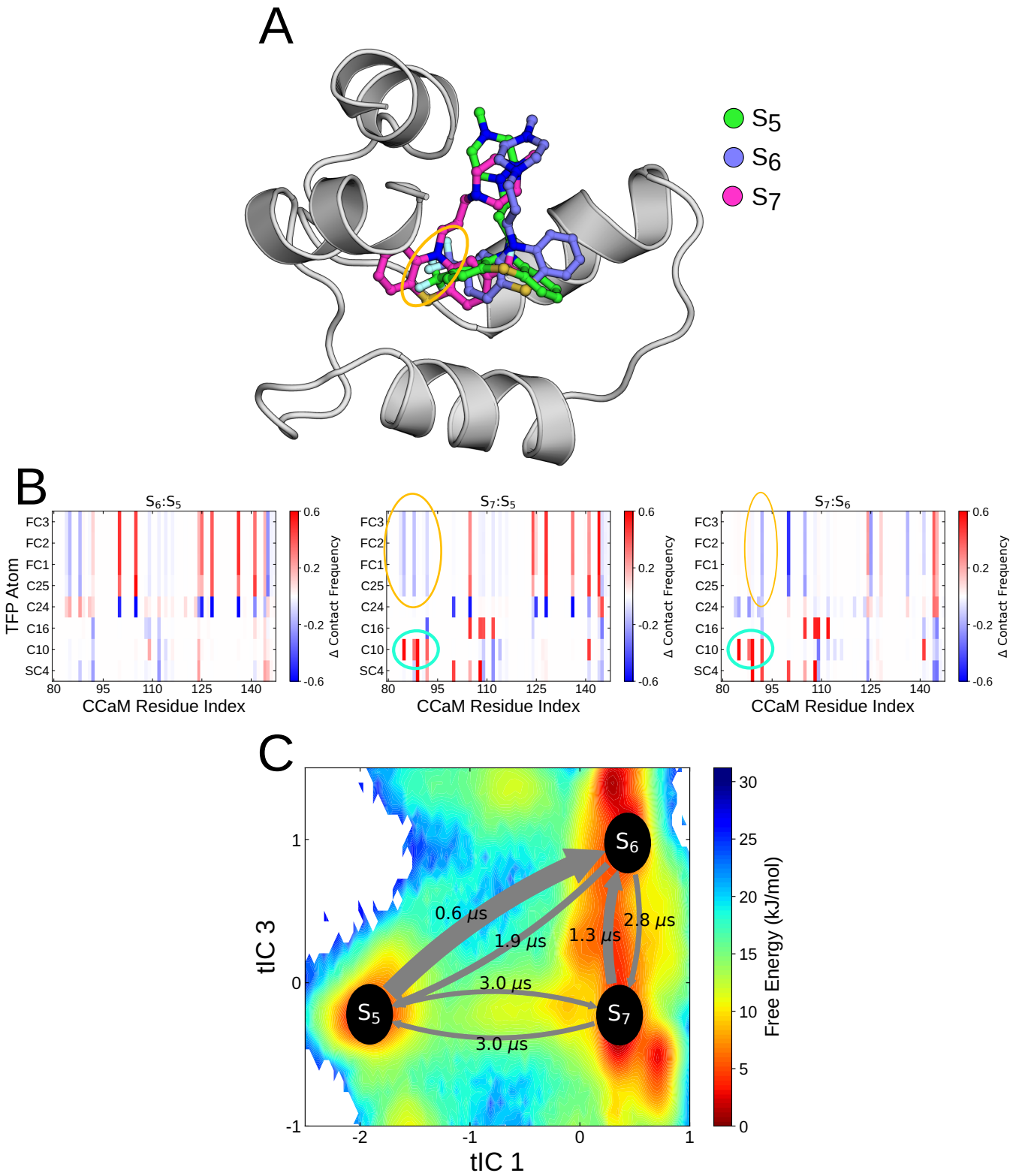

Figure 3

(A) Representative structures of the S_5, S_6 and S_7 macrostates aligned using the Cα C-CaM. The flipped orientation that buries the CF_3 moiety within the S_5 and S_6 macrostates is highlighted using an orange circle. **(B)** Relative contact frequency of the TFP atoms with C-CaM residues compared between the three S_5, S_6 and S_7 macrostates. The flipped orientation that buries the CF_3 moiety within the S_5 and S_6 macrostates is highlighted using an orange circle and the contrasting burying of the C10 atom with S_7 is highlighted using cyan circles. **(C)** Projection of the centres of the three populated macrostates onto the calculated free-energy surface along slowest time-dependent independent components (tIC1 and tIC3). Transition rates representing the mean first passage times between the macrostates are illustrated as arrows.

Clearly, state S_7 represents a separate binding pose compared to S_5 and S_6. A distinct difference between S_5 and S_6 is, however, difficult to assess by visual inspection. Moreover, unlike the static representative structures, the ensembles of microstates provide a dynamic picture of residue movements and interactions with TFP. To account for the conformational heterogeneity within metastable core-states, we therefore analyzed the ensemble of frames assigned to each state. By doing so, we set out to reveal subtle, but important, differences in the interactions between TFP and CaM. First, we validated the pose difference across metastable states by studying the state-dependent interactions between TFP atoms and C-CaM residues. The difference in contact frequencies between the three metastable states is shown in Fig. 3B. Consistent with a flipped orientation between S_5/S_6 and S_7, the C10 and C25 atoms located at opposite edges display similar delta contact frequencies within the S_7:S_5 and S_7:S_6 differences. In S_5 and S_6 , the C25 atom and the Fluorines are buried to interact with the F89/F92 residues at the base of the hydrophobic pocket (orange circle). Alternatively, in S_7, the C10 atom slots into the hydrophobic pocket and interacts with the buried aromatic residues (cyan circle). The differences between the S_5 and S_6 binding poses, however, remained difficult to characterize. Indeed, these differences are subtle, and the orientation of the phenothiazine tricyclic is identical in both states. The atomistic contact analysis instead suggests that the main difference between the two binding-modes primarily stems from the interactions between C24 and the residues making up the binding pocket.

To depict the dynamics of interconversion between TFP binding modes, we calculate the mean first passage times (MFPT) [62] between the three major macrostates. The results demonstrate that the S_5 and S_6 states interchange on the sub-microsecond timescale while exchanges with the S_7 state occur on a significantly longer timescale, consistent with the structural differences observed above. Interestingly, the MFPTs also show that a transition to S_7 from the S_5 state with the buried TFP typically requires a longer timescale than a transition from the less buried TFP pose in state S_6. This hints that although S_6 is globally more similar to S_5, it may represent an intermediate state with subtle features similar to both S_7 and S_5.

TFP binding-modes alter the C-CaM binding pocket by affecting local interactions at a Ca^{2+}-binding site

With the MSM, we were able to identify three major macrostates which agree with the experimentally resolved TFP-bound configurations. However, the backbone conformations of C-CaM in the representative structures are nearly identical, with a Cα RMSD of only 1.71 Å. Hence, we hypothesize that the effect of TFP binding-modes on C-CaM conformation is instead manifested through subtle state-dependent changes of residue interactions. To pinpoint CaM residues with distinct interaction patterns in the three states, we utilized machine learning and explainable artificial intelligence [59]. Briefly, we used a random forest (RF) classifier to recognize the three major states based on internal coordinates and metastable state-assignment, and then extracted the per-residue accumulated feature importance (see 'Materials and Methods').

Figures 4A-C show the state-dependent RF importance profiles. The Kullback-Leibler divergence approach validates these importance profiles by yielding profiles which share the main features picked up by RF (SI Fig. 7). As expected for S_5/S_6 and S_7 states with opposing orientations, the RF models identify M144 and F92 residues as important for discerning the states. This result is consistent with the C10 atom oriented towards these residues in the respective states [25]. In fact, both F92 and M144 are highly conserved residues [63], often interacting with target-proteins [7]. Earlier work showed that exposure of F92 to solvent favors deep binding to target-proteins [5], and that the F92A mutant is linked to a drop in ion affinity [64]. In addition to F92, we also note that F141, another highly conserved residue [63], has a relatively high importance in the S_7 profile. Not only is F141L a known long QT syndrome (LQTS) mutation which also decreases Ca^{2+}-affinity [65,66], but the interaction and stacking of F141 and F92 has been shown to associate with the transition from apo- to holo-like conformations [13]. Moreover, we observe an increased importance of V108 in S_7 (Fig. 4C). This residue is indeed important for the packing and repacking of these aromatic residues during the transition from apo- to holo-like conformations [13]. It was also found important to characterize the overall conformational ensemble of C-CaM [59]. Finally, we note that N111 has a relatively high importance in the S_7 profile (Fig. 4C). This residue has also been suggested to change interactions upon Ca^{2+}-binding [67]. These results indicate that the flipping of TFP at the C-CaM binding site causes rearrangements of hydrophobic residues like the transition between holo and apo-like CaM conformations.

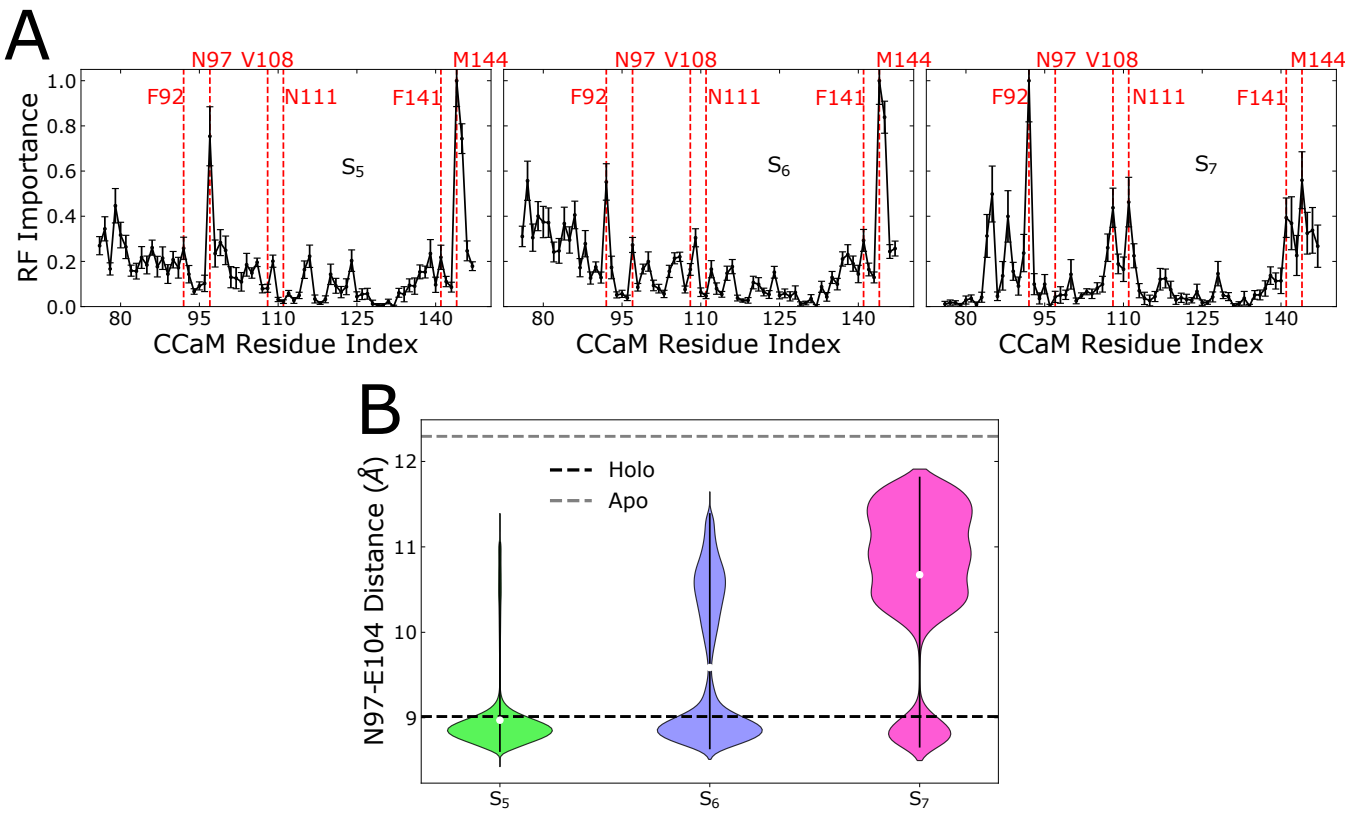

Figure 4

(A) The per-residue importance in discerning the S_5, S_6 and S_7 macrostates calculated using the supervised Random Forest method. Plots represent the mean values calculated from five-fold cross-validation and the standard deviations are plotted as error bars. Physiologically important residues and those with high importance values are illustrated using red dotted lines. **(B)** Distance between center-of-mass of the N97

and E104 residues making up the Ca^{2+} binding site within the S_5, S_6 and S_7 macrostates. Each violinplot spans the 5th and 95th percentile of distances and is weighted by the Markov state model probabilities. The median value for each macrostate is represented as a white dot. The inter-residue distances calculated from *apo*- (PDB: 1CFD [9]) and *holo*- (PDB: 1CLL [8]) calmodulin structures are shown as grey and black dotted lines respectively.

Next, we detect differences in the importance profiles of S_5 and S_6 (Figs. 4A-B). We find that N97 is important for recognizing S_5, while F92 appears to change its interaction pattern in S_6. Interestingly, the N97I mutant has been shown to lead to a reorientation of the hydrophobic domain [3], while the N97S mutant causes LQTS [66]. The latter has specifically been shown to affect the activation of the voltage-gated ion channel KCNQ1 likely due to the interactions between CaM and the voltage-sensor domain during channel activation [68,69]. The estimated importance of F92 is likely due to its position within the hydrophobic pocket and interaction with TFP (Fig. 3A). N97, on the other hand, is on a loop distant from the hydrophobic pocket, and instead coordinates one of the bound Ca^{2+} ions (Fig. 1B). To assess the state-dependence of residue interactions within this Ca^{2+}-binding site, we computed the distribution of distances between the N97 and its interacting residue E104 for each state (Fig. 4D). In the S_5 state, the residues are stable at a spatially close distance, a characteristic of holo CaM. The flipped TFP pose in state S_7, however, disrupts the ion-binding site and leads to a more *apo*-like conformation with a larger separation of the residues. Incidentally, TFP binding in C-CaM has been suggested to markedly reduce the Ca^{2+} affinity [32], possibly due to such disruptions at the ion binding site. However, consistent with the higher affinity of the second Ca^{2+}-binding site [70], the TFP binding modes display no noticeable disruption on its stability (D129-D133 – SI Fig. 8).

Flipping of the TFP molecule affects the hydrophobic pocket and is associated with a changed β-sheet structure content

As mentioned above, the process of binding and unbinding Ca^{2+} is intrinsically associated with CaM conformational changes. In the presence of ions, the lobe undergoes a 'repacking' of hydrophobic residues which is characterized by the movement of the F92 residue. This repacking results in a stacking of four aromatic residues (SI Fig. 9) [13]. Motivated by the findings from the state-dependent RF importance profiles, we therefore investigate how the different TFP binding-modes affect the hydrophobic pocket. To characterize this, we studied the interactions between F92, M124 and F141 – three residues from different helices whose sidechains subtend into the pocket . In fact, M124, together with M109, M144 and M145, make up the set of highly conserved methionines [63]which often participate in target-protein binding [11,12].

Figure 5A shows the hydrophobic packing along two inter-residue distances – the F92-F141 and F92-M124 distances. In contrast to the other two states, the slotting of the C10 atom with its aromatic ring in S_7 prevents the otherwise advantageous stacking interactions of F92 with the other aromatic residues. This leads to the adoption of a conformation reminiscent of the crystallographic apo-CaM (Fig. 5A). However, the observed binding mode-dependent effects are local. As such, the packing of the methionines lining the pocket is left unaffected (SI Fig. 10).

Viewing the S_7 as a state with local apo-like structural features is consistent with the findings that the transition from apo- to holo-like conformation is mainly characterized by the repacking of aromatic residues rather than changed interactions between the methionines [13,71]. In summary, our results suggest that the F92, F141 and N97 residues play particularly significant roles in the coupling between the hydrophobic domain and Ca^{2+}-site, and that the orientation of TFP controls the switch between these apo- and holo-like interactions.

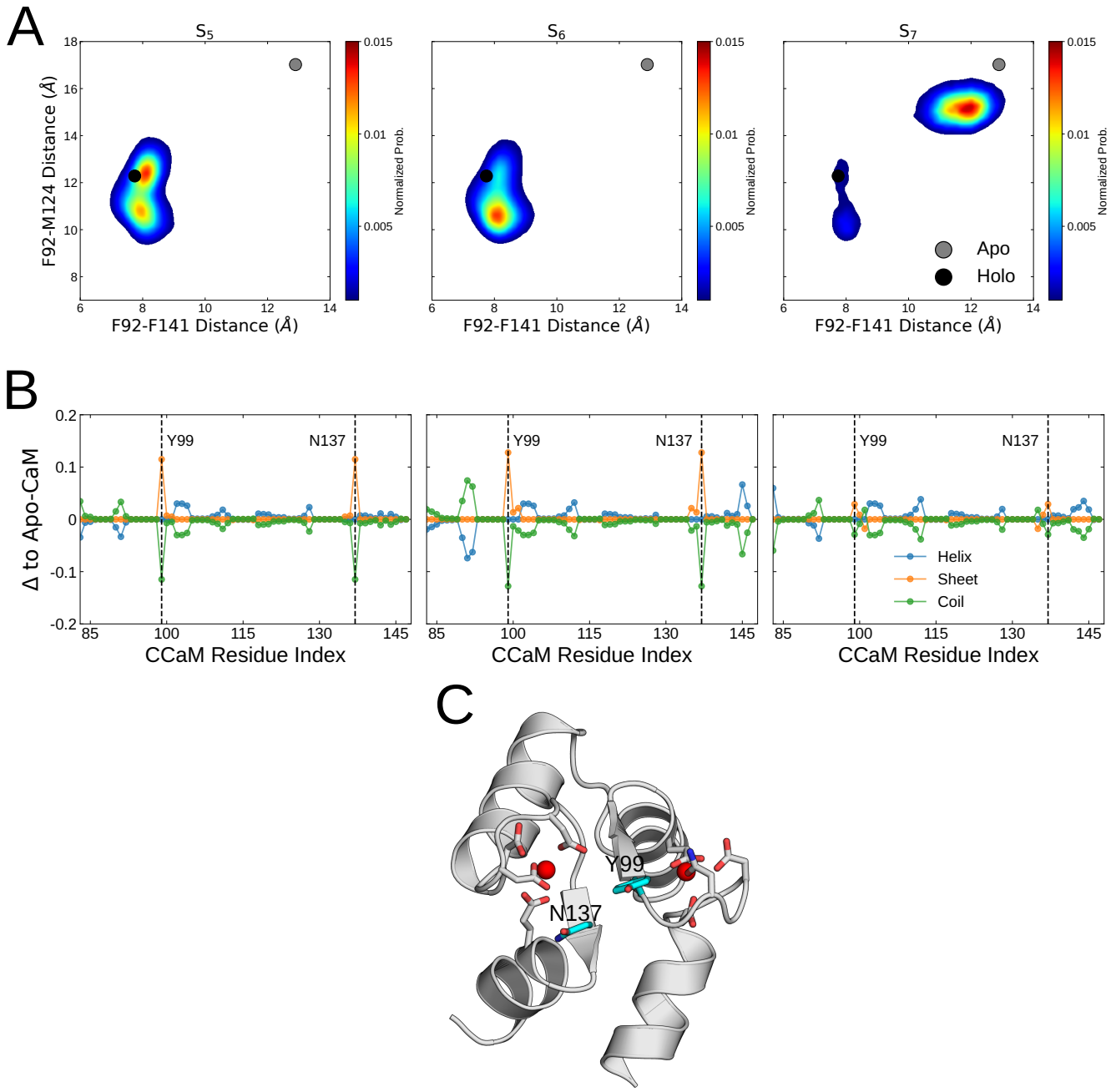

Figure 5

(A) Inter-residue distances between the sidechains of the F92, M124 and F141 within the S_5, S_6 and S_7 macrostates. The two-dimensional histograms are weighted by their Markov state model probabilities. The inter-residue distances calculated from *apo*- (PDB: 1CFD [9]) and *holo*- (PDB: 1CLL [8]) calmodulin structures are shown as grey and black dots respectively. **(B)** The per-residue secondary structure of C-CaM within the three populated macrostates compared to their behaviour within simulations of *apo*-CaM. The Y99 and N137 residues displaying enhanced sheet behaviour are illustrated as dotted lines. **(C)** The molecular structure of C-CaM illustrating the Y99-N137 antiparallel sheet situated adjacent to the two Ca^{2+} ion binding sites. The bound Ca^{2+} ions are shown as red spheres.

The two Ca^{2+} sites in C-CaM are cooperative with ion-binding at one interface resulting in up to a 10 kJ/mol enhancement in binding at the other interface [5,6]. Moreover, binding Ca^{2+} at the two sites yields a shift in the position of the β-sheet structure within the lobe [5]. To investigate whether the state-dependent TFP-binding can also affect the apo- and holo-like hallmarks of secondary structure content, we compared the per-residue secondary structure frequency in each state with that of the drug-unbound CaM ensemble (Fig. 5B) [5].

The $_{98}$GYISA$_{102}$ and $_{134}$GQVNY$_{138}$ loops are adjacent to the two ion-sites and transiently form antiparallel β-sheet structures [72]. Specifically, the Y99 and N137 residues participate more frequently in sheet structures in the S_5 and S_6 states (Fig. 5B). Recent work has shown that Y99 and N137 are important for the functional interactions between CaM and the activated-open state of KCNQ1 [69]. Together, the results provide a hypothesis on the mechanism of Calmodulin inhibition by the varying TFP binding modes. Apart from obstructing the hydrophobic pocket, in the S_5 and S_6 states, TFP could potentially hinder interactions between CaM and target-proteins by locking the crucial Y99 and N137 residues in a β-sheet structure. Conversely, within the S_7 state, TFP functions by enabling a local switch to an *apo*-like state characterized by differential interactions of the aromatic F89, F92 and F141 residues.

Conclusions

The rationalizing of protein-ligand interactions forms the basis of drug design. The structural heterogeneity of CaM, which forms the basis of its function, however, makes it a challenging drug target [73]. The inhibition of CaM requires a drug to bind across its conformational ensemble, including transitioning between *apo*- and *holo-like* states. To understand the molecular aspects of such an inhibition, we performed extensive unbiased and adaptively sampled MD simulations of TFP-bound (holo) CaM to build an MSM from which metastable TFP-binding modes could be extracted. The results demonstrate that TFP is a heterogeneous inhibitor which acts by blocking the different CaM binding pockets via various binding poses. The burial of the halogen moiety into the hydrophobic pocket, for example, correlates with C-CaM adopting subtle *holo*-like features around a Ca^{2+} binding site, while the 'flipped' TFP-binding mode could be attributed to apo-like features in this region. Moreover, we observed a general apo-like secondary structure content in C-CaM due to TFP binding. This, together with the observed state-dependent packing of hydrophobic residues in C-CaM, hints that TFP may prevent target-protein binding through subtle yet distinct blocking mechanisms. The different binding-poses may affect the coupling between the C-CaM β-sheet structure, a Ca^{2+}-binding site and the packing of aromatic residues.

An extension of the current work may address the construction of MSM models for understanding TFP binding to N-CaM. Since N-CaM is less dynamic than C-CaM [16], a comparison can shed further light on the inhibitory mechanism of TFP. Nonetheless, the notion that TFP uses different binding poses to block CaM is in line with the conformational heterogeneity and binding-promiscuity of CaM. This work thus deepens our understanding of how one drug can inhibit such

highly flexible binding pockets. The results presented here may serve as a stepping-stone towards a full understanding of the pharmacological modulation of CaM, with implications on signaling pathways and associated diseases.

Author contributions

Conceptualization AMW, AS, LDe ; Data Curation AMW, AS, LDa, AA, AB ; Formal Analysis AMW, AS, LDa, AA, AB ; Funding Acquisition LDe ; Investigation AS, AMW, LDa, AA, AB, LDe ; Methodology AS AMW ; Project Administration AS, AMW, LD ; Resources LDe ; Software AS ; Supervision AMW, LDe; Validation AMW, AS, LDa, AA, AB, LDe ; Visualization AMW, AS, AB ; Writing – Original Draft Preparation AMW, AS ; Writing – Review & Editing AMW, LDa, AA, AB, LDe

Acknowledgements

A.M.W. was funded by a Swedish e-research center (SeRC) COVID-19 transition grant. A.S. is funded by Marie-Sklodowska Curie Fellowship Lipopeutics (Grant Number 898762). LD acknowledges SciLifeLab, the Gustafsson foundation and the Swedish Research Council (VR 2018-04905 and 2019-02433) for funding. The MD simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) on Beskow at the PDC Center for High Performance Computing (PDC-HPC). We further wish to thank Magnus Lundborg for assisting in ligand parameterization.

Data Availability Statement

The data needed to reproduce the work herein can be retrieved from the Zenodo database with DOI 10.5281/zenodo.6626415.

References

1. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: New perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research. 2017;45. doi:10.1093/nar/gkw1092
2. Berridge MJ, Bootman MD, Roderick HL. Calcium signalling: Dynamics, homeostasis and remodelling. Nature Reviews Molecular Cell Biology. 2003. doi:10.1038/nrm1155
3. Tidow H, Nissen P. Structural diversity of calmodulin binding to its target sites. FEBS Journal. 2013. doi:10.1111/febs.12296
4. Mahling R, Rahlf CR, Hansen SC, Hayden MR, Shea MA. Ca2+-saturated calmodulin binds tightly to the N-terminal domain of A-type fibroblast growth factor homologous factors. Journal of Biological Chemistry. 2021;296. doi:10.1016/j.jbc.2021.100458
5. Westerlund AM, Delemotte L. Effect of Ca2+on the promiscuous target-protein binding of calmodulin. PLoS Computational Biology. 2018;14. doi:10.1371/journal.pcbi.1006072
6. Smith DMA, Straatsma TP, Squier TC. Retention of conformational entropy upon calmodulin binding to target peptides is driven by transient salt bridges. Biophysical Journal. 2012;103. doi:10.1016/j.bpj.2012.08.037
7. Villarroel A, Taglialatela M, Bernardo-Seisdedos G, Alaimo A, Agirre J, Alberdi A, et al. The ever changing moods of calmodulin: How structural plasticity entails transductional adaptability. Journal of Molecular Biology. 2014. doi:10.1016/j.jmb.2014.05.016
8. Chattopadhyaya R, Meador WE, Means AR, Quiocho FA. Calmodulin structure refined at 1.7 Å resolution. Journal of Molecular Biology. 1992;228. doi:10.1016/0022-2836(92)90324-D
9. Kuboniwa H, Tjandra N, Grzesiek S, Ren H, Klee CB, Bax A. Solution structure of calcium-free calmodulin. Nature Structural Biology. 1995;2. doi:10.1038/nsb0995-768
10. Linse S, Helmersson A, Forsen S. Calcium binding to calmodulin and its globular domains. Journal of Biological Chemistry. 1991;266. doi:10.1016/s0021-9258(18)92938-8
11. Siivari K, Zhang M, Palmer AG, Vogel HJ. NMR studies of the methionine methyl groups in calmodulin. FEBS Letters. 1995;366. doi:10.1016/0014-5793(95)00504-3
12. Gopalakrishna R, Anderson WB. The effects of chemical modification of calmodulin on Ca2+-induced exposure of a hydrophobic region. Separation of active and inactive forms of calmodulin. BBA - Molecular Cell Research. 1985;844. doi:10.1016/0167-4889(85)90099-0
13. Shukla D, Peck A, Pande VS. Conformational heterogeneity of the calmodulin binding interface. Nature Communications. 2016;7. doi:10.1038/ncomms10910
14. Westerlund AM, Delemotte L. Inflecs: Clustering free energy landscapes with gaussian mixtures. Journal of Chemical Theory and Computation. 2019. doi:10.1021/acs.jctc.9b00454
15. Jeon J, Yau WM, Tycko R. Millisecond Time-Resolved Solid-State NMR Reveals a Two-Stage Molecular Mechanism for Formation of Complexes between Calmodulin and a Target Peptide from Myosin Light Chain Kinase. J Am Chem Soc. 2020;142. doi:10.1021/jacs.0c11156
16. Tjandra N, Kuboniwa H, Ren H, Bax A. Rotational Dynamics of Calcium-Free Calmodulin Studied by 15N-NMR Relaxation Measurements. European Journal of Biochemistry. 1995;230. doi:10.1111/j.1432-1033.1995.1014g.x
17. Westerlund AM, Harpole TJ, Blau C, Delemotte L. Inference of Calmodulin's Ca2+-Dependent Free Energy Landscapes via Gaussian Mixture Model Validation. Journal of Chemical Theory and Computation. 2018;14. doi:10.1021/acs.jctc.7b00346
18. Lucchesi PA, Scheid CR. Effects of the anti-calmodulin drugs calmidazolium and trifluoperazine on 45Ca transport in plasmalemmal vesicles from gastric smooth muscle. Cell Calcium. 1988;9. doi:10.1016/0143-4160(88)90028-0
19. Weiss B, Prozialeck WC, Wallace TL. Interaction of drugs with calmodulin. Biochemical, pharmacological and clinical implications. Biochemical Pharmacology. 1982;31. doi:10.1016/0006-2952(82)90104-6

20. Manoharan GB, Okutachi S, Abankwa D. Potential of phenothiazines to synergistically block calmodulin and reactivate PP2A in cancer cells. PLOS ONE. 2022;17: e0268635-. Available: https://doi.org/10.1371/journal.pone.0268635
21. Weiss B, Prozialeck W, Cimino M, Sellinger Barnette M, Wallace TL. PHARMACOLOGICAL REGULATION OF CALMODULIN. Ann N Y Acad Sci. 1980;356. doi:10.1111/j.1749-6632.1980.tb29621.x
22. Cook WJ, Walter LJ, Walter MR. Drug Binding by Calmodulin: Crystal Structure of a Calmodulin-Trifluoperazine Complex. Biochemistry. 1994;33. doi:10.1021/bi00255a006
23. Vandonselaar M, Hickie RA, Quail JW, Delbaere LTJ. Trifluoperazine-induced conformational change in Ca2+-calmodulin. Nature Structural Biology. 1994;1. doi:10.1038/nsb1194-795
24. Vertessy BG, Harmat V, Böcskei Z, Náray-Szabó G, Orosz F, Ovádi J. Simultaneous binding of drugs with different chemical structures to Ca2+-calmodulin: Crystallographic and spectroscopic studies. Biochemistry. 1998;37. doi:10.1021/bi980795a
25. Feldkamp MD, Gakhar L, Pandey N, Shea MA. Opposing orientations of the anti-psychotic drug trifluoperazine selected by alternate conformations of M144 in calmodulin. Proteins: Structure, Function and Bioinformatics. 2015;83. doi:10.1002/prot.24781
26. Matsushima N, Hayashi N, Jinbo Y, Izumi Y. Ca2+-bound calmodulin forms a compact globular structure on binding four trifluoperazine molecules in solution. Biochemical Journal. 2000;347. doi:10.1042/0264-6021:3470211
27. Osawa M, Kuwamoto S, Izumi Y, Yap KL, Ikura M, Shibanuma T, et al. Evidence for calmodulin inter-domain compaction in solution induced by W-7 binding. FEBS Letters. 1999;442. doi:10.1016/S0014-5793(98)01637-8
28. Matsushima N, Hayashi N, Watanabe N, Jinbo Y, Izumi Y. Binding of trifluoperazine to apocalmodulin revealed by a combination of small-angle X-ray scattering and nuclear magnetic resonance. Journal of Applied Crystallography. 2007;40. doi:10.1107/S0021889807002117
29. Marshak DR, Lukas TJ, Watterson DM. Drug-Protein Interactions: Binding of Chlorpromazine to Calmodulin, Calmodulin Fragments, and Related Calcium Binding Proteins. Biochemistry. 1985;24. doi:10.1021/bi00322a020
30. Pande VS, Beauchamp K, Bowman GR. Everything you wanted to know about Markov State Models but were afraid to ask. Methods. 2010;52: 99–105. doi:10.1016/j.ymeth.2010.06.002
31. Chodera JD, Noé F. Markov state models of biomolecular conformational dynamics. Current Opinion in Structural Biology. 2014;25: 135–144. doi:10.1016/j.sbi.2014.04.002
32. Feldkamp MD, O'Donnell SE, Yu L, Shea MA. Allosteric effects of the antipsychotic drug trifluoperazine on the energetics of calcium binding by calmodulin. Proteins: Structure, Function and Bioinformatics. 2010;78. doi:10.1002/prot.22739
33. Babu YS, Bugg CE, Cook WJ. Structure of calmodulin refined at 2.2 Å resolution. Journal of Molecular Biology. 1988;204. doi:10.1016/0022-2836(88)90608-0
34. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. Journal of Chemical Physics. 1984;81: 3684. doi:10.1063/1.448118
35. Wang L, Friesner RA, Berne BJ. Replica exchange with solute scaling: A more efficient version of replica exchange with solute tempering (REST2). Journal of Physical Chemistry B. 2011;115: 9431–9438. doi:10.1021/jp204407d
36. Bussi G. Hamiltonian replica exchange in GROMACS: A flexible implementation. Molecular Physics. 2014;112: 379–384. doi:10.1080/00268976.2013.824126
37. Patriksson A, van der Spoel D. A temperature predictor for parallel tempering simulations. Physical Chemistry Chemical Physics. 2008;10: 2073–2077. doi:10.1039/b716554d
38. Weber JK, Pande VS. Characterization and rapid sampling of protein folding Markov state model topologies. Journal of Chemical Theory and Computation. 2011;7. doi:10.1021/ct2004484
39. Zimmerman MI, Porter JR, Sun X, Silva RR, Bowman GR. Choice of Adaptive Sampling Strategy Impacts State Discovery, Transition Probabilities, and the Apparent Mechanism of Conformational Changes. Journal of Chemical Theory and Computation. 2018;14. doi:10.1021/acs.jctc.8b00500

40. Abraham MJ, Murtola T, Schulz R, Pall S, Smith JC, Hess B, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1: 19–25. doi:10.1016/j.softx.2015.06.001
41. Bonomi M, Branduardi D, Bussi G, Camilloni C, Provasi D, Raiteri P, et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. Computer Physics Communications. 2009;180: 1961–1972. doi:10.1016/j.cpc.2009.05.011
42. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: New feathers for an old bird. Computer Physics Communications. 2014;185: 604–613. doi:10.1016/j.cpc.2013.09.018
43. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. Journal of Chemical Physics. 1993;98: 10089. doi:10.1063/1.464397
44. Hess B. P-LINCS: A parallel linear constraint solver for molecular simulation. Journal of Chemical Theory and Computation. 2008;4: 116–122. doi:10.1021/ct700200b
45. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied Physics. 1981;52: 7182–7190. doi:10.1063/1.328693
46. Hoover WG, Holian BL. Kinetic moments method for the canonical ensemble distribution. Physics Letters, Section A: General, Atomic and Solid State Physics. 1996;211. doi:10.1016/0375-9601(95)00973-6
47. Huang J, Mackerell AD. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. Journal of Computational Chemistry. 2013;34. doi:10.1002/jcc.23354
48. Liao J, Marinelli F, Lee C, Huang Y, Faraldo-Gómez JD, Jiang Y. Mechanism of extracellular ion exchange and binding-site occlusion in a sodium/calcium exchanger. Nature Structural and Molecular Biology. 2016;23. doi:10.1038/nsmb.3230
49. Lundborg M, Lindahl E. Automatic GROMACS topology generation and comparisons of force fields for solvation free energy calculations. J Phys Chem B. 2015;119: 810–23. doi:10.1021/jp505332p
50. Noé F, Clementi C. Kinetic Distance and Kinetic Maps from Molecular Dynamics Simulation. Journal of Chemical Theory and Computation. 2015;11. doi:10.1021/acs.jctc.5b00553
51. Wu H, Noé F. Variational Approach for Learning Markov Processes from Time Series Data. Journal of Nonlinear Science. 2020;30. doi:10.1007/s00332-019-09567-y
52. Pérez-Hernández G, Paul F, Giorgino T, de Fabritiis G, Noé F. Identification of slow molecular order parameters for Markov model construction. Journal of Chemical Physics. 2013;139. doi:10.1063/1.4811489
53. McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, et al. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. Biophysical Journal. 2015;109: 1528–1532. doi:10.1016/j.bpj.2015.08.015
54. Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, et al. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. Journal of Chemical Theory and Computation. 2015;11: 5525–5542. doi:10.1021/acs.jctc.5b00743
55. Swope WC, Pitera JW, Suits F. Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. Journal of Physical Chemistry B. 2004;108. doi:10.1021/jp037421y
56. Swope WC, Pitera JW, Suits F, Pitman M, Eleftheriou M, Fitch BG, et al. Describing protein folding kinetics by molecular dynamics simulations. 2. Example applications to alanine dipeptide and a β-hairpin peptide. Journal of Physical Chemistry B. 2004;108. doi:10.1021/jp037422q
57. Deuflhard P, Weber M. Robust Perron cluster analysis in conformation dynamics. Linear Algebra and Its Applications. 2005;398. doi:10.1016/j.laa.2004.10.026
58. Röblitz S, Weber M. Fuzzy spectral clustering by PCCA+: Application to Markov state models and data classification. Advances in Data Analysis and Classification. 2013;7. doi:10.1007/s11634-013-0134-6
59. Fleetwood O, Kasimova MA, Westerlund AM, Delemotte L. Molecular Insights from Conformational Ensembles via Machine Learning. Biophysical Journal. 2020;118. doi:10.1016/j.bpj.2019.12.016
60. Pedregosa F, Michel V, Grisel O, Blondel M, Prettenhofer P, Weiss R, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12.

61. Breiman L. Random forests. Machine Learning. 2001;45. doi:10.1023/A:1010933404324
62. Polizzi NF, Therien MJ, Beratan DN. Mean First-Passage Times in Biology. Israel Journal of Chemistry. 2016. doi:10.1002/ijch.201600040
63. Halling DB, Liebeskind BJ, Hall AW, Aldrich RW. Conserved properties of individual Ca2+-binding sites in calmodulin. Proc Natl Acad Sci U S A. 2016;113. doi:10.1073/pnas.1600385113
64. Wang K, Holt C, Lu J, Brohus M, Larsen KT, Overgaard MT, et al. Arrhythmia mutations in calmodulin cause conformational changes that affect interactions with the cardiac voltage-gated calcium channel. Proc Natl Acad Sci U S A. 2018;115. doi:10.1073/pnas.1808733115
65. Crotti L, Johnson CN, Graf E, de Ferrari GM, Cuneo BF, Ovadia M, et al. Calmodulin mutations associated with recurrent cardiac arrest in infants. Circulation. 2013;127. doi:10.1161/CIRCULATIONAHA.112.001216
66. Limpitikul WB, Dick IE, Joshi-Mukherjee R, Overgaard MT, George AL, Yue DT. Calmodulin mutations associated with long QT syndrome prevent inactivation of cardiac L-type Ca2+ currents and promote proarrhythmic behavior in ventricular myocytes. Journal of Molecular and Cellular Cardiology. 2014;74. doi:10.1016/j.yjmcc.2014.04.022
67. Shimoyama H, Takeda-Shitaka M. Residue-residue interactions regulating the Ca2+-induced EF-hand conformation changes in calmodulin. Journal of Biochemistry. 2017;162. doi:10.1093/jb/mvx025
68. Sun J, MacKinnon R. Cryo-EM Structure of a KCNQ1/CaM Complex Reveals Insights into Congenital Long QT Syndrome. Cell. 2017;169. doi:10.1016/j.cell.2017.05.019
69. Kang PW, Westerlund AM, Shi J, White KMF, Dou AK, Cui AH, et al. Calmodulin acts as a state-dependent switch to control a cardiac potassium channel opening. Science Advances. 2020;6. doi:10.1126/sciadv.abd6798
70. Malmendal A, Evenäs J, Forsén S, Akke M. Structural dynamics in the C-terminal domain of calmodulin at low calcium levels. Journal of Molecular Biology. 1999;293. doi:10.1006/jmbi.1999.3188
71. Kelly KL, Dalton SR, Wai RB, Ramchandani K, Xu RJ, Linse S, et al. Conformational Ensembles of Calmodulin Revealed by Nonperturbing Site-Specific Vibrational Probe Groups. Journal of Physical Chemistry A. 2018;122. doi:10.1021/acs.jpca.8b00475
72. Finn BE, Forsén S. The evolving model of calmodulin structure,function and activation. Structure. 1995;3. doi:10.1016/S0969-2126(01)00130-7
73. Menyhard D, Keseru G, Naray-Szabo G. Calmodulin in Complex with Proteins and Small Molecule Ligands: Operating with the Element of Surprise; Implications for Structure-Based Drug Design. Current Computer Aided-Drug Design. 2009;5. doi:10.2174/157340909789577874