Toward Understanding Calmodulin Plasticity by Molecular Dynamics

Abstract

Aim: Calmodulin interacts in many different ways with its ligands. We aim to shed light on its plasticity analysing the changes followed by the linker region and the relative position of the lobes using conventional Molecular Dynamics (cMD), accelerated MD (aMD) and scaled MD (sMD). **Materials & Methods**: Three different structures of calmodulin are compared, obtaining a total of 2.5 μs of molecular dynamics, which have been analysed using the principal component analysis and clustering methodologies **Results**: sMD simulations reach conformations that cMD is not able to, without compromising the stability of the protein. On the other hand, aMD requires optimization of the setup parameters to be useful. **Conclusion**: SMD is useful to study flexible proteins, highlighting those factors that justify its promiscuity.

Keywords: Scaled Molecular Dynamics, calmodulin, CaM plasticity

1. Introduction

Calmodulin (CaM) is a small (16.7 kDa), ubiquitous acidic protein involved in many processes within the cell [1]. It has been described to act as a calcium sensor, as it can bind two calcium ions modifying its shape, hence acquiring the *holo* structure. Originally CaM was thought to need these calcium ions to regulate other proteins. It has been demonstrated that the *apo* structure (unbound to calcium ions) does also bind to other proteins [2].

CaM can bind to, at least, 30 different proteins such as nitric oxide synthase (NOS) [3], phosphodiesterase myosin light chain kinase (MLCK) or K-Ras4B [4,5], all of them known as calmodulin binding proteins (CaMBPs). Certain proteins pertaining to this group share domains or patterns (1-5-8-14 CaM binding motif) and bind to CaM through a similar mechanism; some others, however, do not [6]. The key to understand how a single protein can modify itself to interact with such a large (and heterogeneous) group of proteins lays in its own structure: two lobes connected by a linker region. Each lobe consists of an EF-hand motif (helix-loop-

helix). Both lobes can modify their shape mildly, exposing hydrophobic domains when bound to calcium ions, but what really allows CaM to acquire the needed conformations is its linker region; a long extended alpha helix that can bend in order to display the N and C lobes as necessary [7]. Thus, we will focus our study in the analysis of these three basic regions.

Even though several structures of complexes of CaM with a CaMBPs are available since a few years ago [8], the subtleties of the interaction between CaM and its targets remain unclear. With computational techniques on their heydays and CaM still in the spotlight thanks to the diverse roles it seems to play [9], diverse studies have been conducted mainly focused on its interactions with other proteins highlighting the contribution of the N and C lobes of CaM to binding [3,10] or in the effect of calcium bonding/removal [11]. Up to date it remains unknown if CaM has a wide pool of conformations or if its targets modify its structure in order to help the protein to reach certain conformations. It seems that a mixture of both mechanisms would fit with current available data, that is, there exists a wide pool of conformations of CaM initially accessible, which are lightly modified by CaMBPs [12,13].

The linker region of CaM has been studied previously with Molecular Dynamics (MD) [14,15], attempting to describe the relevance of the residues of the linker region or the behaviour of the lobes when binding calcium ions. Nevertheless, to the best of our knowledge, an exhaustive study of CaM flexibility has not been tackled yet. In the present work, we have performed accumulated 2.5 μs of simulations, using three diverse methodologies: conventional Molecular Dynamics (cMD), scaled Molecular Dynamics (sMD) [16] and accelerated Molecular Dynamics (aMD) [17]. Both aMD and sMD are new methodologies that enhance the sampling of cMD by adding an energy boost. They have already been applied successfully to describe the conformational behaviour of different proteins [18,19]. All these methodologies have been essayed on three different systems in order to determine their different capacity to sample the CaM conformational space determined by the behaviour of its linker region. The three selected initial systems correspond to calmodulin in diverse shapes according to the structure adopted by the linker region and with or without a peptide target bound.

2. Methods

The calculations described in the present work were carried out at molecular mechanics level using the Amber ff99SB [20] force field with the AMBER v14 [21] suite of programs. The calcium parameters were downloaded from the Bryce group database (http://research.bmh.manchester.ac.uk/bryce/amber). RMSD, clustering and dihedral angle analyses were performed using the cpptraj program [22].

2.1 Preparation of CaM and CaM/HIV-1 peptide complex

Three systems containing calmodulin have been studied in this work. For two of them the three-dimensional structure was extracted from the Protein Data Bank (http://www.rcsb.org/pdb) with access codes 1CLL and 2MGU respectively. The 1CLL structure is composed of only calmodulin and calcium ions and it was selected because it is the first X-ray crystallographic structure of calmodulin with a completely extended central helix structure, therefore representing an ideal starting point to analyse what happens when other factors, like solvation or dynamics, are considered. The 2MGU code corresponds to an NMR structure that includes calmodulin, calcium ions and a 36 amino acids long oligopeptide extracted from the binding domain of the HIV-1 matrix protein (from now on 2MGUpep). The third system was obtained by removing the above-mentioned peptide from the 2MGU complex (from now on 2MGU), which will permit to analyse how the free structure evolves from a point far from the ideal free situation. After an initial step where all the structures were protonated using the Protonate function of MOE [23] with a pH of 7, we prepared the three systems (1CLL, 2MGU and 2MGUpep) following the same procedure: all complexes were placed in a cubic periodic box filled with TIP3P [24] water molecules, imposing a minimal distance of 15 Å between the protein and the box walls. Water molecules closer than 2.2 Å to the protein were removed and neutralizing counterions were added at positions of lowest electrostatic potential using the LEaP module of AmberTools15 [21].

2.2 Molecular Dynamics of CaM and CaM/HIV-1 peptide complex

All the calculation were done using de GPU version of the PMEMD module of AMBER v14 (programme) suite [21]. First, each system was relaxed following a multistep procedure using the conjugate gradient (CG) algorithm to minimize the initial structures. In the first step only water molecules and counterions were allowed to move keeping fixed the rest of the system with a force constant of 10 kcal mol^{-1} $Å^{-1}$. During the second step the side chains of CaM are also minimized keeping the backbone atoms fixed with a decreasing force constant of 10, 5, 1 and 0.5 kcal mol^{-1} $Å^{-1}$ (to remove the restrain). In the last step the whole system was minimized without restrictions. For the two first steps 5000 iterations of CG were performed, while 10000 were used for the last step.

Once the complexes were minimized the final structures were used as starting points for MD simulations. Trajectories were carried out in the NVT ensemble at a constant temperature of 300 K. The Langevin [25,26] thermal bath with a time coupling constant of 2 ps^{-1} was used to keep fixed the temperature. The SHAKE algorithm [27] was used to constraint all the bonds involving hydrogen atoms in order to (be able) to use a 2 fs time integration step in all the simulations. Long-range electrostatic interactions were treated with the particle-mesh Ewald method [26] with a cut-off of 10 Å, also used for the non-bonded hydrophobic interactions.

After minimization, structures were heated to 300K step wise at a rate of 30K every 20 ps, restraining the protein backbone atoms with a force constant of 0.5 kcal mol^{-1} $Å^{-1}$. Next, a 200 ps MD simulation in the NPT ensemble was performed without any restraint to increase the density of each system. Then, a short MD run of 2 ns length within the NVT assembly was done to allow small structural readjustments. Finally, in order to obtain a better sampling of the conformational space, [28] three 250 ns MDs were obtained for each system.

2.3 Accelerated and Scaled Molecular Dynamics

An enhanced conformational sampling of our systems was pursued through the use of aMD [17] and sMD [16]. The aMD simulations provide a boost energy

ΔV(r) when the potential energy of the system V(r) is below a threshold energy level E, as shown in eqn (1). The boost potency depends on the parameter α.

$$\Delta V(r) = \frac{(E - V(r))^2}{\alpha + (E - V(r))} \qquad (1)$$

For the aMD calculations we used a boost for both the total potential and the dihedral force field terms. The two parameters needed (E and α) for each boost were defined taking into account the number of residues in the system (N_{res}) and the number of atoms in the system (N_{atoms}) as shown in eqns (2) to (5):

$$E_{Dihed} = \langle V_{Dihed}(r) \rangle + 2.5 \times N_{res} \qquad (2)$$

$$\alpha_{Dihed} = 1.25 \times N_{res} \qquad (3)$$

$$E_{tot} = \langle V_{tot}(r) \rangle + 0.17 \times N_{atoms} \qquad (4)$$

$$\alpha_{tot} = 0.17 \times N_{atoms} \qquad (5)$$

As a starting point for this accelerated molecular dynamics run, we used the coordinates obtained after 5 ns of conventional molecular dynamics and the average value for the dihedral potential energy $\langle V_{Dihed}(r) \rangle$ and for the total potential energy $\langle V_{tot}(r) \rangle$ obtained from these 5 ns of cMD.

The sMD approach, instead of adding a boost potential as in aMD, modifies the potential energy surface (PES) of the system (by) scaling V(r) by a factor λ that ranges from 0 to 1: $V^*(r) = \lambda V(r)$.

2.4 Principal component analysis

In order to have the same topology in all the systems studied, we first prepared them to have exactly the same number of atoms (only calmodulin protein with calcium ions). All trajectories were then superimposed to the first NMR structure of 2MGU (without the peptide) using as reference residues 121-125 (which belong to an α-helix) of CaM, which is the region with the lowest RMSF of the protein. Principal component spaces were constructed from all the simulations available, one regarding the N-ter lobe that comprises residues 4 to 64 (N-ter) and another studying the behaviour of the helix of the linker region including residues 65 to 92 (Helix). The trajectories of each set of simulations were then projected onto the first

and second principal components (PCs) of each space, together with all the NMR structures of 2MGU.

2.5 Potential of Mean Force (PMF)

A two-dimensional histogram was constructed with a bin width and height of 1 or 2 Å, although different values were tested to assess the stability of the results. In the reweighting process each frame k with coordinates (PC1, PC2) is assigned to its corresponding bin ij = (i,j) = ($PC1_i$, $PC2_j$), with a weight $w_k = \exp(\beta \Delta V_k)$, where ΔV_k is the boost potential of frame k. Thus, the reweighted histogram at bin ij was obtained as:

$$H_{ij} = \sum_{k=1}^{K} w_k = \sum_{k=1}^{K} \exp(\beta \Delta V_k) \qquad (6)$$

where the sum k extends over all frames whose coordinates fall in the ij bin. Note that for conventional molecular dynamics $w_k = 1$ and $H_{ij} = K = N_{ij}$. That is, H_{ij} represents the population of bin ij. In this work the exponential was approximated by a 10 order Maclaurin series. Finally, the two –dimensional free energy profile is determined by

$$W_{ij} = -k_B \mathrm{T} \ln(H_{ij}) + W_o \qquad (7)$$

where W_0 is chosen such that the minimum of the free energy profiles is set to zero.

In the sMD case, we recover the canonical distribution of populations p(r) by a simple reweighting $p(r) = p^*(r)^{1/\lambda}$.

3. Results and discussion

3.1 Comparison of the different MD methodologies essayed

Before starting the analysis of CaM flexibility, a detailed study of the extent to which the different methodologies essayed explore the protein conformational space appeared to be necessary. In order to clarify this matter, we performed four cMD, three sMD and three aMD 250 ns runs for the 2MGU system, obtaining a total of 2.5 μs of molecular dynamics. Then, the coordinates of all the structures for each type of MD were projected onto the plane defined by the two principal components with highest eigenvalues (PC1 and PC2) and an approximation to the free energy surface (FES) was obtained.

Fig. 1 displays the FES obtained after combining the different trajectories for each

type of MD, focusing in the N-ter region. As it can be seen, even though all the simulations were able to sample similar regions, sMD and aMD have access to a wider region of the PC space, especially for the largest PC2 values. On the other hand, both cMD and sMD clearly present the most stable structures around (1:-120,0), where 1: stands for the label for that point, and -120,0 for its PC1, PC2 coordinates in the graph. This minimum is also present in the aMD profile but it is less pronounced. This last methodology also presents another minimum around (3:80,180), a region sparsely sampled by the other two methodologies. cMD and sMD also share a minimum around (2:-125,170) that is not present in the aMD FES. Another minimum at (4:200,60) is only present in the sMD FES. An analysis of the results for the Helix region (see Fig. 2) allows to confirm the enhanced sampling of the PC space by both sMD and aMD as compared to cMD. In this case, both cMD and sMD present two principal minima, around (1:30,-20) and around (2:40,60) for the first method, and one around (2:50,50), near the second cMD minimum, and another at (3:-100,-20), for the second method. Nevertheless, the energy profile obtained with aMD is very different, presenting two distinct minima that extend over a wide region centered at (4:-20,80) and (5:80,20), which are not found with the remaining methods. These results suggest that the use of a boosted method allows a deeper analysis of CaM plasticity, difficult to achieve with cMD. On the other hand, although an extensive work should be conducted to understand why (the) aMD method renders (so) different results with respect to the other two methods, the most likely explanation is that the parameters we used to modify the potential energy surface allow the exploration of structures with much higher energy that are not easily accessible with the other two approaches and which could hide the true energy profile. For this reason, we decided to focus only on cMD and sMD for other analyses.

The performance of cMD and sMD methods to sample the conformational space of CaM can be analysed not only in terms of the amount of conformational space explored during a predefined period of time, as described previously, but also by the speed at which a particular conformational sampling is obtained. Thus, we compared the free energy profiles at two different times of both the N-ter and the linker regions of CaM, analysing how early the patterns appeared and focusing on the sampled regions and the minima of the final pattern.

Fig. 3 displays the FES obtained for the 2MGU system after 150 ns of accumulated MD for both the N-ter and Helix regions. As it can be seen by comparing Figs. 1 and 3, the energy profile of the N-ter region of CaM looks very similar after 150 or 250 ns of molecular dynamics in both the cMD and sMD cases. Even though the cMD profile seems to change a little bit more than the sMD one when the MD run length increases moving forward in time, both approaches increase the population of the same region of the PC space. Moreover, the most populated basin at (1:-120,10) is always present. However, the behavior of cMD and sMD for the Helix region as the simulation time increases differs. Hence, four 150 ns cMD trajectories are still sampling the first minimum around coordinates (1:50, -40), with no population on the minimum at (2:40,60), while three 150 ns sMD three trajectories quickly explore the second minimum with coordinates (2:30,50).

Results for the 2MGUpep system confirm these conclusions. Thus, Fig. 4 shows the energy profile of this system for the N-ter region at two simulation times. As it can be seen, cMD presents always a minimum around (1:-100,50) and another minimum at (2:200,110). At the end of the MD runs this last minimum becomes more populated and another less populated minimum appears around (3:200,-110). However, the general profile remains very similar. The FES for the sMD simulations remains practically identical at both times with the three minima described for the cMD similarly populated. In this system, the Helix region behaves similarly to the N-ter region. As it can be seen from Fig. 5, cMD presents two minima at (1:10,0) and (2:-20,-60) PC coordinates that also appear in sMD at both times. Moreover, a third minimum around (3:-70,20), changing slightly along the time for cMD and widely for sMD, can be seen, although the general profile remains very similar.

Finally, Fig. 6 allows to see that the 1CLL system presents the most pronounced differences between cMD and sMD results for both the N-ter and Helix regions. For the N-ter region, cMD finds one minimum around (1:-30,-170) which also appears with sMD. However, while cMD remains sampling around this minimum during the whole run, sMD visits two different minima at (2:210,90) and (3:20,200). The Helix region shows a similar behaviour, with cMD sampling only the (1:0,-60) minimum,

while sMD also explores a second minimum at (2:-40,70).

Considering all the previous results it seems clear that given the flexibility of the system it is difficult to obtain a realistic energy profile without a certain amount of calculations. However, sMD proved to be useful to reduce the time of calculation needed to achieve a defined degree of conformational sampling and to sample a larger conformational space.

3.2 1CLL, 2MGU and 2MGUpep sample different regions

Although long molecular dynamics starting from the 1CLL (*apo* structure) and 2MGU (structure of CaM obtained by removing the peptide of the original 2MGU PDB code structure) structures should lead to the same stable region of the conformational space, their FES present significant differences that seem to be something less important for the sMD calculations, where the covered conformational space is larger (Fig. S1 and S2). For both the N-term and Helix regions, the sampled space is larger for 2MGU than for 1CLL, as expected due to its different starting points.

We will start first with results from cMD simulations. Regarding the effect of removing the peptide from the experimental structure, it can be observed that the 2MGU system samples N-ter positions near the minimum in 1CLL, although without having a clear energy minimum around this conformation (Fig. S1). Moreover, the 2MGUpep system cannot effectively sample this minimum. Although there was a great resemblance in the sampling of the 2MGU and 2MGUpep systems, it could be noted that depleting the peptide bound of the original structure led to the loss of two energy minima. This suggests that two conformations are stabilized due to the binding of the peptide. In fact, it could also be seen that the presence of this peptide limited the amount of conformations the simulation could sample, as demonstrated by the diminished sampled area compared with the 2MGU system.

The Helix linker region was also analysed and significantly different distributions among the three systems could be observed (Fig. S2). The main feature

of the 1CLL structure is that it has a complete α-helix linker region, unlike 2MGU and 2MGUpep, which exhibit partially unstructured linker regions. Thus, 1CLL simulations are far more constrained than those for the other systems due to this feature. 2MGU was the most relaxed system, as it could sample more regions than the others. It is important to note that the linker helix did not get fully structured at any time of the full trajectory. Yet again, 2MGUpep showed different energy minima, which were possibly generated by the binding of the peptide, since it may have stabilized certain conformations.

The sMD data set was studied as well, and it led to similar conclusions. As shown in Fig. S1, the N-ter analysis reports results similar to those from cMD simulations as regards the number of energy minima and their location (2MGUpep had two more minima again). However, there was a significant improvement in the sampling of 1CLL and 2MGU since both systems managed to reach new conformations that were not explored with the cMD approach. This conclusion is even more apparent in 1CLL, as a whole new region was sampled in both the N-ter and the linker region energy profiles. As confirmed by the clusterizations (read below) this enhanced performance was due to the accelerated restructuration of the linker region in the sMD simulations, clearly seen when comparing the results of the analysis of the linker region behaviour between cMD and sMD.

In order to assess the completeness of our sampling, we compared the obtained trajectories with all the structures in the PDB file of 2MGUpep. As shown in Fig. 7, cMD explores all the experimental conformations acquired experimentally for the N-ter region, albeit, 2MGUpep and 2MGU sampled more thoroughly the region corresponding to such structures, while the 1CLL sampling is slightly less exhaustive. Results for the Helix region (and) with sMD lead to the same conclusions. For the N-ter region, we also performed an analysis of the RMSD value along the trajectory by the generation of a histogram obtained by comparison of all the frames of the cMD trajectories to all the experimental structures. Experimental structures have lower values of RMSD compared to 2MGUpep than when compared with the other systems **(**Fig. S3**),** confirming the aforementioned statement. Furthermore, 2MGU retained some similarity, but RMSD values for 1CLL are the largest.

To further analyse the conformational space sampled by the three systems and to gain insight on CaM plasticity, a clusterization was made so as to obtain representative structures of the most important conformations both in cMD and sMD. Clusterizations were carried out analysing the residues of the linker region, as it has been described as a major component regarding the plasticity of CaM. After clusterization was performed we selected the clusters with at least a 10% of population. Afterwards RMSD values were calculated when superposing the representatives of cMD with the representatives of sMD. Even though some cMD cluster representatives are similar to their counterparts in the sMD simulations, the majority of them have a RMSD value above 4 Å. It is to be noted that both data sets (cMD and sMD) shared a similar representative from the 1CLL system that belongs to the structure of CaM with a full α-helix like linker region (Fig. 8a). The RMSD value obtained after superimposing this representative is pretty low, confirming the similarity of these structures (Fig. 8b). Neither 2MGU nor 2MGUpep exhibit this cluster, since both of them fail to achieve a reconstruction (as an α-helix shape) of their linker regions.

In order to compare the clusters of the different systems within a set of simulations (either cMD or sMD) we analysed the RMSD value between their representatives, either for the linker region (superposing only residues 64 to 90) or for the whole protein and selected those couples of representatives with a RMSD value below 6 Å (threshold 1 Å) for the linker region and below 12 Å (threshold 2 Å) for the α carbons of the whole protein (Table S1 and Table S2). Finally, we visually studied these couples.

First, the classical simulations were examined. Even though the couples of 1CLL/2MGU yield more matching structures, two couples of 2MGU/2MGUpep are surprisingly similar, with RMSD values for the whole protein below 7 Å (Table S1). Thus, 2MGU presents similar structures to both 1CLL and 2MGUpep, but the most resembling representatives belong to 2MGU/2MGUpep pairs. On the other hand, 1CLL and 2MGUpep present few couples surpassing the thresholds and with high RMSD values. The most similar representatives of the 2MGU/2MGUpep simulations

are shown in Fig. S4.

Since sMD provides an energy boost, the simulations generated more diverse representatives, as the structures sampled a wider variety of conformations, which resulted in higher RMSD values (Table S2). In spite of the conformational diversity generated, almost half of the 1CLL/2MGU couples still managed to surpass the thresholds. Interestingly, the representatives for both 1CLL/2MGUpep and 2MGU/2MGUpep lost most of their similarity, yielding few couples that passed the thresholds applied. This phenomenon is likely related to the ability of the linker region to explore new conformations for both 1CLL and 2MGU thanks to the energy boost provided by the sMD. Due to this extra push 1CLL explored regions similar to 2MGU, but 2MGU was also driven further away from the structures of 2MGUpep. However, one 2MGU/2MGUpep couple still conserves similarity, thus proving that 2MGU can act as a link between the other two systems, as it can keep its similarity with both.

Our simulations were able to link the three systems depending on every situation, and also to reflect the importance of the linker region. Conventional molecular dynamics proved useful to relate simulations with similar starting points (2MGU/2MGUpep) as sMD helped with its boost to link more dissimilar systems as 1CLL/2MGU. All in all, our data seems to indicate that there is a pool of conformations of CaM that link them. This fact could help this exceptionally flexible protein to adopt several structures which could, in turn, interact with diverse proteins.

3.3 Central residues of liner region are important for the plasticity of CaM

Since the linker region had already proven its relevance in CaM dynamism, its role was further studied by analysing the dihedral angles of this region. The focus was set on the central residues, as they are prone to bend in order to provide flexibility to the lobes. Residues from 70 to 91 were examined by obtaining the corresponding Ramachandran plots presenting the 3 systems with both essayed methodologies. The dihedral angles of the NMR structures of the PDB protein with

code 2MGU were also added.

Firstly, cMD simulations were analysed. As shown in Fig. S5, the initial and last residues of the helix presented no significant differences compared with the experimental values (residues 70-71/84-91) or between them. Thus, the sampled regions of the three systems are superimposed. However, the central residues provided very different results. As can be seen in Fig. 9, there is a great variability depending on the system studied. Both 1CLL and 2MGU presented similar profiles, even though 2MGU had more freedom to explore different conformations and, as a result, there were certain angles 2MGU could reach that 1CLL could not (Fig. 9). A clear example are the Φ/Ψ angles of residues 73, 74, 75, 78, 79 and 83, where 2MGU managed to sample several angles different to those of 1CLL. The simulations of 2MGUpep presented quite different profiles. For instance, there were some restrictions on certain residues when compared to 2MGU or 1CLL (residues 73, 74, 75, 76 or 83), but other residues were able to sample angles the other systems were not. For example, the region explored by residues 78 to 80 in the 2MGUpep simulations was greatly broadened. Interestingly, this wider exploration in the cMD simulations resulted in the sampling of dihedral angles that are naturally present in CaM structure when bound to the HIV-1 matrix protein, as shown by the NMR structures, also depicted in the Ramachandran plots of Fig. 9. The clearest example is residue 78, whose dihedral angles for the NMR structures spread all over the plot and only the 2MGUpep simulations were able to track them down. Altogether, these results support the idea that the performed cMD simulations are in concordance with the experimental results.

As commented previously, the energy boost in the sMD simulations leads to a wider sampling in all three systems. This broader exploration resulted in the reduction of the differences between 1CLL and 2MGU, as 1CLL sampled regions it was not able to reach with cMD thanks to the additional energy. The simulations of 1CLL and 2MGU presented really similar plots for residues 73 to 81 (Fig. 10), which supports the idea that sMD leads to more resemblance between 1CLL and 2MGU (as shown previously with the clusterizations).

Interestingly, the expansion of the dihedral angles visited was remarkably selective in the 2MGUpep system, since not all residues presented the same

response to the energy boost. In the case of residues 78 to 80 it led, as expected, to a broadening of the dihedral angles reached. Strikingly, the additional energy had no effect on residues 72, 73, 74, 77, 82 and 83 of the 2MGUpep simulations, as the additional movement supplied by the sMD did not enlarge the sampled area, even though it expanded the angles sampled by 2MGU and 1CLL. This result highlights the fact that, even though sMD enhances the sampling, it respects the present restrictions of the system (in the 2MGUpep system, the presence of a restrictive peptide).

All in all, the central amino acids of the linker region presented a wide variability in the angles in which they bended, especially 73 to 81. Similarly to what is described by Kukic *et al.* [15] our data suggest a model in which these central residues provide a major part of the malleability of CaM. However, we concluded that not only residues 76, 77, 79, 80 and 81 are highly flexible, but also 73, 74, 75 and 78. This might be due to the features of our simulations, which are unrestricted and can sample throughout more time (a total of 1 μs for cMD and 750 ns for sMD). The high motility of these residues may explain why the linker region provides such a huge pool of conformations to CaM.

4. Conclusions

Our data provide evidence that sMD are a reliable source to study flexible proteins such as calmodulin. Our results point that this methodology can obtain results similar to those from cMD, or even outperform them by requiring less time of calculation. Thus, sMD simulations have allowed to reach certain conformations cMD could not explore by providing a boost of energy that allowed a further destructuration of the linker region. Furthermore, as the dihedral angles study shown, sMD sticks to the intrinsic restrictions the peptide imposed on CaM. Thus, these simulations allowed a further sampling of conformations without compromising the stability of the protein or the reliability of the results obtained. However, aMD was not useful for the systems herein studied with the parameters used.

The diverse simulations performed revealed how the three systems under study

can interconvert. 2MGU and 2MGUpep are obviously strongly related for sharing the original conformation, but at the same time have great differences in their energy profiles due to the presence of the peptide in 2MGUpep. 1CLL can also be linked to 2MGU as, hypothetically, they will eventually reach the same point although they initially sample different areas. The use of sMD, as seen through the backbone dihedral angles study allow both systems to gain resemblance and explore similar conformations. The relationship between these three structures has also been shown through a clusterization analysis, which reflected the links between 2MGU, 2MGUpep and 1CLL by revealing that they share certain structures with a high percentage of relevance throughout their trajectories. This conclusion is actually logical because these systems should be prone to become the same, as they are exactly the same protein in the cell.

Our data supports the idea that CaM has a vast pool of conformations; two of them being the described by 2MGU and 1CLL, and certain structures can be further modified by the interaction with the diverse targets of CaM (Fig.11). This model would explain why CaM can interact with such a large group of proteins taking advantage of its plasticity, which seems to be obtained through residues in the linker region. These residues have some preferential backbone dihedral angles when not bound to a target (1CLL/2MGU), but they can change in order to provide a better interaction with other proteins, providing this way a mechanistical explanation for the proposed model.

Future perspective

Calmodulin can bind to more than 30 proteins because it can modify itself to interact with this large group of proteins. Thus, the possibility of generating new inhibitors using information extracted from the CaM/ CaMBPs complexes is a very interesting and novel approach. The use of potent graphic processing units together with new sampling methodologies such as scaled molecular dynamics will permit the selective design of compounds disrupting the protein-protein surface.

Summary points

- The conventional molecular dynamics method does not allow an adequate sampling of conformational space
- Accelerated molecular dynamics method requires optimization of the setup parameters to be useful.

- The scaled molecular dynamics approach is judged the best.
- The linker region of calmodulin can adopt multiple conformations and may be responsible for promiscuous recognition of calmodulin binding proteins

References

1. Marshall CB, Nishikawa T, Osawa M, Stathopulos PB, Ikura M. Calmodulin and STIM proteins: Two major calcium sensors in the cytoplasm and endoplasmic reticulum. *Biochem. Biophys. Res. Comm.*, 460(1), 5-21 (2015).
2. Jurado LA, Chockalingam PS, Jarrett HW. Apocalmodulin. *Physiol. Rev.*, 79(3), 661-682 (1999).
   ** A good starting point to understant Calmodulin plasticity.
3. Piazza M, Guillemette JG, Dieckmann T. Dynamics of nitric oxide synthase-calmodulin interactions at physiological calcium concentrations. *Biochem.*, 54(11), 1989-2000 (2015).
4. Alvarez-Moya B, López-Alcalá C, Drosten M, Bachs O, Agell N. K-Ras4B phosphorylation at Ser181 is inhibited by calmodulin and modulates K-Ras activity and function. *Oncogene*, 29, 5911 (2010).
5. Lopez-Alcala C, Alvarez-Moya B, Villalonga P, Calvo M, Bachs O, Agell N. Identification of essential interacting elements in K-Ras/calmodulin binding and its role in K-Ras localization. *J Biol Chem*, 283(16), 10621-10631 (2008).
6. Villarroel A, Taglialatela M, Bernardo-Seisdedos G *et al.* The ever changing moods of calmodulin: how structural plasticity entails transductional adaptability. *J. Mol. Biol.*, 426(15), 2717-2735 (2014).
7. Persechini A, Kretsinger RH. The Central Helix of Calmodulin Functions as a Flexible Tether. *J Biol Chem*, 263(25), 12175-12178 (1988).
   ** Very useful to understant Calmodulin plasticity
8. Vlach J, Samal AB, Saad JS. Solution structure of calmodulin bound to the binding domain of the HIV-1 matrix protein. *J Biol Chem*, 289(12), 8697-8705 (2014).
9. Nussinov R, Muratcioglu S, Tsai CJ, Jang H, Gursoy A, Keskin O. The Key Role of Calmodulin in KRAS-Driven Adenocarcinomas. *Mol. Cancer Res.*, 13(9), 1265-1273 (2015).
10. Garrido E, Lazaro J, Jaumot M, Agell N, Rubio-Martinez J. Modeling and subtleties of K-Ras and Calmodulin interaction. *PLoS Comput. Biol.*, 14(10), e1006552 (2018).
    ** The first structure of Calmodulin-KRas complex
11. Negi S. Effect of Calcium Ion Removal, Ionic Strength, and Temperature on the Conformation Change in Calmodulin Protein at Physiological pH. *J. Biophys.*, 2014, 329703 (2014).
12. Meireles L, Gur M, Bakan A, Bahar I. Pre-existing soft modes of motion uniquely defined by native contact topology facilitate ligand binding to proteins. *Protein Sci.*, 20(10), 1645-1658 (2011).
13. Bakan A, Bahar I. The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. *P Natl Acad Sci USA*, 106(34), 14349-14354 (2009).
    ** Very good description of the induce fit effect

14. Tamura K, Hayashi S. Linear Response Path Following: A Molecular Dynamics Method To Simulate Global Conformational Changes of Protein upon Ligand Binding. *J. Chem. Theory Comput.*, 11(7), 2900-2917 (2015).
15. Kukic P, Camilloni C, Cavalli A, Vendruscolo M. Determination of the individual roles of the linker residues in the interdomain motions of calmodulin using NMR chemical shifts. *J. Mol. Biol.*, 426(8), 1826-1838 (2014).
** An experimental approximation to the Calmodulin dynamics
16. Sinko W, Miao Y, de Oliveira CA, McCammon JA. Population based reweighting of scaled molecular dynamics. *J. Phys. Chem. B*, 117(42), 12759-12768 (2013).
17. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.*, 120(24), 11919-11929 (2004).
18. Miao Y, McCammon JA. Unconstrained Enhanced Sampling for Free Energy Calculations of Biomolecules: A Review. *Mol. Simul.*, 42(13), 1046-1055 (2016).
19. Gomez-Gutierrez P, Rubio-Martinez J, Perez JJ. Identification of Potential Small Molecule Binding Pockets in p38alpha MAP Kinase. *J. Chem. Inf. Model.*, 57(10), 2566-2574 (2017).
20. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, 65(3), 712-725 (2006).
21. Case DA, Babin V, Berryman JT *et al. AMBER 14* (University of California, San Francisco, University of California, San Francisco, 2014).
22. Roe DR, Cheatham TE, 3rd. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.*, 9(7), 3084-3095 (2013).
23. *Molecular Operating Environment (MOE) 2009.1. Chemical Computing Group Inc., 1010 Sherboke St. West, Suite #90, Montreal, QC, Canada, H3A 2R7, 2013*.
24. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.*, 79(2), 926-935 (1983).
25. Adelman SA, Garrison BJ. Generalized Langevin Theory for Gas-Solid Processes - Dynamical Solid Models. *J. Chem. Phys.*, 65(9), 3751-3761 (1976).
26. Davidchack RL, Handel R, Tretyakov MV. Langevin thermostat for rigid body dynamics. *J. Chem. Phys.*, 130(23), 234101 (2009).
27. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.*, 23(3), 327-341 (1977).
28. Perez JJ, Tomas MS, Rubio-Martinez J. Assessment of the Sampling Performance of Multiple-Copy Dynamics versus a Unique Trajectory. *J. Chem. Inf. Model.*, 56(10), 1950-1962 (2016).

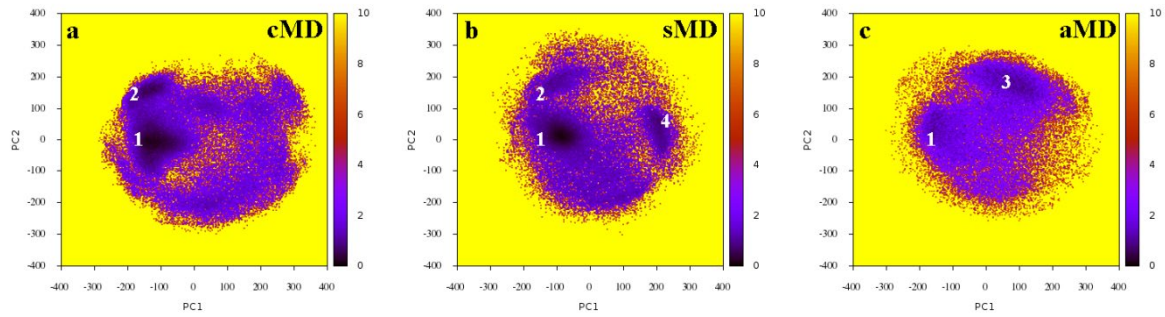


Fig. 1. FES of the N-ter region of 2MGU simulations obtained after combining the different trajectories for each type of MD. a) Conventional MD (cMD). b) Scaled MD (sMD). c) Accelerated MD (aMD).

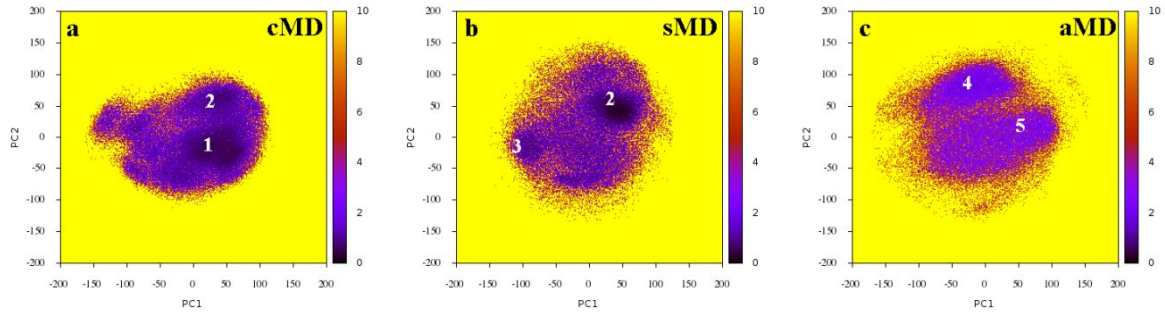


Fig. 2. FES of the Helix region of 2MGU simulations obtained after combining the different trajectories for each type of MD. a) Conventional MD (cMD). b) Scaled MD (sMD). c) Accelerated MD (aMD).

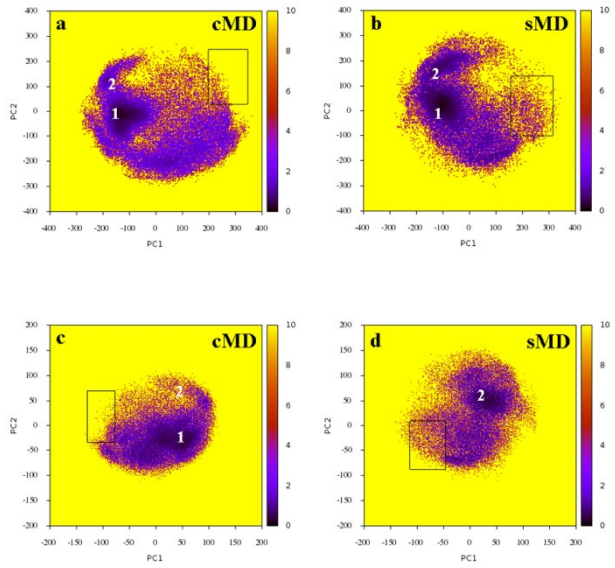


Fig. 3. FES obtained for the 2MGU system after 150 ns of accumulated MD for both the N-ter and Helix regions. a) N-Ter region with cMD b) N-Ter region with sMD. c) Helix region with cMD d) Helix region with sMD.

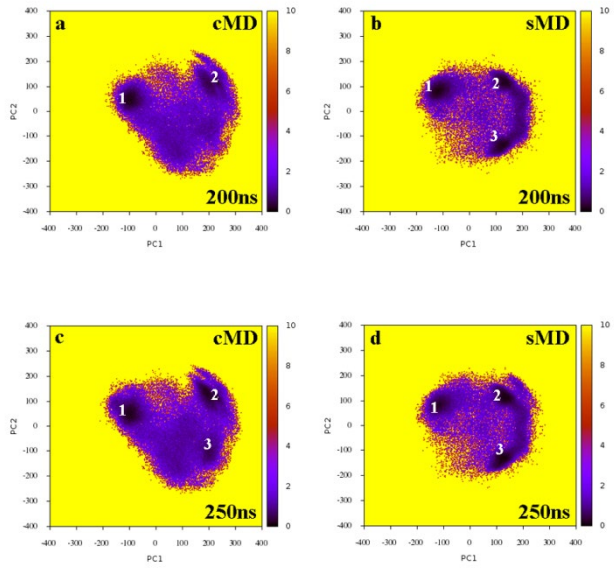


Fig. 4. FES obtained for the 2MGUpep system for the N-ter region at two different times of the accumulated molecular simulations. a) 200 ns of cMD, b) 200 ns of sMD c) 250ns of cMD, d) 250 ns of sMD.

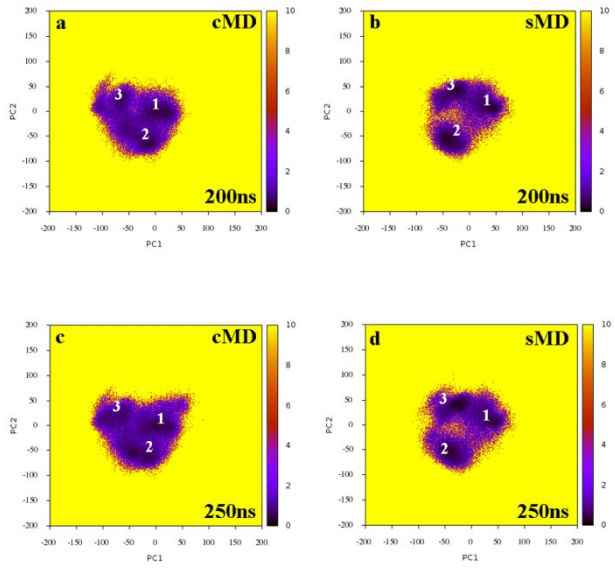


Fig. 5. FES obtained for the 2MGUpep system for the Helix region at two different times of the accumulated molecular simulations. a) 200ns of cMD, b) 200 ns of sMD c) 250 ns of cMD d) 250 ns of sMD.

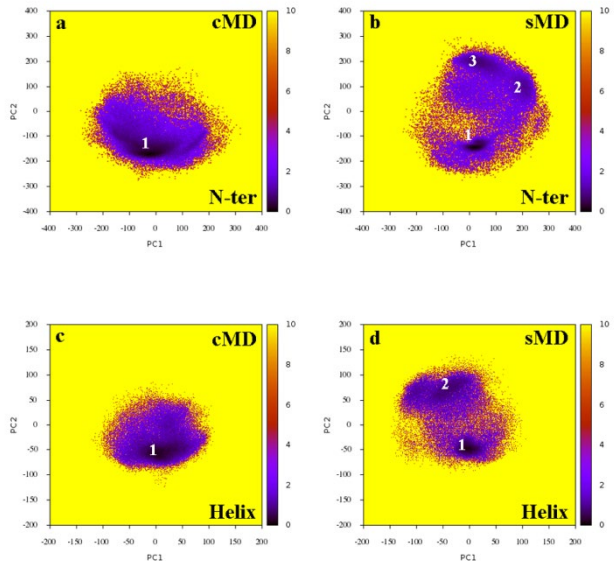


Fig. 6. FES obtained for the 1CLL system after 250 ns of accumulated MD for both the N-ter and Helix regions. a) N-Ter region with cMD, b) N-Ter region with sMD, c) Helix region with cMD, d) Helix region with sMD.

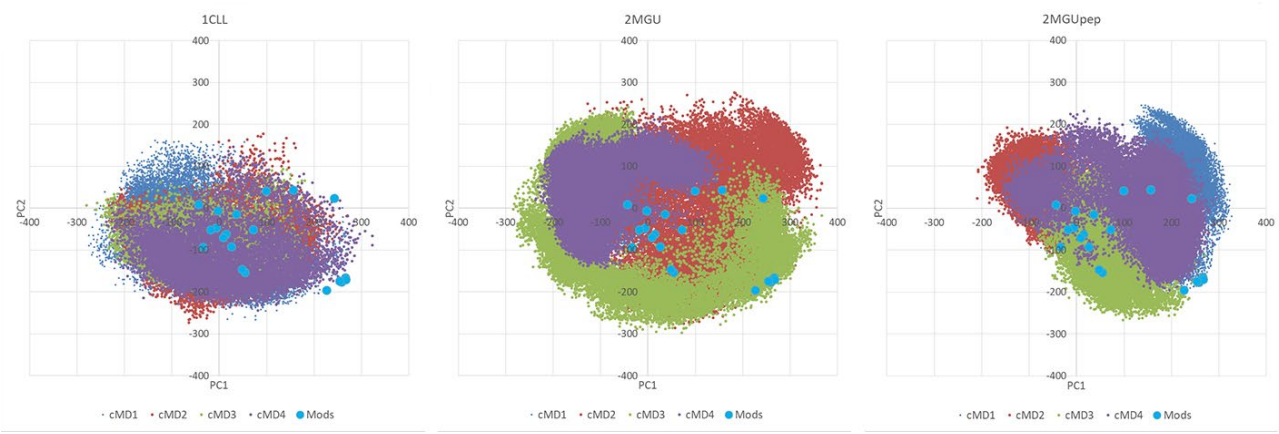


Fig. 7. Projection upon the principal component space of the experimental structures of CaM from PDB code 2MGU and the cMD trajectories of a) 2MGU, b) 2MGUpep and c) 1CLL. Each colour represents one of the four cMD simulations, while the bold blue dots represent the experimental structures.

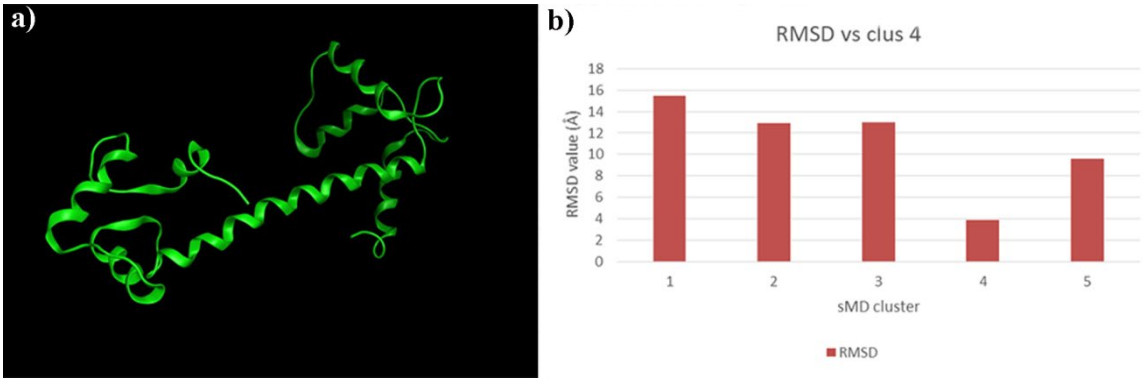


Fig. 8. Exclusive representative for the 1CLL simulations. a) Common representative between cMD and sMD of 1CLL after the clustarization. b) RMSD values when superimposing the extended linker region of the cluster of cMD with the cluster representatives of the sMD simulations. Number 4 is the cluster of sMD in common with cMD.

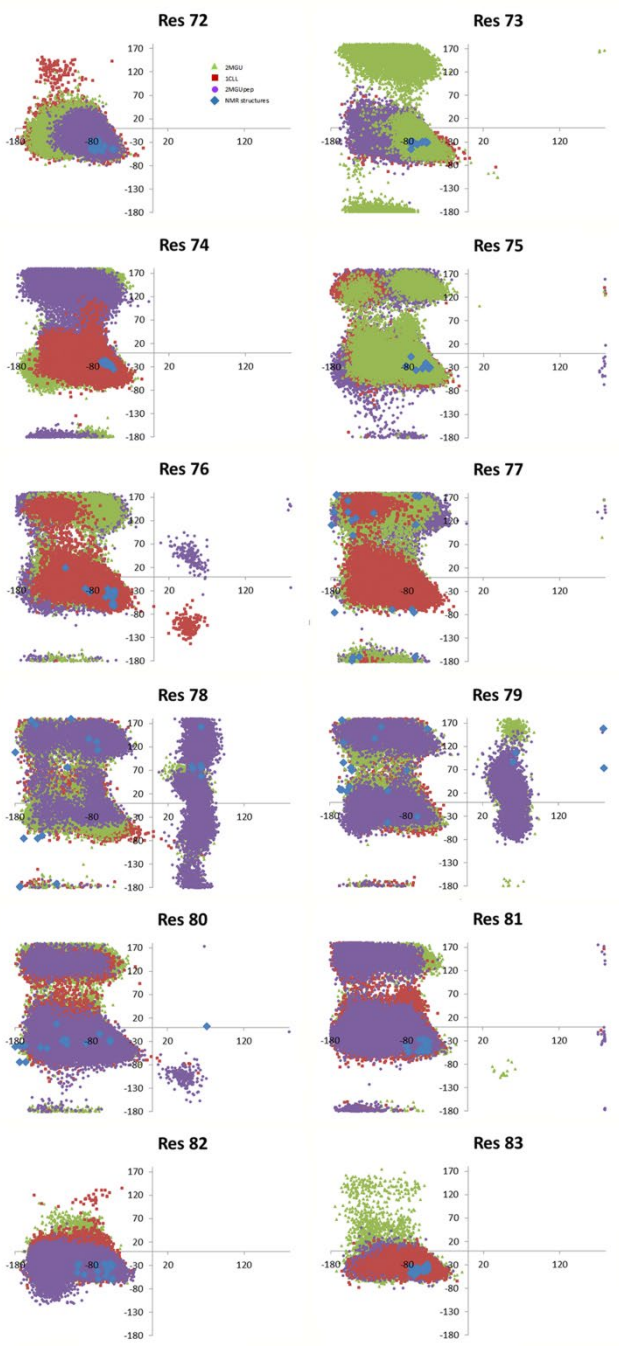


Fig. 9. Backbone dihedral angles for the residues of the linker region using cMD. The horizontal axis represents the Φ angle, while the vertical axis reflects the Ψ angle. 2MGU simulations in green, 2MGUpep in purple, 1CLL in red and experimental values from RMN structures with PDB code 2MGU in bold blue.

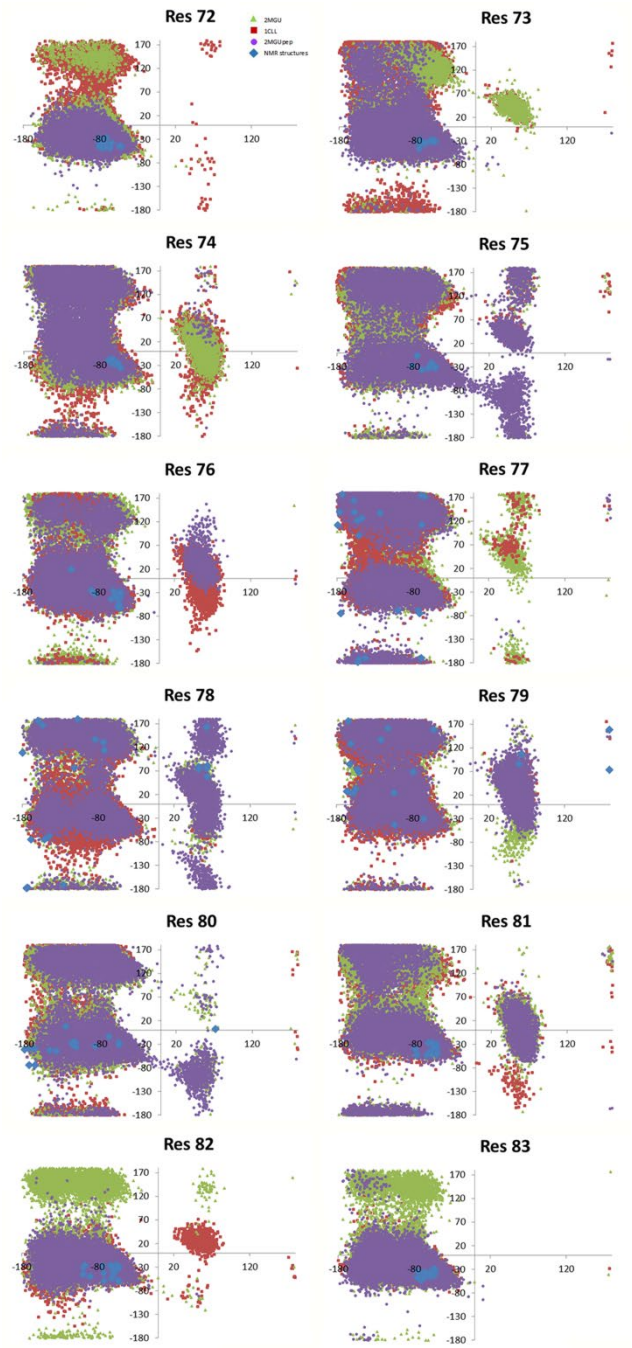


Fig. 10. Backbone dihedral angles for the residues of the linker region using sMD. The horizontal axis represents the Φ angle, while the vertical axis reflects the Ψ angle. 2MGU simulations in green, 2MGUpep in purple, 1CLL in red and experimental values from RMN structures with PDB code 2MGU in bold blue.

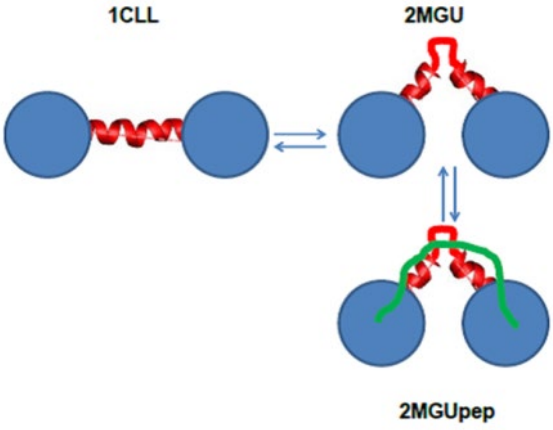
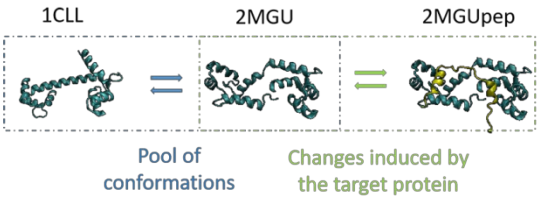


Fig. 11. Model suggested for the pool of conformations of CaM. The structure of unbound CaM can vary through the movements of its linker region, and can be further modified by the interaction with a CaMBP. Terminal lobes presented as blue circles; linker region in red and a peptide in green.

TOC



CaM a very flexible protein has been studied taking advantage of the new available methodologies such as scaled and accelerated molecular dynamics. The results will be important to determine how this protein behaves when binding to its partners, as many new roles have been recently found for this protein.