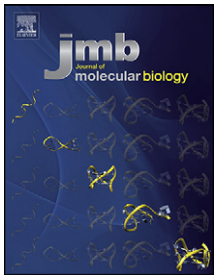

Determination of the Individual Roles of the Linker Residues in the Interdomain Motions of Calmodulin Using NMR Chemical Shifts

Predrag Kukic, Carlo Camilloni, Andrea Cavalli and Michele Vendruscolo

Department of Chemistry, *University of Cambridge, Cambridge CB2 1EW, UK*

Correspondence to Michele Vendruscolo: *mv245@cam.ac.ukz*
http://dx.doi.org/10.1016/j.jmb.2014.02.002
Edited by A. G. Palmer

Abstract

Many protein molecules are formed by two or more domains whose structures and dynamics are closely related to their biological functions. It is thus important to develop methods to determine the structural properties of these multidomain proteins. Here, we characterize the interdomain motions in the calcium-bound state of calmodulin (Ca^{2+}-CaM) using NMR chemical shifts as replica-averaged structural restraints in molecular dynamics simulations. We find that the conformational fluctuations of the interdomain linker, which are largely responsible for the overall interdomain motions of CaM, can be well described by exploiting the information provided by chemical shifts. We thus identify 10 residues in the interdomain linker region that change their conformations upon substrate binding. Five of these residues (Met76, Lys77, Thr79, Asp80 and Ser81) are highly flexible and cover the range of conformations observed in the substrate-bound state, while the remaining five (Arg74, Lys75, Asp78, Glu82 and Glu83) are much more rigid and do not populate conformations typical of the substrate-bound form. The ensemble of conformations representing the Ca^{2+}-CaM state obtained in this study is in good agreement with residual dipolar coupling, paramagnetic resonance enhancement, small-angle X-ray scattering and fluorescence resonance energy transfer measurements, which were not used as restraints in the calculations. These results provide initial evidence that chemical shifts can be used to characterize the conformational fluctuations of multidomain proteins.

Introduction

It has been estimated that about two-thirds of prokaryote proteins and 80% of eukaryote proteins contain two or more domains [1]. The functions of these multidomain proteins are closely associated with their domain composition and interdomain rearrangements [2,3], and multidomain proteins belonging to the same family tend to undergo similar functionally relevant conformational changes [4]. Since the lengths and sequences of the flexible linkers that connect these domains have allosteric and functional roles a [2–15], an accurate characterization of the interdomain rearrangements is key to understand the molecular mechanisms of binding. While X-ray crystallography provides high-resolution information about the structures of the individual domains and about the interdomain arrangements in specific states, a range of other techniques, including cryo-electron microscopy, small-angle X-ray scattering (SAXS), fluorescence resonance energy transfer (FRET) and NMR spectroscopy, can be used to obtain further information about the conformations and the dynamics of multidomain proteins [16,17].

In the present work, we explore the possibility of using NMR chemical shifts, the most accurately and readily measured NMR observables, to study interdomain motions in proteins. This idea follows from the recognition that chemical shifts can be used to reproduce conformational fluctuations of proteins at a resolution that can be comparable to that provided by more standard NMR methods [18–24]. In particular, chemical shifts have been employed to characterize the interdomain motions of ribonuclease A [19,20] by using an approach in which the time- and ensemble-averaged nature of the chemical shifts was accounted for through the use of structural restraints in replica-averaged molecular dynamics (MD) simulations [25,26].

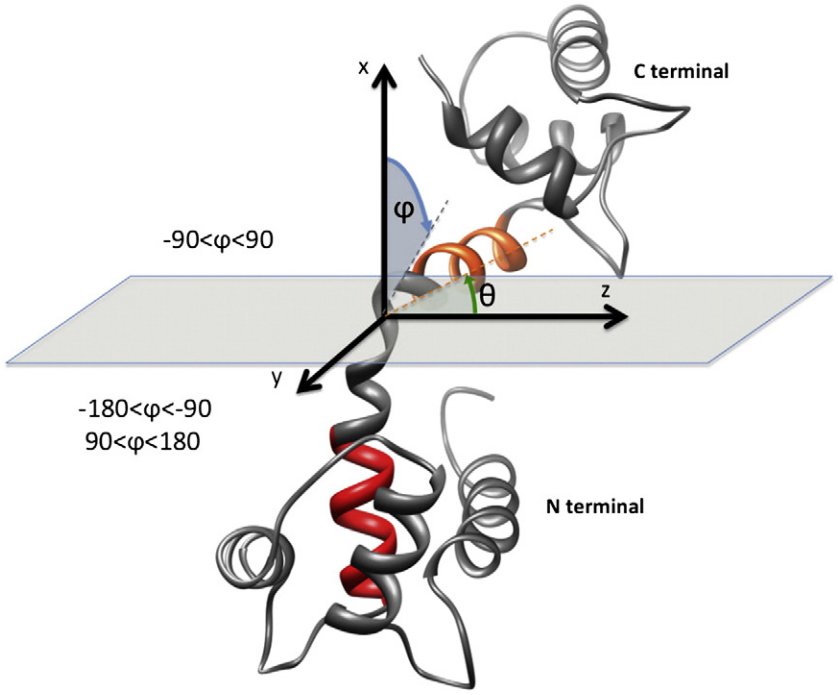

Fig. 1. Illustration of the coordinate system adopted in this work. The position of α-helix V (orange) relative to α-helix IV (red) is given in terms of spherical coordinates (ϕ and θ angles).

This type of approach is adopted in the present study with the aim of probing interdomain motions of calmodulin (CALcium-MODULated proteIN, or CaM) in its calcium-bound state (Ca^{2+}-CaM). Ca^{2+}-CaM is chosen as a model system since it exhibits interdomain motions of relatively large amplitude [27–37] and thus represents a challenging test for any method aimed at characterizing large-scale interdomain rearrangements in proteins. On the basis of residual dipolar coupling (RDC) and paramagnetic resonance enhancement (PRE) measurements, previous results have indicated that Ca^{2+}-CaM explores a very wide region of the interdomain orientations and that the interdomain linker region is important in determining such aspect [29–31,33]. Here we exploit the ability of chemical shifts to provide detailed information about the conformational fluctuations of proteins at the amino acid level to identify the specific roles of individual residues in the linker in determining the interdomain motions. We thus found 10 residues in the interdomain linker region that change their conformations upon substrate binding. Five of these residues (Met76, Lys77, Thr79, Asp80 and Ser81) are highly flexible and cover the range of conformations observed in the substrate-bound state, while the remaining five (Arg74, Lys75, Asp78, Glu82 and Glu83) are much more rigid and do not populate conformations typical of the substrate-bound form.

Results and Discussion

Generation and characterization of the Ca^{2+}-CaM conformational ensemble

We used MD simulations with replica-averaged chemical shift restraints [19,20] to generate two ensembles (CS ensembles) of conformations representing the conformational fluctuations of, respectively, *Drosophila melanogaster* and human Ca^{2+}-CaM, using chemical shifts available from the literature [34,35]. This ensemble was superimposed on the axis of α-helix IV (residues 65–74) in the N-terminal domain (NTD). In this superimposed ensemble, the probability distribution of α-helix V (residues 83–91) in the C-terminal domain (CTD) describes the relative orientation of the CTD in respect to the NTD (see Fig. 1 for the definition of the coordinate system). The relative orientations of the CTDs of *D. melanogaster* Ca^{2+}-CaM are shown in Fig. 2. The ϕ angle samples values ranging from −80° to 70°, whereas the θ angle covers values ranging from 20° to 160° (Fig. 2). Almost identical results are obtained for human Ca^{2+}-CaM (Fig. S1). In principle, interdomain orientations should be defined by three rotational degrees of freedom, for example, in terms of the three Euler angles. Here, we have adopted the angles described in Fig. 1 in order to enable a comparison between the present results and the

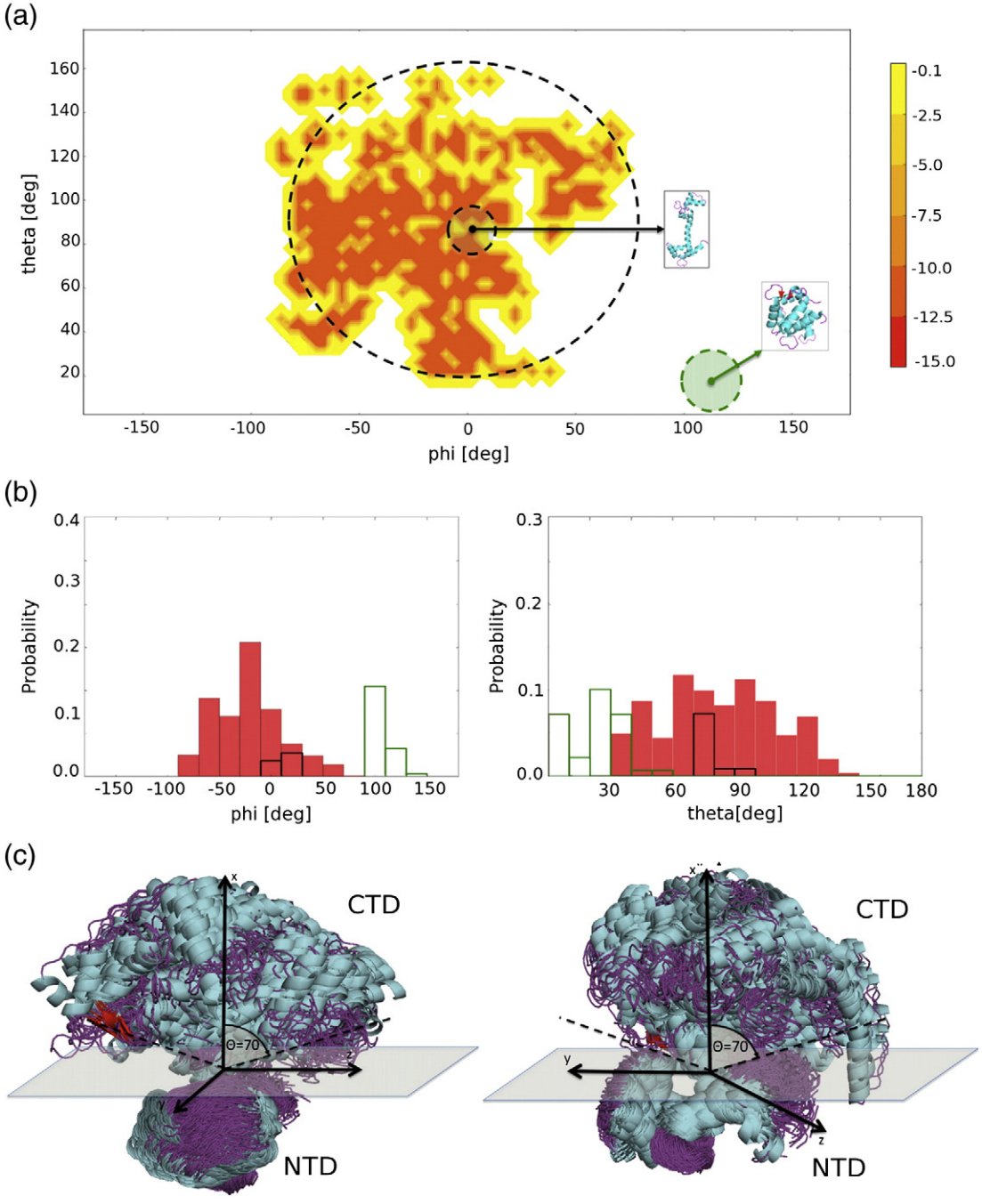

Fig. 2. Analysis of interdomain motions of *D. melanogaster* Ca^{2+}-CaM described using chemical shift restraints (the CS ensemble). (a) Free-energy landscape with respect to the ϕ and θ angles that define the position of α-helix V relative to α-helix IV (see Fig. 1); the free energy as a function of the ϕ and θ angles was obtained as $-k_B T \ln H(\phi, \theta)$, where $H(\phi, \theta)$ is the number of times a conformation with specific ϕ and θ angles was sampled during the simulations, and the color bar values are given in kilojoules per mole. The positions of the extended conformations of Ca^{2+}-CaM present in the PDB are depicted by the small black circle, whereas those of the peptide-bound conformations are depicted by the green circle. (b) Comparison of the distributions of the ϕ and θ angles in the CS ensemble (red bars) and in the X-ray structures of CaM deposited in the PDB (Ca^{2+}-CaM structures are shown in black; Ca^{2+}-CaM structures bound to substrates are shown in green); the distributions corresponding to the X-ray structures are downscaled by a factor 10 for clarity. (c) Representation of the CTD structures in the CS ensemble superimposed onto the NTD.

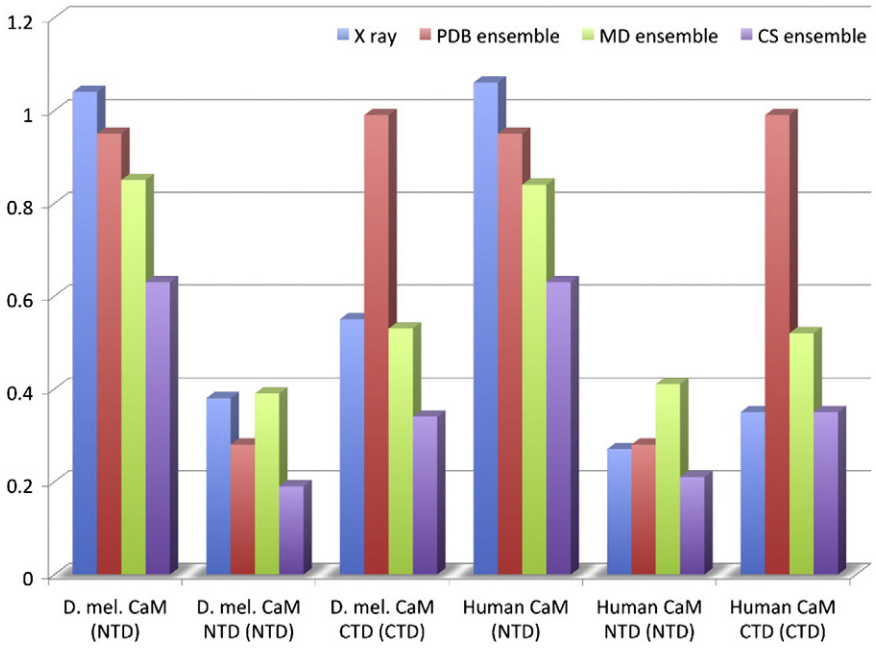

Fig. 3. Validation of the CS ensemble using RDC measurements in terms of *Q*-factors. Purple, CS ensemble; blue, X-ray structures (PDB IDs: 1CLL and 4CLN) of human and *D. melanogaster* Ca^{2+}-CaM; red, the ensemble of Ca^{2+}-CaM structures from the PDB (PDB ensemble; see Methods); green, the control ensemble obtained using the same protocol as the CS ensemble but without chemical shift restraints (MD ensemble; see Methods). In the *Q*-factor calculations, the entire protein, NTD and CTD are depicted separately. Domains used for the calculation of the alignment tensors are given in parentheses; thus, for example, "D. mel. CaM (NTD)" indicates the results for the full-length *D. melanogaster* Ca^{2+}-CaM structure in which the alignment was calculated from the NTD RDCs.

corresponding ones obtained for Ca^{2+}-CaM using RDC and PRE measurements [31,33].

The interdomain motions sampled across the two variants of Ca^{2+}-CaM analyzed in this work are similar and can be described by an elliptical cone with a 70° semiangle. This cone includes the extended conformations of Ca^{2+}-CaM present in the PDB (small black circle, $\phi = 0°$ and $\theta = 90°$ in Fig. 2). By contrast, the fully closed conformations of Ca^{2+}-CaM bound to its substrates (green circle, $\phi = 110°$ and $\theta = 10°$) are not present in our ensemble. These results are in agreement with the conclusions obtained from previous experimental studies that employed RDC, pseudocontact shift and SAXS measurements and the maximum occurrence computational approach [29,30] or PRE-derived distance restraints [31,37]. Those ensembles, as the ones discussed here, sample elongated forms of the protein that correspond to a cone of similar properties; the semiangle of the cone reported in the previous study was approximated to be between 50° and 80°. All the ensembles are slightly leaned toward the second metal binding site in the NTD ($\theta < 90°$) [29]. At least some of the differences in the interdomain motions across the variants can be attributed to the slightly different conditions that were used in the experiments (see Fig. S2 for differences in secondary structure populations inferred from the two sets of measured chemical shifts).

Validation of the conformational ensembles using PRE, RDC and SAXS measurements

Relatively compact conformations resembling those populated by Ca^{2+}-CaM upon substrate binding have been recently reported at low populations in studies in which PRE measurements were used to generate ensembles of structures [31,33]. PRE-based approaches are particularly suitable in the case of CaM since they enable the detection of conformations with low statistical weights. Fully compact conformations characteristic of the Ca^{2+}-CaM in complex with its substrates, however, were not detected even in these PRE-based studies (Fig. S4) [31,33]. The presence of relatively compact conformations was also observed using MD simulations with S^2 order parameters restraints [32]. In addition to the results shown in Fig. S4, in order to further test whether such relatively compact conformations are also present in the ensemble that we generated, we back-calculated the PRE intensities (Fig. S3), finding a good agreement between the regions reported to be in contact [31]. A control simulation carried out using the same protocol but without the chemical shift restraints (MD ensemble; see Methods) is not in equally good agreement with the PRE measurements (Fig. S3). For example, the contacts between the spin label at position 128 and the regions of residues 15–20 and

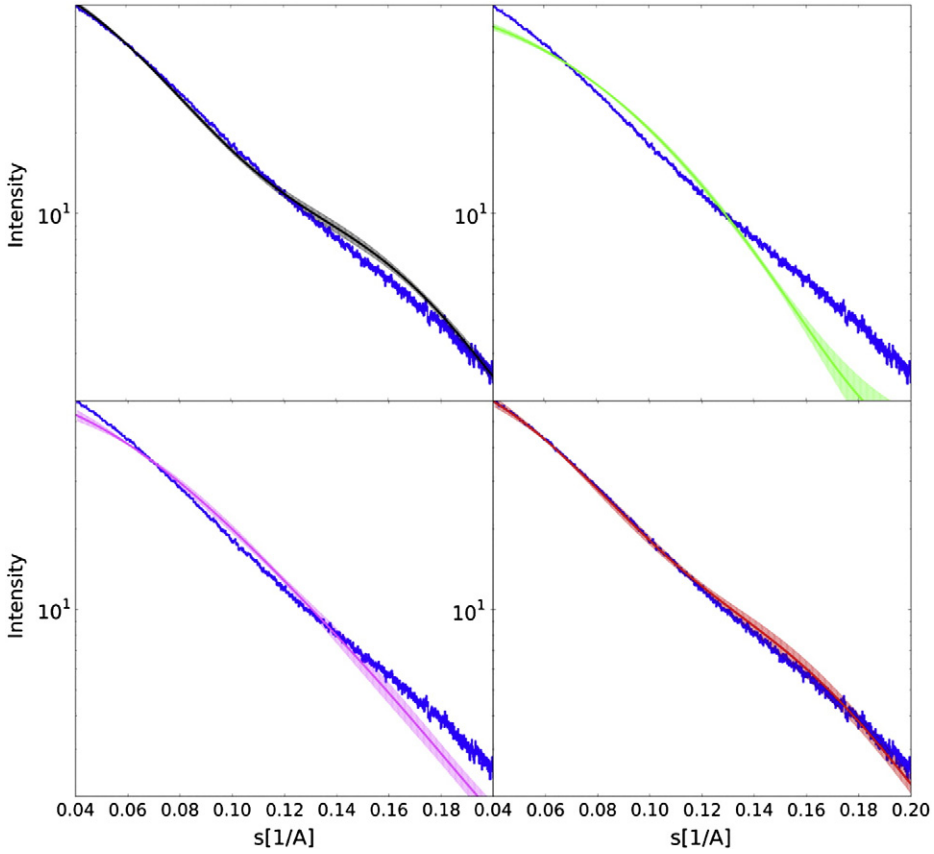

Fig. 4. Validation of the CS ensemble using SAXS measurements. Comparison between experimental [30] (blue line) and calculated SAXS profiles with (a) elongated X-ray structures of Ca^{2+}-CaM (black line) in the PDB, (b) a compact X-ray structure of Ca^{2+}-CaM (PDB ID: 1PRW; green line), (c) the Ca^{2+}-CaM structures in the MD ensemble (a control ensemble obtained using the same protocol as the CS ensemble but without chemical shift restraints) (pink line) and (d) the Ca^{2+}-CaM structures in the CS ensemble (red line).

75–80 are seen in the CS ensemble, but not in the MD ensemble (Fig. S3b); the heights of the PRE profiles of these contacts in the CS ensemble are comparable with those obtained when the PRE are not used as validation but imposed as restraints in the simulations (see Fig. 4d in Ref. [31]). Furthermore, the distribution of the ϕ and θ angles from the PRE-based ensemble was compared to the corresponding distribution obtained from the CS ensemble (Fig. S4a), finding similarly compact conformations in the two ensembles. In addition, the compact conformations from the two ensembles showed similar distances to the peptide-bound form (Fig. S4b). These results indicate that, although the errors on the estimates of the populations of the different states increase for decreasing populations (i.e., the errors in the estimates of the corresponding free energies increase with the free energy itself), the populations that we obtained in this study are relatively accurate at least down to 5–10% populations, consistently with previous estimates [23,24].

The structural ensembles determined in this study (CS ensembles) were validated against experimentally measured N-HN RDCs of human Ca^{2+}-CaM [36], which were not used as restraints in the calculations (Fig. 3). When the alignment tensors are calculated from the NTD and then used in the prediction of the RDCs of the CTD, the *Q*-factor is 0.63, which is lower than the value of 1.06 obtained from a single representative X-ray and the value of 0.95 obtained from the ensemble of existing PDB structures (PDB ensemble; see Methods). The improvement is also significant when only individual domains are evaluated against the experimental RDCs (Fig. 3). The control MD ensemble is not in equally good agreement with the RDCs (Fig. 3). Although the *Q*-factor for the CS ensemble is higher than the typical values (around 0.2) obtained when the RDCs are used as structural restraints, it is comparable with those of structures and ensembles determined by other NMR measurements such as in particular nuclear Overhauser enhancements.

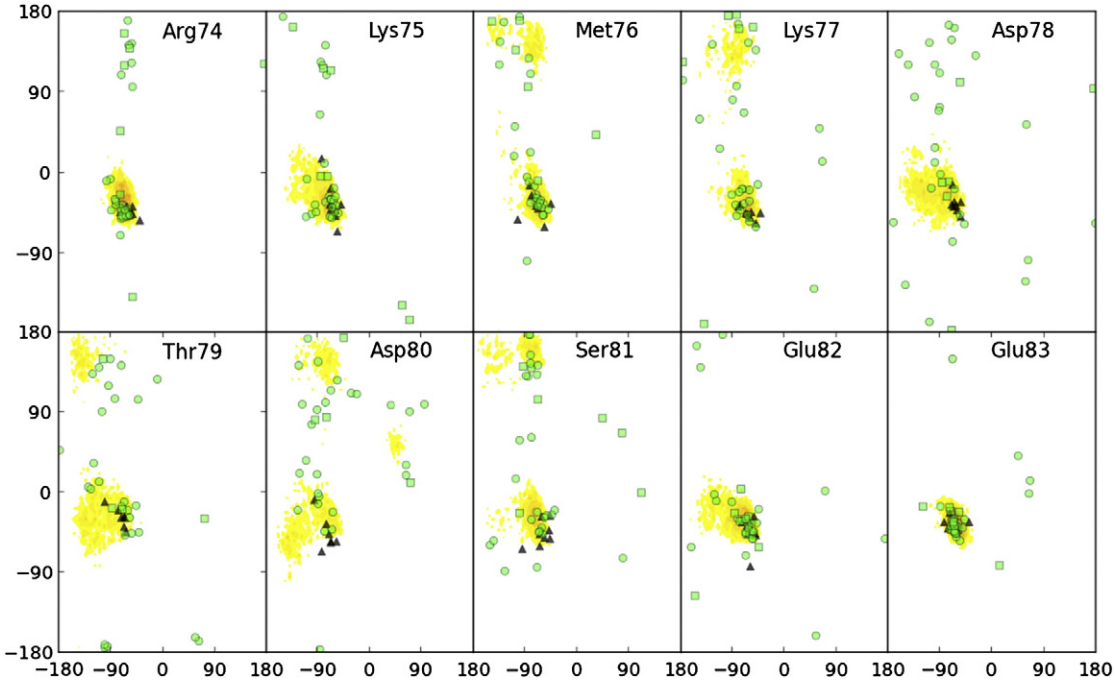

Fig. 5. Ramachandran maps of the residues in the interdomain linker of *D. melanogaster* Ca^{2+}-CaM. Yellow-to-orange regions represent the histogram of the values in the conformational ensemble determined in this work (the CS ensemble; see Methods), black triangles represent the Ca^{2+}-CaM conformations in the PDB (the PDB ensemble; see Methods) and green circles represent substrate-bound conformations of Ca^{2+}-CaM in the PDB. Only the 10 residues (Arg74 to Glu83) in the linker region that change conformation between the substrate-free and substrate-bound structures in the PDB are depicted here; all the other residues are shown in Figs. S6–S9. Five residues (Met76, Lys77, Thr79, Asp80 and Ser81) show great flexibility in the Ca^{2+}-CaM ensemble determined in this work, which is not observed in the Ca^{2+}-CaM conformations in the PDB ensemble (black triangles), but cover a range of conformations in the substrate-bound state structures in the PDB (green circles). By contrast, the other five residues (Arg74, Lys75, Asp78, Glu82 and Glu83) are not very flexible in the Ca^{2+}-CaM state, but they change in conformation upon substrate binding. For comparison, the corresponding dihedral angle distributions in the control simulations carried out using the same protocol but without chemical shift restraints (the MD ensemble; see Methods) do not show multiple basins (Fig. S10).

The CS ensemble of structures was also validated against SAXS and FRET measurements [30,38,39]. Our results indicate that, while both the elongated (Fig. 4a) and the compact (Fig. 4b) structures of Ca^{2+}-CaM are not compatible with the SAXS profile, the CS ensemble is highly consistent with it (Fig. 4d). By contrast, the control simulation (MD ensemble) is not in good agreement with the SAXS profile (Fig. 4c). Similar conclusion can be drawn from the FRET measurements between residue 35 in the NTD and residue 110 in the CTD (Fig. S5). A large majority of distances between the donor and acceptor ranges between 30 Å and 40 Å, which is the characteristic of structures in between the elongated form and the compact form. Even though the size and orientation of the FRET probes precludes accurate prediction of the distance from the ensembles, the distance distribution calculated from the CS ensemble (Fig. S5b) shows considerable better agreement than the corresponding distribution calculated from the control MD simulations (Fig. S5a).

Analysis of the conformational fluctuations of the interdomain linker

An analysis of the backbone dihedral angle distributions in the *D. melanogaster* Ca^{2+}-CaM ensemble determined in this work indicates that most of these distributions are rather narrow (Fig. 5 and Figs. S6–S9). The only ones that are highly heterogeneous are those in the central region of the linker (residues 74–83; Fig. 5); similar results are obtained for the human Ca^{2+}-CaM ensemble. By contrast, the control simulation (MD ensemble) resulted in dihedral angle distributions without multiple basins (Fig. S10), indicating that the use of chemical shifts offers additional information with respect to that provided by the force field used in the simulation.

These results are consistent with previous conclusions based on ^{15}N relaxation data [27,28], as well as on the PRE [31,33], RDC [36] and SAXS [30] measurements discussed above, that the interdomain motions of Ca^{2+}-CaM are mainly associated

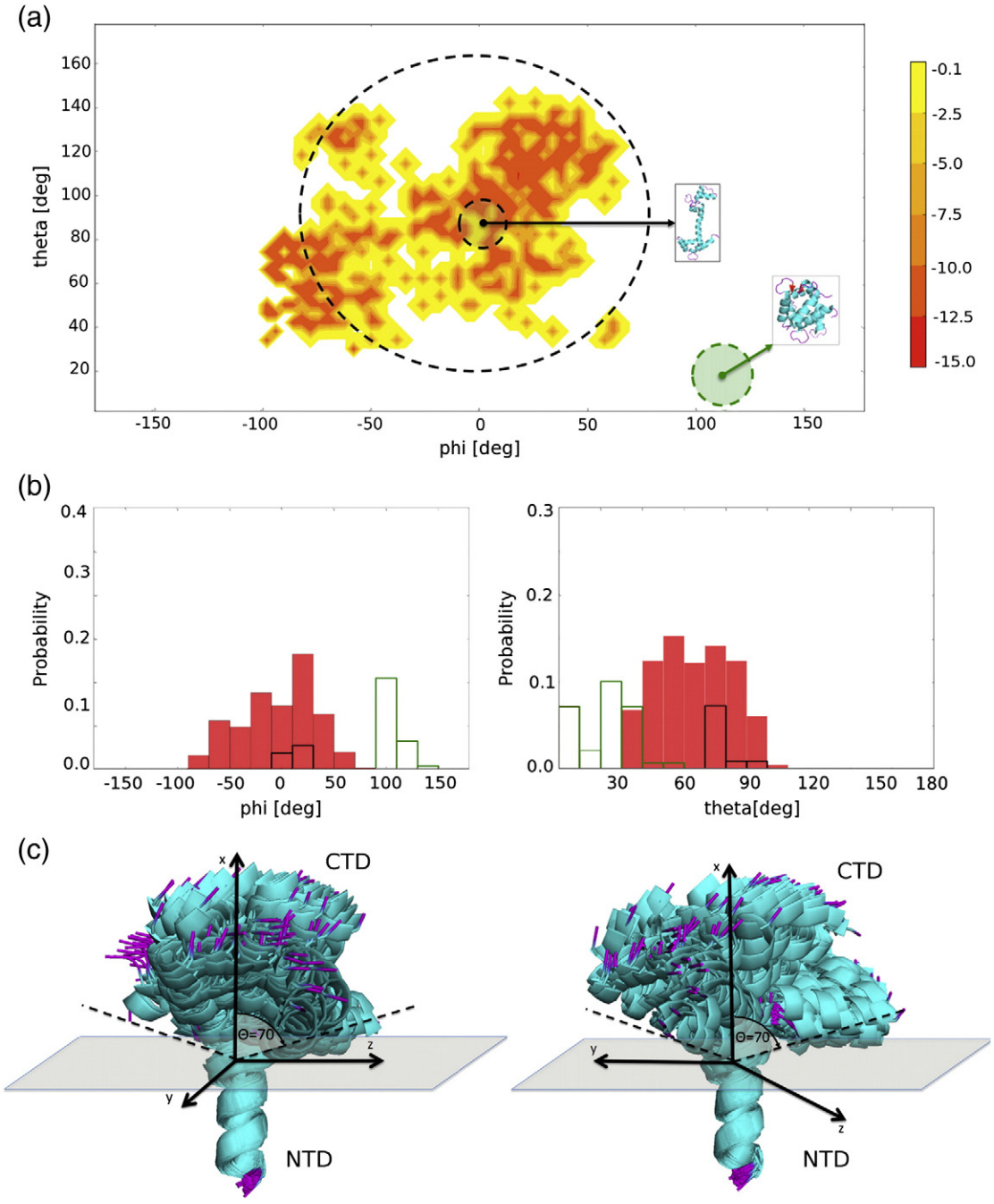

Fig. 6. Analysis of the conformational space sampled by a 27-residue peptide that corresponds to the interdomain linker (residues 65–91). The ensemble of conformations representing the structural fluctuations of this peptide is generated using experimental chemical shifts of *D. melanogaster* Ca^{2+}-CaM, with the peptide capped at both ends. See Fig. 2 for more details.

with the conformational fluctuations of the interdomain linker and that mutations in the linker have a crucial impact on the ability of calmodulin to bind its partners; indeed, it was recently reported that mutation of the five linker residues Lys77–Ser81 into alanine residues decreases the binding affinity of Ca^{2+}-CaM for skeletal muscle MLCK peptide from K_d = 426 pM to K_d = 780 pM [37,40]. The strong dependence of chemical shifts on the backbone dihedral angles, however, enabled us to characterize the distributions of the backbone dihedral angles in greater detail than before. To illustrate this point, we carried out a survey of the values of the dihedral angles found in the Ca^{2+}-CaM structures (Fig. 5,

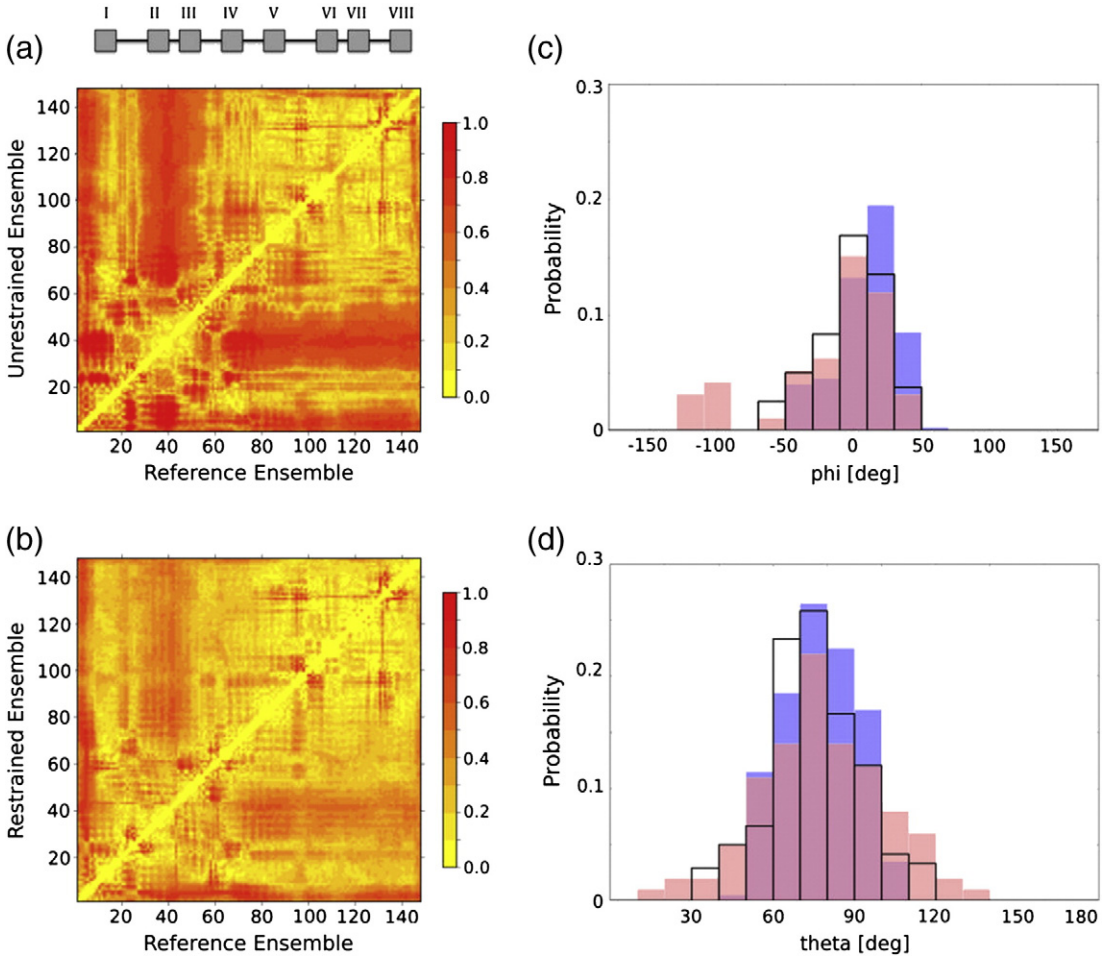

Fig. 7. Validation of the CS ensemble with the reference ensemble method (see Methods). S matrix of the (a) unrestrained ensemble and (b) two-replica restrained ensemble with respect to the distance distribution of backbone C^{α} atoms in the reference ensemble. Distributions of the (c) ϕ and (d) θ angles (see Fig. 1) in the reference (blue bars), unrestrained (red bars) and restrained ensembles (black bars).

black triangles) and the substrate-bound CaM (Fig. 5, green circles) available in the PDB. A comparison of these distributions and those corresponding to the CS ensemble demonstrates that five residues (Met76, Lys77, Thr79, Asp80 and Ser81) show great flexibility in the CS ensemble, which is not observed in the Ca^{2+}-CaM conformations in the PDB. These five residues cover a range of conformations in the substrate-bound state structures in the PDB. By contrast, the other five residues in the linker (Arg74, Lys75, Asp78, Glu82 and Glu83) are not very flexible in the Ca^{2+}-CaM state, but they change in conformation upon substrate binding.

In order to further investigate the behavior of the interdomain linker, we carried out replica-averaged MD simulations of a 27-residue peptide corresponding to the central linker using the chemical shifts of the corresponding residues (residues 65–91; Fig. 6). The comparison of the results for the full-length protein (Fig. 2) and for the isolated peptide (Fig. 6) reveals that the flexibility of the central linker is an intrinsic property of this region, which is not influenced greatly by the remainder of the protein, a result previously suggested by van der Spoel et al. [41] from unrestrained MD simulations of the capped interdomain linker. The differences in the distributions of the ϕ and θ angles obtained from the central α-helix analysis with respect that of the full-length protein are likely to be caused by the complex modes of capping of the central α-helix (Fig. S11). At the N-terminal end of the α-helix, the capping is carried out by a hydrogen bond between the carbonyl oxygen of Glu64 and the NH group of Glu67 and a hydrogen bond between the carbonyl oxygen of Glu67 and the NH group of Glu64. At the C-terminal end of the α-helix, the capping is carried out by a hydrogen bond between the carboxyl oxygen of Glu96 and the NH group of Glu96. Moreover, additional capping is carried out by hydrophobic interactions between N-terminal part of the central α-helix and highly hydrophobic helix I in the

NTD and between C-terminal part of the central α-helix and highly hydrophobic helix VIII in the CTD. Both hydrogen bond and hydrophobic modes of capping of the central α-helix were preserved over the whole course of MD simulation with the full-length protein.

Validation of the conformational ensembles with the reference ensemble method

We have previously shown that the incorporation of chemical shifts as replica-averaged structural restraints in MD simulations using Eq. (1) generates ensembles of conformations that do not significantly depend on the underlying force field [19]—if two different force fields are used in the two separated restrained simulations, the two resulting restrained ensembles are much more similar with each other than the two corresponding unrestrained ensembles [19]. This result is consistent with the observation that the use of replica-averaged restraints represents an implementation of the maximum entropy principle [25,26].

To further demonstrate the ability of NMR chemical shifts incorporated as structural restraints in a force field to provide an accurate representation of the dynamics of Ca^{2+}-CaM, we carried out the test of reference ensemble [42,43]. In this test, a so-called "reference ensemble" of conformations is generated at first by unrestrained MD simulations, in this case, using the CHARMM22* force field [44]. NMR chemical shifts are then back-calculated from this reference ensemble using Sparta+ [45] and employed as structural restraints in MD simulations using a second force field, in this case, AMBER99SB*-ILDN-Q [46]; without the use of restraints, this second force field corresponds to an ensemble (the "unrestrained ensemble") that is different from the reference ensemble. The aim of this procedure is to test the ability of NMR chemical shifts to reconstruct the distribution of structures of the reference ensemble, independently of the underlying force field [42,43]. This protocol allows for an objective cross-validation analysis since the atomic coordinates of conformations to be reconstructed are known exactly; thus, the structural heterogeneity obtained from the restrained simulations can be compared with high accuracy to that obtained from the reference ensemble [42,43].

Firstly, we evaluate the ability of the restrained and unrestrained simulations to reproduce chemical shift values corresponding to the reference ensemble. Our results indicate that the chemical shifts from the restrained simulations match more closely those obtained from the reference simulations than those obtained from the unrestrained simulations that use AMBER99SB*-ILDN-Q [46] without chemical shift restraints (Table S1 and Fig. S12a). Different numbers of replicas were tested [43]. After this initial test, we assessed the ability of the restrained simulations to reproduce the structural heterogeneity of the reference ensemble by comparing inter-residue distance distributions in the two ensembles. For this purpose, we generated 148 S × 148 S matrices [43] (Fig. 7a and b; Figs. S13 and S14). The S matrix provides a characterization of both the local and the global structural similarities of two ensembles with a value of 0 corresponding to identical distributions and a value of 1 corresponding to completely non-overlapping distributions of distances.

The S matrix calculated from the reference and unrestrained ensembles reveals a significant diversity between the two ensembles (Fig. 7a). The differences in their distance distributions result from the different parameterization of the CHARMM22* and AMBER99SB*-ILDN-Q force fields. The main ones concern inter-residue distances between the NTD (residues 1–74) and the CTD (residues 83–148). When the chemical shift restraints, however, are added to the AMBER99SB*-ILDN-Q force field, an overall increase in the similarity in the distance distributions is observed across the full-length protein (Fig. 7b). The best results are obtained when two replicas are used (Table S2 and Fig. S13). In this case, the enforcement of chemical shift restraints in the simulations results in an average *S* score between the reference and restrained ensembles of 0.29 and an RMSD of 6.5 Å. This level of accuracy demonstrates that the use of restraints makes the restrained ensemble identical within statistical errors with the reference ensemble, as it is comparable to the accuracy found by comparing the two halves of the reference simulations (Table S2 and Fig. S14).

To further verify whether the conformational fluctuations of the restrained ensemble also closely reproduce that of the reference ensemble, we compared the distribution of relative motions between NTD and CTD of Ca^{2+}-CaM in the reference, unrestrained and restrained ensembles (Fig. 7c and d). The distribution of the position of α-helix V was considerably different in the unrestrained ensemble in comparison to that of the reference ensemble. However, the addition of chemical shift restraints averaged over two replicas in AMBER99SB*-ILDN-Q force field was able to drive this distribution toward the distribution present in the reference ensemble (Fig. 7c and d).

Similar conclusions were reached by applying the reference ensemble test to the isolated peptide corresponding to the central linker (Fig. 6; Table S3 and Figs. S12b and S15). The *S* score between the reference and unrestrained ensemble decreased from 0.33 to 0.25 when the force field was complemented with the chemical shift restraints. Moreover, the interdomain orientations in the restrained simulations were more similar to that of reference than the unrestrained ensembles (Fig. S16). Thus, the inclusion of ensemble-averaged chemical shift restraints into AMBER99SB*-ILDN-Q force field was able to overcome the parameterization differences with the

CHARMM22* force field and hence to generate an ensemble of Ca^{2+}-CaM and its central linker conformations with almost identical interdomain orientation distribution; the distribution compatible with the given set of chemical shifts restraints.

In conclusion, we have used chemical shifts as replica-averaged structural restraints in MD simulations to determine ensembles of structures representing the interdomain motions of calmodulin in its calcium-loaded state (Ca^{2+}-CaM). As such motions are very well characterized experimentally [27–37], we have been able to present an extensive validation of the ensembles determined here with RDC [36], PRE [31], SAXS [30] and FRET [38,39] measurements, which were not used as restraints in the calculations. By exploiting the great sensitivity of chemical shifts to the conformational properties of proteins, we have defined the Ramachandran maps of the residues within the interdomain linker, which reveal their specific contributions to the interdomain motions experienced by this protein in its calcium-loaded state. As the type of methodology that we have described is general, these results provide initial support for the use of NMR chemical shifts to study interdomain motions of multidomain proteins in solution.

Methods

We studied the human and *D. melanogaster* CaM variants, for which experimental chemical shift data are available [29,34]. The amino acid sequence of *D. melanogaster* CaM is identical with that of human CaM except for three residues in the CTD (Y99F, Q143T and A147S). Their representative X-ray structures, 4CLN [47] and 1CLL [48], have a C^{α} RMSD of 0.75 Å.

The CS ensembles

Two ensembles of structures (CS ensembles) representing the structure and dynamics of the two variants were generated using MD simulations with replica-averaged chemical shift restraints [19,20]. In this procedure, we used a force field obtained by adding to the AMBER99SB*-ILDN-Q force field [46] a chemical shift-based energy term defined as

$$E_{CS} = \sum_{i=1}^{148}\sum_{j=1}^{6} E_{ij}\left(\delta_{ij}^{calc} - \delta_{ij}^{exp}\right) \quad (1)$$

where E_{ij} is a chemical shift-based energy term corresponding to an atom of type j (e.g., H^{α}, HN, N, C^{α}, C^{β}, C′) and to the ith residue in the protein [18–20,49]. The experimental chemical shifts are denoted by δ_{ij}^{exp}, and their corresponding calculated values δ_{ij}^{calc} are obtained as averages over two replicas in annealing cycles [19,20]. The inclusion of replica-averaged chemical shifts into the force field generates an ensemble of Ca^{2+}-CaM compatible with the given set of NMR chemical shifts in the sense of the maximum entropy principle [25,26].

The MD ensembles

In order to control for the effects of the chemical shift restraints, we calculated two additional ensembles of structures (MD ensembles) representing the structure and dynamics of the two variants (human and D. *melanogaster*) of Ca^{2+}-CaM with the same protocol used for the CS ensembles but without the chemical shift restraints.

The PDB ensemble

In addition to the CS and MD ensembles, we created an ensemble of structures of Ca^{2+}-CaM structures available in the PDB (PDB ensemble). PDB ensembles of this type have been shown to match NMR measurements, including RDCs, better than individual structures [50]. The PDB ensemble consisted of 10 X-ray structures (1CLM, 1CLL, 1EXR, 1OOJ, 1OSA, 1PRW, 1UP5, 3CLN, 3IFK and 4CLN).

MD simulations

All MD simulations were performed in explicit solvent using GROMACS [51] package. In all trajectories, the starting coordinates were derived from the crystal structure of Ca^{2+}-CaM from *D. melanogaster* (4CLN [47]) and the crystal structure of Ca^{2+}-CaM from *Homo sapiens* (1CLL [48]). The starting coordinates in simulations with the central linker alone were obtained from the coordinates of residues 65–91 by adding the ACE and NME caps to the N- and C-termini. The protein and the linker were initially solvated in a water box that extends 12 Å from their surfaces. The net charge of the system was neutralized by adding Na^{+} ions. The system was evolved with a time step of 2 fs by constraining the fast-bonded modes using LINCS [51]. Van der Waals interactions were accounted for using a cutoff of 12 Å. The particle mesh Ewald method [52] with a grid spacing of 1.09 Å was used for the electrostatic contribution to non-bonded interactions.

Reference ensemble calculations

In the reference ensemble test [42,43] that we carried out, the reference ensemble consisted of 480 structures. The ensemble was generated using the CHARMM22* force field [44] and the TIP3P explicit water model [53] in a series of annealing cycles between 300 K and 380 K. The system was initially heated from 300 K to 380 K for 100 ps, then kept at constant high temperature of 380 K for 100 ps, cooled down to 300 K for 300 ps and finally kept at constant low temperature of 300 K for 100 ps. The total simulation time was 30 ns. The final structures were extracted only from low constant temperature frames.

Unrestrained and restrained ensemble calculations

The unrestrained and restrained ensembles consisted of 480 structures. The structures were generated following the same procedure as for the reference ensemble but using the AMBER99SB*-ILDN-Q [46] force field with the TIP3P water model [53]. The chemical shift restraints [Eq. (1)] were implemented into GROMACS [51] using the

chemical shift predictor CamShift [18,49] and the PLUMED package [54].

Determination of structural ensembles using experimental chemical shifts

The ensembles representing the dynamics of, respectively, *D. melanogaster* and human Ca^{2+}-CaM were derived using experimentally measured chemical shifts available from the literature [34,35]. These ensembles were obtained by enforcing the chemical shift restraints over two replicas of MD simulations [Eq. (1)], as this number provided the best results in the reference ensemble test described in the results section. Calculations were carried out in the explicit solvent using the same annealing protocol described for the unrestrained and restrained ensembles with the AMBER99SB*-ILDN-Q [46] force field and the TIP3P water model [53]. The simulations for the two-replica system were carried out for 60 ns. The final ensemble consisted of 960 structures.

Comparison of structural ensembles using S matrices

In an S matrix [43], each entry, s_{ij}, represents the difference between the distributions P^{ref} and P^{res} of the distances between pairs of C^{α} atoms in the two ensembles to be compared

$$s_{ij} = \frac{1}{2}\sum_k \left| P^{\mathrm{ref}}_{ij,k} - P^{\mathrm{res}}_{ij,k} \right| \quad (2)$$

Here, k runs over the bins used to characterize the distributions; the subscripts ij indicate that the distributions P^{ref} and P^{res} refer to the C^{α} atom pair of residues i and j.

Acknowledgements

We are grateful to Dr. Nick Anthis and Dr. Marius Clore for sending us the ensemble of structures that they determined and that we used in Fig. S4. This study was funded by Biotechnology and Biological Sciences Research Council (P.K. and M.V.), The Federation of European Biochemical Societies (C.C.) and the European Union (C.C.).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2014.02.002.

Received 23 November 2013;
Received in revised form 18 January 2014;
Accepted 4 February 2014
Available online 12 February 2014

Keywords:
NMR spectroscopy;
molecular dynamics simulations;
replica-averaged structural restraints

Abbreviations used:
RDC, residual dipolar coupling; PRE, paramagnetic resonance enhancement; SAXS, small-angle X-ray scattering; FRET, fluorescence resonance energy transfer; NTD, N-terminal domain; CTD, C-terminal domain; MD, molecular dynamics.

References

[1] Apic G, Gough J, Teichmann SA. Domain combinations in archaeal, eubacterial and eukaryotic proteomes. J Mol Biol 2001;310:311–25.
[2] Ma BY, Tsai CJ, Haliloglu T, Nussinov R. Dynamic allostery: linkers are not merely flexible. Structure 2011;19:907–17.
[3] Kuriyan J, Eisenberg D. The origin of protein interactions and allostery in colocalization. Nature 2007;450:983–90.
[4] Wriggers W, Chakravarty S, Jennings PA. Control of protein functional dynamics by peptide linkers. Biopolymers 2005;80:736–46.
[5] Nussinov R. How do dynamic cellular signals travel long distances? Mol BioSyst 2012;8:22–6.
[6] Zhuravleva A, Clerico EM, Gierasch LM. An interdomain energetic tug-of-war creates the allosterically active state in Hsp70 molecular chaperones. Cell 2012;151:1296–307.
[7] Sarkar P, Reichman C, Saleh T, Birge RB, Kalodimos CG. Proline *cis-trans* isomerization controls autoinhibition of a signaling protein. Mol Cell 2007;25:413–26.
[8] Wang XQ, Wu C, Anh Vu JES, Dahlquist FW. Computational and experimental analyses reveal the essential roles of interdomain linkers in the biological function of chemotaxis histidine kinase chea. J Am Chem Soc 2012;134:16107–10.
[9] Akimoto M, Selvaratnam R, McNicholl ET, Verma G, Taylor SS, Melacini G. Signaling through dynamic linkers as revealed by PKA. Proc Natl Acad Sci USA 2013;110:14231–6.
[10] Russo L, Maestre-Martinez M, Wolff S, Becker S, Griesinger C. Interdomain dynamics explored by paramagnetic NMR. J Am Chem Soc 2013;135:17111–20.
[11] Yuwen T, Post CB, Skrynnikov NR. Domain cooperativity in multidomain proteins: what can we learn from molecular alignment in anisotropic media? J Biomol NMR 2011;51:131–50.
[12] Zhuravleva A, Gierasch LM. Allosteric signal transmission in the nucleotide-binding domain of 70-kDa heat shock protein (Hsp70) molecular chaperones. Proc Natl Acad Sci U S A 2011;108:6987–92.
[13] Chen K, Tjandra N. Determining interdomain structure and dynamics of a retroviral capsid protein in the presence of oligomerization: implication for structural transition in capsid assembly. Biochemistry 2013;52:5365–71.
[14] Maciejewski M, Tjandra N, Barlow PN. Estimation of interdomain flexibility of N-terminus of factor h using residual dipolar couplings. Biochemistry 2011;50:8138–49.
[15] Takayama Y, Schwieters CD, Grishaev A, Ghirlando R, Clore GM. Combined use of residual dipolar couplings and solution X-ray scattering to rapidly probe rigid-body conformational transitions in a non-phosphorylatable active-site mutant of the 128 kDa enzyme I dimer. J Am Chem Soc 2011;133:424–7.
[16] Ward AB, Sali A, Wilson IA. Integrative structural biology. Science 2013;339:913–5.
[17] Robinson CV, Sali A, Baumeister W. The molecular sociology of the cell. Nature 2007;450:973–82.

[18] Robustelli P, Kohlhoff K, Cavalli A, Vendruscolo M. Using NMR chemical shifts as structural restraints in molecular dynamics simulations of proteins. Structure 2010;18:923–33.
[19] Camilloni C, Robustelli P, De Simone A, Cavalli A, Vendruscolo M. Characterization of the conformational equilibrium between the two major substates of RNase a using NMR chemical shifts. J Am Chem Soc 2012;134:3968–71.
[20] Camilloni C, Cavalli A, Vendruscolo M. Assessment of the use of NMR chemical shifts as replica-averaged structural restraints in molecular dynamics simulations to characterize the dynamics of proteins. J Phys Chem B 2013;117:1838–43.
[21] Jensen MR, Salmon L, Nodet G, Blackledge M. Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. J Am Chem Soc 2010;132:1270–2.
[22] Robustelli P, Stafford KA, Palmer AG. Interpreting protein structural dynamics from NMR chemical shifts. J Am Chem Soc 2012;134:6365–74.
[23] Camilloni C, Cavalli A, Vendruscolo M. Replica-averaged metadynamics. J Chem Theory Comput 2013;9:5610–7.
[24] Granata D, Camilloni C, Vendruscolo M, Laio A. Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics. Proc Natl Acad Sci U S A 2013;110:6817–22.
[25] Cavalli A, Camilloni C, Vendruscolo M. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. J Chem Phys 2013;138:094112.
[26] Roux B, Weare J. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. J Chem Phys 2013;138:084107.
[27] Barbato G, Ikura M, Kay LE, Pastor RW, Bax A. Backbone dynamics of calmodulin studied by ^{15}N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. Biochemistry 1992;31:5269–78.
[28] Chang SL, Szabo A, Tjandra N. Temperature dependence of domain motions of calmodulin probed by NMR relaxation at multiple fields. J Am Chem Soc 2003;125:11379–84.
[29] Bertini I, Del Bianco C, Gelis I, Katsaros N, Luchinat C, Parigi G, et al. Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. Proc Natl Acad Sci U S A 2004;101:6841–6.
[30] Bertini I, Giachetti A, Luchinat C, Parigi G, Petoukhov MV, Pierattelli R, et al. Conformational space of flexible biological macromolecules from average data. J Am Chem Soc 2010;132:13553–8.
[31] Anthis NJ, Doucleff M, Clore GM. Transient, sparsely populated compact states of apo and calcium-loaded calmodulin probed by paramagnetic relaxation enhancement: interplay of conformational selection and induced fit. J Am Chem Soc 2011;133:18966–74.
[32] Gsponer J, Christodoulou J, Cavalli A, Bui JM, Richter B, Dobson CM, et al. A coupled equilibrium shift mechanism in calmodulin-mediated signal transduction. Structure 2008;16:736–46.
[33] Bertini I, Luchinat C, Nagulapalli M, Parigi G, Ravera E. Paramagnetic relaxation enhancement for the characterization of the conformational heterogeneity in two-domain proteins. Phys Chem Chem Phys 2012;14:9149–56.
[34] Ikura M, Kay LE, Bax A. A novel approach for sequential assignment of ^{1}H, ^{13}C, and ^{15}N spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. Biochemistry 1990;29:4659–67.
[35] Ikura M, Spera S, Barbato G, Kay LE, Krinks M, Bax A. Secondary structure and side-chain proton and carbon-13 resonance assignments of calmodulin in solution by heteronuclear multidimensional NMR spectroscopy. Biochemistry 1991;30:9216–28.
[36] Chou J, Li S, Klee C, Bax A. Solution structure of Ca^{2+}-calmodulin reveals flexible hand-like properties of its domains. Nat Struct Biol 2001;8:990–7.
[37] Anthis NJ, Clore GM. The length of the calmodulin linker determines the extent of transient interdomain association and target affinity. J Am Chem Soc 2013;135:9648–51.
[38] Slaughter BD, Allen MW, Unruh JR, Urbauer RJB, Johnson CK. Single-molecule resonance energy transfer and fluorescence correlation spectroscopy of calmodulin in solution. J Phys Chem B 2004;108:10388–97.
[39] Slaughter BD, Urbauer RJB, Urbauer JL, Johnson CK. Mechanism of calmodulin recognition of the binding domain of isoform 1b of the plasma membrane Ca^{2+}-ATPase: kinetic pathway and effects of methionine oxidation. Biochemistry 2007;46:4045–54.
[40] VanBerkum MFA, George SE, Means AR. Calmodulin activation of target enzymes: consequences of deletions in the central helix. J Biol Chem 1990;265:3750–6.
[41] van der Spoel D, de Groot B, Hayward S, Berendensen H, Vogel H. Bending of the calmodulin central helix: a theoretical study. Protein Sci 1996;5:2044–53.
[42] Kuriyan J, Petsko GA, Levy RM, Karplus M. Effect of anisotropy and anharmonicity on protein crystallographic refinement: an evaluation by molecular dynamics. J Mol Biol 1986;190:227–54.
[43] Richter B, Gsponer J, Varnai P, Salvatella X, Vendruscolo M. The MUMO (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. J Biomol NMR 2007;37:117–35.
[44] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4:187–217.
[45] Shen Y, Bax A. Sparta+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. J Biomol NMR 2010;48:13–22.
[46] Best RB, Hummer G. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. J Phys Chem B 2009;113:9004–15.
[47] Taylor D, Sack J, Maune J, Beckingham K, Quiocho F. Structure of a recombinant calmodulin from *Drosophila melanogaster* refined at 2.2-Å resolution. J Biol Chem 1991;266:21375–80.
[48] Chattopadhyaya R, Meador W, Means A, Quiocho F. Calmodulin structure refined at 1.7 Å resolution. J Mol Biol 1992;228:1177–92.
[49] Kohlhoff KJ, Robustelli P, Cavalli A, Salvatella X, Vendruscolo M. Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. J Am Chem Soc 2009;131:13894–5.
[50] Best RB, Lindorff-Larsen K, DePristo MA, Vendruscolo M. Relation between native ensembles and experimental structures of proteins. Proc Natl Acad Sci U S A 2006;103:10901–6.
[51] Hess B, Kutzner C, van der Spoel D, Lindahl E. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 2008;4:435–47.
[52] Essmann U, Perera L, Berkowitz M, Darden T, Lee H, Pedersen L. A smooth particle mesh Ewald method. J Chem Phys 1995;103:8577–93.

[53] Jorgensen W, Chandrasekhar J, Madura J, Impey R, Klein M. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–35.

[54] 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. Comp Phys Comm 2009;180:1961–72.