BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

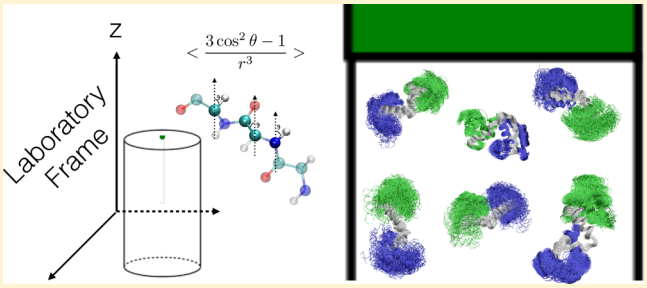
Using Pseudocontact Shifts and Residual Dipolar Couplings as Exact NMR Restraints for the Determination of Protein Structural Ensembles

Carlo Camilloni and Michele Vendruscolo*

Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom

Supporting Information

ABSTRACT: Nuclear magnetic resonance (NMR) spectroscopy provides detailed information about the structure and dynamics of proteins by exploiting the conformational dependence of the magnetic properties of certain atomic nuclei. The mapping between NMR measurements and molecular structures, however, often requires approximated descriptions based on the fitting of a number of parameters, thus reducing the quality of the information available from the experiments. To improve on this limitation, we show here that it is possible to use pseudocontact shifts and residual dipolar couplings as "exact" NMR restraints. We implement this strategy by using a replica-averaging method and illustrate its application by calculating an ensemble of structures representing the dynamics of the two-domain protein calmodulin.



The determination of structural ensembles representing the conformational fluctuations of proteins is a major goal in structural biology because it enables the observation of multiple substates populated by these molecules.[1−9] Determining these structural ensembles, however, is a challenging task that can require the integration of multiple sources of information from experiment and theory.[1−9]

Recently, it has been shown that it is possible to integrate these different sources in the framework of statistical mechanics by using the maximum entropy principle.[10−13] In this approach, one modifies the a priori knowledge about the proteins (i.e., their physicochemical properties as described by molecular mechanics force fields) using data acquired from experimental measurements. The simulations thus sample a conformational space corresponding to the maximum entropy perturbation of the force field that matches the experimental data. This strategy is particularly useful in NMR spectroscopy, in particular, for NOEs and chemical shifts.[11−13] An important aspect in all these approaches is the construction of a mapping between experimental observables and the protein conformations. Some observables can be approximately described using interatomic distances (e.g., NOE-derived distances[3,5,11−13]), while for others the mappings are more complicated functions of the atomic coordinates (e.g., chemical shifts[14,15]). The approximations in these mappings can reduce the structural information made available by the experimental measurements.

Among NMR observables, some not only are a function of the conformation of the molecule under observation but also depend on the interaction of the molecule itself with its environment. This is typically the case of residual dipolar couplings (RDCs), where a shift in the one-bond *J*-coupling is the result of the breaking of the rotational symmetry of a molecule due to the presence of an alignment medium.[16,17] This shift is proportional to the fraction of molecules whose average angle, ϑ_i, of the bond with respect to the static magnetic field is nonzero

$$D_i = -\frac{\mu_0 \gamma_1 \gamma_2 \hbar}{8\pi^3} \left\langle \frac{3\cos^2 \vartheta_i - 1}{r_i^3} \right\rangle \quad (1)$$

where μ_0 is the magnetic constant, γ_1 and γ_2 are the gyromagnetic moments of the nuclei, and $\hbar$ is the Planck constant.

Another NMR observable that depends on the interaction between a protein and its environment is the pseudocontact shift (PCS).[2,18,19] In this case, the presence of a metal ion susceptible to anisotropic magnetization determines a change in the chemical shift of a nucleus depending on its distance, r_i, from the metal ion and on the angle, ϑ_i, between the distance vector and the static magnetic field

$$\Delta\delta_i = -\chi \left\langle \frac{3\cos^2 \vartheta_i - 1}{r_i^3} \right\rangle \quad (2)$$

where χ is a constant that accounts for the magnetic susceptibility of the protein.[18−21] Here the interaction of the metal ion with the external magnetic field generates an additional magnetic field that breaks the rotational symmetry

Received: October 18, 2015
Revised: December 1, 2015
Published: December 1, 2015

DOI: 10.1021/acs.biochem.5b01138
Biochemistry 2015, 54, 7470−7476

of the metal ion. If the metal ion is bound to a protein then also the rotational symmetry of the protein due to its tumbling is broken, and each atomic nucleus of the protein feels the additional magnetic field to an extent proportional to its distance from the metal ion and to the local symmetry breaking, as indeed molecules not bound to the metal ion are practically not affected.

To introduce a description of this type of observable that does not depend on the environment, it is common to introduce a reference framework where the average effect of the environment on a conformation is given in terms of an alignment tensor.[16,17] This alignment tensor can either be determined from the data themselves[22] or by modeling the interaction of the molecule with the environment.[23] While the former approach assumes that the system is a rigid body,[22] which complicates the study of the dynamics, the latter can, at least in principle, describe correctly the dynamics of the system. In particular for charged media, however, the modeling of the interactions can be quite challenging.

An alternative to the introduction of an alignment tensor is to work in the laboratory reference frame, with the NMR static magnetic field defining the z direction, and use the NMR measurements to model at the same time the protein and its interactions with the environment.[24] This "tensor free" approach has been implemented using RDCs as restraints in replica-averaged molecular dynamics simulations, where averaged observables over multiple replicas of a protein are restrained with respect to the corresponding experimental data.[24−26] It can also be recast in an approach where the direction of the field is an addition degree of freedom that can be optimized using the maximum entropy principle thus allowing the use of multiple alignment media.[27]

In this approach RDC and PCS restraints can be considered "exact", at least to the extent to which it is possible to assume that eqs 1 and 2 describe exactly the phenomena under observation and as far as the maximum entropy principle provides an accurate framework to infer the structural ensembles underlying the experimental data. Indeed we are here not limited by the approximations resulting either from a rigid body assumption or from an inaccurate description of the interaction between the medium and the system. Furthermore, the restraints do not depend on any additional parameter; in particular, they can be applied using the correlation between the observables calculated using, respectively, eqs 1 and 2, averaged over the replicas, and the experimental data. The use of the correlation allows one to compare the observables calculated using eqs 1 and 2 with the experimental data up to a multiplicative constant whose specific value does not affect the results of the calculations. We note, however, that the structural ensembles resulting from "exact" structural restraints should not be considered as "exact" themselves, even if our implementation of eqs 1 and 2 does not involve free parameters, because the calculations required to obtain the ensembles require other approximations, including in particular the use of a specific force field in the molecular dynamics simulations.

Having recently applied this approach to RDCs measured in an alignment medium,[24] we show here that the approach is more general by extending it to the use of PCSs and RDCs measured in the presence of paramagnetic metal ions, by using the case of the two-domain protein calmodulin in its calcium-bound state[2,28−30] (Ca^{2+}−CaM).

METHODS

Molecular Dynamics Simulations. Molecular dynamics simulations of the calcium-bound state of calmodulin (Ca^{2+}−CaM) were performed using the CHARMM22* force field in explicit solvent. All the simulations were run in GROMACS[31] using PLUMED 2.[32] The van der Waals interactions and short-range electrostatic effects were implemented with a cutoff at 1.0 nm; long-range electrostatic effects were treated with the particle mesh Ewald method[33] on a mesh of 0.12 nm. All simulations were carried out in the isothermal−isobaric ensemble by thermosetting the system with the Bussi thermostat[34] and controlling the pressure with the Parrinello−Rahman barostat.[35] The starting conformation was taken from an available X-ray structure[36] (PDB code 1CLL). This structure was solvated with 17 400 water molecules and 16 Na^+ ions in a dodecahedron box of 550 nm^3 of volume. A 30 ns preliminary simulation was used to select 16 starting conformations.

Replica-Averaged Metadynamics (RAM) Simulations. RAM simulations[37] were performed using either pseudocontact shifts (PCSs) alone or PCSs and residual dipolar couplings (RDCs) as replica-averaged restraints, while metadynamics was implemented as well-tempered bias exchange.[38] In RAM simulations, the number of replicas is such that the larger is their number the better is the approximation of the maximum entropy principle[11] and the sampling of the conformational space due to the higher number of replicas used for bias exchange.[39] Sixteen replicas of the system were simulated in parallel at 305 K with a restraint applied on the correlation between the experimental and replica-averaged back-calculated PCSs and on the correlation between experimental and replica-averaged back-calculated RDCs. The starting conformations were selected as described above by extracting a frame every 2 ns from preliminary 30 ns long simulations.

For each PCS and each replica, we calculate both the distance, r_i, between the metal ion and the affected nucleus i and the angle ϑ_i between the distance vector and the z-axis (see eq 1). Using 16 replicas, we calculate the replica-averaged PCS as

$$\Delta\delta = -\frac{1}{16}\sum_{i=1}^{16}\frac{3\cos^2\vartheta_i - 1}{r_i^3} \tag{3}$$

This same approach is used for RDCs, where distances and angles are calculated for each bond vector as (see eq 2)

$$D = -\frac{1}{16}\frac{\mu_0\gamma_1\gamma_2\hbar}{8\pi^3}\sum_{i=1}^{16}\frac{3\cos^2\vartheta_i - 1}{r_i^3} \tag{4}$$

Given all the PCSs or RDCs, we then calculate their correlations with the experimental data. For the PCSs, a linear restraint is applied on the correlation as

$$V_{PCS} = -K_{PCS}[\rho(\boldsymbol{\Delta\delta}^{calcd}, \boldsymbol{\Delta\delta}^{expt}) - 1] \tag{5}$$

and for the RDCs, a similar restraints is applied as

$$V_{RDC} = -K_{RDC}[\rho(\mathbf{D}^{calcd}, \mathbf{D}^{expt}) - 1] \tag{6}$$

where $\rho(\boldsymbol{\Delta\delta}^{calcd},\boldsymbol{\Delta\delta}^{expt})$ is the coefficient of correlation between calculated ($\boldsymbol{\Delta\delta}^{calcd}$) and experimental ($\boldsymbol{\Delta\delta}^{expt}$) PCS values, and $\rho(\mathbf{D}^{calcd},\mathbf{D}^{expt})$ is the coefficient of correlation between calculated ($\mathbf{D}^{calcd}$) and experimental ($\mathbf{D}^{expt}$) RDC values.

DOI: 10.1021/acs.biochem.5b01138
Biochemistry 2015, 54, 7470−7476

A preliminary 10 ns long simulation was used to linearly increase the restraint force up to 10 000 kJ/mol, in such a way to maximize the correlation before starting the bias-exchange metadynamics sampling. This force constant has been chosen as the largest for which the simulation was stable, in agreement with the maximum entropy principle prescription.[11,12]

We also note that the use of PCSs and RDCs as exact NMR restraints can be implemented in other methods to use NMR data in combination with molecular simulations.[3,9,29,30] In particular, the replica-averaging tensor-free approach can be used to optimize the orientations of multiple rigid structures in a way that reproduces the common alignment tensor.[26]

No Free Parameters Are Required in the Simulations. *Scaling Factor.* In many standard approaches to use NMR data as structural restraints in the simulations, the matching between experimental and calculated values is enforced by minimizing the sum of the squares of the differences between them

$$V_{\mathrm{NMR}} = -K_{\mathrm{NMR}} \sum_i (D_i^{\mathrm{calcd}} - D_i^{\mathrm{expt}})^2 \quad (7)$$

where the sum runs over the number of available experimental data points, D_i^{calcd} and D_i^{expt} are, respectively, the calculated and experimental data points, and K_{NMR} is the strength of the restraint term, which in the maximum entropy principle approach should be taken as large as the numerical stability of the simulations allows. This approach requires a scaling factor in the calculation of the D_i^{calcd} values; otherwise eq 7 can never become zero. By contrast in eqs 5 and 6, the restraint term involves a correlation between experimental and calculated values, and hence it can be minimized without introducing a scaling factor.[24−26] This procedure is of particular importance when leading with flexible systems where the scaling factor can be a function of the conformation, because more extended structures can be more aligned than compact ones.

Rhombic Alignment. A rhombic alignment can be present when there are dynamics in the external degrees of freedom. In case of RDCs measured in the presence of an alignment medium, these dynamics result in different interchanging orientations of the protein; in the case of PCSs, the paramagnetic center can have multiple alignments with the magnetic field. In the replica method that we use in this work, the replicas distribute themselves automatically between the different orientations under the effect of the restraint term (eq 6). Hence there is no need to define a rhombic parameter to weight the different orientations.[24−26]

Back-Calculation of NMR Observables. By imposing PCSs and RDCs as structural restraints using the approach described here, we not only provide information on the relative distances between atoms but also on the orientation of interatomic vectors with respect to a predefined z direction, thus resulting in a breaking of the rotational symmetry of the molecule. As a consequence in the resulting ensemble, the absolute positions of each conformation are meaningful and needed to back calculate PCSs and RDCs directly from eqs 1 and 2 a posteriori.

Implementation of the Collective Variables. Bias-exchange metadynamics was set up using one collective variable (CV) per replica, five collective variables were selected in order to enhance the sampling of the relative motion of the two domains and applied on three different replicas, and a single replica was left unbiased. The selected CVs were (1) the ALPHABETA calculated over the φ and ψ angles for residues between 74 and 83, (2) the DIHCOR calculated over the ψ angles for successive residues between 74 and 83, (3) the ALPHARMSD (α-helical content) for residues 63 to 81, (4) the ALPHARMSD (α-helical content) for residues 77 to 94, and (5) the DHENERGY, Debye−Hückel energy, between the two domains calculated using only the charged side-chains belonging to the two domains. The choice of the above CVs was based on the previous observation of the role played by linker residues, by the observation that the C-terminal α-helix of the NTD and the N-terminal α-helix of CTD can be found formed to a different extent in the PDB, and by observing the large amount of charged residues present in both domains. Gaussians deposition was performed with an initial rate of 0.125 kJ/(mol·ps), a bias-factor of 10, and σ values set to 0.2, 0.1, 0.11, 0.11, and 0.5 for the above five CVs, respectively.

Furthermore, in order to limit the extent of accessible space along each collective variable and correctly treat the problem of the borders, we set the bias as constant outside a defined interval for each CV,[40] because it has been shown that this approach lead to a correct reconstruction of a one-dimensional free energy landscape inside the chosen range; intervals were set to 0−20, 0−10, 4−14, 4−14, and −10−36 for the five CVs, respectively. Each replica have been evolved for 130 ns, with exchange trials every 50 ps. The control unrestrained simulation has been performed with the same bias-exchange setup.

Convergence of the Simulations. The convergence of the sampling was assessed by monitoring the differences of the free energies in the range between 0 and 25 kJ/mol at increasing simulation length during the simulations and between replicas biased on the same CV. After the first 90 ns, the free energy landscapes are stable within less than 1 kJ/mol, suggesting that all the relevant minima in the landscape have been found, and the average changes in the free energy landscapes over the last 40 ns of simulations are below 1 kJ/mol. Replica-exchange is optimal with an average exchange probability per replica of 35% and with all replicas exchanging relatively uniformly among them. Each replica samples multiple regions of the CV space thus enabling the sampling of multiple minima and suggesting an overall ergodic behavior. Taken together these results suggest that the free energies that we obtained from the RAM simulations are on average correct within less than 1 kJ/mol (Figures S1−S3). The reweighting of the trajectories has been performed using the weighted histogram analysis (WHAM).[41,42]

Availability of the Method. All above methods are available in PLUMED 2[32] (www.plumed.org). A sample script is provided in the Supporting Information.

Sketch-Map Analysis. Sketch-map collective variables have been calculated using as input the φ and ψ dihedral angles for residues from 63 to 94 to keep into account the system from the NTD C-terminal to the CTD N-terminal α-helix.

RESULTS

Structural Ensembles from Replica-Averaged Metadynamics Simulations. We performed replica-averaged metadynamics (RAM) simulations[37] of Ca^{2+}−CaM in explicit solvent (see Methods). The averaging of the experimental data was performed over 16 replicas of the protein, and the sampling was enhanced using metadynamics along five collective variables each applied on three different replicas leaving one replica without any metadynamics bias (see Methods). We generated three ensembles: a control unrestrained ensemble

DOI: 10.1021/acs.biochem.5b01138
Biochemistry 2015, 54, 7470−7476

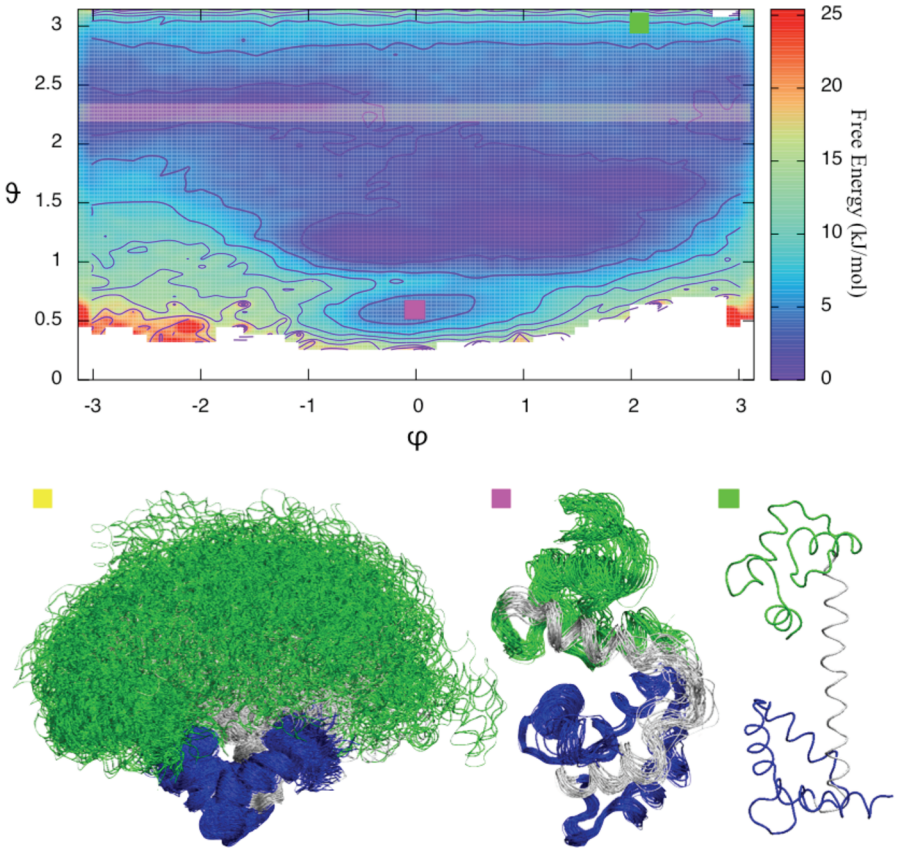


Figure 1. Free energy surface (in kJ/mol) of the Ca^{2+}–CaM ensemble generated using PCSs and RDCs (PCS+RDC ensemble) shown as a function of two angles, ϑ and ϕ (see text). Structures in the low free energy region (yellow) at ϑ around 2.2 rad are shown at the bottom left, compact structures of low free energy (magenta) are shown at the bottom center, while a high free energy structure resembling crystal structures (green) is shown at the bottom right. Structures are represented as tubes colored in blue (NTD), green (CTD), and white (linker region). All structures are aligned on the NTD.

(MD ensemble) without structural restraints, an ensemble in which PCS restraints were added to the force field (PCS ensemble), and an ensemble where both the PCSs and RDCs measured in the presence of Tb^{3+} as a paramagnetic metal[2,29] were used as restraints (PCS+RDC ensemble). The structures in the PCS+RDC ensemble are available as Supporting Information in PDB format together with their statistical weights in order to enable statistical averages to be calculated. The Tb^{3+} data set consists of 32 PCSs for residues in the N-terminal domain (NTD), 82 PCSs for residues in the C-terminal domain (CTD), and 32 RDCs for bonds in the CTD.

Validation of the Structural Ensembles. To validate our results, we used an additional data set of PCSs and RDCs measured in the presence of another paramagnetic ion (Tm^{3+}), which consists of 82 PCSs for residues in the NTD, 82 PCSs for residues in the CTD and 30 RDCs for bonds in the CTD. The correlation between the Tb^{3+} and Tm^{3+} RDC data sets is 0.5 with an overlap of 80% between the nuclei for which a measurement is available (Figure S4). The PCSs have a correlation of 0.9 for the N-terminal domain with just a 30% overlap between the data sets and a correlation of 0.6 with an overlap of 80% for the C-terminal domain (Figure S4). The ensembles are also validated using small-angle X-ray scattering (SAXS) data[29] (Figure 3).

NMR-Derived Free Energy Landscapes of Calmodulin. The extent of the conformational space explored by calmodulin is illustrated by the free energy landscape of the PCS+RDC ensemble (Figure 1). This free energy landscape is calculated as a function of the ϑ and ϕ angles,[43] which are the spherical angles measured for the axis of α-helix V defined using as a reference the axis α-helix IV. Consistently with previous results,[2,29,30,43] this ensemble is characterized by a relatively flat distribution of structures for the ϕ angle in the region of ϑ between 1.5 and 3 rad (i.e., 85° and 170°), indicating that when the two domains are relatively open they can move almost independently from each other. Instead in more compact conformations (i.e., ϑ between 0.5 and a 1.5 rad; 30° to 85°) the two α-helices can only move in a narrower range of values for ϕ. The average orientation is in this case ϑ = 1.9 rad and ϕ = 0.2 rad, that is, with the two α-helices bent at a 110° angle, a value compatible with that previously found using the data used in this work together with other restraints.[2,29]

To assess the information content of eqs 1 and 2 for PCSs and RDCs, we compared the free energy landscape of the PCS +RDC ensemble with those of the PCS ensemble and of the MD ensemble (Figure 2). To measure the differences between these free energy landscapes, we used the Kullback–Leibler divergence (KL). The free energy landscape of the MD ensemble shows significant differences from those of both the PCS and the PCS+RDC ensembles, indicating that the addition of the experimental restraints introduces relevant information. By contrast, the free energy landscapes of the PCS and the PCS +RDC ensembles are identical within errors. In particular, the MD ensemble contains a large number of relatively close conformations, while the PCS and PCS+RDC ensembles are on average more open. These results suggest that the PCSs

DOI: 10.1021/acs.biochem.5b01138
Biochemistry 2015, 54, 7470–7476

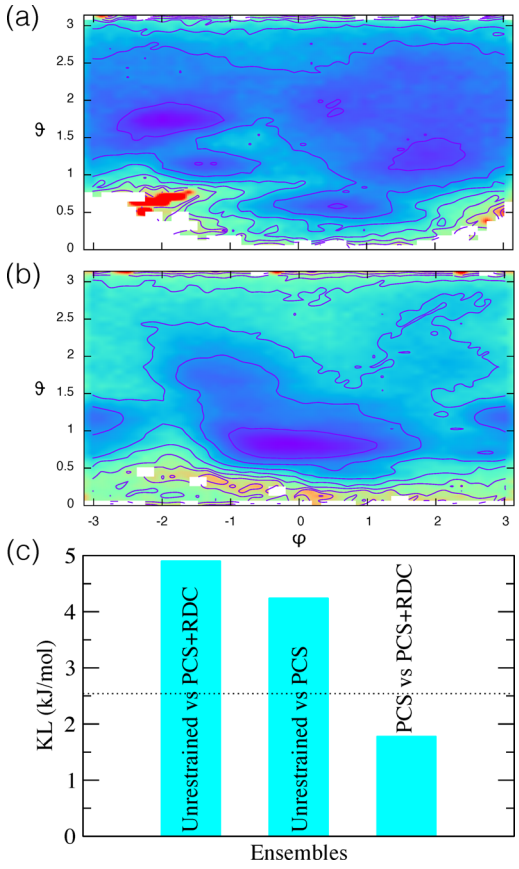


Figure 2. Comparison of the free energy surfaces of the PCS and MD ensembles. (a,b) Free energy surfaces (in kJ/mol) and as a function of ϑ and ϕ (see Figure 1) of the PCS ensemble (a) and the MD ensemble (b). (c) The differences between these free energy surfaces and that of the PCS+RDC ensemble are measured using the Kullback–Leibler divergence, which has been multiplied by k_BT to compare it with the thermal energy shown by the dotted line representing the energy scale at the experimental temperature (2.54 kJ/mol).

alone are sufficient to determine the main dynamical features of Ca^{2+}–CaM. We note that these results are consistent with other recent ones where conformationally averaged RDCs caused by steric alignment and of both RDCs and pseudocontact shifts caused by paramagnetic alignment contain a similar amount of information.[44]

Comparison of the Structural Ensembles. The differences between the three ensembles should be put in the context of their agreement with the data used as restraints as well as other independent data (Figure 3). The PCS+RDC ensemble is that in best agreement with all the available data. While the internal dynamics of the NTD is well captured already by the force field alone in the MD ensemble, it is clear that the relative motions of the two domains can be described more accurately by the inclusion of PCSs and RDCs restraints. In particular, when looking at the SAXS intensity (calculated with Crysol[29]), we conclude that PCSs alone provide already sufficient information to obtain an ensemble where the relative motions of the two domains are described accurately. We also point out that when PCSs and RDCs are used as restraints, it is possible to back-calculate them directly without assumptions using eqs 1 and 2, while otherwise (i.e., for data sets not included in the simulations and or for unrestrained simulations) one should use the single value decomposition method.[22]

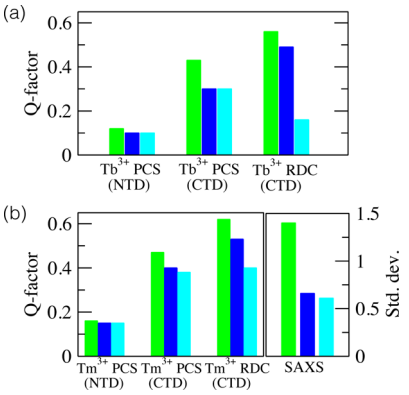


Figure 3. Comparisons between back-calculated and experimental PCS, RDC, and SAXS values. Results are reported as Q-factors or standard deviations for the three ensembles described in this work: MD ensemble (green), PCS ensemble (dark blue), and PCS+RDC ensemble (light blue). (a) Results for the data used as restraints in the PCS+RDC ensembles. (b) Results for the data not used as restraints: PCSs and RDCs measured in Tm^{3+} and SAXS.[2,29]

Substates in the Structural Ensembles. To illustrate further the richness of the dynamics of Ca^{2+}–CaM, which cannot be fully described in terms of the two angles used in Figures 1 and 2, we analyzed the ensemble using a dimensional reduction technique, namely, sketch-map.[45] We obtained in this way a free energy landscape by using as input all the ϕ and ψ backbone dihedral angles of the extended linker region from residues 63 to 94 (Figure 4). From this analysis, we found that the ensemble is characterized by a large number of similarly populated microstates, which accurately represent the rich conformational heterogeneity of the Ca^{2+}–CaM state of calmodulin. These microstates illustrate the extent of dynamics of calmodulin compatible with the available experimental data and show how it is possible to rationalize these dynamics in terms of the equilibrium between multiple, relatively homogeneous substates. To analyze and validate the substates identified by the sketch-map analysis, we projected on the same two collective variables five PDB structures of Ca^{2+}–CaM bound to different peptides (2F3Y, 2LTL, 2M0J, 2M0K, and 2MG5) and a PDB structure for the free state of Ca^{2+}–CaM (1CLL) (Figure 4). We found that the PDB structure of the free state falls inside the most populated basin and that the ensemble of conformations belonging to this basin represents the structural fluctuations around the X-ray structure. Indeed we found that the minimum RMSD calculated over all the backbone atoms between conformations of this basin and the X-ray structure is only 1.5 Å (Table S1). On the contrary, all the bound PDB structures fall in high free energy regions of the plot but always near different local minima (Figure 4). In this case, it is possible to show that the conformations belonging to the local minima are the most similar to the PDB structures, with a minimum RMSD of ∼5 Å. Indeed in the bound state PDB structures, the two domains are always more closed (in contact) than the conformations found in the nearest local minima (Table S1). These features are not captured to the same extent by the unrestrained simulation (Table S1). These results demonstrate once again the important role played by linker residues in determining the binding properties of Ca^{2+}–CaM; furthermore they support the result, also found using chemical shifts restraints, that the binding of Ca^{2+}–CaM involves substrate-bound-like conformations populated in the free state.[30,43]

DOI: 10.1021/acs.biochem.5b01138
Biochemistry 2015, 54, 7470–7476

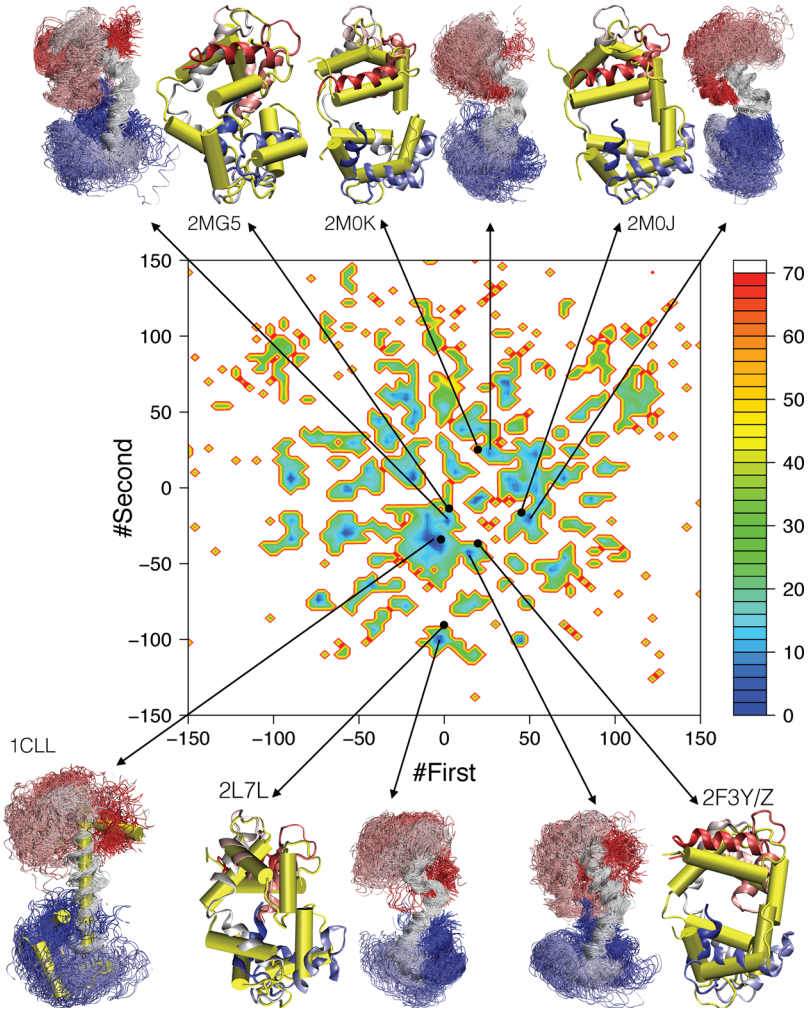


Figure 4. Sketch-map representation[45] of the PCS+RDC ensemble. From the free energy surface (in kJ/mol), as a function of two coordinates determined by the sketch-map dimensional reduction algorithm,[45] one can identify a small set of structurally homogeneous substates. The color code is the same as in Figure 1. On the same coordinates are represented five PDB structures of Ca^{2+}−CaM bound to different peptides (2F3Y, 2LTL, 2M0J, 2M0K, and 2MG5) and a PDB structure for the free state of Ca^{2+}−CaM (1CLL). All the bound structures falls in high free energy regions next to local minima; the most similar conformation from the local minima (and from the ensemble) is superposed to the PDB structure. The PDB structure of the free state falls inside the free energy minimum corresponding to the ground state.

CONCLUSIONS

We have shown that it is possible to use PCSs and RDCs as exact structural restraints in molecular dynamics simulations. The approach that we have presented can be used to generate structural ensembles in the maximum entropy framework to provide an accurate representation of the dynamics of proteins associated with equilibrium structural fluctuations.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.5b01138.

Methods, additional figures, and PCS+RDC ensemble in PDB format with statistical weights (PDF)

PCS+RDC ensemble in PDB format with statistical weights (XLXS)

PCS+RDC ensemble in PDB format with statistical weights (XLXS)

PCS+RDC ensemble in PDB format with statistical weights (XLXS)

PCS+RDC ensemble in PDB format with statistical weights (XLXS)

PCS+RDC ensemble in PDB format with statistical weights (XLXS)

AUTHOR INFORMATION

Corresponding Author

*Prof. Michele Vendruscolo. E-mail: mv245@cam.ac.uk. Phone: +44 1223 763873.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Clore, G. M., and Schwieters, C. D. (2004) How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation? *J. Am. Chem. Soc. 126*, 2923−2938.

(2) Bertini, I., Del Bianco, C., Gelis, I., Katsaros, N., Luchinat, C., Parigi, G., Peana, M., Provenzani, A., and Zoroddu, M. A. (2004) Experimentally exploring the conformational space sampled by domain

DOI: 10.1021/acs.biochem.5b01138
Biochemistry 2015, 54, 7470−7476

reorientation in calmodulin. *Proc. Natl. Acad. Sci. U. S. A. 101*, 6841−6846.
(3) Lindorff-Larsen, K., Best, R. B., DePristo, M. A., Dobson, C. M., and Vendruscolo, M. (2005) Simultaneous determination of protein structure and dynamics. *Nature 433*, 128−132.
(4) Mittermaier, A., and Kay, L. E. (2006) New tools provide new insights in NMR studies of protein dynamics. *Science 312*, 224−228.
(5) Vögeli, B., Segawa, T. F., Leitz, D., Sobol, A., Choutko, A., Trzesniak, D., van Gunsteren, W., and Riek, R. (2009) Exact distances and internal dynamics of perdeuterated ubiquitin from NOE buildups. *J. Am. Chem. Soc. 131*, 17215−17225.
(6) Boehr, D. D., Nussinov, R., and Wright, P. E. (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol. 5*, 789−796.
(7) Tzeng, S.-R., and Kalodimos, C. G. (2011) Protein dynamics and allostery: an NMR view. *Curr. Opin. Struct. Biol. 21*, 62−67.
(8) Lewandowski, J. R., Halse, M. E., Blackledge, M., and Emsley, L. (2015) Direct observation of hierarchical protein dynamics. *Science 348*, 578−581.
(9) Jensen, M. R., Zweckstetter, M., Huang, J.-r., and Blackledge, M. (2014) Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem. Rev. 114*, 6632−6660.
(10) Pitera, J. W., and Chodera, J. D. (2012) On the use of experimental observations to bias simulated ensembles. *J. Chem. Theory Comput. 8*, 3445−3451.
(11) Cavalli, A., Camilloni, C., and Vendruscolo, M. (2013) Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J. Chem. Phys. 138*, 094112.
(12) Roux, B., and Weare, J. (2013) On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *J. Chem. Phys. 138*, 084107.
(13) Boomsma, W., Ferkinghoff-Borg, J., and Lindorff-Larsen, K. (2014) Combining experiments and simulations using the maximum entropy principle. *PLoS Comput. Biol. 10*, e1003406.
(14) Cavalli, A., Salvatella, X., Dobson, C. M., and Vendruscolo, M. (2007) Protein structure determination from NMR chemical shifts. *Proc. Natl. Acad. Sci. U. S. A. 104*, 9615−9620.
(15) Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J. M., Liu, G., Eletsky, A., Wu, Y., Singarapu, K. K., Lemak, A., et al. (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. U. S. A. 105*, 4685−4690.
(16) Tolman, J., Flanagan, J., Kennedy, M. A., and Prestegard, J. (1995) Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc. Natl. Acad. Sci. U. S. A. 92*, 9279−9283.
(17) Tjandra, N., and Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science 278*, 1111−1114.
(18) Bleaney, B. (1972) Nuclear magnetic resonance shifts in solution due to lanthanide ions. *J. Magn. Reson. 8*, 91−100.
(19) Mayo, B. (1973) Lanthanide shift reagents in nuclear magnetic resonance spectroscopy. *Chem. Soc. Rev. 2*, 49−74.
(20) Bertini, I., Felli, I. C., and Luchinat, C. (1998) High magnetic field consequences on the NMR hyperfine shifts in solution. *J. Magn. Reson. 134*, 360−364.
(21) Bertini, I., Luchinat, C., and Parigi, G. (2002) Magnetic susceptibility in paramagnetic NMR. *Prog. Nucl. Magn. Reson. Spectrosc. 40*, 249−273.
(22) Losonczi, J. A., Andrec, M., Fischer, M. W., and Prestegard, J. H. (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson. 138*, 334−342.
(23) Zweckstetter, M., and Bax, A. (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. *J. Am. Chem. Soc. 122*, 3791−3792.
(24) Camilloni, C., and Vendruscolo, M. (2015) A Tensor-Free Method for the Structural and Dynamical Refinement of Proteins using Residual Dipolar Couplings. *J. Phys. Chem. B 119*, 653−661.
(25) Wirz, L. N., and Allison, J. R. (2015) Comment on "A tensor-free method for the structural and dynamic refinement of proteins using residual dipolar couplings. *J. Phys. Chem. B 119*, 8223−8224.
(26) Camilloni, C., and Vendruscolo, M. (2015) Reply to "Comment on 'A Tensor-Free Method for the Structural and Dynamic Refinement of Proteins using Residual Dipolar Couplings'. *J. Phys. Chem. B 119*, 8225−8226.
(27) Olsson, S., Ekonomiuk, D., Sgrignani, J., and Cavalli, A. (2015) Molecular dynamics of biomolecules through direct analysis of dipolar couplings. *J. Am. Chem. Soc. 137*, 6270−6278.
(28) Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W., and Bax, A. (1992) Backbone dynamics of calmodulin studied by nitrogen-15 relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry 31*, 5269−5278.
(29) Bertini, I., Giachetti, A., Luchinat, C., Parigi, G., Petoukhov, M. V., Pierattelli, R., Ravera, E., and Svergun, D. I. (2010) Conformational space of flexible biological macromolecules from average data. *J. Am. Chem. Soc. 132*, 13553−13558.
(30) Anthis, N. J., Doucleff, M., and Clore, G. M. (2011) 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. 133*, 18966−18974.
(31) Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., et al. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics 29*, 845−854.
(32) Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., and Bussi, G. (2014) PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun. 185*, 604−613.
(33) Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N⊕ log (N) method for Ewald sums in large systems. *J. Chem. Phys. 98*, 10089−10092.
(34) Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys. 126*, 014101.
(35) Parrinello, M., and Rahman, A. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys. 52*, 7182−7190.
(36) Chattopadhyaya, R., Meador, W. E., Means, A. R., and Quiocho, F. A. (1992) Calmodulin structure refined at 1.7 Å resolution. *J. Mol. Biol. 228*, 1177−1192.
(37) Camilloni, C., Cavalli, A., and Vendruscolo, M. (2013) Replica-averaged metadynamics. *J. Chem. Theory Comput. 9*, 5610−5617.
(38) Piana, S., and Laio, A. (2007) A bias-exchange approach to protein folding. *J. Phys. Chem. B 111*, 4553−4559.
(39) Camilloni, C., and Vendruscolo, M. (2014) Statistical mechanics of the denatured state of a protein using replica-averaged metadynamics. *J. Am. Chem. Soc. 136*, 8982−8991.
(40) Baftizadeh, F., Cossio, P., Pietrucci, F., and Laio, A. (2012) Protein folding and ligand-enzyme binding from bias-exchange metadynamics simulations. *Curr. Phys. Chem. 2*, 79−91.
(41) Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., and Kollman, P. A. (1995) Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem. 16*, 1339−1350.
(42) Marinelli, F., Pietrucci, F., Laio, A., and Piana, S. (2009) A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations. *PLoS Comput. Biol. 5*, e1000452.
(43) Kukic, P., Camilloni, C., Cavalli, A., and Vendruscolo, M. (2014) Determination of the individual roles of the linker residues in the interdomain motions of calmodulin using NMR chemical shifts. *J. Mol. Biol. 426*, 1826−1838.
(44) Andralojć, W., Berlin, K., Fushman, D., Luchinat, C., Parigi, G., Ravera, E., and Sgheri, L. (2015) Information content of long-range NMR data for the characterization of conformational heterogeneity. *J. Biomol. NMR 62*, 353−371.
(45) Ceriotti, M., Tribello, G. A., and Parrinello, M. (2011) Simplifying the representation of complex free-energy landscapes using sketch-map. *Proc. Natl. Acad. Sci. U. S. A. 108*, 13023−13028.

DOI: 10.1021/acs.biochem.5b01138
Biochemistry 2015, 54, 7470−7476