Exploring the natural conformational changes of the C-terminal domain of calmodulin

J. Elezgaray,* G. Marcou, and Y. H. Sanejouand
Centre de Recherche Paul Pascal, Avenue Schweitzer, 33600 Pessac, France
(Received 8 April 2002; published 23 September 2002)

Several experimental results suggest that the Ca^{2+}-loaded C-terminal domain of calmodulin (or some of its mutants) exhibits conformational changes triggered solely by thermal fluctuations. The time scales involved are in the 10^{-6}–10^{-3} s range. Here we develop a theoretical method to explore this type of motions based on a modified version of molecular dynamics algorithm where the secondary structure motifs are held fixed. In this version, increasing the temperature enhances the sampling of conformations with locally fixed secondary structures. From the temperature dependence of the transition rate between various conformational states, we obtain characteristic times that are consistent with those observed experimentally.

DOI: 10.1103/PhysRevE.66.031908 PACS number(s): 87.15.He

I. INTRODUCTION

Protein function is not based on static structure, but rather on structural dynamics. Conformational changes have been linked to function in many cases, including binding, catalysis and protein–nucleic-acid interactions. Experimental evidences for such conformational fluctuations are available since the 1970s, when the first measurements of hydrogen exchange and fluorescence quenching were reported. An accurate, yet static information can be obtained through crystallographic data. Typical examples include the binding of a substrate or, in a more general way, the transmission of conformational changes to a distant site, as in allosteric mechanisms. Very often, the internal motions of proteins involve a collective character. For instance, in the case of citrate synthase, the binding of coenzyme *A* induces an 18° rotation of the small domain around an axis located near residue 274, resulting in the so-called hinge-bending motion [1–3]. This results in the closure of the substrate binding site.

An interesting question is whether these collective motions arise from the coupling with an exterior agent (ligand), or they are already present as an intrinsic motion of the protein. An increasing number of experimental evidences show that the second possibility is actually realized in proteins such as triosphosphate isomerase [4] and lysozyme [5]. The study of such conformational fluctuations is hindered by the lack of experimental techniques that yield detailed information on the structural changes involved. On the other hand, numerical methods based on molecular dynamics simulations with atomic resolution can only yield partial views on this dynamic process as the expected time scales ($>10^{-6}$ s) are out of reach of the present day computing capabilities. Let us mention that the use of discontinuous potentials [6] allows dynamics on the millisecond time scale, although these simulations are still limited by the accuracy of ''square-well'' potentials.

The goal of this paper is the introduction of a numerical method that, provided some assumptions are verified, gives access to such a time scale. The underlying idea is quite simple and is based on the experimental observation that the secondary structure of a fluctuating protein is well conserved even in the cases where several conformations are visited. Therefore, it is expected that the addition of constraints that fix, locally, these structures will not significantly perturb the protein dynamics at room temperature and, at the same time, will allow to speed up the conformational exchange by increasing the temperature of the simulation. A similar approach has already been considered for the case of triosphosphate isomerase [7]. The present method can be considered as a generalization of some of the ideas in Ref. [7]. To test the method, we will consider the case of the C-terminal domain of calmodulin, a small protein that displays natural conformational fluctuations, as explained in the following section.

II. A FEW FACTS ABOUT CALMODULIN

Calmodulin (CaM) belongs to the EF-hand family of proteins, the members of which act as Ca^{2+} sensors, coupling transient increases in cellular calcium concentration with a variety of cellular activities [8–10]. These range from cell cycle control to muscle contraction and are involved in diseases such as cancer and Alzheimer's disease [11,12]. Calmodulin is highly conserved (100% identity among vertebrates) and interacts with more than 100 different target proteins, including kinases, phosphatases, cyclases, etc. [9,13]. The crystal structure of Ca^{2+} bound CaM shows a dumbell-shaped structure, with two globular domains (the N and C domains) linked by a long helix. Each of these domains consists of two EF-hand motifs [14,15] and fixes two calcium ions. Upon binding, each domain goes from a ''closed'' conformation [16,17] where the two helices of the EF-hand motifs are almost antiparallel to an ''open'' conformation, where these helices are nearly perpendicular [18,19] (see Fig. 1). One of the effects of this change is the exposure to the solvent of hydrophobic patches that form the binding sites of target proteins [20]. This opening is clear from the x-ray structure, but has been contested on the grounds of both experimental and numerical data for the N-terminal domain. In Ref. [21], it was concluded that the crystal packing stabilizes the exposure of the hydrophobic patches. On the other hand, from the NMR data obtained by Evenäs and co-workers [22,23], the calcium-loaded C-terminal domain

*Electronic address: elez@crpp.u-bordeaux.fr

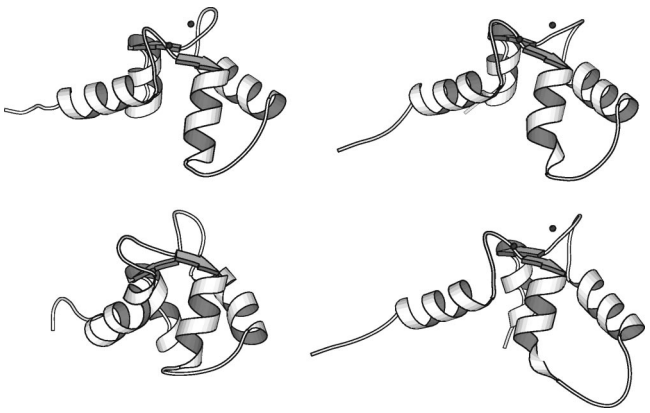

FIG. 1. The conformational changes of the C-terminal domain of calmodulin. Left, as observed through NMR experiments (top, the open calcium-loaded form; bottom, the closed calcium-free form). Right, as observed during our molecular dynamics simulation of the calcium-loaded form (top, before the conformational change takes place, at $t=3{,}6$ ns; bottom, after it occurred, at $t=4$ ns).

seems to expose these hydrophobic patches to solvent. In this sense, the C and N terminals may not play a symmetric role. This is not in contradiction with the experimentally observed fact that the N- and C-terminal domains are largely independent of one another. For instance, the sum of the Ca^{2+} affinities of each separated domain compares well with the sum of the same affinity for the whole protein [24]. In addition, the three-dimensional structure [25] of the recombinant C-terminal fragment M76-K148 (obtained by NMR) is very close to that observed in the CaM crystallographic structure. These observations, put together, justify the study of each individual domain, rather than the full CaM protein.

Recently, a number of studies [22,23] have focused on several mutants (E104Q and E140Q) of the C-terminal fragment M76-K148 (TR2C, ''tryptic fragment 2 in the presence of $CaCl_2$'') and shown that the calcium-loaded fragment undergoes conformational exchanges between several, not clearly defined, substates on a submillisecond time scale. The conformational exchange contributions to the tranverse autorelaxation rate constant correlate well with the chemical shift difference between the apo and Ca^{2+}-saturated states of the wild-type protein, thus indicating that at least part of the substates explored correspond well to the ''closed'' conformation of the TR2C domain. From the same data, it seems clear that the secondary structures of the explored substates are well defined and no unfolding is involved in these conformational changes. As already noticed, similar behavior has been observed for a small number of proteins, such as lysozyme or triosphospate isomerase [7,26], indicating that the conformational changes observed upon substrate binding are, in fact, natural motions of the protein, enhanced by the presence of the substrate.

Several molecular dynamics studies have considered CaM, mostly concerning the relative orientation of the N- and C-terminal domains, or the flexibility of the central helix [27–29]. Two 4 ns trajectories of calmodulin were reported in Ref. [30], where it was found that the C- and N-terminal domains behave mostly like rigid bodies, with fluctuations of interdomain distances within a 7 Å range and of interdomain angles by up to 60°. No open-closed transition of the C-terminal domain was reported. In particular, the backbone rms of the C-terminal domain did not exceed 2.5 Å.

More recently, a 3 ns trajectory [21] of the calcium-loaded N-terminal domain showed a transition from the open to the close configuration. The authors concluded that the N-terminal domain is most of the time in the closed configuration, exhibiting transitions to the open state that favor target binding. In order to further characterize the conformational changes of calcium-loaded TR2C, we introduce in the following section a simulation protocol that allows atomic description and at the same time the observation of millisecond events. Inspired by the fact that the secondary structure of the calcium-loaded TR2C is well defined even in the intermediary states [22,23], we consider molecular dynamics trajectories in which the secondary structure of TR2C is fixed (although some flexibility can be introduced as well) and take into account some modeling of the solvent. Clearly, these simplifications are not enough to reach (with present day computational resources) millisecond-long trajectories. In order to accelerate the exchange processes, we artificially increase the temperature. The hypothesis that the observed conformational changes follow an Arrhenius law allows the extrapolation of the observed dynamics to physiological temperatures. In spite of the high temperatures we consider, TR2C preserves well its tertiary structure (in particular, it remains compact). We do observe conformational changes that bear some resemblance to the open-closed transition. However, in terms of hydrophobic exposed surface, the net effect is the inverse: more hydrophobic patches get exposed to the solvent during these transitions, thus facilitating the interactions with a potential ligand. In this scenario, the net result is analogous to the one suggested in Ref. [21], namely, from time to time, each domain of CaM increases temporarily its hydrophobic surface. The difference between the N- and C-terminal domains would be that the predominant conformation is the closed one for the former and the open one for the later.

III. METHODS

All energy minimization and molecular dynamics calculations were performed with the CHARMM program [31]. The simulations reported below used the standard PARAM19 charges and force constants. We checked that qualitatively similar results were obtained with the all-hydrogen PARAM27 parameter set. Electrostatic interactions with a constant dielectric function ($\epsilon=1$) where smoothly turned off at 12 Å. We used the same cutoff distance for the van der Waals interactions. The nonbonded list was updated by a heuristic procedure (default option in CHARMM: updates are performed only if atoms move by more than half of the width of the buffer region, 1 Å, since the last update).

The initial coordinates for TR2C and the two calcium ions are taken from the crystallographic structure [19] of the calcium-loaded calmodulin (pdb code: 1CLL). Water molecules are taken from a pure water (TIP3) simulation

TABLE I. A description of the substructuring used to discretize TR2C. The left column gives the secondary structure element (helix or loop), the right column contains the corresponding sequence of residues.

Nter loop	MET LYS ASP THR ASP SER
Helix 1	GLU GLU GLU ILE ARG GLU ALA PHE ARG VAL PHE
Loop 1	ASP LYS ASP GLY ASN GLY TYR ILE SER ALA ALA GLU
Helix 2	LEU ARG HIS
Loop 2	VAL MET THR ASN LEU GLY GLU LYS LEU THR ASP
Helix 3	GLU GLU VAL ASP GLU MET ILE ARG GLU
Loop 3	ALA ASP ILE ASP GLY ASP GLY GLN VAL ASN TYR GLU
Helix 4	GLU PHE VAL GLN MET
Cter loop	MET THR ALA

(tip256.crd file in CHARMM). This box was replicated in space to fully hydrate TR2C. The main problem encountered in these simulations was the choice of an appropriate model for the solvent effects. We tried unsuccessfully the primary shell model of Beglov and Roux, and the ACE model [32] as implemented in the CHARMM27 code. In both cases, long enough (typically 5 ns long) standard simulations of the solvent model and TR2C at room temperature lead to unrealistic conformations of the protein, with an rms of the alpha carbons with respect to the crystallographic structure larger than 4 Å. This was not the case with the quartic droplet model, where the water molecules are constrained through a spherical polynomial potential of order four. In our case, water molecules further than 28 Å from the center of the system were deleted, resulting in a set of 2443 solvent molecules. No counterions were added. The resulting structure was minimized and was the initial condition of the dynamic simulations. The alpha-carbon rms deviation from the crystallographic structure was 1.1 Å.

TR2C was simulated using a recently developed method [33] where parts of the protein are considered as partially flexible or even rigid bodies. In all the simulations reported here, the secondary structure elements (four α helices and two short β strands) were considered as rigid bodies (see Table I), as well as the water molecules. Within the loops connecting them, each peptidic plane and the lateral chains of the amino acids were considered as rigid. In other words, within loops, only the relative motions between amino acid side chains and peptidic planes are allowed. We therefore integrated the equations of motion of a total of $83+2443$ rigid bodies, using the standard Lobatto algorithm with a time step of 5 fs. The temperature was controlled through a Langevin bath, with a coupling constant of $\beta=1$ ps^{-1}. This coupling is not intended to mimic the coupling to any solvent, but rather used as a straightforward temperature-control algorithm that yields the Boltzmann statistics. The resulting algorithm is close to the MBO(N)D algorithm [34] implemented in CHARMM, although it also works with parallel architectures and also differs at the level of the dynamical equations, as explained in Ref. [33]. For the sake of completeness, we give in the Appendix the evolution equations of the present method.

IV. RESULTS

As stated in the Introduction, our goal is to simulate the dynamics of TR2C at the millisecond time scale. The level of details we are seeking here prevents us to reach such a goal using today's computational resources. There are several methods that can be used to improve the sampling of the conformational space. Most of them rely on biased simulations (umbrella sampling method), where the molecule is constrained to sample uniformly regions defined by one or several coordinates [35]. This increases the number of transitions between the different important conformations, and thereby reduces the simulation time needed to characterize the system. In this paper, we will be using a different point of view, where we rely only on the increase in temperature to accelerate the conformational sampling. From experimental data, we expect that relevant dynamical events do not disturb much the local secondary structure, which will be kept fixed subsequently. In the following, we will focus on high temperature (~ 800°K) simulations, and argue that, below some critical temperature, the observed behavior can be extrapolated to physiological temperatures. For $T=800$ K, we considered three different trajectories with the same initial coordinates, but several initial velocities, and we explicitly checked that the conclusions given below are independent of the initial conditions.

In Fig. 2, we compare the time evolution of, respectively, (a) the angle between helices 1 and 2, (b) between helices 3 and 4, (c) the rms deviation from the minimized structure restricted to the α carbons, (d) the accessible surface restricted to hydrophobic residues (ACSH) and (e) the difference δACS between the total accessible surface and ACSH, as observed in a 5 ns trajectory at temperature $T=800$ K. Most of the time the molecule stays close to the minimized structure (the rms is less than 2.5 Å). The fact that the molecule does not unfold at this temperature is certainly due to the constraints imposed on the secondary structures. In other words, the usual temperature-induced unfolding of a protein

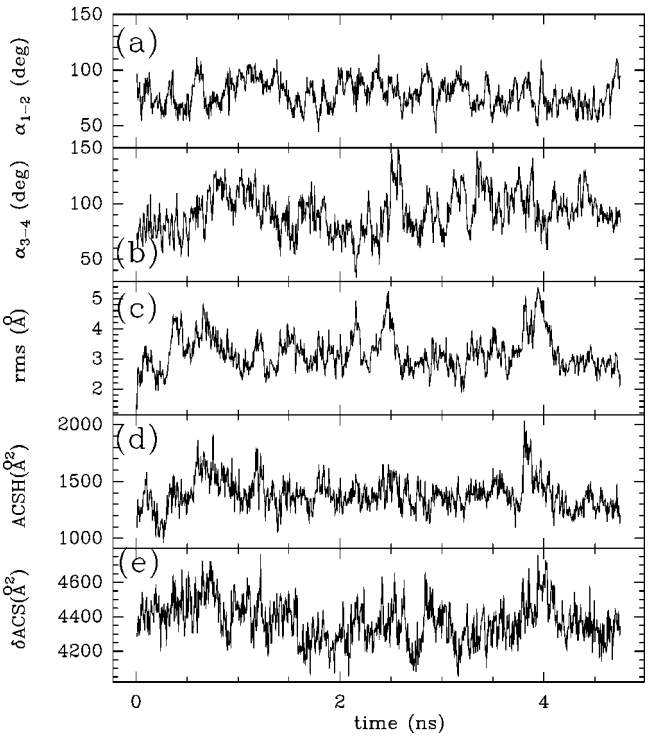

FIG. 2. The 800 K trajectory: (a) angle α_{1-2} between helices 1 and 2, (b) angle α_{3-4} between helices 3 and 4, (c) rms (Å) deviation from the minimized structure (only the α carbons are taken into account), (d) accessible surface restricted to hydrophobic residues ACSH (Å^2), and (e) the difference δACS between the total accessible surface ACST and ACSH (Å^2).

proceeds locally, through the breaking of hydrogen bonds, while the unfolding of a constrained molecule is much more difficult, since it implies the breaking of the global packing.

From Fig. 2, it is clear that the molecule can adopt a whole range of alternative conformations with rms around 4 Å or even higher, and significantly populated as well. Another striking feature of the simulation is the clear correlation between the bursts in rms and the increase of the accessible surface around hydrophobic residues. As Fig. 2(e) shows, the fluctuations are also significantly larger for hydrophobic than for hydrophilic residues.

Let us recall that one of the differences between the open and closed forms of TR2C is the fact that the angle between helices 1 and 2 (α_{1-2}) or 3 and 4 (α_{3-4}) increases from $\sim$90° (open) to $\sim$120° (closed) with an overall 4–5 Å rms between the two forms. This is coupled to a burying of hydrophobic residues. Although some correlation between the burst in rms and the increase of the angles $\alpha_{1,2}$ and $\alpha_{3,4}$ exists, it is clear from the above remarks on the ACSH that the transitions found in our simulations are not of the open-closed type.

Figure 3 shows that the distance from the calcium ion to the first EF hand, although it fluctuates more than in standard room temperature molecular dynamics simulations, remains mostly stationary, indicating that the transitions are not related to the unbinding of the calcium ion. Eight oxygens are found in the coordination sphere of the calcium ion that is less than 3.0 Å away, while the bidentate coordination of the

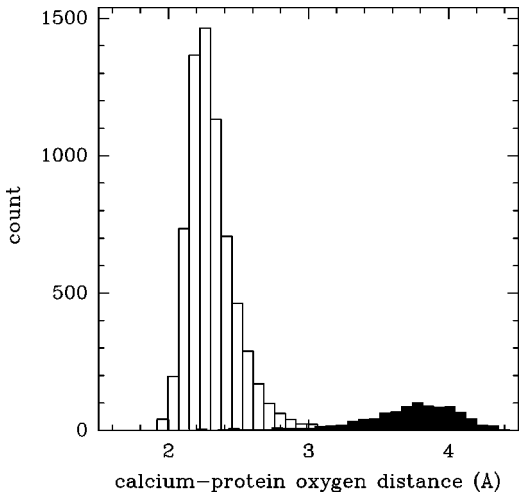

FIG. 3. Behavior of the first calcium-binding site during the 800 K simulation. The aspartate in the first relative position in the EF loop (Asp 93) remains monodentate, one of its two carboxylate oxygens pointing always away from the calcium (black histogram).

glutamate in the 12th relative position in the EF loop (Glu104) is maintained. Note that in previous simulations [36] performed with similar parameter sets, eight oxygen atoms were also systematically observed in the calcium coordination sphere, instead of seven in all known calcium-loaded EF hands. It seems fair to wonder why calcium does not dissociate at $T=800$ K. According to the transition state theory, when the activation energy is much larger than the thermal energy, a rate constant can be well approximated by [37]

$$k=\kappa\nu_c\exp^{-\Delta G/k_BT}, \tag{1}$$

where ν_c is the harmonic frequency of the reactant along the reaction coordinate at the bottom of the well, ΔG is the activation energy, k_BT is the thermal energy and κ, the transmission coefficient, is smaller than one when tuneling effects are negligible. Thus, setting $\kappa=1$ and $\nu_c=k_BT/h$, where h is the Planck's constant, that is, setting the harmonic frequency to the highest value expected at a given temperature, yields a good estimate for the lower bound of ΔG, once the rate constant is known.

Since k_{off}, the calcium-off rate constant of the C-terminal domain of calmodulin, is 37 s^{-1} at room temperature [38], the corresponding activation energy, ΔG_{off}, is ΔG_{off} $\sim$15.5 kcal/mole, and the order of the magnitude of the calcium-off rate constant is $\sim 10^9$ s^{-1} at 800 K, that is, calcium ions are indeed expected to remain bound to the C-terminal domain of calmodulin at least 1 ns at 800 K. The nonobservation of a calcium dissociation during the 800 K trajectory is therefore not really incompatible with the available experimental data, however it could also reflect a poor parametrization of the calcium-protein interactions.

It is also of interest to compare the accessible surface values observed along this trajectory and the same values computed from the NMR structures. In the apo, calcium-free state (pdb code 1cmf), the accessible surface corresponding to the hydrophobic residues (all the residues) is 2600 Å^2

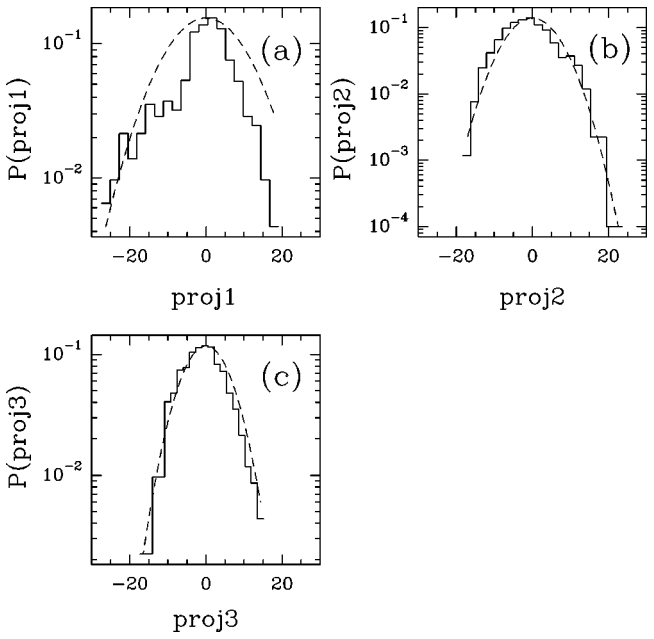

FIG. 4. Probability density of the (a) first, (b) second, and (c) third principal components, as computed from the 800 K trajectory. For each distribution, the dashed line corresponds to a Gaussian distribution with the same mean and variance.

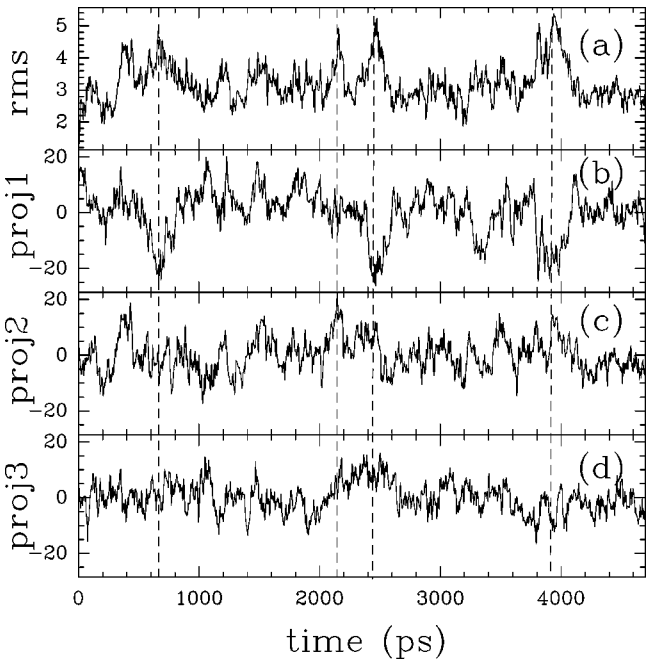

FIG. 5. Principal component analysis of the 800 K trajectory. (a) The rms (Å) deviation from the minimized structure; (b),(c),(d) time series of the first, second, and third principal component.

(4900 $Å^2$). In the holo, calcium-loaded state (pdb code 1cmg), the corresponding values are 3000 and 5540 $Å^2$. Notice that the crystallographic structure of TR2C obtained from that of the calcium-loaded CaM yields 2965 and 5160 $Å^2$, indicating that the crystal packing and may be the influence of the N-terminal part leave almost intact the hydrophobic accessible surface, but sligthly decreases the total surface. Our simulation indicates that the average ACSH is that found by NMR, although large fluctuations, both positive and negative, are possible.

In order to further analyze the bursting phenomenon, we did perform a principal component analysis [39] of the trajectory at 800 K. In Fig. 4 are shown the histograms of the values taken by the first three principal components. The log-lin representation has been used in order to emphasize the Gaussian (or not) character of these projections. The logarithm of the second and third projections clearly look much parabolic than the first one, indicating that these components can be well approximated as random Gaussian variables. However, solely from this data, it is difficult to conclude about the existence of two well-defined states, as hypothetized in Ref. [22]. Such a scenario would imply the existence of two maxima in some of these histograms with a gap between the maxima significantly larger than the variance of the other components. The histograms in Fig. 4 show that this is not the case. However, a more careful examination of these projections shows that the two-state hypothesis holds as a first approximation. In fact, as can be seen in Fig. 5, most of the rms bursts are well correlated with the extreme negative values of the first principal component, although not all the bursts can be accounted for in this way. For instance, the burst occurring at around 2.4 ns is clearly related to a large fluctuation of the first component, whereas the one occurring at around 2.2 ns is more related to a sudden increase in the second component. The deviation from gaussianity of the first principal component is therefore not an artifact of poor sampling, but rather a sign that this component displays coherent motions.

As stated in the Introduction, one of the initial goals was to compare our results against the NMR data of Ref. [22]. We have emphasized the fact that the two-state model for the bursting transition is not incompatible with the results from the principal component analysis. It seems therefore plausible to apply an Arrhenius law to fit the dependence of the average time $\langle \tau \rangle$ spent in the open state (i.e., the ground state in the presence of calcium) versus the temperature. As a measure of this average time, we consider the time intervals longer than 25 ps where the rms with respect to the minimized structure is less than 3 Å. The rms threshold value 3 Å was found to be a good compromise between a sensible definition of a state being ''close'' to the ground state and obtaining meaningful statistics. In Fig. 6 we plotted the average

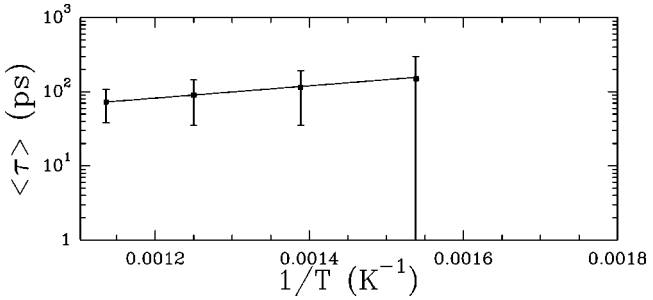

FIG. 6. Average time life $\langle \tau \rangle$ spent in the open state as a function of the inverse temperature.

length of these intervals as a function of the inverse of the temperature. The length of the trajectories used to estimate $\langle\tau\rangle$ was 5 ns. Obviously, the lower the temperature, the larger the error in this estimates is. Nevertheless, these data suggest an extrapolated average life time in the open state of around 30 ns, again compatible with the NMR data of Ref. [22].

In conclusion, our analysis strongly relies on the observed fact that, at $T=300$ K, the secondary structures are close to rigid (in the nanosecond range from the molecular dynamics simulations reported in the Introduction) or at least well conserved (from the NMR experiments of Evenas *et al.*). Formally, imposing constraints on the secondary structure amounts to replace the Hamiltonian H associated with the force field by a constrained Hamiltonian H_c, including the appropriate Lagrange factors. Our main hypothesis is then that the system protein+solvent visits in a similar way the conformational space under the action of H or H_c: the probability $p(C,T,H)$ of obtaining a given conformation in the unconstrained model is close to the same probability computed with H_c, when $T=300$ K. Now, the Arrhenius extrapolation based on trajectories generated with H_c at high values of T yields an estimate of the ratio between $p(C',T,H_c)$ and $p(C,T,H_c)$, the probability to be in two conformations C and C'. According to our hypothesis, $p(C',T=300,H_c)/p(C,T=300,H_c)\sim p(C',T=300,H)/p(C,T=300,H)$. Of course, this is only a heuristic argument: making it more rigorous is, in our opinion, extremely difficult and only the confrontation with new experimental data will give additional support to the approach.

V. CONCLUSIONS

Molecular dynamics simulations with local constraints were used to characterize the intramolecular dynamics of the calcium-loaded C-terminal fragment of calmodulin. High temperature trajectories suggest that TR2C explores regions of the conformational space located quite far in the rms sense from the ground, open state. It seems clear that these regions correspond to an increase of the accessible surface around hydrophobic residues. However, the connection with the open-closed transition is not completely evident, in particular the observed variations of the angles between each pair of helices do not correspond to those expected in that transition. Extrapolation to physiological temperatures of the measured transition rates tend to agree with the NMR data of Ref. [22], even though longer trajectories will be needed to give additional support to this conclusion. Interestingly, the principal component analysis of the high temperature trajectories clearly differentiates the dynamic role of the first principal component from the others, the former being non-Gaussian. This is probably not at odds with the conclusions of Ref. [22], who suggested that the conformational exchange is not well represented by a single, global process. As shown in Fig. 3, the amplitude of the fluctuations associated with the second and third principal components are comparable to that of the first. In this sense, even if the first component is singled out as a non-Gaussian variable, its own dynamics is significantly perturbed by the noise of the other components. Let us emphasize that the Arrhenius extrapolation is only the simplest hypothesis that could be used to fit our data. More complex scenarios such as the "rugged energy landscape" advocated in Ref. [22], where the temperature dependence of the conformational exchange constant includes $1/T^2$ terms, are probably more realistic. However, taking into account the quality of the fit obtained with the simple Arrhenius law, we doubt this would bring additional improvement. As stressed before, the success of the Arrhenius model is probably due to the constraints applied to the secondary structure that probably smooth the conformational space sampled by the protein. It also will be of interest to quantify the effect of the rigid-body constraints imposed on the secondary structures. For instance, one might wonder whether these constraints could create artificial barriers for tertiary structure rearrangements. The above extrapolation to physiological temperature shows that these are probably not very important. Simulations using the methods in Ref. [33] are under way to further quantify this point.

ACKNOWLEDGMENTS

The authors acknowledge computer resources from Pôle M3PEC (U. Bordeaux I), and from IDRIS (under Contract No. 990871).

APPENDIX: EVOLUTION EQUATIONS

The purpose of this section is the derivation of the evolution equations that we used to generate the high temperature trajectories. We assume that, for each body (either parts of the protein or solvent molecules), the position of each atom i with mass m_i can be written as

$$X_i(t)=T(t)+R(t)\left(X_i^0+\frac{1}{\sqrt{m_i}}\sum_{n=1}^{N'}c_n(t)\varphi_n(i)\right), \quad \text{(A1)}$$

where $R(t)$ $[T(t)]$ is a rotation matrix (a translation vector), both being time dependent. X^0 is a reference configuration (for instance, some configuration of minimum energy) and the vectors φ_n account for some flexibility of the body. In previous work [33], the vectors φ_n were normal modes. In the present work, this contribution is missing, still we include it here for the sake of completeness.

Here, we derive the evolution equations for the amplitudes $c_n(t)$, rotation $R(t)$ and translation $T(t)$. In this appendix, we will use the notation $\dot{a}\equiv da/dt$. We start from the expression (A1) and derive twice with respect to time, obtaining

$$\Gamma_i-\Gamma_g=\dot{\Omega}\wedge x_i+\Omega\wedge(\Omega\wedge x_i)+2\Omega\wedge R\left[\frac{1}{\sqrt{m_i}}\sum_n d_n\varphi_n(i)\right]$$
$$+R\frac{1}{\sqrt{m_i}}\sum_n \ddot{d}_n\varphi_n(i), \quad \text{(A2)}$$

where

$$\dot{c_n} \equiv d_n, \qquad x_i \equiv X_i - X_g, \qquad \Gamma_i = \ddot{x}_i, \qquad \Gamma_g = \ddot{X}_g,$$

X_g being the position of the center of mass, and Ω the angular velocity. Equation (A2) is easily obtained by repeatedly applying the relation

$$\dot{A} = \dot{A}_{moving} + \Omega \wedge A$$

relating the time derivatives of a time dependent vector A in a fixed reference ($\dot{A}$) and a moving reference $\dot{A}_{moving}$. We rewrite Eq. (A2) as

$$0 = \ddot{\Sigma}_i - Z_i + \dot{\Omega} \wedge \sqrt{m_i} x_i, \tag{A3}$$

with

$$Z_i = \sqrt{m_i}(\Gamma_i - \Gamma_g) - \sqrt{m_i}\Omega \wedge (\Omega \wedge x_i) - 2\Omega \wedge R \sum_n d_n \varphi_n(i),$$

and

$$\ddot{\Sigma}_i = R \sum_n \ddot{c}_n \varphi_n(i).$$

A simple counting shows that the system (A3) is ill posed, as it has less unknowns $(N'+3)$ than equations $(3N_a)$. A simple way out is to define the unknowns $\ddot{c}_n$ and $\dot{\Omega}$ as given by the least-squares solution of Eq. (A2), that is, they minimize the quantity

$$\frac{1}{2} \sum_i (-Z_i + \ddot{\Sigma}_i + \dot{\Omega} \wedge \sqrt{m_i} x_i)^2.$$

It is then straightforward to obtain

$$\mathbf{I}\dot{\Omega} = -\sum_i \sqrt{m_i}(Z_i \wedge x_i) + \sum_i \left[\sqrt{m_i} \sum_n \ddot{c}_n R \varphi_n(i) \wedge x_i \right], \tag{A4}$$

$$\ddot{c}_n = \sum_i [Z_i R \varphi_n(i)] - \dot{\Omega} \widetilde{\varphi_n}, \tag{A5}$$

where

$$\widetilde{\varphi}_n \equiv \sum_i \sqrt{m_i} x_i \wedge R \varphi_n(i),$$

and $\mathbf{I}$ is the instantaneous inertia matrix, i.e.,

$$\mathbf{I}_{11} = \sum_i x_i^2 x_i^2 + x_i^3 x_i^3, \qquad \mathbf{I}_{12} = -\sum_i x_i^1 x_i^2, \ldots \tag{A6}$$

with $x_i = (x_i^1, x_i^2, x_i^3)$.

The main difference with respect to the MBO(N)D method [34] is the choice of the rotation matrix. In MBO(N)D, $R(t)$ defines a rotation with respect to the previous body (the protein is considered as a chain of flexible bodies), in our approach $R(t)$ defines an absolute rotation, with no reference to other bodies. Accordingly, in MBO(N)D, the evolution equations of the different parts are coupled, whereas in our method, they are uncoupled. The other main difference is the use of rescaled short range forces, as explained in Ref. [33]. This difference is crucial without substructuring.

[1] S. Remington, G. Weigand, and R. Huber, J. Mol. Biol. **158**, 111 (1982).
[2] R. Hubert and W. Bennett, Biopolymers **22**, 261 (1983).
[3] G. Wiegand and S. Remington, Annu. Rev. Biophys. Biophys. Chem. **15**, 97 (1986).
[4] J.C. Williams and A.E. McDermott, Biochemistry **34**, 8309 (1995).
[5] H.S. Mchaourab, K.J. Oh, C.J. Fang, and W.L. Hubbell, Biochemistry **36**, 307 (1997).
[6] A. Voegler Smith and C.K. Hall, Proteins **44**, 344 (2001); **44**, 376 (2001).
[7] P. Derreumaux and T. Schlick, Biophys. J. **74**, 72 (1998).
[8] H.J. Vogel, Biochem. Cell Biol. **72**, 357 (1994).
[9] A. Crivici and M. Ikura, Annu. Rev. Biophys. Biomol. Struct. **24**, 85 (1995).
[10] J. Evenäs, A. Malmendal, and S. Forsén, Curr. Opin. Chem. Biol. **2**, 293 (1998).
[11] P. Vito, E. Lacana, and L.D. Adamio, Science **271**, 521 (1996).
[12] W. Hait and J. Lazo, J. Clin. Oncol. **4**, 994 (1986).
[13] M. R. Celio, T. Pauls, and B. Schwaller, *Guidebook to the Calcium-Binding Proteins* (Oxford University Press, Oxford, 1996).
[14] R.H. Kretsinger and C.E. Nockolds, J. Biol. Chem. **248**, 3313 (1973).
[15] R.H. Kretsinger, Cold Spring Harbor Symp. Quant. Biol. **52**, 499 (1987).
[16] H. Kuboniwa, N. Tjandra, S. Grzesiek, H. Ren, C.B. Klee, and A. Bax, Nat. Struct. Biol. **2**, 768 (1995).
[17] M. Zhang, T. Tanaka, and M. Ikura, Nat. Struct. Biol. **2**, 758 (1995).
[18] Y.S. Babu, C.E. Bugg, and W.J. Cook, J. Mol. Biol. **204**, 191 (1988).
[19] R. Chattopadhayaya, W.E. Meador, A.R. Means, and F.A. Quiocho, J. Mol. Biol. **228**, 1177 (1992).
[20] M. Ikura, G.M. Clore, A.M. Gronenborn, G. Zhu, and A. Bax, Science **25**, 632 (1992).
[21] D. Vigil, S.C. Gallagher, J. Trewhella, and A.E. García, Biophys. J. **80**, 2082 (2001).
[22] J. Evenäs, S. Forsén, A. Malmendal, and M. Akke, J. Mol. Biol. **289**, 603 (1999).
[23] A. Malmendal, J. Evenäs, S. Forsén, and M. Akke, J. Mol. Biol. **293**, 883 (1999).
[24] S. Linse, A. Helmersson, and S. Forsén, J. Biol. Chem. **266**, 8050 (1991).
[25] B.E. Finn, J. Evenäs, T. Drakenberg, J.P. Waltho, E. Thulin, and S. Forsén, Nat. Struct. Biol. **2**, 777 (1995).
[26] J.C. Williams and A.E. McDermott, Biochemistry **34**, 8309 (1995).

[27] J.-L. Pascual-Ahuir, E. Mehler, and H. Weinstein, Mol. Eng. **1**, 231 (1991).
[28] H. Weinstein and E.L. Mehler, Annu. Rev. Physiol. **56**, 213 (1994).
[29] W. Wriggers, E. Mehler, F. Petici, H. Weinstein, and K. Schulten, Biophys. J. **74**, 1622 (1998).
[30] C. Yang, G.S. Jas, and K. Kuczera, J. Biomol. Struct. Dyn. **19**, 247 (2001).
[31] B.R. Brooks, E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[32] M. Schaefer and M. Karplus, J. Phys. Chem. **100**, 1578 (1996).
[33] J. Elezgaray, G. Marcou, and Y.H. Sanejouand, Theor. Chem. Acc. **106**, 62 (2001).
[34] H.M. Chun, C.E. Padilla, D.N. Chin, M. Watanabe, V.I. Karlov, H.E. Alper, K. Soosar, K.B. Blair, O.M. Becker, L.S. Caves, R. Nagle, D.N. Haney, and B.N. Farmer, J. Comput. Chem. **21**, 159 (2000).
[35] C. Bartels, M. Schaefer, and M. Karplus, J. Chem. Phys. **111**, 8048 (1999).
[36] D. Allouche, J. Parello, and Y.H. Sanejouand, J. Mol. Biol. **285**, 857 (1999).
[37] J. E. Straub, *Computational Biochemistry and Biophysics* (Marcel Dekker, New York, 2001), pp. 199–220.
[38] S.R. Martin, A. Teleman, P.M. Bayley, T. Drakenberg, and S. Forsen, Eur. J. Biochem. **151**, 543 (1985).
[39] A. Kitao, F. Hirata, and N. Go, J. Chem. Phys. **158**, 447 (1991).