PROTEINS: Structure, Function, and Genetics 50:648–656 (2003)

Thermal Unfolding Simulations of Apo-Calmodulin Using Leap-Dynamics

Jens Kleinjung,[1*] Franca Fraternali,[1] Stephen R. Martin,[2] and Peter M. Bayley[2]
[1]*Division of Mathematical Biology, National Institute for Medical Research, London, United Kingdom*
[2]*Division of Physical Biochemistry, National Institute for Medical Research, London, United Kingdom*

ABSTRACT **The simulation method leap-dynamics (LD) has been applied to protein thermal unfolding simulations to investigate domain-specific unfolding behavior. Thermal unfolding simulations of the 148-residue protein apo-calmodulin with implicit solvent were performed at temperatures 290 K, 325 K, and 360 K and compared with the corresponding molecular dynamics trajectories in terms of a number of calculated conformational parameters. The main experimental results of unfolding are reproduced in showing the lower stability of the C-domain: at 290 K, both the N- and C-domains are essentially stable; at 325 K, the C-domain unfolds, whereas the N-domain remains folded; and at 360 K, both domains unfold extensively. This behavior could not be reproduced by molecular dynamics simulations alone under the same conditions. These results show an encouraging degree of convergence between experiment and LD simulation. The simulations are able to describe the overall plasticity of the apo-calmodulin structure and to reveal details such as reversible folding/unfolding events within single helices. The results show that by using the combined application of a fast and efficient sampling routine with a detailed molecular dynamics force field, unfolding simulations of proteins at atomic resolution are within the scope of current computational power. Proteins 2003;50:648–656.** © 2003 Wiley-Liss, Inc.

Key words: molecular dynamics; far-UVCD; implicit solvent; conformational transitions; enhanced sampling

INTRODUCTION

It is generally accepted that evolution has selected for those protein sequences that fold rapidly to a stable and functional structure.[1,2] The process of protein folding to the native structure relies on a dynamic self-recognition mechanism that is driven by the balance of intramolecular and solvation forces. Although folding is a highly complicated process, the main stabilizing force of the folded conformation has been attributed to the burial of hydrophobic side-chains in a core, thereby exchanging unfavorable hydrophobic-solvent interactions for more favorable combinations of hydrophobic-hydrophobic and solvent-solvent interactions.[3,4] Reversible thermal unfolding/refolding is a common experimental approach used in the study of protein folding events. As the temperature rises, the increase in kinetic energy and chain entropy overcomes the stabilizing intramolecular forces and drives the protein toward random coil configurations.[5,6] Although considerable progress has been made toward a detailed understanding of protein folding, many aspects are still not fully understood; most important is the exact balance of the many energetic/entropic terms and the accessibility of conformational space for an ensemble of conformations.[7–9] However, such questions have mainly been addressed in the context of peptides and small proteins undergoing reversible folding/unfolding processes, using a number of theoretical and experimental approaches. As a theoretical approach, molecular dynamics (MD) allows simulations of the internal motions of proteins at the atomic level and can provide a detailed model description of the underlying dynamic processes of proteins and also for their conformational transitions.[10–17]

Protein dynamics involving conformational equilibria and folding processes occur on a wide variety of timescales from picoseconds to minutes.[18] Current computational resources and force-field accuracy[19] allow for sampling of folding/refolding events of small to medium sized peptides by means of MD.[20] However, because of their larger size, proteins generate a far more severe sampling problem owing to the vast number of degrees of freedom and a large number of energy barriers above k_BT (where k_B is the Boltzmann constant and T is the temperature).[13,21,22] The strength of MD lies in the energy conservation of the algorithm. The generated ensemble contains only conformations on or close to a constant energy hypersurface. However, energy conservation can trap systems in local energy minima and significant conformational changes may be missed. Thus, the main reason for inefficient sampling in MD is the relatively small conformational and energetic fluctuations that occur naturally during afford-

Abbreviations: apo-calmodulin, apo (Ca^{2+}-free) calmodulin; EM, energy minimization; far-UVCD, far ultraviolet circular dichroism; LD, leap-dynamics; MD, molecular dynamics; RMSD, root-mean-square deviation.

*Correspond to: Jens Kleinjung, Division of Mathematical Biology, National Institute for Medical Research, Mill Hill London, United Kingdom NW7 1AA. E-mail: jkleinj@nimr.mrc.ac.uk

Received 29 May 2002; Accepted 1 October 2002

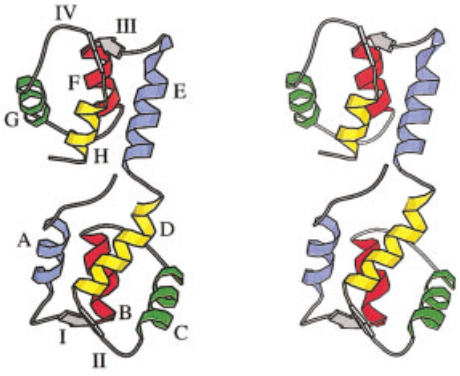


Fig. 1. Stereo ribbon representation of apo-calmodulin (PDB entry 1CFD).[31] N-domain (helices A–D) and C-domain (helices E–H) form a four-helix bundle stabilized by short antiparallel β-strands between the (empty) Ca^{2+}-binding loops (sites I–IV). The color scheme illustrates the relationship of helices located in N- and C-domain.

able simulation times (i.e., in the range of nanoseconds to microseconds).

Alternative sampling strategies for improved sampling efficiency have been proposed.[23–26] We have described a novel approach to molecular simulations called leap-dynamics (LD).[27] Several all-α proteins have been studied by unfolding (e.g., ribonuclease H,[28] monomeric λ repressor,[29] and engrailed homeodomain protein).[30] Apo-calmodulin was chosen for this study, because of its interesting property that it is composed of two similar domains that show different unfolding behavior. Apo-calmodulin is a two-domain protein,[31–33] in which each domain consists of a similar four-helix bundle stabilized by a short β-strand between two antiparallel loop sequences and with a flexible linker sequence between the N- and C-domains (Fig. 1). Structural and functional aspects of calmodulin have been extensively reviewed,[34,35] and several MD studies have been performed on the folded apo-calmodulin domains,[36–39] but unfolding simulations have so far not been attempted. An important characteristic of the calmodulin structure is its functionally relevant plasticity.[40,41] The two domains of calmodulin have 69% sequence identity and adopt an identical fold, but they show distinct temperature unfolding behavior. Spectroscopic and calorimetric studies reveal transition midpoints (T_m) of approximately 55°C (328 K) (N-domain) and 42°C (315 K) (C-domain).[42–45] The low-transition temperatures make apo-calmodulin an interesting subject for thermal unfolding studies, because this temperature range is readily accessible to standard CD and NMR studies without the need for the addition of chemical denaturants, pH changes, or mutations. The different transition temperatures of the domains pose a specific criterion for the predictive capability of molecular simulations.

RESULTS

Unfolding Trajectories

Four independent LD trajectories of apo-calmodulin at three different temperatures (290 K, 325 K, and 360 K), each comprising 3 ns of MD were performed for simulating

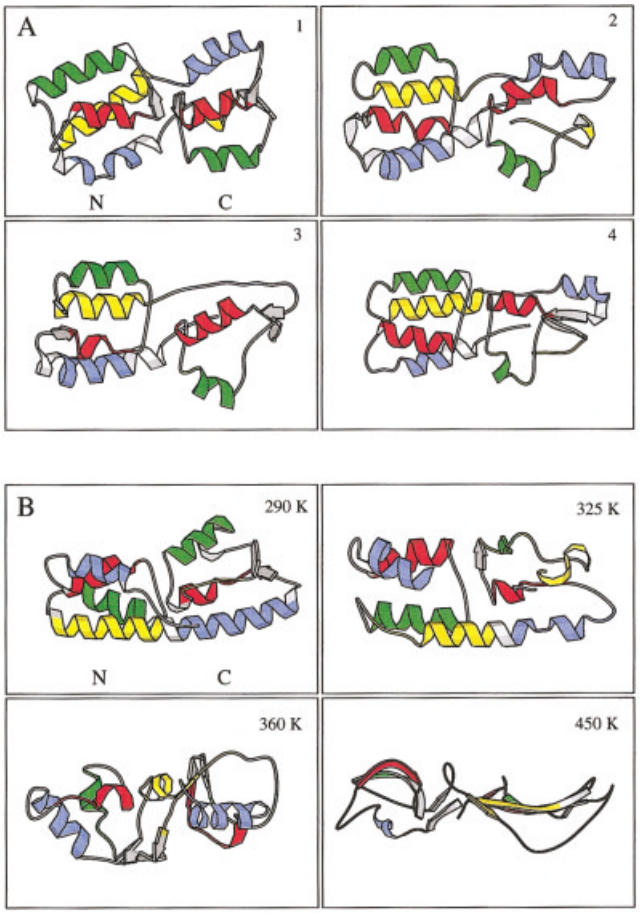


Fig. 2. **A**: Conformers of a LD trajectory of apo-calmodulin at 325 K. Note the distinctive unfolding of the C-domain. **B**: Final conformers of the LD trajectories at 290 K, 325 K, 360 K, and 450 K.

the differential unfolding of the N- and C-domain of apo-calmodulin, as described in detail in Materials and Methods. MD simulations of 3 ns length at these temperatures were run under the same conditions for comparison. The simulation temperatures were chosen to cover the temperature range of far-UVCD thermal unfolding experiments presented here. An additional simulation at 450 K was performed to explore sampling under extreme unfolding conditions. An illustration of some typical conformers along a single 325 K LD trajectory is given in Figure 2(A). Figure 2(B) shows the final conformers at the end of trajectories at different temperatures, showing the relative stability of both domains at 290 K, the distinctive unfolding of the C-domain at 325 K, and unfolding of both domains at 360 K and 450 K. The variation in sampling is illustrated in Figure 3, where the helical content of apo-calmodulin in two of the four trajectories at 325 K and 360 K is plotted. The trend of the simulations is similar, but there is considerable local variation between the two LD trajectories.

The unfolding of apo-calmodulin is monitored here by four properties: (i) helical content as number of residues in helix conformation (Fig. 4, top); (ii) number of effective contacts as conserved contacts relative to a reference state

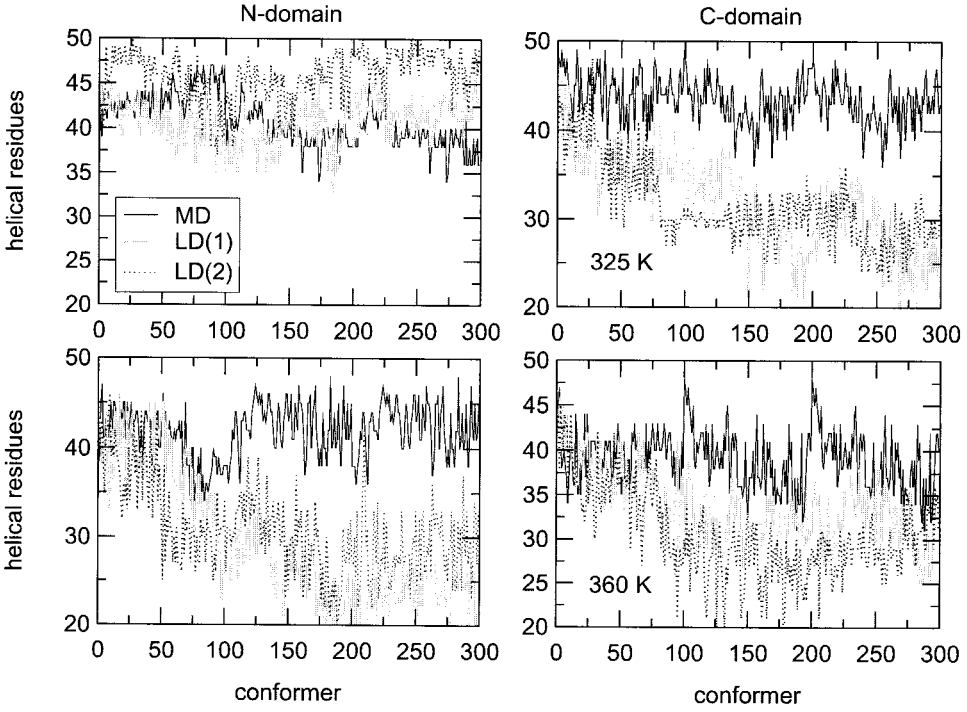


Fig. 3. Two example LD trajectories (dotted, gray) of apo-calmodulin unfolding at 325 K (**top**) and 360 K (**bottom**) represented by the helical content. The trajectory of the MD simulations under the same conditions (black) reveal insufficient sampling, showing a stable C-domain at 325 K and a stable N-domain at 360 K.

(Fig. 4, bottom); (iii) solvent-accessible surface area (SASA) (Fig. 5, top); and (iv) the RMSD relative to the starting conformation (Fig. 5, bottom). Because our main interest was to test whether LD is capable of characterizing domain-specific unfolding behavior, trajectories have been plotted separately on left panels for the N-domain and on right panels for the C-domain. It should be stressed that the plotted curves for a given temperature are compiled by averaging four independent LD trajectories. Because the secondary structure of apo-calmodulin consists mainly of α-helices, the helical content is a sensitive measure of the unfolding process. At 290 K, the N-domain equilibrates in LD simulations at a helical content of 43 residues, compared to 39 in the MD simulation (Fig. 4, top and Table I). The more flexible C-domain loses about 3 helical residues compared to the MD simulation and equilibrates at a helical content of 32 residues. At 325 K, the N-domain remains largely intact, although it does show a loss of about 2 helical residues relative to the equilibrium value. Partial unfolding of the C-domain, which is above the thermal transition point of the C-domain, is apparent in a loss of 9 helical residues. Extension of the trajectories by 100 additional conformations confirmed that the end point had really been reached (data not shown). At 360 K, above the thermal transition point of both domains, the N-domain unfolds with an additional loss of about 12 residues in helical conformation, and the helical content of the C-domain diminishes by another 3 residues.

The number of atom-atom contacts is a measure of local conformational changes, which is expressed as the "effective contacts" along the trajectory (Fig. 4, bottom) that are in common with those found under equilibrium conditions; here we have chosen the last 100 conformations of the 290 K reference MD simulation. The loss of effective contacts generally reflects changes of the interaction pattern within the structure. The behavior is very similar to that observed in helical content. The N-domain retains more contacts at 290 K and 325 K and it unfolds at 360 K, whereas the C-domain already unfolds at 325 K. The absolute number of effective contacts lost is larger in the N-domain because its core contains more residues than that of the C-domain.

Solvent accessibility has been evaluated, because it is directly coupled to the solvation energy of the molecule. During unfolding, buried and predominantly hydrophobic surface elements become exposed to the solvent. At 325 K, hydrophobic interactions seem to keep most of the core residues buried despite considerable loss of secondary structure content (see above). A significant increase in solvent accessible surface occurs only at the end of the 360 K trajectory (Fig. 5, top). Although the change of exposed surface area is limited, the contribution of the implicit solvation term to the overall forces is sufficient to support the unfolding at the experimental temperatures.

Backbone RMSD values reflect the spatial deviation of the trajectory conformations from the starting conformation. The domains equilibrate at an RMSD value of 2.7 Å (N-domain) and 2.5 Å (C-domain) from the starting conformation. These are comparatively high values, which reflect the high degree of flexibility in apo-calmodulin. Motions of helices relative to each other account for a greater part of the RMSD value. However, unfolding is very distinct in the C-domain at 325 K and 360 K with

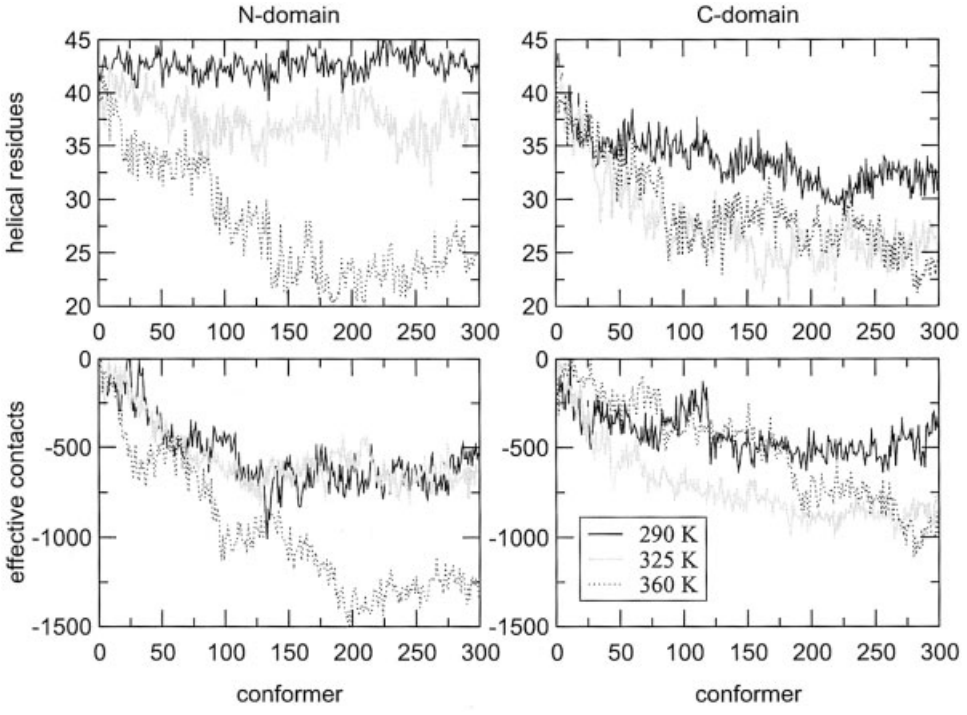


Fig. 4. Trajectories of helix content (**top**) and effective contacts (**bottom**) of apo-calmodulin at 290 K (black), 325 K (gray), and 360 K (dotted) generated by LD. Curves for N-domain (left) and C-domain (right) are presented separately for clarity. The helix content of the N-domain is roughly constant at 290 K and 325 K, but its unfolding at 360 K is clearly visible with a loss of 10–20 helical residues. The folded C-domain contains about 32 helical residues, and unfolding can be seen at 325 K and 360 K with a concomitant loss of about 10 helical residues. The behavior of the effective contacts is similar to that of the helical residues. The N-domain unfolds visibly at 360 K, whereas the C-domain unfolds already at 325 K. Note that the abscissa are in units of conformer numbers rather than simulated time because the leaps in LD sampling lead to a discontinuous time domain.

RMSD values amounting to about 4 Å. The N-domain adopts intermediate RMSD values around 3.4 Å at 325 K, but above the thermal transition point at 360 K unfolding leads to an average RMSD of 4.4 Å.

Sampling Performance

An illustration of the sampling performance of LD is given in Figure 3. Two example trajectories of the LD simulations at 325 K (top) and 360 K (bottom) are represented by the the helical content of apo-calmodulin (gray and dotted). The MD trajectories performed under the same conditions are plotted for comparison (black lines). There are considerable differences between the LD trajectories at several stages of the unfolding process, indicating that the stochastic nature of LD explores multiple pathways toward the unfolded state, nevertheless eventually yielding comparable degrees of unfolding. The trajectories converge at conformers with similar helical content, which suggests that the final state is roughly independent of the unfolding pathway as expected. However, the different sampling behavior also necessitates simulating several LD trajectories and to average them to generate a representative result (as in Fig. 4 and 5). The MD trajectories imply a stable C-domain at 325 K and a stable N-domain at 360 K, both in contradiction to experimental data. We note, however, that the results of a recent temperature-jump study of apo-calmodulin allow us to estimate the half-time for the unfolding of the C-domain to be of the order of 250 ns at 360 K, which is outside the range of the MD simulation time of this study. For many proteins, the transition state is close to the native state, and therefore, the kinetics of refolding are dominated by a large negative heat capacity change. This results in the characteristic observation that plots of refolding rate versus temperature go through a maximum. In apo-calmodulin, the maximum refolding rate is 6000 s^{-1} at 307 K.[46]

It should be stressed that LD does not predict transition states per se. Although some of the conformations along the trajectory might be energetically higher than the starting conformation and thus can be considered as representing transient conformations, they do not necessarily represent the most favorable path from the folded to the unfolded state. LD serves rather as a dynamic framework in which the molecule explores pathways to energetically more favorable states, where only the initial conformer and an ensemble of final conformers should be regarded as representative.

Role of Solvation in the Energetics of Unfolding

Two factors are particularly important in unfolding simulations: (i) the temperature dependence of solvation energetics and (ii) the solvent-induced electrostatic shield-

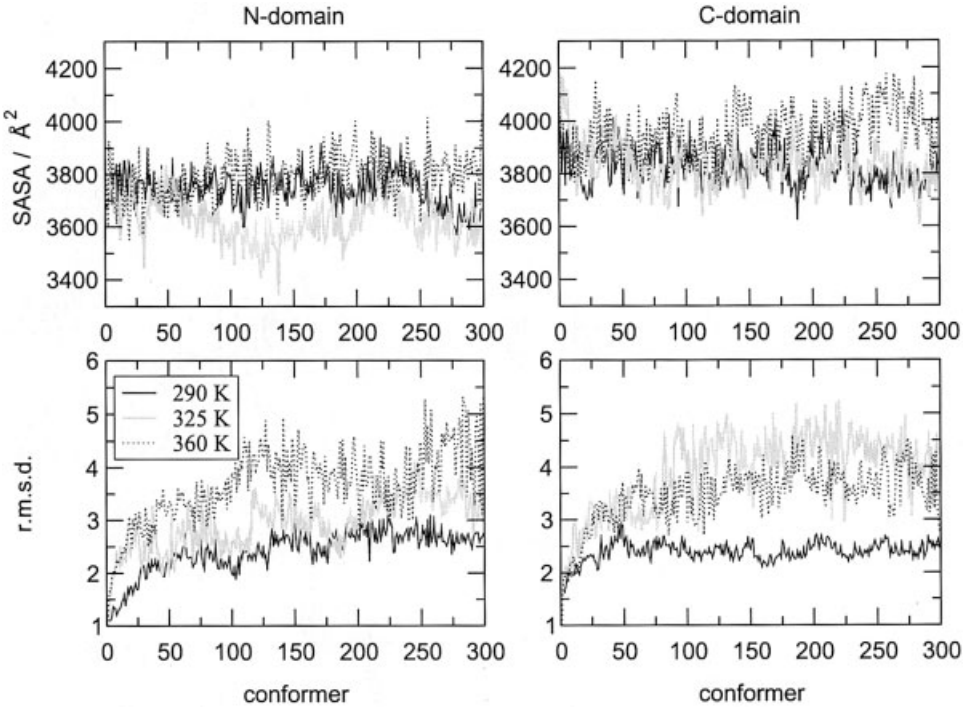


Fig. 5. Trajectories of hydrophobic SASA (**top**) and RMSD values (**bottom**) of apo-calmodulin. SASAs of the N-domain at different temperatures show no clear distinction between the different temperatures. Unfolding of the C-domain leads to markedly increased SASA at 360 K only. RMSD values of the N-domain reflect increasing mobility or partial unfolding at 325 K and unfolding at 360 K. The C-domain reveals a stable trajectory in RMSD at 290 K and apparent unfolding at the two higher temperatures.

TABLE I. Properties of apo-calmodulin along LD Trajectories (LD) of N-domain (top) and C-domain (bottom) Averaged Over the Last 25 Conformations of Four LD Trajectories per Temperature[†]

Property	290 K^{MD}	290 K^{LD}	325 K^{LD}	360 K^{LD}
N-domain				
Helix content[a]	39 ± 2	42.7 ± 3.5	37.4 ± 5.7	25.3 ± 5.8
Effective contacts[b]	(9000)	−567 ± 64	−657 ± 57	−1248 ± 57
$SASA_{pho}$[c]/Å^2	3497 ± 93	3664 ± 54	3600 ± 57	3818 ± 77
RMSD[d]	–	2.7 ± 0.1	3.4 ± 0.2	4.4 ± 0.6
E[e]/kJ · mol^{-1}	2524	5132 ± 19	5092 ± 27	5023 ± 33
C-domain				
Helix content	35 ± 3	32 ± 6.9	26.3 ± 3.3	23.5 ± 6.9
Effective contacts	(8500)	−440 ± 83	−842 ± 44	−976 ± 65
$SASA_{phob}$	3738 ± 84	3781 ± 48	3801 ± 69	4026 ± 88
RMSD	–	2.5 ± 0.1	4.2 ± 0.4	3.7 ± 0.5
E/kJ · mol^{-1}	2076	5513 ± 26	5407 ± 25	5324 ± 39

[†]For comparison, MD equilibrium conformations (MD) are reported averaged over the last 100 conformations.
[a]The number of residues in helical conformation.
[b]The loss of contacts in common with the reference conformations. The number in parentheses indicates the number of reference contacts.
[c]Hydrophobic solvent-accessible surface area as calculated in implicit solvation.
[d]Root mean square deviation values of backbone atoms.
[e]Potential energy of the domain in the GROMOS96 force field using implicit solvation.

ing of dipoles. Both factors are difficult to treat computationally.

Most thermal unfolding simulation studies have used temperatures in the range of 400 K–500 K to accelerate the unfolding process. Explicit water models have generally been parametrized at ambient conditions (i.e., 300 K and 1 bar). Water changes its properties in simulations quite dramatically above 400 K, and this is likely to affect protein unfolding by adding excess contributions to the free energy of the protein.[47] The rather subtle and compen-

sating effects of the temperature dependence of helical hydrogen bonding have recently been addressed experimentally.[48] The implicit solvation model used in this study has been calibrated by using a set of folded proteins at 300 K.[49,50] For the LD simulation temperatures from 290 K to 360 K, because the properties of the implicit solvation model are temperature independent, no excess contributions from the solvation are expected. On the contrary, thermal energy transfer from the solvent to the protein is nonexistent in implicit solvent.

Although dipolar shielding is an intrinsic feature of explicit water models, which has even been modeled by using an additional term to treat polarization,[51] the implicit solvation model used here does not account for this effect. It is plausible that unfolded conformers are slightly disfavored in implicit solvent compared to explicit water, because NH and CO dipoles are not electrostatically shielded. The shielding effect of water has been approximated in a recent study by a combination of a distance-dependent dielectric constant and reduced charges for ionic groups.[52] To explore the persistence of the residual secondary structure elements, a simulation at artificially high temperature (450 K) was undertaken here, and this does indeed yield unfolding to a residual helical content of 12%, similar to the end point of the experimental unfolding at 75°C (see below). This result suggests the extent of calculated residual structure may depend on the chosen description of solvation. Despite these limitations, the implicit model used here has been successfully applied (albeit in combination with a different force field) in several unfolding studies,[53,54] and residual structure has been found, by combined theoretical and experimental means, in many unfolded proteins (e.g., barnase,[55] engrailed homedomain,[30] and SH3.[53])

Comparison With Experimental Data

Far-UVCD has been extensively used to estimate the degree of unfolding of proteins as a function of pH, denaturant, and temperature. In calmodulin, a combination of near and far-UVCD, absorption, and fluorescence allows the distinctive behavior of the two domains to be identified.[45] Analysis of the far-UVCD thermal unfolding curve of apo-calmodulin yields values for transition temperatures T_m and enthalpies of unfolding ΔH_m for the C-domain of 43.5 ± 0.2° (316.5 K) and 31.2 ± 0.4 kcal/mol^{-1}, and for the N-domain of 57.5 ± 0.2° (330.5 K) and 44.6 ± 0.5 kcal/mol^{-1}, respectively. The corresponding computed unfolding curves for the N- and C-domains in intact apo-calmodulin are shown in Figure 6(A). Figure 6(B and C) show the far-UV CD spectra and the extracted helical content of apo-calmodulin as a function of temperature, using the methods described by Sreerama and Woody.[56] The helical content of the final conformers of the LD simulations is offset relative to the experimental value at higher temperatures, but the the final conformers of the 450 K trajectory reveals a helical content of 12%, close to the experimental value. For completeness, we show less reliable values from single trajectories at 5°C intervals.

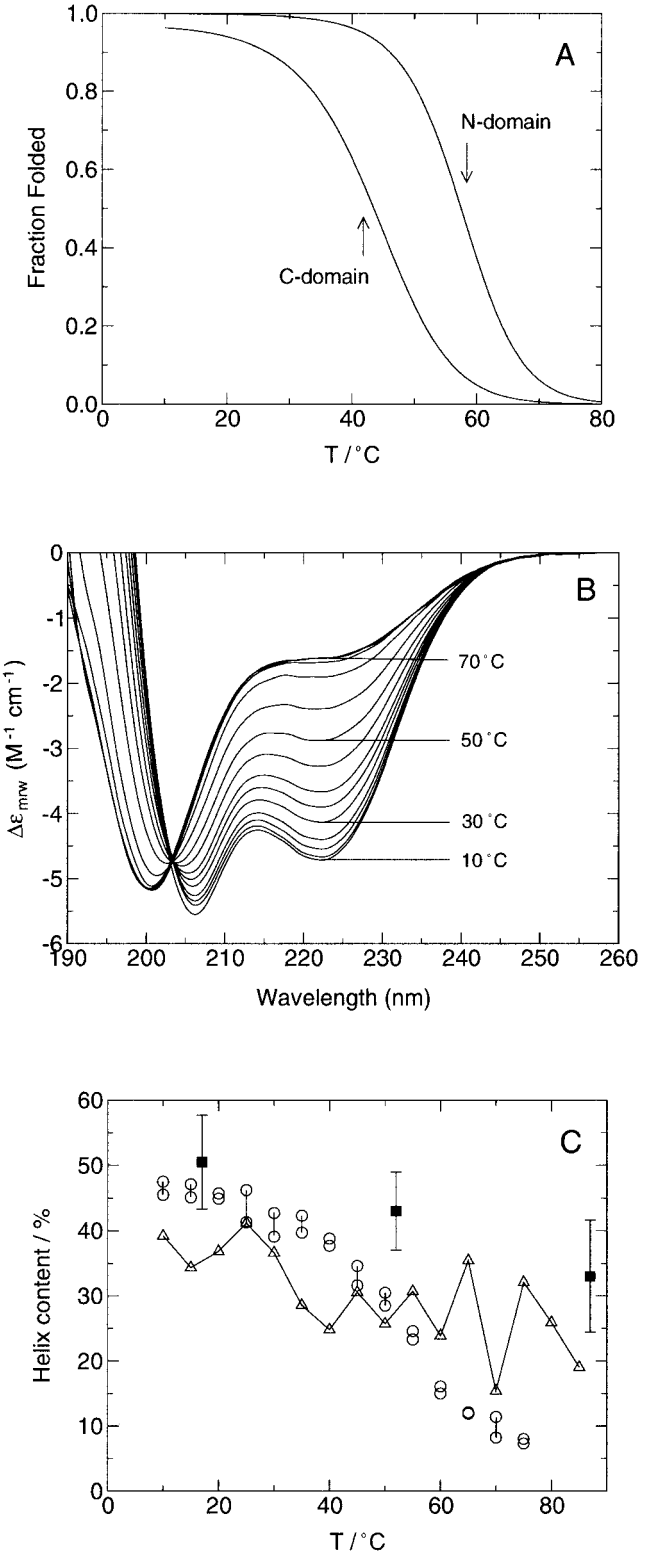
DISCUSSION

Monitored properties reveal that the LD trajectories at experimental temperatures discriminate between the relatively subtle unfolding behavior of the two domains in apo-calmodulin, in agreement with experimental findings. The LD simulations of apo-calmodulin thermal unfolding show that protein unfolding simulations under experimental conditions can yield improved agreement with experimental results when nonclassical sampling is used to enhance the exploration of conformational space. The most striking result is that the LD simulations show a clear distinction between the behavior of the two domains of apo-calmodulin in the temperature sensitivity of the unfolding, reproducing experimental unfolding results. This performance is even more significant considering the high-sequence identity and structural similarity between the domains.

Significant fluctuations of secondary structure elements are observed in individual trajectories in the simulations of the folded state. The main localized processes observed to accompany unfolding are as follows: loss of β-sheets between loops, fraying of helices particularly at the termini of domains and dynamic changes of helix length, relative dislocation of EF-hands within a domain, helix unwinding, repacking of unfolded segments, and partial β-sheet formation. In addition, significant interactions occur between the two domains. The first visible event is loss of the β-strands, but this is a reversible process with the short β-sheet reappearing during the trajectories as long as the domain remains in an overall folded state. This β-sheet is part of a cycle of aligned donor-acceptor pairs, which results in structural stabilization of the molecule.[57,58] The importance of this structural element for the overall stability of the molecule is shown by the mutation of (Ile or Val) residues at position 8 of the loop, leading to a major destabilization of the C-domain.[43] Further LD simulations are currently being applied to the V136G mutant[59,60] to evaluate the potential predictive capability of the method for protein-engineering applications. The flexible helices of the C-terminus show time-dependent variation in length, and the interdomain sequence is occasionally seen to form a temporarily continuous helix with the adjacent helices D and E. Helix G is energetically less stabilized by noncovalent interactions than the other helices and extremely mobile, whereas helix H is susceptible to unwinding from the C-terminal end. Both factors may lead to the disruption of the hydrophobic core and unfolding at low temperature.

There are, of course, inevitable limitations in the simulation of the complete unfolding process of a protein molecule of this size. Ideally, folding/unfolding should be shown to be a reversible process as in the definitive studies of peptide unfolding in organic solvents.[20] However, two major obstacles prevent us thus far from achieving similar results with proteins. First, simulations of processes in (explicit) water are much slower than those in organic solvents owing to the large number of solvent atoms, the tight hydrogen network, and strong solvent-solvent and solvent-solute interactions. Second, the experimentally

observed kinetics of protein folding tend to be slow relative to current simulation timescales. The use of implicit solvation resolves some of the difficulties connected to the strong water interactions, and it permits the nonclassical leaps without the need for long equilibration times. Additional improvements of solvation terms and sampling strategies may be expected to further enhance the general methods for the simulation of the reversible unfolding of proteins.



MATERIALS AND METHODS

Protein Expression and Purification

Drosophila melanogaster wild-type calmodulin was expressed in *Escherichia coli* and purified as described previously.[43]

Optical Spectroscopy and Thermodynamic Analysis

Far-UVCD spectra were recorded on a Jasco J-715 spectropolarimeter as described elsewhere[45] and analyzed for secondary structure content.[56]

Molecular Simulations

The MD and LD calculations were performed on Intel PII processors running on the Linux 2.2.14 kernel. Protein coordinates for apo-calmodulin were taken from the Brookhaven Protein Data Bank (PDB entry 1CFD). The GROMOS96 package was used for energy minimization (EM) and MD simulations.[61] MD simulations were performed for 3 ns at each temperature 290 K, 325 K, and 360 K, applying weak coupling to a temperature bath and using a relative relaxation time of 0.1. The SHAKE algorithm[62] was used, enabling timesteps of 2 fs. EM was performed by using the steepest descent method. Implicit solvent was incorporated by a mean force for water interactions,[49,50] using three interaction parameters for apolar, polar, and charged atoms and modified treatment of pseudo-ionic side-chains as described elsewhere.[27] Secondary structure calculation and hydrogen bond assessment were performed by using the program DSSP.[63] RMSD differences were taken from SAP structural superpositions.[64] Effective contacts were calculated on the basis of neighbor lists created by the PRONBL routine of GROMOS96, and energies were obtained from the PROMD routine. A contact between two residues was counted when the distance of any two atoms was $\leq$5 Å. Solvent-accessible surface areas were computed by applying routines of the implicit solvation model.[49,50]

LD Simulations

The LD scheme[27] consists of a conformer-generating procedure using the program CONCOORD and a refine-

Fig. 6. Comparison of apo-calmodulin thermal unfolding transition of N- and C-domain measured by far-UVCD spectroscopy and calculated helical content. **A**: Theoretical unfolding curves for N- and C-domain based on thermodynamic data from CD. Arrows indicate the calculated transition midpoints (T_m). **B**: Far-UVCD spectra of the thermal unfolding of apo-calmodulin from 10 to 85°C in intervals of 5°C (curves from 70 to 85°C are superposed). **C**: Percentage helical content of apo-calmodulin calculated from far-UVCD data (circles). Note the residual helical content of about 10% at high temperature. The average helical content (with standard deviation interval) of the LD conformations is plotted for the three simulation temperatures 290 K, 325 K, and 360 K (squares). Unfolding to a residual helix content of 12% was observed at a simulation temperature of 450 K. Values from a series of single trajectories are shown as triangles.

ment procedure using classical EM and MD in implicit solvent.

The CONCOORD method[65] is a distance geometry procedure designed to generate conformers from a starting structure that are representing its most important collective degrees of freedom. The generated conformers have randomized atomic coordinates that satisfy rule-based distance bounds derived from the starting structure. Structural differences between the generated conformers have been proven to be similar to those observed from experiment and from conventional MD simulations. One of the test molecules for the validation of the CONCOORD method was calmodulin.

A CONCOORD-generated conformer is subjected to a refinement procedure consisting of a sequence of EM (50 steps), MD (10 ps), and EM (500 steps). The refined conformer is then resubmitted to the CONCOORD procedure to start the next cycle; because of the randomization of coordinates, we call this transition a "leap". The number of cycles of the entire scheme was set here to 300. To limit conformational deviations, selection criteria were applied to CONCOORD-generated conformers. The acceptance criteria were twofold: (i) a new conformer was accepted if the RMSD value to the previous conformer was below a specified threshold (here 5 Å) and (ii) a Metropolis selection scheme was applied to ensure sampling of a Boltzmann distribution around a given energy threshold (here the average energy of the already sampled ensemble).

In the LD approach, generation of a trajectory is considered as a stochastic process (i.e., as a Markov chain). Within such a stochastic framework, a motion is defined as the outcome of trials (new conformations), which depend only on the immediately preceeding trial (previous conformation). The conformational search is not directed by an artificial potential or external constraints. Compared to conventional MD, the conformational changes imposed by LD lead to enhanced sampling, and the dynamic evolution follows a statistically broad path rather than a single trajectory. In the limiting case $\lim_{\mathrm{RMSD}\to 0}$, the LD scheme is equivalent to classical MD, enabling an evaluation of the perturbation effect on sampling efficiency. There is no temporal correlation between any two conformers that are generated by LD; therefore, the accessible space defined by the distance bounds is more efficiently sampled than by procedures in which a temporal correlation exists. This finding distinguishes the LD procedure from high-temperature unfolding schemes that sample and collect conformers at artificially high temperatures. In the LD approach, the refinement phase (and data point collection) is performed after the system has relaxed at the experimental temperature.

The LD procedure was automatically performed by a Perl script. Four trajectories consisting of 300 conformations were simulated for apo-calmodulin at temperatures 290 K, 325 K, and 360 K, summing up to a total MD simulation time of 3 ns per trajectory. These trajectories were analyzed in detail. A separate set of 16 trajectories (identically configured to the previous ones) was calculated for the simulated unfolding curve in Figure 6(C) to match each of the temperature points of the CD spectra. The secondary content of the last 25 conformers at each temperature was averaged and plotted as percentage of helical residues.

ACKNOWLEDGMENTS

The authors are grateful to P. Browne for the preparation of calmodulin samples.

REFERENCES

1. Dill KA. Dominant forces in protein folding. Biochemistry 1990;29: 7133–1755.
2. Fersht AR, Daggett V. Protein folding and unfolding at atomic resolution. Cell 2002;108:573–582.
3. Kauzmann W. Some factors in the interpretation of protein denaturation. Adv Protein Chem 1959;14:1.
4. Tanford C. How protein chemists learned about the hydrophobic factor. Protein Sci 1997;6:1358–1366.
5. Schellman JA. The thermodynamic stability of proteins. Annu Rev Biophys Biophys Chem 1987;16:115–137.
6. Smith LJ, Fiebig K, Schwalbe H, Dobson CM. The concept of a random coil: residual structure in peptides and denatured proteins. Fold Design 1996;1:95–106.
7. Dill KA. Polymer principles of protein folding. Protein Sci 1999;8: 1166–1180.
8. Dobson C, Karplus M. The fundamentals of protein folding: bringing together theory and experiment. Curr Opin Struct Biol 1999;9:92–101.
9. Grantcharova V, Alm EJ, Baker D, Horwich AL. Mechanism of protein folding. Curr Opin Struct Biol 2001;11:70–82.
10. Lazaridis T, Karplus M. "New view" of protein folding reconciled with the old through multiple unfolding simulations. Science 1997;278:1928–1931.
11. Marti-Renom MA, Stote RH, Querol E, Aviles FX, Karplus M. Refolding of potato carboxypeptidase inhibitor by molecular dynamics simulations with disulfide bond constraints. J Mol Biol 1998;284:145–172.
12. Brooks CL. Simulations of protein folding and unfolding. Curr Opin Struct Biol 1998;8:222–226.
13. Duan Y, Kollman A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 1998;282:740–744.
14. Tsai J, Levitt M, Baker D. Hierarchy of structure loss in MD simulations of src SH3 domain unfolding. J Mol Biol 1999;291:215–255.
15. Paci E, Karplus M. Unfolding proteins by external forces and temperature: the importance of topology and energetics. Proc Natl Acad Sci USA 2000;97:6521–6526.
16. Daggett V. Molecular dynamics simulations of protein unfolding/folding. Methods Mol Biol 2001;168:215–247.
17. Cavalli A, Ferrara P, Caflisch A. Weak temperature dependence of the free energy surface and folding pathways of structured peptides. Proteins 2002;47:305–314.
18. Brooks C, Karplus M, Petitt B. Proteins: a theoretical perspective of dynamics, structure, and thermodynamics. Adv Chem Phys 1988;71:1–259.
19. Stocker U, and van Gunsteren WF. Molecular dynamics simulation of hen egg white lysozyme: a test of the GROMOS96 force field against nuclear magnetic resonance data. Proteins 2000;40:145–153.
20. Daura X, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Reversible peptide folding in solution by molecular dynamics simulation. J Mol Biol 1998;280:925–932.
21. Straub JE, Thirumalai D. Exploring the energy landscape in proteins. Proc Natl Acad USA 1993;90:809–813.
22. Kitao A, Hayward S, Go N. Energy landscape of a native protein: jumping-among-minima model. Proteins 1998;33:496–517.
23. Wu X, Wang S. Self-guided molecular dynamics. J Phys Chem B 1998;102:7238–7250.
24. Elber R, Meller J, Olender R. Stochastic path approach to compute atomically detailed trajectories: application to the folding of C peptide. Phys Chem B 1999;103:899–911.
25. Crooks GE, Chandler D. Efficient transition path sampling for

nonequilibrium stochastic dynamics. Phys Rev E 2001;64: 026109,1–026109,4.
26. Feldman HJ, Hogue CWV. Probabilistic sampling of protein conformations: new hope for brute force? Proteins 2002;8:8–23.
27. Kleinjung J, Bayley P, Fraternali F. Leap-dynamics: efficient sampling of conformational space of proteins and peptides in solution. FEBS Lett 2000;470:257–262.
28. Chamberlain AK, Marqusee S. Touring the landscapes: partially folded proteins examined by hydrogen exchange. Structure 1997;5: 859–863.
29. Munoz V, Eaton WA. A simple model for calculating the kinetics of protein folding from three-dimensional structures. Proc Natl Acad Sci USA 1999;96:11311–11316.
30. Mayor U, Johnson CM, Daggett V, Fersht AR. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. Proc Natl Acad Sci USA 2000;97:13518–13522.
31. Kuboniwa H, Tjandra N, Grzesiek S, Ren H, Klee CB, Bax A. Solution structure of calcium-free calmodulin. Nat Struct Biol 1995;2:768.
32. Zhang M, Tanaka T, Ikura M. Calcium-induced conformational transition revealed by the solution structure of apo-calmodulin. Nat Struct Biol 1995;2:758–767.
33. Urbauer JL, Short JH, Wand AJ. Structural analysis of a novel interaction by calmodulin: high-affinity binding of a peptide in the absence of calcium. Biochemistry 1995;34:8099–8109.
34. Crivici A, Ikura M. Molecular and structural basis of target recognition by calmodulin. Annu Rev Biophys Biomol Struct 1995;24:85–116.
35. Jurado LA, Chockalingam PS, Jarrett HW. Apocalmodulin. Physiol Rev 1999;79:661–682.
36. Wriggers W, Mehler E, Pitici F, Weinstein H, Schulten K. Structure and dynamics of calmodulin in solution. Biophys J 1998;74:1622–1639.
37. Malmendal A, Evenäs J, Forsen S, Akke M. Structural dynamics in the C-terminal domain of calmodulin at low calcium levels. J Mol Biol 1999;293:883–899.
38. Evenäs J, Malmendal A, Akke M. Dynamics of the transition between open and closed conformations in a calmodulin C-terminal domain mutant. Structure 2001;9:185–195.
39. Vigil D, Gallagher SC, Trewhella J, Garcia AE. Functional dynamics of the hydrophobic cleft in the N-domain of calmodulin. Biophys J 2001;80:2082–2092.
40. Wilson MA, Brunger AT. The 1.0 Å crystal structure of Ca^{2+}-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. J Mol Biol 2000;301:1237–1256.
41. Chou JJ, Li S, Klee CB, Bax C. Solution structure of Ca^{2+}-calmodulin reveals flexible hand-like properties of its domains. Nat Struct Biol 2001;8:990–997.
42. Tsalkova TN, Privalov PL. Thermodynamic study of domain organization in troponin C and calmodulin. J Mol Biol 1985;181: 533–544.
43. Browne JP, Strom M, Martin SM, Bayley PM. The role of β-sheet interactions in domain stability, folding, and target recognition reactions of calmodulin. Biochemistry 1997;36:9550–9561.
44. Sorensen BR, Shea MA. Interactions between domains of apo-calmodulin alter calcium binding and stability. Biochemistry 1998;37:4244–4253.
45. Masino L, Martin S, Bayley PM. Ligand binding and thermodynamic stability of a multi-domain protein, calmodulin. Protein Sci 2000;9:1519–1529.
46. Rabl CR, Martin SR, Neumann E, Bayley PM. Temperature-jump study of the stability of apo-calmodulin. Biophys Chem 2002;101–102:553–564. Forthcoming.
47. Walser R, Mark AE, van Gunsteren WF. On the temperature and pressure dependence of a range of properties of a type of commonly used in high-temperature protein unfolding simulations. Biophys J 2000;78:2752–2760.
48. Taylor JW, Greenfield NJ, Wu B, Privalov PL. A calorimetric study of the folding -unfolding of an α-helix with covalently closed N- and C-terminal loops. J Mol Biol 1999;291:965–976.
49. Fraternali F, van Gunsteren WF. An efficient mean solvation force model for use in molecular dynamics simulations of proteins in aqueous solution. J Mol Biol 1996;256:939–948.
50. Fraternali F, Cavallo L. Parameter optimised surfaces (POPS): analysis of key interactions and conformational changes in the ribosome. Nucleic Acids Res 2002;30:2950–2960.
51. Sharp KA. Inclusion of solvent effects in molecular mechanics force fields. In: van Gunsteren W, Wilkinson A, editors. Computer simulations of biomolecular systems. Leiden, The Netherlands: ESCOM Science Publishers B.V.; 1993. p 147–160.
52. Lazaridis T, Karplus M. Effective energy function for proteins in solution. Proteins 1999;35:133–152.
53. Gsponer J, Caflish A. Role of native topology investigated by multiple unfolding simulations of four SH3 domains. J Mol Biol 2001;309:285–298.
54. Ferrara P, Apostolakis J, Caflisch A. Evaluation of a fast implicit solvent model for molecular dynamics simulations. Proteins 2002; 46:24–33.
55. Bond CJ, Wong KB, Clarke J, Fersht AR, Daggett V. Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway. Proc Natl Acad Sci USA 1997;94:13409–13413.
56. Sreerama N, Woody RW. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON and COSSTR methods with an expanded reference set. Anal Biochem 2000;287:252–260.
57. Biekofsky RR, Martin SR, Browne JP, Bayley PM, Feeney J. Ca^{2+} coordination to backbone carbonyl oxygen atoms in calmodulin and other EF-hand proteins: ^{15}N chemical shifts as probes for monitoring individual-site Ca^{2+} coordination. Biochemistry 1998; 37:7617–7629.
58. Biekofsky RR, Feeney J. Cooperative cyclic interactions involved in metal binding to pairs of sites in EF-hand proteins. FEBS Lett 1998;439:101–106.
59. Fefeu S, Biekofsky RR, Martin SR, Bayley PM, McCormick J, Feeney J. Calcium-induced refolding of the calmodulin V136G mutant: interaction between the two globular domains. Biochemistry 2000;39:15920–15931.
60. Biekofsky RR, Martin SR, McCormick JE, Masino L, Fefeu S, Bayley PM, Feeney J. Thermal stability of calmodulin and mutants studied by ^{1}H-^{15} N HSQC NMR measurements of selectivity labeled ^{15}N Ile proteins. Biochemistry 2002;41:6850–6859.
61. van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott W, Tironi I. Biomolecular simulations: the GROMOS96 manual and user guide. Zürich, Groningen: BIOMOS b.v., Laboratory of Physical Chemistry, ETH Zentrum, CH-8092 Zürich vdf Hochschulverlag AG; 1996.
62. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of alkanes. J Comput Phys 1977;23:327–341.
63. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
64. Taylor WR. Protein structure comparison using iterated double dynamic programming. Protein Sci 1999;8:54–665.
65. de Groot BL, van Aalten DMF, Scheek RM, Amadei A, Vriend G, Berendsen HJC. Prediction of protein conformational freedom from distance constraints. Proteins 1997;29:240–251.