Characterization of the Protein Unfolding Processes Induced by Urea and Temperature

Alessandro Guerini Rocco,* Luca Mollica,† Piero Ricchiuto,* António M. Baptista,‡ Elisabetta Gianazza,* and Ivano Eberini*

*Gruppo di Studio per la Proteomica e la Struttura delle Proteine, Dipartimento di Scienze Farmacologiche, Università degli Studi di Milano, Milano, Italy; †Dulbecco Telethon Institute c/o S. Raffaele Scientific Institute, Biomolecular NMR Laboratory, Milano, Italy; and ‡Instituto de Tecnologia Quimica e Biologica, Universidade Nova de Lisboa, Oeiras, Portugal

ABSTRACT Correct folding is critical for the biological activities of proteins. As a contribution to a better understanding of the protein (un)folding problem, we studied the effect of temperature and of urea on peptostreptococcal Protein L destructuration. We performed standard molecular dynamics simulations at 300 K, 350 K, 400 K, and 480 K, both in 10 M urea and in water. Protein L followed at least two alternative unfolding pathways. Urea caused the loss of secondary structure acting preferentially on the β-sheets, while leaving the α-helices almost intact; on the contrary, high temperature preserved the β-sheets and led to a complete loss of the α-helices. These data suggest that urea and high temperature act through different unfolding mechanisms, and protein secondary motives reveal a differential sensitivity to various denaturant treatments. As further validation of our results, replica-exchange molecular dynamics simulations of the temperature-induced unfolding process in the presence of urea were performed. This set of simulations allowed us to compute the thermodynamical parameters of the process and confirmed that, in the configurational space of Protein L unfolding, both of the above pathways are accessible, although to a different relative extent.

INTRODUCTION

The relevance of protein folding has been recognized for many years. Recent studies show that pathological conditions may result from a protein misfolding process: Alzheimer's and Creutzfeldt-Jacob's disease, cystic fibrosis, and some cancer (1–4). Most proteins, such as enzymes and receptors, fulfill their biological activity only when correctly and completely folded. The folding process often starts when protein translation is not yet completed: the protein N-terminus begins to fold while the C-terminus is still being synthesized by the ribosomal machinery (5). Specialized cellular proteins called chaperones sometimes assist in the process.

If compared with the time simulated by standard molecular dynamics techniques, folding is complex and slow (some folding events are completed within seconds, while the complete folding process can span up to minutes), hence it is difficult to study with computational tools. For these reasons, several scientists are approaching this topic focusing their attention on the unfolding process, which is useful to identify key structural features that keep the proteins folded. To enhance the rate of the unfolding process, both the experimental and the computational protocols in use resort to high temperature, high pressure, low pH, or chemical denaturants (i.e.: guanidinium chloride or urea) (6–10).

Chaotropic agents like urea upset the established hydrogen-bonding pattern: for proteins, this results in a loss of high-order structure. A few molecular dynamics (MD) simulations of proteins in explicit aqueous solvent containing urea at high concentration have been reported in the literature (11–17). All these studies have focused on understanding the molecular mechanism of urea-induced denaturation, which is not completely elucidated, despite urea being in general use for protein structural studies. Two hypothetical mechanisms have been proposed for its chaotropic mechanism. According to the first, urea upsets the hydrogen-bonding network of the solvent around hydrophobic side chains, providing a better solvation environment for nonpolar amino acids; according to the second, urea interacts directly with the protein, competing with intramolecular hydrogen bonding (16,17).

In this investigation, we first built, in silico, a 10 M urea solution and assessed its simulated physicochemical properties with reference to the experimental features. We chose as test molecule the light-chain immunoglobulin-binding domain of Protein L from *Peptostreptococcus magnus*, a small peptide whose structure is well characterized (18–21). As seen in Fig. 1 (*middle left*), the peptide features three secondary structural elements: the first β-hairpin, the 15 amino-acid-long α-helix, and the second β-hairpin, with the overall topology of an ubiquitinlike roll (CATH Classification, Ver. 3.0.0).

We performed several standard MD simulations on Protein L at various temperatures both in water and in a 10 M urea box. Furthermore, replica-exchange molecular dynamics (REMD) (22) simulations were carried out to improve the sampling of the configurational space of Protein L unfolding in comparison with standard MD. A more reliable sampling of the structures explored during the unfolding process could be ensured, because REMD simulations permit bridging between states for which interconversion is difficult and are thus able to easily overcome energy barriers.

Submitted June 19, 2007, and accepted for publication November 5, 2007.

Address reprint requests to Ivano Erini, Tel.: 39-02-5031-8395, E-mail: ivano.eberini@unimi.it.

Editor: Angel E. Garcia.

0006-3495/08/03/2241/11 $2.00

doi: 10.1529/biophysj.107.115535

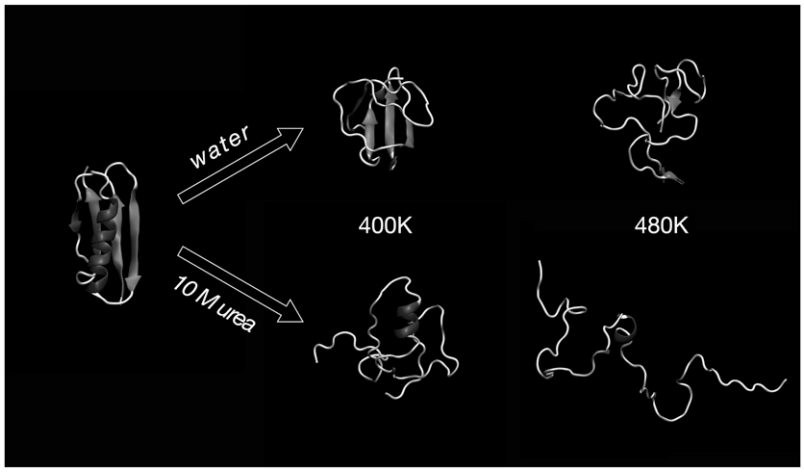

FIGURE 1 Mean NMR structure of Protein L (*left*), and structures at the end of standard MD, at 400 K (*middle*) and at 480 K (*right*), as simulated in pure water (*top row*) or in a 10 M urea solution (*bottom row*). Protein structure is rendered (with VMD) as α-carbon ribbon.

One of the greatest challenges in the computational study of the leading events in protein folding/unfolding is to obtain a reliable and accurate distribution of configurational states. Standard MD simulations, carried out at increasing temperature, provide continuous, realistic pathways of the unfolding process at atomic resolution. Moreover, the differences in temperature between different runs may be set arbitrarily, hence a wide range can be explored (if unsystematically). However, in ordinary simulations at low temperature, the systems may be trapped in local minima, or the events observed during a single simulation could be single occurrences without any statistical relevance. Hence, an efficient sampling of the phase space is not insured and the simulation could, in principle, explore only a small portion of the potential energy surface.

An approach suggested in the late 1990s for the correct statistical treatment of this problem features an implementation of the standard Cartesian space MD in the REMD scheme (22) (also referred to as the multiple Markov chain method or parallel tempering (23)).

In this implementation, many copies of the same system are run simultaneously and independently at different temperatures (noninteracting replicas of the same system). Periodically the replicas are exchanged with a transition probability ruled by the Metropolis criterion, and hence governed by the product of Boltzmann factors (22). In this way, a robust algorithm is provided for a correct exploration of the configurational space of the protein through a random walk performed in energy space. Moreover, the weight factors for obtaining the normal distribution of space are known a priori using a very simple dynamic routine (24).

The REMD method has been extensively used on biomolecular systems because of its low computational cost and of its effectiveness in providing a statistical mechanics value to MD simulations. The main applications of REMD are in the study of protein folding (25–27) and of protein aggregation (28). In this report, we use REMD for analyzing temperature-induced unfolding of a protein in an explicit solvent other than water, i.e., 10 M urea.

MATERIALS AND METHODS

Building and equilibrating the 10 M urea solvation box

We built a molecule of urea with the Builder module of the Insight II suite (Accelrys, San Diego, CA) running on SGI Fuel. The urea topology was taken from Smith et al. (29) and adapted for our use with the GROMACS program by Alessandra Villa (personal communication). Then we set up a $6 \times 6 \times 6$ nm^3 cubic box containing 160 molecules of urea and 480 of simple point charge (SPC) model of water. The system was subjected to energy minimization with the steepest-descent algorithm down to a maximum gradient of 2000 kJ/mol $\times$ nm^{-1}, and simulated for 1 ns, annealing from 300 to 0 K under an isotropic pressure of 100 bar. The system was then relaxed for 1 ns at standard pressure, heating from 0 to 300 K, and simulated for another 1 ns at 300 K. At the end of this equilibration procedure, the size of the box was $3 \times 3 \times 3$ nm^3. We performed two further 10-ns MD simulations in NVT ensemble at 300 and 310 K starting from the relaxed box, using the Nosé-Hoover coupling algorithm. We then applied the fluctuation-dissipation theorem (30), which states that h (computed as in Eq. 1) has the value of unity at equilibrium of

$$h = \frac{\langle E^2 \rangle - \langle E \rangle^2}{(RT)^2 C_V}, \tag{1}$$

in which the energy fluctuations can be obtained directly from the MD simulations. The specific heat C_V (see in Eq. 2) is estimated from a linear interpolation between the average energies of two simulations performed at two different temperatures:

$$C_V = \frac{d\langle E \rangle}{dT} = \frac{\langle E \rangle_{310} - \langle E \rangle_{300}}{R\Delta T}. \tag{2}$$

Standard MD in water and in 10 M urea

We removed the first 17 unstructured N-terminal amino acids from the mean NMR structure of light-chain immunoglobulin-binding domain of Protein L from *Peptostreptococcus magnus* (PDB ID: 2PTL) (18–21). The truncated protein (Protein L for short) was put in a cubic $5.6 \times 5.6 \times 5.6$ nm^3 box and

solvated with SPC water or with the 10 M urea solution described above. The system was neutralized by adding 1 Na^+ ion. Energy minimization was carried out as described above, then the system was simulated along the following steps: 1), 100 ps of position-restrained MD, with an isotropic force (1000 kJ/mol × nm^{-2}) applied to all protein atoms, to allow for solvent relaxation; and 2), unrestrained MD at various temperatures (as discussed later; conditions of the various simulations listed in Table 1). All the simulations were performed at 1 bar with a coupling constant of 1.0 ps for pressure and 0.1 ps for temperature, in both cases applying the Berendsen weak coupling algorithm (31). The time step for the integration algorithm was set at two femtoseconds and all bonds were constrained by the LINear Constraint Solver algorithm. The GROMACS 3.2.1 program (32–34) with the GROMOS96 force field (ffG43a1) was used, and fast particle-mesh Ewald electrostatics (35,36) was applied with the following parameters: distance for the Lennard-Jones cutoff = 0.9 nm; distance for the Coulomb cutoff = 0.9 nm; maximum spacing for the Fast Fourier Transform grid = 0.12 nm; and a cubic interpolation order. The RMSD was always calculated using the first timeframe as reference structure and fitting the α-carbon sets of all the subsequent structures. On all the trajectories, we calculated the frequencies of the pairs: number of intramolecular hydrogen bonds and gyration radius, number of native contacts and gyration radius, RMSD and number of amino acids in secondary structure (37). The results were rendered as density plots in two-dimensional space using the R Project for Statistical Computing software package (http://www.r-project.org), applying a probability coverage of 80%. All the simulations were carried out on SUN Ultra 40 workstations running Linux CentOS 4.3.

REMD

The REMD simulation was performed in NPT ensemble with the GROMACS 3.3 program on the same system (protein coordinates, urea, water molecules, and number of ions) used for the simulations described in the previous paragraph. Even if widely used in REMD simulations (24,38), the NVT ensemble less closely agrees with experimental data and at high temperature could generate artifacts due to a large increase in pressure values (25). The time for each replica simulation was 30 ns: this duration was chosen to ensure that the majority of the replicated simulations presented a very low content of structured residues, and that the protein-solvent energy profile of all the replicas reached a plateau. Twenty-two replicas were run, with the temperatures 300, 304, 308.1, 312.3, 316.6, 321, 325.5, 330, 334.7, 339.4, 344.3, 349.2, 354.3, 359.4, 364.7, 370, 375.4, 381, 386.6, 392.3, 398.1, and 404 K, respectively. This selection corresponds to an extensively discussed and widely used protocol ensuring a random walk in energy and conformational space (24). The temperature range covered in this simulation setup corresponds to the conditions most often adopted in protein folding/unfolding experiments (25,39). Temperature differences between 4 and 6 K avoid low exchange rate between replicas at the highest temperatures due to the decreasing overlap of energy states in NPT ensemble, as extensively discussed by Siebert et al. (25). An exchange was attempted every 500 steps (corresponding to 1 ps) and acceptance ratio was computed according to the Metropolis criterion of

$$\varphi(R_1 \leftrightarrow R_2) = \min[1, e^{-\Delta}], \Delta = (\beta_2 - \beta_1)(U_1 - U_2), \quad (3)$$

where φ ($R_1 \leftrightarrow R_2$) is the transition probability between two neighbor replicas (computed as the minimum between the percentages of 1 and $e^{-\Delta}$) and $\beta = \frac{1}{k_B}T$, where k_B is the Boltzmann constant (25). The exchange probability (computed as the ratio between the successful exchanges and the total number of trials) varied between 8% and 15% for each pair of neighbor replicas; the average exchange probability was 11 ± 2%, a level ensuring an efficient exploration of the conformational space (at least for polypeptide chains) (22). The REMD calculations were performed on an IBM eServer BladeCenter equipped with 28 double-processor JS20 Blades (56 PPC64 2.8 GHz processors) and Myrinet communication system.

RESULTS AND DISCUSSION

Standard MD simulations

We validated the 10 M urea box comparing computational parameters obtained from our MD simulations to experimental data for a 10 M urea solution. The density calculated for our system is 1134.53 $\frac{kg}{m^3}$, whereas the experimental values are 1147.0 $\frac{kg}{m^3}$ (40) and 1157.7 $\frac{kg}{m^3}$ (41). The measurement of the system density and its comparison with experimental data can be useful to evaluate the quality of the selected molecular descriptors. The differences in literature data can possibly be ascribed to instrumental and systematic errors and to the near-saturation concentration of the solution (40,41). The computed density of our system is slightly lower than the experimental values, and this deviation can be referred to the peculiar features of the SPC water model (29). Conversely, in their MD simulations on a system very similar to ours, Smith et al. (29) measured a slightly higher density value (40).

Our 10 M urea solvation box was built while modifying system pressure and temperature. To check for its thorough equilibration, we applied the fluctuation-dissipation theorem, and calculated h-values (for the definition, see Materials and Methods) during 10 ns of MD simulation: it ranged between 0.96 and 1.03 (data not shown). This result suggests that a necessary condition for the thermodynamic equilibrium was already satisfied at the onset of the MD. In fact, in the last step of the urea box preparation, the system is relaxed for 1 ns at standard pressure and increasing temperature, then simulated for 1 ns at 300 K. This protocol allows for the efficient equilibration of water and urea molecules.

TABLE 1

Simulation	Water	Urea (conc.)	Na^+	Time	T (K)	Type	Ensemble
1	5522	0	1	30 ns	300 K	Standard MD	NPT
2	5522	0	1	30 ns	350 K	Standard MD	NPT
3	5522	0	1	30 ns	400 K	Standard MD	NPT
4	5522	0	1	10 ns	480 K	Standard MD	NPT
5	3384	816 (10 M)	1	30 ns	300 K	Standard MD	NPT
6	3384	816 (10 M)	1	30 ns	350 K	Standard MD	NPT
7	3384	816 (10 M)	1	30 ns	400 K	Standard MD	NPT
8	3384	816 (10 M)	1	10 ns	480 K	Standard MD	NPT
9	3384	816 (10 M)	1	30 ns	300–404 K	REMD	NPT

We carried out eight standard MD simulations (30 ns, except for the simulations at 480 K, which had a simulation time of 10 ns) at different temperatures (300 K, 350 K, 400 K, and 480 K), both in water and in 10 M urea. We monitored several parameters to investigate the different unfolding mechanisms experienced by Protein L under these conditions.

The time evolution of the structural changes of all the MD simulations is plotted in Fig. 2. In water, at the lowest tested temperatures (300 and 350 K, in Fig. 2, *a* and *c*), the secondary structure of the protein is only minimally perturbed. At higher temperatures, however (Fig. 2, *e* and *g*), a loss of structural elements becomes most evident; the structures recorded at the end of the simulation time at 400 and 480 K are shown in the upper row of Fig. 1. The destructuration of α-helix in the MD simulation carried out at 400 K (Fig. 2 *e*) appears massive and highly cooperative, with sudden onset, and completion in <5 ns, whereas under the same conditions only 16% of β-structure is lost. At 480 K (Fig. 2 *g*) all of the α-structure is lost within 3 ns, whereas the β-structure content decreases by $\sim$30%, most of which was during the first nanoseconds of the simulation.

In 10 M urea, both the mode of protein unfolding and extent of structural loss are entirely different from water. At 300 K (Fig. 2 *b*), the structural variation during the simulation time is negligible; at 350 K (Fig. 2 *d*), only the β-structure content decreases by $\sim$20%. Massive loss of structure is instead shown by the matrices in Fig. 2, *f* and *h*, for 400 and 480 K MD, respectively. The β-structure drops to zero between 12 and 20 ns of simulation at 400 K (Fig. 2 *f*), and after 4 ns at 480 K (Fig. 2 *h*). Conversely, the content in α-structure progressively decreases with time, down to a final reduction of $\sim$66% at 400 K (Fig. 2 *f*), while at 480 K the α-helix is completely lost in 2 ns (Fig. 2 *h*). Fig. 1 (*bottom row*) shows the structure of Protein L at the end of the simulation in urea at 400 K (*middle*) and at 480 K (*right*).

In pure water simulations, the water radial distribution function (RDF), plotted in Fig. 3 *a*, approaches 1 around a radial distance of 2 nm. In 10 M urea simulations, the RDF (Fig. 3 *b*) of water and urea with respect to the protein are significantly different. Urea molecules are tightly packed around the protein: at all temperatures, the value of the distribution function equals 1 already at a radial distance of 0.4 nm, that of water at $\sim$2.4 nm. In agreement with observations reported by previous articles (11,13,16,42), the RDF between protein and solvents (Fig. 3, *a* and *b*) show that urea accumulates around the protein in comparison with the bulk of the solution. Because of the presence of urea instead of water in close contact with the protein, the number of hydrogen bonds between protein and water is remarkably lower in the simulations in 10 M urea (see later in this subsection). Indeed, in the simulation in 10 M urea, the RDF between water and Protein L tends to become more and more similar to pure water with increasing temperature, while in the water-only simulations the RDF of water with respect to the protein is similar at all temperatures. As a result, the number of protein-water hydrogen bonds increases with temperature, whereas the number of hydrogen bonds between protein and urea decreases. The time-dependent modification of water and urea RDFs is shown in the insets of Fig. 3 *b*: they are much more similar to one another at the beginning than at the end of the simulation.

The presence of high concentrations of urea strongly affects intra- and intermolecular interactions, as shown by Fig. 4, panel *b* versus panel *a*. Temperature considerably affects the number of hydrogen bonds, as reported in Fig. 4, *a* and *b*. The decrease of protein-urea hydrogen bonds with the increase of temperature is quite similar to the one for protein-water bonds in the water-only simulations, maybe because of the homogeneity of the closest solvation shells, made up mostly by urea in the first case, and by water in the second. Furthermore, at each simulation temperature, the number of protein-urea hydrogen bonds increases with time, whereas the number of intramolecular hydrogen bonds decreases, even if only at 400 K and 480 K to a remarkable extent (results not shown).

Urea leads to an increase of both hydrophobic and hydrophilic solvent-accessible surface (SAS) of the protein, computed with the g_sas GROMACS tool (43,44), see Fig. 3, *c–f*; furthermore, simulations in 10 M urea both at 400 and at 480 K are characterized by a time-dependent increase of SAS (Fig. 3, *d* and *f*). Maximal increase of SAS is recorded for the 10 M urea simulation at 480 K (Fig. 3, *d* and *f*). As already mentioned, one of the hypotheses about the urea-mediated unfolding mechanism is linked to an increased solvation both for hydrophobic and hydrophilic amino acids. In our simulations at the lower temperatures (i.e., 300 and 350 K), urea induces an increase of the hydrophilic SAS value with respect to water simulations. On the other hand, at the higher temperatures (i.e., 400 and 480 K), this effect is more noticeable on the hydrophobic SAS. All our data are in agreement with Tirado-Rives et al. (11), Zhang et al. (13), and Smith et al. (17), suggesting that SAS increase is a relevant aspect for urea-induced unfolding. This data on the effect of urea on hydrogen-bonding pattern and protein surface availability suggest that both of the proposed mechanisms can take part in the unfolding process.

To represent the free energy surface of the protein unfolding process, according to Zhang et al. (37), number of occurrences was recorded for pairs of reaction parameters of each MD simulation. This is: intramolecular hydrogen-bond number and radius of gyration (in Fig. 5); number of native contacts and gyration radius (in Supplementary Material Fig. SA); RMSD; and number of amino acids in secondary structure (in Supplementary Material Fig. SB).

Fig. 5 compares the density plots for the reaction coordinate pair (intramolecular hydrogen bonds and gyration radius) for the simulations in water (*black*) and urea (*green*) at 300 K (Fig. 5 *a*) and 480 K (Fig. 5 *b*). Eighty-percent probability coverage area is much larger in water at higher than at lower temperature; area is further extended in the presence of

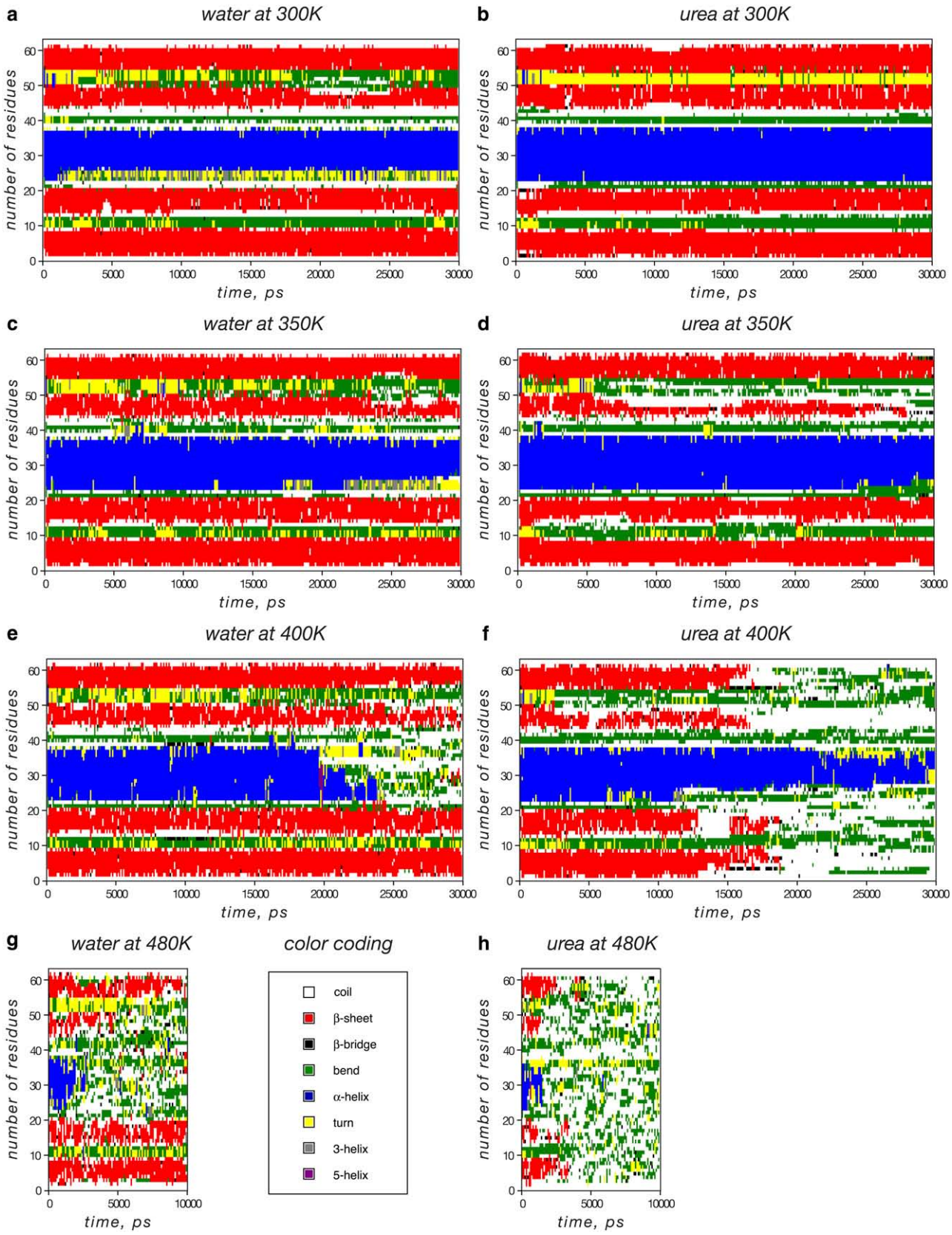

FIGURE 2 Existence matrices for specified secondary structures for simulations in water (*left*) and 10 M urea (*right*) at increasing temperatures. Secondary structure elements are classified according to the program DSSP (59). Types of secondary structure are color-coded as reported in the legend between *g* and *h*.

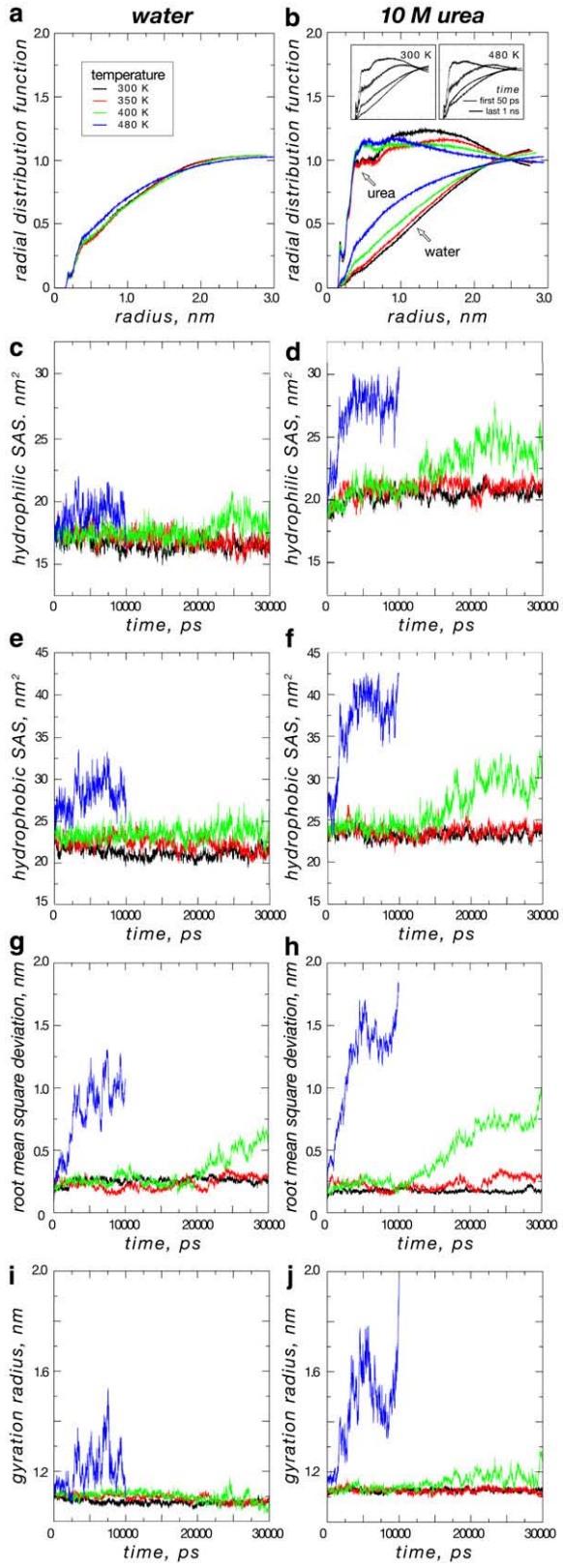

FIGURE 3 (*a* and *b*) Radial distribution function of the mass center of solvent molecules with respect to Protein L surface on the last 1 ns of standard MD in water (*a*) or in 10 M urea (*b*; water and urea plots are marked with *arrows*). Time-dependent variation of the RDF in 10 M urea at 300 K (*b*, *left inset*) and at 480 K (*b*, *right inset*): first 50 ps, gray line; last 1 ns, black line. (*c* and *d*) Hydrophilic SAS of Protein L during standard MD simulations in water (*c*) or in 10 M urea (*d*). (*e* and *f*) Hydrophobic SAS of Protein L during standard MD simulations in water (*e*) or in 10 M urea (*f*). (*g* and *h*) RMSD of Protein L during standard MD simulations in water (*g*) or in 10 M urea (*h*). (*i* and *j*) Gyration radius of Protein L during standard MD simulations in water (*i*) or in 10 M urea (*j*). Temperature-dependence of the plots is color-coded (see *inset* in *a*): 300 K, black; 350 K, red; 400 K, green; 480 K, blue.

urea. The newly represented species (higher versus lower temperature and urea versus water) are characterized by a lower number of intramolecular hydrogen bonds and a larger gyration radius; these obviously correspond to nonnative conformations involving a varying extent of unfolding. In standard MD simulations at 300 K (Fig. 5 *a*), urea causes reduction of hydrogen-bond number, and minimal variation of gyration radius. The chaotropic effect becomes more and more evident at higher temperatures. At 480 K (Fig. 5), the protein in urea has a higher gyration radius and a much lower number of hydrogen bonds than in water at the same temperature. The structures sampled from the widest green area in Fig. 5 *b* do not show any residual secondary structure (Fig. 1). All these observations suggest that the number of hydrogen bonds and the gyration radius are good reaction coordinates, useful to shape the energetical landscape and monitor conformational rearrangement during the protein unfolding process.

This conclusion is strengthened by the inspection of the plots for the two further pairs of reaction coordinates in Supplementary Material Figs. SA and SB, for both of which the same mutual relationships apply.

Taking the RMSD on the α-carbons as a marker of unfolding, only in the simulations at very high temperature ($\geq$400 K) can Protein L be said to unfold during the allotted simulation time (Fig. 3, *g* and *h*). As demonstrated by REMD data (see later in the ''REMD simulations'' subsection), unfolding would have eventually occurred in urea also at lower temperature (350 K), had the simulation time been longer than 30 ns. At a given temperature, the extent of unfolding assessed by RMSD is higher in 10 M urea than in water; as expected, urea enhances unfolding, resulting in an RMSD increase from 0.6 nm to 1.0 nm at 400 K and from 1.1 nm to 2.2 nm at 480 K. In Fig. 3, *i* and *j*, gyration radius is reported for all the run MD simulations.

The RMSD (Fig. 3, *g* and *h*), the gyration radius (Fig. 3, *i* and *j*), the number of intramolecular hydrogen bonds (Fig. 4), and the detailed structural content (Fig. 2) may all be followed as markers of unfolding. At any temperature, all of these features point out, as expected, a much higher extent of unfolding in 10 M urea than in water. Complete destructuration (namely, 100% loss of secondary structure content) is only attained in urea at the highest temperature (480 K): indeed, this is the only simulation in which a substantial increase of the radius of gyration is recorded. During the allotted simulation time of 30 ns, no major changes in

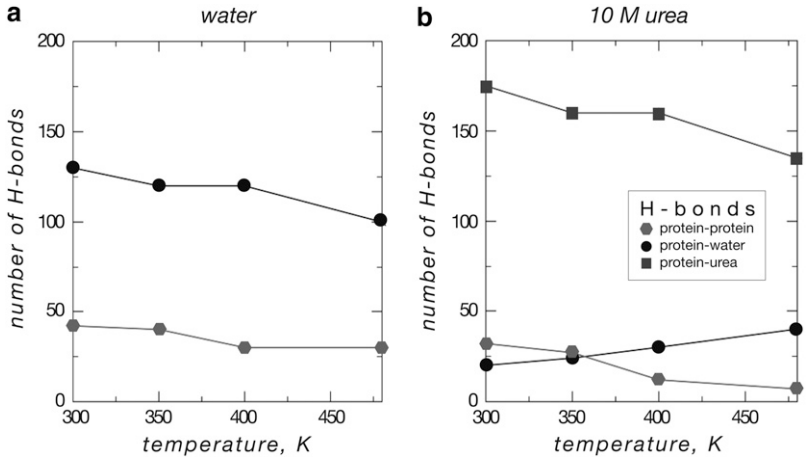

FIGURE 4 Number of hydrogen bonds after 30 ns of standard MD on Protein L in water (*a*) or in 10 M urea (*b*) as a function of the simulation temperature.

structural content are observed for Protein L below 350 K in either water or urea. Unfolding is detectable, in the simulations in 10 M urea at 350 and 400 K, as a loss of residues in β-sheet, and in the MD in water at 400 and 480 K, as a loss of the α-helix. From this evidence, we conclude that unfolding takes alternative pathways, depending on the experimental conditions: urea preferentially causes the loss of the β-sheet, whereas high temperature unfolds the α-helix. The comparison between MD simulations carried out at the same temperature in water or in urea point out a different stability of secondary structure motives: indeed, the presence of the chaotropic agent stabilizes the helical folding, whereas the strand folding seems quite sensitive to the same denaturing conditions.

We compared literature data with the above findings on urea and/or temperature-driven protein unfolding obtained by classical and/or replica exchange MD. Tirado-Rives et al. (11) reported that in the MD simulation of the unfolding of barnase, the β-sheet is affected by urea to a much larger extent than the α-helices. Differently, Tobi et al. (42), who carried out MD simulations of a blocked valine peptide in urea, observed a stronger hydrogen bonding of the peptide with urea than with water, with preference for helical conformation. And Bennion et al. (14), in an MD simulation of chymotrypsin inhibitor 2 in urea, observed a residual native helical structure, whereas the β-structure was completely lost.

An interesting set of articles from Garcia's research group (39,45,46) presented data on protein folding/unfolding obtained by REMD. The authors computed the free energy, entropy, and enthalpy surfaces as a function of structural reaction coordinates. The computed energy landscapes allowed them to follow and describe the folding mechanism for the helical fragment B of *Staphylococcus aureus* protein A (45). An extension of REMD, in which pressure and temperature were modified with a Monte Carlo method, was applied to study the unfolding of a 20-amino-acid α-helical peptide. No modification of the protein secondary structure was observed by varying the pressure, whereas water coordination to the backbone carbonyls of the protein was affected (46). In all these articles (39,45,46), the authors showed that REMD is a computational approach useful to study the entire phase diagram of macromolecules, and to characterize the effect of pressure and temperature on the molecular mechanisms of protein unfolding.

The urea-induced unfolding of proteins has been extensively studied by Dagget and colleagues using standard and replica exchange MD (14,47). The unfolding of chymotrypsin inhibitor 2 in 8 M urea at 333 K showed the complete disruption of the β-structure, while the protein maintained a residual native helical structure (14). The relative stability of helical motives reported by these authors had already been described by Tirado-Rives et al. for barnase (11), as discussed above. Similarly, Zhou et al. (48) performed a very long MD simulation of hen egg-white lysozyme and of its mutant W62G in 8 M urea at different pH values. Both the native and the mutant forms showed the persistence of some α-helical structure during the whole simulation time.

These partially conflicting data on the different sensitivity of helices and sheets to urea and/or temperature suggest that the molecular mechanism of unfolding can be peptide-specific, and/or that different force fields for protein and urea parameterization only partially describe their molecular properties. Eleftheriou et al. (49) simulated the thermal denaturation of hen egg-white lysozyme and of its mutant W62G with the OPLSAA and the CHARMM force fields. Despite their conclusion that the two force fields produced qualitatively similar results, the simulation based on OPLSAA at 500 K and the one based on CHARMM at 400 K showed roughly the same RMSD values, suggesting that different force fields can describe different potential energy surfaces for the same molecular entity. Furthermore, the protein unfolded with OPLSAA presented some nonnative short β-strands, while the same macromolecule unfolded with CHARMM showed more of a random coil.

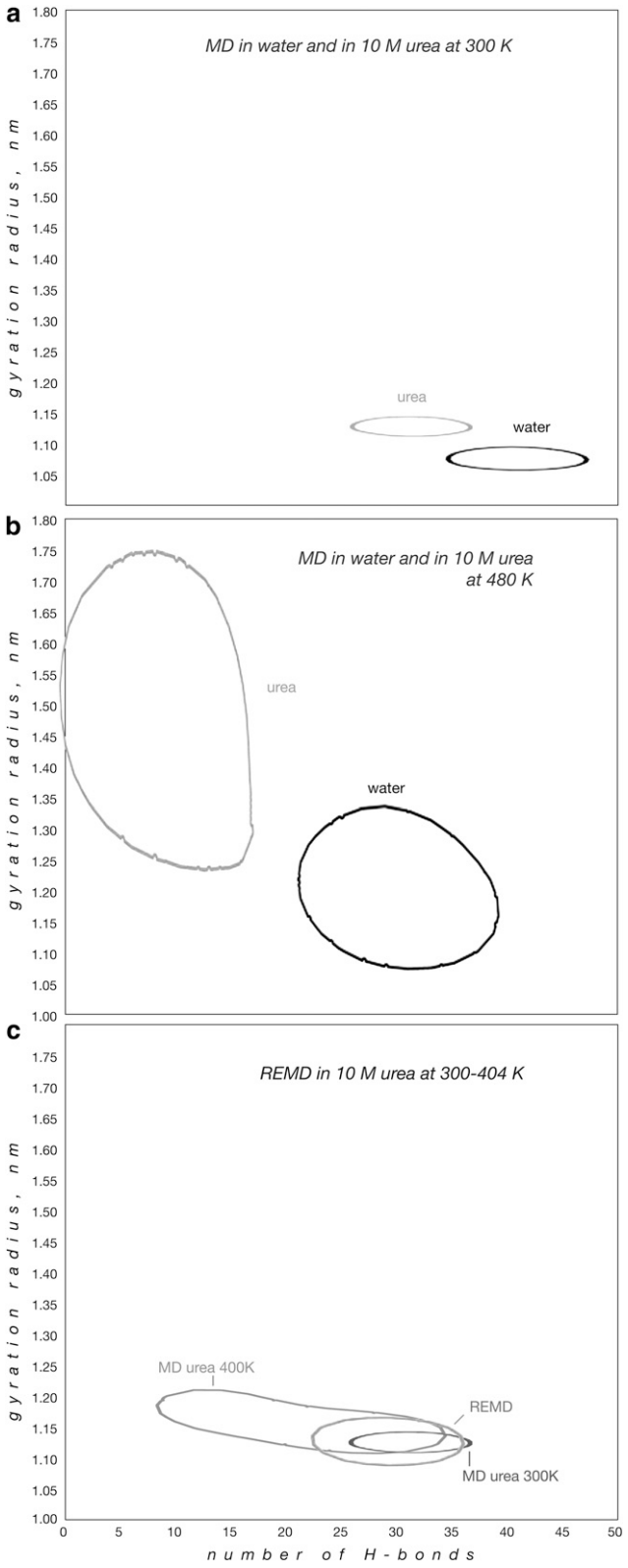

FIGURE 5 Density two-dimensional plots for number of hydrogen bonds and gyration radius. (*a* and *b*) Plots for the MD simulations in water and 10 M urea at 400 K (*a*) and 480 K (*b*). (*c*) Plot for the REMD simulations in 10 M urea at 300–404 K and standard MD at 300 and 400 K. In all plots the enclosed areas represent 80% of the cumulative probability.

REMD simulations

To obtain a representative trajectory of the whole conformational space sampled in the urea-driven unfolding of Protein L, all the trajectories of the replicas simulated in 10 M urea between 300 and 404 K were concatenated. From this collection of states, a set of clustered structures was obtained, using hierarchical clustering with the average linkage method (50), based upon a backbone RMSD threshold value of 0.2 nm. A total of 125 different average structures (representative of each cluster) were extracted disregarding fully unfolded states. Among these, the representative 76 conformations observed $\geq$2 times were classified according to their secondary structure score: helical and sheet content were represented by the percentage of residues retaining their native structure. The threshold for structure differentiation was set at 10% of the structural content difference. Approximately one-third of the examined structures were characterized by an equal amount of β-sheet and α-helix. Conversely, distribution of structures with dominant helix or sheet content was highly asymmetrical: α-helical prevailed in $\sim$20% of the conformations, β-sheet in $\sim$50%. Fig. 6 shows two statistically relevant conformations (both have been detected more than three times, hence their occurrence cannot be considered a rare event) corresponding to the extremes of either structural pattern. Fig. 6 *a* represents a structure with a poorly defined β-sheet and a certain amount of helical content; Fig. 6 *b* represents a structure with no residual helical content and an almost intact extended β-sheet motif.

REMD data confirm that the full configurational space of Protein L unfolding can be viewed as composed by distinct but parallel pathways: one entailing the unfolding of the α-helix before the involvement of the β-sheet, the other taking the opposite path. These configurations are populated to a very different extent, the mostly-β-sheet structure being $\sim$2.5 times more frequent than the one with dominant helical content. This observation proves the statistical relevance of one of the main findings of standard MD, namely the coexistence of two distinct pathways within the full energy

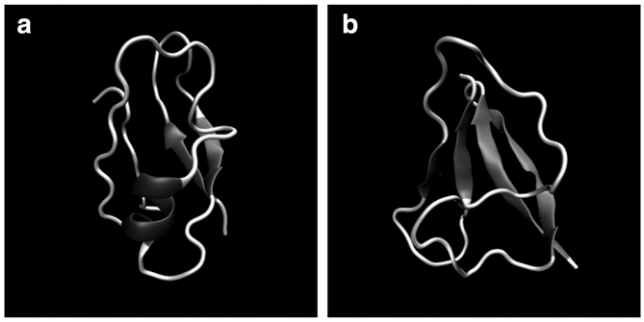

FIGURE 6 Representative conformations of Protein L from the REMD simulation. (*a*) A structure with a poorly defined β-sheet and a certain amount of helical content. (*b*) A structure with no residual helical content and an almost intact extended β-sheet motif.

landscape of Protein L unfolding. In agreement with Baker et al. (21), in all the structures characterized by the presence of a helical element and almost no β-sheet, the α-helix stretch consistently includes residues 27–33. This computational finding confirms the relevance of this region for the correct folding of Protein L: moreover, this finding is in good agreement with the work of Bennion et al. (14) about the presence of a residual α-helical motif in chymotrypsin inhibitor 2 in 8 M urea.

Again, a snapshot of the whole energetic surface accessible during the early events of protein unfolding was obtained plotting all of the above structures as a function of hydrogen-bond number and gyration radius. The resulting density plot in Fig. 5 *c* (*green*) is overlaid to the 80% probability areas for the standard MD simulation at 300 K in 10 M urea (*blue*) and at 400 K (*red*).

As suggested by Pascheck et al. (39), the free energy for urea-driven Protein L unfolding can be computed using the relationship for each replica of

$$\Delta G = -RT \ln K_{\mathrm{eq}} = -RT \ln \frac{1-X}{X}, \qquad (4)$$

in which X represents the reaction coordinate of the process. According to the cited reference, at each temperature the X variable is set equal to the percentage of residues that retain their native fold. Hence, a van 't Hoff plot can be obtained from the data of all the replicas, in which the value $-R \ln K_{\mathrm{eq}}$ is plotted versus $\frac{1}{T}$. The van 't Hoff plot shown in Fig. 7 is only referred to the α-helical content of the protein; it allows the evaluation of the thermodynamic parameters for a limited region of the system under investigation. In Fig. 7, the values of $-R \ln K_{\mathrm{eq}}$ computed with the method presented above are reported as a function of $1/T$. Despite the scatter of the data, a linear trend can be identified: in this way, the enthalpy and the entropy of the reaction can be computed under the assumption that in the range of investigated temperatures the enthalpy shows no dependence from temperature. The slope of the trend line for this plot and its y-axis intercept hence provide $\Delta H = -13.8$ kJ/mol and $\Delta S = -45.6$ J/mol for Protein L unfolding. Interestingly, using the percentage of folded residues of the whole protein as the reaction coordinate (instead of the percentage of folded residues of the helix), the computed thermodynamic parameters are very similar ($\Delta H = -9.3$ kJ/mol and $\Delta S = -40.7$ J/mol, data not shown). On the contrary, no trend can be identified using only the β-sheet structural elements (not shown). Due to the presence of too many effects that cannot be separated, a simple two-state model could be too simple for a correct interpretation of the overall unfolding mechanism of the protein, hence data based on the percentage of the whole set of structured residues could be misleading about the process thermodynamics. In the following, only data relative to the α-helical region will be considered.

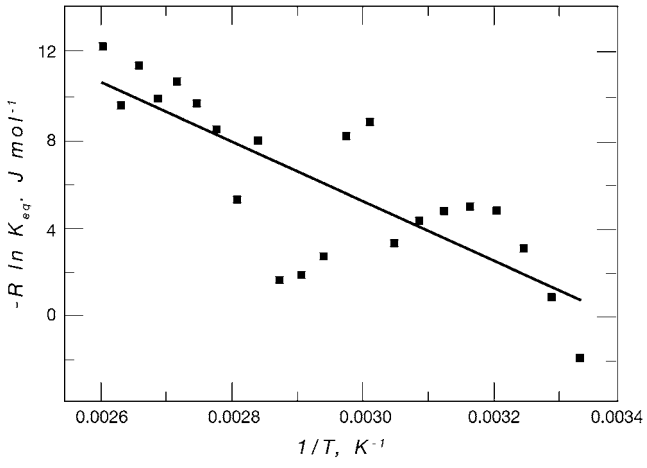

FIGURE 7 Evaluation of thermodynamic parameters for Protein L unfolding in 10 M urea through a van 't Hoff regression analysis, in which ΔG, obtained from all the REMD data, is plotted versus $\frac{1}{T}$.

The value computed for ΔH suggests an exothermic character of the urea-driven unfolding of Protein L: as reported elsewhere (51), an exothermic character could be to some extent explained by the release of heat associated to the interaction between protein (specifically, its hydrophobic side chains, (52)) and urea molecules. This effect very often is masked in standard calorimetric experiments by the strongly endothermic character of the dilution of the denaturant. Our simulations are performed at constant urea concentration without interference from a dilution process. The exothermic character of unfolding has already been observed for some other proteins in the past (51,53). Moreover, the melting temperature value computed with the proposed model is ~302 K, in agreement with the observation that around room temperature the urea-driven unfolding reaction occurs spontaneously without having to provide any heat from an external source. Interestingly, the variation of entropy is negative: this finding, taken together with enthalpy variation, suggests that in urea the high temperature could have the ability to partially stabilize the α-helical structure. The linearity of the plot implies a small change in heat capacity; in comparison to Pascheck et al. (39), we make a wider approximation by considering the heat capacity of the unfolding reaction constant and this possibly results in the computation of thermodynamic quantities with a lower degree of accuracy. Despite this approximation, the thermodynamic parameters we obtain are comparable to some experimental (54,55) and theoretical (39) data recorded for other small proteins. The absolute values of ΔH and ΔS are similar to the results of the theoretical treatment for the C-terminal fragment of Protein G (39). Indeed, sequence length and peptide fold are similar although a different solvent is considered. For Protein G, the entropic contribution appears higher, indicating a stronger contribution to the unfolding by the hydrophobic effect; conversely, enthalpy is comparable to Protein L unfolding, but changed in sign due to the presence of a chaotropic agent. As suggested by Paschek and Garcia (39), the computed parameters can indeed differ from the experi-

mental values: nevertheless, the residual enthalpy values we compute ~ -0.5 kJ/mol $\times$ res^{-1} are similar to (or of the same order of magnitude as) the ones obtained experimentally for other proteins with a similar helical fold, such as Naf-BBL (55) and hbSBD (56), despite different experimental conditions.

The comparison between the free-energy landscapes explored during standard and replica-exchange MD suggests that, for Protein L, standard simulations achieve a sampling efficiency similar to REMD (Fig. 5 *c*), in contrast with typical reports on MD versus REMD assessments (57). For this reason, our REMD simulation seems to be more useful to give a statistical validation of standard MD data rather than to further explore the energy landscape between 300 and 400 K. In a recent article, Beck et al. (58) compared data obtained from protein unfolding simulated both by REMD and by conventional MD. The authors showed that conventional MD can provide a more reliable estimate of the protein melting temperature than REMD simulations for much less than half of the computational cost. Furthermore, only the trajectory obtained by conventional MD can be used to describe conformational transitions and to analyze their kinetics. The same authors suggested also some reasons for the superior computational efficiency and ability to reproduce experimental data of conventional MD. REMD is preferable to overcome dynamical bottlenecks, which, however, are relevant only at low temperatures: for this reason, a set of conventional simulations at different temperatures can provide comparable or better results than REMD.

In summary, standard MD shows that, in the configurational space of Protein L, unfolding is accessible through different pathways, although to a different relative extent. In 30 ns of simulation time, we could observe the complete unfolding of Protein L only at high temperature. Further investigations, supported by more extensive computing resources, shall allow observing the urea-driven unfolding process without the influence of high temperature.

SUPPLEMENTARY MATERIAL

To view all of the supplemental files associated with this article, visit www.biophysj.org.

The authors are grateful to Guido Tiana and Giorgio Colombo for fruitful discussion.

This investigation was supported in part by a grant from Ministero dell'Università e della Ricerca (FIRB No. 2003: ''Il riconoscimento molecolare nelle interazioni proteina-ligando, proteina-proteina e proteina-superficie: Sviluppo di approcci sperimentali e computazionali integrati per lo studio di sistemi di interesse farmaceutico''). A.G.R. and P.R. were recipients of fellowships from the Ministero dell'Università e della Ricerca grant.

REFERENCES

1. Temussi, P. A., L. Masino, and A. Pastore. 2003. From Alzheimer to Huntington: why is a structural understanding so difficult? *EMBO J.* 22:355–361.
2. Chiti, F., and C. M. Dobson. 2006. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* 75:333–366.
3. Chaudhuri, T. K., and S. Paul. 2006. Protein-misfolding diseases and chaperone-based therapeutic approaches. *FEBS J.* 273:1331–1349.
4. Gregersen, N. 2006. Protein misfolding disorders: pathogenesis and intervention. *J. Inherit. Metab. Dis.* 29:456–470.
5. Fedorov, A. N., and T. O. Baldwin. 1997. Cotranslational protein folding. *J. Biol. Chem.* 272:32715–32718.
6. Tanford, C. 1968. Protein denaturation. Part A; Part B: the transition from native to denatured state. *Adv. Protein Chem.* 23:121–282.
7. Tanford, C. 1969. Protein denaturation. Part C: theoretical models for the mechanism of denaturation. *Adv. Protein Chem.* 24:1–95.
8. Polverini, E., M. Fornabaio, A. Fasano, G. Carlone, P. Riccio, and P. Cavatorta. 2006. The pH-dependent unfolding mechanism of P2 myelin protein: an experimental and computational study. *J. Struct. Biol.* 153:253–263.
9. Pang, J., and R. K. Allemann. 2007. Molecular dynamics simulation of thermal unfolding of *Thermatoga maritima* DHFR. *Phys. Chem. Chem. Phys.* 9:711–718.
10. O'Brien, E. P., R. I. Dima, B. Brooks, and D. Thirumalai. 2007. Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: lessons for protein denaturation mechanism. *J. Am. Chem. Soc.* 129:7346–7353.
11. Tirado-Rives, J., M. Orozco, and W. L. Jorgensen. 1997. Molecular dynamics simulations of the unfolding of barnase in water and 8 M aqueous urea. *Biochemistry.* 36:7313–7329.
12. Caflisch, A., and M. Karplus. 1999. Structural details of urea binding to barnase: a molecular dynamics analysis. *Structure.* 7:477–488.
13. Zhang, Z., Y. Zhu, and Y. Shi. 2001. Molecular dynamics simulations of urea and thermal-induced denaturation of S-peptide analogue. *Biophys. Chem.* 89:145–162.
14. Bennion, B. J., and V. Daggett. 2003. The molecular basis for the chemical denaturation of proteins by urea. *Proc. Natl. Acad. Sci. USA.* 100: 5142–5147.
15. Bennion, B. J., and V. Daggett. 2004. Counteraction of urea-induced protein denaturation by trimethylamine N-oxide: a chemical chaperone at atomic resolution. *Proc. Natl. Acad. Sci. USA.* 101:6433–6438.
16. Caballero-Herrera, A., K. Nordstrand, K. D. Berndt, and L. Nilsson. 2005. Effect of urea on peptide conformation in water: molecular dynamics and experimental characterization. *Biophys. J.* 89:842–857.
17. Smith, L. J., R. M. Jones, and W. F. van Gunsteren. 2005. Characterization of the denaturation of human α-lactalbumin in urea by molecular dynamics simulations. *Proteins.* 58:439–449.
18. Scalley, M. L., Q. Yi, H. Gu, A. McCormack, J. R. Yates 3rd, and D. Baker. 1997. Kinetics of folding of the IgG binding domain of peptostreptococcal protein L. *Biochemistry.* 36:3373–3382.
19. Plaxco, K. W., K. T. Simons, and D. Baker. 1998. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* 277:985–994.
20. Kim, D. E., C. Fisher, and D. Baker. 2000. A breakdown of symmetry in the folding transition state of protein L. *J. Mol. Biol.* 298:971–984.
21. Yi, Q., M. L. Scalley-Kim, E. J. Alm, and D. Baker. 2000. NMR characterization of residual structure in the denatured state of protein L. *J. Mol. Biol.* 299:1341–1351.
22. Sugita, Y., and Y. Okamoto. 1999. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314:141–151.
23. Tesi, M. C., E. J. J. van Rensburg, E. Orlandini, and S. G. Whittington. 1996. Monte Carlo study of the interacting self-avoiding walk model in three dimensions. *J. Stat. Phys.* 82:155–181.
24. Mitsutake, A., Y. Sugita, and Y. Okamoto. 2001. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers.* 60: 96–123.
25. Seibert, M. M., A. Patriksson, B. Hess, and D. van der Spoel. 2005. Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. *J. Mol. Biol.* 354:173–183.

26. Juraszek, J., and P. G. Bolhuis. 2006. Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proc. Natl. Acad. Sci. USA.* 103:15859–15864.
27. Jang, S., E. Kim, and Y. Pak. 2006. Direct folding simulation of α-helices and β-hairpins based on a single all-atom force field with an implicit solvation model. *Proteins.* 66:53–60.
28. Cecchini, M., F. Rao, M. Seeber, and A. Caflisch. 2004. Replica exchange molecular dynamics simulations of amyloid peptide aggregation. *J. Chem. Phys.* 121:10748–10756.
29. Smith, L. J., H. J. Berendsen, and W. F. van Gunsteren. 2004. Computer simulation of urea-water mixtures: a test of force field parameters for use in biomolecular simulation. *J. Phys. Chem. B.* 108:1065–1071.
30. Huang, K. 1987. Statistical Mechanics. Wiley, New York.
31. Berendsen, H. J. C., J. P. M. Postma, A. DiNola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
32. Berendsen, H. J. C., D. van der Spoel, and R. van Drunen. 1995. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91:43–56.
33. Lindahl, E., B. Hess, and D. van der Spoel. 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* 7:306–317.
34. van der Spoel, D., E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen, and H. J. C. Berendsen. 2006. GROMACS User Manual, V. 3.3. Department of Biophysical Chemistry, University of Groningen, Groningen, The Netherlands.
35. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
36. Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. 1995. A smooth particle mesh Ewald potential. *J. Chem. Phys.* 103:8577–8592.
37. Zhang, W., C. Wu, and Y. Duan. 2005. Convergence of replica exchange molecular dynamics. *J. Chem. Phys.* 123:154105.
38. Sugita, Y., and Y. Okamoto. 2000. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chem. Phys. Lett.* 329:261–270.
39. Paschek, D., and A. E. Garcia. 2004. Reversible temperature and pressure denaturation of a protein fragment: a replica exchange molecular dynamics simulation study. *Phys. Rev. Lett.* 93:238105.
40. Kawahara, K., and C. Tanford. 1966. Viscosity and density of aqueous solutions of urea and guanidine hydrochloride. *J. Biol. Chem.* 241: 3228–3232.
41. Lide, D. R., editor. 2000. Handbook of Chemistry and Physics, 81st Ed. CRC Press, Boca Raton, FL.
42. Tobi, D., R. Elber, and D. Thirumalai. 2003. The dominant interaction between peptide and urea is electrostatic in nature: a molecular dynamics simulation study. *Biopolymers.* 68:359–369.
43. Eisenberg, D., and A. D. McLachlan. 1986. Solvation energy in protein folding and binding. *Nature.* 319:199–203.
44. Eisenhaber, F., P. Lijnzaad, P. Argos, C. Sander, and M. Scharf. 1995. The double cube lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.* 16:273–284.
45. Garcia, A. E., and J. N. Onuchic. 2003. Folding a protein in a computer: an atomic description of the folding/unfolding of protein A. *Proc. Natl. Acad. Sci. USA.* 100:13898–13903.
46. Paschek, D., S. Gnanakaran, and A. E. Garcia. 2005. Simulations of the pressure and temperature unfolding of an α-helical peptide. *Proc. Natl. Acad. Sci. USA.* 102:6765–6770.
47. Day, R., and V. Daggett. 2005. Sensitivity of the folding/unfolding transition state ensemble of chymotrypsin inhibitor 2 to changes in temperature and solvent. *Protein Sci.* 14:1242–1252.
48. Zhou, R., M. Eleftheriou, A. K. Royyuru, and B. J. Berne. 2007. Destruction of long-range interactions by a single mutation in lysozyme. *Proc. Natl. Acad. Sci. USA.* 104:5824–5829.
49. Eleftheriou, M., R. S. Germain, A. K. Royyuru, and R. Zhou. 2006. Thermal denaturing of mutant lysozyme with both the OPLSAA and the CHARMM force fields. *J. Am. Chem. Soc.* 128:13388–13395.
50. Jain, A., and R. Dubes. 1988. Algorithms for Clustering Data. Prentice Hall, Englewood Cliffs, NJ.
51. Zolkiewski, M., N. J. Nosworthy, and A. Ginsburg. 1995. Urea-induced dissociation and unfolding of dodecameric glutamine synthetase from *Escherichia coli*: calorimetric and spectral studies. *Protein Sci.* 4:1544–1552.
52. Scholtz, J. M., D. Barrick, E. J. York, J. M. Stewart, and R. L. Baldwin. 1995. Urea unfolding of peptide helices as a model for interpreting protein unfolding. *Proc. Natl. Acad. Sci. USA.* 92:185–189.
53. Andrade, M. I., M. N. Jones, and H. A. Skinner. 1976. The enthalpy of interaction of urea with some globular proteins. *Eur. J. Biochem.* 66:127–131.
54. Robic, S., M. Guzman-Casado, J. M. Sanchez-Ruiz, and S. Marqusee. 2003. Role of residual structure in the unfolded state of a thermophilic protein. *Proc. Natl. Acad. Sci. USA.* 100:11345–11349.
55. Sadqi, M., D. Fushman, and V. Munoz. 2006. Atom-by-atom analysis of global downhill protein folding. *Nature.* 442:317–321.
56. Kouza, M., C. F. Chang, S. Hayryan, T. H. Yu, M. S. Li, T. H. Huang, and C. K. Hu. 2005. Folding of the protein domain hbSBD. *Biophys. J.* 89:3353–3361.
57. Rao, F., and A. Caflisch. 2003. Replica exchange molecular dynamics simulations of reversible folding. *J. Chem. Phys.* 119:4035–4042.
58. Beck, D. A., G. W. White, and V. Daggett. 2007. Exploring the energy landscape of protein folding using replica-exchange and conventional molecular dynamics simulations. *J. Struct. Biol.* 157:514–523.
59. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 12:2577–2637.