Simulating protein unfolding under pressure with a coarse-grained model

Ramiro Perezzan and Antonio Rey

Citation: J. Chem. Phys. **137**, 185102 (2012); doi: 10.1063/1.4765057
View online: http://dx.doi.org/10.1063/1.4765057
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v137/i18
Published by the American Institute of Physics.

Simulating protein unfolding under pressure with a coarse-grained model

Ramiro Perezzan and Antonio Rey[a)]
Departamento de Química Física I, Facultad de Ciencias Químicas, Universidad Complutense, E-28040 Madrid, Spain

(Received 28 August 2012; accepted 17 October 2012; published online 9 November 2012)

We describe and test a coarse-grained molecular model for the simulation of the effects of pressure on the folding/unfolding transition of proteins. The model is a structure-based one, which takes into account the desolvation barrier for the formation of the native contacts. The pressure is taken into account in a qualitative, mean field approach, acting on the parameters describing the native stabilizing interactions. The model has been tested by simulating the thermodynamic and structural behavior of protein GB1 with a parallel tempering Monte Carlo algorithm. At low effective pressures, the model reproduces the standard two-state thermal transition between the native and denatured states. However, at large pressures a new state appears. Its structural characteristics have been analyzed, showing that it corresponds to a swollen version of the native structure. This swollen state is at equilibrium with the native state at low temperatures, but gradually transforms into the thermally denatured state as temperature is increased. Therefore, our model predicts a downhill transition between the swollen and the denatured states. The analysis of the model permits us to obtain a phase diagram for the pressure-temperature behavior of the simulated system, which is compatible with the known elliptical shape of this diagram for real proteins. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4765057]

I. INTRODUCTION

From the beginning of the 20th century, it is known the effect of pressure on the native conformation of proteins. A hydrostatic pressure of about 3 kbar can unfold a protein,[1] and therefore it was thought that high pressures can be somehow similar to the effect of temperature on the stability of these biopolymers.[1] Nevertheless, the difficult task to carry out accurate experimental measurements at high pressures, especially in a biological environment, has made the pressure effects to be traditionally less known and understood than other properties which also denature natural proteins, as temperature, pH or chemical agents. In the last decade, however, the continuous development of biophysical techniques has led several research groups to pursue the study of pressure effects on proteins, from different perspectives. On one hand, there is a fundamental need to understand the interactions responsible for the stability of folded proteins,[2,3] and the pressure represents in this sense a novel tool, which provides new insights into the study of the folding funnel. These may result from the speed increment in the transitions between the folded and denatured states,[4,5] or from the opening of folding pathways which are energetically unaccessible at room pressure.[6] On the other hand, there are also several practical applications of pressure in the fields of pharmacology,[7] food industry,[8,9] or even medicine,[10] since high pressures can inhibit protein aggregation[11] related to the development of neurodegenerative diseases such as Alzheimer or Parkinson.[12,13]

The pressure-temperature behavior for the stability of folded proteins has been known for some years, and has an elliptical shape in many cases.[14] The situation may be different in some occasions, depending specially on the values of the system compressibility,[15] but the ellipsoidal phase-diagram is considered rather general. The folded state occupies the internal area of the ellipse, and can be lost by increasing (or decreasing) temperature, pressure, or a combination of both. It is very important to mention that, although the unfolded state is sometimes considered as "everything which is not native," both experimental and numerical simulation studies have shown that there are differences between protein unfolded states resulting from a pressure increase or from a temperature increase.[3,16] For single domain proteins, the temperature unfolding process is considered to be cooperative for many proteins, and both the tertiary and the secondary structures of the native conformation are essentially lost in the unfolded state. On the other hand, a pressure unfolded protein shows a larger radius of gyration than the native state, but the secondary structure, or at least a substantial fraction of it, is preserved. Therefore, this unfolded state keeps a certain globular shape, but it is swollen in comparison to the native conformation.[17]

According to recent discoveries,[3,6,13] an increase in the hydrostatic pressure of an aqueous protein solution produces an effective force which tends to inject the solvent molecules inside the protein internal cavities and void volumes. From that point of view, proteins with more or larger cavities tend to be more destabilized by the effect of pressure.[18] Sometimes, this has been considered as a softening in the interactions responsible for the stability of the folded conformation,[19] mainly the hydrophobic interactions.[20] As an alternative, purely thermodynamic, point of view, it has been realized that the swollen system resulting from this injection, considering

a)E-mail: jsbach@quim.ucm.es.

the protein and the solvent altogether,[21] has a smaller partial molar volume than the "dry-core" folded protein surrounded by water,[15,22] and therefore an increase in pressure produces a denaturation of the protein (see also Ref. 18 and references therein).

Given this situation, computational models which use molecular simulations to analyze the effect of pressure on protein stability have mostly used detailed, all-atom models with explicit solvent and molecular dynamics techniques.[18,23–25] These models, as accurate as they are, may be also too detailed and computationally costly to properly consider (i.e., with enough statistical accuracy) processes that happen in long time scales, as the full folding of a protein.

Coarse-grained models, on the other hand, allow to explore in a reasonable computational time the folding/unfolding equilibrium of a protein.[26] The loss of the atomic details in the simulation results becomes, in this sense, an affordable compromise, especially when the simple model is well defined for the specific purpose of a given investigation. Although the pressure unfolding of proteins, as described in the paragraph above related to cavities, seems to be incompatible with molecular models where the packing details have to be carefully considered,[18] it is still possible to obtain interesting physical insights of this process by using reduced representations of the system and mean field potentials. They have been commonly used, with different levels of approach, to describe the stabilizing interactions responsible for a protein native conformation,[27–29] mainly the hydrophobic interactions and their relation to the solvation of the hydrophobic residues,[30,31] with the solvent being implicitly considered. The resulting potential of mean force between two hydrophobic residues, obtained when the degrees of freedom of the surrounding water are integrated out, shows two attractive basins: one, at short distances, corresponds to the direct contact between the interacting units; another one, at a larger distance, represents the interaction between the same units mediated by a single water molecule. These two minima appear at a distance of ∼3 Å, which roughly corresponds to the diameter of a water molecule. The minima are separated by a region of positive potential energy, commonly named *desolvation barrier*, which represents the energy needed to expel the last water molecule in the approaching of the considered hydrophobic units.[32]

Although most of these approximate potentials have been developed to study protein folding at room pressure, some results have also arisen showing how they are influenced by pressure increases up to several thousand bars.[33] Essentially, the pressure induces changes in the relative depth of the minima and the height of the barrier.[25,34] Therefore, it seems natural as a first approach, at the level of a coarse-grained model, to simulate the effect of pressure by a change in the different contributions to the potential of mean force. In this work we show the development of this type of model. We have used the ideas mentioned above to design a structure-based model[35,36] whose stabilizing interactions (among residues which form a contact in the native structure) are made pressure-dependent through the strategy just described. Of course, we cannot establish a real pressure scale, but we will be able to gradually increase both pressure and temperature to analyze the characteristics of the resulting folding/unfolding transition. Since in this work we are interested in its thermodynamic and structural features, we have used a Monte Carlo sampling procedure, which allows us a simple mathematical definition of the interaction potential described (we do not need the potential to be differentiable, as it happens in molecular or stochastic dynamics simulations). The model, as required, is numerically efficient, and has allowed us to analyze in detail the influence of pressure on the stability of the native state, and to find pressure unfolded states different from those resulting from thermal denaturation.

II. METHODS

A. The model

In this work, we use an off-lattice coarse-grained model, where every residue in the polypeptide chain is represented by a single interacting unit, centered at its α carbon position. The distance between bonded residues is kept constant at a value of 3.8 Å.

Since we are using a structure-based model, we need to know the contact map of the native conformation of the selected protein in order to define the model interactions. We consider that all i–j interactions with $|i - j| = 2$ or 3 define native contacts, as they correspond to virtual bond and torsion angles in the model. For residues with $|i - j| \geq 4$, a native contact is considered when, in the protein conformation taken from the Protein Data Bank (PDB),[37] the shortest distance between any heavy atom belonging to either residue is less than 4.5 Å. In previous work from our group, where folding was studied at room pressure as a function of temperature alone,[38–40] a simple attractive potential with the mathematical form of a truncated harmonic well was defined for this interaction. Now, we are going to use a more complex potential, to include what has been called[29] an additional *solvent separated minimum* (ssm), and a *desolvation barrier* (db). The resulting potential is shown in Figure 1. To keep the model as simple as possible, all the individual curves correspond to harmonic potentials. The attractive well at short distances, which we call

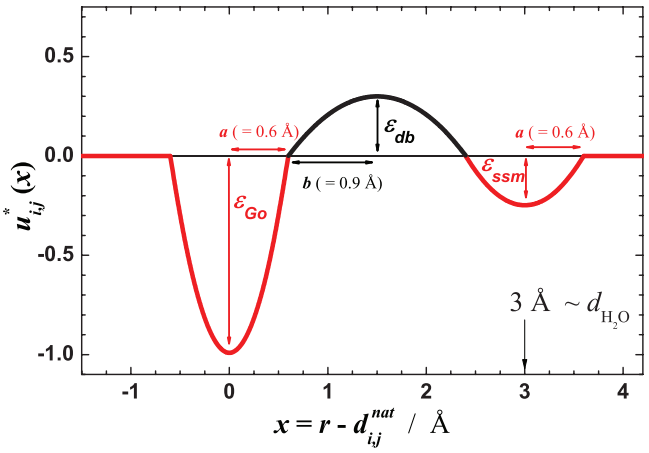

FIG. 1. Interaction potential between pairs of residues that form a native contact in the PDB structure. The short range repulsion is not shown.

the *Go minimum* (from the seminal work of Gō on the use of structure-based models[35,36]), corresponds to the direct contact between two residues. It has a minimum of depth $-\epsilon_{Go}$ centered at the native distance between the α carbons of the interacting residues in the native conformation. The second attractive well, centered at a distance increased in 3 Å for the same interacting pair, has a depth $-\epsilon_{ssm}$, and the same width as the Go minimum. The desolvation barrier in between has a height ϵ_{db}, and its width is defined so that the full potential is continuous. All the harmonic curves are truncated where the potential becomes null. Therefore, the mathematical definition for the interaction potential in our model for a pair of residues ij, located at a distance r, which are in contact in the native conformation at a distance d_{ij}^{nat}, is

$$u_{ij}(r) = \begin{cases} \frac{\epsilon_{Go}}{a^2}\left[\left(r - d_{ij}^{nat}\right)^2 - a^2\right], & d_{ij}^{nat} - a < r \le d_{ij}^{nat} + a \\ -\frac{\epsilon_{db}}{b^2}\left[\left(r - d_{ij}^{nat} - a - b\right)^2 - b^2\right], & d_{ij}^{nat} + a < r \le d_{ij}^{nat} + 2b + a \\ \frac{\epsilon_{ssm}}{a^2}\left[\left(r - d_{ij}^{nat} - 2a - 2b\right)^2 - a^2\right], & d_{ij}^{nat} + a + 2b < r \le d_{ij}^{nat} + 3a + 2b \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Even though the global expression is a bit cumbersome, it is still very simple if compared to similar solvation potentials used in dynamic simulations.[29,32] The well width a in Eq. (1) and in Figure 1 has been taken as $a = 0.6$ Å, from our previous analysis of the influence of this value on the folding transition at room pressure.[38] The value of b, as previously mentioned, is then fixed at the value $b = 0.9$ Å to make the full potential continuous, given that the two minima for a given interacting pair are separated by a distance of 3 Å.

The expression above and the curve in Figure 1 do not show the repulsive contribution that keeps the excluded volume of the model. Since we are computing Monte Carlo simulations, we just use a hard-sphere potential that avoids any conformation where the distance between any pair of residues becomes less than 4.2 Å, a value previously adjusted in our group.[38] However, in a few cases we have found native contacts whose α-carbons become very close in the native state (mostly in β-sheets), close or even below that threshold. In these occasions, we shift the position of this repulsive barrier to $d_{ij}^{nat} - a$, to warrant that the full attractive basin of the Go minimum is included in the considered potential. It is interesting to mention that using a repulsive barrier which is mostly decoupled from the attractive interactions, something which is not possible with frequently used Lennard-Jones potentials, has been recently claimed as a clear improvement in structure-based potentials.[41,42] The repulsive interaction is the only one present between residues which do not form a native contact in the PDB conformation.

From the ideas about the dependence of the potential of mean force representing the hydrophobic interactions on pressure, it seems clear that we could represent a pressure increase by a relative stabilization of the ssm minimum with respect to the Go one. Although the desolvation barrier would be also affected,[25] in this first work dealing with thermodynamic and structural properties we have focused on the effect of the attractive basins. So, in all the results presented here, we have used $\epsilon_{Go} = 1$, which therefore corresponds to our energy unit, $\epsilon_{db} = 0.3$, and values of ϵ_{ssm} ranging from very small values as 0 or 0.25 (which would correspond to room pressure) up to 1.4. For the largest values of ϵ_{ssm}, the ssm minimum is energetically more stable than the Go one, which would favor conformations with a swollen volume, representing the injection of water molecules inside the hydrophobic protein core occurring at high pressure. We should mention that, in the original works about the dependence of hydrophobic interactions on pressure,[25,34] the ssm minimum never becomes as deep as the direct contact minimum is. In this work, we have had to go beyond that limit in order to reach the desired behavior. This is needed to cover the details left behind by the coarse-grained model.

There is an important factor, however, which has to be taken into account: not all the native contacts registered in a contact map are of hydrophobic nature alone. A clear example comes from residues which are close along the sequence ($|i - j| < 4$) and are therefore usually also close in space. Other cases would correspond to residues linked by hydrogen bonds, whose nature is rather different from the hydrophobic interaction. In those cases, a potential as that of Figure 1 would not be sensible at all. Therefore, in our model we have defined different classes of contacts, which are then subjected to modified versions of the potential. *Local contacts* are those between residues with $|i - j| < 4$. In these cases, only the Go minimum and the excluded volume are kept in the interaction energy (i.e., for them $\epsilon_{db} = \epsilon_{ssm} = 0$). The same happens to our *hydrogen bonded contacts*. If we detect the presence of a native backbone hydrogen bond between two residues in the PDB conformation (according to a simple model recently developed in our group[43,44]), we only keep the Go part in Eq. (1) and the short range repulsion. This helps to preserve the secondary structure elements in the pressure unfolded conformations (see below). Finally, the remaining native contacts are labeled as *long range contacts*, and the full potential of Eq. (1), plus the excluded volume term, are used for their interactions, with the fixed values of ϵ_{Go} and ϵ_{db} mentioned in the previous paragraph, and the different values of ϵ_{ssm} spanned in this work.

It is important to mention that, for a given protein, the use of the potential introduced here, with the different types of contacts, creates a certain frustration into the system as a function of pressure. At low pressures, when $\epsilon_{Go} \ll \epsilon_{ssm}$, the native conformation is a well defined global minimum for the system. However, when the two values become comparable, and moreover when $\epsilon_{ssm} > \epsilon_{Go}$, the situation is far more complicated. As we have said, local and hydrogen bond contacts would still try to stay at the Go well (the only one existing for these contacts at any pressure), while the individual long range contacts would be more stable if the distance between the pair of residues is enlarged until the ssm well is reached. But this is, in many cases, impossible without breaking the molecule, or even if this was allowed. Under these conditions, there is not a clear unique stable conformation for a chain molecule, and the system will have to find a compromise for the protein ensemble of conformations as a function of temperature and pressure (ϵ_{ssm}). This is one of the points that make the results presented in this work particularly interesting.

B. Simulation algorithm

As previously mentioned, to obtain efficient results of the thermodynamic and structural features of the transitions from the folded state as a function of both temperature and pressure, we have used a Monte Carlo procedure previously developed in our group, which has given us the possibility to study different proteins at room pressure.[38–40,43] In every simulation, pressure has been kept "constant" (i.e., the value of ϵ_{ssm} is not modified along every individual trajectory). The algorithm uses a parallel tempering simulation technique,[45] where several replicas of the system are simultaneously sampled at different temperatures, with occasional exchanges between replicas at contiguous temperatures. This way, we try to avoid the system to become trapped in local minima, something especially important given the additional frustration imposed by the potential definition in this work. The number of temperatures is adapted depending on this situation as a function of ϵ_{ssm}, and for the results presented here it has been between 42 and 50. At every simulation, starting from an extended conformation, we equilibrate the system for a number of 2×10^6 to 5×10^6 Monte Carlo cycles at every temperature, followed by 5×10^6 to 20×10^6 additional cycles for production. A Monte Carlo cycle involves N tries to move the N residues of our model polypeptide chain (the individual moves are described elsewhere[38]). For the protein considered here (see below), the calculation of a single parallel tempering trajectory (for a given ϵ_{ssm}) has taken between 8 and 12 CPU hours in a single processor. To ensure a proper statistical meaning of the simulation results, every full simulation is repeated 5–9 times independently, starting from different seed numbers. The results presented in Sec. III correspond to averages or accumulated results from all these independent simulations.

III. RESULTS AND DISCUSSION

As we did in our development of the structure based model at room pressure,[38] we have used as a study test the folding of the B1 domain of protein G from *Streptococcus*, whose PDB code is 2GB1.[46] It is a small protein, with $N = 56$ residues, with two β-hairpins associated into a β-sheet, which packs against an α-helix located in the middle of the sequence. The folding of this protein at room pressure happens as a cooperative, two-state process, which has been proved both experimentally[47,48] and computationally.[49,50] There are also some structural results for this protein at relatively high pressures[51] (about 2 kbar), where the native structure is started to be affected by the injection of a water molecules inside one of its cavities.

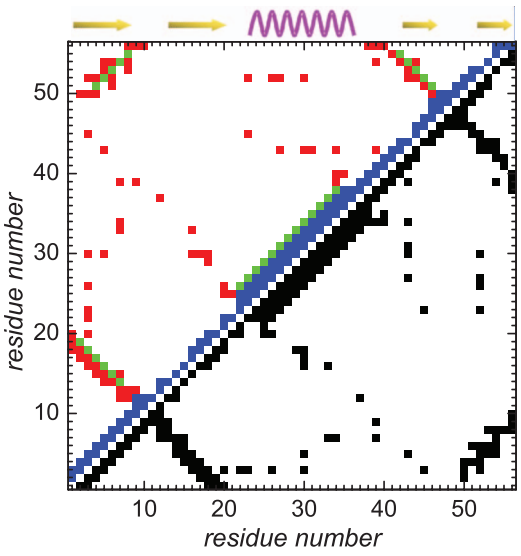

FIG. 2. Contact map for the GB1 protein. The lower-right triangle shows the native contacts, in black color, according to the cut-off used. The upper-left triangle shows the classification of the contacts: in green, *hydrogen bond contacts*; in blue, *local contacts*; and in red, *long range contacts*.

The contact map we have calculated for the PDB structure of this protein is shown in Figure 2. The lower-right triangle shows the contacts satisfying the cut-off criterion mentioned in Sec. II. In the upper-left triangle, we have colored the same set of contacts according to our classification as local (118 contacts, in blue), hydrogen-bonded (29, in green), or long range (79, in red), given that the interactions for them are different, as already described. The hydrogen-bond contacts have been located as i–$i + 4$ contacts in the helical region[44] (since i–$i + 3$ contacts are considered already as *local*). They also appear along the contacts among neighbor β-strands in the β-sheet.

As already mentioned, for this protein we have kept constant the values $\epsilon_{Go} = 1$ and $\epsilon_{db} = 0.3$, while ϵ_{ssm} has taken values of 0, 0.25, 0.5, 0.75, 1.0, 1.1, 1.2, 1.3, and 1.4, corresponding to pressures from room conditions to high values on the order of several thousand bars (since at these pressures experiments show that the native conformation is not stable any more for globular proteins[6]).

A. Folding/unfolding characteristics under pressure

From our simulation results, we can readily obtain the heat capacity curves as a function of temperature from the energy fluctuations of the system conformations sampled. These

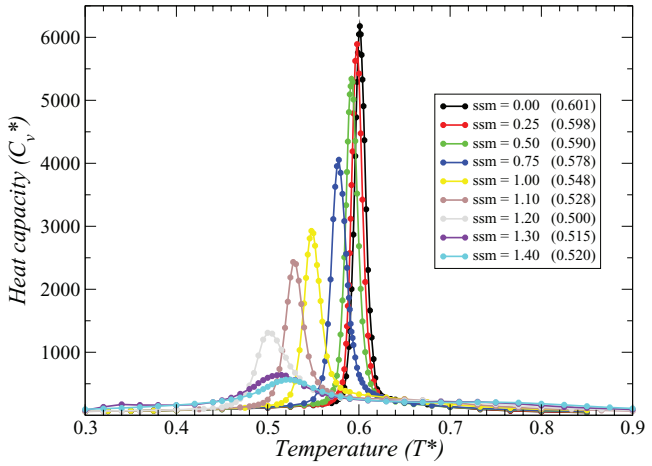

FIG. 3. Heat capacity curves as a function of temperature (both in reduced units), for the different values of ϵ_{ssm}, as indicated in the legend. The values in parentheses show the reduced temperature at the peak of each curve.

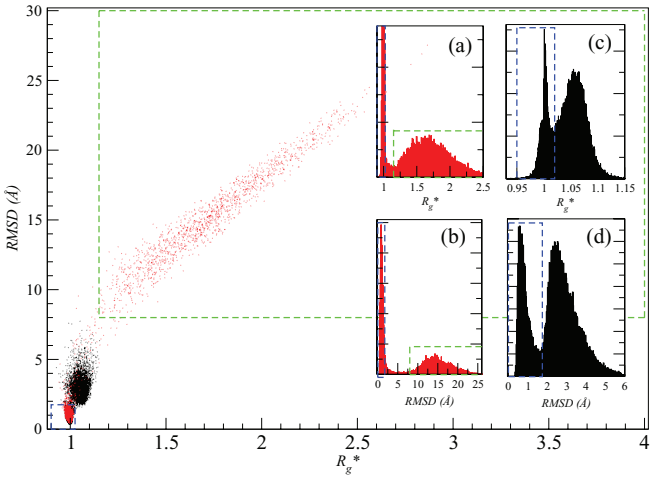

FIG. 4. Representation of RMSD against $R_g^* = R_g/R_g^{\text{PDB}}$ for the conformations recorded along the simulations with $\epsilon_{ssm} = 0.0$ (in red) and $\epsilon_{ssm} = 1.2$ (in black), at their corresponding transition temperatures (0.601 and 0.500, respectively). (a) and (b) Frequency histograms for RMSD and R_g^*, respectively, for $\epsilon_{ssm} = 0.0$. (c) and (d) Frequency histograms for RMSD and R_g^*, respectively, for $\epsilon_{ssm} = 1.2$. The dashed lines indicate areas selected to characterize different states (see text for details).

curves are presented in Figure 3, for the nine values of ϵ_{ssm} computed. In all the cases, the curves show a well defined absolute maximum, which is usually assigned to the transition temperature, T_m^* (the temperatures in this work are presented in reduced units, coherent with the consideration of ϵ_{Go} as our energy unit). In general, an increase in pressure leads to a smaller transition temperature. More importantly, the heat capacity curve becomes clearly wider (and therefore lower) as the pressure is increased. This implies a decrease on the cooperative features of the transition. In addition, for the highest pressures, the heat capacity curves are very broad and swallow, and the transition temperature increases now with ϵ_{ssm}, probably indicating a completely different type of transition. There are even several shoulders in these curves, indicating a richer structural (and energetic) landscape under these high pressure conditions.

To analyze in detail the different transitions observed as a function of pressure, we show in Figure 4 the behavior of one representative case at atmospheric pressure (specifically, that computed with $\epsilon_{ssm} = 0$, red results in the figure) and another one with rather high pressure ($\epsilon_{ssm} = 1.2$, black results in the figure). The full graph in Figure 4 shows the different conformations sampled in every case's transition temperature. Each individual conformation recorded (just a tiny fraction of all those sampled) is represented as a dot corresponding to the root mean square deviation (RMSD) between that conformation and the PDB one, in the vertical axis, and the radius of gyration R_g in the horizontal axis, which for the sake of clarity is represented as $R_g^* = R_g/R_g^{\text{PDB}}$ (in the PDB conformation, $R_g = 10.1$ Å, as computed from the α-carbons). At room pressure (red spots), the transition temperature shows a clear bimodal distribution of conformations. This is quantitatively reflected in the insets at Figures 4(a) and 4(b), where the distributions of both structural properties at T_m^* are shown. We can observe a narrow and intense peak, corresponding to folded conformations, with $R_g^* \cong 1$ and RMSD ~ 1 Å, just reflecting the small thermal fluctuations existing at this temperature in the native state. Then, there is also a wide distribution of conformations, with large RMSD values and a considerable growing in the protein size, as reflected by the ratio of the radii of gyration. They correspond to the unfolded conformations of the denatured state. Both states show a region between them, in any of the recorded structural properties, which is scarcely populated, indicating a high free energy barrier between the native and denatured states. Therefore, the transition at room pressure shows for the model the characteristic two-state features which correspond to this protein. The results for other small values of ϵ_{ssm} (data not shown) present a similar behavior.

On the other hand, when the pressure is raised up to high values, the behavior found is quite different. The black data in Figure 4, corresponding to the results found with $\epsilon_{ssm} = 1.2$ at its transition temperature, still show a bimodal distribution. Moreover, the characteristics of the native state are essentially the same found at room pressure. Actually, the black spots in the large graph of the figure just lie behind the cloud of red spots for small values of R_g^* and RMSD, and can be missed if not carefully observed. The histograms in the insets at Figures 4(c) and 4(d) show again the narrow peak corresponding to the conformations of the native state (notice the larger expanded horizontal scale in these graphs, as compared to those from insets (a) and (b)). The difference at this pressure comes then from the second cloud of points. It represents a distribution of conformations with relatively small values of RMSD (2–6 Å) and whose size is only 5%–10% larger than the native conformation, according to the radius of gyration. This situation looks as a relatively distorted, swollen version of the native state, with characteristics which are very distinct from the denatured state mentioned above at low pressure. The bimodal distributions still show characteristics of a two state transition, although the larger population existing in this case at intermediate values of RMSD and R_g^* are indicative of a small free energy barrier for this transition, in comparison to that existing between the native and the denatured states

at low pressures. Recent sophisticated experiments using very fast pressure jumps have shown that the transition from a pressure unfolded protein to its native state happens indeed at very high speeds.[52]

Therefore, we have found that the peak of the heat capacity curves reflects the equilibrium between the native, N, and the denatured, D, states at low pressures, but a different equilibrium between the native and a swollen, S, state at high pressures. The distributions, both monodimensional and bidimensional, shown in Figure 4 allow us also to define a range of structural parameters to characterize the conformations belonging to the different states. Thus, the blue dashed rectangles would frame values of RMSD and R_g^* characteristic of conformations belonging to the native state ($R_g^* < 1.025$ and RMSD < 1.75 Å), while the green rectangle would frame conformations belonging to the denatured state ($R_g^* > 1.15$ and RMSD > 8 Å). In a standard thermal two-state transition at room pressure, conformations outside these two states would show a negligible population at any temperature, which could be possibly useful to characterize the transition state ensemble. At high pressures, however, we have found a significant population in this region, with well defined characteristics, which corresponds to what we have named the swollen state, S, according to its global features.

We have checked that the boundaries for the N and D states are pretty consistent across the different temperatures and part of the range of pressures simulated (as far as these states are uniquely defined, see below). Therefore, we have used these ranges to analyze the relative populations of the different states appearing along our simulations. The results are shown as a function of temperature in Figure 5, for all the pressures (values of ϵ_{ssm}) considered in this work. The first row of plots corresponds to low or moderate pressures, where the two-state transition between N and D is the only relevant situation. The population of S is negligible or very small, and never becomes the larger population at any of the considered temperatures. The curves for the populations of N and D against T^* show the characteristic sigmoidal shape, with a very steep change at the transition region, showing also here the characteristics of a cooperative transition. In the second row of graphs we present intermediate values of ϵ_{ssm}, which would already correspond to high pressures. As it previously happened only to the N and D states, here also the population of the swollen state S becomes the major one, even the only significant one, in a certain range of temperatures, which grows broader the larger the pressure is. The transition N–S is still quite abrupt, but the transition between S and D shows curves with smoother variations. The situation becomes even more evident at the highest values of ϵ_{ssm}, corresponding to very high pressures, shown in the last row of plots in Figure 5. For $\epsilon_{ssm} > 1.2$ in our model, the native state population is only significant at very small temperatures, far from the interesting range. Therefore, the swollen state would be stable at the relevant small temperatures, gradually transforming into the denatured state as the temperature is raised.

These different characteristics of both transitions can be shown in a more evident way by analyzing the distribution of structural parameters R_g^* and RMSD as a function of temperature. This is shown, for $\epsilon_{ssm} = 1.2$, in Figure 6. The left column plots correspond to temperatures up to the transition temperature or slightly above it ($T_m^* = 0.5$ in these conditions). The plots at the right column, with scales spanning larger ranges, correspond to temperatures above T_m^* in all the cases. Below T_m^*, we clearly observe the conformations of the native state at low temperatures, and two well separated clouds when we come closer to the transition region, indicating an equilibrium between the N and S states, where the population shifts from N to S in a cooperative way as the temperature is risen. While the N state keeps its structural characteristics as the temperature changes, the distribution for the S states becomes

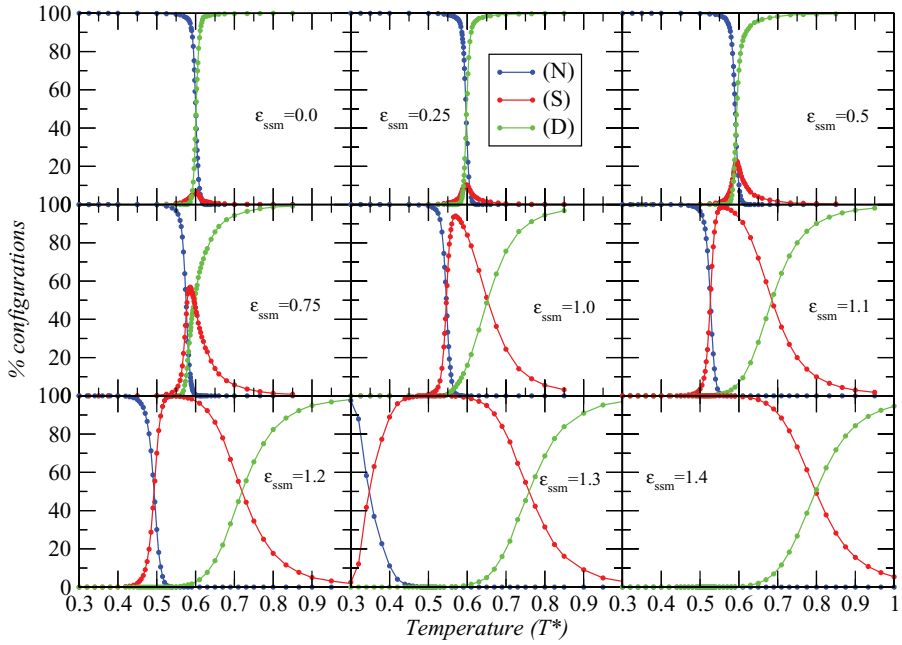

FIG. 5. Thermal evolution of the populations of the three defined states: native (N), swollen (S), and denatured (D) in terms of the value of ϵ_{ssm}.

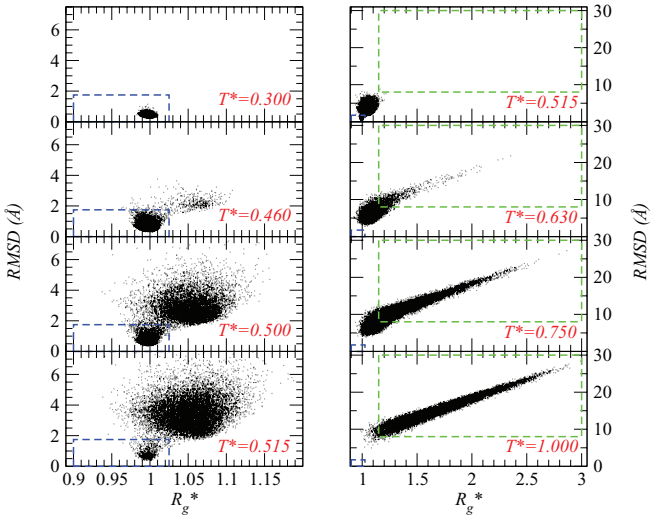

FIG. 6. Representation of RMSD against R_g^* for the conformations recorded in the simulations with $\epsilon_{ssm} = 1.2$ and different temperatures, as indicated.

broader the higher the temperature becomes. In the right column plots, we see that this feature is further enhanced. There is a fully continuous transition between the S and D states as the temperature increases. This implies that there is not a free energy barrier between these two states. We should then consider the S–D transition, appearing at high pressures, as a downhill transition, as those reported for some small proteins at room pressure.[53] Therefore, the population results shown in the high pressure conditions of Figure 5 have to be considered with a word of caution, since the S and D states cannot be strictly defined as independent cases. This explains the smooth behavior of the population for these "two states" already commented on.

B. Detailed analysis of the swollen state

In addition to the properties of the global system (the conformations as a whole) considered in Sec. III A, we have also carried out a finer analysis, considering the individual native contacts which may or may not appear in the conformations recorded along the simulations. We have focused here mostly on those belonging to the swollen state S, which is characteristic of the pressure influence on the folding/unfolding transition. For every native contact included in the contact map of this protein (Figure 2), we analyze in any conformation if the contact is present. When the full potential of Figure 1 is used, we discriminate if the contact is formed because the distance between the residues is inside the range corresponding to the Go attractive well, the ssm well, or neither. This way, we can compute a frequency contact map, which indicates the fraction of the simulated conformations in a given set of conditions where every contact is present, in any of the two attractive basins considered. Some representative frequency contact

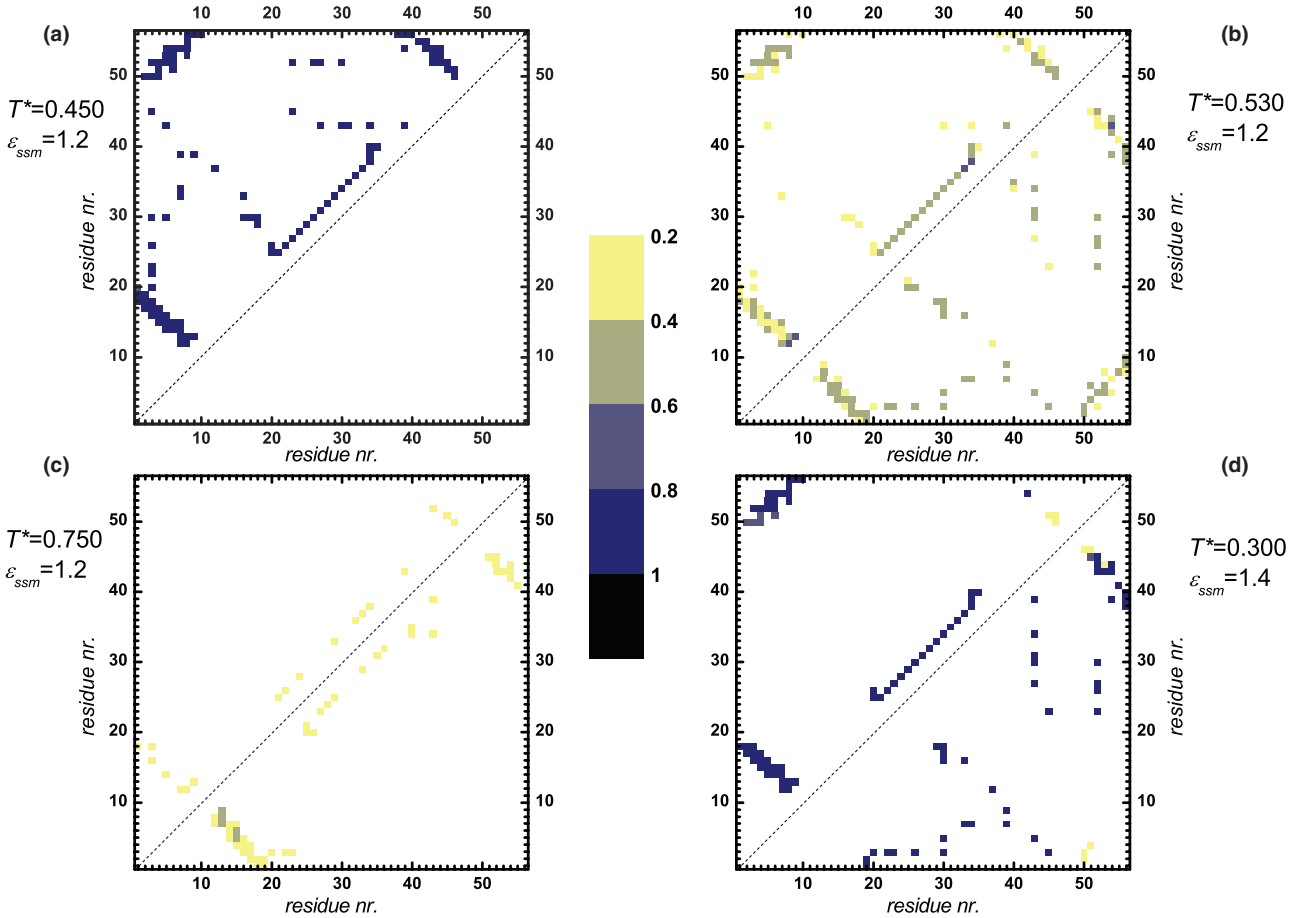

FIG. 7. Frequency contact maps under different conditions, as indicated in each panel. In every map, the upper-left triangle shows the contacts at the Go minimum, while the lower-right triangle shows the contacts at the ssm minimum.

maps are plotted in Figure 7. In every map, the upper-left triangle shows contacts inside the Go attractive well, while the lower-right triangle shows contacts inside the ssm basin (previously named as "pseudocontacts"[32]). The map in panel (a) is shown just as a test of the procedure. It is computed on the conformations sampled at a temperature below T_m^* for $\epsilon_{ssm} = 1.2$, where only the state N is populated. As expected, all the native contacts are close to the Go minimum in essentially all the conformations of this state. In panel (b) we show a situation slightly above the transition temperature for the same pressure, where S is by far the most populated state. Now, we can see that many native contacts are partitioned between both energy minima. The internal contacts of the α-helix are mostly preserved in the Go minimum, since its i–$i + 4$ contacts have been classified as "hydrogen-bond" contacts in our model, and therefore they do not have the ssm minimum in their interactions. In the contacts belonging to the β-sheet, however, the situation is more complex, and reflects the frustration introduced into the system by large ϵ_{ssm} values, as previously mentioned. So, even though some of the native contacts have been labeled as "hydrogen-bond" and only interact through the Go attraction, other contacts (in a larger number) affecting neighbor residues are also stabilized by the ssm attraction, which is now energetically more intense than the Go value. Therefore, the ensemble of conformations is more diverse in this case, as reflected in the average values shown in these frequency maps. If the temperature is further raised, we come to the situation shown in panel (c), which corresponds to conformations belonging both to the S and D states (since, as we have shown, one cannot distinguish between them in these conditions). Only a few native contacts appear in either attractive well, and they always show a very low, marginal population. This is slightly more abundant in the ssm region than in the Go one, but the average values are so small that this fact seems to be rather irrelevant, and probably only reflects the larger volume considered to define a contact as present when the framing distances are larger.

Finally, in panel (d) we show the results at low temperature and very high pressure ($\epsilon_{ssm} = 1.4$). This map somehow shows the successful behavior of the interaction model we have defined in this work. At these low temperatures, all the conformations of the trajectory correspond to the swollen state S, as we defined it in Sec. III A. We can see that the helix is correctly formed, and three of the four strands in the β-sheet also keep their contacts at the Go minimum. On the other hand, the contacts among residues belonging to different secondary structure elements (mostly those between the helix and the β-sheet) appear with a high population in the ssm attractive well. This means that, by especially considering in our interaction scheme those contacts which clearly do not correspond to hydrophobic contacts, even at the simple level employed in this work, our simulation model is able to find a state with secondary structure very similar to the native conformation, but with a larger radius of gyration as it corresponds to a swollen version of the former. At this low temperature, the parallel tempering procedure almost produces an energy minimization of the system, which results in a very regular set of conformations. At more relevant, larger temperatures, the thermal fluctuations create wider structural fluctuations, as those reflected in panel (b) (for a slightly lower pressure), which would better represent the diversity of conformations in a pressure induced unfolded state, with features very different from the temperature induced one.

It may be interesting to mention that, at this moment, we cannot directly compare these results to the experimental evidence existing for the same protein at high pressures.[51] The experiments show a less distorted structure at a pressure which is probably less than that corresponding to our results at large values of ϵ_{ssm}. According to the nuclear magnetic resonance experiments reported,[51] only one water molecule has been included into the hydrophobic core of the structure at 2 kbar. Probably, at higher pressures the larger structural fluctuations in the pressure unfolded state make it difficult to obtain relevant information about the structure present. This just remarks the need for further research, both on the experimental and simulation sides, on the effect of pressure on proteins conformational space.

IV. SUMMARY AND CONCLUSIONS

In this work, we show the performance of a very simple simulation model we have introduced in order to study the pressure unfolding of proteins at a coarse-grained level. The model uses only the α carbons of the protein, without any explicit consideration of the surrounding water molecules. The interactions of the system, on the other hand, use the results of potentials of mean force which have into account the solvation effects on hydrophobic interactions, together with a mean field approach of the pressure effects on the interaction parameters. Therefore, we have used a structure-based potential, and modified the relative depth of the energy terms corresponding to direct contacts and solvent separated contacts to reproduce, in a qualitative way, an increase in the system pressure.

We have used parallel tempering Monte Carlo simulations to obtain the most relevant structural and thermodynamic features of the unfolding processes occurring at different pressures for protein GB1. To this end, we have first obtained the global properties of the system, as the transition temperature from the folded state to the unfolded state. But we have also checked that the unfolded state one finds at room pressure and high temperature is very different from that we have found at high pressures and relatively low temperatures. Under these latter conditions, an ensemble structurally similar to the native state (as derived from its relatively low RMSD value) but with a slightly larger radius of gyration appears, which we have named the swollen state. Moreover, this ensemble shows a real two-state equilibrium with the native state in a certain range of temperature and pressures, but gradually transforms into the denatured state arising from thermal unfolding at room pressure when the temperature is progressively increased. Therefore, the swollen state shows the distinctive features of a downhill transition when transforming into the denatured state.

As a summary of all these results, in Figure 8 we show a sketchy phase diagram showing our *ad hoc* pressure scale (in the form of the energetic parameter ϵ_{ssm}) against the reduced temperature of the system. The red solid circles show the po-

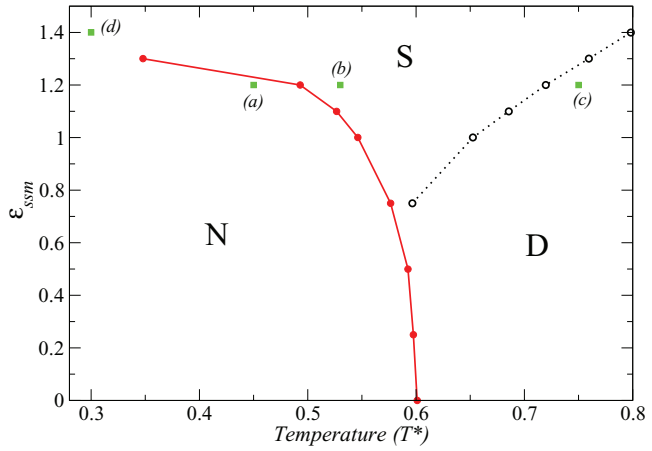

FIG. 8. Representation of ϵ_{ssm} against the mid-point temperature of the two transitions discussed in this work. The red solid line shows the boundary line corresponding to the N–D or N–S equilibrium, while the black dotted line shows the approximate center of the downhill S–D transition. Green dots show the situation of the frequency contact maps in Fig. 7.

sitions of T_m^* for the given pressure conditions. As previously discussed, they correspond to a real cooperative transition, although with a small free energy barrier, reflecting an equilibrium between the native N and the swollen S states. The red line could then be considered as a real phase boundary in this diagram. Interestingly, the shape of this line would correspond to one part of the elliptical phase diagram found for protein unfolding as a function of temperature and pressure, the only region attainable for such a simple model as ours. For example, there is no way we can reproduce the cold denaturation of a protein with our model, since explicit water is not taken into account.

The black open circles in Figure 8 correspond to an approximate transition temperature between the swollen and the denatured states. We have computed them from the crossing of the S and D populations in Figure 5. As we have already mentioned, these two states are indeed only one, whose properties gradually transform from one end to the other, in a continuous, downhill-type transition. Therefore, the dashed line in Figure 8 does not correspond to a real boundary between two different states, and cannot be considered as a phase boundary in the diagram.

The green dots in Figure 8 indicate the situation of the systems analyzed in the frequency maps shown in Figure 7, with the letters in parentheses showing the specific panels in Figure 7. As previously discussed, panel (b) in Figure 7 is the most interesting one in this study, since it corresponds to the real (thermally distorted) swollen state, although panel (d) shows the "ideal" structure of this swollen state after an energy minimization.

As a final summary, we have proved that our designed potential creates a swollen state which is stable at low or moderate temperatures and high pressures. The detailed structural characterization of this state shows that the elements of secondary structure are relatively well preserved, but the hydrophobic contacts appear at distances larger than those in the native state, due to the injection of water molecules inside the protein hydrophobic core. Of course, in our coarse-grained model water is not explicitly considered; this effect is reproduced instead by a change in the energetic minimum occupied by the interacting residues, from the Go well to the solvent separated minimum. The careful definition of some contacts where this change is not allowed preserves the nice characteristics of this swollen ensemble, compatible with the global observed features of pressure unfolded proteins.

ACKNOWLEDGMENTS

This work has been partially supported by Comunidad Autónoma de Madrid (Grant No. S2009/PPQ-1551) and the Spanish Ministerio de Ciencia e Innovación (Grant No. FIS2009-13364-C02-02). R.P. acknowledges a researcher stipend from Comunidad Autónoma de Madrid.

[1] P. Bridgman, J. Biol. Chem. **19**, 511 (1914).
[2] S. Rick, J. Phys. Chem. B **104**, 6884 (2000).
[3] Y. Harano, T. Yoshidome, and M. Kinoshita, J. Chem. Phys. **129**, 145103 (2008).
[4] N. Hillson, J. N. Onuchic, and A. E. Garcia, Proc. Natl. Acad. Sci. U.S.A. **96**, 14848 (1999).
[5] U. Mayor, N. Guydosh, C. Johnson, J. Grossmann, S. Sato, G. Jas, S. Freund, D. Alonso, V. Daggett, and A. Fersht, Nature (London) **421**, 863 (2003).
[6] J. Silva, D. Foguel, and C. Royer, Trends Biochem. Sci. **26**, 612 (2001).
[7] J. Silva, D. Foguel, M. Suarez, A. Gomes, and A. Oliveira, J. Phys.: Condens. Matter **16**, S929 (2004).
[8] D. Knorr, V. Heinz, and R. Buckow, Biochim. Biophys. Acta **1764**, 619 (2006).
[9] X. Liu, J. Ning, and S. Clark, Front. Chem. Eng. China **3**, 436 (2009).
[10] F. Meersman and C. Dobson, Biochim. Biophys. Acta **1764**, 452 (2006).
[11] D. Foguel, C. Robinson, P. de Sousa, J. Silva, and A. Robinson, Biotechnol. Bioeng. **63**, 552 (1999).
[12] C. Dobson, Nature (London) **426**, 884 (2003).
[13] F. Meersman, C. Dobson, and K. Heremans, Chem. Soc. Rev. **35**, 908 (2006).
[14] F. Meersman, L. Smeller, and K. Heremans, Biochim. Biophys. Acta **1764**, 346 (2006).
[15] C. Royer, Biochim. Biophys. Acta **1595**, 201 (2002).
[16] S. Sarupria, T. Ghosh, A. E. Garcia, and S. Garde, Proteins **78**, 1641 (2010).
[17] T. Yoshidome, Entropy **12**, 1632 (2010).
[18] J. Roche, J. A. Caro, D. R. Norberto, P. Barthe, C. Roumestand, J. L. Schlessman, A. E. Garcia, B. Garcia-Moreno E., and C. A. Royer, Proc. Natl. Acad. Sci. U.S.A. **109**, 6945 (2012).
[19] B. Boonyaratanakornkit, C. Park, and D. Clark, Biochim. Biophys. Acta **1595**, 235 (2002).
[20] J. R. Grigera and A. N. McCarthy, Biophys. J. **98**, 1626 (2010).
[21] M. Kinoshita, Int. J. Mol. Sci. **10**, 1064 (2009).
[22] T. V. Chalikian and R. B. Macgregor, Jr., J. Mol. Biol. **394**, 834 (2009).
[23] D. Paschek and A. E. Garcia, Phys. Rev. Lett. **93**, 238105 (2004).
[24] D. Paschek, S. Hempel, and A. E. Garcia, Proc. Natl. Acad. Sci. U.S.A. **105**, 17754 (2008).
[25] T. Ghosh, A. E. Garcia, and S. Garde, J. Am. Chem. Soc. **123**, 10997 (2001).
[26] V. Tozzini, Curr. Opin. Struct. Biol. **15**, 144 (2005).
[27] C. Czaplewski, S. Rodziewicz-Motowidlo, A. Liwo, D. Ripoll, R. Wawak, and H. Scheraga, Protein Sci. **9**, 1235 (2000).
[28] M. Moghaddam, S. Shimizu, and H. S. Chan, J. Am. Chem. Soc. **127**, 303 (2005).
[29] Z. Liu and H. S. Chan, Phys. Biol. **2**, S75 (2005).
[30] H. Kaya and H. S. Chan, J. Mol. Biol. **326**, 911 (2003).
[31] D. Asthagiri, S. Merchant, and L. R. Pratt, J. Chem. Phys. **128**, 244512 (2008).
[32] M. Cheung, A. E. Garcia, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **99**, 685 (2002).
[33] G. Hummer, S. Garde, A. E. Garcia, M. Paulaitis, and L. Pratt, Proc. Natl. Acad. Sci. U.S.A. **95**, 1552 (1998).
[34] T. Ghosh, A. E. Garcia, and S. Garde, J. Chem. Phys. **116**, 2480 (2002).
[35] H. Taketomi, Y. Ueda, and N. Gō, Int. J. Pept. Protein Res. **7**, 445 (1975).

[36]N. Gō, J. Stat. Phys. **30**, 413 (1983).
[37]H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, and P. Bourne, Nucleic Acids Res. **28**, 235 (2000).
[38]L. Prieto, D. de Sancho, and A. Rey, J. Chem. Phys. **123**, 154903 (2005).
[39]L. Prieto and A. Rey, J. Chem. Phys. **127**, 175101 (2007).
[40]M. Larriva, P. Pruscolini, and A. Rey, Proteins **78**, 73 (2010).
[41]H. Lammert, A. Schug, and J. N. Onuchic, Proteins **77**, 881 (2009).
[42]J. K. Noel, P. C. Whitford, and J. N. Onuchic, J. Phys. Chem. B **116**, 8692 (2012).
[43]M. Enciso and A. Rey, J. Chem. Phys. **132**, 235101 (2010).
[44]M. Enciso and A. Rey, Biophys. J. **101**, 1474 (2011).
[45]U. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).
[46]A. Gronenborn, D. Filpula, N. Essig, A. Achari, M. Whitlow, P. Wingfield, and G. Clore, Science **253**, 657 (1991).
[47]D. Idiyatullin, I. Nesmelova, V. Daragan, and K. Mayo, J. Mol. Biol. **325**, 149 (2003).
[48]K. Ding, J. Louis, and A. Gronenborn, J. Mol. Biol. **335**, 1299 (2004).
[49]J. Shimada and E. Shakhnovich, Proc. Natl. Acad. Sci. U.S.A. **99**, 11175 (2002).
[50]P. Derreumaux, J. Chem. Phys. **119**, 4940 (2003).
[51]D. J. Wilton, R. B. Tunnicliffe, Y. O. Kamatari, K. Akasaka, and M. P. Williamson, Proteins **71**, 1432 (2008).
[52]C. Dumont, T. Emilsson, and M. Gruebele, Nat. Methods **6**, 515 (2009).
[53]V. Muñoz, Annu. Rev. Biophys. Biomol. Struct. **36**, 395 (2007).