

## Thermodynamics of Gō-type models for protein folding

Lidia Prieto, David de Sancho, and Antonio Rey<sup>a)</sup>

*Departamento de Química Física, Facultad de Ciencias Químicas, Universidad Complutense, E-28040 Madrid, Spain*

(Received 12 August 2005; accepted 22 August 2005; published online 18 October 2005)

Gō-type potentials, based on the inter-residue contacts present in the native structure of a protein, are frequently used to predict dynamic and structural features of the folding pathways through computer simulations. However, the mathematical form used to define the model interactions includes several arbitrary choices, whose consequences are not usually analyzed. In this work, we use a simple off-lattice protein model and a parallel tempering Monte Carlo simulation technique to carry out such analysis, centered in the thermodynamic characteristics of the folding transition. We show how the definition of a native contact has a deep impact on the presence of simple or complex transitions, with or without thermodynamic intermediates. In addition, we have checked that the width of the attractive wells has a profound effect on the free-energy barrier between the folded and unfolded states, mainly through its influence on the entropy of the denatured state.

© 2005 American Institute of Physics. [DOI: 10.1063/1.2064888]

### I. INTRODUCTION

The protein folding problem, i.e., the determination of the three-dimensional native structure of a protein from its amino acid sequence, has been the aim of intense experimental and theoretical works for the last decades. The advancements in both the temporal resolution for experimental measurements and the simulation techniques have also focused a great attention into the folding process itself, the set of conformational changes leading from the structurally wide unfolded ensemble to the narrower native ensemble, and in its thermodynamic and kinetic characterization.<sup>1</sup>

Given the complexity of the protein-solvent system, and the span in distance and time scales of the folding process, simplified representations of the protein geometry and its interactions are often used in the theoretical and simulation consideration of the folding transition. Among them, the so-called Gō-type models deserve some special consideration.<sup>2,3</sup> Initially proposed about thirty years ago as a lattice representation of a protein, they have been “reinvented” in the last years, providing exciting data on the influence of the native topology in the folding pathways. Thus, and without any hope to be comprehensive in the references, they have been used with all combinations of Monte Carlo,<sup>4–6</sup> molecular dynamics<sup>7–9</sup> or Langevin dynamics,<sup>10,11</sup> or even pure theoretical expansions,<sup>12,13</sup> to provide insights in the transient intermediates appearing along the folding pathways, or the cooperative character of the transition, among other topics.

However, not much attention has been devoted to the details of the model itself. In its initial lattice proposal, the definition of a contact is straightforward, but the same cannot be claimed in continuous off-lattice models, at any resolution level. Clearly, it is somehow arbitrary to decide whether two residues are in contact or not in continuous space, specially

in simplified models where not all the protein atoms are taken into account. Recently, some studies have been carried out warning about possible problems with this contact definition. For example, Cieplak and Hoang<sup>14</sup> have used molecular dynamics to analyze the range of contact interactions and its influence in the foldability and folding times of a crambin model. Other works have also questioned the validity of Gō-type models in reproducing the cooperativity of protein folding transitions,<sup>15,16</sup> considering also in one case the problem with contacts definition.<sup>16</sup>

Other problem arises if one considers the mathematical form of the interaction potential. The Gō models are based on the native structure, i.e., a contact between two residues is given a favorable (attractive) contribution to the system energy if the same contact is also present in the native structure. In lattice models, this just means to define some energy contribution  $-\epsilon$  to be added when such contacts appear. Again, in off-lattice models the problem is more complex, since a continuous potential has to be defined, implying the “invention” of a distance dependence without a physical basis. Depending on the type of simulations (see references above), different analytical expressions have been used for the energy terms, without really analyzing, in the best of our knowledge, the possible intrinsic consequences of a given choice.

In this work, we try to perform such an investigation, centered here in the thermodynamic characteristics of the folding transition. These thermodynamic properties have been obviously considered in the previous studies as well,<sup>5,16–20</sup> but without a complete analysis on the model details which, as we shall show here, can clearly influence the simulation results and their interpretation. Thus, we shall show how the presence or absence of thermodynamic intermediates, or the height of the free-energy barrier between the folded and the unfolded states of the model, can be determined by the precise definition of the interaction potential.

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: jsbach@quim.ucm.es

## II. PROTEIN MODEL AND INTERACTIONS

In this work, we use a  $C^\alpha$  representation of the protein, i.e., every residue is represented by a hard sphere centered at its  $C^\alpha$ . The distance between contiguous beads in the primary sequence is kept constant at 3.8 Å, corresponding to a trans peptide bond.

In addition to a hard repulsion between beads, which avoids unphysical conformations, we have used a Gō potential, which defines an attractive interaction between every pair of amino acids which are in contact in the native structure. Precisely, the definition of what a contact is and the shape of the potential to represent this attraction is the primary concern of this work, as mentioned in the Introduction. In spite of the artificial nature of a Gō potential, many authors have used physically based potentials for their mathematical expressions. Therefore, it is rather common to find harmonic springs centered at the native distance or angle for virtual bond lengths and bond angles, and Lennard-Jones-type potentials, with several sets of exponents, for interactions among non-neighbor residues.<sup>7,14,16</sup> Other authors have used square-well potentials,<sup>18</sup> whose widths are also taken from the standard considerations of physical attractions. While these mathematical expressions are well established in different force fields, and are probably very convenient if one tries to adapt commercial or freely available software packages, it is difficult to justify its mathematical form in a potential which lacks any physical nature.

Therefore, we have chosen a simple form for our potential which allows us to carry out this type of analysis. In order to do so, we have defined Gō-type attractions between residues  $i$  and  $j$  which form a contact (*vide infra*) in the native structure through a harmonic well centered at the native distance  $d_{ij}^{\text{nat}}$ , and truncated at different distances  $a$  relative to this center. Therefore, the mathematical definition of this potential for two residues at a distance  $r$  is

$$u_{ij}(r) = \begin{cases} w_{ij}[(r - d_{ij}^{\text{nat}})^2 - a^2], & d_{ij}^{\text{nat}} - a < r < d_{ij}^{\text{nat}} + a \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

a definition which warrants a continuous potential, i.e.,  $u_{ij}(d_{ij}^{\text{nat}} \pm a) = 0$ . The weights  $w_{ij}$  serve two different purposes. First place, they are obviously null ( $w_{ij} = 0$ ) if residues  $i$  and  $j$  are not in contact in the native structure. When the native contact exists,  $w_{ij}$  takes a value which controls the well depth. In this work, we have considered all contributions equal, as it corresponds to a “true” Gō-type potential, without any sequence information. Therefore, for residues participating in native contacts,  $w_{ij} = 1/a^2$ , which gives an energy  $u_{ij} = -1$  at the well minimum for all the individual contributions. The energies and other thermodynamic quantities along this work are all expressed in dimensionless (reduced) units.

The truncation of the potential contributions favors disordered unfolded states at high temperatures. If the harmonic well extended to large deviations relative to the native distances, only collapsed states would be favored at any tem-

perature, resulting in a hardly justifiable folding-unfolding transition. Therefore, in this work we study only small values of the distance  $a$ , less than or equal to 1 Å.

In the potential, we have included all types of non-neighbor distances. For  $j = i \pm 2$ , we control the virtual bond angles of our model. For these contributions, we always consider  $w_{ij} \neq 0$ . The same happens in our model for  $j = i \pm 3$ , i.e., for the virtual torsion angles of the model. In this contribution, we introduce a small modification in Eq. (1). A Gō-type potential, based only on distances, is not able to discriminate between a conformation and its mirror image. To avoid this fact, we have introduced the local chirality at the level of these distances. Thus, both  $r$  and  $d_{ij}^{\text{nat}}$  for contributions between  $i$  and  $i \pm 3$  residues have a sign (positive or negative), equal to the sign of the triple (scalar) product of the three vectors defining the virtual torsion angle. This way, the mirror image of the native structure is disfavored.

For  $j \geq i + 4$  or  $j \leq i - 4$ , we introduce a second parameter in the potential considered in this work, controlling the definition of a “contact” in the native structure, and therefore the weights  $w_{ij}$  of the potential. As mentioned in the Introduction, this definition has always a certain arbitrary component in off-lattice models. Classic textbooks on protein structure used to display contact maps, as an alternative representation of protein structures, marking with spots distances between  $C^\alpha$  located at less than 10 Å, sometimes even larger distances.<sup>21</sup> On the other hand, recent studies employ shorter distances, ranging from, say, 6.5 to 8.5 Å between  $C^\alpha$ , to define a contact in the native structure.<sup>14</sup> In this work we use a parameter  $d_{\text{max}}$  to define a contact, and analyze its influence in the results of the folding process for the model. Therefore, for contacts between residues  $i$  and  $j$  with  $|i - j| \geq 4$ , we define  $w_{ij} = 0$  when  $d_{ij}^{\text{nat}} > d_{\text{max}}$ , and the weight mentioned above ( $w_{ij} = 1/a^2$ ) if the contact is present according to the choice of  $d_{\text{max}}$ . Here, we have considered two different values for it, the “classical” value of 10 Å and a lower value of 7.5 Å, in the middle of the “newer” range. We want to anticipate that it is not reasonable to expect a single optimum value of  $d_{\text{max}}$  valid for any protein. The specific packing in every protein structure will most probably result in a different choice for the threshold distance for the definition of a native contact. Thus, our aim is not only to check the effects of  $d_{\text{max}}$ , but to try to establish at least a semiquantitative criterion to determine its best value for a given native structure.

## III. SIMULATION ALGORITHM

Since we want to study the thermodynamic characteristics of the folding-unfolding transition for our model, we have used a simulation method which should be able to properly reproduce the equilibrium distribution of states for the system at every temperature under consideration. Therefore, we have used a parallel tempering Monte Carlo simulation algorithm.<sup>22</sup> This way, we can simultaneously study the large conformational changes characterizing the unfolded state at temperatures above the transition, and the small fluctuations around the native state at low temperatures, and at the same time we minimize the possibility to become trapped in local minima which would spoil the thermodynamic analysis. We

should mention that, since the folding transition may have, as we shall see, features of a first-order thermodynamic transition, a large number of temperatures has to be used in order to warrant a proper overlap of the energy distributions at neighbor temperatures around the transition midpoint, and the adequate wandering of the replicas through the different temperatures. Otherwise, the replica exchange which characterizes the parallel tempering technique has huge rejection ratios, and the method becomes useless.<sup>23</sup>

For the chain model, we use end moves for the vectors defining the virtual bonds for the two extreme units, and spike moves which preserve virtual bond lengths for the internal units. Also, we have included the possibility of a “shifting move,” consisting of an end-type movement for an internal unit, followed by a shifting of the remaining of the chain to keep the connectivity. The three types of moves have small amplitudes, and they are therefore quite adequate to properly sample compact states as those appearing at temperatures below the folding transition of the polypeptide chain model. After every individual move, the conformation is checked for overlappings and, if none is present, its energy is calculated according to the potential defined in Sec. II. A standard Metropolis test decides whether the new conformation is accepted or rejected.

A set of  $N$  local move trials, with  $N$  being the number of residues of the simulated protein, defines one of our “Monte Carlo cycles.” Every several thousand of these cycles, an exchange of replicas is tried between contiguous temperatures. The exact number of temperatures for every parallel tempering simulation depends on the transition characteristics, as deduced from preliminary relatively short simulations, and ranges in this work from 12 to 30, a large number due to the transition complexities mentioned above. The full simulation consists of several (5 to 20) million Monte Carlo cycles (at every temperature), after  $(2-5) \times 10^6$  equilibration cycles. Every simulation requires, for the system simulated (see below), about 2–4 days in single processor machines (2.8 GHz CPU). To warrant the correct sampling corresponding to a system in equilibrium, at least five independent simulations, starting from different seed numbers, are computed for every system (set of potential parameters). The results presented in Sec. IV correspond to statistical averages over the sampling at every temperature, and over the five independent runs.

#### IV. NUMERICAL RESULTS AND DISCUSSION

We are checking in this work the effects of the parameters defining a structure-based potential on the thermodynamic aspects of the folding transition. Therefore, we have chosen a protein whose transition is thoroughly characterized from a thermodynamic point of view. Our choice has been the immunoglobulin binding domain of streptococcal protein G [Protein Data Bank<sup>24</sup> (PDB) code 2gb1,<sup>25</sup> which we shall use in the remaining of the paper to name the protein]. It has two  $\beta$  hairpins forming a  $\beta$  sheet and a central  $\alpha$  helix, arranged in a highly symmetric form, in a total of 56 amino acids. In addition to its small size, very convenient from the computational side, its folding process has been the subject

of numerous detailed analysis, both experimental<sup>26–30</sup> and from different theoretical techniques,<sup>31–34</sup> mainly devoted to the folding dynamics and pathways. The main conclusions are that this protein folds in a cooperative, all-or-none manner, without detectable thermodynamic intermediates, although some kinetic intermediates have been characterized. From the thermodynamic point of view, a first-order folding-unfolding transition, with a high (given its stability) free-energy barrier, is the experimental outcome for this protein.

To characterize the type of transition we get with our G $\bar{o}$ -type model and the different potential parameters, we have calculated in first place the heat capacity of our system at every temperature, readily computable from the energy fluctuations. The results for this property, in reduced units, are shown in Fig. 1. For any value of  $d_{\max}$  or  $a$ , the transition is clearly appreciable as a peak in the heat capacity, which shifts towards higher temperatures when  $a$  grows. We shall comment on that fact later. Here, the most important feature is the dramatic effect of  $d_{\max}$  on the peak itself. While for the largest value,  $d_{\max}=10 \text{ \AA}$ , we observe single, well-defined peaks, for the smaller cutoff  $d_{\max}=7.5 \text{ \AA}$  we clearly observe a transition with two peaks, neatly distinguishable for small values of the potential width  $a$ , and overlapping providing a complex peak profile for higher values of  $a$ . This is clearly indicating the presence of a thermodynamic intermediate when we use a small cutoff distance for the definition of a native contact. As a matter of fact, the observation of snapshots of the simulated conformations for  $d_{\max}=7.5 \text{ \AA}$  shows at intermediate temperatures a state in which the  $\beta$  sheet is formed, but the helix is unfolded. At lower temperatures, the helix correctly forms, and the full native conformation is the equilibrium state in this case, as it is in all the others we have considered in this work, thus validating our conformational sampling and simulation method.

Although some studies have detected the early transient formation of one of the  $\beta$  hairpins with respect to the other structural elements in 2gb1, we are not dealing with dynamic intermediates along the folding pathway. Our heat capacity results for  $d_{\max}=7.5 \text{ \AA}$  represent a thermodynamic intermediate, which is not observed in real experiments for this protein, and which disappears when the value of  $d_{\max}$  is increased. In order to try to understand the reason for this behavior, we show in Fig. 2 several contact maps for the native structure of this protein, computed from different criteria. The first two panels, (a) and (b), show the distances among  $C^\alpha$ , truncated at 10 and 7.5  $\text{\AA}$ , respectively. In any of them it is easy to visualize the native protein structure: the central helix, with its thick cloud of spots along the map diagonal; the two  $\beta$  hairpins flanking the helix along the sequence, with the antiparallel contacts between strands 1 and 2, on one hand, and 3 and 4, on the other; and the parallel contacts between strands 1 and 4, which define the four member  $\beta$  sheet. Apart for numerous, but small, details, the main difference between maps (a) and (b) is the lack of contacts between the helix and the strands in the  $\beta$  sheet when the cutoff distance is reduced to 7.5  $\text{\AA}$ . This explains the apparent reduced stability of the helix for this choice of  $d_{\max}$ , since it is then only defined through local interactions.

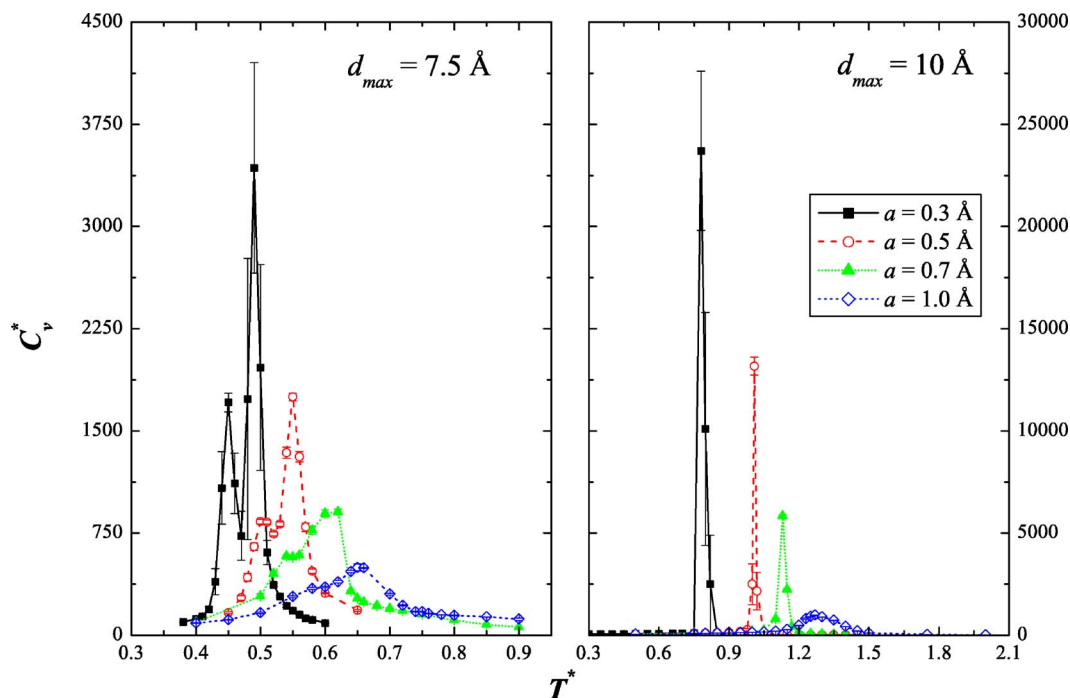


FIG. 1. (Color online) Heat capacity (in reduced units) obtained from our simulations of 2gb1 for different choices of the potential width  $a$  and the contact definition cutoff  $d_{\max}$ .

Thus, the intermediate we have found for one of our choices of the contact definition is a natural consequence of the potential used.

An interesting question should immediately arise: Is there a way to know which the proper choice of  $d_{\max}$  is to define a Gō-type potential for folding in off-lattice models? In order to try to answer this question, we have to go beyond the simplicities of the model. In real protein structures, the contacts are usually defined through the packing of side-chain atoms, which one does not explicitly include in a reduced representation of the polypeptide chain. However, since structure-based potentials always start from the experimentally obtained structure, these contacts are available. In panel (c) of Fig. 2 we show the contact map we obtain taking into account all the heavy atoms in every residue. The distance represented for residues  $i$  and  $j$  in this map is the smallest one from all the pairs of distances among heavy atoms in residue  $i$  and those in residue  $j$ . We have used a distance cutoff of 4.5 Å in this map, slightly larger than twice the average van der Waals radius for heavy atoms in proteins. Although this map does not completely match any of the other two maps in the figure, it is quite clear that it includes contacts between the helix and the four strands in the  $\beta$  sheet. At the level of a model which ignores the sequence, the presence of both local and long-range (along the sequence) contacts may cause this protein to fold in a cooperative manner, without thermodynamic intermediates reflecting an artificial relative stabilization of certain structural elements against others, which would result from an incorrect definition of the contact map, and therefore of a Gō-type folding potential. We could surely refine our choice of  $d_{\max}$ , since 10 Å seems to be larger than the needed value when comparing with the heavy atom contact map. We have not done that here since we believe our present results clearly

state the important point we want to point out. Only by properly including the relevant real contacts in the simplified interaction scheme can the model provide results compatible with the experimental observations. Otherwise, artifacts appear.

Therefore, we have seen that  $d_{\max}$ , one of the two parameters controlling the definition of a Gō-type potential, has a profound effect on the thermodynamics of the folding transition for the model and, one can infer, in the folding pathway as well. We start now to consider the effect of the potential width  $a$ . We have already seen in Fig. 1 that the values of  $a$  influence the broadness of the transition and its middle-point temperature, shifting it towards larger values for larger widths. In a naive analysis, this is a bizarre result. In the native structure, every individual contribution to the potential is at its minimum, irrespective from the value of  $a$ . Therefore, the native conformation of 2gb1 has the same energy for any of the potentials, once the number of contacts has been defined. If the energy of the native state was the only contribution to the protein stability, as it is sometimes considered, all the heat capacity curves should present their maxima at the same temperature, something that is not happening. Therefore, a more rigorous thermodynamic analysis of the simulation results has to be accomplished.

To do so, we have computed the energy histograms found by the simulation algorithm at every temperature. In Fig. 3 we show the results for the four potential widths considered in this work,  $d_{\max}=10$  Å, and three temperatures: the folding temperature  $T_m$ , obtained from the maximum of the heat capacity curves, and two additional temperatures around it. For every value of  $a$  we find a single state for temperatures above and below  $T_m$ . For  $T_f < T_m$ , this state corresponds to fluctuations around the native state of the protein, with large negative values of energy, as it corresponds to most of

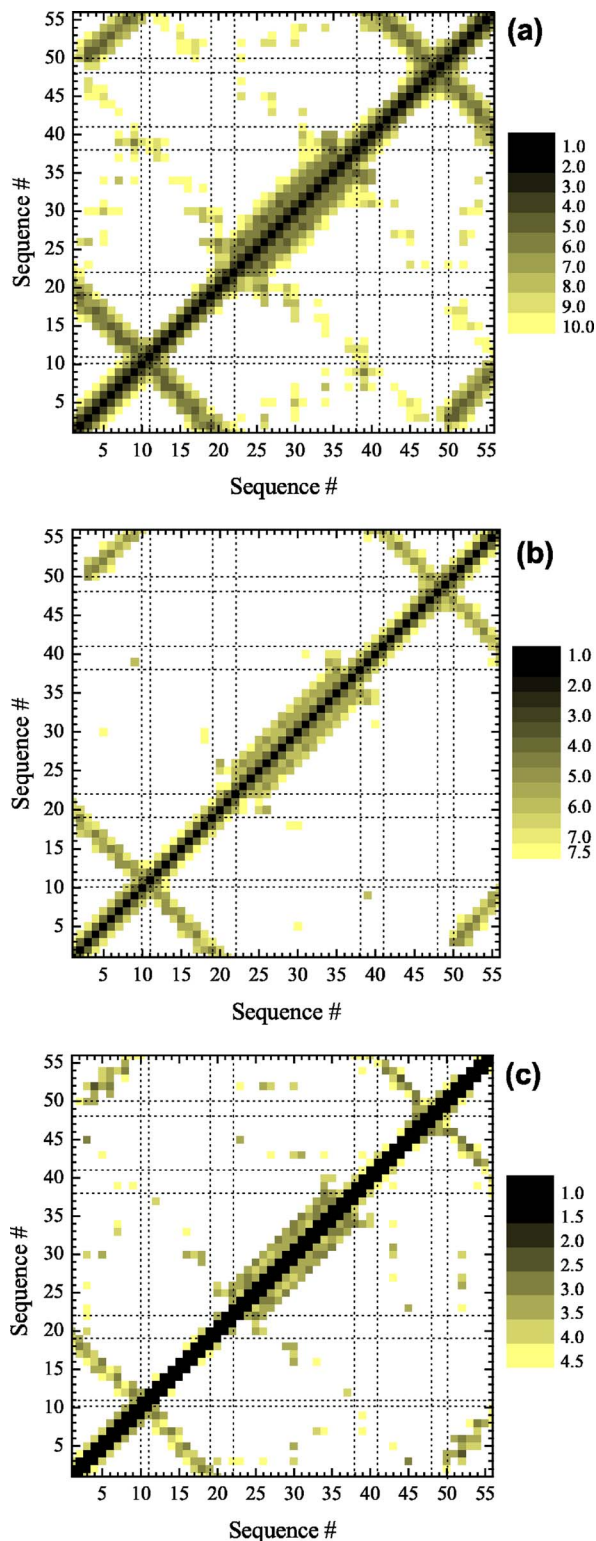


FIG. 2. (Color online) Contact maps for the native structure of protein 2gb1, calculated from: (a) distances between  $C^\alpha$  atoms, truncated at 10 Å; (b) distances between  $C^\alpha$  atoms, truncated at 7.5 Å; and (c) minimum distances between any heavy atom in every residue, truncated at 4.5 Å. Dotted lines separate elements of secondary structure. Legends in angstrom.

the native contacts being reproduced. For  $T_u > T_m$ , on the other hand, we get another single peak distribution at higher energies, corresponding to the denatured state. Its energies are still negative (our potential does not include positive con-

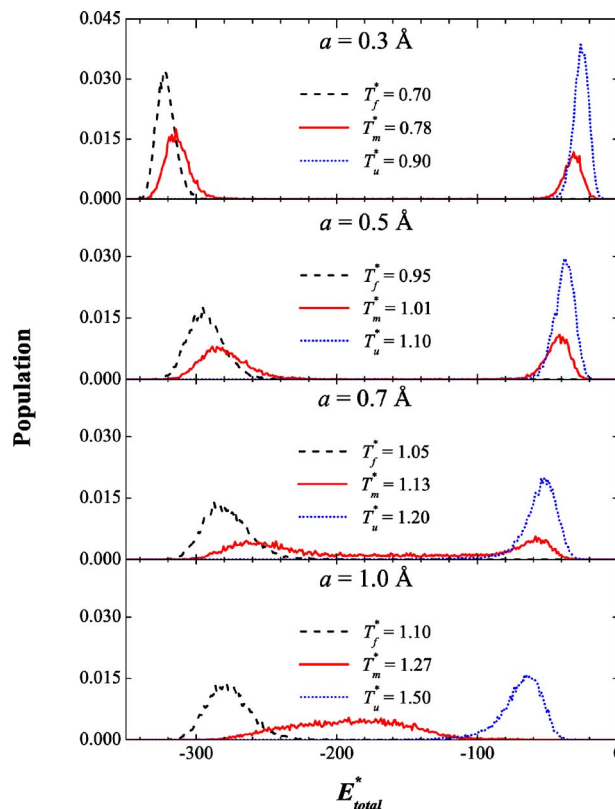


FIG. 3. (Color online) Reduced energy histograms, computed from the simulations, for different choices of the potential width parameter  $a$ , and three temperatures below ( $T_f$ ), equal ( $T_m$ ) or above ( $T_u$ ) the transition temperature.  $d_{\max} = 10$  Å.

tributions, apart from hard-sphere repulsions), corresponding to sporadic contacts between native neighbors or natively local structures.

The most interesting conclusions come from the energy histograms at the equilibrium temperature  $T_m$ . While for small values of the potential width  $a$  they show a bimodal distribution, with both peaks of the native and denatured states, for the largest  $a$  the potential creates a single peak distribution, centered at intermediate energies. This represents fundamental differences in the folding process. A histogram with two peaks at  $T_m$  corresponds to the usual view of the folding process for single domain proteins as a “two-state” or first-order thermodynamic transition. The native and denatured states are separated by a free-energy barrier, which creates the all-or-none, cooperative transition, which traditionally characterizes protein folding.<sup>35</sup> On the other hand, a histogram with a single peak at  $T_m$  corresponds to what has been recently termed “downhill” folding, or folding without significant barriers.<sup>36,37</sup> One of the states gradually converts into the other as conditions (temperature, in our case) are changed, and therefore only one structural ensemble is populated at each temperature. The question whether this scenario may be compatible with real proteins is still under controversy.<sup>38,39</sup> What seems to be clear is that it does not correspond to the folding of 2gb1. Since the presence or absence of free-energy barriers severely affects the dynamics of the folding process, a control of the interaction potential employed in the simulations seems then to be of primary

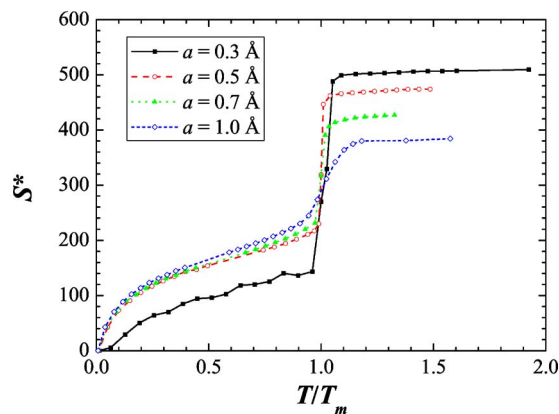


FIG. 4. (Color online) Reduced absolute entropies against the reduced temperatures for different choices of the potential width parameter  $a$ .  $d_{\max} = 10$  Å.

importance. For 2gb1, our results clearly indicate that  $a = 1$  Å, apparently not too large deviation from native distances for a coarse-grained model, creates a barrierless transition which does not reproduce the experimental findings for this protein.

That does not imply that  $a$  should be as small as possible. One of the advantages of the potential we are using is that it makes our simple protein model to show a rather narrow unfolding transition, even too narrow in some cases. According to the experimental calorimetric results of Ref. 26, 2gb1 unfolds at acid pH at about 65 °C, with the peak in the transition heat capacity spanning about  $\pm 20$  °C from this point. In a reduced temperature scale  $T/T_m$  (with  $T$  in Kelvin), that range would mean a transition which spans  $\pm 5\%$  from its middle point (in a rough estimate, assuming an approximately symmetric peak after baseline subtraction). In the same scale, our simulation results for  $d_{\max} = 10$  Å indicate that for  $a = 0.3$  Å the transition extends less than 1% from its middle point, about 2% for  $a = 0.5$  Å, close to 5% for  $a = 0.7$  Å, and more than 15% for  $a = 1.0$  Å. Of course, these numbers have to be considered with great caution, since it is difficult to properly define the beginning and end of the transition from a heat-capacity curve. Also, experimental results correspond to  $C_p$ , while simulation results compute  $C_v$ . Thus, we cannot claim from our simulation results for a single protein a fixed “good” value for  $a$  which could be extrapolated to other proteins. What we want to make clear is that the character of the transition is severely influenced by the width of the attractive wells, something independent from the protein contacts, and therefore subject to another arbitrary choice of the model details by the researcher.

From the energy histograms calculated from the simulation results, we have used the multiple histogram technique<sup>40,41</sup> to compute also the absolute entropies for the system at every temperature. These absolute entropies, as a function of the reduced temperature  $T/T_m$ , are shown in Fig. 4. Again, we observe the different broadnesses of the transition for the various choices of  $a$  we have already commented on. What we want to point out in this figure are the different values of the absolute entropy that the potential rises for the different values of  $a$ . The entropy of the native state does not

TABLE I. Dimensionless thermodynamic properties for the folding transition between the native and the unfolded states of our model, for different values of the potential width  $a$ .  $d_{\max} = 10$  Å.

$a$ (Å)	$T_m^a$	$\Delta E_m^*$	$\Delta S_m^*$	$T_m^b$
0.3	0.78	279	347	0.80
0.5	1.01	235	233	1.01
0.7	1.13	155	137	1.13
1.0	1.27	...	...	...

<sup>a</sup>From the maximum of the heat-capacity curves.

<sup>b</sup>From the ratio  $\Delta E_m^*/\Delta S_m^*$ .

seem to reflect any important change depending on  $a$ , with the exception of  $a = 0.3$  Å, with a smaller value due to the very narrow attractive wells, which allow very small fluctuations. Anyway, the entropy results for the native state are as expected. For the unfolded state, on the other hand, the absolute entropy clearly reaches lower values the larger  $a$  is. In principle, one should expect that the unfolded state would be probably the worst ensemble represented by a Gō-type potential, based on the native structure, and without any sequence information in it. But that fact does not preclude its clear influence on the model behavior, which again has to be understood and taken into account when using this kind of potentials. The absolute energies show a similar behavior, with lower (more negative) values in the unfolded state for larger values of  $a$ , as it may be seen in the high-temperature peak of Fig. 3. This is consistent with the fact that a wider potential can capture more sporadic contacts than a narrow one, therefore explaining the energy differences of the denatured state. In a funnel view of the folding landscape,<sup>42</sup> as it would correspond to a Gō-type model with little frustration, lower levels of energy correspond to less entropies as well, as we numerically get from our simulations. From a quantitative point of view, however, the differences are not compensated. In Table I we present the calculated values of  $\Delta S$  and  $\Delta E$  between the folded and unfolded states at the transition temperature  $T_m$ , when there is a two-state transition (i.e., for values of  $a \neq 1.0$  Å in our model). Both sets of values decrease when  $a$  increases, but the steeper reduction of the entropy change makes the transition temperature grow, as we have seen from Fig. 1. As a consistency check of our entropy calculations, we also compute in the last column of Table I the transition temperature from the equilibrium condition  $\Delta F = \Delta E - T_m \Delta S = 0$ . The results we get precisely coincide, within the numerical precision of our results, with the values of  $T_m$  obtained from the heat-capacity curves. There is only a small difference for the narrowest potential  $a = 0.3$  Å, but the excessively abrupt nature of the transition for this value of  $a$ , which makes its detailed analysis very difficult, is surely the responsible for this small difference.

In conclusion, we see how the influence of the potential width on the denatured state affects the thermodynamic characteristics of the transition in yet another way. The displacement of the folding temperature may induce wrong conclusions on the stability of the native state for the protein model. Only through an entropy (or free energy) analysis and the

exploration of the unfolded state, frequently forgotten in this kind of simulation studies, can the results be properly interpreted.

## V. SUMMARY AND CONCLUSIONS

In this work, we have used a simple model of protein 2gb1 based on  $C^\alpha$  to analyze the influence of the parameters defining a Gō-type potential on the thermodynamic characteristics of the simulated folding-unfolding transition. Specifically, we have considered two parameters:  $d_{\max}$ , the cutoff distance for the definition of a contact in the native structure; and  $a$ , the width of the potential well which represents the attraction between two residues subjected to this contribution.

We have found that the value of  $d_{\max}$  is of critical importance, since it determines the contact map itself, which lies on the bottom of the potential definition. An inadequate value (too small, in our case) may artificially create regions in the protein which, according to the energy function, become less stable than others. If that happens, the folding transition may happen through a series of thermodynamic intermediates, and a complex analysis of different types of transitions could be possibly made (collapse, disordered-ordered, etc.<sup>17</sup>). This confirms previous results.<sup>20</sup> But we also find that, with a contact map properly chosen, simple models based on  $C^\alpha$  alone can be used to simulate a “correct” cooperative folding transition, in qualitative agreement with the experimental data, without the need for more complex representations of the protein chain.<sup>8,19</sup> Also, we have found that, at least for 2gb1, a contact map based on the protein heavy atoms can be used to properly adjust the best value of  $d_{\max}$ , or at least to discard potentially dangerous values, as it happens to be  $d_{\max}=7.5$  Å for this protein.

Equally important, but for different reasons, is the width potential parameter  $a$ . In a small range variation, from 0.3 to 1.0 Å, which represents the deviations from native distances still falling inside the attractive well, we obtain a clear variation in the free-energy barrier between the folded and the unfolded states, which becomes negligible (barrierless) for the largest value of  $a$  we have considered. Thus, through a careful tuning of this parameter it seems to be possible to go from a first-order, all-or-none, very abrupt transition to a downhill folding transition. The relevance of this change to the folding pathway and dynamics, though not explicitly considered in this work due to the simulation technique employed, is obvious. But this type of thermodynamic analysis is not customarily done prior to the dynamic analysis of folding trajectories, even though the different forms used for the attractive wells (usually Lennard-Jones potentials with 12-10 or 12-6 exponents, and thus different widths) could probably shift the type of transition from one side to the other. We are not claiming that previous simulations using Gō-type potentials are wrong, but it might be interesting anyway to have the types of checks suggested in this work performed to warrant a better predictive character for these types of models. (Incidentally, we want to point out that the influence of the potential width on the free-energy barrier is not a consequence of the specific shape of the potential. We

have obtained some results with attractive square wells of different widths, obtaining the same conclusions.)

We want also to point out that the changes in the free-energy barriers we have observed are mainly due to the effects of the potential details on the denatured state of the protein model. This state tends to be ignored, given that its definition in a native-based potential without sequence information is far from complete, if not clearly wrong. Nevertheless, we have observed that both its energy and specially its entropy are clearly influenced by the width of the attractive wells, what translates into the folding transition thermodynamic characteristics themselves.

Altogether, we believe that this type of work can contribute to a better understanding of the behavior of simplified models of protein folding, and to a more correct interpretation of the simulation results they provide.

## ACKNOWLEDGMENTS

This work was in part supported by Grant No. BQU2002-04626-C02-01 from Ministerio de Ciencia y Tecnología, Spain. One of the authors (L.P.) acknowledges a grant from Spanish Ministerio de Educación y Ciencia, and another author (D.S.) a grant from the Spanish Ministerio de Ciencia y Tecnología. We also want to thank Professor Juan J. Ruiz Lorenzo, from Univ. Extremadura, for his kind help regarding parallel tempering algorithms.

<sup>1</sup>V. S. Pande, Proc. Natl. Acad. Sci. U.S.A. **100**, 3555 (2003).

<sup>2</sup>H. Taketomi, Y. Ueda, and N. Gō, Int. J. Pept. Protein Res. **7**, 445 (1975).

<sup>3</sup>N. Gō, J. Stat. Phys. **30**, 413 (1983).

<sup>4</sup>V. S. Pande and D. S. Rokhsar, Proc. Natl. Acad. Sci. U.S.A. **95**, 1490 (1998).

<sup>5</sup>J. Shimada, E. L. Kussell, and E. I. Shakhnovich, J. Mol. Biol. **308**, 79 (2001).

<sup>6</sup>J. Wang and W. Wang, J. Chem. Phys. **118**, 2952 (2003).

<sup>7</sup>C. Clementi, H. Nymeyer, and J. N. Onuchic, J. Mol. Biol. **298**, 937 (2000).

<sup>8</sup>A. Linhananta and Y. Zhou, J. Chem. Phys. **117**, 8983 (2002).

<sup>9</sup>C. Clementi, A. E. García, and J. N. Onuchic, J. Mol. Biol. **326**, 933 (2003).

<sup>10</sup>T. H. Hoang and M. Cieplak, J. Chem. Phys. **113**, 8319 (2000).

<sup>11</sup>B. Erman, Biophys. J. **81**, 3534 (2001).

<sup>12</sup>S. Takada and P. G. Wolynes, J. Chem. Phys. **107**, 9585 (1997).

<sup>13</sup>C. Clementi and S. S. Plotkin, Protein Sci. **13**, 1750 (2004).

<sup>14</sup>M. Cieplak and T. H. Hoang, Int. J. Mod. Phys. C **13**, 1231 (2002).

<sup>15</sup>E. Paci, M. Vendruscolo, and M. Karplus, Biophys. J. **83**, 3032 (2002).

<sup>16</sup>H. Kaya and H. S. Chan, J. Mol. Biol. **326**, 911 (2003).

<sup>17</sup>Y. Zhou and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **94**, 14429 (1997).

<sup>18</sup>Y. Zhou and M. Karplus, J. Mol. Biol. **293**, 917 (1999).

<sup>19</sup>Y. Zhou and A. Linhananta, J. Phys. Chem. B **106**, 1481 (2002).

<sup>20</sup>H. Jang, C. K. Hall, and Y. Zhou, Biophys. J. **82**, 646 (2002).

<sup>21</sup>T. E. Creighton, *Proteins. Structures and Molecular Properties*, 2nd ed. (Freeman, New York, 1993).

<sup>22</sup>U. H. E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).

<sup>23</sup>N. Rathore, M. Chopra, and J. J. de Pablo, J. Chem. Phys. **122**, 024111 (2005).

<sup>24</sup>H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, Nucleic Acids Res. **28**, 235 (2000).

<sup>25</sup>A. M. Gronenborn, D. R. Filpula, N. Z. Essig, A. Achari, M. Whitlow, P. T. Wingfield, and G. M. Clore, Science **253**, 657 (1991).

- <sup>26</sup>P. Alexander, S. Fahnestock, T. Lee, J. Orban, and P. Bryan, *Biochemistry* **31**, 3597 (1992).
- <sup>27</sup>P. Alexander, J. Orban, and P. Bryan, *Biochemistry* **31**, 7243 (1992).
- <sup>28</sup>F. J. Blanco, M. A. Jiménez, A. Pineda, M. Rico, J. Santoro, and J. L. Nieto, *Biochemistry* **33**, 6004 (1994).
- <sup>29</sup>K. Ding, J. M. Louis, and A. M. Gronenborg, *J. Mol. Biol.* **335**, 1299 (2004).
- <sup>30</sup>D. Idiyatullin, I. Nesmelova, V. A. Daragan, and K. H. Mayo, *J. Mol. Biol.* **325**, 149 (2003).
- <sup>31</sup>E. L. McCallister, E. Alm, and D. Baker, *Nat. Struct. Biol.* **7**, 669 (2000).
- <sup>32</sup>J. Shimada and E. I. Shakhnovich, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11175 (2002).
- <sup>33</sup>P. Derreumaux, *J. Chem. Phys.* **119**, 4940 (2003).
- <sup>34</sup>S. Y. Lee, Y. Fujitsuka, D. H. Kim, and S. Takada, *Proteins* **55**, 128 (2004).
- <sup>35</sup>S. E. Jackson and A. R. Fersht, *Biochemistry* **30**, 10428 (1991).
- <sup>36</sup>V. Muñoz, *Int. J. Quantum Chem.* **90**, 1522 (2002).
- <sup>37</sup>F. Y. Oliva and V. Muñoz, *J. Am. Chem. Soc.* **126**, 8596 (2004).
- <sup>38</sup>N. S. Ferguson, P. J. Sharpe, T. D. Sharpe, S. Sato, and A. R. Fersht, *J. Mol. Biol.* **344**, 295 (2004).
- <sup>39</sup>A. N. Naganathan, R. Pérez-Jiménez, J. M. Sánchez-Ruiz, and V. Muñoz, *Biochemistry* **44**, 7435 (2005).
- <sup>40</sup>A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988).
- <sup>41</sup>D. Gront, A. Kolinski, and J. Skolnick, *J. Chem. Phys.* **115**, 1569 (2001).
- <sup>42</sup>J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, *Proteins* **21**, 167 (1995).