Entropy reduction effect imposed by hydrogen bond formation on protein folding cooperativity: Evidence from a hydrophobic minimalist model

Marco Aurélio A. Barbosa,[1,2] Leandro G. Garcia,[1] and Antônio F. Pereira de Araújo[1,*]
[1]*Laboratório de Biologia Teórica, Departamento de Biologia Celular, Universidade de Brasília, Brasília-DF 70910-900, Brazil*
[2]*Instituto de Física, Departamento de Física Geral, Universidade de São Paulo, São Paulo-SP 05508-900, Brazil*
(Received 28 February 2004; revised manuscript received 24 August 2005; published 1 November 2005)

Conformational restrictions imposed by hydrogen bond formation during protein folding are investigated by Monte Carlo simulations of a non-native-centric, two-dimensional, hydrophobic model in which the formation of favorable contacts is coupled to an effective reduction in lattice coordination. This scheme is intended to mimic the requirement that polar backbone groups of real proteins must form hydrogen bonds concomitantly to their burial inside the apolar protein core. In addition to the square lattice, with $z=3$ conformations per monomer, we use extensions in which diagonal step vectors are allowed, resulting in $z=5$ and $z=7$. Thermodynamics are governed by the hydrophobic energy function, according to which hydrophobic monomers tend to make contacts unspecifically while the reverse is true for hydrophilic monomers, with the additional restriction that only contacts between monomers adopting one of $z_h<z$ local conformations contribute to the energy, where z_h is the number of local conformations assumed to be compatible with hydrogen bond formation. The folding transition abruptness and van't Hoff-to-calorimetric-enthalpy ratio are found to increase dramatically by this simple and physically motivated mechanism. The observed increase in folding cooperativity is correlated to an increase in the convexity of the underlying microcanonical conformational entropy as a function of energy. Preliminary simulations in three dimensions, even though using a smaller relative reduction in lattice effective coordination $z_h/z=4/5$, display a slight increase in cooperativity for a hydrophobic model of 40 monomers and a more pronounced increase in cooperativity for a native-centric Go-model with the same native conformation, suggesting that this purely entropic effect is not an artifact of dimensionality and is likely to be of fundamental importance in the theoretical understanding of folding cooperativity.

DOI: 10.1103/PhysRevE.72.051903 PACS number(s): 87.15.Cc, 87.10.+e

I. INTRODUCTION

One of the most remarkable experimental observations on folding thermodynamics of small proteins is the cooperative, two-state character of the process. Global properties that reflect different aspects of the three-dimensional organization of the molecule, like amount of secondary structure, rigid asymmetrical environment around aromatic side chains, burial of tryptophane residues, etc., vary concomitantly within a relatively narrow range of a given control parameter, such as temperature, pressure or denaturant concentration. Moreover, the van't Hoff enthalpy ΔH_{vH}, obtained from the temperature derivative of an effective equilibrium constant taken directly from the value of monitored properties under the assumption of two-state behavior, agrees reasonably well with the actual enthalpy of denaturation obtained calorimetrically, ΔH_{cal} [1]. The value of different monitored properties therefore indeed reflects, to a good approximation, the equilibrium constant between two macroscopically distinguishable thermodynamic states: the native and the denatured macroscopic states. The native state is thermodynamically dominant, i.e., has a lower free energy, at physiological temperature while the denatured state dominates at high temperatures. Around the transition temperature the system absorbs a significant amount of heat, implying abrupt changes in enthalpy and entropy. Since the entropy is the temperature derivative of the Gibbs free energy it is natural to consider the folding process as analogous to a first-order, or two-state, phase transition, as defined by a discontinuity in the first derivative of the thermodynamic potential [2,3].

Simple minimalist protein models, which take chain connectivity and excluded volume into account explicitly, have been able to reproduce protein folding two-state behavior in the sense that the equilibrium population of microstates shifts, as the temperature is increased, from a unimodal distribution dominated by low energy conformations, corresponding to the macroscopic native state, to unimodal distribution dominated by high energy conformations, corresponding to the denatured macroscopic state, passing through, around the transition temperature, a bimodal distribution where both macroscopic states are distinguishable (e.g., Ref. [4], reviewed in Refs. [5–7]). In other words, around the transition temperature there are two free energy minima separated by a free energy maximum along the energy coordinate or, equivalently, there is a convex region in the microscopic entropy as a function of energy [3,8]. It has become apparent, however, that the transition observed in these models tends to be less abrupt than in real proteins and that the ratio between model quantities intended to be analogous to the van't Hoff and calorimetric folding enthapies, with reasonable interpretations of experimental calorimetric data, is usually far below 1 [9–11]. Additionally, it is not trivial to introduce physically motivated terms to normally used energy functions in order to improve cooperativity [11,12].

*Electronic address: aaraujo@unb.br

In the present study we investigate possible entropic effects on protein folding thermodynamic cooperativity of the well known requirement of hydrogen bond formation between backbone carbonyl and nitrogen polar groups as they become buried inside the protein apolar core during folding (e.g., Ref. [13]). We imagine that the number of local backbone conformations adopted by apolar monomers participating in favorable hydrophobic interactions must be smaller than the total number of possible local conformations, since they must be compatible with hydrogen bond formation of adjacent backbone polar groups. This simple idea is implemented in lattice models in which the formation of hydrophobic contacts is coupled to an effective reduction in lattice coordination.

II. METHODS

We use the two-dimensional square lattice, with $z=3$ local conformations per monomer, as well as lattice extensions with diagonal step vectors (covalent bonds) of the form $\{\pm 1, \pm 1\}$ resulting in $z=5$ (when angles between adjacent bonds smaller than 90° are forbidden) and $z=7$. Favorable hydrophobic interactions can only occur between monomers which happen to be in one of $z_h \leqslant z$ local conformations compatible with hydrogen bond formation. The hydrophobic energy function [14], consisting of the sum over monomers of their hydrophobicities multiplied by the number of contacts they are making, now becomes

$$E(\vec{h},\vec{c}') = -\sum_i h_i c_i' = -\vec{h}\cdot\vec{c}', \tag{1}$$

where $\vec{c}'=\{c_1',\ldots,c_N'\}$ represents the number of *viable* contacts made by each monomer in a given structure, where the prime reminds the viability restriction that only contacts involving monomers in one of the z_h local conformations are counted, and $\vec{h}=\{h_1,\ldots,h_N\}$ is the sequence of hydrophobicities, as usual. Note that in this simple implementation the hydrogen bond restriction is actually imposed on all nonneutral interactions, not necessarily hydrophobic.

Three two-dimensional conformations of 24 monomers shown in Figs. 1(a)–1(c) are used as native structures. The structure shown in Fig. 1(a), which will be called n24D2a, has been used in previous studies and was originally selected because it is structurally segregated, i.e., most of its monomers are completely buried or completely exposed, a property already shown, both in two and three dimensions, to be crucial for appropriate native structures to the hydrophobic function [14–16]. Hydrophobicities are either 1 (hydrophobic) or −1 (hydrophilic) for monomers that are respectively buried or exposed in the native structure while the hydrophobicity of both neutral extremities is 0. The resulting native energy is $E^*=-24$. Structures shown in Figs. 1(b) and 1(c), n24D2b and n24D2c, correspond to different contact matrices (the set of all contacts), contact vectors (the number of contacts made by each monomer) although having the same segregation, compactness and energy, for their designed sequences, as n24D2a. Their use is intended to provide an indication of the eventual dependence of our results on topological details of the native structure.

Two sets of local conformations for a given monomer are considered to be compatible with hydrogen bond formation. In the first set, step vectors (covalent bonds) connecting the monomer to its neighbors along the sequence must be in the square lattice, giving $z_h=3$. Since most adjacent step vectors form right angles in the native structures, we also use a set of hydrogen bond conformations where adjacent step vectors must form an angle of 90°, giving $z_h=2$, except for monomers not satisfying this restriction in the native structure. Effective lattice coordination reductions, as measured by z_h/z, ranged therefore from 2/7 (the largest reduction) to 3/3 (the original model with no reduction). Standard Monte Carlo simulations, as detailed elsewhere [4,14,17], were performed but with a move set consisting of end moves and "generalized" kink and crankshaft moves, resulting from the exchange between adjacent and next to adjacent step vectors, respectively. Application to shorter, completely enumerable, chains confirmed that correct thermodynamics are reproduced by this move set not only for $z=3$ but also for $z=5$ and $z=7$. It is important to emphasize that the resulting two-dimensional model is not "native-centric," or "teleological" [9], since it has no explicit bias, either global or local, towards the native structure.

The density of states for each model, $g(E)$, representing the number of model conformations as a function of energy E, was estimated from long Monte Carlo trajectories by standard histogram techniques [18,19]. Heat capacity curves obtained from $g(E)$ were then used to compute two folding quality parameters: the cooperativity index proposed by Klimov and Thirumalai, Ω_c [20], which measures the transition sharpness, or abruptness, and Chan's κ_2 [9], which is intended to be analogous to the ratio between experimental van't Hoff and calorimetric enthalpies and is a measure therefore of the so called calorimetric cooperativity. In our notation [16], with the energy itself taken as order parameter,

$$\Omega_c = \frac{T_f^2 C_V(T_f)}{(-E^*)\Delta T}, \tag{2}$$

where ΔT is the width of the heat capacity peak at its half maximal height, and

$$\kappa_2 = \frac{\Delta H_{\mathrm{vH}}}{\Delta H_{\mathrm{cal}}} = \frac{2T_f\sqrt{C_V(T_f)}}{-E^*}, \tag{3}$$

where the calorimetric enthalpy is $\Delta H_{\mathrm{cal}}=-E^*=24$ in all two-dimensional models since they have the same native energy, −24, and the same average energy, 0, at high temperature.

We have also performed simulations with a chain of 40 monomers in three dimensions whose native structure, called n40D3a and shown in Fig. 1(d), was already used in previous studies [15,21]. The lattice has not been extended in this case, however, and only standard moves have been used. Lattice coordination reduction is therefore less pronounced, with $z_h/z=4/5$, where local conformations corresponding to an angle of 90° between adjacent covalent bonds are assumed to be compatible with hydrogen bond formation. As already described for the two-dimensional model, this restriction is not applied to monomers between covalent bonds forming an angle of 180° in the native structure. In addition

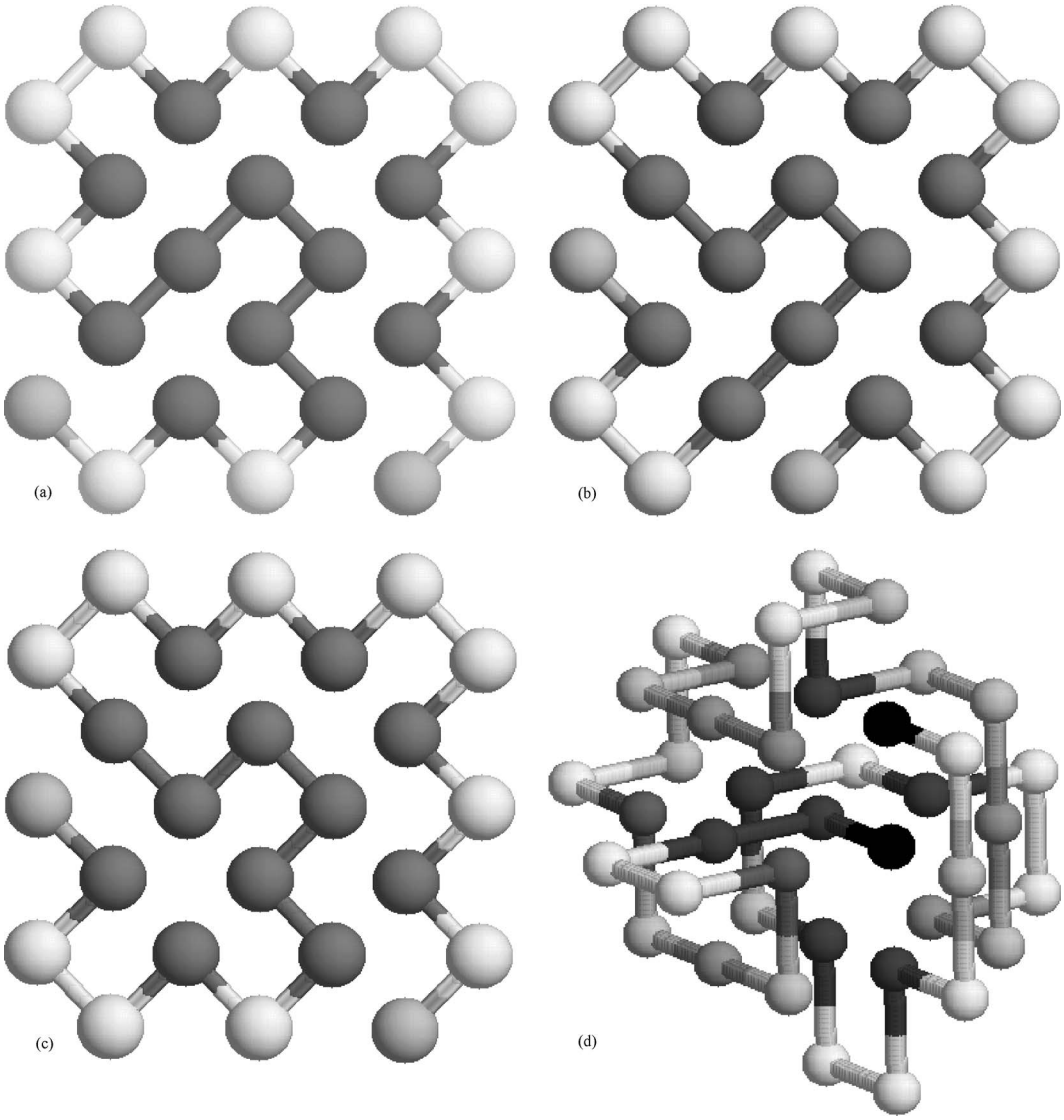

FIG. 1. Two-dimensional conformations of 24 monomers used as native structures in the simulations: n24D2a (a), n24d2b (b), and n24D2c (c). Buried monomers (dark gray) are hydrophobic, exposed monomers (white) are hydrophilic and both extremities (light gray) are neutral. Hydrophobicity values of 1, −1, and 0, for hydrophobic, hydrophilic, and neutral monomers, respectively, result in the same native energy $E^*=-24$. The conformation of 40 monomers shown in (d) was used in all simulations in three dimensions. Monomer hydrophobicities, represented by different tones of gray, were obtained from the deviation in the number of native contacts from their average value in the native structure, resulting in $E^*=-105$. For the Go potential all 33 native contacts have energy −1 while non-native contacts are neutral and $E^*=-33$.

to the hydrophobic energy function, which was previously described to result in a collapse transition above the folding temperature and, as a consequence, in poor calorimetric cooperativity, we have also used a native-centric Go-type energy function for which cooperativity is more pronounced [21].

III. RESULTS AND DISCUSSION

Figure 2 shows heat capacity curves for all six combinations of z and z_h for n24D2a (a), n24D2b (b), and n24D2c (c). Folding transition temperatures T_f, defined by the position of absorption peaks, and corresponding heat capacity maxima $C_V(T_f)$ are shown in Table I for all plots shown in Fig. 2. Corresponding cooperativity parameters Ω_c and κ_2 are shown in the same table. The general behavior is similar for the three two-dimensional native conformations. For fixed z_h, $C_V(T_f)$ becomes larger but T_f becomes smaller as z increases. For fixed z, both T_f, only slightly for $z=5$ and $z=7$, and $C_V(T_f)$ increase when z_h decreases. The quantity

$$C_V(T_f) = \frac{E^{*2}\kappa_2^2}{4T_f^2}, \tag{4}$$

obtained directly from Eq. (3), is also plotted in Fig. 2, with $E^*=-24$, for $\kappa_2=0.6$, $\kappa_2=0.8$, and $\kappa_2=1.0$. These curves permit direct visualization of κ_2 values for different heat capacity peaks and emphasize opposite effects on calorimetric co-

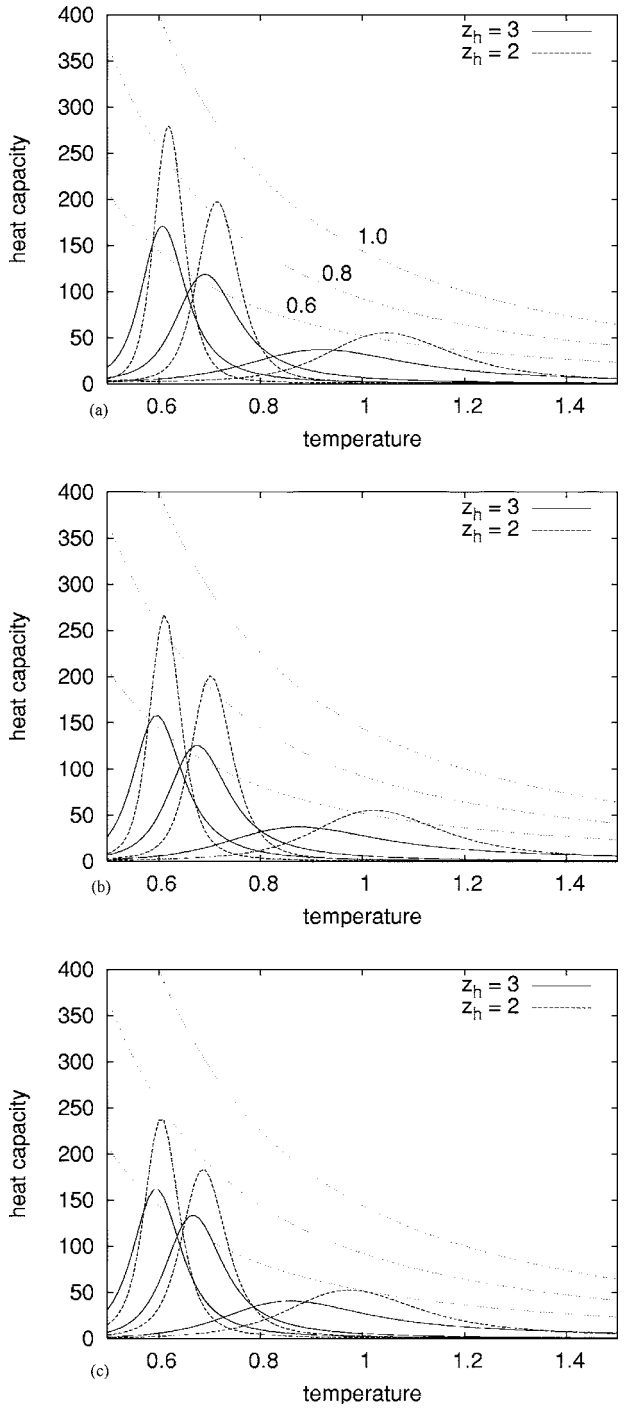

FIG. 2. Heat capacity curves for n24D2a (a), n24D2b (b), and n24D2c (c), obtained by the simple histogram technique from long Monte Carlo trajectories of 10^8 or 10^9 time steps (1 time step equals $N=24$ move attempts), recorded at 10^4 time steps intervals, at temperatures close to T_f. Appropriate conformational sampling was checked by small control simulations at different temperatures. Heat capacity curves are plotted with two types of lines, depending on z_h. For each z_h, z increases from 3 to 5 and then to 7 as the heat capacity maximum shifts to lower temperatures. Monotonically decreasing curves represent the expression of Eq. (4) for κ_2 values of 0.6, 0.8 by 1.0, as labeled in (a), and indicate the height of absorption peaks that would correspond to the given κ_2 value.

TABLE I. For each combination of z and z_h, the table shows T_f and $C_V(T_f)$ obtained directly from the heat capacity curves shown in Fig. 2 and cooperativity parameters κ_2 and Ω_c for the three two-dimensional native conformations under consideration.

z/z_h		3/3	5/3	7/3	3/2	5/2	7/2
T_f	n24D2a	0.92	0.69	0.61	1.05	0.72	0.62
	n24D2b	0.87	0.68	0.60	1.02	0.70	0.61
	n24D2c	0.86	0.67	0.60	0.97	0.69	0.61
κ_2	n24D2a	0.47	0.63	0.66	0.65	0.84	0.86
	n24D2b	0.44	0.63	0.62	0.63	0.83	0.83
	n24D2c	0.46	0.64	0.63	0.59	0.77	0.78
Ω_c	n24D2a	3.3	16.3	25.2	8.5	44.0	64.6
	n24D2b	3.1	16.9	20.1	8.0	42.9	57.0
	n24D2c	3.5	18.4	21.0	6.7	34.7	45.0

operativity of the smaller T_f and larger $C_V(T_f)$ resulting from an increase of z.

Small differences are observed in the exact numerical values, however. Ω_c, which ranges from 10 to 100 in real two-state proteins [20], increases in the case of n24D2a from $\Omega_c(3/3)=3.3$, for the original model with $z=z_h=3$, to $\Omega_c(7/2)=64.6$, for $z=7$ and $z_h=2$ while κ_2 becomes as high as $\kappa_2(7/2)=0.86$ for this most cooperative model, significantly closer to unity than the original value of $\kappa_2(3/3)=0.47$. Folding transition temperatures and cooperativity parameters are consistently lower for n24D2b when compared to n24D2a with the same z/z_h ratio. Maybe more significantly, calorimetric cooperativity does not always increase with z/z_h, as could be suggested by the results for n24D2a. For n24D2b $\kappa_2(7/2)=\kappa_2(5/2)=0.83$ and $\kappa_2(5/3)=0.63 > \kappa_2(7/3)=0.62$. It appears therefore that there is an optimal z/z_h combination which happens to be slightly dependent on native topology. Since the conformational entropy of real unfolded proteins has been estimated from calorimetric data to correspond to around eight conformations per monomer [1] while backbone hydrogen bonds can be formed in α-helical or β-strand local conformations, it is interesting to note the z/z_h values in the present study do not seem unrealistic. Folding temperatures for n24D2c tend to be very similar to the corresponding temperatures for n24D2b but its cooperativity parameters for $z_h=3$ tend to be higher, in some cases being actually even higher than corresponding values for n24D2a [$\kappa_2(5/3)=0.64$, $\Omega_c(3/3)=3.5$, and $\Omega_c(5/3)=18.4$]. For $z_h=2$, however, cooperativity parameters have consistently the lowest values among the three conformations for n24D2c. This observation certainly results from the fact that three monomers are connected by covalent bonds with an angle of 180° in n24D2c [Fig. 1(c)] and are therefore restricted as in $z_h=3$ instead of $z_h=2$, while for the other two conformations this is the case for only one monomer [Figs. 1(a) and 1(b)].

The hydrogen bond effect can be seen more clearly in the entropy function obtained directly from the density of states of the models, $S(E)=\ln(g(E))$, as shown in Fig. 3 for n24D2a. The most obvious effect of increasing z is the ex-

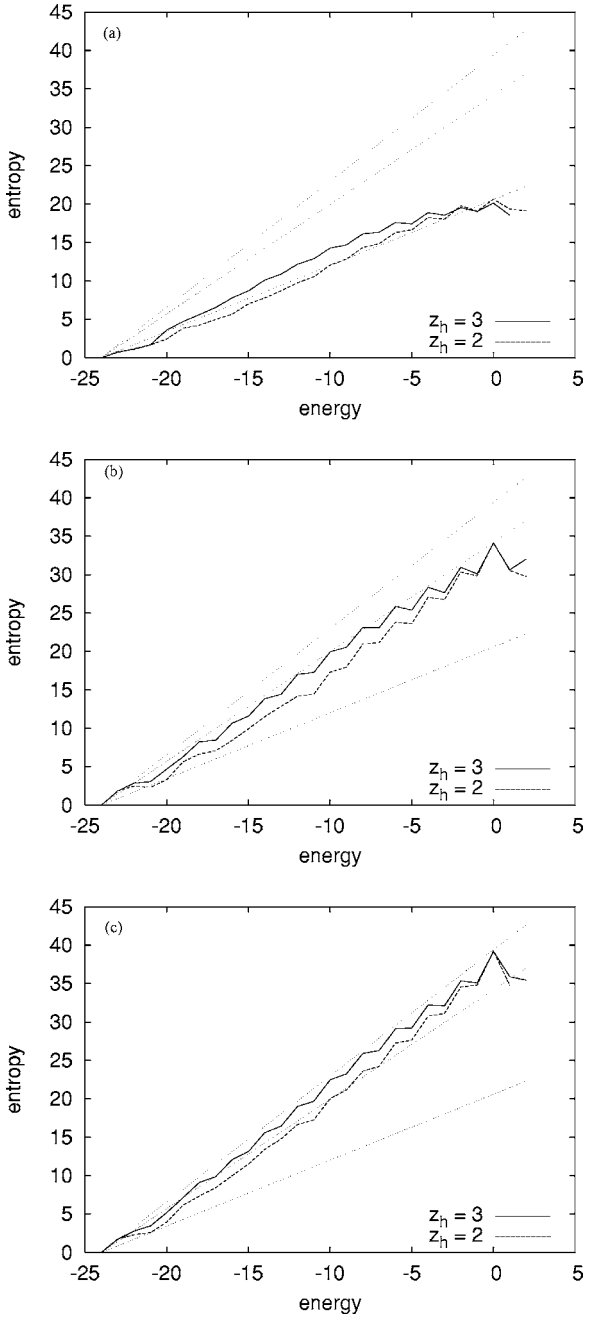

FIG. 3. Microscopic entropies, $S(E)=\ln(g(E))$, for n24D2a obtained by the simple histogram technique and used to generate the heat capacity curves shown in Fig. 2(a). Each panel contains two curves corresponding to $z_h=3$ and $z_h=2$ and a single value of $z=3$ (a), $z=5$ (b), and $z=7$ (c). Straight lines connecting points $(E^*=-24,0)$ and $(E=0,S(0))$, for $z_h=2$, which actually would be essentially identical for $z_h=3$, facilitate the visualization of the increase in concavity as z_h decreases. Repetition of three straight lines in all diagrams facilitates the comparison between different values of z.

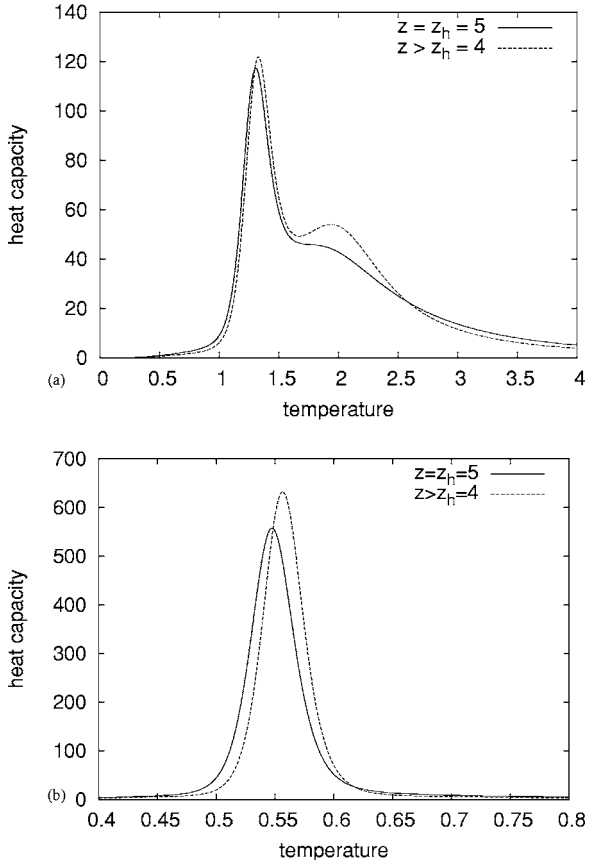

FIG. 4. Heat capacity curves for the three-dimensional model with native conformation n40D3a and hydrophobic (a) and Go (b) potentials. These curves were obtained from long trajectories of up to 4×10^9 time steps (1 time step=N=40 move attempts) at different temperatures by standard histogram techniques. Each panel contains two curves corresponding to the presence or absence of effective lattice coordination reduction. Note the difference in scale between the two panels.

pected increase in the maximal entropy, which happens to correspond to unfolded conformations with $E=0$ in these models, and a directly related decrease in T_f, which in itself contributes to a decrease in cooperativity. Since the entropy of low energy conformations is not affected as much, however, the net result is a significant increase in the convexity of the curve, a sufficiently sharper heat capacity peak and an overall increase in calorimetric cooperativity. This distinctive entropic effect on low energy conformations, when compared to their high energy counterparts, is similar to the entropic reduction of low energy conformations previously observed in a homopolymer model with a more explicit, spinlike consideration of hydrogen bonds [22]. The stronger temperature dependence of Ω_c on z when compared to κ_2, as seen in Table I, is consistent with the major dependence of the transition sharpness on the entropy of the unfolded state [10]. The behavior of the calorimetric cooperativity, on the other hand, appears to be more significantly affected by the convexity itself. Straight lines connecting the points $(E^*=-24,\ S(-24)=0)$ and $(E=0,S(0))$ are shown in order to facilitate visualization of convexities. Their inclinations would correspond to the inverse transition temperatures of perfectly two-state transitions between two macroscopic states with $E=-24$ and $E=0$. We note that the most important feature of the density of states for the cooperative models is not the number of conformations in itself but the convexity of the microscopic entropy curve, $S(E)=\ln(g(E))$. In

particular, the number of conformations does not have to be 0 for essentially all intermediate energy values between $E=E^*$ and $E=0$, as could be suggested by a parabolic (concave) microscopic entropy derived from a hypothetical Gaussian density of states [10], but can actually increase continuously as it is almost unavoidable in any polymeric system.

Figure 4 shows heat capacity curves for the three-dimensional model with the hydrophobic (a) and Go (b) energy functions. As previously described [15,21], the shoulder on the heat capacity curve for the hydrophobic energy function (a) corresponds to a compaction (or expansion) transition at a higher temperature than the folding transition. Calorimetric cooperativity is correspondingly small, with $\kappa_2 \approx 0.27$, because a significant amount of heat is absorbed by the unfolded state while the chain expands [21]. As seen by the second curve in the same panel, lattice coordination reduction does not significantly affect the folding transition peak, and there is only a slight increase in the cooperativity parameter $\kappa_2 \approx 0.28$. The compaction transition shoulder, however, is more significantly affected and becomes more pronounced. This result is very informative since it indicates that the implemented lattice coordination reduction mainly affects compaction, as could actually be expected, and not the folding transition in itself. When both transitions occur concomitantly, however, as for the two-dimensional models described above, an increase in the cooperativity of the compaction transition would automatically result in a corresponding increase in the folding transition. This hypothesis is also consistent with a more significant increase in calorimetric cooperativity upon lattice coordination reduction for the Go potential with the same three-dimensional native structure, as shown in Fig. 4(b). In this case folding and compaction occur concomitantly [21] and cooperativity increases from $\kappa_2 \approx 0.79$ to $\kappa_2 \approx 0.85$.

IV. CONCLUSION

We have investigated the effect of lattice coordination reduction upon chain compaction on the folding cooperativity of minimalist hydrophobic protein models. Coordination reduction is intended to mimic a purely entropic effect resulting from the requirement that polar groups in the protein backbone must form hydrogen bonds concomitantly to their burial inside the apolar internal region of the globular structure. The effective dependence of monomer hydrophobicity on local conformation can be considered to take into some account many-body interactions, being similar in this general sense to the environment-dependent hydrogen bond enthalpies investigated by Chan and co-workers [12]. The explicit modification of conformational entropy additionally suggests that the observed effect might be related to so called "capillarity models," also studied by Chan [10] after ideas previously developed by Freire and co-workers [23], according to which the polymer chain consists of *a priori* cooperative segments. A more extensive investigation, possibly involving sequence optimization [24], different structures, chain lengths, and z_h/z combinations, would be useful to determine the limits of coordination reduction on thermodynamic folding cooperativity and to clarify possible relations with these previously proposed schemes. The corresponding effect on folding kinetics [25] must also be investigated. Additionally, since lattice coordination reduction appears to increase folding cooperativity mainly through its effect on chain compaction, as could actually be expected, studies using even simpler, homopolymeric models are likely to provide interesting insight. Our present results, however, already show that this very simple and physically motivated mechanism can dramatically affect the behavior of minimalist models, suggesting that an analogous entropic effect caused by hydrogen bond formation is likely to be of fundamental importance in any theoretical attempt to understand protein folding cooperativity.

ACKNOWLEDGMENTS

This research has been supported by the Brazilian Government agency "Conselho Nacional de Pesquisa" (CNPq). M.A.A.B. has been supported by FAPESP.

[1] P. L. Privalov, *Protein Folding*, edited by T. E. Creighton, (Freeman, New York, 1992).
[2] M. Karplus and E. I. Shakhnovich, in *Protein Folding*, edited by T. E. Creighton (Freeman, New York, 1992).
[3] A. F. Pereira de Araújo, Protein & Peptide Letters **12**, 223 (2005).
[4] A. F. Pereira de Araújo and T. C. Pochapsky, Folding Des. **1**, 299 (1996).
[5] K. A. Dill, S. Bronberg, K. Yue, K. M. Fiebig, D. Yee, P. D. Thomas, and H. S. Chan, Protein Sci. **4**, 561 (1995).
[6] E. I. Shakhnovich, Curr. Opin. Struct. Biol. **7**, 29 (1997).
[7] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).
[8] M. Hao and H. A. Scheraga, J. Mol. Biol. **277**, 973 (1998).
[9] H. Kaya and H. S. Chan, Proteins **40**, 637 (2000).
[10] H. S. Chan, Proteins **40**, 543 (2000).
[11] H. S. Chan, S. Shimizu, and H. Kaya, Methods Enzymol. **380**, 350 (2004).
[12] H. Kaya and H. S. Chan, Phys. Rev. Lett. **85**, 4823 (2000).
[13] S. Miller, J. Janin, A. M. Lesk, and C. Chotia, J. Mol. Biol. **196**, 641 (1987).
[14] A. F. Pereira de Araújo, Proc. Natl. Acad. Sci. U.S.A. **96**, 12482 (1999).
[15] L. G. Garcia, W. L. Treptow, and A. F. Pereira de Araújo, Phys. Rev. E **64**, 011912 (2001).
[16] Marco Aurélio A. Barbosa and A. F. Pereira de Araújo, Phys. Rev. E **67**, 051919 (2003).
[17] W. L. Treptow, M. A. A. Barbosa, L. G. Garcia, and A. F. Pereira de Araújo, Proteins **49**, 167 (2002).
[18] N. D. Socci and J. N. Onuchic, J. Chem. Phys. **103**, 4732

(1995).
[19] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
[20] D. K. Klimov and D. Thirumalai, Folding Des. **3**, 127 (1998).
[21] L. G. Garcia and A. F. Pereira de Araújo, Proteins (to be published).
[22] J. Borg, M. H. Jensen, K. Sneppen, and G. Tiana, Phys. Rev. Lett. **86**, 1031 (2001).
[23] E. Freire and K. P. Murphy, J. Mol. Biol. **222**, 687 (1991).
[24] A. F. Pereira de Araújo, J. Chem. Phys. **114**, 570 (2001).
[25] H. Kaya and H. S. Chan, Proteins **52**, 510 (2003).