Alternative hydrogen bond implementations produce opposite effects on collapse cooperativity of lattice homopolypeptide models

Gustavo M. N. Fleury
Instituto de Física, Universidade de Brasília, Brasília-DF 70910-900, Brazil
and Laboratório de Biologia Teórica, Departamento de Biologia Celular, Universidade de Brasília, Brasília-DF 70910-900, Brazil

Marco A. A. Barbosa
Instituto de Física Departamento de Física Geral, Universidade de São Paulo, São Paulo-SP 05508-900, Brazil
and Laboratório de Biologia Teórica, Departamento de Biologia Celular, Universidade de Brasília, Brasília-DF 70910-900, Brazil

Antônio F. Pereira de Araújo*
Laboratório de Biologia Teórica, Departamento de Biologia Celular, Universidade de Brasília, Brasília-DF 70910-900, Brazil
(Received 1 June 2007; published 20 November 2007)

We use complete enumeration of self-avoiding chains of up to $N=26$ monomers in two-dimensional lattices to investigate the effect of alternative implementations of backbone hydrogen bonds on the cooperativity of homopolypeptide collapse. Following a recent study on protein folding models, we use the square lattice with $z=3$ local conformations per monomer and lattice extensions containing diagonal steps which result in $z=5$ or $z=7$ and assume that only a subset of $z_h<z$ local conformations is compatible with hydrogen bond formation. As previously observed in heteropolymeric folding, a significant increase in cooperativity, as measured by κ_2 values, results from the coupling between hydrogen bonds and hydrophobic interactions, in such a way that hydrophobic contacts are favorable only when contacting monomers are involved in hydrogen bond formation. For some z/z_h combinations the energy distribution is bimodal at the collapse transition temperature. The situation can be regarded as if all hydrophobic contacts actually decrease the energy by the same amount, $2h$, with the addition of an energetic *increase*, $\epsilon_2=h$, as a penalty for each contacting monomer not satisfying the hydrogen bond condition. Cooperativity is little affected and might even decrease, however, when hydrogen bonds produce a *decrease* in energy by the same amount, $\epsilon_1=h$, for each bonding monomer. For the more general situation when the hydrogen bond effect is not equal, in modulus, to the hydrophobic interaction, i.e., $\epsilon_2\neq h$ or $\epsilon_1\neq h$, we observe a pronounced increase in κ_2 for small ϵ_2, with a maximum around $\epsilon_2/h\approx 1.5$, followed by a gradual decrease to a limiting value at large ϵ_2. The opposite behavior is observed when ϵ_1 is varied. The observed qualitative difference is shown to arise from opposite effects on the convexity of the total density of states of the system when subdensities corresponding to different numbers of hydrogen bonds are differently favored as opposed to the case when subdensities corresponding to different numbers of contacting monomers not forming hydrogen bonds are differently unfavored. Potential implications for the cooperativity of protein folding and protein unspecific collapse are discussed.

DOI: 10.1103/PhysRevE.76.051914 PACS number(s): 87.15.Cc, 87.10.+e

I. INTRODUCTION

We have recently observed that a significant increase in folding cooperativity can arise in protein lattice models when the formation of favorable hydrophobic interactions is coupled to an effective reduction in lattice coordination [1]. This scheme was intended to mimic the reduction in local conformational entropy resulting from the requirement that backbone polar groups must form hydrogen bonds upon insertion into the apolar protein core during the folding process. It was suggested, therefore, that hydrogen bonds might contribute significantly to protein folding two-state behavior. Since the underlying entropic effect is independent of sequence, a similar result could be expected to arise even for homopolymeric collapse or for the rapid unspecific contraction observed for some proteins upon a sudden environmental change from denaturing to folding conditions. Accordingly, the increase in cooperativity was actually more pronounced for collapse than for folding itself in one of the models for which the two transitions were not concomitant [1].

For several simple models investigated during the last decades, with no explicit consideration of hydrogen bonds or secondary structure, homopolymeric coil to globule transitions were found to be continuous, e.g., [2], as well as the initial unspecific collapse observed in some folding heteropolymeric models, e.g., [3–5], which is consistent with the theoretical prediction of a not cooperative, one-state, second-order collapse transition for flexible homopolymers [6–8]. It has also been known, both from theory and simulations, that collapse becomes a cooperative, two-state, first-order transition for "stiff" polymers [6–10] and it is unclear how stiff a more realistic polypeptide model could be. Simulations with the "210" (or "knight") lattice, which is intended to be geometrically consistent with protein secondary structure [11], have also indicated, however, that the homopolypeptide collapse transition, when induced by a combination of uniformly attractive hydrophobic interactions with hydrogen bonds favoring α-helix formation, could be reasonably sharp for some choices of parameters but not two state [12]. Langevin dynamic simulations of off-lattice, fold-

ing heteropolypeptide models, with explicit consideration of favorable hydrogen bonds, have also found the transition from denatured to compact non-native conformations to be not cooperative [13]. It can be considered somewhat surprising, therefore, that experimentally observed time courses of rapid collapse for real proteins [14], and even some nonfolding heteropolypeptides [15], which only became technically possible during the last decade, were found to be consistent with a cooperative, barrier-crossing process, slower than expected for the continuous collapse of flexible polymers.

In the present study we perform an extensive investigation of the effect of alternative hydrogen bond implementations on collapse cooperativity of two-dimensional, exact, lattice homopolypeptide models. We are particularly interested in eventual qualitative thermodynamic differences resulting from distinct assumptions about the dominant net enthalpic contribution of hydrogen bond formation for the stability protein structures. An interesting debate on this issue has taken place in the literature for many years, as reviewed in [16]. Easily implementable, energetically favorable models of hydrogen bonds [12,13] implicitly assume that internal bonds in proteins, between backbone polar groups, are enthalpically more favorable than external bonds, between these groups and water, and do not account for the cost of burying unpaired polar groups. Alternatively, as in the implementation used in our previous study [1], it might be assumed that internal and external bonds are enthalpically similar but the burial of unpaired polar groups is significantly unfavorable [17–19]. Note that both assumptions, although disagreeing about the stability of hydrogen bonds in an aqueous environment, are consistent with enthalpically favorable hydrogen bonds inside the apolar core of protein globules [20] and, therefore, with the plausible hypothesis of a significant reduction in the number of compact relevant conformations imposed by hydrogen bond formation, as corroborated by recent lattice [21,22] and off-lattice simulations [23–25].

We presently observe, however, that very simple models intended to capture these two alternative assumptions might actually display opposite effects regarding the thermodynamic cooperativity of collapse, as measured by the κ_2 parameter proposed by Chan and collaborators [26,27], indicating that this apparently subtle difference might be even more relevant for cooperativity than the geometric details of the polymer chain. The observed qualitative difference is shown to arise from opposite effects on the convexity of the total density of states of the system when subdensities corresponding to different numbers of hydrogen bonds are differently favored as opposed to the case when subdensities corresponding to different numbers of contacting monomers not forming hydrogen bonds are differently unfavored.

II. EXTENDED LATTICE POLYMER MODEL

As in our previous study [1], we use the square lattice with $z=3$ local conformations per monomer and lattice extensions containing diagonal steps which result in $z=5$ and $z=7$ and assume that only a subset of $z_h<z$ local conformations is compatible with hydrogen bond formation (Fig. 1). We use the term "coordination" for this number of possible

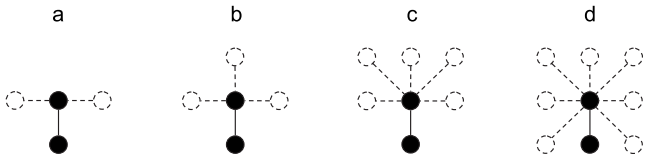

FIG. 1. Possible steps between adjacent beads. In addition to the four vectors $(\pm 1,0)$ and $(0,\pm 1)$ of the usual square lattice, four diagonal vectors $(\pm 1,\pm 1)$ might also be allowed, resulting in variable numbers of local conformations per monomer, or coordination z, as well as two possible distances between adjacent monomers, i.e., 1 and $\sqrt{2}$ lattice units. The figure represents with dotted lines the possible positions for the insertion of a third monomer in a polymer chain, for different values of z used in the present study, after the two previous monomers, represented as filled circles, have already been placed. (a) The lattice with very low coordination $z=2$, with no diagonal steps and adjacent bonds always forming an angle of 90°. (b) The usual square lattice with coordination $z=3$, with no diagonal steps. (c) The extended lattice with coordination $z=5$, with diagonal steps but with no angles between adjacent bonds smaller than 90°. (d) The extended lattice with coordination $z=7$, in which all diagonal steps are allowed with no angle restriction. For a given z/z_h combination, z defines the set of allowed local conformations for each monomer while $z_h<z$ defines the subset of local conformations assumed to be compatible with hydrogen bond formation.

local conformations, z or z_h. For each z/z_h combination, we now consider two alternative definitions of hydrogen bond (α and β) and, for each of them, two alternative forms of energetic contribution to be added to the uniformly attractive hydrophobic interactions. According to definition α, a hydrogen bond is simply counted for each monomer adopting one of the z_h local conformations, independently of the local conformation of any eventual contact partner. It can be imagined that in this case the hydrogen bond interaction is satisfied locally, as in α helices. According to definition β which, in this sense, loosely resembles β sheets, a hydrogen bond is counted only when both monomers forming a contact adopt one of the z_h local conformations. α and β definitions do not differ in the specific z_h local conformations assumed to be compatible with hydrogen bond formation, as would be the case in a more realistic consideration of protein secondary structures. Note that since possible step vectors for the extended lattices are not all equal in length, the distance between two given adjacent beads "fluctuates" between 1 and $\sqrt{2}$ lattice units, resembling therefore other "fluctuating bond" polymer models previously used in protein folding studies, e.g., [28]. The use of variable distances to model a polypeptide system, in which virtual bonds between adjacent C_α atoms are known to be constant, might be considered as a small price to be paid in order to overcome intrinsic limitations of the lattice to otherwise reproduce other properties of real polypeptides expected to be relevant for the problem under investigation, such as geometrically realistic protein backbone conformations [28] or, as in the present study, a large reduction in local conformational entropy upon hydrogen bond formation [1].

The two forms for the energetic contribution from hydrogen bonds are intended to reflect alternative assumptions re-

garding the net enthalpic effect of these interactions in protein structures. In the simplest implementation, which is called E_1 in the present study, the energy of the model is *decreased* by a fixed amount, ϵ_1, for each monomer involved in a hydrogen bond. It is implicitly assumed, therefore, that internal bonds, between protein backbone polar groups, are enthalpically more favorable than external bonds, between these groups and water, by a fixed amount which is independent of their local environment. The unavoidable decrease in the number of available water molecules resulting from monomer burial or, equivalently, the cost of burying unpaired polar groups, is therefore neglected. In the second implementation, which is called E_2, it is assumed that internal and external bonds are enthalpically equivalent but the burial of unpaired polar groups is unfavorable. Accordingly, the energy of the model is *increased* by an amount ϵ_2 for each monomer involved in a contact for which a hydrogen bond is not satisfied. Hydrophobic interactions for each monomer i are always implemented simply as a decrease in energy by the fixed amount h for each contact in which it is involved, independently of hydrogen bond formation.

More specifically, the two energy functions are

$$E_1 = \sum_{i=1}^{N} (-hc_i - \epsilon_1 b_i) = -hC - \epsilon_1 B \tag{1}$$

and

$$E_2 = \sum_{i=1}^{N} (-hc_i + \epsilon_2 c_i'') = -hC + \epsilon_2 C'', \tag{2}$$

where the sum is over all monomers, labeled by i, c_i is the total number of contacts in which monomer i is involved, with $\Sigma_i c_i = C$ being 2 times the total number of contacts, b_i is the number of hydrogen bonds in which it is involved, with $\Sigma_i b_i = B$ being the total number of α hydrogen bonds or 2 times the total number of β hydrogen bonds, and $c_i'' \leq c_i$ is the number of contacts counted in c_i for which a hydrogen bond is *not* satisfied, with $\Sigma_i c_i'' = C''$. The two energy forms can be considered as particular examples of the more general expression

$$E = \sum_{i=1}^{N} (-hc_i - \epsilon_1 b_i + \epsilon_2 c_i'') \tag{3}$$

but we have not investigated the general situation in which both $\epsilon_1 \neq 0$ and $\epsilon_2 \neq 0$. When $\epsilon_1 = \epsilon_2 = 0$ the situation is trivially reduced to the original hydrophobic function $E = -\vec{h} \cdot \vec{c}$ [29,30] which becomes $E = -hC$ for the present homopolymer case.

Combining the two definitions of hydrogen bond with Eqs. (1) and (2) we obtain four model possibilities for each z/z_h combination: α_1, α_2, β_1, and β_2. According to definition α the presence or absence of a hydrogen bond depends only on the local conformation of a single monomer and is independent of any contact formation. There are, therefore, only two possible numbers of hydrogen bonds for each monomer i, $b_i = 0$, and $b_i = 1$. Even for the extended lattices, we only consider monomer j to form a contact with monomer i, $1 < i < N$, if $|j-i| > 1$ and if monomer j occupies one of the four nearest-neighbor sites of the site occupied by monomer i. While for $z=2$ and $z=3$ only two of these sites might be occupied by other monomers than neighboring $i-1$ and $i+1$, which implies $0 \leq c_i \leq 2$, for the extended lattices with $z=5$ and $z=7$ all these four sites might be occupied by nonneighboring monomers, or $0 \leq c_i \leq 4$. The number of contacts for which a hydrogen bond is *not* satisfied for the α definition is either $c_i'' = c_i$, when $b_i = 0$, or $c_i'' = 0$, when $b_i = 1$. Note that only for α_1 there are energetically favorable conformations with no contacts, i.e., when $C=0$ and $B>0$. This situation would be analogous to stable α helices in the absence of long-range interactions. For α_2, on the other hand, $C=0$ implies $E=0$, since $C'' \leq C$.

According to definition β the presence or absence of a hydrogen bond depends on the local conformation of both monomers involved in a contact. In this case b_i is counted as the number of contacts satisfying the hydrogen bond condition in which monomer i is involved and $c_i'' = c_i - b_i$. $b_i \leq c_i$ can, in particular, be larger than 1. Note that if $\epsilon_2 = h$ in Eq. (2) the energy reduces to

$$E_2 = \sum_{i=1}^{N} -h(c_i - c_i'') = \sum_{i=1}^{N} -hc_i' = -hC', \tag{4}$$

where $c_i' \equiv c_i - c_i''$ is the number of "viable" contacts, with $\Sigma_i c_i' = C'$, which corresponds to the expression used in our previous study [1]. It might also be convenient to call the contacts counted in c_i'' as "not viable." Note that only for the β definition the number of viable contacts coincides with the number of hydrogen bonds, i.e., $c_i' = b_i$. For the α definition we have either $c_i' = 0$, when $b_i = 0$, or $c_i' = c_i$, when $b_i = 1$. Also note from Eqs. (1), (2), and (4) that hydrophobic interactions for E_2, but not for E_1, are effectively "coupled" to hydrogen bond formation, in the sense that the hydrophobicity of each monomer effectively depends on whether or not it is forming a hydrogen bond. For this reason we also refer to energy schemes E_1 and E_2 as corresponding to "uncoupled" and "coupled" hydrogen bonds, respectively. α and β hydrogen bonds, both coupled and uncoupled, are illustrated in Fig. 2 for a particular structure of 16 monomers with $z/z_h = 7/3$. An assumed dependence of effective hydrophobicity on hydrogen bond formation has been recently investigated in the context of a geometrically realistic off-lattice protein model [31].

III. RESULTS

In the present study we fix the value of the hydrophobic interaction, h, as the unit of energy and investigate the thermodynamic behavior of the models as a function of the hydrogen bond strength, ϵ_1 or ϵ_2, and chain length, N. Complete enumeration of the conformational space of the four models (α_1, α_2, β_1, and β_2) was performed for all combinations of z/z_h, for chain lengths of sizes up to $N=26$ ($z=3$ and $z=2$), $N=18$ ($z=5$) and $N=16$ ($z=7$). Cooperativity is quantified by κ_2, as proposed by Chan and co-workers [26,27], which is intended to be analogous to the ratio between experimental van't Hoff and calorimetric unfolding enthalpies

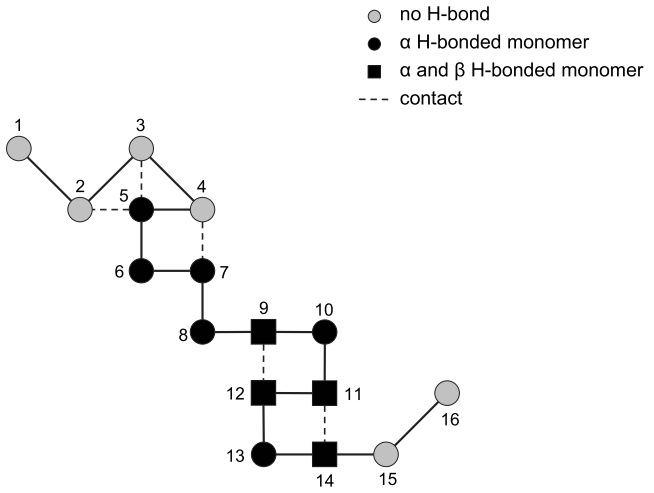

FIG. 2. Illustration of α and β hydrogen bonds, both coupled and uncoupled, for $z/z_h=7/3$ in a conformation of 16 monomers. Diagonal steps are allowed, because $z=7$, but, since $z_h=3$, only monomers 5 to 14, which are connected to both neighbors by nondiagonal steps, are assumed to be compatible with hydrogen bond formation. According to the α definition of hydrogen bond, therefore, one hydrogen bond is counted for each of these monomers, which are colored black. From these 10 compatible monomers only monomers 9, 11, 12, and 14, shown as squares, are also contacting compatible monomers and, therefore, are the only monomers involved in hydrogen bond formation according to the β definition. For this structure, $C=10$ is 2 times the total number of contacts. In order to use Eqs. (1) and (2) to compute the energy of this conformation for $\epsilon_1=h=1$ or $\epsilon_2=h=1$ we note that, for the α definition, $B=10$ and $C''=c_2''+c_3''+c_4''=1+1+1=3$, resulting in $E_1^\alpha=-20$, for uncoupled hydrogen bonds, and $E_2^\alpha=-7$, for coupled hydrogen bonds. For the β definition $B=4$ and $C''=c_2''+c_3''+c_4''+c_5''+c_7''=1+1+1+2+1=6$, resulting in $E_1^\beta=-14$ and $E_2^\beta=-4$ for coupled and uncoupled hydrogen bonds, respectively.

[32]. In our notation, with model energies assumed to correspond to experimental enthalpies, the expression for κ_2 is [1,30],

$$\kappa_2 = \frac{\Delta H_{\mathrm{vH}}}{\Delta H_{\mathrm{cal}}} = \frac{2T_{\max}\sqrt{C_V(T_{\max})}}{\Delta E^*}, \tag{5}$$

where $C_V(T_{\max})$ is the heat capacity at temperature $T_{\max}$, where its peak is located, and the calorimetric enthalpy $\Delta H_{\mathrm{cal}}=\Delta E^*$ is now the difference between the average energy at infinite temperature, which must be computed for each model, and the minimal energy. The heat capacity for some models displays two peaks and in this case two κ_2 values are computed although only the larger value is usually considered for analysis. Note that κ_2 values close to 1, which are considered to resemble proteinlike cooperativity, imply a bimodal energy distribution at the transition temperature. The reverse is not necessarily true, however, when the positions of the energy peaks happen to be significantly dependent on temperature [27,33]. Cooperativity measured by κ_2, as in the present study, has also been called "calorimetric cooperativity" [26,27,30,33] in order to be distinguished from previously used, less restrictive criteria based on the "abruptness" of the transition, e.g., [34].

Figure 3(a) shows collapse cooperativity as a function of chain length for the four lattice variations used in the present study, with no hydrogen bonds. For each coordination value, z, κ_2 tends to decrease with chain length at small N but no clear dependence is observed for $N\gtrsim 15$ for $z=3$ and $z=2$ while for $z=5$ and $z=7$ large conformational spaces have prevented enumeration for $N>18$ and $N>16$, respectively. Cooperativity also tends to increase as z decreases, being as high as 0.6 for $z=2$, with $N=20$. The average number of contacts as a function of temperature for different values of z and fixed chain size, $N=16$, is shown in Fig. 3(b). Heat capacity curves for these four models, normalized through the division by $(\Delta E^*)^2$, are shown in Fig. 3(c). Note, from Eq. (5), that the normalization procedure permits direct comparison between cooperativity values for the different curves. Decreasing lines in the same plot indicate the height of a normalized heat capacity peak that would correspond to specific κ_2 values at different temperatures. Energy distributions, which in the present case are identical in shape to contact distributions, at the temperature of the major heat capacity peak are shown for the same models in Fig. 3(d).

Some oscillations along N are apparent in Fig. 3(a) and they can be partly attributed to the existence of a second peak in the heat capacity curve interfering with the relevant collapse peak at specific chain sizes, an extreme example of which is illustrated by the normalized heat capacity curves for $z=3$ shown in Fig. 4(a). Note that in addition to the primary peak corresponding to chain collapse a secondary peak might appear at a small temperature for some values of N, as illustrated by the curve for $N=20$. When this is the case, a corresponding second cooperativity value, smaller than the collapse cooperativity from the major peak, can also be computed. For $N=9$ and $N=12$, however, the primary peak is masked by this secondary peak, virtually disappearing in the first case and becoming a shoulder in the second, and only the cooperativity of the secondary peak can be computed. As seen in Fig. 4(b), κ_2 values observed for these chain sizes are unexpectedly low when compared to the rest of the curve shown in Fig. 3(a), which displays only the largest cooperativity value for each N, but are consistent with values for the secondary peak.

In order to investigate the effect of hydrogen bond implementations, α_1, α_2, β_1, and β_2, for different z/z_h combinations, on collapse cooperativity we computed κ_2 values for these models with varying N and fixed $\epsilon_1=1$ or $\epsilon_2=1$ and also with varying hydrogen bond strength and fixed N. On the left-hand side of Fig. 5 we observe the behavior of κ_2 as a function of N for (a) $z/z_h=7/3$ and (c) $z/z_h=5/3$, with different lines corresponding to different hydrogen bond implementations. In order to facilitate the comparison between different curves, cooperativity for simple collapse, with no hydrogen bonds, in lattices corresponding to z (lower dotted line) and z_h (upper dotted line), reproduced from Fig. 3(a), are also plotted in these two panels. For this somewhat arbitrary choice of hydrogen bond "strength" the β_2 implementation is consistently the most cooperative, followed by α_2, β_1, and α_1. For $z/z_h=7/3$, as seen in Fig. 5(a), κ_2 is around 0.6 for the β_2 implementation and $N=16$, which rep-

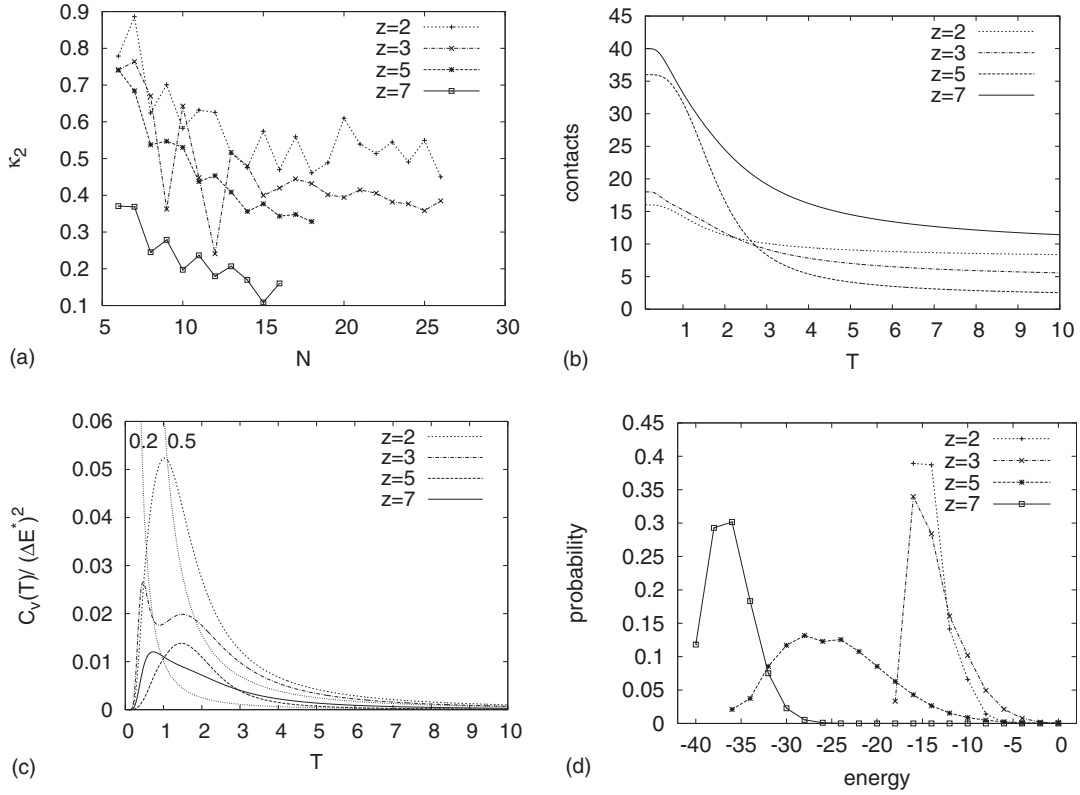

FIG. 3. Thermodynamic characterization of collapse for simple homopolymers, without hydrogen bonds, for different coordination's z. (a) κ_2 dependence on chain size with different lines corresponding to different coordination's z. (b) The average number of contacts as a function of temperature, (c) the normalized heat capacity, $C_V(T)/(\Delta E^*)^2$, as a function of temperature, and the equilibrium energy distribution at $T_{\max}$, the temperature of the primary heat capacity maximum is shown for different values of z and fixed $N=16$. Decreasing lines in (c) indicate the height of a normalized heat capacity corresponding to the indicated cooperativity values.

resents a significant increase when compared to values lower than 0.2 observed in the simple collapse for $z=7$ and the same chain size, or to values around 0.4 observed for the square lattice with $z=3$. κ_2 values around 0.4 for $N=16$ are also observed for the α_2 implementation. For the β_1 implementation the whole curve is very similar to the dotted line corresponding to the simple collapse for $z=7$ while κ_2 values tend to be even smaller than this low reference line for the α_1 implementation. As seen in Fig. 5(c), κ_2 values for large N are also around 0.6 for the β_2 implementation when $z/z_h=5/3$. The general behavior observed for $z/z_h=7/3$ and $z/z_h=5/3$ is also very similar to what was found for lattice reductions $z/z_h=7/2$ and $z/z_h=5/2$, respectively (not shown). Cooperativity for small effective reductions in lattice coordination, either $z/z_h=7/5$ or $z/z_h=3/2$, is always lower than for the original square lattice (not shown).

The effect of hydrogen bond strength on cooperativity for different models is shown on the right-hand side of Fig. 5 for (b) $z/z_h=7/3$ and (d) $z/z_h=5/3$, with different lines in each panel corresponding to different hydrogen bond implementations. The value of N in these two panels corresponds to the maximal chain size for which enumeration was performed for the particular lattice reduction, i.e., (b) $N=16$ and (d) $N=18$. The common initial cooperativity value for all curves, at $\epsilon=0$, corresponds to the lattice with coordination z and no hydrogen bonds. It is apparent that the general behavior of the curves in these two panels is more strongly determined by the energy definitions, E_1 or E_2, corresponding to the absence or presence, respectively, of coupling between hydrogen bonds and hydrophobic interactions, than on the particular definition of hydrogen bond, α or β. For all models with coupled interactions, there is a peak of cooperativity somewhere in the range $1 \lesssim \epsilon_2 \lesssim 3$ followed by a monotonic decrease for large ϵ_2. α_2 hydrogen bonds tend to result in slightly broader and right-shifted peaks when compared to β_2 hydrogen bonds. In the case of uncoupled interactions, on the other hand, maximal cooperativity appears to be approached asymptotically as ϵ_1 increases. The contrasting behavior of coupled versus uncoupled interactions is clearly visualized for (b) $z/z_h=7/3$, in which case sharp cooperativity peaks are observed for β_2 and α_2 hydrogen bonds, reaching values of $\kappa_2 \approx 0.6$ at $\epsilon_2 \approx 1$ and $\kappa_2 \approx 0.45$ at $\epsilon_2 \approx 1.5$, respectively, while no peaks are observed for uncoupled interactions. κ_2 maximal values for large ϵ_1 are around 0.6 and 0.3 for β_1 and α_1 hydrogen bonds, respectively. Similar behavior is observed for (d) $z/z_h=5/3$, but in this case not only β_2 and β_1, but also α_2 shows cooperativity as high as $\kappa_2 \approx 0.6$. Note that cooperativity tends to decrease, or remain constant, in an initial model-dependent range of small hydrogen bond strengths.

Since the highest values of κ_2 observed for these models are never higher than 0.6 and, therefore, still low when compared to proteinlike folding cooperativity of $\kappa_2 \approx 1$, it is important to verify if the large variations in κ_2 values shown in

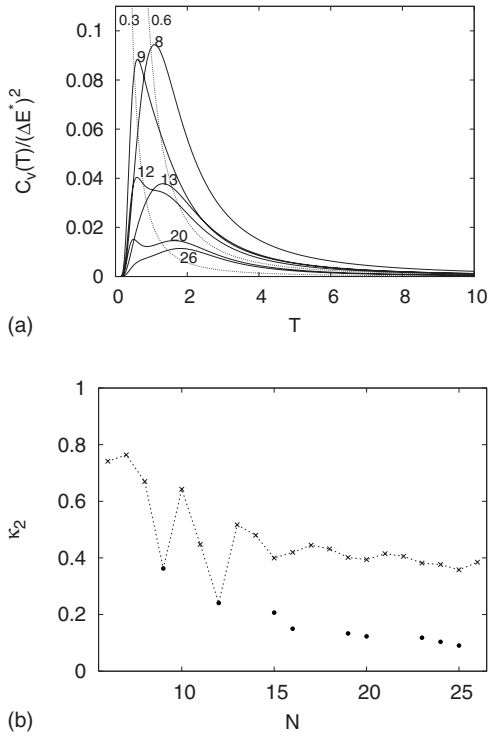

FIG. 4. Cooperativity dependence on chain size for the simple homopolymer, without hydrogen bonds, in the square lattice, corresponding to $z=3$. (a) Normalized heat capacity curves are shown for selected chain lengths, N, as indicated near each curve, with decreasing dotted lines corresponding to the height of normalized heat capacity peaks that would correspond to $\kappa_2=0.3$ and $\kappa_2=0.6$. (b) Calorimetric cooperativity, κ_2, for primary (crosses) and secondary (filled dots) heat capacity peaks as a function of the polymer size. The dashed line connecting the largest values for each N is the line shown in Fig. 3(a) for $z=3$.

Fig. 5 do indeed reflect qualitatively different thermodynamic behaviors, from continuous, one-state, collapse transitions for small κ_2 values, to cooperative, two-state, transitions for large $\kappa_2 \approx 0.6$. Equilibrium energy distributions at increasing temperatures might be used to verify this hypothesis, since one-state transitions, characterized by a continuous shift of a unimodal energy distribution, will be distinguishable from redistribution between modes of a bimodal distribution, which characterizes two-state transitions. In Fig. 6 we show the temperature dependence of the energy equilibrium distributions for the four hydrogen bond implementations, with (a)–(d) $z/z_h=7/3$ and $\epsilon_1=1$ or $\epsilon_2=1$, as well as (e) and (f) β_1 and β_2 with $\epsilon_1=3$ or $\epsilon_2=3$. Note that each panel, therefore, corresponds to a single point in the curves shown in Fig. 5(b), as indicated by matching symbols in both figures. Each panel in Fig. 6 shows the energy distribution at the temperature of maximal heat capacity, $T_{\max}$, as a histogram and five other distributions, from left to right, corresponding to the temperatures at which 10%, 25%, 50%, 75%, and 90% of the total heat of the transition has already been absorbed. The κ_2 value for the corresponding model is also indicated, as well as the matching symbol referring to Fig. 5(b). Consistently with their very low $\kappa_2 < 0.2$ values, it is clear that distributions for (a) α_1 and (c) β_1, for $\epsilon_1=1$, continuously shift to higher energies as temperature increases, as expected for a not cooperative, one-state transition. For (d) β_2, with high $\kappa_2 \approx 0.6$, a two-state transition is apparent from the redistribution among modes of a bimodal distribution. Also note that for (d) β_2, but not for (a) α_1 or (c) β_1, the distribution at $T_{\max}$ is almost coincident with the distribution corresponding to 50% of the heat already absorbed and it is bimodal. This coincidence between these two distributions is also observed for (b) α_2, with $\kappa_2 \approx 0.4$, but in this case it is not bimodal. The increase in κ_2 for β_1 when hydrogen bonds become strong, when compared to hydrophobic interactions, is indeed related to a qualitative change to a two-state transition, as illustrated for (e) $\epsilon_2=3$. The decrease in κ_2 for β_2 in this regime, on the other hand, appears to result primarily from a significant amount of heat still absorbed after the transition has taken place and not from a change to a one-state transition, as illustrated for (f) $\epsilon_2=3$. Low κ_2 values associated to a strong temperature dependence of the high-energy mode in bimodal energy distributions have already been described for protein models [33,35].

IV. DISCUSSION

The effect of alternative implementations of hydrogen bonds on the collapse cooperativity of homopolypeptides was investigated by complete enumeration of two-dimensional self-avoiding chains restricted to different lattices. Despite obvious limitations such as short chain lengths, lack of geometric detail, and space dimensionality, these models are consistent with a significant reduction in local conformational entropy upon hydrogen bond formation, which is an expected property of real hydrogen bonds between different groups of a single polypeptide chain. Additionally, their simplicity permits an extensive investigation of two alternative basic assumptions about the net enthalpic contribution from hydrogen bonds to the stability of protein structures. Interestingly, it is shown that collapse cooperativity might be strongly dependent on these assumptions.

The dependence of collapse cooperativity on lattice coordination z for the simple homopolymer models, with no hydrogen bonds, is likely to be related to chain stiffness, since for a given chain length cooperativity increases as z decreases, as seen in Fig. 3(a), while stiffness might be expected to increase. Other more specific lattice details which, in addition to z, might also contribute to chain stiffness are not able, therefore, to affect the expected order of cooperativity for the models under investigation. The somewhat unexpected order, if based on z alone, of average number of contacts at high temperatures for the four lattices, on the other hand, like the apparently small number for $z=5$ and high number for $z=2$ shown in Fig. 3(b), are likely to reflect more specific details of the lattice, such as the values of possible angles between adjacent bonds.

The level of cooperativity obtained with different hydrogen bond implementations for a given value of lattice coordination reduction, z/z_h, and $\epsilon_1=1$ or $\epsilon_2=1$, does not appear to be trivially related to the cooperativity obtained for the simple homopolymers in lattices with coordination's z or z_h, as shown on the left-hand side of Fig. 5. For models with a

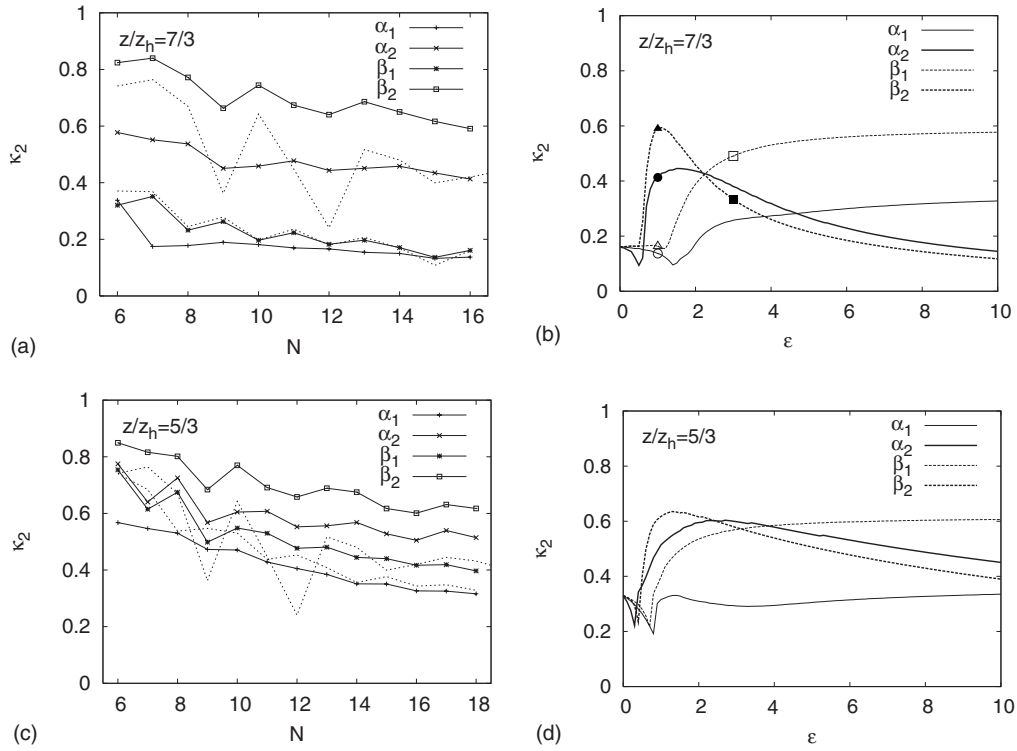

FIG. 5. κ_2 dependence on chain size, N [left-hand side (a) and (c)], and hydrogen bond strength, ϵ_1 or ϵ_2 [right-hand side (b) and (d)], for different hydrogen bond implementations (α_1, α_2, β_1, and β_2). Hydrogen bond strength in (a) and (c) is equal, in modulus, to the hydrophobic interaction, $\epsilon_1=1$ or $\epsilon_2=1$. Chain length is $N=16$ in (b) and $N=18$ in (d). Each panel corresponds to a given lattice reduction: $z/z_h=7/3$ [first row (a) and (b)] and $z/z_h=5/3$ [second row (c) and (d)]. For comparison, curves corresponding to simple homopolymer collapse, without hydrogen bonds, are also shown with dotted lines in (a) and (c) for lattices with coordination z (lower) and z_h (upper). Symbols in (b) indicate the points that were used to calculate energy distributions in Fig. 6 and densities of states in Fig. 7.

large reduction ($z/z_h=7/3$, $z/z_h=7/2$, $z/z_h=5/3$, and $z/z_h=5/2$) β_2 hydrogen bonds result in higher cooperativity than for the simple homopolymer in the lattice with coordination $z=3$ or $z=2$, while for α_2 hydrogen bonds cooperativity tends to be similar or slightly higher than for the lattice with coordination $z=3$ although lower than for the lattice with small coordination $z=2$. Coupled hydrogen bonds, therefore, when $\epsilon_2=1$, clearly tend to increase cooperativity with respect to the collapse of simple homopolymers in the original square lattice. The situation is reversed for uncoupled hydrogen bonds, with $\epsilon_1=1$, since in this case cooperativity tends to be similar or even smaller than for simple flexible homopolymers in high coordination lattices $z=5$ or $z=7$.

The stronger dependence of the qualitative behavior of the system on the energy function definition, coupled versus uncoupled, than on the hydrogen bond definition, α versus β, is even more evident on the effect of hydrogen bond strength on collapse cooperativity shown on the right-hand side of Fig. 5. Also note that in geometrically realistic protein models only the energy definition is likely to be open to choice, since both α and β hydrogen bonds should naturally arise in the same simulation depending on the distance along the sequence between hydrogen bond donor and acceptor. For the situation in which the hydrogen bond effect, ϵ_1 or ϵ_2, is comparable in magnitude to the hydrophobic interaction, coupled hydrogen bonds increase cooperativity while uncoupled hydrogen bonds have little effect or even decrease it, as discussed above. This result is consistent with previous observations of the not-cooperative collapse for polypeptide models with energetically favorable hydrogen bond interactions [12,13]. When the hydrogen bond effect becomes stronger, the situation is reversed, since κ_2 values for coupled hydrogen bonds decrease while they increase for uncoupled hydrogen bonds and, specifically for β_1, reach values as high as the highest values observed for coupled bonds. It is apparent, however, that this situation is likely to be less relevant for realistic models since very strong hydrogen bonds, when compared to hydrophobic interactions, would be likely to prevent collapse, resulting in long single segments of secondary structure, e.g., [12,24].

The connection between κ_2 values, shown in Fig. 5, and the shape of the energy distribution at $T_{\max}$, shown in Fig. 6, can be better understood when it is considered that

$$C_V(T_{\max}) = \left(\frac{\partial E}{\partial T}\right)_V\Bigg|_{T=T_{\max}} = \frac{\sigma_E^2(T_{\max})}{T_{\max}^2} \tag{6}$$

or, from Eq. (5),

$$\kappa_2 = \frac{2\sigma_E(T_{\max})}{\Delta E^*}, \tag{7}$$

where $\sigma_E(T_{\max})$ is the energy standard deviation at $T_{\max}$ [27,35]. It is clear, therefore, that the maximal possible value of $\kappa_2=1$ corresponds to $\sigma_E(T_{\max})=\Delta E^*/2$, or a perfectly bimodal energy distribution at $T_{\max}$, equally divided between

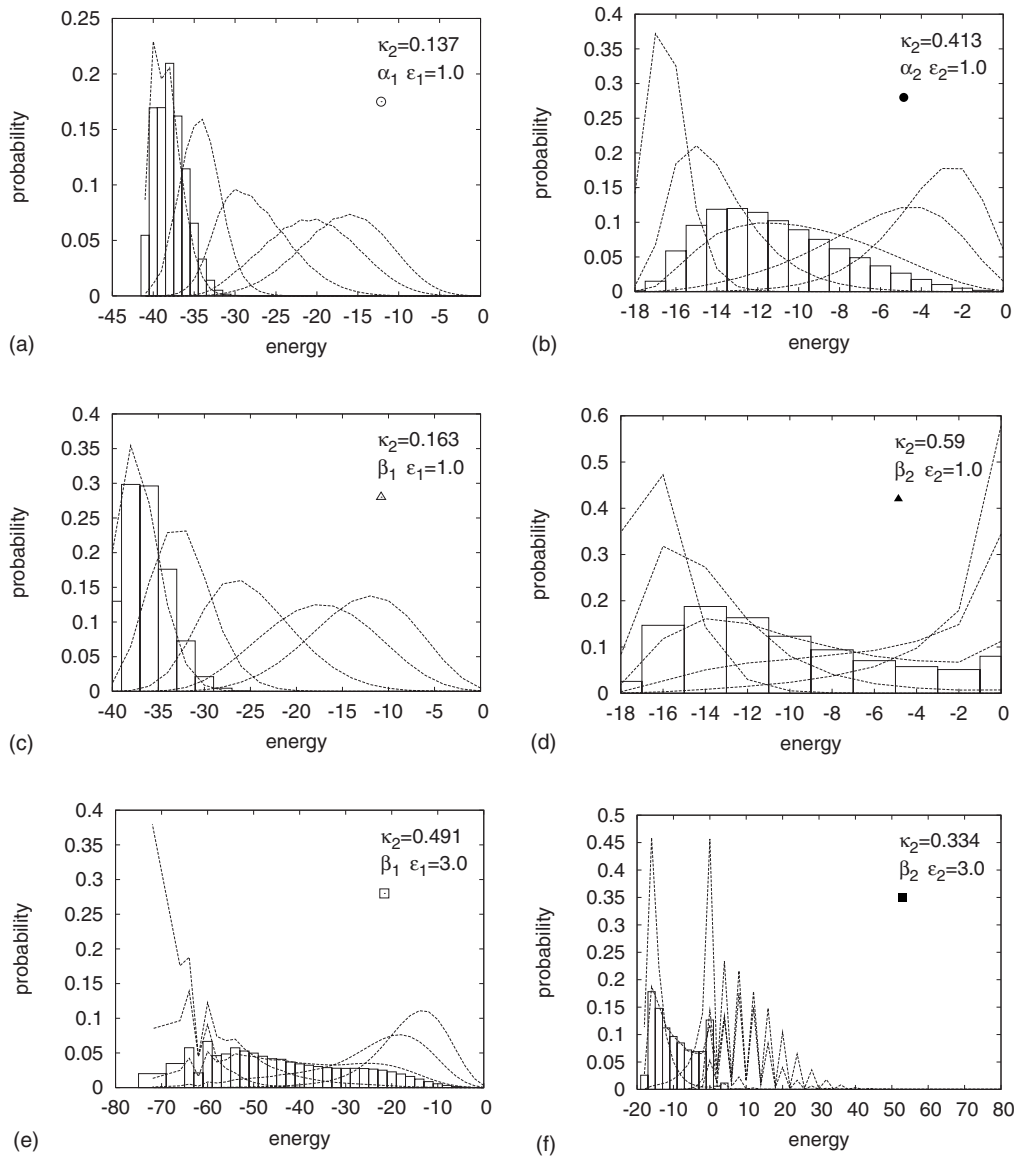

FIG. 6. Equilibrium energy distributions for different hydrogen bond implementations with fixed lattice coordination reduction $z/z_h=7/3$ and chain length $N=16$. In each panel, the energy distribution at $T_{\max}$, the temperature of maximal heat capacity, is shown in boxes while five dashed curves, from left to right, correspond to temperatures at which 10%, 25%, 50%, 75%, and 90% of the total heat of the transition has already been absorbed. The different panels correspond to α_1 with (a) $\epsilon_1=1.0$, α_2 with (b) $\epsilon_2=1.0$, β_1 with (c) $\epsilon_1=1.0$, β_2 with (d) $\epsilon_2=1.0$, β_1 with (e) $\epsilon_1=3.0$, and β_2 with (f) $\epsilon_2=3.0$. κ_2 values for each model are also indicated. Each panel has a symbol matching the corresponding point in the curves shown in Fig. 5(b).

minimal and maximal possible energies, with no conformations with intermediate energy values. The minimal value $\kappa_2=0$, on the other hand, corresponds to all conformations having the same energy, or $\sigma_E=0$. Since the distinction between one-state and two-state transitions is based on the distribution being bimodal or unimodal, it is useful to consider the special value of κ_2 corresponding to a flat energy distribution with all energy levels equally populated at $T_{\max}$, which happens to be, if energy is considered to be a continuous variable, $\kappa_2^{\#}=2\sqrt{(\int_0^1 x^2dx)-(\int_0^1 xdx)^2}$, or $\kappa_2^{\#}=2\sqrt{1/3-1/4}=0.577$. If $\kappa_2>\kappa_2^{\#}$ then the distribution at $T_{\max}$ must be bimodal, as in Fig. 6(d). The reverse is not necessarily true, however, because heat absorption above $T_{\max}$ might increase ΔE^* and result in $\kappa_2<\kappa_2^{\#}$ even if the distribution is clearly bimodal at $T_{\max}$, as in Fig. 6(f) and, to some extent, Fig. 6(e).

The intrinsic qualitative difference between coupled and uncoupled hydrogen bonds must result from their different effects on the logarithm of the density of states of the system, or microcanonical entropy, whose convexity or concavity is intimately connected to the existence or not, respectively, of a cooperative, two-state transition [33,36,37]. Since the number of conformations at any given energy results from the sum of conformations with different numbers of either hydrogen bonds, B, or unviable contacts, C'', the whole density of states can also be considered as the sum of "subdensities," each of them characterized either by the same B or C''. Complete enumeration, for $z/z_h=7/3$, of all conformations for all possible numbers of contacts, C, as well as for the numbers

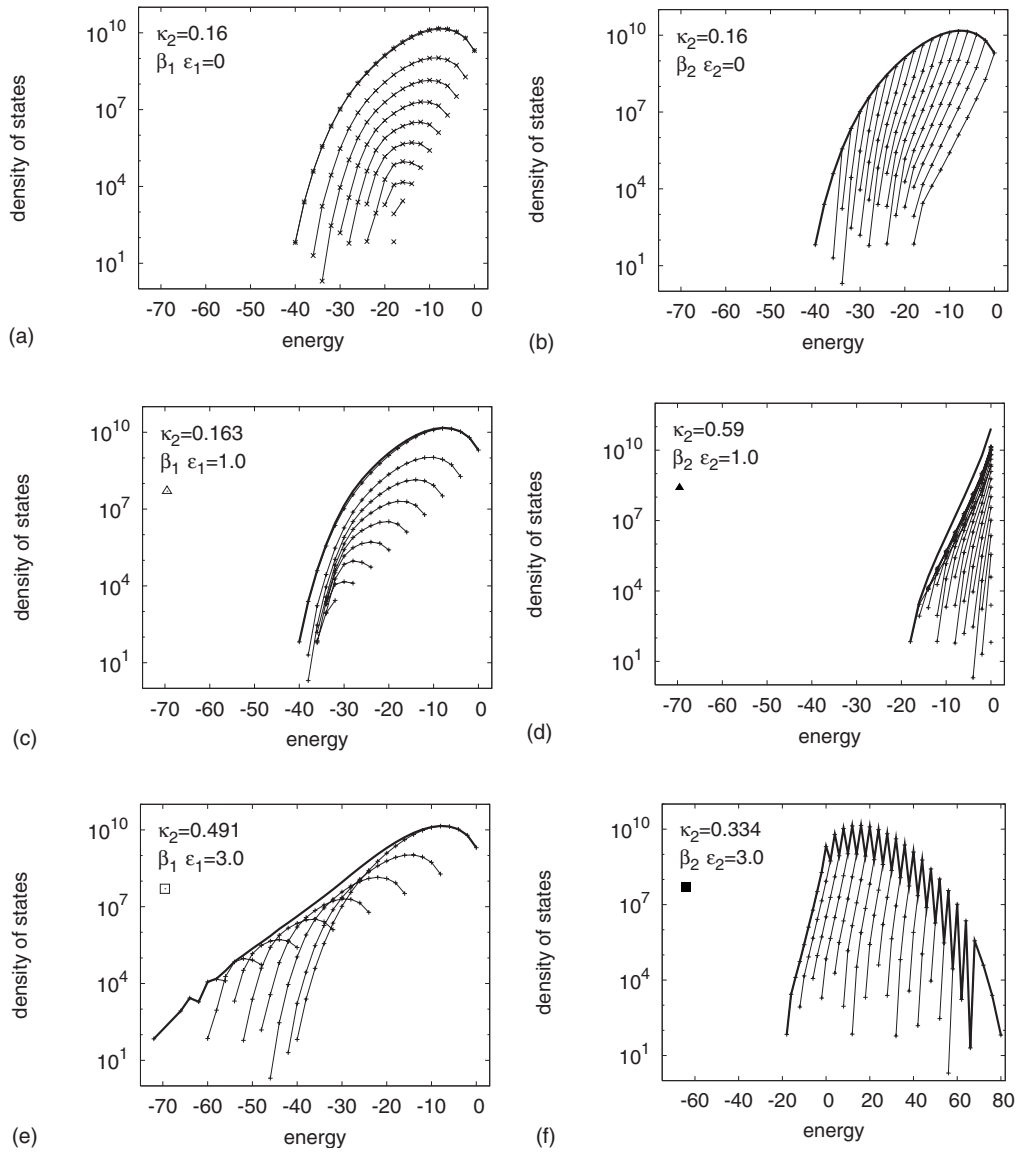

FIG. 7. Total number of conformations as a function of energy, that is, the density of states (thick line), in logarithmic scale, and relevant subdensities (crossed thin lines) for uncoupled (left-hand side) and coupled (right-hand side) β hydrogen bonds, with fixed lattice coordination reduction $z/z_h=7/3$ and chain length $N=16$. Subdensities in (a), (c), and (e) correspond to the number of conformations with a given number of hydrogen bonds, B, which are gradually left shifted as the hydrogen bond strength increases and their energies are *decreased* accordingly by $\epsilon_1 B$. Subdensities in (b), (d), and (f) correspond to the number of conformations with a given number of contacts not satisfying the hydrogen bond condition, C'', which are gradually right shifted as the hydrogen bond strength increases and their energies are *increased* accordingly by $\epsilon_1 C''$. Different panels in each side correspond to different values for the hydrogen bond strength: (a) $\epsilon_1=0.0$ and (b) $\epsilon_2=0.0$; (c) $\epsilon_1=1.0$ and (d) $\epsilon_2=1.0$; (e) $\epsilon_1=3.0$ and (f) $\epsilon_2=3.0$. Different symbols in (c)–(f) match corresponding points in Fig. 5(b), as in Fig. 6.

of hydrogen bonds, B, and of not-viable contacts, C'', shows how the convexity of the total density of states, in logarithmic scale, is differently affected when subdensities corresponding to given numbers of hydrogen bonds are differently shifted to lower energies as ϵ_1 increases in uncoupled hydrogen bonds, as opposed to the case when subdensities corresponding to given numbers of not-viable contacts are differently shifted to higher energies as ϵ_2 is likewise increased in coupled hydrogen bonds (Fig. 7). For uncoupled hydrogen bonds, in which case the energy of all conformations with B hydrogen bonds is *decreased* by $\epsilon_1 B$, the total density of states, shown as a thick line in each panel, becomes even more concave for α_1 (not shown) or is little affected for β_1 when the hydrogen bond strength increases from (a) $\epsilon_1=0$ to (c) $\epsilon_1=1$. For coupled hydrogen bonds, in which case the energy of all conformations with C'' not-viable contacts is *increased* by $\epsilon_2 C''$, the total density of states becomes more convex both for α_2 (not shown) and, particularly, β_2 when ϵ_2 is increased by the same amount (b) and (d). For (e) $\epsilon_1=3$ the density of states becomes more convex, due to the continuous shift of subdensities to lower energies, which is consistent with the increase in its κ_2 value. For (f) $\epsilon_2=3$, on the other hand, the continuous shift of subdensities to higher energies populates previously empty energy levels, increas-

ing the average energy of the system at infinite temperature and, therefore, increasing ΔE^* while decreasing κ_2 [see Eq. (5)].

Simulations with geometrically realistic, three-dimensional models should be useful to verify to what extent the present results are general. An investigation of the effect of coupled versus uncoupled hydrogen bonds on the density and subdensities of states might also turn out to be approachable by approximate analytical models, similar to the ones that have been used to describe protein folding and collapse as a diffusion process along one or a few reaction coordinates, e.g., [38,39]. The present enumeration results, however, already demonstrate that collapse cooperativity might be strongly dependent on the assumption of enthalpically favorable hydrogen bonds, independently of their local environment, as opposed to another, maybe more reasonable, scenario in which internal hydrogen bonds become favorable, or at least more favorable, when buried inside the protein globule. This observation corroborates our previous suggestion that the requirement of hydrogen bond formation upon chain collapse, independently of sequence, might contribute to protein folding two-state behavior when collapse and folding occur concomitantly [1] and is also consistent with a cooperative unspecific protein collapse when collapse and folding are not concomitant or when folding, but not collapse, is abolished by suitable mutations, as suggested by recent experimental results [14,15].

V. CONCLUSION

The present enumeration results demonstrate that, even for geometrically identical polypeptide models, collapse cooperativity might be drastically dependent on basic assumptions underlying the implementation of internal hydrogen bonds. It is apparent that previous simulations suggesting a one-state, continuous transition, might simply reflect the use of uncoupled, energetically favorable bonds. In more realistic implementations it is likely that hydrogen bonds should be coupled to hydrophobic interactions, in such a way that they would only be favorable, or at least more favorable, when buried. It is indicated that in this case collapse cooperativity could be significantly increased and become more consistent with recent experimental observations that suggest a two-state transition for the rapid contraction of some proteins and nonfolding peptides.

ACKNOWLEDGMENTS

This research was supported by CNPq through the Millennium Institute for Structural Biology in Biomedicine and Biotechnology (MISB3). One of the authors (A.F.P.A.) receives support from CNPq and one of the authors (M.A.A.B.) receives support from FAPESP.

[1] Marco Aurelio A. Barbosa, L. G. Garcia, and A. F. Pereira de Araújo, Phys. Rev. E **72**, 051903 (2005).
[2] Y. Zhou, C. K. Hall, and M. Karplus, Phys. Rev. Lett. **77**, 2822 (1996).
[3] E. I. Shakhnovich, G. Farztdinov, A. M. Gutin, and M. Karplus, Phys. Rev. Lett. **67**, 1665 (1991).
[4] L. G. Garcia, W. L. Treptow, and A. F. Pereira de Araújo, Phys. Rev. E **64**, 011912 (2001).
[5] J. Chahine, H. Nymeyer, V. B. P. Leite, N. D. Socci, and J. N. Onuchic, Phys. Rev. Lett. **88**, 168101 (2002).
[6] P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, New York, 1953).
[7] A. Y. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (AIP, New York, 1994).
[8] H. S. Chan and K. A. Dill, Annu. Rev. Biophys. Biophys. Chem. **20**, 447 (1991).
[9] J. P. K. Doye, R. P. Sear, and D. Frenkel, J. Chem. Phys. **108**, 2134 (1998).
[10] H. Zhou, J. Zhou, Z. Ou-Yang, and S. Kumar, Phys. Rev. Lett. **97**, 158302 (2006).
[11] J. Skolnick and A. Kolinski, J. Mol. Biol. **221**, 499 (1991).
[12] J. J. Chou and E. I. Shakhnovich, J. Phys. Chem. B **103**, 2535 (1999).
[13] M. Knott and H. S. Chan, Chem. Phys. **307**, 187 (2004).
[14] S. J. Hagen and W. A. Eaton, J. Mol. Biol. **301**, 1019 (2000).
[15] L. Qiu, C. Zachariah, and S. J. Hagen, Phys. Rev. Lett. **90**, 168103 (2003).
[16] G. D. Rose, P. J. Flemming, J. R. Banavar, and A. Maritan, Proc. Natl. Acad. Sci. U.S.A. **103**, 16623 (2006).
[17] A. Yang and B. Honig, J. Mol. Biol. **252**, 351 (1995).
[18] A. Yang and B. Honig, J. Mol. Biol. **252**, 366 (1995).
[19] H. J. Dyson, P. E. Wright, and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **103**, 13057 (2006).
[20] P. J. Fleming and G. D. Rose, Protein Sci. **14**, 1911 (2005).
[21] J. Borg, M. H. Jensen, K. Sneppen, and G. Tiana, Phys. Rev. Lett. **86**, 1031 (2001).
[22] J. R. Banavar, M. Cieplak, and A. Maritan, Phys. Rev. Lett. **93**, 238101 (2004).
[23] J. R. Banavar, T. X. Hoang, A. Maritan, F. Seno, and A. Trovato, Phys. Rev. E **70**, 041905 (2004).
[24] T. X. Hoang, A. Trovato, F. Seno, J. R. Banavar, and A. Maritan, Proc. Natl. Acad. Sci. U.S.A. **101**, 7960 (2004).
[25] A. Trovato, J. Ferkinghoff-Borg, and M. H. Jensen, Phys. Rev. E **67**, 021805 (2003).
[26] H. Kaya and H. S. Chan, Proteins **40**, 637 (2000).
[27] H. S. Chan, S. Shimizu, and H. Kaya, Methods Enzymol. **380**, 350 (2004).
[28] A. Kolinski, A. Godzik, and J. Skolnick, J. Chem. Phys. **98**, 7420 (1993).
[29] A. F. Pereira de Araújo, Proc. Natl. Acad. Sci. U.S.A. **96**, 12482 (1999).
[30] Marco Aurelio A. Barbosa and A. F. Pereira de Araújo, Phys. Rev. E **67**, 051919 (2003).
[31] A. F. Pereira de Araújo, A. L. C. Gomes, A. A. Bursztyn, and E. I. Shakhnovich, Proteins: Struct., Funct., Bioinf. (to be published).

[32] P. L. Privalov, Adv. Protein Chem. **33**, 167 (1979).
[33] A. F. Pereira de Araújo, Protein Pept. Lett. **12**, 223 (2005).
[34] D. K. Klimov and D. Thirumalai, Folding Des. **3**, 127 (1998).
[35] L. G. Garcia and A. F. Pereira de Araújo, Proteins: Struct., Funct., Bioinf. **62**, 46 (2006).
[36] M. Hao and H. A. Scheraga, J. Phys. Chem. **98**, 4940 (1994).
[37] M. Hao and H. A. Scheraga, J. Mol. Biol. **277**, 973 (1998).
[38] J. D. Bryngelson and P. G. Wolynes, Biopolymers **30**, 177 (1990).
[39] T.-L. Chiu and R. A. Goldstein, J. Chem. Phys. **107**, 4408 (1997).