Steric constraints as folding coadjuvant

M. E. P. Tarragó,[1,*] Luiz F. O. Rocha,[2] R. A. daSilva,[1] and A. Caliri[3,†]
[1]*Universidade de São Paulo, FFCLRP, Departamento de Física e Matemática, Avenida Bandeirantes, 3000, 14040.000 Ribeirão Preto, São Paulo, Brazil*
[2]*Universidade Estadual Paulista, IBILCE, Departamento de Física, Rua Cristovão Colombo 2265, Jardim Nazareth, 15054-000, São José do Rio Preto, São Paulo, Brazil*
[3]*Universidade de São Paulo, FFCLRP, Departamento de Física e Química, Avenida do Café S/N - Monte Alegre, 14040.903 Ribeirão Preto, São Paulo, Brazil*
(Received 30 September 2002; published 10 March 2003)

Through the analyses of the Miyazawa-Jernigan matrix it has been shown that the hydrophobic effect generates the dominant driving force for protein folding. By using both lattice and off-lattice models, it is shown that hydrophobic-type potentials are indeed efficient in inducing the chain through nativelike configurations, but they fail to provide sufficient stability so as to keep the chain in the native state. However, through comparative Monte Carlo simulations, it is shown that hydrophobic potentials and steric constraints are two basic ingredients for the folding process. Specifically, it is shown that suitable pairwise steric constraints introduce strong changes on the configurational activity, whose main consequence is a huge increase in the overall stability condition of the native state; detailed analysis of the effects of steric constraints on the heat capacity and configurational activity are provided. The present results support the view that the folding problem of globular proteins can be approached as a process in which the mechanism to reach the native conformation and the requirements for the globule stability are uncoupled.

DOI: 10.1103/PhysRevE.67.031901 PACS number(s): 87.15.Aa, 87.15.Cc, 82.35.Pq

I. INTRODUCTION

The understanding of the mechanism through which a natural globular protein gets its native conformation [1–3] is of central interest for practical purposes, but it is also an old intellectual challenge for researchers of diverse scientific branches [4]. Although being complex, the folding process has been successfully tackled by minimalist models that, in spite of their contrasting simplicity, are able to capture some fundamental features of the protein folding phenomenon. For instance, it has been theoretically confirmed that some compact configurations, resembling ''proteinlike'' secondary structures, are highly designable, that is, such structures are the native state of many distinct sequences of amino acids [5].

Simple models for macromolecules have a long and successful history [6] but the use of minimalist models in the protein folding problem was mainly motivated by the peculiar hole of the ''hydrophobic forces'' in the folding phenomenon. The application of the concept of hydrophobic bond [7], which is associated to the change in free energy on the transfer of nonpolar residues from the aqueous environment to the interior of proteins, popularized the representation of the hydrophobic energy as the sum on energetic terms associated to pairwise contacting residues of the molecule; particularly, this mapping is exact for some lattice models.

In practice, the folding problem has been simplified up to the point that the configurations of a chain with N monomers are mapped into the set of self-avoiding walks of fixed length $N-1$ in a regular lattice, and its configurational energy for each particular structure $\{\mathbf{r}_s\}$ being written as $E(\{\mathbf{r}_s\}) = \Sigma_{i,j}\varepsilon_{i,j}\Delta(d_{i,j})$, where $\varepsilon_{i,j}$ is the contact energy of the monomer pair (i,j), and $\Delta(d_{i,j})=1$ if the monomers i and j are first neighbors and zero otherwise. Clearly, such simplified models do not intend to describe a particular molecule; rather, they try to produce insights about the folding process itself.

In what refers to its significance to the folding process, hydrophobic or entropic forces have had an important theoretical support. Specifically, a quantitative analysis of structural data of natural proteins concluded that the hydrophobic effect ''generates the dominant driving force for protein folding'' [8]: it was shown that each element M_{ij} of the Miyazawa-Jernigan 20×20 matrix of interresidue contact energies [9] can be reproduced by the following simple equation:

$$M_{ij}=h_i+h_j-C(\delta_i-\delta_j)^2, \tag{1}$$

where h_i is the hydrophobic level of the residue i and δ_i is related to its solubility parameter. The quadratic term $-C(\delta_i-\delta_j)^2$ is directly related to the mixing energy of residues i and j [10], but the sum h_i+h_j is the dominant term [8].

Then, once a suitable hydrophobic scale $\{h_n\}$ has been established, it seems immediate that effective intrachain potentials should be approximated as $\varepsilon_{i,j}=h_i+h_j$. However, with this prescription for the contact energy, the segregation principle, $2\,\varepsilon_{i,j}-\varepsilon_{ii}-\varepsilon_{jj}\geqslant0$, is marginally satisfied through the equal sign, that is, $2\,\varepsilon_{i,j}-\varepsilon_{ii}-\varepsilon_{jj}\equiv0$. Indeed, as it will be clarified later through a Monte Carlo (MC) lattice model, the effective intermonomer potentials of the type $\varepsilon_{i,j}=h_i+h_j$ are efficient for packing and in inducing the chain to

*Permanent address: Pontifícia Universidade Católica do Rio Grande do Sul, Departamento de Física Teórica e Aplicada, FF.
†Corresponding author.

sparsely visit the native configuration, but they fail to provide enough stability to it. Therefore, in what manner extra interactional specificities between the pairs of monomers $\{i,j\}$ should provide such stability? This difficulty may be surmounted if one considers that other factors besides the configurational energy, as steric constraints, play a significant role in the folding process, affecting the folding routes and helping to stabilize the chain in the native state (note that the terms *native conformation* and *native state* are used in this text as synonymous). Then, in this paper, we focus on a simplified lattice model in which suitable hard-core-type constraints $\{c_{i,j}\}$ are added to the intermonomer potential $\{\varepsilon_{i,j}\}$, as introduced in a previous work [11], resulting in the following hydrophobic-type potential,

$$\varepsilon^{*}_{i,j}=h_i+h_j+c_{i,j}\,. \tag{2}$$

The first part of this work consists in analyzing—through the Monte Carlo method—the effects of steric constraints on thermodynamic quantities, such as the heat capacity and configurational activity of the chain. Distinct structures characterized by its contact order χ are used to illustrate that such effects are qualitatively independent of topological attributes of the native structure, although the model used here presents dynamical properties (as folding success and folding speed) that do depend on topological or geometrical parameters [12]. The contact order χ is defined with the chain in the native structure: it is the average sequence separation (in number of residues) between each pair of contacting residues, normalized by the number N of residues of the chain [13]. The values for χ used in this work were chosen to cover, illustratively, the range of all possible values, which for the present model satisfy $0.2381 \leqslant \chi \leqslant 0.4947$.

A model in the continuous space is also considered to attest, in more general circumstances, the aforesaid property of hydrophobic-type potentials, namely, the ability to access configurations presenting common features to the target (native) structure. To accomplish this purpose, the solvent-chain system is represented by a HP-type model in which the amino acid diversity was reduced to a *two*-letter alphabet, representing polar (P) and hydrophobic (H) monomers, at the same time in that all geometrical constraints were practically eliminated; the chain is explicitly exposed to the solvent. The HP pattern of the chains are obtained from real proteins of 35 residues.

II. HYDROPHOBIC MODEL AND STERIC CONSTRAINTS

In the present study, models using contact hydrophobic energy of the type displayed in Eq. (2) will be denominated as hydrophobic models [14]. The lattice model used in this work is composed by a single proteinlike chain constituted by $N=27$ monomers, which are effective residues taken from a repertory of stereochemically different elements; the residues occupy consecutive and distinct sites of a three-dimensional infinity cubic lattice; the interactions are assumed to occur between nearest-neighbor pairs of residues through a set of contact energy $\{\varepsilon_{i,j}=h_i+h_j\}$ and steric constraints $\{c_{i,j}\}$. Together, the set of hydrophobic levels $\{h_n\}$ and steric interactional specificities $\{c_{i,j}\}$ of the residues, constitute a 10-letter alphabet, as shown in Fig. 1. The strength of the interactions $\{h_m\}$ is expressed in units of k_BT (arbitrary energy units); k_B is the Boltzmann constant.

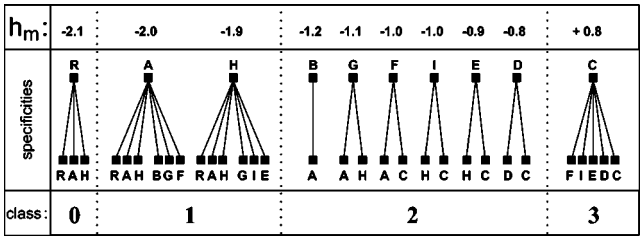

FIG. 1. Steric specificities and hydrophobic level for a 10-letter alphabet, namely, $\{R, A, H, B, G, F, I, E, D, C\}$. The lines connecting pairs of letters indicate the residues allowed to be first neighbors in the cubic lattice. For example, a residue type R can have as a first neighbor only residues of type R, A, and H. Note that the classes of monomers labeled **0** and **2** have higher steric specificities than those labeled **1** and **3**—as indicated at the bottom of the figure. The hydrophobic level for each "residue" is indicated at the top of the figure. When the chain is in the native configuration, these monomers making zero, one, two, and three contacts with the solvent are chosen, respectively, from the classes **0**, **1**, **2** or **3**. Note that there are three major ranges of hydrophobic levels, namely, more hydrophobic about $h=-2.0$; intermediary about $h=-1.0$ (those with more steric specificities); and on the other extreme with $h=+0.8$. The strength of the interactions $\{h_m\}$ are expressed in units of k_BT.

A few of the maximal compact self-avoiding (CSA) configurations (actually cubes $3\times3\times3$) are used as native or target structures; these structures are characterized by their corresponding relative contact order χ. The sequence of residues assigned to each structure is determined through a specific "syntax" that emerges from the constraints $\{c_{i,j}\}$ and from the application of the hydrophobic inside rule; see Ref. [11] for details.

A. Heat capacity and configurational activity for a lattice model

Let us first consider the analysis of the heat capacity and configurational activity for a particular structure, featured by its relative contact order $\chi=0.2381$, to discuss thoroughly the effect of the steric constraints in selecting folding pathways and on the overall globule stability; the monomer's sequence for this case is [*CBCIA ECBCE ADHRH DAECB CEAIC BC*]; see alphabet's details in Fig. 1. Simulations are also carried out for two more cases, characterized by distinct values for χ. All simulations in this study were performed in the time window t_w MC steps, which amounts to 8.1×10^8 configurations.

Without steric constraints, that is, using intermonomers contact potential such as $\varepsilon_{i,j}=h_i+h_j$, the peak of the heat capacity occurs about $k_BT=0.9$ and it is not prominent, as shown by open circles in Fig. 2. This behavior reveals that as the temperature modifies, the system exchanges relatively small amounts of energy with its surroundings. But, if appropriated steric constraints are introduced—now with $\varepsilon^{*}_{i,j}=h_i$

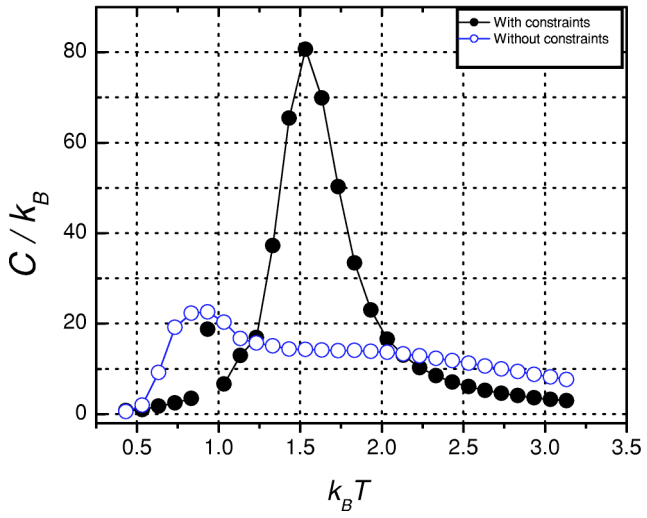

FIG. 2. Heat capacity as a function of k_BT (arbitrary energy units) for the system with constraints (steric specificities) and without constraints shown by solid and open circles, respectively. The remarkable change on the shape of the curve indicates that the chain's configurational activity is substantially distinct for each system. Note that at $k_BT_k=0.93$, for the system with constraints (solid circles) the amount C/k_B depends on the initial conditions (see text); for most of all other values of T the discrepancy between the results for independent simulations is smaller than 3%.

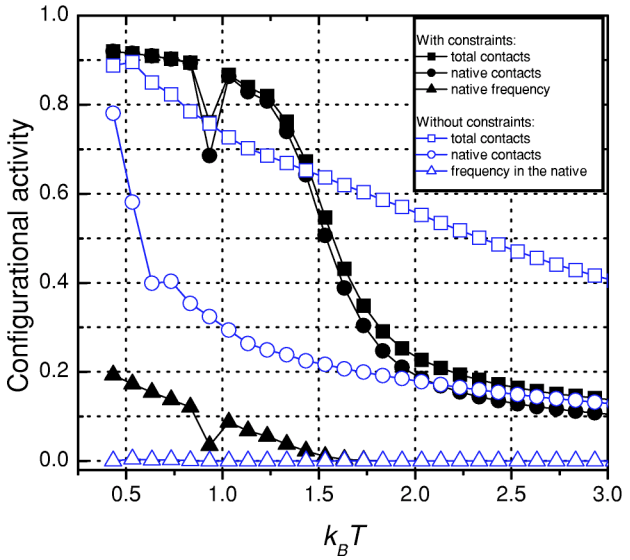

FIG. 3. The chain configurational activity as a function of k_BT. Solid and open marks refer to the system with and without constrains (steric specificities), respectively. The solid and open squares (■ and □) represent the normalized average number of total contacts for the unconstrained (Ψ_u) and constrained (Ψ_c) system, respectively; the solid and open circles (● and ○) represent the normalized average number of native contacts for the constrained ($\Psi_c^{(n)}$) and unconstrained ($\Psi_u^{(n)}$) system, respectively; and the solid and open triangles (▲ and △) are the relative frequencies in the native state for the constrained (Φ_c) and unconstrained (Φ_u) system, respectively.

$+h_j+c_{i,j}$—the curve appearance changes drastically (solid circles): first we note that two distinct temperatures stand out, namely, that corresponding to $k_BT_{\max}\simeq 1.5$ and $T_\kappa < T_{\max}$; the temperature $T_{\max}$ corresponds to the peak of the heat capacity, and at $k_BT_\kappa \simeq 1$ perturbations are observed (note in Fig. 2 the single solid circle out of the curve).

Such fluctuations at T_κ have a singular meaning because the simulations always started with the chain in the native structure (except some checking runs), what corresponds to unfoldinglike computational experiments. We initially describe the system's behavior for temperatures higher and lower than T_k: In the range $T_\kappa < T < T_{\max}$, the chain is always found in the native conformation; although many distinct configurations are visited during the entire simulation time, the native structure acts as an *effective attractor* in the sense that it is systematically visited with a frequency that increases as the temperature T decreases, as shown by Fig. 3. In this temperature range, the thermodynamic results are the same indifferently if the sampling started with the chain in the native configuration or randomly distended, since in the last case the chain reaches the native configuration very rapidly in comparison with the time window t_w. On the other hand, for T below T_k, as expected, the folding process becomes sluggish but if the sampling starts from the native configuration, or near enough to it, the chain stays always bonded to the native conformation. The fluctuations in the internal energy E is then reduced, as it is shown by the small magnitude of the heat capacity $C/k_B=[\langle E^2\rangle-\langle E\rangle^2]/k_B^2T^2$; Fig. 2. But, starting the simulation from a distended random configuration, the chain is easily captured by any of the energetic traps and so the collected configurational sample, during the time t_w, is not statistically meaningful.

However, about $T=T_\kappa$, the system shows a peculiar behavior, not seen for higher or lower temperatures: starting the sampling with the chain in the native structure, the results for all thermodynamic amounts are significantly dependent on the simulation history, what should not be expected for a system at equilibrium. As long as the unconstrained system does not show similar behavior, it implies that just at T_κ there is a synergic interplay between local energy minima and topological restrictions; at these conditions, the temperature seems sufficiently high to populate alternate configurations corresponding to local energy minima, but simultaneously it is low enough to trap the chain, resembling the cold denaturation.

A detailed analysis of the configurational evolution reveals that at $T=T_k$, the transition from the native into an excited (metastable) state occurs in the all-or-none manner, displaying the signature of a first-order phase transition; depending on the series of random numbers employed, this sudden transition happens at a different instant in the time window t_w. The chain becomes much less compact and about 50% of the native contacts (in average) is preserved; the native and excited states have the same pattern of fluctuations, but the latter occurs at significantly ($\simeq 12\%$) higher energy level and presents a much larger number of possible configurations. For temperatures slightly below T_k, the chain is definitively bonded to the native state; on the other hand, a little above T_k thermal fluctuations are high enough to populate these excited states and to overcome eventual energy

barriers, pushing it back and forth around the native conformation. Now, about T_k, a delicate tuning seems to be established: thermal fluctuations are marginally sufficient to populate excited states settled at local minima, but to go from the native to such states, and vice versa, it takes an amount of time of the same magnitude than t_w, because of the small number of paths connecting the native and such excited states (imposed by the steric specificities).

The remarkable transformation on the shape of the heat capacity curve has just one cause, viz., the changes on the chain's configurational space imposed by the steric constraints. Therefore, to follow the details of such alterations, in Fig. 3 some aspects of the configurational activity as a function of the temperature are shown. First, *average relative number of contacts*, namely Ψ, is defined as the average number of contacting first-neighbor monomers divided by 28, which is the total number of contacts for any CSA configuration. The way that Ψ behaves with the temperature when native and non-native contacts are indistinctly considered (that is, the total number of contacts) is represented by Ψ_u for the system without steric constrains (open squares in Fig. 3), and by Ψ_c for the system with steric constraints (solid squares), hereafter, denominated as *unconstrained* and *constrained* system, respectively. In the interval $0.5<k_BT<3.0$, Ψ_u decreases smoothly as the temperature increases, whereas Ψ_c presents an accomplished sigmoidal shape, quickly reducing the number of contacting monomers for values much smaller than the corresponding Ψ_u. This effect of the steric constraints indicates that portion of the conformational space corresponding to globularlike conformations was peculiarly affected: in the temperature region in which the globule is more compact (large Ψ_c), the number of configurations that separate the distended chain configurations from the most compact ones is severely reduced, which explains the sharp peak observed in the heat capacity curve. Such peak indicates that chain's internal energy, and entropy, exhibit a jump, rapidly changing its corresponding amounts for temperatures about $T=T_{\max}$.

Similarly, we turn our attention now to the behavior of the *average relative number of native contacts* for the constrained $\Psi_c^{(n)}$ and unconstrained system $\Psi_u^{(n)}$. For the latter, a relatively low value for the average native contacts is observed for most temperatures, as shown in Fig. 3 (open circles). But for temperatures low enough, when the globule is very compact, namely, for $\Psi_u>0.8$, the average number of native contact is significantly enlarged, with $\Psi_u^{(n)}$ quickly approaching Ψ_u but still $\Psi_u^{(n)}<\Psi_u$. A close look at the configurational evolution along the simulation showed that, even being very compact, that is $\Psi_u>0.8$, the globule shows significant malleability: the amount $\Psi_u^{(n)}$ oscillates intermittently between 15% and 80%, whereas the instantaneous Ψ_u changes continuously from 60% to 100%.

Now, for the constrained system, the number of native contact $\Psi_c^{(n)}$ (solid circles in Fig. 3) closely follows Ψ_c (solid squares). For $T<T_\kappa$, almost all contacts are native, that is, the condition $\Psi_c^{(n)}=\Psi_c$ is practically satisfied; but even for temperatures as high as $k_BT>2$ most of the contacts are native contacts, as displayed by Fig. 3. This result is to be understood as an effect of the steric specificities: for $k_BT>1.5$, the radius of gyration for the constrained system is significantly larger than that for the unconstrained one [15], as depicted in Fig. 4. So, in average, many chain contacts are local contacts, but such contacts are restricted by the steric constraints that favor the native ones because of the design of the sequence. This result also indicates that the steric constraints work as a folding guide, inducing the chain to native contacts, even at higher temperatures above $T_{\max}$.

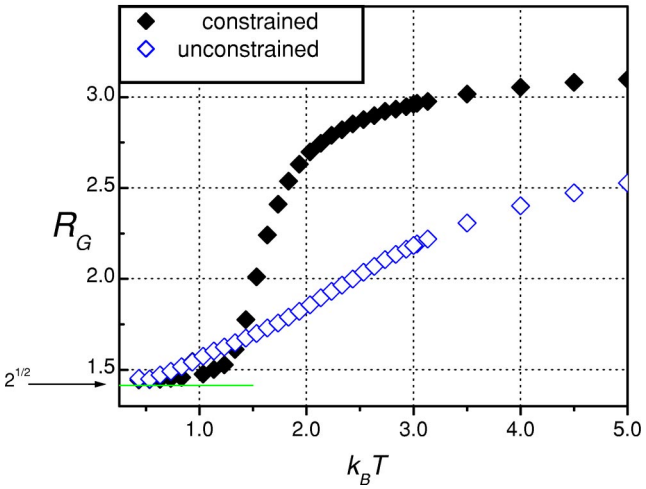

FIG. 4. The radius of gyration R_G as a function of k_BT. For the constrained system, at low temperatures corresponding to $k_BT \lesssim 1.3$, R_G is reduced with respect to the unconstrained system due to the synergism between energetic (local) minima and topological restrictions. But, above $k_BT=1.3$ the Boltzmann factor becomes systematically less influential and so the steric constraints considerably affect the original configurational space, swelling the globule. As the steric specificities do not allow many of the local contacts, this effect persists even for $T\to\infty$.

As a remarkable result, we point out that at the peak of the heat capacity, $k_BT_{\max}\simeq 1.5$, the average number of native contacts approaches 50%, that is $\Psi_c^{(n)}\simeq 1/2$; Fig. 3. Therefore, $T_{\max}$ can be seen as the temperature that separates two distinct behaviors of the configurational activity: below $T_{\max}$ the configurational activity—limited by steric constraints and relatively small thermal fluctuations—defines a compact globular shape for the chain ($\Psi_c>1/2$), and quickly becomes denser for smaller temperatures, while that for increasing temperatures above $T_{\max}$ the chain's globular shape is destroyed because now the distended configurations are statistically more significant.

Finally, we analyze the relative frequency Φ at which the chain is found in the native state. It is the ratio $\Phi=\phi^{(n)}/\phi$ between the number $\phi^{(n)}$ of times the chain was found in the native structure, and the total number ϕ of configurations. For the unconstrained system, the native configuration can eventually be visited but it is unprovided with enough stability, that is, $\Phi_u<10^{-5}$ for all temperatures $T>T_k$; open triangles in Fig. 3. However, the fact that Φ_u is not exactly zero has an important meaning; it suggests that the hydrophobic-type potentials, such as $\varepsilon_{i,j}=h_i+h_j$, are efficient in compacting the chain and reach the native state, although they

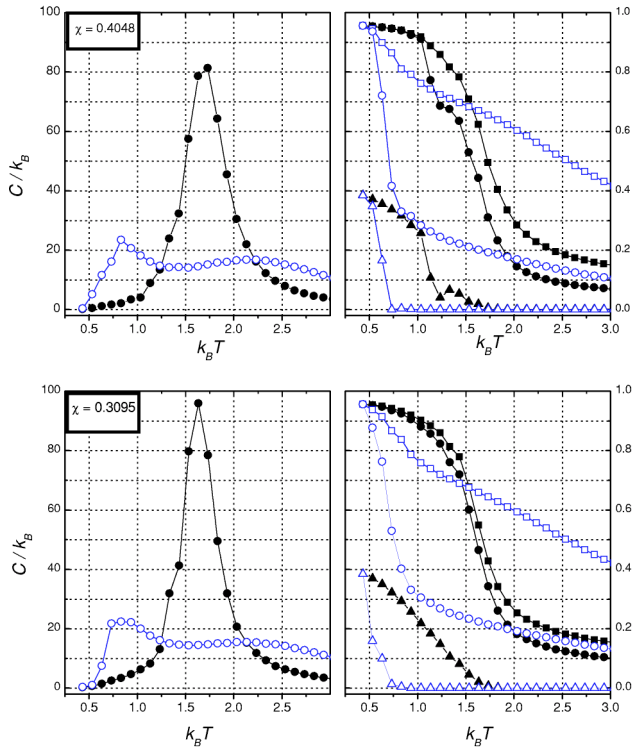

FIG. 5. Heat capacity and configurational activity for two other cases with larger χ. Note that although T_κ and $T_{\max}$ are slightly displaced to higher temperatures, the effect of steric constraints are qualitatively the same as shown in Figs. (2) and (3); the marks and symbols are the same as those used previously.

fail to sustain it properly in that state. However, if appropriated steric interactional specificities are introduced, they work as a type of topological labyrinth for the native configuration. This configurational barrier increases its efficiency as T decreases from $T_{\max}$, and so the relative frequency Φ_c in the native state (solid triangles) assumes significant values, reaching 10% about $k_BT=1$; numerically Φ_c is at least five orders of magnitude larger than Φ_u. Note that just at T_κ, the value Φ_c is smaller than the curve tendency should suggest, which agrees with the comments above about the heat capacity.

As a complement, Fig. 5 shows the heat capacity and the configurational activity for two more structures characterized by different contact orders, namely, $\chi=0.3095$ and $\chi=0.4048$; the corresponding sequence of monomers are [*CBCIA ECBCI AICBC IAICB CIADH RH*] and [*ADHEC GCIAI CBCIA ICBCG CIHRH FC*]; see alphabet's details in Fig. 1.

The results are all qualitatively equivalent to those described above, although for different χ the temperature $T_{\max}$ of the peak of the heat capacity, as well as its values at $T_{\max}$ may change. As a rule, $T_{\max}$ increases slightly with χ, but other topological ingredients also may be influential, as the number of structural patterns resembling secondary structures. Yet with respect to the constrained system discussed here, the time to reach the native state for the first time is smaller about $T_{\max}$, quickly becoming larger as the temperature deviates significantly— above or below —from it.

B. Hydrophobic model for an off-lattice model

As commented above, the fact that Φ_u is not exactly zero suggests that hydrophobic potentials are able to induce the chain to collapse into the nativelike conformation, although not providing the necessary stability to keep the chain bonded to the native state. However, one may yet ask if the native configuration was not incidentally found, considered that the results refer to simulations that started at the native configuration, and there is a relative small number of such compact configurations. Indeed, there are about 10^5 CSA configurations and for appropriated temperatures the native state is found at most once every 10^5 attempted moves, as it was cited above. To clear up this question, we have carried extensively runs of the folding process for the unconstrained system, always starting from a distended configuration. For temperatures, not too far from the peak of the heat capacity, the results have shown that the native state is always reached, but one may yet complain about eventual geometrical bias introduced by the regular lattice. Therefore, to reexamine the mentioned efficiency of the hydrophobic model for an independent system, we introduced an off-lattice model, described as follows.

In this model, the stereochemical diversity of the 20 natural amino acids is reduced to a two-letter alphabet, representing polar (P) and hydrophobic (H) monomers, while that all geometrical constraints were practically eliminated. Technically, the chain-solvent system is represented as a pearl necklace in solution (the solvent is treated explicitly), in which each monomer is represented by a hard sphere of diameter D connected to its neighbors by ideal flexible strings with defined length $D+\varepsilon$, where $\varepsilon\simeq 0.2D$. The 12.565 solvent molecules are also represented by hard spheres of the same diameter D. The magnitude of the specific hydrophobic levels $\{h_n\}$ are equivalent to the one used in the lattice model, that is, each monomer of the chain can have one of two possible values: $h_P=+1$ or $h_H=-2$. The solvent-solvent interaction $e_{s,s'}$, as well the monomer-monomer interaction $e_{m,m'}$ is a hard-core-type potential, while the solvent-monomer interaction, besides the hard-core potential, involves the hydrophobic energy $e_{s,m}=e_0-n_sh_m$, where e_0 is an arbitrary constant, h_m is the hydrophobic level of monomer m (h_P or h_H), and n_s is the number of solvent molecules surrounding it. Note that the energy $e_{s,m}$ increases with n_s if the monomer m is hydrophobic ($h_m=h_H$), and decreases otherwise. Each one of the monomers in the HP sequence corresponds, one by one, to the polar (or nonpolar) attribute of the 35 amino acids of a real protein, and its corresponding three-dimensional (3D) structure was specially chosen as a target configuration. The polar (or nonpolar) attribute of each residue was chosen based on the scale proposed in Ref. [16]

First, the chain packing process was analyzed as a function of the temperature in its global aspect. For this purpose, Fig. 6 shows the behavior of the standard deviation SD_G of the average radius of gyration R_G against k_BT for the last 10^5 MC steps, which corresponds to one fifth of the total time window t_w, namely, $t_w=5\times10^5$ MC steps, which corre-

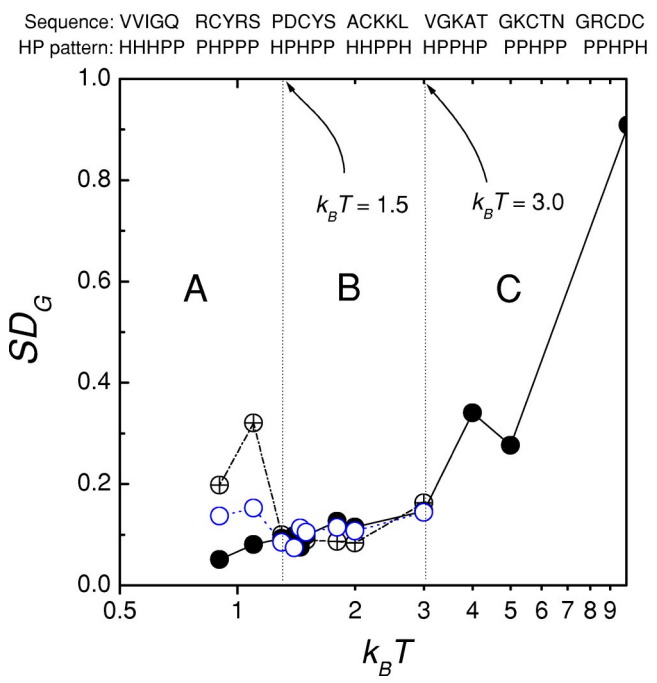

FIG. 6. The behavior of the standard deviation SD_G of the radius of gyration R_G against k_BT. The protein 1tsk sequence (one-letter symbol) and its HP pattern are shown at the top. The physical system as a whole, is represented by a single linear chain of 35 units, surrounded by 12 565 solvent molecules confined in a cubic box. The beads of the chain, as well as the solvent molecules are hard spheres of the same diameter, but the monomer-solvent interaction depends additionally on the hydrophobic attribute of each interacting pair.

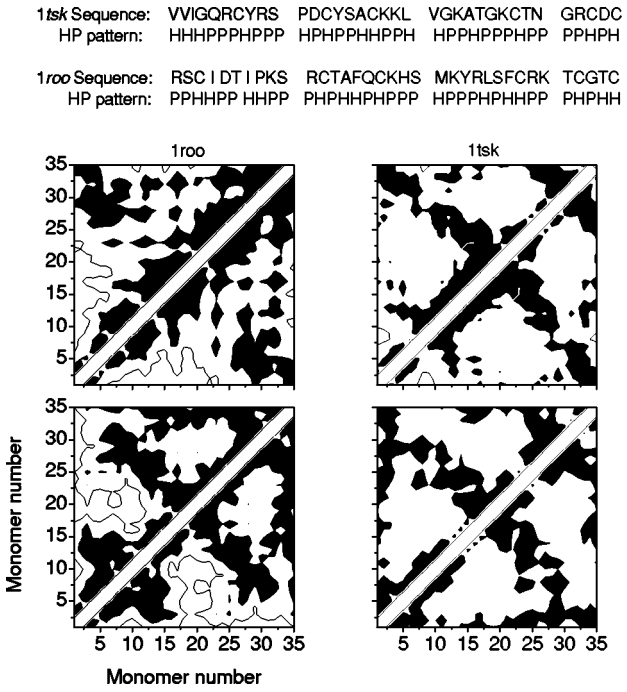

FIG. 7. Contact maps for two globular proteins, namely, 1tsk and 1roo. The amino acid sequence for both cases (one-letter symbol), and the corresponding HP patterns are shown at the top. The spatial scales were properly translated and rescaled in such that distances $d_{i,j}$ fit the range $0 \leqslant d_{ij} \leqslant 1$ (real protein and models). Black and white regions mean distances smaller or equal to 0.3 and larger than 0.3, respectively. Real proteins and models are to be compared by columns. See text for details.

sponds to a total of about 6×10^9 generated configurations. Three distinct regions are identified: For $k_BT < 1.5$ (region *A*), the amount SD_G depends strongly on the initial conditions (results are shown for three independent runs). For $1.5 \leqslant k_BT \leqslant 3.0$ (region *B*), the globule is well defined; the smaller value for SD_G occurs at $k_BT = 1.5$ and then increases slowly up to $k_BT = 3.0$. Finally, for $k_BT > 3.0$ (region *C*), SD_G changes rapidly with the temperature until saturating at $k_BT \gtrsim 5.0$.

For corresponding temperatures starting at $k_BT = 1.5$, the size of the globule can be thermodynamically defined independently of the initial condition: thermal fluctuations are already significantly large to disrupt the nonoptimized hydrophobic contacts and so, independently of the initial conditions, the chain always collapses into a compact globulelike conformation. In the interval $1.5 \leqslant k_BT \leqslant 3.0$, the value of R_G practically does not change with respect to its value at $k_BT = 1.5$: it is about 3% and 5% larger at $k_BT = 2.0$ and $k_BT = 3.0$, respectively. Accordingly, fluctuations of the globule's size begin to increase smoothly and slowly with the temperature until about $k_BT = 3.0$; above $k_BT = 3.0$ huge fluctuations take place. Therefore, all runs with respect to the results reported below were performed at $k_BT = 1.5$.

Figure 7 shows the contact map for two real proteins and for their corresponding models, according to what was discussed above. To minimize the distinct spatial nature between the models and the proteins, two precautions were taken: (i) the intermonomers distances $d_{i,j}$ (center of mass), for protein and model, were properly translated and rescaled to fit the same interval from zero to one, that is, $0 \leqslant d_{i,j} \leqslant 1$; and (ii) black regions in the maps correspond to all distances $d_{i,j}$ satisfying $0 \leqslant d_{i,j} < 0.3$, that is, distances up to 30% of the largest distance (for each case: model and protein), and as white regions for distances $0.3 \leqslant d_{i,j} \leqslant 1$. The contact map for each model corresponds to a particular configuration chosen among ten configurations taken in equal intervals from the last quarter of the simulation time.

Here, of course, the maps must not be conclusively compared with respect to configurational similitudes between proteins and their corresponding models and nor should be expected to be like that, given the severe topological simplifications imposed into the model. However, it is still possible to see specific propensities (even though somewhat distorted) in each model's map, resembling the corresponding real protein's map. To help in the protein-model comparison, it is interesting first to recognize that due to the exclusive virtue of its distinct HP patterns, the two maps corresponding to the models are distinct. And then one looks for the similitudes between the protein's map and its corresponding model's map. As it occurs in the lattice model, here several contacting residues in the real protein (black regions of Fig. 7) have corresponding contacting pairs in the model system; many of them presenting relatively high frequency of contact along the simulation. Although the globule is very compact, it pre-

serves great malleability, as also observed in the lattice model, indicating that the chain is not pinned in any particular configuration. The resemblances between the two maps, protein and model, are recurrent, appearing and disappearing from time to time along the simulation. The chain-model configurations shown in the Fig. 7 were selected for visual purpose only to illustrate our arguments; it has a short lifetime but even with alternations of configurational similarities between the model's results and the native configuration of the corresponding real protein, many pairs of monomers last along the whole simulation time. After the collapse and with the chain's segments already stuck together, it seems ''to breath,'' that is, an incessant succession of swelling and shrinkage of the globule takes place; the chain rambles through the compact configurational subspace, eventually visiting configurations that present more resemblances with the protein's native configuration of its respective real protein.

Therefore, we conclude that the main virtues of hydrophobic-type potentials are (i) efficiency to compact the chain maintaining the globule malleable and, once packed; and (ii) capability to induce the chain through conformations near to the native state, without, however, providing configurational definition to the globule.

III. COMMENTS AND CONCLUSION

In the present work, hydrophobic-type potentials and steric constraints are employed in a simplified model as the two basic ingredients for the folding process. It is shown for a lattice model that contact energy based in such potentials are efficient in packing the chain and in finding the native state, but they fail to provide configurational stability to it. An appropriate set of steric constraints are then added and it is shown that as folding coadjuvant, such steric interactional specificities help to select folding pathways and improve the overall stability condition of the globule in the native configuration. Specifically, through comparisons between two sets of Monte Carlo simulation results, it is shown that suitable steric specificities dramatically change the system's configurational activity. This effect has the following consequences: (i) it transforms the original broad curve of the heat capacity, obtained using a hydrophobic-type potential as pair contact energy, into a peaked and symmetric curve; and (ii) it significantly increases the frequency in which the chain stays in the native state in five or more orders of magnitude. A second (off-lattice) model confirms the effectiveness of the hydrophobic-type potentials in producing a malleable globule and driving the chain through configurations that intermittently approach the native conformation.

Such results suggest that the folding problem of globular protein can be approached as a process in which the mechanism to reach the native conformation and the requirements for the globule stability are uncoupled. In this view, the stereochemical code, namely, the hydrophobic pattern and the steric interactional specificity of each residue, is responsible for the mechanism through which the chain reaches the native state, which must then be considered as a special and unique state. Once in the native state and only in this state, all the energetic and steric factors involved in the process compose themselves in such way to maximize the stability conditions for the globule: most of the hydrogen bonds are now protected from the medium and, as the competition with the solvent is minimized, they effectively contribute to the globule stability; the overall steric complementariness of the residues increases the internal contact area, at the same time it reduces the external contact with the solvent, also producing a net contribution for the energetic stability of the globule, and finally, the steric specificities of the residues work also as hindrances, topologically trapping the chain, as in a 3D puzzle, efficiently helping to maintain the chain in the native conformation.

[1] A. Sali, E. Shakhnovich, and M. Karplus, Nature (London) **369**, 248 (1994).
[2] D. Baker, Nature (London) **405**, 39 (2000).
[3] J. Chahine, H. Nymeyer, V.B.P. Leite, N.D. Socci, and J.N. Onuchic, Phys. Rev. Lett. **88**, 168 101 (2002).
[4] *Protein Folding*, edited by T.E. Creighton (Freeman, New York, 1992).
[5] Hao Li, R. Helling, C. Tang, and N. Wingreen, Science **273**, 666 (1996).
[6] P.-G De Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).
[7] W. Kauzmann, Adv. Protein Chem. **14**, 1 (1959).
[8] Hao Li, C. Tang, and N.S. Wingreen, Phys. Rev. Lett. **79**, 765 (1997).
[9] S. Miyazawa and R.L. Jernigan, J. Mol. Biol. **256**, 632 (1996).
[10] J.H. Hildebrand and R.L. Scott, *The Solubility of Nonelectrolytes* (Reinhold, New York, 1950).
[11] R.A. daSilva, M.A.A. da Silva, and A. Caliri, J. Chem. Phys. **114**, 4235 (2001).
[12] In Ref. [11], it is shown for a minimalist hydrophobic model that the folding success is correlated with topological characteristics of the native state, as contact order and the presence of some structural patterns that resembles secondary structures.
[13] K.W. Plaxco, K.T. Simons, and D. Baker, J. Mol. Biol. **277**, 985 (1998).
[14] A.F. P de Araujo, Proc. Natl. Acad. Sci. U.S.A. **96**, 12 482 (1999).
[15] A. Caliri and M.A.A. daSilva, J. Chem. Phys. **106**, 7856 (1997).
[16] Zi-Hao Wang and H.C. Lee, Phys. Rev. Lett. **84**, 574 (2000).