Effect of local thermal fluctuations on folding kinetics: A study from the perspective of nonextensive statistical mechanics

J. P. Dal Molin,* M. A. A. da Silva, and A. Caliri
Departamento de Física e Química, FCFRP, Universidade de São Paulo, 14040-903 Ribeirão Preto, SP, Brazil
(Received 8 December 2010; revised manuscript received 4 August 2011; published 3 October 2011)

The search through the proteins conformational space is thought as an early independent stage of the folding process, governed mainly by the hydrophobic effect. Because of the nanoscopic size of proteins, we assume that the effects of local thermal fluctuations work like folding assistants, managed by the nonextensive parameter q. Using a 27-mer heteropolymer on a cubic lattice, we obtained—by Monte Carlo simulations—kinetic and thermodynamic amounts (such as the characteristic folding time and the native stability) as a function of temperature T and q for a few distinct native targets. We found that for each native structure, at a specific system temperature T, there exists an optimum q^* that minimizes the folding characteristic time $\tau_{\min}$; for $T = 1$, it is found that q^* lies in the interval 1.15 ± 0.05, even for native structures presenting significantly different topological complexities. The distribution of $\tau_{\min}$ obtained for specific $q > 1$ (nonextensive approach) and temperature T can be fully reproduced for $q = 1$ (Boltzmann approach), but only at higher temperatures $T' > T$. However, assuming that the complete set of proteins of each organism is optimized to work in a narrow range of temperature, we conclude that—for the present problem—the two approaches, namely, $(T, q > 1)$ and $(T > T', q = 1)$, cannot be equivalent; it is not a simple matter of reparametrization. Finally, by associating the nonextensive parameter q with the instantaneous degree of compactness of the globule, q becomes a dynamic variable, self-adjusted along the simulation. The results obtained through the q-variable approach are utterly consistent with those obtained by using a target-tuned parameter q^*. However, in the former approach, q is automatically adjusted by the chain conformational evolution, eliminating the need to seek for a specific optimized value of q for each case. Besides, using the q-variable approach, different target structures are promptly characterized by inherent distributions of q, which reflect the overall complexity of their corresponding native topologies and energy landscapes.

DOI: 10.1103/PhysRevE.84.041903 PACS number(s): 87.15.Cc, 87.15.hm, 87.15.ak, 87.10.Hk

I. INTRODUCTION

The series of events that drive a polypeptide chain into its stable native structure is not yet fully understood. Protein systems involve so numerous complex interactions and remarkable properties that they continuously require new experiments, as well as theoretical and computational approaches [1–5].

Along the folding process of globular proteins, the corresponding native states are found so quickly that, in spite of being a stochastic process, the search through the protein's huge conformational space cannot be processed at random [6]. However, though being so rapid, folding rates (k_f) of different proteins of about the same size span over many orders of magnitude; k_f is a measure of how fast the folding process leads the chain from the unfolded state up to the native structure. Therefore, even for two-state, single-domain proteins, the current theories of folding kinetics remain too limited to accurately explain this experimental observation [7]. Indeed, characterization of the refolding of such simple proteins has shown that factors such as stability and chain length, as proposed by theoretical studies, are apparently not as influential on k_f as the global average sequence separation between contacting residues in the native state, that is, the native structure Contact Order [8].

One reason for not fully understanding the folding process—besides the conflicting interpretations of its fundamentals—may be because the mechanisms for protein folding have been routinely proposed from ensemble-averaged properties [9–11]. Hence, to date, virtually all approaches proposed to investigate the folding process can only partially explain the phenomenon [12]. For instance, the transition state concept provides a satisfactory explanation for two-state kinetics but not for folding reaction rates, while the funnel landscape idea can give insights about the folding rates but not for two-state kinetics [13].

Our central aim in the present work is to show evidence concerning the importance of local thermal fluctuations for the folding kinetics of globular proteins. A typical behavior of small, single-domain, globular proteins is used here as an ideal prototype: Many of such proteins fold via an all-or-nothing process, that is, without detectable intermediates [14]. In the next section, we present a few characteristics of the search step of the folding problem, which are common properties to all proteins. The simplified model employed here, presented in Sec. III, focuses exclusively on the search mechanism and is grounded on the hydrophobic effect. The effects of local thermal fluctuations on a nanoscale structure are treated in the context of nonextensive statistical mechanics; see Sec. IV. Resulting data are analyzed in detail through the dependence of the folding characteristic time τ on the temperature T and the nonextensive parameter q; see Sec. V. Comments and conclusions in Sec. VI are devised according to the premises stated at the end of the next section.

II. GENERAL CHARACTERISTICS OF THE PROTEIN-FOLDING PROCESS

In this section we review a few attributes of the search stage of the folding problem, common to all proteins. Such

*jpdm@fcfrp.usp.br

general properties are used to formulate three premises, which compose general grounds for the folding problem and are incorporated in the model employed in the present work.

A remarkable landmark of the folding process is its robustness. Folding must be similarly processed in a wide temperature range, covering about 100 °C, at the same time that all functional proteins of each individual organism—living in the most different environment—are properly stable near a particular ideal temperature. Actually, living organisms are found in extreme conditions; some of them live in places with temperatures close to freezing water [15], while others are found in environments with boiling water temperatures [16]. Therefore, the search mechanism must work properly over the full range of liquid water temperature, while for each living species, the functional temperature is, in general, restricted to a relatively much smaller range than that. This property of the folding phenomenon is better understood if one considers that the search mechanism is mainly governed by the hydrophobic effect. It has been shown experimentally that the transfer free energy varies slightly in the temperature interval from about zero to 100 °C, at least for small hydrophobic molecules [17,18]. Considering that during the search for the native structure any other conformation of the chain is (logically) unstable, the folding process should be composed by two temporal independent steps [19]: the search mechanism, as the first stage, followed by overall stabilization that begins only when the chain is close enough to its native structure, at the instant when energy and structural requirements, as encoded in the residue sequence, would be associated in a cooperative and productive way.

As a second general property, we recall that proteins present an extraordinarily precise and fast self-organization process [20]. They fold some 10 orders of magnitude faster than the predicted rate of a random search mechanism; thus, although folding rates do appear to have been fully optimized by natural selection, one can think about each protein as an amino acid sequence that was designed to fold as fast as possible. Actually, even though it is unclear whether the high folding rates result from evolutionary optimization, or a consequence of selection for stability, which is certainly a product of natural selection [21], the fact is that protein folding is a general, very fast process. Indeed, it has been shown that the probability of finding a fast-folding sequence, choosing it randomly from the set of all possible sequences, is very small [22]. Furthermore, there is also a physiological aspect related to fast folding: Because there are not enough chaperones to support the folding process of every protein (anyway, chaperones are also proteins), they must fold very rapidly in order to avoid aggregation due to exposing hydrophobic areas of their surface for too long [23,24]. Therefore, it is reasonable to presume that some common peculiarities of the native structures and efficient instructions encoded along the chain will ultimately provide an optimized and very fast folding process.

Finally, we note that globular proteins can be approached as independent nanomachines. This view contrasts with the nature of most currently available experimental data and explains certain inadequate interpretations of the folding problem. The stable appearance and properties of homogeneous macroscopic objects result from the average activity of a very large number of atoms. But, opposing this scenario, nanostructures such as colloidal particles or proteins, in contact with a thermal reservoir (the solvent, in the present case), perceive local thermal fluctuations in a special way [25,26]: Local unbalanced forces continuously shake and deform such nanostrucutures, and its contingent effects cannot be inferred from most available protein kinetics data [6]. So, one could ask how, in different aspects, do local thermal fluctuations affect the folding process [27]? Certainly, new data and ideas emerging from single-molecule experiments will help to illuminate this subject [1]; for instance, data about transition paths at equilibrium, which is observable only for single molecules, permit to obtain crucial mechanistic information, such as folding and unfolding rates [28].

Based on such general characteristics, some hypotheses can be formulated as general grounds for the folding problem, which are fundamental constituents of the model described along the next two sections. Therefore, we assume that (1) the complete folding process is composed by two time- and mechanistically independent steps, namely, the search mechanism, governed mainly by the hydrophobic effect, and the overall productive stabilization; (2) for typical one-domain, two-state globular proteins, instructions encoded in the residues sequence provide a kinetic folding mechanism as fast as possible; and (3) at nanoscale dimensions local thermal fluctuations emerge as a peculiar attribute of the protein molecule enabling an effective folding process.

III. THE LATTICE MODEL

The model presented here is devoted to the first stage of the folding process, the search step. For the sake of computational efficiency, a lattice model is used: A chain of 27 beads occupying consecutive and distinct sites of a three-dimensional infinity cubic lattice, represents a single protein-like chain in solution; effective solvent molecules, which explicitly interact with the chain, fill up the lattice vacant sites. Along the simulation, solvent molecules and chain units exchange their respective sites so that all sites of the lattice always remain fully filled [29]. For each configurational change, only the transfer free energy is taken into account (variations on the system hydrophobic energy); solvent-solvent and residue-residue interactions are represented by hard core-type interactions (excluded volume). For lattice models and uniform-density solvent, this interaction scheme is exactly reproduced using inter-residue potentials with first-neighbor pairwise interaction, namely, $g_{i,j} = h_i + h_j$, where h_i is the hydrophobic level of the ith residue along the chain sequence [30]. Residues are taken from a repertory of ten distinct elements (a 10-letter alphabet), which are characterized by distinct hydrophobic levels $\{h_k\}$ and a set $\{c_{i,j}\}$ of inter-residue steric specificities. The hydrophobic effect, characterized by set $\{h_k\}$, represents the most general and influential chemical factor acting along the folding process [31], while set $\{c_{i,j}\}$ mimics steric specificities of the real residues. Such constraints determine which pairs of residues are allowed to get closer as first neighbors; its main consequence is to select folding and unfolding pathways through the configurational space. Set $\{c_{i,j}\}$ is fixed for each monomer pair; that is, it does not depend on particularities of the native structure [3,19].

The configurational energy $E([\kappa,l])$ of an arbitrary chain configuration ξ is defined by the set $[\kappa,l]$ of N_ξ first-neighbor inter-residue contacts ($0 \leqslant N_\xi \leqslant 28$); that is,

$$E([\kappa,l]) = \sum_{\{i,j\}} (g_{i,j} + c_{i,j})\delta_{(i,j),[\kappa,l]}, \tag{1}$$

where the sum runs over the set of all residues pairs $\{i,j\}$; the factor $\delta_{(i,j),[\kappa,l]} = 1$ if (i,j) belongs to the set $[\kappa,l]$; otherwise $\delta_{(i,j),[\kappa,l]} = 0$.

The present model is not native centric; that is, data from the native structure are not employed to guide the chain along the simulation. Therefore, a rule for residue sequence designing, valid for any target structure representing the native structure, is necessary. The provided syntax is mainly based on the *hydrophobic inside* rule [32] and on the local topological features of the target structure [33].

The native state of models based on the stereochemical potential does not present enough stability because such additive potential, $g_{i,j} = h_i + h_j$, marginally satisfies the segregation principle; that is, $g_{i,i} + g_{j,j} - 2g_{i,j} = 0$. However, adding up steric constraints $\{c_{i,j}\}$ to the hydrophobic potential $g_{i,j}$ [Eq. (1)] helps to select folding and unfolding pathways, which makes the folding process faster and improves the overall stability condition of the globule in the native state [19].

The folding process was simulated through the Metropolis Monte Carlo (MC) method, in which a generalized Boltzmann weight was employed, as described in the next section. The configuration evolution is processed by standard elementary chain moves, namely, crankshaft, corner, and end flips. The process evolves without any reference to the native configuration, except to check when it is found for the first time: For each particular run, the number of MC steps spent to reach the native structure from the initial configuration (the first passage time) is taken as the folding time t for that case.

IV. LOCAL THERMAL FLUCTUATIONS AND THE NONEXTENSIVE STATISTICAL MECHANICS

In this section, in order to motivate the use of nonextensive statistical mechanics, we analyze the energetic effect due to thermal fluctuations on nanometric particles, as a typical protein molecule. Usual thermodynamic and kinetic data about proteins are time-averaged results from the collective behavior of many molecules, something between $\sim 10^{17}$ and 10^{20} molecules/liter. However, when considered individually, thermodynamic properties of proteins surely undergo strong fluctuations. For instance, let us reproduce here a specific thermodynamic calculation [34] for one representative protein of, say, 250 residues, with molecular mass $m = 4 \times 10^{-23}$ kg, which typically shows specific heat capacity ($C_p \cong C_v$) about 0.3 kcal kg^{-1}K^{-1}, at temperature $T = 300$ K. The internal energy fluctuation about the mean for one molecule gives $\Delta U_{\rm rms} \cong 6 \times 10^{-20}$ cal per molecule ($\Delta U_{\rm rms} = (kT^2 mC_v)^{1/2}$); note that $kT \simeq 10^{-21}$ cal, for room temperature. Essentiality, it would be in the same order of magnitude of the typical enthalpy changes on thermal denaturation of proteins—tens of kcal/mol [35]—if all molecules were fluctuating in concert [34]. Fluctuations are individually uncorrelated, and, so, only the average effect of a huge number of macromolecules is perceived. However, if only a single protein is focused, then one can see how local thermal fluctuations can affect—through the above estimated internal energy fluctuations—folding kinetics. This scenario suggests that the folding process has to be explained for a single molecule.

Therefore, let us think about a protein in solution (relaxing toward thermal equilibrium) as a heterogeneous system constituted by particles of different sizes, namely, a large body, the chain, immersed in the solvent, which is constituted by smaller particles (water molecules). In this scenario, the solvent can be seen as a heat reservoir at thermodynamic temperature, let us say, β_0^{-1}. Thermal fluctuations drive stochastically individual protein chains to change their configurations, similar to what occurs with colloidal particles suspended in a fluid (Brownian movement). Accordingly, due to its relatively larger size with respect to water molecules, it is reasonable to assume that the chain configurations evolve at a certain rate that depends on the instant local amount of (available) kinetic energy, including the water molecules impacting the chain. So let us represent the instant local-average of this kinetic energy by β^{-1}, a kind of instantaneous local temperature, which cannot be confused with the thermodynamic temperature β_0^{-1} of the whole system (chain plus solvent). Therefore, considering such heterogeneous system (with a finite component: the chain), one could think about a generalized Boltzmann factor $\exp_q(-\beta_0\epsilon)$ as an integral over all possible locally instantaneous β^{-1}, that is,

$$\exp_q(-\beta_0\epsilon) = \int_0^\infty \exp(-\beta\epsilon) f(\beta)\, d\beta, \tag{2}$$

where the subscript q is a parameter of the distribution $f(\beta)$. It has been shown that if $f(\beta)$ is assumed to be the χ^2 distribution, a special case of the gamma distribution of variable β, present in many common situations [36], the generalized Boltzmann factor becomes [25,26,37]

$$\exp_q(-\beta_0\epsilon) = [1 - (1-q)\beta_0\epsilon]^{\frac{1}{1-q}}, \tag{3}$$

which is the same expression proposed in the context of nonextensive statistical mechanics [38,39]. It is important to note that the χ^2 distribution,

$$f(\beta) = [\Gamma(n/2)]^{-1}\left(\frac{n}{2\beta_0}\right)^{n/2} \beta^{-1+n/2} \exp\left(-\frac{\beta}{2\beta_0}n\right), \tag{4}$$

must be parameterized such that the heat reservoir temperature β_0^{-1} coincides with the average of the β^{-1}; that is, $\beta_0 = \langle\beta\rangle = \int_0^\infty \beta f(\beta) d\beta$. The nonextensive parameter q, set as

$$q = 1 + 2/n, \tag{5}$$

is associated with the relative dispersion of β, according to $q = 1 + (\langle\beta^2\rangle - \beta_0^2)/\beta_0^2$, where n is the number of degrees of freedom. This point will be resumed later in the next section.

In the present work, our interest is on the kinetic behavior of the chain during the search stage of the folding process. Therefore, for MC realizations we assume a generalized transition probability $\exp_q(-\beta_0\Delta\epsilon_{ab}) = \langle\exp(-\beta\Delta\epsilon_{ab})\rangle$ [see Eq. (2)] between the configuration a with energy ϵ_a, and configuration b with energy ϵ_b, that is,

$$\exp_q(-\beta_0\Delta\epsilon_{ab}) = [1 - (1-q)\beta_0\Delta\epsilon_{ab}]^{\frac{1}{1-q}}, \tag{6}$$

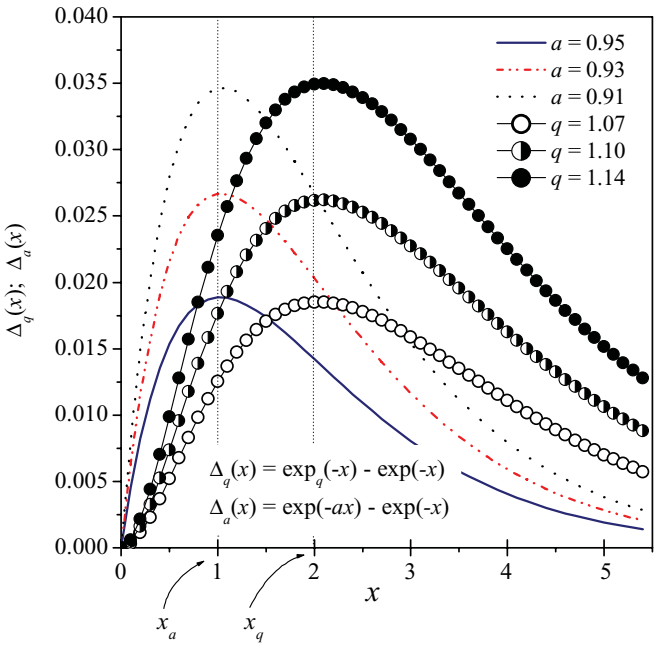

FIG. 1. (Color online) Functional difference between the generalized and the ordinary exponential function: $\Delta_q(x) = \exp_q(-x) - \exp(-x)$, with $q \gtrsim 1$, and between two ordinary exponential functions with the argument of one of them rescaled: $\Delta_a(x) = \exp(-ax) - \exp(-x)$, with $a \lesssim 1$; the parameter a represents a small increment in the system temperature. Their profiles are similar, but, for a and q near to 1, the maximum of $\Delta_a(x)$ and $\Delta_q(x)$ occur at about $x_a = 1$ and $x_q = 2$, respectively. In both cases, x_a and x_q increase very slowly when a and q depart from 1.

where $\Delta\epsilon_{ab} = \epsilon_b - \epsilon_a$. For the sake of adequacy (and to avoid misunderstandings), henceforward the heat reservoir temperature β_0^{-1} will be represented by T, and any reference to local fluctuations should be understood as the effect of local average kinetic energy, as discussed above [in the paragraph just before Eq. (2)].

For $q \gtrsim 1$ the q-exponential and the conventional exponential function approach each other. Actually, one may expect that $\exp_q(-x)$ is effectively equivalent to the conventional exponential function in which its argument has been slightly changed (that is, with the temperature somewhat increased). Comparing the two following difference functions, namely, $\Delta_q(x) = \exp_q(-x) - \exp(-x)$ and $\Delta_a(x) = \exp(-ax) - \exp(-x)$, as shown in Fig. 1 for $a \lesssim 1$ and $q \gtrsim 1$, one sees that their profiles are similar, although the maximum of $\Delta_a(x)$ occurs about $x_a = 1$, and for $\Delta_q(x)$ it is about $x_q = 2$. Actually, the $\lim_{a\to 1}(x_a) = 1$, while $\lim_{q\to 1}(x_q) = 2$; in both cases x_a and x_q increases slowly when a decreases from one and q increases from one.

In the next section, we compare the effect of both approaches on the configurational kinetics through MC simulation, and then we discuss on the possible implications on understanding the folding mechanism. In order to address this issue directly, we consider the folding time t and the folding characteristic time τ (described in the next section) as useful analytical amounts emerging from folding kinetics. The comparison between the two approaches emphasizes the effects of local thermal fluctuations on a class of heterogeneous systems (with nanosized components) in the context of the nonextensive statistical mechanics.

V. RESULTS AND DISCUSSION

Some properties of the lattice model described in Sec. II are very important for the present study. The model is virtually able to fold any 27-mer sequence properly designed (the design rule is provided) for any maximally Compact Self-Avoiding (CSA) configuration. In addition, the model reproduces the observed correlation between folding rates and the complexity of the native structures [33]. We selected three CSA configurations (targets) representing uncorrelated native structures. For each target α, a set $\{p_i\}_\alpha$ of N independent Monte Carlo simulations represents the folding process of a diluted solution of N noninteracting proteins. The MC time t_i spent to find the native structure (first passage time) corresponds to the folding time of the ith protein in $\{p_i\}_\alpha$. So, at the end of N independent runs, one gets a set $\{t_i\}_\alpha$ of N independent folding times for target α. By counting the number of folding times that fall in each time interval $[t, t + \Delta t]$, one finally gets the decay histogram of the number of unfolded proteins as a function of the MC time t. These data are then fitted by one or more exponential functions, giving the specific characteristic folding time τ for that structure; τ corresponds to the inverse of the folding rate. The simulations were carried out in a given range of thermodynamic temperatures T for several values of the nonextensive parameter q, and for three distinct native structures. Each native structure is characterized by its topological complexities, which can be roughly estimated by its structural Contact Order [8].

The effect of parameter q on the folding characteristic time. Table I shows τ as a function of temperature T and nonextensive parameter q, for the intervals $0.9 \leqslant T \leqslant 1.5$ and $1 \leqslant q \leqslant 1.4$. The three target structures are identified as ID 866, 1128, and 36335. In general, τ depends on the structure complexity, and it is a continuous convex function of T and q. A number $N = 150$ of independent runs were used for each pair $(q; T)$. The structure ID 36335 presents higher topological complexity than the other two, a fact reflected in its larger τ.

For any temperature T_i, there is a specific $q_i = q(T_i)$, let us say q_i^*, that minimizes $\tau \to \tau_{\min}$, as emphasized by the bold outlined cells in Table I; better approximations can be achieved by extra refinement of q and a larger set of independents runs. The uncertainty in τ was estimated by the standard deviation of the mean of means, considering 20 distinct samples $\{t_i\}_N$ (with $N = 150$) taken from an extended set of 10^4 independent runs. The uncertainty $\delta\tau$ depends on the pair $(q; T)$; for $N = 150$, the smallest uncertainties, $\simeq 5\%$, occur for the q_i^* values that minimize the characteristic folding time $\tau \cong \tau_{\min}$. Note that for each target structure and specific temperature, the coarse changes $\Delta q = \pm 0.1$ about q_i^* produce variations $\Delta\tau$ on $\tau_{\min}$, which are smaller than the uncertainty $\delta\tau$. This fact is represented in Table I by two or more horizontally adjacent bold outlined cells, meaning that for the involved temperatures and values of q, $\delta\tau \gtrsim \Delta\tau$. On the other hand, when the system approaches the glassy regime ($q = 1$, and $T < 1$) the time spent in metastable states increases substantially, and so $\delta\tau$ is strongly influenced by the size N of set $\{t_i\}_N$ of independent runs.

TABLE I. The characteristic folding (MC) time t for three target (native) structures; the unit MC time used here corresponds to 8100 attempts to move the chain. For each structure and temperature T_i there is a specific $q = q(T_i) \geqslant 1$ that minimizes τ (cells outlined in bold). Due to the higher topological complexity of structure ID 36335, its τ is 5 to 10 times larger than the corresponding values of τ for the two other structures. A set of $N = 150$ independent runs was used to estimate τ for each pair (T,q). The figures were rounded off according to average relative uncertainty $\delta\tau = 10\%$ (two significant figures); see text. At $T = 1$, the interval of values for q that give the smallest folding characteristic times are usually dependent on the complexity of the corresponding native structure (cells outlined in bold, bold figures).

	ID 1128		*q*		CO = 0.29894
T	1	1.1	1.2	1.3	1.4
0.9	560	54	20	22	48
1	130	28	**19**	26	67
1.1	40	22	25	40	130
1.2	39	18	29	65	320
1.3	21	21	52	190	470
1.4	23	30	80	280	1000
1.5	28	47	170	810	2300

ID 866		*q*		CO = 0.27513
1	1.1	1.2	1.3	1.4
330	30	31	30	45
120	**24**	31	36	81
47	26	30	50	130
24	26	36	90	310
21	26	57	190	620
24	36	92	320	1400
26	59	170	710	3100

ID 36335		*q*		CO = 0.30423
1	1.1	1.2	1.3	1.4
8700	870	240	310	470
3800	560	**230**	380	1100
1450	230	210	520	1700
680	200	260	780	3200
460	180	370	1560	…
170	210	…	…	…
200	290	…	…	…

In general, simplified models associate one specific optimal folding temperature $T_{\rm opt}$ for each particular sequence [40] [41], as the temperature that maximizes the folding speed for that sequence. However, having in mind that the proteins of any living organism work optimally in a relatively narrow temperature interval, $\tau_{\rm min}$ is adopted here as the optimum τ, the actual characteristic folding time. This is also in accordance with the assumed hypothesis of "folding as fast as possible" (see Sec. II). As we will comment below, q can be considered as a dynamic variable, dependent on each instantaneous configuration. However, assuming the hypothesis of "folding as fast as possible" allow us to determine q^*, a kind of averaged q, which specifies the actual characteristic folding time $\tau = \tau_{\rm min}$.

Table I is a kind of unrefined encompassing survey of how τ depends on T and q. Although a more detailed investigation is necessary (note, for instance, that at $T = 1$, for the three target structures, q^* lies in the range from $q = 1.1$ up to 1.2), the present scenario suggests that the kinetics of the search mechanism is equally reproduced, regardless whether the chain configurations are relatively weighted by means of the generalized Boltzmann factor at temperature T [see Eq. (3)] or by the conventional Boltzmann factor at some higher temperatures T' with respect to T. This result is readily ratified if, for each q, the behavior of τ is plotted as a function of the translated temperature scale $\mathcal{T}_q = T - T_q^*$, as illustrated in Fig. 2 for structure ID 1128. T_q^* is the temperature in which τ approaches $\tau_{\rm min}$ for that value of q. Essentially, all curves coalesce, behaving in nearly the same way about $\mathcal{T}_q = 0$.

Folding time distribution. A more detailed examination shows that the distributions $\Phi(t)$ of folding times t can be essentially the same for both approaches, that is, using either the generalized $\exp_q(-\Delta\epsilon/kT)$, or the conventional Boltzmann factor $\exp(-\Delta\epsilon/kT')$, with $T' > T$. The result presented in Fig. 3 shows $\Phi(t)$ for the structure ID 866, for different values of T and q; here a much larger number N of independent runs were employed ($N = 10^4$) for each case. In general, the distributions are better fitted with one or more lognormal curves, depending on the temperature.

For $(q;T) = (1;1)$, conventional Boltzmann factor and $T = 1$, the system approaches the glassy regime with manifestation of ergodic difficulties, as indicated by the three peaked curve (Fig. 3, open, smaller circles). As T increases from $T = 1$, the domain of t is accordingly reduced. At $T \cong 1.5$ the distribution presents the smallest domain, namely, $0 < \ln(t) < 5.5$, and as T increases from this point its behavior is reverted: The size of the domain starts to increase again, and the curve peak moves toward a larger t.

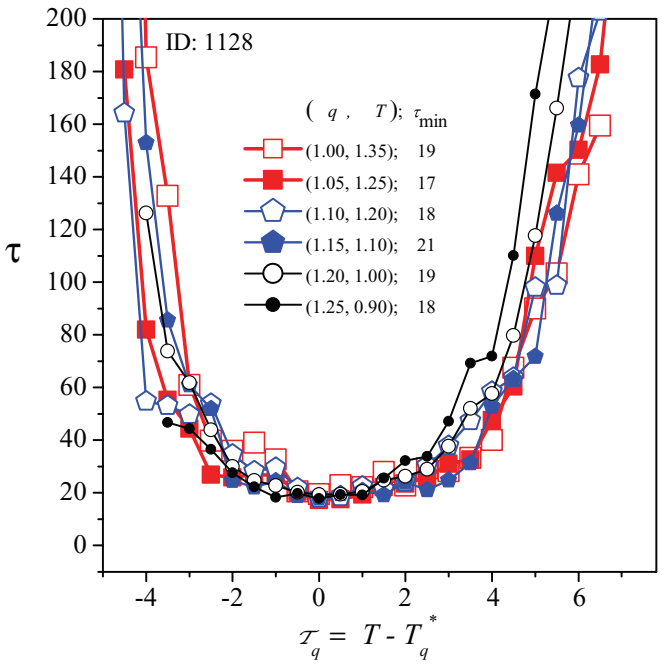

FIG. 2. (Color online) The τ vs T "U shape": Folding characteristic time τ as a function of the translated temperature $\mathcal{T}_q = T - T_q^*$ for several (q,T) combinations. T_q^* is the temperature in which $\tau = \tau_{\rm min}$ for the specified value of q. For each q, the temperature range was covered by increments $\Delta T = 0.05$. All curves behave very similarly around $\mathcal{T}_q = 0$.

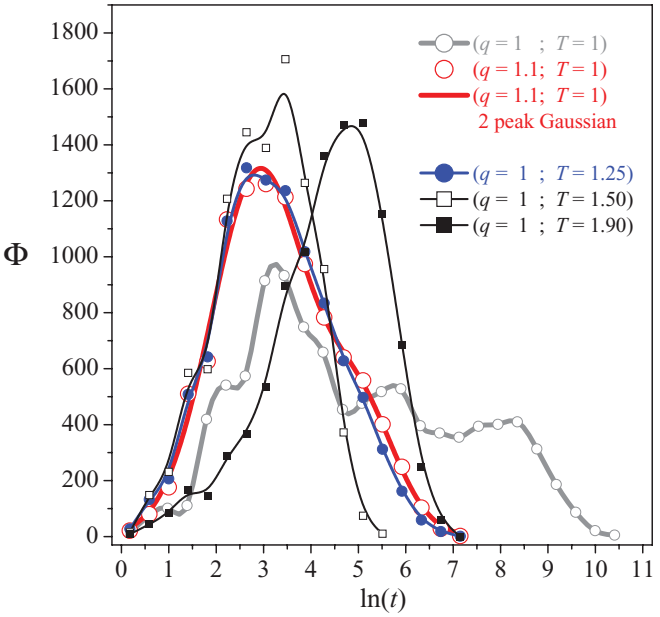

FIG. 3. (Color online) Folding time distributions Φ for structure ID 866. The lognormal distribution for $q = 1.1$ and $T = 1.0$ (larger open circles) was fitted using two peaks Gaussian [continuous thicker line, amplitude: $(A_1; A_2) = (1310 \pm 40; 430 \pm 70)$; width $(w_1; w_2) = (0.49 \pm 0.07; 0.8 \pm 0.3)$, and average $(\mu_1; \mu_2) = (2.91 \pm 0.09; 5.0 \pm 0.2)$]. For $q = 1.0$ the folding time behavior is shown for several temperatures (fitted by basic spline functions). In particular, the folding time distributions for $(q,T) = (1.0,1.25)$ and $(1.1,1.0)$ are practically the same, confirming that the generalized Boltzmann factor $(q > 1)$ at temperature T corresponds to the conventional Boltzmann factor $(q = 1)$ with a certain increase in T. In the region of the smaller folding times $[\ln(t) < 2.5]$ and for temperatures in the interval $1.0 \leqslant T < 1.5$, all curves coalesce, including the case $(q,T) = (1.1,1.0)$. At $T = 1.0$ and $q = 1$ (open gray circles) the system approaches the glassy regime, and manifestation of ergodic difficulties is evident. At temperature $T = 1.5$ the distribution has the lowest domain, that is, $0 < \ln(t) < 5.5$ (open squares), and inasmuch as T increases from this point, the peak of the distribution moves again toward larger t, as illustrated for $T = 1.9$ (full squares).

In the region of the smaller folding times, namely, for $\ln(t) < 2.5$, for the cases with $q = 1$ and T restricted in the interval $1.2 \leqslant T \leqslant 1.5$, all curves present the same behavior (Fig. 3). The meaning for this is that, even with the temperature 25% higher than $T = 1.2$, there exists some starting configurations (random, totally open: without any topological contact) that, combined with certain configurational evolution, can lead the chain very rapidly into the native structure. Note that this is also true for the case $(q; T) = (1.1; 1)$. Remarkably, $(q; T) = (1.1; 1)$, open large circles, and $(q; T) = (1; 1.25)$, full smaller circles (generalized and conventional Boltzmann factor, respectively), give practically the same frequency distribution $\Phi(t)$, implying in the convergence of the folding characteristic time for the two cases, namely, $\tau_{\min} = 24$ and 26, respectively (see Table I and Fig. 3).

Native stability. As pointed above (Sec. III), the additive potential employed here satisfies marginally the segregation principle, and so, once in the native state, it will be stable only at relatively low temperatures. In accordance with the previous

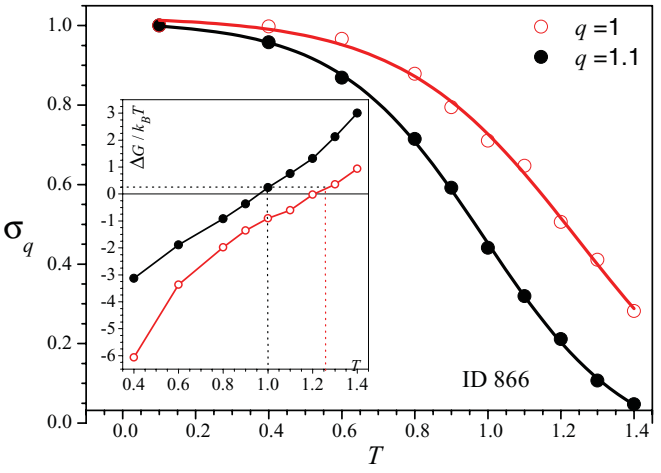

FIG. 4. (Color online) The native stability. The fraction of time in which the chain stays in the native state along the simulation as a function of the temperature T, namely, $\sigma_q(T)$, is shown for $q = 1$ and $q = 1.1$. In the inset, the native stability [actually $\Delta G/(k_B T)$] is shown as a function of the temperature. Note that at $(q; T) = (1; 1.25)$ and $(q; T) = (1.1; 1)$, the native stability is the same: $\Delta G/(k_B T) = 0.25$, as emphasized by dotted lines. This result (for structure ID 866) is a direct consequence of the coalescence of the distribution of folding time Φ for the same two pairs $(q; T)$ used here (as shown in Fig. 3, small full and larger open circles).

results, the native-state stability changes with q: If q increases, the stability diminishes as if the temperature of the whole system was increased. Figure 4 illustrates (target structure ID 866) the relative residence time σ_q in the native state as a function of the temperature T for two cases, namely, $q = 1$ and 1.1. For the present proposal, the native state is defined as any structure with at least 24 out of 28 native contacts: Once the native state is found (actually, to save considerable simulation time, the runs start with the chain already in the native structure), the chain systematically leaves and comes back into the native state. Therefore, the relative residence time $\sigma_q(T)$ gives the fraction of time in which the chain stays in the native state along the simulation. This is achieved by the ratio between the sum of MC time intervals in which the chain stays in the native state and the total simulation MC time. Therefore, interpreting kinetically the ratio $K_{\text{eq}} = [N]/[U]$ in the expression for the stability $\Delta G = -k_B T \ln K_{\text{eq}}$, that is, $K_{\text{eq}} = \sigma_q/(1 - \sigma_q)$, then

$$\Delta G/(k_B T) = -\ln\left[\frac{\sigma_q(T)}{1 - \sigma_q(T)}\right]. \tag{7}$$

The inset of Fig. 4 shows how stability changes with the system temperature T. Note that in perfect agreement with the results shown in Fig. 3 for the characteristic time distribution, the effect of the generalized Boltzmann factor is equivalent to the increase in the system temperature: One finds $G/(k_B T) \cong 0.25$ for $(q; T) = (1.1; 1)$, and about the same 0.25 for $(q; T) = (1; 1.25)$, as indicated by the dotted lines.

The apparent equivalence between the generalized and conventional Boltzmann factor. Although these results corroborate the notion that the net effect of the generalized Boltzmann factor $(q \gtrsim 1)$ is equivalent to increasing the system temperature, this apparent sameness cannot be employed here as a useful

resource because, in real systems, the folding process is optimized in a relatively narrow range of temperature for the total set of proteins of each particular living organism. As already mentioned in Sec. II, the search mechanism operates equally in a large temperature interval, from about zero to 100 °C. However, if proteins are exposed to increasing temperatures, loss of solubility or enzymatic activity is observed over a fairly narrow range. For the set of proteins of each organism there is a working temperature interval $T_\omega - \Delta T_a \lesssim T \lesssim T_\omega + \Delta T_b$ [in general with $(\Delta T_a + \Delta T_b)/T_\omega \lesssim 15\%$; absolute temperature] outside of which its functionality can be seriously reduced or completely lost. Therefore, the system temperature T must be kept as the reference temperature, measured macroscopically, and all thermal characteristics of a nanosize body, in response to the local fluctuations, should be conveniently controlled by the nonextensive parameter q.

Parameter q as a dynamic variable. For any target structure in contact with a specific thermal reservoir (the solvent) at temperature T, it is always possible to adjust q in order to get the optimum τ. But what intrinsic factors would determine a specific q^* value that would induce the fastest folding for that specific protein? In the present case, namely, a chain evolving through the configurational space from an open chain into a compact specific globule, the straightforward idea comes from the observation that the effects of local thermal fluctuations should be dependent on the spatial scale [42,43]. Along the folding process, wrong packing tendencies and specific traps are recurrent, but local thermal fluctuations—on such a nanometric chain—can disassemble them, promoting a rich variety of globule sizes and shapes during the process.

Therefore, let us consider Eq. (5), which allows us to associate the dynamic nature of the number of degrees of freedom n with the parameter q [Eq. (5)]. As the chain degree of compactness changes in the course of time, n changes accordingly: For an open chain we have $q_{\min} = 1 + 2/n_{\max} \to 1$, and for a fully compact globule $q_{\max} = 1 + 2/n_{\min}$. Using $n_{\min} = 1$ as a limiting condition, one gets a kind of upper bond for q, that is, $\sup q = 3$. Therefore, as the chain suffers thermal effects depending on its compactness, it is reasonable to assume that the simulation process is governed by a variable [44] rather than fixed q. The instantaneous q is then functionally linked to the compactness of the globule. For a flexible, linear, open chain, n grows with its length, which is proportional to the number of monomers. But, as packing progresses, the number of inter-residue contacts increases, and for fairly compact globules, n becomes rather a function of the size of the globule surface, which can be roughly estimated by the square of the radius of gyration R_G. Therefore, let us assume that n is proportional to R_G^2, and then, from Eq. (5), one gets $q = 1 + \alpha/R_G^2$. The amount α is determined by the maximum q, a specific $q_{\max}$, which determines the characteristic folding time $\tau = \tau_{\min}$; that is $\tau(q \neq q_{\max}) > \tau_{\min}$. We tested several distinct target structures with different topological complexities, and in all cases $q_{\max} \to 1.33$. This peculiarity allowed us to write a unique expression relating q and the instantaneous R_G, valid, in principle, for all target structures, that is,

$$q = 1 + 2(q_{\max} - 1)/R_G^2. \tag{8}$$

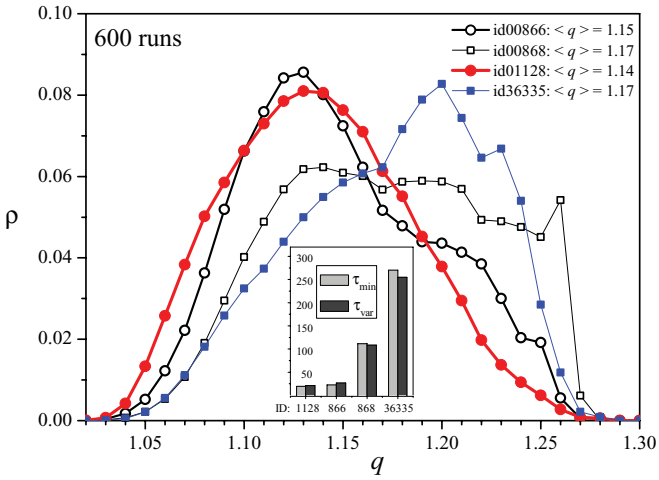

FIG. 5. (Color online) Average normalized distribution ρ of q for four targets; 600 runs were used in each case. For the structure ID 1128 ρ is clearly monomodal, but all other cases present a multimodal behavior. Their distribution averages $\langle q \rangle$ differ only slightly, namely, $\langle q \rangle \simeq 1.15$ for IDs 1128 and 866, and $\langle q \rangle = 1.17$ for IDs 868 and 36335, but their distinct shapes reflect better the topological and energetic landscape complexity. Note the congruence between the values $\langle q \rangle$ (about 1.15) and the optimal values of q in Table I (ranging from 1.1 to 1.2).

Therefore, for an extended chain, R_G^2 approaches to its maximum, and so $q \to 1$, while at the native structure R_G reaches its minimum, which in our case, a compact self-avoiding configuration, it is $\min\left(R_G^2\right) = 2$. It is interesting to note that if $n = 6$ is used in Eq. (5), for the native structure, one gets $q = 1 + 2/6 = q_{\max}$. As the native structure for the present problem is a compact $3 \times 3 \times 3$ cube, the value $n = 6$, as well the $q_{\max} = 4/3$, can be fully understood if one degree of freedom, on average, is associated with each face of the cube.

Figure 5 shows the normalized distribution of q for four target structures. The distribution ρ for structure ID 1128 is monomodal, but multimodal for the other three. The averages, namely, $\langle q \rangle \cong 1.15$ for IDs 1128 and 866, and $\langle q \rangle = 1.17$ for IDs 868 and 36335, differ by less than 2%, although their distribution shapes are significantly dissimilar. Such differences are indicative of distinct kinetics: If the energy landscape is fairly smooth and the number of topological constraints is small, then $\rho(q)$ tends to be monomodal. However, if many energetic traps and topological hindrances are present, producing recurrent wrong compact conformations, then larger values of q are systemically visited, enlarging and producing many peaks in the distribution curve $\rho(q)$. The inset of Fig. 5 illustrates the convergence between the characteristic time τ_{var}, obtained by using q as a dynamic variable, and $\tau_{\min}$ achieved by the optimum fixed $q = q^*$. In order to minimize the uncertainties, the results shown in Fig. 5 were obtained by using $N = 600$ runs for each target structure.

The three-dimensional (3D) configuration and corresponding contact map of the four targets considered here are shown in Fig. 6. The first two configurations, namely, ID 1128 and 866, present only favorable folding configurational patterns similar to α helix and β sheet (contact maps with lines parallel, adjacent to the secondary diagonal, and lines parallel

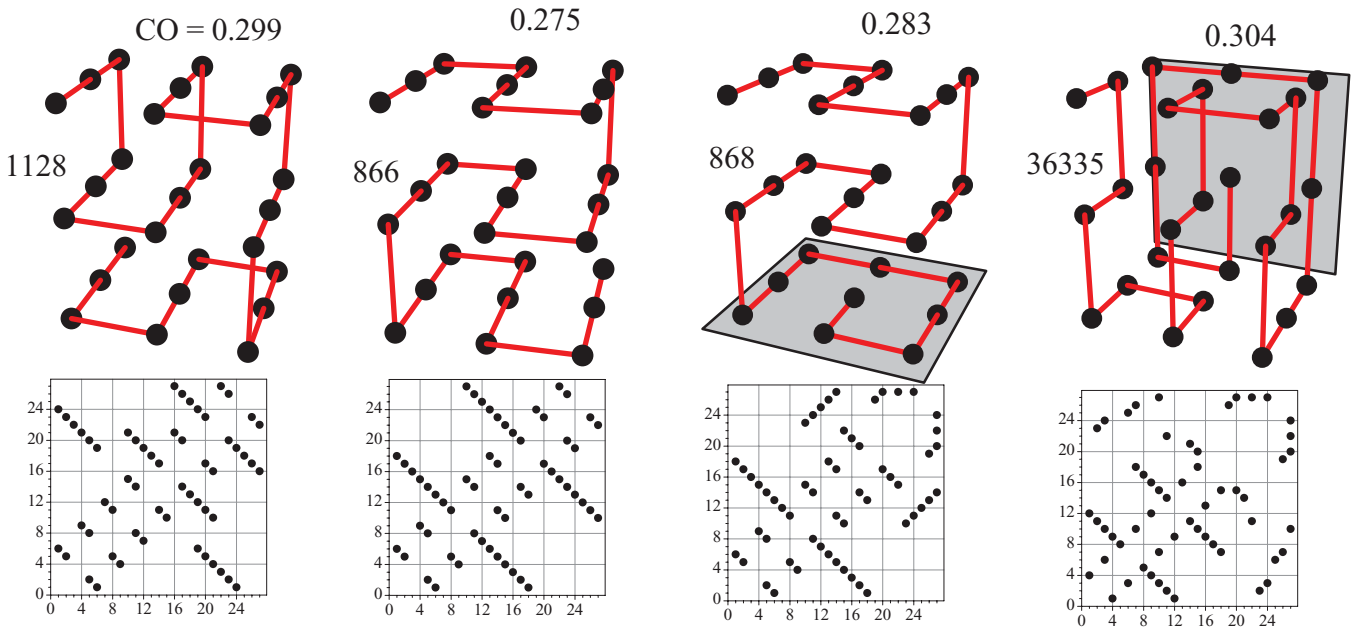

FIG. 6. (Color online) 3D configurations and corresponding contact maps for the four target structure. Each structure is labeled by its ID number (1128, 866, 868, and 36335) and the corresponding Contact Order. Chain segments that constitute folding unfavorable patterns are highlighted by gray planes (see text).

to the principal diagonal); therefore, they present the smaller τ (Fig. 5). Configuration ID 868 is very similar to 866; however, because a typical unfavorable folding pattern takes part of its structure, specifically the chain segment just above the gray plane, its τ is about four times larger the the previous cases. The last structure, ID 36335, is the more complex one, presenting a mix of favorable and unfavorable configurational patterns, easily evidenced by its contact map. Consequently, its τ is one order of magnitude greater than the two first ones. Contact order, as a global estimator of structural complexity, is not sensitive to structural details. However, it is a rough estimator.

VI. FINAL COMMENTS AND CONCLUSIONS

Here we have examined the effect of local thermal fluctuations on protein-folding kinetics through a coarse-grained lattice model. The model hypothesizes that (1) the process by which a protein folds is composed by two independent sequential steps, namely, search and stabilization, and (2) instructions encoded in the residues sequence provide a folding kinetics as fast as possible. The first premise enabled us to place emphasis on the search mechanism as a universal process guided by the hydrophobic force, and the second was used in order to associate the nonextensive parameter q or, in the case of variable q, its distribution, with each native structure.

Our results, based on the comparison between two approaches, namely, nonextensive ($q > 1$) and conventional statistical mechanics ($q = 1$), suggest that suitable thermal fluctuations—adequately achieved only in the nonextensive context—are important causal agents that drive the chain through the fastest possible courses to the native conformation. Although we have demonstrated numerically that the chain kinetics governed by the generalized Boltzmann factor ($q > 1$), at system fixed temperature T, is fully reproduced if the conventional Boltzmann factor is used, at temperatures $T' > T$ (Fig. 3), we argue that the two approaches, namely, $(q > 1; T)$ and $(q = 1; T > T')$, cannot be considered equivalent. Proteins have working markedly different temperatures among distinct organisms, but given that the set of all proteins of any specific organism functions optimally only in a narrow range of temperature, this question cannot be treated as a simple matter of reparametrization. For this specific problem, this apparent equivalence should actually be considered an artifact of the conventional method of statistical mechanics when taken as statistical inference, which was originally conceived to deal with homogeneous systems at equilibrium and in the thermodynamic limit.

The dependence of τ on temperature, shown in Fig. 2, has been commonly and exclusively attributed to peculiarities of the chain sequence; sequences are usually generated and selected for their ability to fold rapidly in a small and specific range of temperature [45]. Under the conventional Boltzmann perspective ($q = 1$), the rough nature of the energy landscapes can trap partially folded protein in non-native local minima for too long. However, a new scenario emerges when local thermal fluctuations experienced by nanoscale structures are associated with their spatial characteristics (such as size and degrees of freedom) by means of the parameter q. Specifically, such as chaperone-assisted folding, well-tuned thermal fluctuations help to disassemble chain segments wrongly collapsed, improving the fastness of the folding process; otherwise, using conventional statistical mechanics, it would be achieved only at extremely high (fatal) temperatures of the heat reservoir. Therefore, at fixed system temperature $T = 1$, a specific q^* that varies slightly around $q = 1.15$ depending on the native structure, minimizes the folding characteristic time, $\tau_{\min}$.

A more refined approach, associating a variable q with the instantaneous degree of compactness of the globule, predicts

a characteristic time τ_{var} essentially the same as that obtained by using a target-tunned parameter q^*, that is, $\tau_{var} \longrightarrow \tau_{min}$. Moreover, it also reveals that the simulation evolution of different target structures—with distinct topological complexities—are characterized by singular distributions of q, and average $\langle q \rangle$ converging to q^*. Such distribution $\rho = \rho(q)$, depending on the roughness of the free-energy landscape and on the nature and number of topological traps, can be monomodal or multimodal, reflecting the overall complexity of the energy landscape and topology of the native structure.

Under the present perspective, we can visualize two main driving forces supporting a continuous process of folding and unfolding during the search stage of the folding process: entropic forces compacting the chain and, on the other hand, local thermal fluctuations tending to open it. This process continues until, eventually, the neighborhood of the native state is reached. At this point, and only under this condition, the native structural peculiarities and chain energetic interactions, as encoded along the chain sequence, would be associated in a cooperative and fully productive way, guaranteeing the overall stability of the globule.

As a final remark we note that the present results clearly point at possible generalizations of the multipurpose Monte Carlo method for chains governed by complex energy landscapes: Using q as a variable self-adjusted by the chain conformational evolution, local thermal fluctuations can be automatically incorporate in the process, eliminating the need to seek specific optimized values q^* for each case.

ACKNOWLEDGMENT

We thank CNPq for funding with Grant No. 142651/2007-9 and PRODOC/CAPES.

[1] S. Kumar and Mai Suan Li, Phys. Rep. **486**, 1 (2010).
[2] D. Thirumalai, E. P. O'Brien, G. Morrison, and C. Hyeon, Annu. Rev. Biophys. **39**, 159 (2010).
[3] J. P. Dal Molin, Marco Antonio Alves da Silva, I. R. da Silva, and A. Caliri, Braz. J. Phys. **39**, 435 (2009).
[4] K. H. Mok, L. T. Kuhn, M. Goez, I. J. Day, J. C. Lin, N. H. Andersen, and P. J. Hore, Nature (London) **447**, 106 (2007).
[5] K. A. Dill, S. B. Ozkan, M. S. Shell, and T. R. Weikl, Annu. Rev. Biophys. **37**, 289 (2008).
[6] C. Levinthal, J. Chim. Phys. **65**, 44 (1968).
[7] B. Gillespie and K. W. Plaxco, Annu. Rev. Biochem. **73**, 837 (2004).
[8] K. W. Plaxco, K. T. Simons, and D. Baker, J. Mol. Biol. **277**, 985 (1998); D. N. Ivankov, S. O. Garbuzynskiy, E. Alm, K. W. Plaxco, D. Baker, and A. V. Finkelstein, Protein Sci. **12**, 2057 (2003).
[9] S. B. Ozkan, G. A. Wu, J. D. Chodera, and K. A. Dill, Proc. Natl. Acad. Sci. USA **104**, 11987 (2007).
[10] G. D. Rose, P. J. Fleming, J. R. Banavar, and A. Maritan, Proc. Natl. Acad. Sci. USA **103**, 16623 (2006).
[11] D. W. Bolen and G. D. Rose, Annu. Rev. Biochem. **77**, 339 (2008).
[12] E. M. Popov, IUBMB Life **47**, 443 (1999).
[13] J. Schonbrun and K. A. Dill, Proc. Natl. Acad. Sci. USA **100**, 12678 (2003).
[14] S. E. Jackson, Folding Design **3**, 81 (1998).
[15] F. Piette, S. D'Amico, C. Struvay, G. Mazzucchelli, J. Renaut, M. L. Tutino, A. Danchin, P. Leprince, and G. Feller, Mol. Microbiol. **76**, 120 (2010); N. J. Russel, Extremophiles **4**, 83 (2000).
[16] P. Wiggins, PLoS ONE **3**, e1406 (2008); M. Zeeb, G. Lipps, H. Lilie, and J. Balbach, J. Mol. Biol. **336**, 227 (2004).
[17] D. Chandler, Nature (London) **417**, 491 (2002).
[18] L. F. O. Rocha, M. E. Tarragó Pinto, and A. Caliri, Braz. J. Phys. **34**, 90 (2004).
[19] M. E. P. Tarragó, L. F. O. Rocha, R. A. da Silva, and A. Caliri, Phys. Rev. E **67**, 031901 (2003).
[20] J. A. Pelesko, *Self-Assembly: The Science of Things That Put Themselves Together* (Chapman and Hall/CRC, Boca Raton, Florida, 2007).
[21] D. E. Kim, Hongdi Gu, and D. Baker, Proc. Natl. Acad. Sci. USA **95**, 4982 (1998).
[22] A. M. Gutin, V. I. Abkevich, and E. I. Shakhnovich, Proc. Natl. Acad. Sci. USA **92**, 1282 (1995).
[23] G. H. Lorimer, FASEB J. **10**, 5 (1996).
[24] H. Ecroyd and J. A. Carver, IUBMB Life **60**, 769 (2008).
[25] C. Beck, Europhys. Lett. **57**, 329 (2002).
[26] A. K. Rajagopal, C. S. Pande, and S. Abe, in *Nano-Scale Materials: From Science to Technology,*Workshop on Nano-Scale Materials—From Science to Technology, APR 05-08, Puri, India (2006), pp. 241–248.
[27] O. J. E. Maroney, Phys. Rev. E **80**, 061141 (2009).
[28] H. S. Chung, J. M. Louis, and W. A. Eaton, Proc. Natl. Acad. Sci. USA **106**, 11837 (2009).
[29] R. A. da Silva, M. A. A. da Silva, and A. Caliri, J. Chem. Phys. **114**, 4235 (2001).
[30] L. F. O. Rocha, I. R. Silva, and A. Caliri, Physica A **388**, 4097 (2009).
[31] H. Li, C. Tang, and N. S. Wingreen, Phys. Rev. Lett. **79**, 765 (1997).
[32] K. A. Dill, Biochemistry **29**, 7133 (1990).
[33] I. R. Silva, L. M. dos Reis, and A. Caliri, J. Chem. Phys. **123**, 154906 (2005).
[34] A. Cooper, Proc. Natl. Acad. Sci. USA **73**, 2740 (1976).
[35] P. L. Privalov, in *Protein Folding*, edited by T. E. Creighton (W. H. Freeman, New York, 1992), p. 83.
[36] N. A. J. Hastings and J. B. Peacock, *Statistical Distributions* (John Wiley & Sons, New York, 2000).
[37] H. Touchette, in *Nonextensive Entropy*: *Interdisciplinary Applications*, edited by M. Gell-Mann and C. Tsallis (Oxford University Press, New York, 2004), p. 159.
[38] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).
[39] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics* (Springer, New York, 2009).

[40] D. M. Travasso, P. F. N. Faisca, and A. Rey, J. Phys. Chem. **33**, 125102 (2010).

[41] A. Imparato, S. Luccioli, and A. Torcini, Prog. Theor. Phys. Suppl. **184**, 339 (2010).

[42] C. Beck, G. S. Lewis, and H. L. Swinney, Phys. Rev. E **63**, 035303(R) (2001).

[43] C. Beck, Phys. Rev. Lett. **87**, 180601 (2001).

[44] G. A. Tsekouras and Constantino Tsallis, Phys. Rev. E **71**, 046144 (2005)

[45] A. R. Dinner, A. Sali, and M. Karplus, Proc. Natl. Acad. Sci. USA **93**, 8356 (1996).