# Frustration and hydrophobicity interplay in protein folding and protein evolution

Leandro C. Oliveira, Ricardo T. H. Silva, Vitor B. P. Leite, and Jorge Chahine

Citation: The Journal of Chemical Physics **125**, 084904 (2006); doi: 10.1063/1.2335638
View online: http://dx.doi.org/10.1063/1.2335638
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/125/8?ver=pdfcov
Published by the AIP Publishing

# Frustration and hydrophobicity interplay in protein folding and protein evolution

Leandro C. Oliveira, Ricardo T. H. Silva, Vitor B. P. Leite,[a),b)] and Jorge Chahine[a),c)]
*Departamento de Física, IBILCE, Universidade Estadual Paulista, São José do Rio Preto, São Paulo 15054-000, Brazil*

(Received 6 April 2006; accepted 17 July 2006; published online 25 August 2006)

A lattice model is used to study mutations and compacting effects on protein folding rates and folding temperature. In the context of protein evolution, we address the question regarding the best scenario for a polypeptide chain to fold: either a fast nonspecific collapse followed by a slow rearrangement to form the native structure or a specific collapse from the unfolded state with the simultaneous formation of the native state. This question is investigated for optimized sequences, whose native state has no frustrated contacts between monomers, and also for mutated sequences, whose native state has some degree of frustration. It is found that the best scenario for folding may depend on the amount of frustration of the native structure. The implication of this result on protein evolution is discussed. © *2006 American Institute of Physics*. [DOI: 10.1063/1.2335638]

## I. INTRODUCTION

Proteins are not random heteropolymers, but have, rather, been selected through evolution.[1] The effect of mutation on the stability of proteins is a crucial issue in protein evolution. Theoretical studies generally emphasize, following Gō's principle of minimal frustration, that good sequences must be optimized.[2] It is accepted that biological proteins have been selected through the natural evolution of the species. It is implied that there must be some tolerance in this optimization while still allowing a protein to exhibit satisfactory thermodynamic stability and efficient folding dynamics. The tolerance to amino acid substitution has been observed experimentally.[3–5] The motivation for the present study is to understand (i) how mutations affect optimally designed protein sequences and (ii) the role of hydrophobicity in the mutation processes.

Much work on hydrophobicity has been done in an attempt to answer the following questions: Do compact conformations due to hydrophobic collapse help protein folding?[6–11] Which scenario would proteins choose in order to fold faster: a fast nonspecific collapse followed by a slow rearrangement to reach the native state or a specific collapse with simultaneous formation of the native state? While studies have shown that some proteins undergo a burst hydrophobic collapse followed by their folding,[12–14] there is experimental evidence that some proteins collapse concomitantly with the formation of their native structure.[15] Some studies[6–8] correlate folding kinetics with four parameters defined as follows: (i) The folding temperature $T_f$ is defined as that where half of the chains are folded, or, in protein models used in simulations, as that at which the probability for finding half of the native contacts is 0.5; (ii) $T_g$ is the glass temperature at which the kinetics is dominated by traps due to many local minima in the energy landscape and at which the kinetics deviates from exponential behavior; (iii) $T_\theta$ is a temperature associated with a nonspecific collapse of the chain, which represents the burst hydrophobic collapse of some proteins; (iv) the stability gap is the difference between the energy of the native structure and all other states. These quantities have been used to define dimensionless parameters, which have revealed a correlation with the folding rates. One of these parameters, defined as $\sigma=(T_\theta-T_f)/T_\theta$, was introduced by Klimov and co-worker,[10,11] who demonstrated that fast folding sequences have small values of $\sigma$, which means that sequences that display a specific collapse with simultaneous formation of the native structure fold faster. This behavior precludes the scenario where the chain collapses to nonspecific structures, after which it rearranges itself slowly to find the native structure. Experimental data from small-angle x-ray scattering and circular dichroism[9] have corroborated the theoretical studies of Klimov and coworker by showing that the proteins which fold the fastest are those associated with small values of $\sigma$. On the other hand, Chiu and Goldstein[8] have shown, through the use of the diffusion equation, that marginally stable proteins fold faster in the presence of a nonspecific interaction that favors compact states. Whether collapse occurs before folding was also a matter of analysis in a theoretical study by Gutin *et al.*[16] Their results suggested that, if an overall attraction among residues dominates, then collapse precedes folding. As regards to the requisites for fast folding, the ratio $T_f/T_g$, proposed by Onuchic *et al.*[7] and Gillespie and Plaxo[17] on simplified models, was shown to be correlated with the folding rates. A different criterion proposed by Sali *et al.*[6] correlates fast folding kinetics with a large stability gap.

The main purpose of the present study is to show that the way that the folding rates change with temperature and also with the degree of frustration of the native state due to mutation depends significantly on the regime of hydrophobicity. This paper is organized as follows. In Sec. II the lattice

[a)]Authors to whom correspondence should be addressed.
[b)]Electronic mail: vleite@ibilce.unesp.br
[c)]Electronic mail: chahine@ibilce.unesp.br

 

model, mutation procedure, and simulation methods are discussed. In Sec. III, starting from a native nonfrustrated sequence, all possible sequences obtained by the mutation procedure are analyzed. Two selected mutated sequences, along with the native one, are studied in detail. In Sec. IV the criteria for folding transition temperature are discussed. In Sec. V the thermodynamic and kinetic behaviors of the selected sequences are described. In Sec. VI, the implications of the results are discussed.

## II. MODEL AND METHODS

The model used for the kinetic simulations has been extensively employed in previous studies.[18–20] The protein is modeled by a 27-length polymer chain (27-mer) on a three-dimensional cubic lattice. The energy of a given three-dimensional protein configuration is associated with the interaction between nonbonded monomers, and it is given by

$$E = N_l E_l + N_n E_n, \tag{1}$$

where $N_l$ is the number of contacts between monomers of the same type (like contacts) and $N_u$ the number of contacts between monomers of different types (unlike contacts). We used as a native sequence a three-letter code sequence ABABBBCBACBABABACACBACAACAB (to which we gave the name native sequence), which forms a well-known nonfrustrated (optimized) native structure.[20,21] We studied the thermodynamics and kinetics of this lattice model at the low- and high-hydrophobicity limits. The hydrophobicity of this model was discussed in detail by Chahine *et al.*[24] The average nonbonded contact energy is proportional to $E_u$ and $E_l$ and the relative frequency of the $u$ and $l$ contacts. The hydrophobicity is normalized by the dispersion in the contact energies, which is associated with the roughness of the energy landscape. In our simulations, $E_l=-1$, $E_u=-3$ regime yields a favorable energy for contact formation, characterizing the occurrence of collapse; this is the high-hydrophobicity (HH) limit. The low-hydrophobicity (LH) limit is defined by $E_l=3$, $E_u=-3$, which shows on average no attraction between monomers. In a $3\times3\times3$ cube conformation, the maximum value of the number of contacts is 28, and so the lowest possible energy is $-84$. This is the energy of the unfrustrated (native) state, which means a state where there are no contacts between monomers of different types. For the 27-mer it is possible to generate all the cube conformation,[21] which makes it possible to verify that the mentioned sequence has a nondegenerate ground state.

The mutated sequences were obtained by the permutation of pairs of different monomers. In this way, the proportion of monomers $A$, $B$, and $C$ is maintained constant. The native structure for the mutated sequences was found among the 103 346 maximally compact conformations. Mutations were classified based on its frustration, gap, and $Z_{\text{score}}$. Frustration $f$ in a sequence is quantified by the number of unfavorable contacts in the native state, since the number of frustrated contacts raises the energy of the native state; frustration has a direct influence on the native state's stability and kinetics,

$$\text{gap} = E_1 - E_0, \tag{2}$$

where $E_0$ is the native state energy and $E_1$ the energy of the first excited states, and

$$Z_{\text{score}} = \frac{\langle E\rangle - E_0}{\sigma}, \tag{3}$$

where $\langle E\rangle$ is the average energy and $\sigma$ the energy standard deviation. These parameters for each sequence were calculated taking into account only the maximally compact structures. Good mutations in principle maximize both gap and $Z_{\text{score}}$. These parameters display good correlation with the stability of the ground state, even when only maximally compact conformations are considered.[22]

In the folding simulations, we used the Monte Carlo algorithm with the standard polymer local lattice moves, which are end, corner flip, and 90° crankshaft moves.[21] In order to find the density of states $\Omega(E,Q,Z)$ we used the single histogram Monte Carlo method.[23] Once the density is known the averages of the quantities of interest such as the mean energy $\langle E\rangle$, the average of the number of all contacts $\langle Z\rangle$, and the average of the number of native contacts $\langle Q\rangle$ can be calculated. Normalized histograms are used as an approximation for the probability and, once the degeneracy of the ground state is known $[\Omega(-84,28,28)=1]$, the free energy is readily obtained, which allows for the determination of the density of states.

## III. MUTATIONS

Starting from the initial native sequence $N$, mutations were obtained by permuting the position of two monomers of different types in the sequence. This procedure follows some criteria. Among many mutated sequences generated by these permutations, it is selected sequences that have the same nondegenerate ground state as that of the native sequence. The feasibility of this procedure is guaranteed by the known maximally compact cube conformations. The mutated sequence is threaded in all the cube conformations, making it possible to verify the degeneracy of the native state. For one permutation there are 236 different sequences. Of these sequences, 206 present a single native state, and 198 (96%) present the same native structure as $N$. For two permutations there are 19 815 different sequences. Of these sequences 8933 present a single native state, and 3991 (45%) present the same native structure as $N$. The mutated sequence score distributions of gap versus $Z_{\text{score}}$ for one and two mutations are shown in Fig. 1.

"Good" and "bad" mutations were classified based on $f$, gap, and $Z_{\text{score}}$. In the process of evolution of a protein, it is expected that good mutations fold satisfactorily, with comparable folding times and folding temperatures to the native nonfrustrated sequence. Following this reasoning, mutations with considerably longer folding time and a less stable native state are unlikely to survive the natural selection process, since they do not satisfy this folding criterion. The distributions of gap versus $Z_{\text{score}}$ and $f$ vs $Z_{\text{score}}$ are shown in Fig. 1. Two good mutations, in which $f$ equals 3 and 6 and with the same native structure as the native one, were selected and

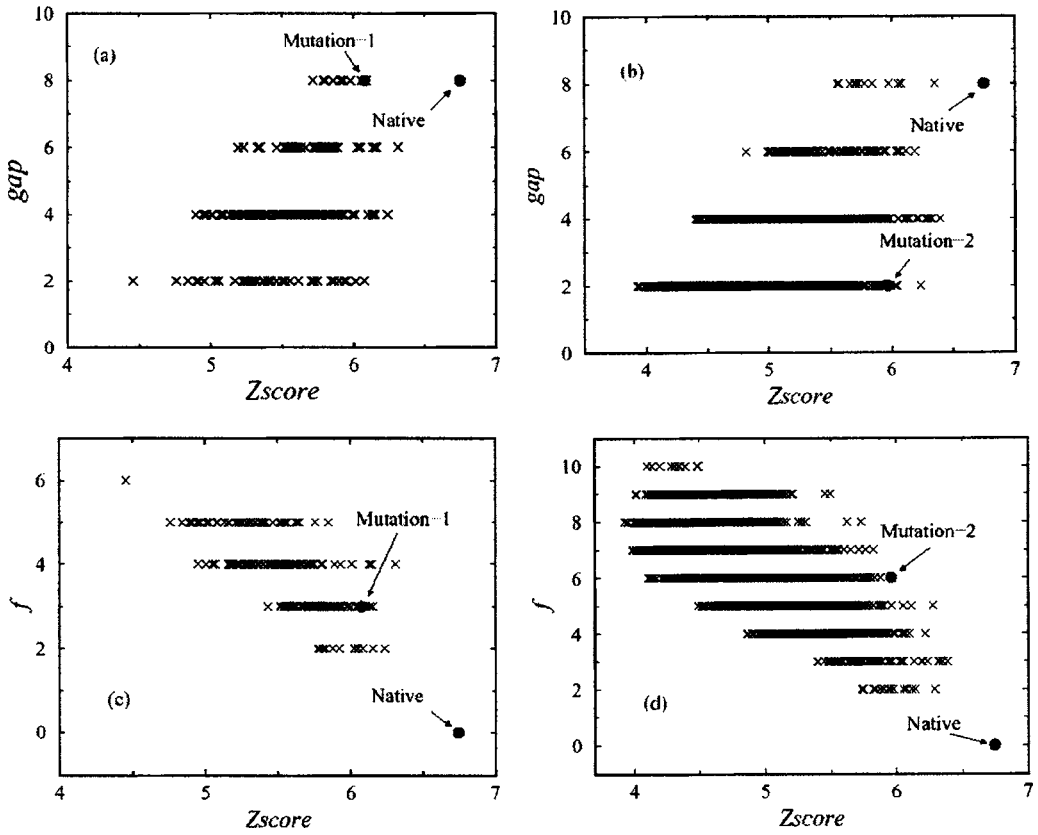


FIG. 1. Gap vs $Z_{score}$ is shown for mutation 1(a) and mutation 2(b); the number of frustrated contacts $f$ vs $Z_{score}$ is shown for mutation 1(c) and mutation 2(d). The native and selected good mutations are marked. Mutation 1 and mutation 2 have, respectively, three and six frustrated contacts in the native conformation.

their thermodynamic and kinetic behaviors were compared with those of the native one. From the single mutation distribution [Figs. 1(a) and 1(c)], a good sequence with three frustrated contacts was ABABBBCBACBABACA-CACBABAACAB (mutation 1). From the double mutation distribution [Figs. 1(b) and 1(d)], a good sequence with six frustrated contacts was CBABBBCBCCBABABA-CACBAAAAAAB (mutation 2). If the argument of robust good mutations is reasonable and, from the distributions of Fig. 1, one would expect that the selected sequences (mutation 1 and mutation 2) should provide reasonable folding features. For constant values of gap and $f$, both mutations have high $Z_{score}$ compared to the average $Z_{score}$.

## IV. CRITERIA FOR FOLDING TEMPERATURE

Three criteria have been used in simulation studies of protein models in order to define the folding temperature: (i) $P_{nat}$ is the probability to find the native structure calculated from histogram techniques,[23] and the folding temperature is that which makes the probability equal to 0.5; (ii) $\langle Q \rangle$ is the average number of native contacts between the monomers that form the chain normalized by the total number of native contacts in the native conformation, in which case the folding temperature is that which turns the value of $\langle Q \rangle$ to 0.5; (iii) $\langle Q^* Q \rangle - \langle Q \rangle \langle Q \rangle$ is the fluctuation of $\langle Q \rangle$, where the folding temperature is that for which the fluctuation has a peak. Since $Q$ is related to the energy, this criterion is similar to the peak of the heat capacity (as a function of the temperature) to find the folding temperature. We performed simulations for two different sequences: the native one, whose folded structure has no frustration, and the mutation 2 sequence, whose folded structure has six frustrated contacts. The reason for conducting such simulations was due to the possible dependence of the difference in folding temperatures of the two sequences $\Delta T$, according to the three criteria. Once this dependence is calculated, for computational reasons, we chose the criterion that produced the lowest $\Delta T$. The mutated sequence may show a very low folding temperature, which may lead to a high computational time to fold the chain. Figure 2 shows the results for the two sequences, the native (nat) and mutation 2 (mut) in the two regimes of hydrophobicity (LH and HH). The lower portion of the figure refers to the native sequence. Curve (c) is the native sequence in the HH regime and (d) is the native sequence in the LH regime. The higher portion formed by curves (a) and (b) shows the corresponding results for the mutated sequence. When frustration is absent the three criteria give nearly similar values for the folding temperature as shown by curves (c) and (d). The folding temperatures differ by at most 6% for each case. These differences increase by a small amount for the mutated sequence in the HH regime as shown by curve (a). As for the LH regime, the differences are significant when mutation is introduced, as shown by curve (b), which also shows that criterion (2) (represented by the dotted line) produces the smallest $\Delta T$, when comparing the folding temperatures between plots (b) and (d). This criterion makes use of $\langle Q \rangle$, the

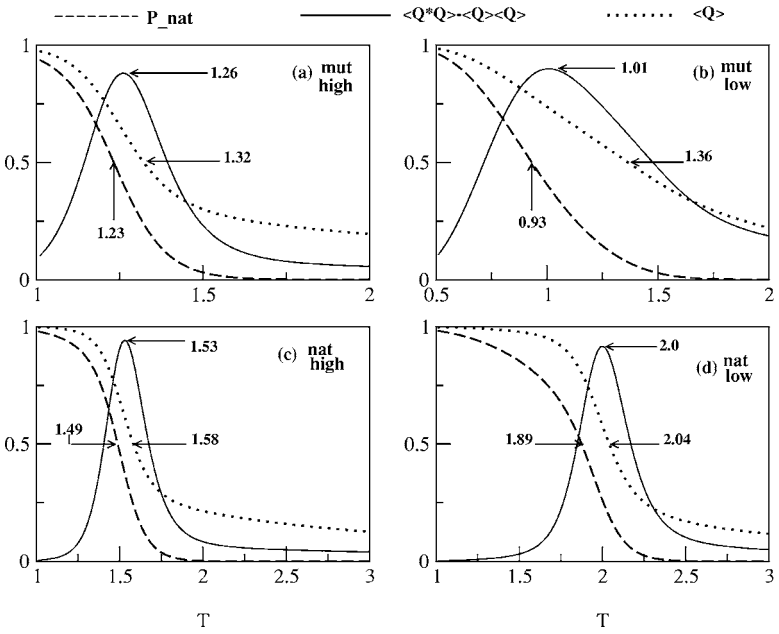


FIG. 2. The figures show the three different criteria for defining the folding temperature for the two sequences, the native (nat) and mutation 2 (mut) in the two regimes of hydrophobicity (LH and HH). Criterion $P_{\mathrm{nat}}$ (with dashed line) is the probability of finding the native state. Criterion $\langle Q^{*}Q\rangle - \langle Q\rangle^{*}\langle Q\rangle$ (continuous line) is the fluctuation of the number of native contacts. Finally criterion $\langle Q\rangle$ (dotted line) is the average of the native contact, which is the average number of native contacts divided by the total number of native contacts in the native structure (28). Curve (c) is the native sequence in the HH regime and (d) is the native sequence in the LH one. Curves (a) and (b) show the corresponding results for the mutated sequence. In the absence of frustration the three criteria give nearly similar values for the folding temperatures as shown by curves (c) and (d). In the presence of frustration the differences between those temperatures increase by a small amount for the mutated sequence in the HH regime as shown by curve (a), but increase significantly for the LH limit.

normalized average number of native contacts, which is 0.5 at the folding temperature. This is also the order parameter used by Gutin *et al.*[16]

Monte Carlo simulations and histogram techniques were used to study the behavior of the native sequence in the two regimes of hydrophobicity. The occurrence of a collapse transition which is not related to folding, i.e., a nonspecific collapse, is studied through the parameter $\langle Z\rangle$, which is the average number of any contacts, native or not. Figure 3(a) shows $\langle Z\rangle$ and its fluctuation as a function of the temperature, for the LH regime. The midtransition for $\langle Z\rangle$ and the peak in the fluctuation occur at temperatures close to $T=2.0$. In this regime the chain collapses directly to the native structure. Figure 3(b) shows the corresponding quantities for the HH regime. Now the midtransition for $\langle Z\rangle$ and the peak in its fluctuation occur for temperature $T=3.0$, which is significantly different from the temperatures shown in Fig. 2(c) (the dotted and continuous lines) for the native quantities $\langle Q\rangle$ and $\langle Q^{*}Q\rangle - \langle Q\rangle^{*}\langle Q\rangle$. The increase of $\langle Z\rangle$ and a peak in the fluctuation at a temperature much higher than the folding temperature are indicative of a nonspecific collapse of the chain. This conclusion correlates with the results of Ref. 16: when the overall interaction between monomers is attractive (HH), folding is preceded by a collapse, which does not occur for the LH regime where the overall interaction is nearly zero or slightly repulsive. Collapse was studied in detail for the mutated sequences at HH, and it was observed that the collapse transition always occurs at the same temperature around 2.8.

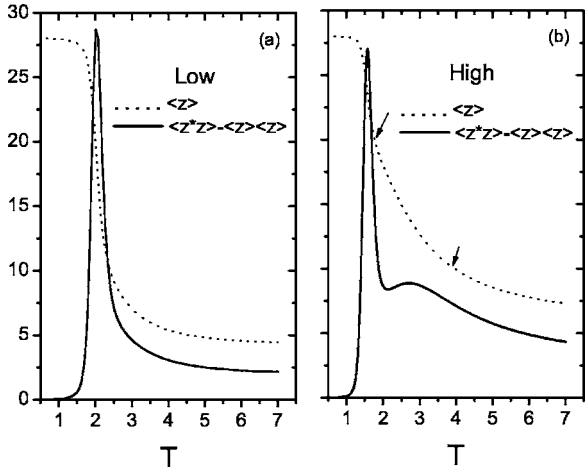


FIG. 3. The curves show the average number of contacts $\langle Z\rangle$ (right axis) and its fluctuation (left axis) as a function of the temperature for LH and HH. For the HH limit, the arrows show a fast increase in $\langle Z\rangle$ which causes the emergence of the smaller peak. Below $T=2$, the value of $\langle Z\rangle$ experiences an even faster increase due to the folding transition, which is related to the higher peak of the specific heat. The curves for the LH limit show a single transition related to the folding of the chain.

## V. THERMODYNAMICS AND KINETICS

This section sets out the main results of this work, which are related to the folding thermodynamics and kinetics of native and mutated sequences in the two regimes of hydrophobicity. The transition temperatures $T_f$ and $T_g$ and the folding parameter $T_f/T_g$ were calculated for native, mutation 1 ($f=3$), and mutation 2 ($f=6$) sequences at HH and LH regimes. For both hydrophobicity regimes $T_g$ does not depend on the degree of frustration, which is expected, since $T_g$ is associated with the roughness of the landscape and does not depend on the sequence details. $T_f$ decreases significantly with $f$ at LH, and remains approximately constant at HH. The summary of these results is shown through the ratio $T_f/T_g$ in Fig. 4. From this figure it is evident that frustration strongly affects the stability and foldability of a sequence at LH, which suggests that in this regime an optimized sequence is not robust with respect to mutations. That is not the case for the HH regime, which shows no dependence on degree of frustration. The kinetic results reinforce this evidence.

The kinetics of three sequences was studied by measuring their folding time ($\tau$), which was calculated by perform-

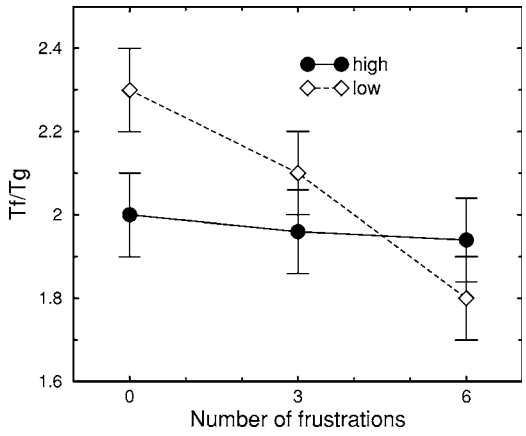


FIG. 4. Folding parameter $T_f/T_g$ as a function of the number of frustrated contacts $f$ in the native conformation and hydrophobicity. $f=0$, 3, and 6 correspond to native, mutation 1, and mutation 2 sequences, respectively.

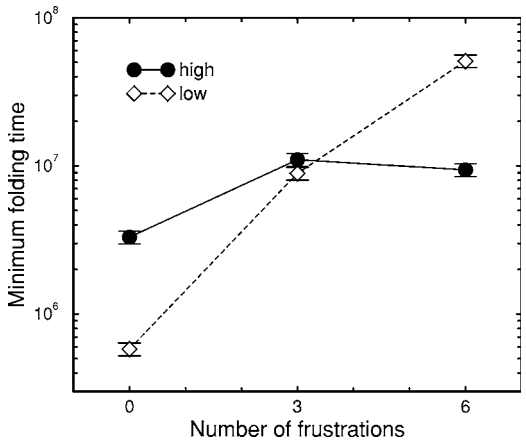


FIG. 6. Minimum folding time of each curve of Fig. 5 as a function of frustration $f$ and hydrophobicity. $f=0$, 3, and 6 correspond to native, mutation 1, and mutation 2 sequences, respectively.

ing at least 100 independent runs to reach the native state. Twelve values for $\tau$ were calculated in the range of temperature $1.2<T<3.0$. The results are shown in Fig. 5, where the continuous lines are only a visual guide. Since there is a scaling factor between the temperatures and the energy parameters, the temperature for each sequence and regime of hydrophobicity is normalized by its respective folding transition temperature, $T/T_f$. By doing this, it is possible to compare their kinetics in the same range of temperature, which means that their folding temperature occurs at the same relative value $T/T_f=1$. Consistent with previous studies,[25,26] folding mean first passage time, or simply folding time, as a function of temperature has a U-shaped dependence for all sequences and regime of hydrophobicity. $\tau$ increases as frustration increases in a very obvious way at LH regime. At HH regime, $\tau$ remains approximately constant. For the native sequence, the kinetics at LH is much faster than for the HH. For a mutation 1 sequence ($f=3$), the kinetics in HH and LH regimes have similar rates. For mutation 2 ($f=6$) the kinetics at the LH limit is slower than in the HH limit, and an inversion is seen from Fig. 5(a)–5(c). The results are summarized in Fig. 6, where the minimum folding time $\tau_{\min}$ in Fig. 5 is shown as a function of frustration and hydrophobicity. Minimum folding time is a good overall parameter for the kinetics. While $\tau_{\min}$ is approximately constant at HH, at LH $\tau_{\min}$ varies by about two orders of magnitude.

## VI. DISCUSSION

With regard to the folding criteria, the value of the folding temperature is nearly independent of the criteria when the sequence is optimally designed or a small amount of frustration is present in the native state. In the present model, this means that less than 10% of all native contacts are unfavorable. When this amount is increased, the low hydrophobic limit indicates a strong dependence on the chosen folding criteria, whereas the high hydrophobic limit is less sensitive to the choice of the criterion. In the HH limit, characterized by an overall attraction between monomers, folding is preceded by a nonspecific collapse, which is absent at the other limit. This result is in agreement with previous studies of Gutin *et al.*[16] As for the kinetic aspect, it is crucial to study two regimes of stability: the higher stability regime for the native sequence and the lower stability regime for the mu-

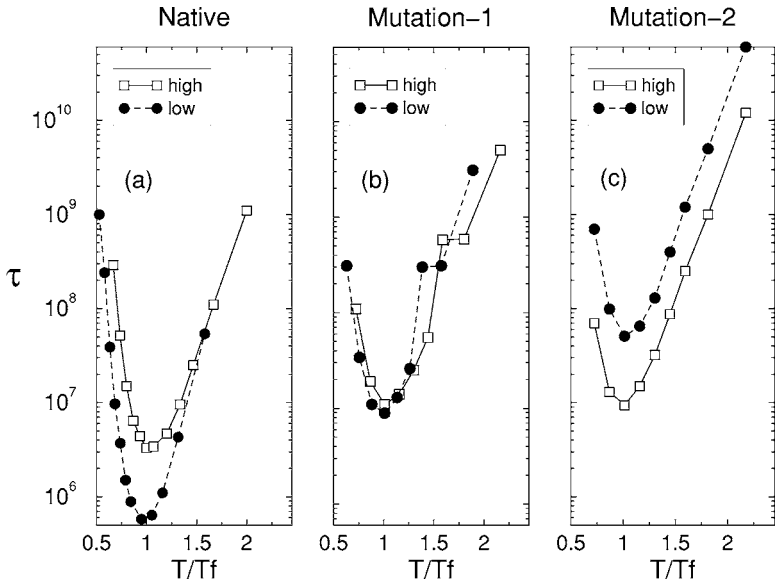


FIG. 5. Folding times for the native (a), mutation 1 (b), and mutation 2 (c) sequences as a function of temperature at the two limits of hydrophobicity. For a better comparison, the temperature for each sequence and regime of hydrophobicity is normalized by its respective folding transition temperature, $T/T_f$.

tated sequences. The criterion of Thirumalai and Klimov[11] states that the chain folds faster when the parameter $\sigma=(T_\theta - T_f)/T_\theta$ is relatively small which is consistent with $T_\theta \cong T_f$ or the absence of a nonspecific collapse. In the absence of frustration in the native state, i.e., strongly optimized sequence and high stability, the present results are also in agreement with the above criterion. On the other hand, for a lower stability regime, due to a certain amount of frustration in the native state of the protein, which in this model means that 20% of all native contacts are unfavorable, the simulations in Fig. 5 indicate faster folding rates for an overall attraction between monomers (HH limit). For this lower stability regime, our results are in agreement with the theoretical results of Chiu and Goldstein[8] who studied the kinetic behavior of marginally stable proteins. They have demonstrated that, for marginal protein stability, compaction, induced by nonspecific interactions, leads to the increase of the folding rates. The results of Figs. 5 and 6 also suggest that, when stability is reduced, the stronger attractive interaction responsible for the chain collapse produces faster folding rates.

Our results suggest that optimally designed sequences, characterized by high stability, will fold faster at the LH limit characterized by the absence of a nonspecific collapse. For less optimized sequences and lower stability, the HH limit, characterized by the presence of a nonspecific collapse preceding folding, provides faster folding rates. The first case correlates with previous results of Thirumalai and Klimov,[11] and the second correlates with the results of Chiu and Goldstein.[8]

The nonspecific collapse of the chain may be a way of overcoming frustration existing in the native state and leads us to speculate on issues regarding the evolution of proteins. An evolutionary process that would make sequences optimally designed would remove the nonspecific collapse that precedes folding. On the other hand, if evolutionary pressures did not result in strongly optimized sequences, Figs. 4 and 5 suggest a scenario where nonspecific collapse precedes folding in order to make the folding process faster. Collapse would help folding when some degree of frustration is present in the native state. If the time for optimizing sequences in the huge space of sequences ($20^N$ for $N$ amino acids) would be prohibitively large, especially for large proteins, the resulting sequences, not strongly optimized, would first collapse and then fold in a biologically relevant time; thus, the necessity to further optimize the sequence would no longer exist. These sequences could be the result of the least effort in the evolutionary process of choosing polypeptide chains that could fold in some functional structures. Particularly, for large proteins, the space of sequences would be so vast that it would make strongly optimized sequences improbable. This would imply that large proteins are likely to collapse before folding into their native state. Among all the large proteins whose sequences are not strongly optimized, those that collapse before folding would predominate as suggested by Fig. 5(c). Also, at the HH limit, the folding process is less sensitive to mutations that worsen the sequence design. If we look at the continuous line in Figs. 5 (HH) we see that the increase in folding time is moderate from Fig. 5(a) and 5(b) and is nearly absent from Fig. 5(b) and 5(c). At the opposite limit, the LH one, represented by the dotted lines, a continuous increase in the folding times is observed, suggesting a strong dependence on mutations. In addition, sequences that experience a nonspecific collapse before folding would prevail after mutations, not only due to the faster kinetics but also because the folding temperature $T_f$ is less sensitive to mutation.

In short, sequences that collapse concomitantly with the formation of their native structure were strongly optimized during evolution and have the fastest rates. Among the sequences which were not strongly optimized by evolution (probably large proteins), those which collapse before folding have faster rates and would be predominant. Thus, the chain collapse would be the result of three features: (i) Sequences not strongly optimized imply an amount of frustration in the native structure, which, in the terms of the present study, means that 20% (or more) of all the native contacts are unfavorable; (ii) for this type of sequence, the HH limit (collapse) provides faster folding rates; (iii) protein evolution seeks optimal biological relevant time rather than the maximal folding rates. By this last statement we mean that collapse could be the result of an evolutionary process that was sufficient for the protein to achieve its biological functions and necessary stability. In this scenario, as far as foldability and stability are concerned, there is no need for further sequence optimization that would remove collapse from the chain folding.

The present results contribute to the debate on the intriguing subject of hydrophobic collapse. Of course, the statements regarding protein evolution should be considered more as plausible explanations than conclusions, which cannot be drawn from this simple study.

## ACKNOWLEDGMENTS

The authors are grateful to Dr. José Nelson Onuchic for helpful discussions. Two of the authors (L.C.O. and R.T.H.S.) were supported by CAPES, Brazil. Two of the authors (V.B.P.L. and J.C.) were partially supported by the Brazilian agency CNPq. One of the authors (V.B.P.L.) was supported by FAPESP, Brazil.

[1] A. Frauenfelder, J. Deisenhofer, and P. Wolynes, *Simplicity and Complexity in Proteins and Nucleic Acids* (Dahlem University Press, Berlin, 1999).
[2] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins: Struct., Funct., Genet. **21**, 167 (1995).
[3] D. Rennell, S. E. Bouvier, L. W. Hardy, and A. R. Poteete, J. Mol. Biol. **222**, 67 (1991).
[4] J. Sondek and D. Shortle, Proteins **13**, 132 (1992).
[5] L. H. Weaver, M. G. Grütter, S. J. Remington, T. M. Gray, N. W. Isaacs, and B. W. Matthews, J. Mol. Evol. **21**, 97 (1985).
[6] A. Sali, E. Shakhnovich, and M. Karplus, J. Mol. Biol. **235**, 1614 (1994).
[7] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).
[8] T. L. Chiu and R. A. Goldstein, J. Chem. Phys. **107**, 4408 (1997).
[9] I. S. Millet, L. E. Townsley, F. Chiti, S. Doniach, and K. W. Plaxo, Biochemistry **41**, 321 (2002).
[10] D. K. Klimov and D. Thirumalai, Phys. Rev. Lett. **76**, 4070 (1996).
[11] D. Thirumalai and D. K. Klimov, Curr. Opin. Struct. Biol. **9**, 197 (1999); C. J. Camacho and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **90**, 6369 (1993).
[12] T. Kiefhaber, Proc. Natl. Acad. Sci. U.S.A. **92**, 9029 (1995).

[13] M. Mucke and F. X. Schmid, J. Mol. Biol. **239**, 713 (1994).
[14] S. Khorasanizadeh, I. D. Peters, T. R. Butt, and H. Roder, Biochemistry **32**, 7054 (1993).
[15] K. W. Plaxo, I. S. Millet, D. J. Segel, S. Doniach, and D. Baker, Nat. Struct. Biol. **6**, 554 (1999).
[16] A. M. Gutin, V. I. Abkevich, and E. I. Shakhnovich, Biochemistry **34**, 3066 (1995).
[17] B. Gillespie and K. W. Plaxo, Proc. Natl. Acad. Sci. U.S.A. **96**, 12014 (2000).
[18] P. E. Leopold, M. Nontal, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **89**, 8721 (1992).
[19] A. Li, R. Helling, C. Tang, and N. Wingreen, Science **273**, 666 (1996).
[20] N. D. Socci and J. N. Onuchic, J. Chem. Phys. **101**, 1519 (1994).
[21] E. Shakhnovich and A. Gutin, J. Chem. Phys. **93**, 5967 (1990).
[22] R. I. Dima, J. R. Banavar, M. Cieplak, and A. Maritan, Proc. Natl. Acad. Sci. U.S.A. **96**, 4904 (1999).
[23] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988).
[24] J. Chahine, H. Nymeyer, V. Leite, N. Socci, and J. N. Onuchic, Phys. Rev. Lett. **88**, 168101 (2002).
[25] V. B. P. Leite, J. N. Onuchic, G. Stell, and J. Wang, Biophys. J. **87**, 3633 (2004).
[26] N. D. Socci and J. N. Onuchic, J. Chem. Phys. **103**, 4732 (1995).