Thermodynamics and kinetics of a Gō proteinlike heteropolymer model with two-state folding characteristics

Anna Kallias[a)]
Institut für Theoretische Physik, Universität Leipzig, Postfach 100 920, D-04009 Leipzig, Germany and Centre for Theoretical Sciences (NTZ), Universität Leipzig, Postfach 100 920, D-04009 Leipzig, Germany

Michael Bachmann[b)]
Institut für Theoretische Physik, Universität Leipzig, Postfach 100 920, D-04009 Leipzig, Germany, Centre for Theoretical Sciences (NTZ), Universität Leipzig, Postfach 100 920, D-04009 Leipzig, Germany, and Computational Biology and Biological Physics Group, Department of Theoretical Physics, Lund University, Sölvegatan 14A, SE-223 62 Lund, Sweden

Wolfhard Janke[c)]
Institut für Theoretische Physik, Universität Leipzig, Postfach 100 920, D-04009 Leipzig, Germany and Centre for Theoretical Sciences (NTZ), Universität Leipzig, Postfach 100 920, D-04009 Leipzig, Germany

(Received 14 September 2007; accepted 14 November 2007; published online 7 February 2008)

We present results of Monte Carlo computer simulations of a coarse-grained hydrophobic-polar Gō-like heteropolymer model and discuss thermodynamic properties and kinetics of an exemplified heteropolymer, exhibiting two-state folding behavior. It turns out that general, characteristic folding features of realistic proteins with a single free-energy barrier can also be observed in this simplified model, where the folding transition is primarily driven by the hydrophobic force. © *2008 American Institute of Physics*. [DOI: 10.1063/1.2822287]

I. INTRODUCTION

Spontaneous protein folding is a dynamic process, which starts after the generation of the DNA-encoded amino acid sequence in the ribosome and is in many cases finished when the functional conformation, the native fold, is formed. As this process takes microseconds to seconds, a dynamical computational analysis of an appropriate microscopic model, which could lead to a better understanding of the conformational transitions accompanying folding[1] is extremely demanding. Since protein folding is a thermodynamic process at finite temperature, a certain folding trajectory in the free-energy landscape is influenced by Brownian collisions with surrounding solvent molecules. Therefore, it is more favorable to study the kinetics of the folding process by averaging over an appropriate ensemble of trajectories.

A significant problem is that the complexity of detailed semiclassical microscopic models based on force fields and solvent parameter sets (or explicit solvent) rules out molecular dynamics (MD) in many cases and, therefore, Markovian Monte Carlo (MC) dynamics is a frequently used method for such kinetic studies. It is obvious, however, that the time scale provided by MC is not directly comparable with the time scale of the folding process. It is widely believed that the folding path of a protein is strongly correlated with contact ordering,[2] i.e., the order of the successive contact formation between residues and, therefore, long-range correlations and memory effects can significantly influence the kinetics.

A few years ago, experimental evidence was found that classes of proteins show particularly simple folding characteristics: single exponential and two-state folding.[3,4] In the two-state folding process, which is in the focus of the present study, the peptide is either in an unfolded, denatured state or it possesses a nativelike, folded structure. In contrast to the barrier-free single exponential folding, there exists an unstable transition state to be passed in the two-state folding process. Due to the comparatively simple folding characteristics, strongly simplified, effective models were established. Knowledge-based models of Gō-like type[5–8] were investigated in numerous recent studies.[9–18] In Gō-like models the native fold must be known and is taken as input for the energy function. The energy of an actual conformation depends on its structural deviation from the native fold (e.g., by counting the number of already established native contacts). By definition, the energy is minimal, if conformation and native fold are identical in all degrees of freedom involved in the model. The simplicity of the model entails reduced computational complexity and MD simulations, e.g., based on Langevin dynamics,[16] can also be successfully performed.

In this paper, we follow a different approach. We also study a Gō-like model, but it is based on a minimalistic coarse-grained hydrophobic-polar representation of the heteropolymer. The basic idea behind coarse graining is the introduction of a mesoscopic length scale, i.e., the reduction of microscopic details, in order to classify heteropolymers with respect to their folding characteristics. The assumption is that, if there is some sort of universality in folding processes, then an effective model should allow for a general description of the qualitative folding behavior. In fact, in a recent work we could show that two-state folding, folding through

[a)]Present address: Volkswagen AG, Postfach 1778, D-38436 Wolfsburg, Germany. Electronic mail: anna.kallias@itp.uni-leipzig.de.
[b)]Electronic mail: michael.bachmann@itp.uni-leipzig.de.
[c)]Electronic mail: wolfhard.janke@itp.uni-leipzig.de. Homepage: http://www.physik.uni-leipzig.de/CQT.html.

 © 2008 American Institute of Physics

Downloaded 18 Feb 2008 to 139.18.9.115. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/jcp/copyright.jsp

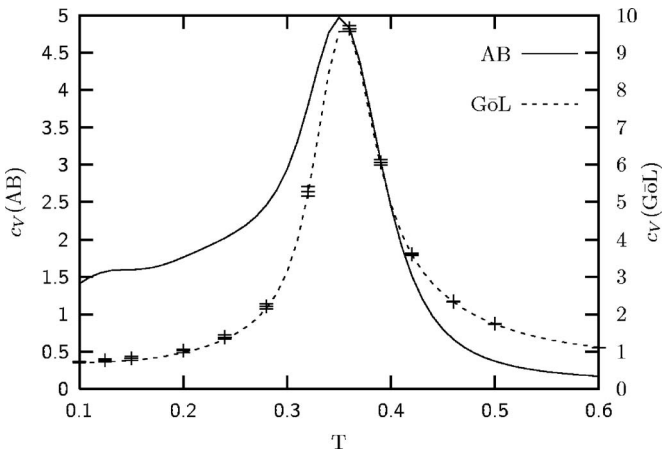

FIG. 1. Specific heats for the peptide with sequence $\mathcal{S}$ as obtained with the AB model (Ref. 28) and the gauged GōL model as defined in Eq. (3).

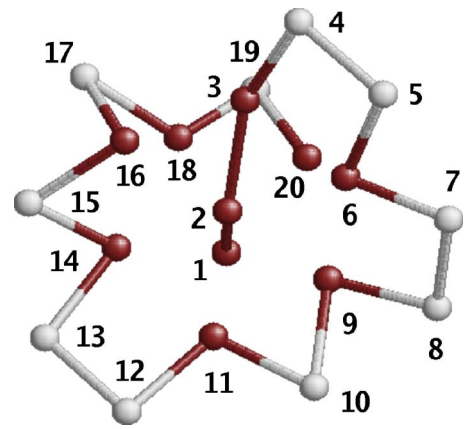

FIG. 2. (Color online) Global-energy minimum of sequence $\mathcal{S}$ in the AB model with $E_{\mathrm{AB}} \approx -19.3$. Dark monomers are hydrophobic (A) and light residues polar (B).

intermediates, and metastability are inherent tertiary folding processes which can also be found in folding studies[19,20] of a simple hydrophobic-polar heteropolymer model.[21] In this context, it is quite interesting that secondary structures are also intrinsic geometries of polymerlike objects, even on a mesoscopic scale.[23–26]

II. MODEL AND METHODS

In the following, we consider the hydrophobic-polar heteropolymer sequence $\mathcal{S}=A_3B_2AB_2ABAB_2ABABABA$, where the A's indicate hydrophobic monomers and the B's polar (hydrophilic) residues. For our comparative model study, we employ the physically motivated AB model[21,22] and a knowledge-based model of Gō type,[5,6,11] which is referred to as the Gō-like (GōL) model throughout the paper.

In the AB model, the energy of a conformation $\mathbf{R}=\{\mathbf{r}_1,\ldots,\mathbf{r}_N\}$ with N monomers, where $\mathbf{r}_i$ denotes the spatial location of the ith monomer, is given by

$$E_{\mathrm{AB}}(\mathbf{R}) = \frac{1}{4}\sum_{k=1}^{N-2}(1-\cos\vartheta_k) + 4\sum_{i=1}^{N-2}\sum_{j=i+2}^{N}\left(\frac{1}{r_{ij}^{12}} - \frac{C(\sigma_i,\sigma_j)}{r_{ij}^{6}}\right), \tag{1}$$

where the first term is the bending energy and the sum runs over the $(N-2)$ bending angles ϑ_k $(0 \leqslant \vartheta_k \leqslant \pi)$ between monomers k, $k+1$, and $k+2$ of successive covalent bond vectors which have length unity. The Lennard-Jones-type potentials depend on the types of the interacting monomers ($\sigma_i = A,B$) and on their spatial distance r_{ij} and represent the influence of the specific AB sequence on the energy. The long-range behavior is attractive for pairs of like monomers and repulsive for AB pairs of monomers,

$$C(\sigma_i,\sigma_j) = \begin{cases} +1, & \sigma_i,\sigma_j = A \\ +1/2, & \sigma_i,\sigma_j = B \\ -1/2, & \sigma_i \neq \sigma_j. \end{cases} \tag{2}$$

In this model, the AB heteropolymer with sequence $\mathcal{S}$ experiences a hydrophobic collapse transition which is signalized by the peak in the specific-heat curve plotted in Fig. 1. The transition between random coils and nativelike hydrophobic-core conformations is a (pseudo)phase separation process and the folding transition of two-state (folded/unfolded) type. In particular, for such systems it is known from model studies of realistic amino acid sequences that knowledge-based Gō-type models reveal kinetic aspects of folding and unfolding processes reasonably well.[11,16]

To perform kinetic studies of $\mathcal{S}$ in the simplified hydrophobic-polar approach as well, we use the (putative) global-energy minimum, identified in energy-landscape paving (ELP) minimizations[27] of the AB model,[28] as input for the definition of a hydrophobic-polar GōL model. The (putative) native conformation $\mathbf{R}^{(0)}$ is shown in Fig. 2 and its energy is $E_{\mathrm{AB}} \approx -19.3$ in the units of the AB model [Eq. (1)]. In GōL models, the "energy" of a given conformation is related to its similarity with the ground state. This means that "energy" in the GōL model plays rather the role of a similarity or "order" parameter and is, therefore, not a potential energy in the usual physical sense (as there is no physical force associated with it). Denoting the $(N-2)$ bending angles of the global-energy minimum conformation by $\vartheta_k^{(0)}$, its $(N-3)$ torsional angles as $\varphi_l^{(0)}$, and the intermonomer distances by $r_{ij}^{(0)}$, we define the GōL model according to the representation in Ref. 11 as

$$\begin{aligned} E_{\mathrm{G\sigma L}}(\mathbf{R})/\varepsilon = {} & K_\vartheta \sum_{k=1}^{N-2}(\vartheta_k - \vartheta_k^{(0)})^2 \\ & + \sum_{n=1,3}\sum_{l=1}^{N-3} K_\varphi^{(n)}\{1-\cos(n[\varphi_l - \varphi_l^{(0)}])\} \\ & + \sum_{i<j-1}^{\text{native}}\left(5\left[\frac{r_{ij}^{(0)}}{r_{ij}}\right]^{12} - 6\left[\frac{r_{ij}^{(0)}}{r_{ij}}\right]^{10}\right) \\ & + \sum_{i<j-1}^{\text{non-native}} \frac{1}{r_{ij}^{12}}. \end{aligned} \tag{3}$$

The last two sums run over all pairs of nonbonded monomers. In the case the pair (i,j) forms a native contact, i.e., $r_{ij} < r_{\mathrm{cut}}$ and $r_{ij}^{(0)} < r_{\mathrm{cut}}$, the monomers experience a short-range 10-12 Lennard-Jones attraction, while for non-native

Downloaded 18 Feb 2008 to 139.18.9.115. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/jcp/copyright.jsp

contacts an overall repulsive $1/r^{12}$ contribution is taken into account. The constants $K_\vartheta=20$, $K_\varphi^{(1)}=0.5$, and $K_\varphi^{(3)}=0.25$ weigh the relative strengths of the angular energy contributions. The values were adjusted to have a reasonable coincidence of the peak temperature of the specific heat compared to the results obtained with the AB model[28] (see Fig. 1). The free overall energy scale ε was chosen such that $E_{\mathrm{AB}}(\mathbf{R}^{(0)})=E_{\mathrm{G\sigma L}}(\mathbf{R}^{(0)})=-\varepsilon n_{\mathrm{tot}}$, where n_{tot} is the total number of native contacts. The definition of a native contact requires the introduction of a cutoff radius, which we chose as $r_{\mathrm{cut}}=1.14$. This entails $n_{\mathrm{tot}}=20$ for the native conformation $\mathbf{R}^{(0)}$, and therefore $\varepsilon\approx 0.966$. The results are not very sensitive on the choice of the cutoff radius, provided it is not too small.

For the thermodynamic analyses, we performed parallel tempering simulations[29–31] and for the kinetic studies, a standard implementation of the Metropolis Monte Carlo method was used. Chain updates were performed employing the spherical-cap updates as described and parametrized in Ref. 28. A sweep consists of $(N-1)$ sequential bond vector update trials.

III. THERMODYNAMICS

For the study of the folding transition, the introduction of an effective parameter which uniquely describes the macrostate of the ensemble of heteropolymer conformations is useful. A widely used measure is the contact number $q(\mathbf{R})$ which is for a given conformation $\mathbf{R}$ simply defined as the fraction of the already formed native contacts $n(\mathbf{R})$ in conformation $\mathbf{R}$ and the total number of native contacts n_{tot} in the final fold, i.e., $q(\mathbf{R})=n(\mathbf{R})/n_{\mathrm{tot}}$. Then, the statistical ensemble average of this quantity $\langle q(\mathbf{R})\rangle$ at a given temperature characterizes its macrostate. Roughly, for a two-state folder, if $\langle q(\mathbf{R})\rangle>0.5$, nativelike conformations dominate the statistical ensemble. If less than half the total number of contacts is formed, the heteropolymer tends to reside in the pseudophase of denatured conformations. Note that folding transitions are not sharp phase transitions in the thermodynamic sense as the heteropolymer sequence is of finite length and cannot be extended.

Another suitable parameter is the angular overlap parameter Q (Ref. 28) which has been proven to be extremely useful in the characterization of heteropolymer folding channels in the AB model.[19,20] If all virtual bond and torsion angles of two conformations, to be compared with each other, coincide, it is unity and between 0 and 1 otherwise. The advantage is that it is particularly useful for classifying intermediate or metastable structures with stable, but nonnative contacts. In Fig. 3, the fluctuations of both parameters, i.e., $d\langle q\rangle/dT$ and $d\langle Q\rangle/dT$, respectively, are plotted as functions of the temperature for the GōL model of sequence $\mathcal{S}$. We clearly see that the temperature region of conformational activity as signalized by these two "order" parameters coincides with the thermally active region indicated by the peak of the specific heat, which is also shown for comparison. The folding temperature, i.e., the temperature of maximum activity, is $T_f\approx 0.36$.

The classification of the heteropolymer with sequence $\mathcal{S}$ as two-state folder arises from the analysis of the free-energy

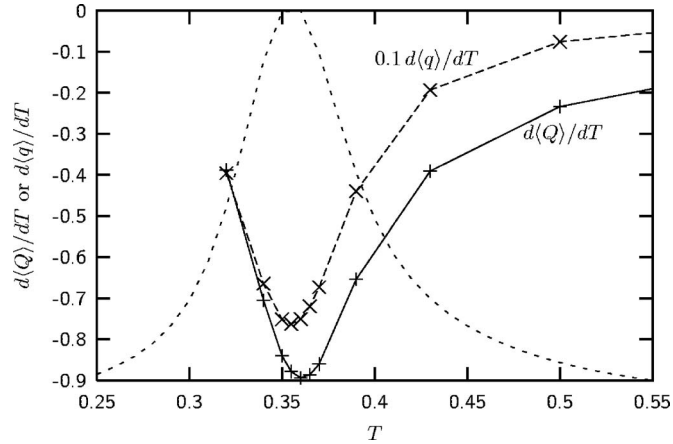

FIG. 3. Fluctuations of the similarity parameter q and the angular overlap parameter Q as functions of the temperature. For comparison, the specific heat (dotted line) is also plotted in this figure.

landscape. We assume that q is a suitable parameter that describes the macrostate of the system adequately. Considering this parameter as a constraint, we can formally average out the conformational degrees of freedom and the probability for a conformation in a macrostate with contact parameter q' reads

$$p(q')=\langle\delta(q'-q(\mathbf{R}))\rangle=\frac{1}{Z}\int\mathcal{D}\mathbf{R}\,\delta(q'-q(\mathbf{R}))e^{-E_{\mathrm{G\sigma L}}(\mathbf{R})/k_BT}, \tag{4}$$

where Z denotes the unrestricted partition function. The integral measure is simply $\mathcal{D}\mathbf{R}=\Pi_{i=1}^{N}[d^3r_i]\Pi_{i=1}^{N-1}[\delta(r_{ii+1}-1)]$. Expression (4) can be used to define a free energy as function of the q parameter by

$$F(q)=-k_BT\ln p(q). \tag{5}$$

Since the value of q is a qualitative measure for the macrostate the system resides in, the minimum of the free energy at a given temperature T is related to the dominant macrostate in the canonical ensemble at this temperature. Actually, as can be read off from Fig. 4, the folding transition of the heteropolymer with sequence $\mathcal{S}$ is a phase-separation

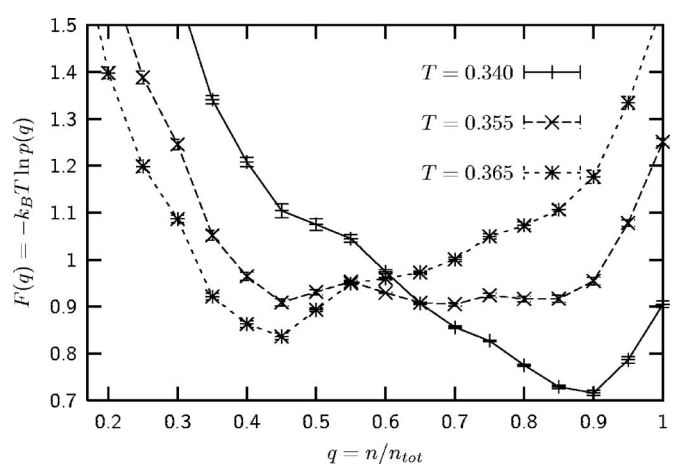

FIG. 4. Free-energy landscape $F(q)$ of $\mathcal{S}$ for different temperatures close to the transition point.

Downloaded 18 Feb 2008 to 139.18.9.115. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/jcp/copyright.jsp

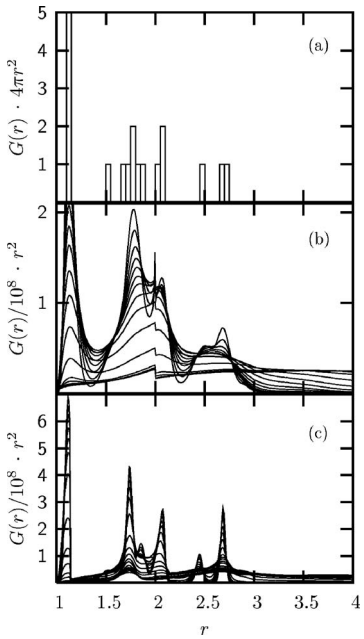

FIG. 5. Radial distribution functions of $\mathcal{S}$ (a) for the lowest-energy conformation (see Fig. 2), (b) at 16 temperatures in the interval $T\in[0.1,1.5]$ (curves from top to bottom) employing the AB model, and (c) at 18 temperatures in the interval $T\in[0.1,1.5]$ for the GōL model.

process, i.e., at the transition point close to $T_f\approx 0.36$, the folded and the denatured pseudophase coexist and the transition state barrier possesses a local maximum close to $q\approx 0.5$, as expected for typical two-state folding characteristics.

A useful measure of structure formation is the radial distribution function and its dependence on temperature. Due to translational invariance in the three-dimensional space, three degrees of freedom contribute only a volume factor to the partition function. Therefore, we utilize this by fixing the position of the first monomer, $\mathbf{r}_1=\mathbf{0}$. Thus, we measure radial distances r of the other monomers with respect to the first one and we define the radial distribution function as

$$G(r)=\frac{1}{4\pi r^2}\left\langle \sum_{i=3}^{N}\delta(|\mathbf{r}_i|-r)\right\rangle. \tag{6}$$

Note that we have excluded $i=2$ as, by definition, the virtual covalent bonds are rigid and, therefore, $r_{12}=1$. Actually, from our definition, $4\pi\int_0^\infty dr r^2 G(r)=N-2$. In Fig. 5, the radial distribution functions of $\mathcal{S}$ are shown for (a) the lowest-energy conformation and the ensembles at different temperatures in the (b) AB model and (c) GōL model. Although the global-energy minimum conformation shown in Fig. 2 obviously does not form a regular crystal structure, the Lennard-Jones-type interactions in combination with the rigid virtual bonds induce preferable substructures, favoring, e.g., bond angles of 60°, 90°, and 120°.[20] Therefore, the peaks of the radial distances from the first monomer for the global-energy minimum conformation in Fig. 5(a) can be partly explained by these local segments. The first peak at $r=2^{1/6}\approx 1.12$ is related to the minimum potential distance between two nonbonded A monomers (and the reference monomer at $\mathbf{r}_1=0$ is of type A). Also the location of other peaks can be deduced from Fig. 2 by similar geometric arguments.

The temperature dependence of the peak evolution is shown for the AB model in Fig. 5(b) and for the GōL model in Fig. 5(c). As a first result, we see that there is nice coincidence not only in the location of the peaks, but also in the fact that there is no indication of intermediary, weakly stable conformations. Actually, the folding characteristics is similar in both models, at least from the thermodynamic point of view. At high temperatures, random-coil conformations dominate: There is no significant structuring in the radial distribution function. Passing the folding transition temperature, local, planar structures form first, before the tertiary, three-dimensional ordering towards the native conformation occurs.

IV. KINETICS

The advantage of the simple GōL model for two-state folding is that it likewise enables kinetic studies of folding *and* unfolding events. In fact, this is the main purpose of this knowledge-based model, because kinetics studies of physically motivated models are typically computationally extremely demanding. This is hence unfortunately also the case for folding studies employing the AB model. A striking example is shown in Fig. 6 where for a folding event the gray scale coded average probability of native-contact formation is plotted for the original AB model [Fig. 6(a)] and the GōL model [Fig. 6(b)]. The average was taken over a "time" window $[t_{\mathrm{MC}}-\Delta t, t_{\mathrm{MC}}+\Delta t]$ with $\Delta t=1000$ MC steps. The darker a bar, the higher the probability that the associated contact is formed. As the simulation was carried out in the folding regime, the probability of native-bond formation increases with the number of total MC steps. It is, however, obvious that the folding is much slower using the AB model, whereas in the GōL model most of the native contacts have already been formed after 80 000 MC steps. The native conformation of the considered sequence partly possesses kind of zigzag structures or "turns" (see Fig. 2). The folding of these segments is particularly simple and the probability of the formation of the associated native contacts of monomers i with monomers $i+2$ or $i+3$ increases much faster than for the other contacts. But even in this case, the GōL kinetics is unbeatable.

This example exhibits the dilemma of physics-based models in studying kinetic aspects of structure formation by means of computer simulations. It is a notoriously difficult problem, because the time scale of molecular dynamics is typically too small to see folding events, but also Markov chain Monte Carlo dynamics of physical models is typically too slow. Just for kinetic aspects, where absolute time scales are widely irrelevant, knowledge-based GōL models are an alternative that allow for an increase in the sampling efficiency, at least for systems with simple folding characteristics (as, for example, two-state folding). Actually, for such models an adequate statistical sampling of ensembles close to the transition state can be achieved by sophisticated Monte Carlo methods where efficiency is typically gained by simulating generalized ensembles at the expense of an artificial

Downloaded 18 Feb 2008 to 139.18.9.115. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/jcp/copyright.jsp

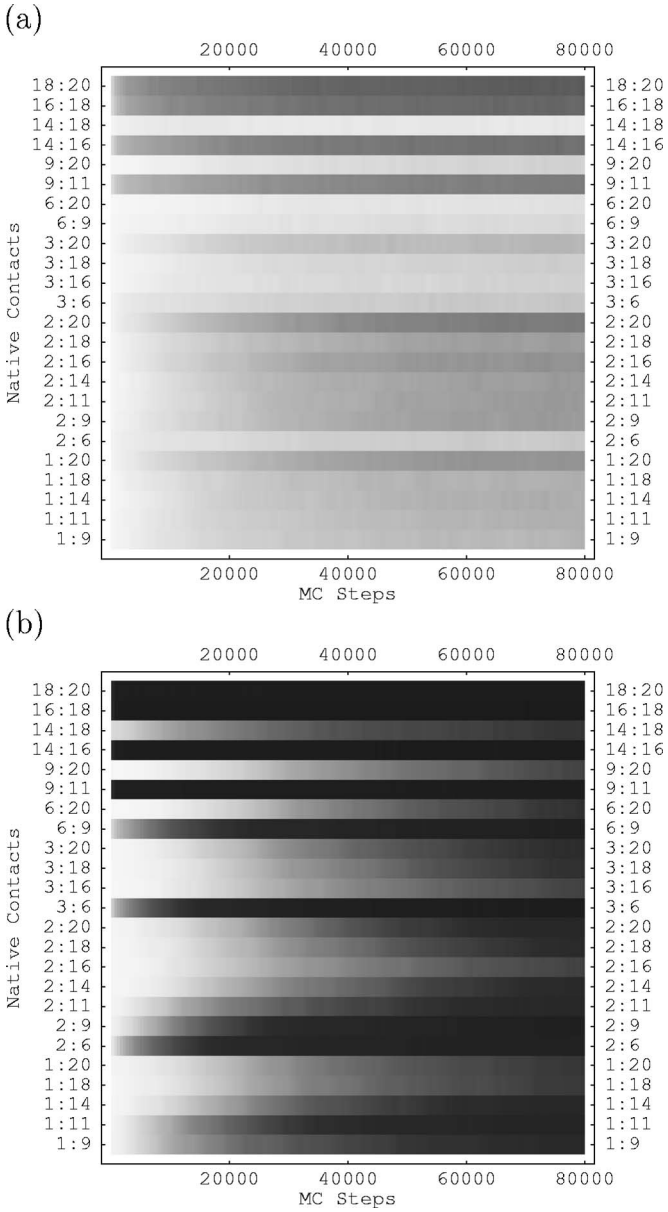

FIG. 6. Gray scale coded averages of the probabilities for native-contact formation of sequence $\mathcal{S}$ as a function of Monte Carlo time for the (a) AB and the (b) GōL model, averaged over 1000 folding events. The probability is calculated as a combined temporal average over 2000 MC steps in the ensemble of the 1000 folding events. The darker the bars are, the higher is the probability that the associated contact has formed. The labels of the pair contacts refer to the numbering in the native conformation shown in Fig. 2.

dynamics. Therefore, it is widely believed that free-energy-driven dynamics, as it is relevant for protein folding, can reasonably be provided only by Boltzmann-Markov chains of conformational updates. One possibility to achieve this is the application of a conventional Metropolis Monte Carlo method. Although the overall time scale is left open, this method allows for comparative folding (unfolding) studies at various temperatures. The kinetics of the folding (unfolding) transitions is thereby obtained by averaging over sufficiently many folding (unfolding) trajectories.

In Fig. 7, snapshots of a single GōL folding event of the heteropolymer with sequence $\mathcal{S}$ at $T=0.3$ are shown at different times t_{MC}. The gray scale of the monomers encodes the variance of the monomer positions, $\sigma^2_{\mathbf{r}_i}=\overline{\mathbf{r}_i^2}-\overline{\mathbf{r}_i}^2$, averaged over the MC time interval $[t_{\mathrm{MC}}-1000, t_{\mathrm{MC}}+1000]$. The first monomer is fixed and, therefore, not moving. The higher the mobility of a monomer in this time interval, the brighter is its color. For $\sigma^2_{\mathbf{r}_i}<0.1$, the monomer is rendered in black, and for $\sigma^2_{\mathbf{r}_i}>1$ in white. The gray scales are linearly interpolated in between these boundaries. Although there are periods of relaxation and local unfolding, a stable intermediate conformation is not present and the folding process is a relatively "smooth" process. This is also confirmed by the more quantitative analysis of the same folding event in Fig. 8, where the temporal averages of the similarity parameters q and Q and of the energy E are shown. Since the temperature lies sufficiently far below the folding temperature ($T_f\approx 0.36$), the free-energy landscape does not exhibit substantial barriers which hinder the folding process.

Nonetheless, the chevron plot shown in Fig. 9 exhibits a rollover which means that the folding characteristics is not perfectly of two-state type, in which case the folding (unfolding) branches would be almost linear.[16] In this plot, the temperature dependence of the mean first passage time τ_{MFP} is presented. We define τ_{MFP} as the average number of MC steps that are necessary to form at least 13 native contacts in the folding simulations, starting from a random conformation. In the unfolding simulations, we start from the native state and τ_{MFP} is the number of MC steps required to reach a conformation with less than seven native contacts, i.e., 13 native contacts are broken. In all simulations performed at different temperatures, τ_{MFP} is averaged from the first passage times of a few hundred respective folding and unfolding trajectories. Assuming a linear dependence at least in the transition state region, τ_{MFP} is directly related to exponential folding and unfolding rates $k_{f,u}\approx 1/\tau^{f,u}_{\mathrm{MFP}}\sim\exp(-\varepsilon_{f,u}/k_BT)$, respectively, where the constants $\varepsilon_{f,u}$ determine the kinetic folding (unfolding) propensities. The dashed lines in Fig. 9 are tangents to the logarithmic folding and unfolding curves at the transition state. The slopes are the folding (unfolding) propensities and have in our case values of $\varepsilon_f\approx -1.32$ and $\varepsilon_u\approx 5.0$.

In this variant of the chevron plot, which is similar to the presentations discussed in Refs. 16 and 32, the temperature T mimics the effect of the denaturant concentration that is in experimental studies the more generic external control parameter. The hypothetic intersection point of the folding and unfolding branches defines the transition state. The transition state temperature estimated from this analysis coincides very nicely with the folding temperature $T_f\approx 0.36$ as identified in our discussion of the thermodynamic properties of the system. This result also demonstrates that the description of the folding and unfolding transitions from the kinetic point of view is not only qualitatively but even quantitatively consistent with the thermodynamic results from the canonical-ensemble analysis.

V. MESOSCOPIC HETEROPOLYMERS VERSUS REAL PROTEINS

It is not the purpose of this manuscript to perform a coarse-grained analysis of the two-state folding characteris-

Downloaded 18 Feb 2008 to 139.18.9.115. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/jcp/copyright.jsp

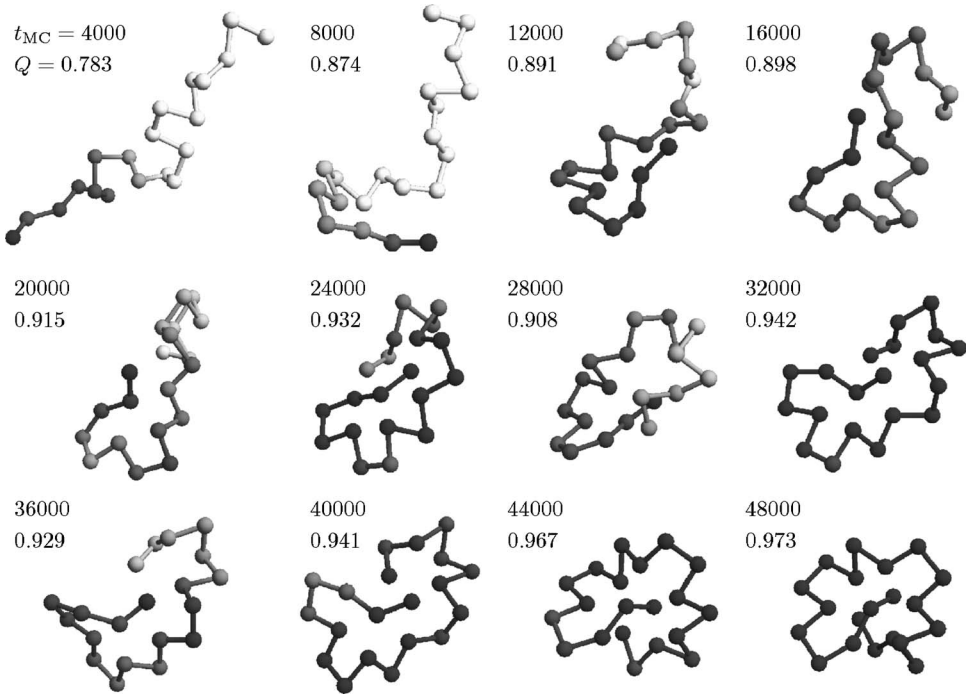

FIG. 7. Single folding event for the heteropolymer with sequence $\mathcal{S}$ in the GōL model at $T=0.3<T_f\approx 0.36$. The description of the gray scale code is given in the text.

tics of a specific, real protein. Rather, we have shown that—although also only exemplified for a single sequence—it is actually useful to study thermodynamic and kinetic properties of mesoscopic peptide models without introducing atomic details. The main focus of such models is pointed towards general features of the folding transition (measured in terms of "order" parameters being specific for the corresponding transition, such as, e.g., the contact and overlap parameters investigated in our present study) that are common to a number of proteins behaving qualitatively similarly. It is then furthermore assumed that these proteins can be grouped into classes of certain folding characteristics. The sequence $\mathcal{S}$ used in our paper is not obtained from a one-to-one hydrophobic-polar transcription of a real amino acid sequence. We do not think that such a mapping is particularly useful. Rather, $\mathcal{S}$ is considered as a representative that exhibits two-state folding characteristics in the coarse-grained model considered here. This implies, in general, that the classification of peptide folding behaviors is not necessarily connected with detailed atomic correlations and particular contact ordering. It is rather an intrinsic property of proteinlike heteropolymers and can thus already be discovered employing models on mesoscopic scales.[19,20]

Several proteins are known to be two-state folders and their folding transitions exhibit the features we have also seen in our present coarse-grained model study. A famous example is Chymotrypsin Inhibitor 2 (CI2),[3] one of the first proteins, where two-state folding characteristics has experi-

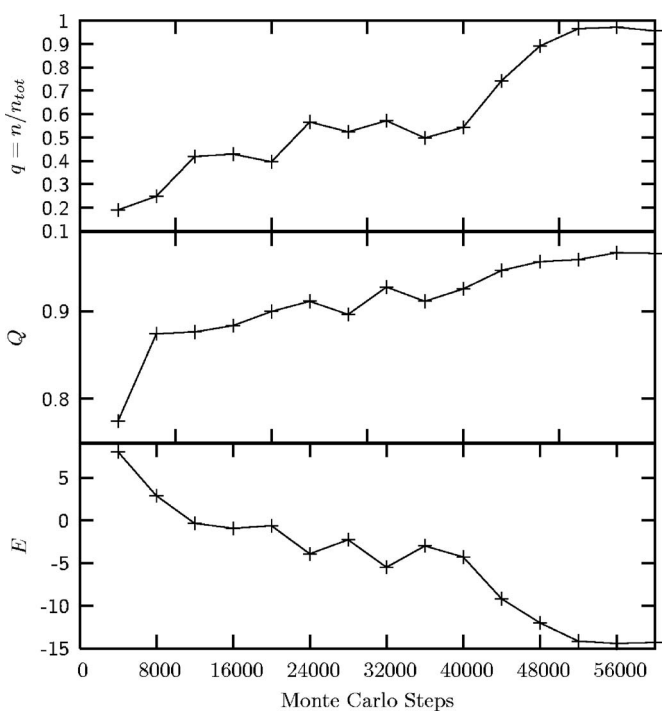

FIG. 8. Temporal averages of the native-contact and the overlap similarity parameters q and Q, respectively, and the energy for the folding event shown in Fig. 7. The temporal averages are calculated every 4000 MC steps over the time interval $[t_{MC}-1000, t_{MC}+1000]$.

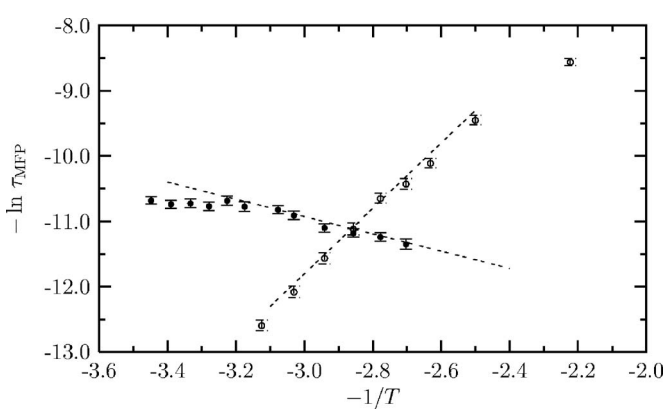

FIG. 9. Chevron plot of the mean first passage times from folding (●) and unfolding (○) events at different temperatures. The hypothetic intersection point corresponds to the transition state.

Downloaded 18 Feb 2008 to 139.18.9.115. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/jcp/copyright.jsp

mentally been identified. Clear signals of a first-order-like folding-unfolding transition were also seen in computational Gō model analyses of that peptide.[11,16] It is clear, however, that a precise characterization of the transition state ensemble, which is required for a better understanding of the folding (or unfolding) process of a specific peptide [e.g., secondary-structure formation or disruption in CI2 (Ref. 15)], is not possible. In the models used in our study, for example, only tertiary folding aspects based on hydrophobic-core formation are considered as being relevant. Nonetheless, as shown in our study of sequence $\mathcal{S}$, macroscopic quantities or cooperativity parameters manifest a qualitatively similar behavior of the heteropolymer considered here, compared to real two-state folders.

VI. SUMMARY

In this study, we have analyzed the folding thermodynamics and kinetics of a knowledge-based hydrophobic-polar Gō-like (GōL) model for heteropolymers at a mesoscopic scale. For a given sequence, the native conformation and the peak temperature of the specific heat as obtained from the physically motivated AB model were used to parametrize the GōL model. Since the chosen sequence was known to exhibit two-state-like folding characteristics in the AB model, the definition of the corresponding GōL model is appropriate for kinetic folding studies, which are very time consuming with the AB model itself. The reason is that the Metropolis Monte Carlo method we used for the kinetic studies is responsible for trapping effects and slowing down the Markovian dynamics of the physical model. Therefore, folding events are difficult to analyze in the AB model. This is the reason why for statistical analyses of this model general-ensemble methods are typically employed.[19,28] Unfortunately, these methods are not suitable in studying kinetic aspects in a fixed-temperature ensemble.

This work focuses on the question how kinetic aspects of two-state folding behavior of realistic proteins can also be identified in a strongly simplified mesoscopic knowledge-based model, at least qualitatively. In Refs. 19 and 20, we could show that it is possible to reveal statistical folding properties of different folding characteristics employing the coarse-grained AB model. Here, we have found that the corresponding mesoscopic GōL model allows for the qualitative analysis of kinetic properties of proteinlike heteropolymers with two-state folding behavior. Thermodynamic and kinetic properties were even found to be quantitatively consistent.

The advantage of such simplified, coarse-grained models is that they enable a more global, generalized view on the physics of conformational transitions accompanying protein folding processes. Our results show, in particular, that characteristic folding behaviors are not necessarily specific to microscopic details, but also an intrinsic property of hydrophobic-polar heteropolymers in general.

ACKNOWLEDGMENTS

This work is partially supported by the DFG (German Science Foundation) grant under Contract No. JA 483/24-1. M.B. thanks the DFG and Wenner-Gren Foundation for support by research fellowships. We acknowledge support by the DAAD-STINT Personnel Exchange Programme. We are also grateful to the John von Neumann Institute for Computing (NIC), Forschungszentrum Jülich, for the computer time Grant No. hlz11.

[1] K. A. Dill and H. S. Chan, Nat. Struct. Biol. **4**, 10 (1997).
[2] H. Zhou and Y. Zhou, Biophys. J. **82**, 458 (2002).
[3] S. E. Jackson and A. R. Fersht, Biochemistry **30**, 10428 (1991).
[4] A. R. Fersht, *Structure and Mechanisms in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (Freeman, New York, 1999).
[5] Y. Ueda, H. Taketomi, and N. Gō, Int. J. Pept. Protein Res. **7**, 445 (1975).
[6] N. Gō, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983).
[7] S. Takada, Proc. Natl. Acad. Sci. U.S.A. **96**, 11698 (1999).
[8] T. Head-Gordon and S. Brown, Curr. Opin. Struct. Biol. **13**, 160 (2003).
[9] V. S. Pande and D. S. Rokhsar, Proc. Natl. Acad. Sci. U.S.A. **96**, 1273 (1999).
[10] J.-E. Shea, J. N. Onuchic, and C. L. Brooks III, Proc. Natl. Acad. Sci. U.S.A. **96**, 12512 (1999).
[11] C. Clementi, H. Nymeyer, and J. Onuchic, J. Mol. Biol. **298**, 937 (2000).
[12] S. B. Ozkan, I. Bahar, and K. A. Dill, Nat. Struct. Biol. **8**, 765 (2001).
[13] M. Cieplak and T. X. Hoang, Proteins: Struct., Funct., Genet. **44**, 20 (2001).
[14] N. Koga and S. Takada, J. Mol. Biol. **313**, 171 (2001).
[15] L. Li and E. I. Shakhnovich, Proc. Natl. Acad. Sci. U.S.A. **98**, 13014 (2001).
[16] H. Kaya and H. S. Chan, J. Mol. Biol. **326**, 911 (2003).
[17] H. Kaya and H. S. Chan, Phys. Rev. Lett. **90**, 258104 (2003).
[18] J. Schonbrun and K. A. Dill, Proc. Natl. Acad. Sci. U.S.A. **100**, 12678 (2003).
[19] S. Schnabel, M. Bachmann, and W. Janke, Phys. Rev. Lett. **98**, 048103 (2007).
[20] S. Schnabel, M. Bachmann, and W. Janke, J. Chem. Phys. **126**, 105102 (2007).
[21] F. H. Stillinger, T. Head-Gordon, and C. L. Hirshfeld, Phys. Rev. E **48**, 1469 (1993).
[22] F. H. Stillinger and T. Head-Gordon, Phys. Rev. E **52**, 2872 (1995).
[23] J. R. Banavar and A. Maritan, Rev. Mod. Phys. **75**, 23 (2003).
[24] J. R. Banavar, A. Flammini, D. Marenduzzo, A. Maritan, and A. Trovato, J. Phys.: Condens. Matter **15**, S1787 (2003).
[25] J. R. Banavar, T. X. Hoang, A. Maritan, F. Seno, and A. Trovato, Phys. Rev. E **70**, 041905 (2004).
[26] T. Neuhaus, O. Zimmermann, and U. H. E. Hansmann, Phys. Rev. E **75**, 051803 (2007).
[27] U. H. E. Hansmann and L. T. Wille, Phys. Rev. Lett. **88**, 068105 (2002).
[28] M. Bachmann, H. Arkın, and W. Janke, Phys. Rev. E **71**, 031906 (2005).
[29] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
[30] K. Hukushima, H. Takayama, and K. Nemoto, Int. J. Mod. Phys. C **7**, 337 (1996).
[31] C. J. Geyer, in *Computing Science and Statistics*, Proceedings of the 23rd Symposium on the Interface, edited by E. M. Keramidas (Interface Foundation, Fairfax Station, 1991), p. 156.
[32] H. S. Chan and K. A. Dill, Proteins: Struct., Funct., Genet. **30**, 2 (1998).

Downloaded 18 Feb 2008 to 139.18.9.115. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/jcp/copyright.jsp