Comparison of Sequence-Based and Structure-Based Energy Functions for the Reversible Folding of a Peptide

Andrea Cavalli,* Michele Vendruscolo,† and Emanuele Paci*
*Biochemisches Institut der Universität Zürich, Zürich, Switzerland; and †Department of Chemistry, University of Cambridge, Cambridge, United Kingdom

ABSTRACT We used computer simulations to compare the reversible folding of a 20-residue peptide, as described by sequence-based and structure-based energy functions. Sequence-based energy functions are transferable and can be used to describe the behavior of different proteins, since interactions are defined between atomic species. Conversely, structure-based energy functions are not transferable, since the interactions are defined relative to the native conformation, which is assumed to correspond to the global minimum of the energy. Our results indicate that the sequence-based and the structure-based descriptions are in qualitative agreement in characterizing the two-state behavior of the peptide that we studied. We also found, however, that several equilibrium properties, including the free-energy landscape, can be significantly different in the various models. These results suggest that the fact that a model describes the native state of a polypeptide chain does not necessarily imply that the thermodynamic and kinetic properties will also be reproduced correctly.

INTRODUCTION

Determining how a protein folds to a stable native structure is a problem of great importance in biophysics, molecular biology, and medicine (Onuchic et al., 1997; Pande et al., 2000; Dinner et al., 2000; Thirumalai et al., 2002; Fersht and Daggett, 2002). All-atom computer simulations are a uniquely powerful technique for describing the structure and the dynamics of proteins and thus they provide an accurate framework for interpreting experimental measurements (Dinner et al., 2000; Fersht and Daggett, 2002). In the last several years, in addition, progress in understanding the physical basis of the process of protein folding has also been made from the study of simple models, such as lattice proteins with empirical pairwise interactions (Šali et al., 1994; Onuchic et al., 1997; Pande et al., 2000; Thirumalai et al., 2002) and Gō models (Taketomi et al., 1975; Zhou and Karplus, 1999; Micheletti et al., 1999; Clementi et al., 2003; Shimada et al., 2001; Karanicolas and Brooks, 2002, 2003) where the interactions present in the native state, assumed to be known, are defined to be stronger than all other interactions. Gō models have been used to study the stability of the native state of proteins (Kaya and Chan, 2002), the transition state for folding (Shimada et al., 2001; Ding et al., 2002a; Clementi et al., 2003; Karanicolas and Brooks, 2003), and the process of protein aggregation (Ding et al., 2002b).

Computer simulations are capable, at least in principle, of reconstructing with accuracy the complete free-energy landscape of proteins. This task is an ambitious one, however, since it requires the generation of very long trajectories from which the equilibrium properties of a system can be determined. Therefore, free-energy landscapes have, most often, been determined by the use of simple protein models (Dinner et al., 2000; Karanicolas and Brooks, 2003). More recently, advances in computer technology have made it possible to use all-atom molecular dynamics simulations to characterize the free-energy landscape of several polypeptide chains, including the two 20-residue-designed peptides GSGS (Ferrara and Caflisch, 2000) and Betanova (Bursulaya and Brooks, 2000), the C-terminal β-hairpin of protein G (Dinner et al., 1999; Zhou and Zhou, 2002), protein A (Ghosh et al., 2002), and the src-SH3 domain (Shea et al., 2002).

The study presented here compares in detail the thermodynamic properties of the GSGS peptide (TWIQNGSTKWYQNGSTKIYT) (de Alba et al., 1999), as determined by using a sequence-based (transferable) energy function (TEF) and various structure-based (nontransferable) Gō-like energy functions (GEFs). One of the GEF models that we studied is an all-atom model that has the same degrees of freedom of the TEF; the other GEF models are frequently used models based on coarse-grained descriptions of the polypeptide chain that use only C_α atoms. Reversible folding of the GSGS peptide, a necessary condition for determining equilibrium properties, can be achieved by using all such models.

In structure-based models the parameters are chosen so that the native state of the particular polypeptide chain under study corresponds to the overall minimum of the energy. In most cases, however, the parameters are not optimized to reproduce also the experimental measurements on the thermodynamics and kinetics of the system. Indeed, in several studies where this type of model was used, the folding mechanism was shown to depend strongly on the

Submitted October 29, 2004, and accepted for publication February 8, 2005.

Address reprint requests to Dr. Emanuele Paci, University of Zurich, Dept. of Biochemistry, Winterthurestrasse 190, Zurich, 8057, Switzerland. Tel.: 41-1-635-5559; E-mail: paci@bioc.unizh.ch.

Emanuele Paci's present address is the Institute of Molecular Biophysics, School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, UK.

0006-3495/05/05/3158/09 $2.00

doi: 10.1529/biophysj.104.055335

values of the parameters (Zhou and Karplus, 1999; Zhou and Linhananta, 2002; Zhou et al., 2003). Similarly, it is a very difficult task to define the values of the parameters of sequence-based models to simultaneously reproduce all the structural, thermodynamic, and kinetic observations made experimentally. This problem was illustrated clearly in a recent study by Alan Fersht and co-workers that showed that the folding behavior of protein A was predicted in different ways by the various models used (Sato et al., 2004). The observation that models having the correct native state may not describe the kinetic properties correctly is complemented in the present study by a quantitative comparison between the thermodynamic properties of the GSGS peptide as obtained by using several different foldable models.

METHODS

All-atom GEF simulations

All-atom GEF (aa-GEF) simulations were performed with an all-atom Monte Carlo (MC) package, *almost* (the package *almost* is available on http://open-almost.org). Each atom was represented as a hard sphere, with a van der Waals radius (r) scaled by a factor $\lambda_1 = 0.7$. For two atoms, A and B, at a distance R, the energy $E(A, B)$ was computed as (Shimada et al., 2001)

$$E(A,B) = \begin{cases} \infty, & R < \sigma \\ \Delta(A,B), & \sigma \le R \le \lambda_2\sigma \\ 0, & R > \lambda_2\sigma \end{cases}, \qquad (1)$$

where $\sigma = \lambda_1(r_A + r_B)$ is the hard-core distance, $\lambda_2 = 1.65$ a scaling factor that controls the width of the well, and $\Delta(A, B) = E_n = -1$, if A and B are in contact in the native reference structure and E_{nn} otherwise. We considered three values for non-native energy E_{nn} (0, 0.5, and 1). The total energy was computed as $E = \sum E(A, B)$ where the sum was extended to all-atoms pairs, excluding those of successive residues along the chain and all backbone-backbone contacts. We used two types of MC moves. The first was a rotation of a side-chain rotatable bond by an angle drawn from a Gaussian distribution with zero mean and standard deviation of 0.1 rad. The second type of move was a concerted rotation of the backbone ϕ–ψ angles of four successive residues, for which we used a variant of the algorithm of Favrin et al. (2001). In all simulations the acceptance ratios were close to 40%.

C_α GEF simulations

We used several different Gō models based on a C_α-only description of the polypeptide chain. In these models, interactions between C_α pairs are attractive for native contacts. The models differ in the functional form of the interactions, which are always attractive for native contacts, i.e., pairs of C_α atoms that are less than 8 Å in the native structure. In one model, the interaction between pairs making a native contact has a square-well form (C_α-SW-GEF): the interaction is -1 for distances between 0.9 d_N and 1.2 d_N, where d_N is the distance in the native structure, zero for distances more than 1.2 d_N, and infinity for distances less than 0.9 d_N. In this case, we also studied the effect of different types of non-native interactions—repulsive, neutral, or mildly attractive. For non-native contacts we considered three different values for the interaction E_{nn} (-0.1, 0.5, 1), defined for distances between 0.9 d_{NN} and 1.2 d_{NN}. The distance for a pair of C_α atoms is computed as follows. For each atom i, we define its repulsion radius $d_R(i)$ to be the distance to the first atom j not making a native contact with it; the non-native distance is then

$$d_{NN}(i,j) = \frac{1}{2}(d_R(i) + d_R(j)). \qquad (2)$$

Simulations were performed with the Monte Carlo (MC) package, *almost*.

We also considered the improved C_α Gō model, C_α-KB-GEF, proposed by Karanicolas and Brooks (2002). In this model, residues separated in sequence by three or more bonds, and which are in contact in the native reference structure, are subject to an interaction energy of the form

$$V = \varepsilon\left[12\left(\frac{\sigma}{r_{ij}}\right)^{12} - 18\left(\frac{\sigma}{r_{ij}}\right)^{10} + 4\left(\frac{\sigma}{r_{ij}}\right)^{6}\right]. \qquad (3)$$

An important feature of this model, which is different from all the other Gō models considered here, is that both the strength (ε_{ij}) and the range on native interactions (σ_{ij}) are determined from the all-atom native structure using detailed interactions (hydrogen bonds, etc.). Residues not defined as native contacts were subject to repulsive potential of the form

$$V_{ij} = \varepsilon_{ij}\left(\frac{\sigma}{r_{ij}}\right)^{12}. \qquad (4)$$

Another original feature of the C_α-KB-GEF force field is a sequence-specific term related to the backbone dihedral angles of the protein (Karanicolas and Brooks, 2002). C_α-KB-GEF simulations were performed using Langevin molecular dynamics using the program CHARMM (Brooks et al., 1983); the appropriate topology and parameter files were generated using the web server http://mmtsb.scripps.edu/.

Molecular dynamics simulations

Molecular dynamics (MD) TEF simulations were performed with the program CHARMM (Brooks et al., 1983), using an implicit solvation model based on the solvent-accessible surface (Ferrara et al., 2002). This model has not been optimized using the structure of this specific peptide, and it has been shown to reversibly fold various peptides to their respective experimental structures (α-helical or β-sheet; see Ferrara et al., 2002; Ferrara and Caflisch, 2000; Hiltpold et al., 2000). In the cases of the GSGS peptide, during long MD simulations a triple-stranded β-sheet conformation satisfying the experimental NOE-derived distances (de Alba et al., 1999) is highly populated. A reference native structure was generated by extracting a low energy structure from a low temperature simulation started from the native state. This structure satisfies all the 26 experimental NOE-derived distances and was subsequently energy-minimized. The resulting structure was also assumed as the native reference structure for all the GEF models used in this work.

Thermodynamic properties

Since we compare how different models (C_α and all-atom) describe the conformational properties of the GSGS peptide, only those properties depending on the C_α positions are used here. We thus define the radius of gyration and the root mean-square deviation (RMSD) from the native structure in terms of C_α atoms. We also monitor the number of contacts; all the pairs of C_α atoms less than 8 Å apart and separated by more than 3 residues in the sequence are considered to be in contact. The total number of contacts in the reference structure of the GSGS peptide is 40; 20 of these contacts are between strands 1 and 2, and 19 between strands 2 and 3.

The number of non-native contacts is the total number of contacts minus the number of native ones. To estimate the number of folding events, we count the number of times the peptide satisfies the specific conditions for being folded, starting from any conformation that satisfies the specific condition for being unfolded. We assume that when more than 30 native

contacts are present and the RMSD from the native reference structure is less than 2.5 Å, the peptide is native; when less than 12 native contacts are present and the RMSD from the native structure is more than 5 Å, the peptide is assumed to be unfolded. These criteria are derived from the typical bimodal distributions of both the number of native contacts and RMSD from the native structure observed for the models considered in this work (see below).

RESULTS

Molecular dynamics simulations

The equilibrium behavior of the GSGS peptide was studied by performing four MD simulations for a duration ranging from 2.7 to 4.4 μs (12.7 μs in total), started from the folded structure, at a temperature of 330 K (Cavalli et al., 2003; Paci et al., 2003). Over the total simulation time of 12.7 μs, 78 folding and unfolding events were observed.

All-atom GEF simulations

The reference native structure of the peptide contains 360 interatomic native contacts. In the aa-GEF case we performed five simulations, each of 10^9 MC steps, for three different values of E_{nn} (0, 0.5, and 1).

As in the TEF simulations, reversible transitions between completely folded and completely unfolded conformations are observed. For example, we observed 35 folding/unfolding events over the total 5×10^9 MC steps for the case $E_{nn} = 1$ at $T = 1.52$.

C_α-GEF simulations

C_α-KB-GEF simulations, using Langevin molecular dynamics, were performed at $T = 344$ K; we observed about 4000 folding/unfolding events during a 20-μs simulation. All the other C_α-GEF models were sampled with MC using the program *almost*. The C_α-SW-GEF was sampled with 10^9 MC steps; for example, for the case $E_{nn} = 1$, we counted 317 folding/unfolding events.

Determination of the melting temperature

All simulations were performed close to the melting temperature T_m. We used two different methods to compute T_m. For all the Gō models, several simulations in a broad range of temperatures were performed. The specific heat was then computed by combining the simulations with the weighted histogram algorithm method (Ferrenberg et al., 1995). For the TEF model, the density of states $\Omega(E)$ was computed using the Wang-Landau method (Wang and Landau, 2001a,b). Then, the specific heat was obtained from

$$C_v = \frac{1}{kT^2}(\langle E^2\rangle - \langle E\rangle^2), \tag{5}$$

where

$$\langle E\rangle = \int dE\Omega(E)e^{\frac{E}{kT}}. \tag{6}$$

The specific heat as a function of the temperature, for the various models, is shown in Fig. 1.

Comparison of the thermodynamic properties of TEF and GEF models

The time-series of the RMSD from the native structure for the models studied exhibits a typical two-state behavior with fast transitions between a folded and an unfolded state (Fig. 2). An analogous behavior is observed for other macro-

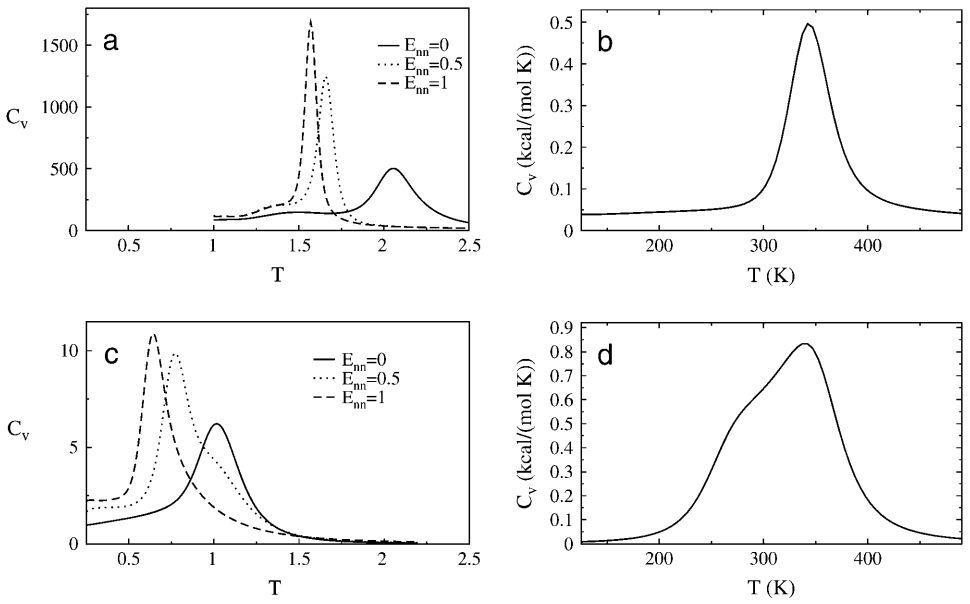

FIGURE 1 Specific heat as a function of temperature for (*a*) the aa-GEF model, (*b*) the C_α-KB-Gō model, (*c*) the C_α-SW-GEF model, and (*d*) the TEF model. In the cases *a* and *b*, temperatures are given in units of ϵ, where ϵ is the magnitude of the interaction for a native contact.

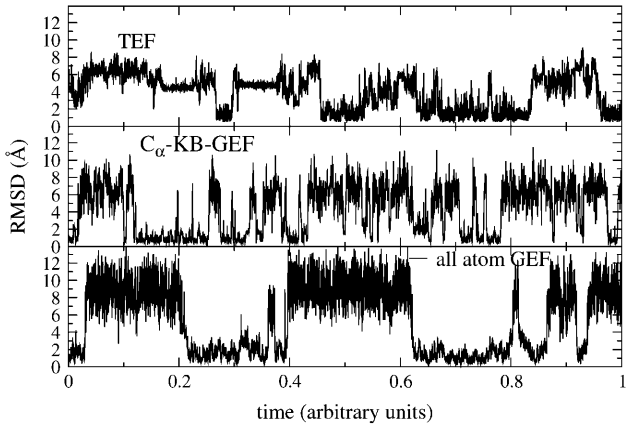

FIGURE 2 Time-series of the C_α-RMSD from the reference structure in the TEF and in two GEF models. The timescale corresponds to 0.6 μs for the TEF model, 300 ns for the C_α-KB-GEF model, and 10^9 MC steps for the aa-GEF model.

scopic variables, such as, for example, the number of native contacts.

The distribution of several quantities obtained by MD or MC simulations is shown in Fig. 3; these properties are the number of native contacts N, the number of non-native contacts N_{nn}, the radius of gyration R_g, and the RMSD from the reference structure. The distributions of the number of native contacts and of the RMSD have, in all cases, two peaks. We use these two peaks to identify the native and the unfolded regions (see Methods). In the TEF model, we observe a rather extended native-state basin, whereas the unfolded state is formed by partially structured substates, as also indicated by the presence of at least two peaks both in the distribution of N and in that of the RMSD; also, the distribution of the RMSD drops to zero at about 7 Å. For the aa-GEF model the distributions of N and RMSD show a rather broad native peak whereas the unfolded state is characterized by a number of native contacts close to zero and a value for the RMSD between 5 and 12 Å. In the aa-GEF case, the behavior depends on E_{nn} (see also Zhou and Karplus, 1999). For $E_{nn} = 1$, the behavior is first-order-like, and the separation between the native and the unfolded states is clearly observed in both the distributions of Q and C_α-RMSD. This type of behavior becomes less well-pronounced for lower values of E_{nn} and almost disappears for $E_{nn} = 0$. In the C_α-KB-GEF case, the distributions of native contacts and of the RMSD from the native state have two well-defined peaks that correspond to the native and unfolded states, respectively. The C_α-SW-GEF model is characterized, in the case of $E_{nn} = 1$, by a rather narrow native state, in terms of both the distributions of N and RMSD; whereas in the unfolded state, it has a peak at about 20 native contacts and a broad distribution of RMSD between 3 and 10 Å. Interestingly, the typical two-state behavior and a well-defined native state with a population of about 50% is also observed when non-native interactions are set to zero or even to a slightly negative value ($E_{nn} = -0.1$).

The distributions of non-native contacts and of the radius of gyration are also shown in Fig. 3. These quantities are particularly interesting because their behavior is not expected to be related to that of the number of native contacts or of the RMSD, which are better suited to characterize the native state. The number of non-native contacts, N_{nn}, has, for all models, a peak around zero and decreases rapidly for an increasing number of non-native contacts. The probability of finding a large number of non-native contacts (e.g., more than 10) is negligible in all GEF models, whereas it is sizable for the TEF model. The distribution of R_g can be used to distinguish the TEF model from the GEF models, since, in the former case, the distribution is unimodal and narrowly peaked at about 7 Å (99% of the conformations have an R_g between 6 and 8.5 Å); i.e., all conformations sampled have a comparable degree of compactness. This compactness might be slightly overestimated due to the use of the EEF1 model, which does not include explicit hydrogen bonds and hydrophobic interactions with the solvent; this suggestion was also made in a comparison between explicit and various implicit solvent models (Bursulaya and Brooks, 2000). In all the GEF models the distribution of R_g has a peak about 7.5 Å, which is the value in the native state, but also a second, broader peak at a larger value of R_g, indicative of a highly expanded unfolded state. In particular, for the aa-GEF model, in the case of strong repulsive non-native interactions ($E_{nn} = 1$), the distribution of R_g indicates that the unfolded state is made up by structures with a R_g between 9 and 16 Å; these sizes are loosely related to the magnitude of the non-native interactions and tend to decrease when E_{nn} is less than 1. For the C_α-GEF models, and particularly for the C_α-KB model, the distribution of R_g is narrower and closer to the TEF case. Interestingly, for the C_α-SW model, the effect of slightly attractive non-native interactions does not dramatically change the thermodynamical properties of the system, as shown by the distributions in Fig. 3 *d*.

A closer view of the conformations populated in the various models studied is provided by the two-dimensional histogram of N_{12} and N_{23}, which provides information about the degree of formation of the two native β-hairpins. The probability of conformations with a given number of contacts N_{12} and N_{23} (Fig. 4) shows remarkable differences between the TEF and the GEF cases and between the C_α and the all-atom GEF cases. In the TEF case, the simultaneous formation of the two hairpins is rather unlikely, whereas conformations in which only one of the two hairpins is entirely formed correspond to local minima on the free-energy surface; the free-energy surface is not symmetric and the formation of contacts N_{23} is slightly more favorable. This is not the case for the aa-GEF model; here symmetric conformations, in which the same number of native contacts is present in each hairpin, are more favorable than asymmetric ones. We also observe that the native minimum in the

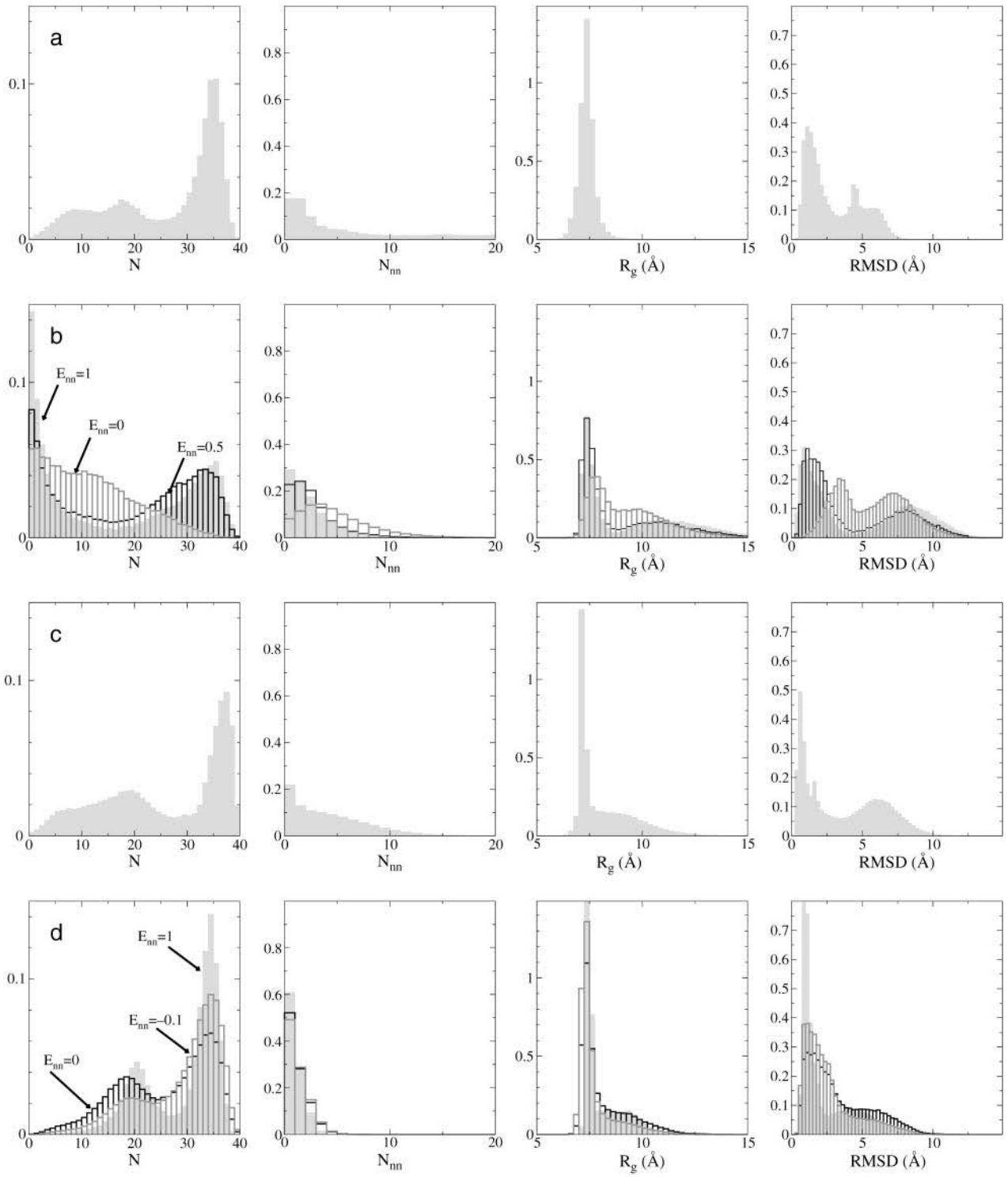

FIGURE 3 Comparison of the distributions of N, N_{nn}, R_g, and RMSD for all the models considered here. (*a*) TEF; (*b*) aa-GEF; (*c*) C_α-KB-GEF; and (*d*) C_α-SW-GEF. The normalized probability density is reported on the *y* axis.

GEF case is broad and not very close to the minimum energy conformation (i.e., the conformation where all the native contacts are formed). In the C_α-GEF models some of the features of the TEF are preserved; for example, at variance with the aa-GEF case, the folded conformation corresponds to a narrow minimum in free energy. All C_α-GEF models, and the KB model in particular, have an asymmetric free-energy landscape due to a high probability of finding conformations

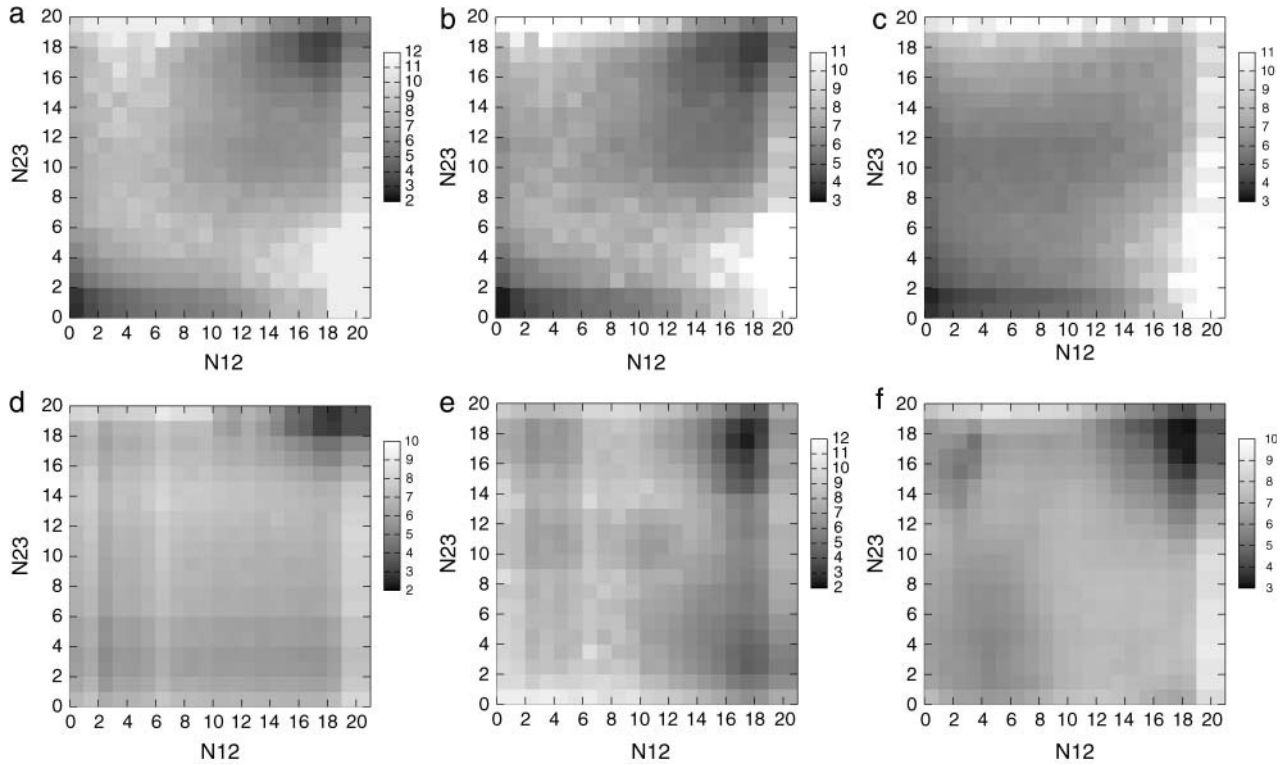

FIGURE 4 Free energy (calculated as $-\ln(N_{n,m}/N_{0,0})$, where $N_{n,m}$ is the number of conformations with $N_{12} = n$ and $N_{23} = m$) as a function of the number of native contacts between β-strands 1 and 2 (N_{12}) and β-strands 2 and 3 (N_{23}), in k_BT units. (*a*) aa-GEF, $E_{nn} = 1$; (*b*) aa-GEF, $E_{nn} = 0.5$; (*c*) aa-GEF, $E_{nn} = 0$; (*d*) C_α-KB-GEF; (*e*) C_α-SW-GEF; and (*f*) TEF.

where the hairpin 1–2 is formed and the hairpin 2–3 is not. Vice versa, in the TEF model a well-populated metastable state consists of conformations where only the hairpin 2–3 is formed.

The average energy for given number of contacts N_{12} and N_{23} is shown in Fig. 5, *a–f*. To compare different models, the energy is reported in units of k_BT_m. For all the Gō models the energy is symmetric and has a deep minimum in correspondence of $N_{12} = 19$ and $N_{23} = 20$, i.e., when both hairpins are formed. For the TEF model (Fig. 5 *f*), the energy surface is slightly different, and shows that conformations where strands 2–3 are formed are energetically very favorable. It is also relevant that, in the case of the TEF model, the largest energy difference is about 30 (in k_BT_m units), whereas it is considerably larger for the GEF models; it is of about 70 for the C_α-KB-GEF model, 180 for the aa-GEF model ($E_{nn} = 1$), and 45 for the C_α-SW-GEF. In the TEF model the energy difference between the native and the most unfolded conformations is thus considerably lower than that of any GEF model.

The total number of contacts as a function of the number of native contacts is shown in Fig. 6. In the TEF case, the unfolded state (i.e., $Q < 15$, see Fig. 3) has, on average, several non-native contacts that stabilize compact conformations. Folding occurs through the rearrangement of non-native interactions into native ones. In the GEF case, instead, non-native interactions are always highly improbable. Interestingly, at least in the range of temperatures that we studied, this feature is not related to the presence of a term in the energy function that disfavors non-native contacts; there is a trend to form more non-native contacts as E_{nn} decreases, but even when E_{nn} is zero or slightly negative, non-native contacts are negligible relative to the TEF case.

Comparison of the TEF and GEF results show that, although there are qualitative similarities, there are also significant differences in the thermodynamic properties of these various models. The unfolded state is, in the GEF cases, formed by structures with very few native contacts and a very large RMSD from the native state, i.e., it mainly comprises random-coil-like conformations. In the TEF case instead we observe that the unfolded state is closer to the native state in terms of RMSD, and some residual native structure is present. Another important difference concerns the native state. Although the energy surfaces are similar, with a well-defined minimum in correspondence to conformations with all the native contacts formed, the free-energy surfaces are different. In the GEF case, the minimum of the free energy corresponding to the native state is rather broad and shifted from the minimum of the energy. These results suggest that entropic effects in the native basin may be different in GEF and TEF models, probably because in the GEF models all native interactions have the same energy, and therefore

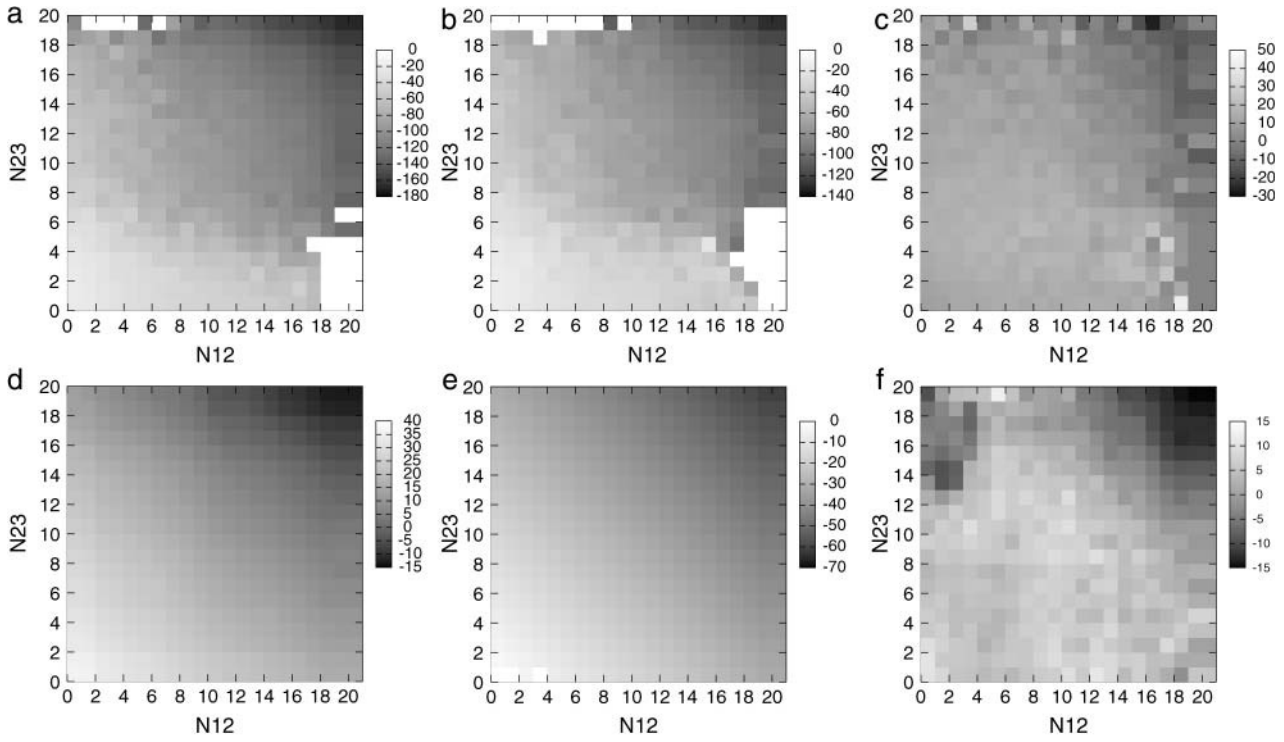

FIGURE 5 Average energy (in k_BT_m units) as a function of the number of native contacts between β-strands 1 and 2 (N_{12}) and β-strands 2 and 3 (N_{23}). (*a*) aa-GEF, $E_{nn} = 1$; (*b*) aa-GEF, $E_{nn} = 0.5$; (*c*) aa-GEF, $E_{nn} = 0$; (*d*) C_α-KB-GEF; (*e*) C_α-SW-GEF; and (*f*) TEF.

slightly expanded states within the native basin have a favorable entropy due to the many equivalent ways in which is possible to lose a few contacts. The results obtained with the C_α-KB-GEF indicate that the use of Gō-like models defined at a more coarse-grained level is less affected by this problem, and possibly better suited to represent the native free-energy minimum (see e.g., Zhou and Karplus, 1999; Karanicolas and Brooks, 2003).

The definition of the native contact map is an important aspect of any GEF simulation. Owing to thermal fluctuations, the native state is best represented by an ensemble of contact maps. In a GEF model, however, a single reference contact map should be chosen. We explored the dependence of the results on this choice by repeating the calculations for two additional contact maps, for the all-atom Gō models studied here. The variations in contact maps are particularly significant in the case of the small peptide considered here for which the number of all-atom contacts ranges between 260 and 360 for all the structures with $N = 40$ characteristic contacts in the TEF native simulation. Since some properties of GEF models may be sensitive to the choice of the reference contact map, we repeated the calculations for two

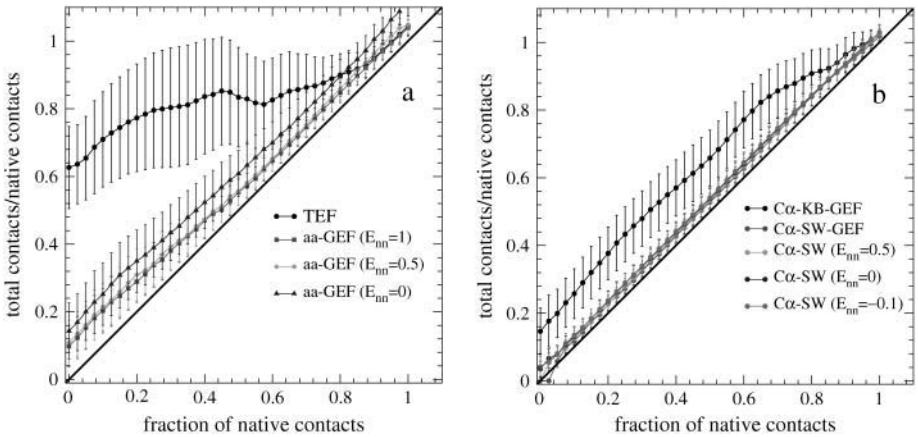

FIGURE 6 Number of native contacts divided by the total number of contacts as a function of fraction of native contacts Q. Results for the TEF and the various GEF models are shown. For all the GEF models, non-native contacts are essentially absent.

additional aa-GEF models that differ by the choice of the native contact map. Alternative contact maps were derived from the structure with the largest and the smallest number of contacts among the structures explored in a native simulation. Although the changes in the total number of contacts in the native state, and in the position of the maxima in the distributions of RMSD and Q are not negligible (data not shown), the general features of the free-energy landscape of a GEF model, such as the broadness of the native state and the high degree of disorder of the unfolded state, do not depend on the choice of the reference contact map.

CONCLUSIONS

We have compared the conformations sampled by a 20-residue peptide using a transferable, sequence-dependent model (TEF), and various Gō-like, structure-dependent models (GEF). Our results indicate that the TEF and GEF models we studied have different free-energy landscapes. These differences are particularly significant for the all-atom GEF model where interactions between pairs of atom which are in contact in the native state are equal for all pairs. In addition, in this case, a repulsive interaction between pairs that are not in contact in the native state seems to be a requirement to obtain a two-state behavior. Simpler models, where only C_α atoms are considered, give results which are, at least in certain respects, closer to the transferable, sequence-dependent potential. This is particularly the case for the C_α-KB-GEF model, where interactions are weighed according to the detailed interatomic interactions in the native state. Favorable non-native interactions are also absent in this model, but, possibly because of the presence of some long-range interactions, non-native contacts are observed in the unfolded state. As a consequence, the unfolded state is more compact than for other Gō-like models, even if not as compact as in the TEF case.

In this article we have compared the thermodynamic properties of various sequence-based and structure-based models, and only briefly discussed how these models describe the folding behavior of this peptide. Kinetic properties, however, could, in general, be expected to be different for different free-energy surfaces. A recent experimental study of the folding protein A (Sato et al., 2004) has discussed several theoretical predictions of the folding mechanism obtained with different models, and showed that they lead to different descriptions of the folding pathway. Taken together, these results suggest that a given native fold may be encoded by models with very different thermodynamic and kinetic behaviors.

We are grateful to Amedeo Caflisch for stimulating discussions and useful comments on this article.

We acknowledge the financial support from the Royal Society and the Leverhulme Trust (to M.V.) and the Forschungskredit der Universität Zürich (to E.P.).

REFERENCES

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* 4:187–217.

Bursulaya, B. D., and C. L. Brooks III. 2000. Comparative study of the folding free-energy landscape of a three-stranded β-sheet protein with explicit and implicit solvent models. *J. Phys. Chem. B.* 104:12378–12383.

Cavalli, A., U. Haberthür, E. Paci, and A. Caflisch. 2003. Fast protein folding on downhill energy landscape. *Protein Sci.* 12:1801–1803.

Clementi, C., A. E. Garcia, and J. N. Onuchic. 2003. Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: all-atom representation study of protein L. *J. Mol. Biol.* 326:933–954.

de Alba, E., J. Santoro, M. Rico, and M. A. Jimenez. 1999. De novo design of a monomeric three-stranded antiparallel β-sheet. *Protein Sci.* 8:854–865.

Ding, F., N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich. 2002a. Direct molecular dynamics observation of protein folding transition state ensemble. *Biophys. J.* 83:3525–3532.

Ding, F., N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich. 2002b. Molecular dynamics simulation of the SH3 domain aggregation suggests a generic amyloidogenesis mechanism. *J. Mol. Biol.* 324:851–857.

Dinner, A. R., V. I. Abkevich, E. I. Shakhnovich, and M. Karplus. 1999. Factors that affect the folding ability of proteins. *Proteins.* 35:34–40.

Dinner, A. R., A. Šali, L. J. Smith, C. M. Dobson, and M. Karplus. 2000. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem. Sci.* 25:331–339.

Favrin, G., A. Irbäck, and F. Sjunnesson. 2001. Monte Carlo update for chain molecules: biased Gaussian steps in torsional space. *J. Chem. Phys.* 114:8154–8158.

Ferrara, P., and A. Caflisch. 2000. Folding simulations of a three-stranded antiparallel β-sheet peptide. *Proc. Natl. Acad. Sci. USA.* 97:10780–10785.

Ferrara, P., J. Apostolakis, and A. Caflisch. 2002. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins.* 46:24–33.

Ferrenberg, A. M., D. P. Landau, and R. H. Swendsen. 1995. Statistical errors in histogram reweighting. *Phys. Rev. E.* 51:5092–5100.

Fersht, A. R., and V. Daggett. 2002. Protein folding and unfolding at atomic resolution. *Cell.* 108:573–582.

Ghosh, A., R. Elber, and H. A. Scheraga. 2002. An atomically detailed study of the folding pathways of protein A with the stochastic difference equation. *Proc. Natl. Acad. Sci. USA.* 99:10394–10398.

Hiltpold, A., P. Ferrara, J. Gsponer, and A. Caflisch. 2000. Free-energy surface of the helical peptide Y(MEARA)(6). *J. Phys. Chem. B.* 104: 10080–10086.

Karanicolas, J., and C. L. Brooks III. 2002. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* 11:2351–2361.

Karanicolas, J., and C. L. Brooks III. 2003. Improved Go-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *J. Mol. Biol.* 334:309–325.

Kaya, H., and H. S. Chan. 2002. Towards a consistent modeling of protein thermodynamic and kinetic cooperativity: how applicable is the transition state picture to folding and unfolding? *J. Mol. Biol.* 315:899–909.

Micheletti, C., J. R. Banavar, A. Maritan, and F. Seno. 1999. Protein structures and optimal folding from a geometrical variational principle. *Phys. Rev. Lett.* 82:3372–3376.

Onuchic, J. N., Z. Luthey-Schulten, and P. G. Wolynes. 1997. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48:545–600.

Paci, E., A. Cavalli, M. Vendruscolo, and A. Caflisch. 2003. Analysis of the distributed computing approach applied to the folding of a small β-peptide. *Proc. Natl. Acad. Sci. USA.* 100:8217–8222.

Pande, V. S., A. Y. Grosberg, and T. Tanaka. 2000. Heteropolymer freezing and design: toward physical models of protein folding. *Rev. Mod. Phys.* 72:259–314.

Sato, S., T. L. Religa, V. Daggett, and A. R. Fersht. 2004. Testing protein-folding simulations by experiment: B domain of protein A. *Proc. Natl. Acad. Sci. USA.* 101:6952–6956.

Shea, J. E., J. N. Onuchic, and C. L. Brooks III. 2002. Probing the folding free-energy landscape of the Src-SH3 protein domain. *Proc. Natl. Acad. Sci. USA.* 99:16064–16068.

Shimada, J., E. L. Kussell, and E. I. Shakhnovich. 2001. The folding thermodynamics and kinetics of crambin using an all-atom Monte Carlo simulation. *J. Mol. Biol.* 308:79–95.

Taketomi, H., Y. Ueda, and N. Gō. 1975. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* 7:445–459.

Thirumalai, D., D. K. Klimov, and R. Dima. 2002. Insights into specific problems in protein folding using simple concepts. *Adv. Chem. Phys.* 120:36–76.

Šali, A., E. Shakhnovich, and M. Karplus. 1994. How does a protein fold? *Nature.* 369:248–251.

Wang, F. G., and D. P. Landau. 2001a. Determining the density of states for classical statistical models: a random walk algorithm to produce a flat histogram. *Phys. Rev. E.* 64:056101.

Wang, F. G., and D. P. Landau. 2001b. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* 86: 2050–2053.

Zhou, H., and Y. Zhou. 2002. Folding rate prediction using total contact distance. *Biophys. J.* 82:458–463.

Zhou, Y., and M. Karplus. 1999. Interpreting the folding kinetics of helical proteins. *Nature.* 401:400–403.

Zhou, Y., and A. Linhananta. 2002. Role of hydrophilic and hydrophobic contacts in folding of the second β-hairpin fragment of protein G: molecular dynamics simulation studies of an all-atom model. *Proteins.* 47:154–162.

Zhou, Y., C. Zhang, G. Stell, and J. Wang. 2003. Temperature dependence of the distribution of the first passage time: results from discontinuous molecular dynamics simulations of an all-atom model of the second β-hairpin fragment of protein G. *J. Am. Chem. Soc.* 125:6300–6305.