Numerical Study of the Entropy Loss of Dimerization and the Folding Thermodynamics of the GCN4 Leucine Zipper

Jorge Viñals,* Andrzej Kolinski,*† and Jeffrey Skolnick*
*Laboratory of Computational Genomics, Donald Danforth Plant Science Center, St. Louis, Missouri 63132 USA, and †Department of Chemistry, University of Warsaw, 02-093 Warsaw, Poland

ABSTRACT A lattice-based model of a protein and the Monte Carlo simulation method are used to calculate the entropy loss of dimerization of the GCN4 leucine zipper. In the representation used, a protein is a sequence of interaction centers arranged on a cubic lattice, with effective interaction potentials that are both of physical and statistical nature. The Monte Carlo simulation method is then used to sample the partition functions of both the monomer and dimer forms as a function of temperature. A method is described to estimate the entropy loss upon dimerization, a quantity that enters the free energy difference between monomer and dimer, and the corresponding dimerization reaction constant. As expected, but contrary to previous numerical studies, we find that the entropy loss of dimerization is a strong function of energy (or temperature), except in the limit of large energies in which the motion of the two dimer chains becomes largely uncorrelated. At the monomer-dimer transition temperature we find that the entropy loss of dimerization is approximately five times smaller than the value that would result from ideal gas statistics, a result that is qualitatively consistent with a recent experimental determination of the entropy loss of dimerization of a synthetic peptide that also forms a two-stranded α-helical coiled coil.

INTRODUCTION

Reduced models of a protein have been shown to provide a possible route for the estimation of the free energy of dimerization of relatively short coiled coils (Vieth et al., 1995, 1996; Mohanty et al., 1999). A key step in the calculation of the free energy of dimerization concerns the entropy loss upon bringing two monomer chains together to form the dimer. A practical method for the numerical estimation of this entropy loss by Monte Carlo simulation is discussed in this paper.

The focus of our work is on the calculation of free energies of dimerization, and in particular a re-analysis of prior computational research on the folding thermodynamics of the GCN4 leucine zipper (Vieth et al., 1996; Mohanty et al., 1999). Leucine zippers belong to the class of structural motifs that are known as coiled coils. Generically, they comprise right-handed α-helices wrapped around each other with a small left-handed super-helical twist (Crick, 1953). While leucine zippers can exist in both monomeric or dimeric form, GCN4 forms a dimer in both the crystalline phase (O'Shea et al., 1991), and in solution (d'Avignon et al., 1998).

Numerical calculations of the free energy difference between the monomeric and dimeric forms of GCN4 have already been given by Vieth et al. (1996) and Mohanty et al. (1999). In the former case, the free energy of the monomer was estimated by transfer matrix methods, whereas the entropy change of dimerization was estimated by Monte Carlo methods. To obtain the latter, two monomers were placed in a parallel configuration and in registry with one another. As we discuss further below, the actual configuration space volume available to the dimer was substantially underestimated due to the restriction that the alignment of the two monomers introduced. This numerical analysis of the dimerization equilibrium was later extended by Mohanty et al. (1999) by using the Entropy Sampling Monte Carlo method of Hao and Scheraga (1994), although they used the same values of the entropy loss of dimerization that have been given by Vieth et al. (1996). Several thermodynamic quantities were then computed, including the dimer fraction as a function of temperature and monomer concentration, the helical content of the monomer and dimer, and an analysis of the existence of possible folding intermediates was given.

In our present work, we use the Replica Exchange Monte Carlo method (Swendsen and Wang, 1986; Geyer, 1992; Hukushima and Nemoto, 1996) to obtain the canonical distribution of both monomer and dimer forms separately. According to this method, one considers r independent Monte Carlo simulations at a set of constant temperatures $T_1, T_2, \ldots, T_r$. At fixed intervals, two configurations at different temperatures ("replicas") are chosen at random, and their respective temperatures exchanged with probability that preserves detailed balance as given by the canonical probability distribution. By exchanging configurations that have been equilibrated at different temperatures, this method is more efficient in sampling complex configuration spaces than the standard Monte Carlo method. Reweighting methods as first introduced by Ferrenberg and Swendsen (1989a, b) are then used to calculate the free energy as a function of temperature. Energy histograms collected at the set of temperatures T_i are combined ("reweighted") to minimize the statistical error in the resulting density of states. As a byproduct of the calculation, the method yields the thermodynamic free energy $F_i = F(T_i)$ at the set of temperatures T_i. Because the simulations for the monomer and

Submitted April 1, 2002, and accepted for publication July 3, 2002.

Address reprint requests to Jorge Vinals, Florida State University, Tallahassee, FL 32306-4120. Tel.: 850-644-1010; E-mail: jvinals@danforthcenter.org.

0006-3495/02/11/2801/11 $2.00

dimer are carried out separately, we finally describe how to place the free energies of the monomer and dimer on the same relative scale so that the free energy difference between them and the corresponding dimerization constant can be computed.

The calculation of the entropy loss of dimerization that we present in this paper is free from any assumptions made in previous treatments. Given the reduced model of the protein and the related set of interaction potentials, there are no further restrictions in the region of configuration space sampled in the simulations. Furthermore, we show that the entropy loss is a strong function of energy (or temperature) so that approximate treatments that factorize contributions to the partition function arising from internal degrees of freedom of the chains and contributions from relative chain translational and rotational degrees of freedom are not correct in general. We find, for example, that the volume available to the second chain of the dimer relative to the first chain had been previously underestimated by about two orders of magnitude (Vieth at al., 1996), a factor that corresponds to overestimating the entropy loss of dimerization by $\sim 4.5\ k_B$, where k_B is Boltzmann's constant. Finally, and although we do not take the solvent explicitly into account in our calculations, we compare the values of the entropy loss of dimerization that we obtain with experimental values obtained for the case of a short synthetic peptide that forms a two-stranded, α-helical coiled coil (Yu et al., 2001). In this latter work, it was found that the standard entropy change of dimerization is $\sim -5\ k_B$, or a tenth of the value predicted by ideal gas statistics ($\simeq -50\ k_B$, at standard temperature and concentration). Our calculations show a similar reduction, although our estimate of the entropy loss at the transition temperature is only 20–25% of the ideal gas value, or a factor of two larger than the corresponding experimental ratio. We finally note that the method presented in this paper to compute the entropy loss of dimerization is not restricted to our particular protein model, but that it can be straightforwardly extended to any number of computational schemes that use a more refined description of the peptide chain.

THE PROTEIN MODEL

Our analysis is based on a reduced model in which a protein is represented by a sequence of virtual bonds connecting effective particles (Kolinski and Skolnick, 1994, 1996). Each particle represents one side chain, but is assumed to be located at the center of mass of the side chain in question plus the corresponding backbone α-carbon. The effective particles are embedded in a regular cubic lattice of fixed spacing that allows for a fairly accurate representation of the backbone of known protein structures. This geometric part of the model has been checked against all structures contained in the protein data bank (Bernstein et al., 1977). The observed root-mean-squared deviation between the lattice representation of any protein and its resolved structure is typically below 0.8 Å (Kolinski et al., 1999). The actual resolution of the model is of course lower, typically of the order of 2 Å for small proteins.

Effective interactions (force fields) are introduced among the particles that include generic (sequence-independent) and sequence-specific contributions. The potentials associated with the generic type of interactions are defined to introduce a bias toward reasonable secondary structures. One such potential is introduced to account for the fact that proteins exhibit a characteristic bimodal distribution of neighbor residue distances, especially between the ith and $(i + 4)$th residues. Configurations corresponding to the larger distances in the distribution are associated with proteins that exhibit either β-type or expanded coils, whereas shorter distances correspond to helices and turns (Kolinski et al., 1999). A second generic interaction further introduces a bias toward certain packing structures such as helices and β-type states. The first potential produces the required stiffness of the polypeptide chain, whereas the second provides for cooperative packing.

Sequence-specific interactions are of three types and include short-range interactions, long-range, pairwise interactions, and many body interactions. The short-range pairwise potentials are of statistical origin and are fitted to non-homologous reference structures that do not include any coiled coils. This is accomplished by considering the known distances between pairs of amino acids that are separated by one through four bonds along the chain. Chirality is also introduced by considering an interaction between the ith and $(i + 3)$rd and $(i + 6)$th bonds to produce the correct pitch. Long-range interactions are also of statistical origin and are assumed to depend not only on relative distances between the effective particles, but also on the relative orientation between the corresponding side chains (for example, residues of opposite charges are attractive when the corresponding bonds are parallel to each other, whereas the interaction is weak or repulsive when they are anti-parallel).

Finally, although multi-body interactions are implicitly included in the pairwise potentials determined from inter-residue distances and bond angles, two additional terms are added to model hydrophobic interactions and the known probability of a residue to have a given number of parallel and anti-parallel contacts. The hydrophobic potential is estimated from the surface exposure of a given side chain, i.e., of all possible contacts of a side chain, those that are not effectively occupied by contacts with neighboring chains. The second multi-body potential introduces a bias toward known propensities of various amino acids to pack their side chains in a parallel or anti-parallel orientation. Recent applications of the methodology include the improvement of threading based structure prediction (Kolinski et al., 1999; Kihara et al., 2001), and direct ab initio folding studies (Kolinski et al., 2000). Further details can be found in Appendix A.

MONOMER-DIMER EQUILIBRIUM AND MONTE CARLO METHOD

At constant temperature and in thermodynamic equilibrium, the concentration of freely associating monomers and dissociating dimers is governed by the reaction constant, $K = [D]/[M]^2$, where $[M]$ and $[D]$ are the concentrations of monomers and dimers, respectively. The total concentration of individual chains c_0 (assuming that the only species in the system of volume V are monomers and dimers, but no solvent) is $c_0 = 2[D] + [M]$. The corresponding fractions are $x_M = [M]/c_0$ and $x_D = 2[D]/c_0$. With these definitions, we have $K = x_D/(2c_0x_M^2)$. We therefore have $x_M = (\sqrt{-1 + 1 + 8Kc_0})/4Kc_0$ and $x_D = 1 - x_M$. To relate these quantities to microscopic variables, it is useful to introduce the equilibrium constant $K_{MD} = c_0K = x_D/2x_M^2$, which can be expressed in terms of the canonical partition functions of both monomer and dimer as (Mayer and Mayer, 1963),

$$K_{MD} = N\frac{Z_D}{Z_M^2} = N\frac{Q_D}{\sigma_D Q_M^2}, \tag{1}$$

where N is the total number of chains (i.e., two chains per dimer), $c_0 = N/V$ the chain concentration in a system of volume V, and Z_M and Z_D the canonical partition functions of the monomer and dimer. The momenta degrees of freedom of both monomer and dimer can be integrated out from their partition functions, and the respective kinetic contributions cancel. We have therefore introduced the configurational partition functions Q_M and Q_D, as well as the symmetry factor σ_D that takes into account the indistinguishability of the two chains that form the dimer ($\sigma_D = 2!$ in the present case) (Mayer and Mayer, 1963). The purpose of the present paper is an estimation of Q_M and Q_D by the Monte Carlo method.

Although it is in principle possible in to estimate the ratio of configurational partition functions in Eq. 1 by direct simulation involving coexisting monomers and dimers that transform into each other by association or dissociation, we have found that this is not practical. First, proteins are large molecules (even in the reduced representation used in our work and described in Appendix A), and only a small number of them can be accommodated in a numerical computation. Second, the protein models used lack very long-range interactions, and hence there are large entropic contributions to the free energies that are difficult to sample accurately over the large configurational space of a protein. We instead follow the approach of Vieth et al. (1996) and Mohanty et al. (1999), which consists of two steps: an independent computation of the free energies of the monomer and dimer forms, followed by a transformation to a common reference state so that the free energy difference between the two can be estimated.

The configurational partition functions Q_M and Q_D are estimated by the Replica Exchange Monte Carlo method (Swendsen and Wang, 1986). Either a single monomer or a single dimer is placed in the computational volume, and a set of canonical Monte Carlo simulations are performed at a set of prescribed neighboring temperatures. By conducting simulations involving only one monomer or one dimer we are explicitly assuming that at physiological concentrations monomer-dimer or dimer-dimer interactions can be neglected. The simulation to obtain Q_D involves two identical monomers constrained during the course of the simulation to states in which there is at least one inter-chain contact.

We consider a set of r independent canonical simulations conducted in parallel at a set of neighboring inverse temperatures $\beta_i = 1/k_BT_i$, ($i = 1, \ldots, r$), where k_B is Boltzmann's constant. At fixed intervals during the course of the simulation, two configurations at different temperatures ("replicas") are chosen at random, and their respective temperatures exchanged with probability defined to preserve detailed balance as given by the canonical probability distribution. Additional details about the so-called Replica Exchange Monte Carlo method are given by Swendsen and Wang (1986). At each inverse temperature β_i, a sample of n_i statistically independent configurations is collected, and the corresponding energy histograms $h_i(E)$ calculated with some arbitrary energy binning ΔE. The histograms are then reweighted by simultaneously solving (Ferrenberg and Swendsen, 1989a),

$$p(E, \beta) = \frac{(\sum_{i=1}^{r} h_i(E))e^{-\beta E}}{\sum_{i=1}^{r} n_i e^{-\beta_i E + f_i}}, \tag{2}$$

$$e^{-f_i} = \Delta E \sum_E p(E, \beta_i), \tag{3}$$

where $f_i = \beta_i F(T_i)$ is the dimensionless thermodynamic free energy at temperature T_i. The solution of this system of equations yields an optimized estimate of the probability distribution of the energy states $p(E, \beta)$ at any temperature β, as well as the thermodynamic free energies f_i. As is standard, free energy and canonical partition function are related through $f_i = -\ln Q(\beta_i)$ (recall that the momenta degrees of freedom are not considered in the simulation, and hence the free energy obtained is related to the configurational partition function only). Also, for later reference, we note that the configurational density of states $W(E)$ is given by,

$$W(E) = p(E, \beta)e^{\beta E},$$

and therefore the configurational entropy can be obtained as,

$$\frac{S(E)}{k_B} = \ln(W(E)\Delta E).$$

Two sets of simulations are performed to yield monomer $\{f_{M,i}\}$ and dimer $\{f_{D,i}\}$ free energies at the set of inverse temperatures $\{\beta_i\}$. Note that each set is known up to an arbitrary additive constant.

In the remainder of this section we describe a number of transformations of the configurational partition functions Q_M

and Q_D either for computational convenience or for the calculation of the proper reference state. Given the assumed integration of particle momenta, both Q_M and Q_D are expressed in terms of individual particle coordinates, and have dimensions of V^M and V^{2M}, respectively, where M is the number of amino acids in the monomer. Elimination of rigid translation or rotation degrees of freedom from the configurational partition functions, for example, is usually accomplished by introducing the center of mass and principal axes of inertia of the molecule, and then relative coordinates for the individual particles. This requires either integrals over the corresponding canonical momenta or the explicit consideration in the coordinate integrals of the appropriate transformation Jacobians. Because the Jacobians are configuration-dependent, we follow instead earlier work by Vieth et al. (1995), and conduct all our transformations on single particle coordinates alone, thus obviating the need to introduce complicated Jacobian functions. For example, the elimination of the degrees of freedom associated with uniform translations is accomplished by eliminating the motion of particle 1 in one of the chains. The elimination of rigid rotation is partially accomplished by disallowing the rotation of the bond between particles 1 and 2 of one of the chains. Finally, in our estimate of entropy losses on dimerization (next section), we exclusively use phase space volumes of single particle coordinates to maintain the necessary dimensional consistency of Eq. 1 after all the transformations of both Q_M and Q_D that are described in this section.

The configurational partition functions Q_M and Q_D are independent of the location of the molecule and of its orientation. The statistical accuracy of the simulation is greatly increased if those degrees of freedom that correspond to rigid translations and rotations are eliminated. The translational degree of freedom is eliminated by fixing the location of the first particle in the monomer chain, or of chain one in the dimer. Because the partition function is independent of this particle's location, this coordinate can be integrated out to yield a factor of V to the configurational partition function. The state of rigid rotation of the molecule can be specified by three angles, or the orientation of one axis plus a rotation around this axis. The orientation of the axis is defined by its azimuth $\alpha \in (-\pi, \pi)$ and a polar angle $\theta \in (0, \pi)$. The rotation around this axis is given by a third angle $\gamma \in (-\pi, \pi)$. Therefore, the corresponding element of volume in configuration space is given by the triple integral $\int d\alpha\ d\cos(\beta)\ d\gamma = 8\pi^2$. This value can also be exactly factored out from the configurational partition function. However, given that the Monte Carlo method used employs multiple bond transitions, we have found it convenient to proceed somewhat differently. We disallow the Monte Carlo transition that corresponds to a two-bond change at the N-terminus, and therefore effectively eliminate the motion of particles 1 and 2 of the chain. While fixing the location of particle 1 still allows an exact calculation of the partition function, eliminating the motion of particle 2 introduces two approximations. The first one involves the factorization of the configuration space volume of particle 2. Because the motion of this particle is not independent of the motion of the rest of the chain, this factorization is only approximate. The degrees of freedom that are eliminated include the orientation of the axis defined by the bond vector between particles 1 and 2, plus fluctuations in bond vector length. The elimination of the rotation is exact because the partition function is independent of the orientation of this axis, and yields a factor of $\int d\alpha\ d\cos(\beta) = 4\pi$ to the overall partition function. Factorization of the configuration space volume associated with bond length fluctuations is only approximate. We write,

$$Q \simeq V V_2^{(1)} Q' \simeq V \frac{4\pi}{3} (R_{max}^3 - R_{min}^3) Q', \qquad (4)$$

where Q' is the partition function that is actually computed during the Monte Carlo simulation, and $V_2^{(1)}$ is the *constant* accessible volume for particle 2, including rotation of α and β of the first bond. The second approximation made involves the assumption that the second particle is free to move within a spherical shell centered in the first particle, of inner radius R_{min} = 4.35 Å and outer radius R_{max} = 7.94 Å. These two values are the smallest and largest bond distances allowed in the lattice model used. Finally, note that the computed partition function Q' still contains a factor of 2π corresponding to the angle γ, the unrestricted rigid rotation of the molecule around the axis defined by the bond vector between particles 1 and 2 that is not eliminated during the simulation.

Reference state calculation

To introduce a common scale for the monomer and dimer free energy sets $\{f_{M,i}\}$ and $\{f_{D,i}\}$, we follow the method of Vieth et al. (1996). At sufficiently high energies, one may assume that inter-chain interactions are negligible, and that the internal motions within each dimer chain are well-approximated by those of the monomer. By internal motions one means those included in the internal partition function, once collective translation and rotation degrees of freedom have been eliminated. At high energies, particle motions within each chain become uncorrelated with those of the other chain and therefore the probability distribution corresponding to the internal degrees of freedom of the dimer factors $p_{D,int} \simeq p_{M,int}^2$. Therefore, and only in this limit, the entropy of the dimer is approximately twice that of the monomer when the energy of the dimer is twice the energy of the monomer. This fact allows one to place the entropies from both monomer and dimer simulations in the same reference state at high energy, and hence to compute free energy differences between the two.

We briefly summarize here the steps taken for both monomer and dimer. The calculation of the internal entropy of the monomer is straightforward. In the dimer case, however, the contributions from the internal modes have to be separated

from other degrees of freedom related to the relative position and orientation of the two chains. Because in the dimer simulation the two chains are constrained to have at least one contact, the computed entropies of the dimer at high energies still contain entropy losses due to this constraint that have to be estimated and subtracted to compute its internal entropy.

Monomer

The monomer partition function Q'_M computed by the Monte Carlo method still contains the contribution of one degree of freedom that is not associated with the internal motions of the particles, and that corresponds to a rigid rotation of the molecule around the axis defined by the bond between effective particles 1 and 2. From the Monte Carlo simulation and reweighting, we obtain the probability density $p'(E, \beta)$ that corresponds to the partition function Q'_M in Eq. 4. The corresponding entropy is,

$$\frac{S'_M(E)}{k_B} = \ln(p'(E, \beta)\Delta E) + \beta E,$$

a quantity that is independent of the inverse temperature β. The internal entropy follows by subtracting the entropy of rotation of the azimuth angle of bond 2–3, and by adding the estimate given above for the radial part of the first bond,

$$\frac{S_{M,int}(E)}{k_B} = \ln(p'(E, \beta)\Delta E) + \beta E - \ln(2\pi) + \ln\frac{(R^3_{max} - R^3_{min})}{3}. \quad (5)$$

Dimer

The dimer simulation is conducted by fixing the positions of particles 1 and 2 of one of the chains. Therefore, a Monte Carlo estimate is obtained for Q'_D as given in Eq. 4, As was the case for the monomer, we first define the entropy as estimated from the simulation by

$$\frac{S'_D(E)}{k_B} = \ln(p'(E, \beta)\Delta E) + \beta E.$$

where $p'(E, \beta)$ corresponds to Q'_D above. At sufficiently high energies, where the internal degrees of freedom of each chain are expected to become independent of the relative position and orientation of both chains, the total conformational density of states factors into a product involving the various contributions. In terms of the entropy, this factorization leads to the decomposition,

$$\frac{S_{D,int}(E)}{k_B} = \ln(p'(E, \beta)\Delta E) + \beta E - \ln(2\pi) + \ln\frac{(R^3_{max} - R^3_{min})}{3} - \ln V_1^{(2)}(E) - \ln(\varphi_\alpha(E)\varphi_{\cos\beta}(E)\varphi_\gamma(E)). \quad (6)$$

The quantity $V_1^{(2)}(E)$ is the accessible volume of particle 1 of chain 2 at energy E, and hence yields the accessible volume loss of dimerization. Its estimate during the Monte Carlo run is one of the main topics of this paper. Because the motion of the second chain relative to the first is constrained so that the number of inter-chain contacts is greater than zero, this volume will be in general much less than V. The quantity $\varphi_\alpha(E)$ is the configuration volume available for the azimuth of bond 1 of chain 2, $\varphi_{\cos\beta}$ for the cosine of the polar angle of bond 1 of chain 2, and φ_γ for the azimuth of bond 2 of chain 2. The product of the three represents the loss of rotational configuration space volume due to the formation of a dimer. If the second chain were to rotate freely relative to the first, we would have $\varphi_\alpha(E)\varphi_{\cos\beta}(E)\varphi_\gamma(E) = 8\pi^2$. The value found is less that this upper bound, but it approaches $8\pi^2$ as the energy of the dimer is increased. As was the case with $V_1^{(2)}(E)$, these three quantities also need to be estimated during the simulation.

Configuration space volume estimation

The configuration space volumes $V_1^{(2)}(E)$, $\varphi_\alpha(E)$, $\varphi_{\cos\beta}(E)$, and $\varphi_\gamma(E)$ have been estimated by using the method of coincidences (Ma, 1985). Let Γ be the volume of a certain region of configuration space. Consider a finite sample of configurations that are uniformly distributed in Γ, and let $\Gamma_0 \ll \Gamma$ be a small coarse-graining volume in configuration space. The method involves computing the coincidence rate R that a pair of configurations in the sample belongs to the same coarse-graining volume. If the configurations are uniformly distributed in Γ, the probability of a coincidence is $R = \Gamma_0/\Gamma$. Therefore an estimate of $R = n_c/n_t$, where n_t is the total number of pairs in the sample, and n_c the total number of coincidences given Γ_0, allows an estimation of Γ.

To satisfy the conditions of the method, we first group all configurations (regardless of their temperature) according to their energy. Because all the configurations with the same energy are expected to occur with equal probability, we calculate the coincidence rate $R(E)$ to estimate $\Gamma(E)$ for the various magnitudes of interest ($V_1^{(2)}$, φ_α,$\varphi_{\cos\beta}$, and φ_γ).

We next note some limitations in the accuracy of the method. If Γ_0 is not much smaller than Γ, error is introduced as Γ_0 will not generate a good covering set of Γ, and it is likely that the method will overestimate the size of the region Γ. However, if Γ_0 is too small the number of coincidences will be small, and the statistical error in the determination of R is large. There is a third source of error associated with the sample size at each energy or, equivalently, the total number of pairs $n_t(E)$ (Ma, 1985). The number of coincidences can be estimated as,

$$n_c(E) \sim \frac{1}{2}\left(\frac{n_t(E)}{k}\right)^2\frac{\Gamma_0}{\Gamma},$$

with Γ_0 such that all $n_t(E)$ pairs have been distributed among k groups. Therefore,

$$\Gamma \sim \frac{1}{2}\left\langle \frac{n_t(E)}{k} \right\rangle^2 \frac{\Gamma_0}{n_c(E)},$$

so that for a fixed minimum $n_c(E)$ to ensure adequate statistics of the coincidence rate, the estimated value of Γ is bounded by $n_t^2(E)$. Therefore sufficiently large samples are needed at each energy if the corresponding value of Γ is large. We will further illustrate these limitations in the next section.

Reference entropy difference

In order to place both the monomer and dimer in the same reference state, we require that in the limit of high E,

$$S_{D,int}(E) + S_0 = 2S_{M,int}(E/2), \tag{7}$$

where S_0 is a *constant,* independent of E, $S_{D,int}(E)$ is given by Eq. 6, and $S_{M,int}(E)$ by Eq. 5. The quantity S_0 is determined numerically as shown in the next section.

Once the constant S_0 has been determined, the free energy and partition function of the dimer are re-scaled according to

$$f_i'' = f_i' - S_0, \quad Q_D'' = Q_D' e^{S_0}. \tag{8}$$

We can now compute the equilibrium constant K_{MD} by substituting Eq. 4 for both monomer and dimer into Eq. 1, but using the rescaled dimer partition function Q_D'' defined in Eq. 8 instead of Q_D',

$$K_{MD} = \frac{N}{VV_2^{(1)}} \frac{Q_D''}{\sigma_D Q_M'^2} = \frac{Ne^{S_0} Q_D'}{VV_2^{(1)} \sigma_D Q_M'^2}. \tag{9}$$

With this re-definition of the dimer partition function, both Q_D'' and Q_M' refer to the same reference state, and hence absolute values of K_{MD} can be given.

In terms of the free energies of the monomer and dimer that are obtained from the Monte Carlo calculation after reweighting (Eqs. 2 and 3), Eq. 9 leads to,

$$\ln K_{MD}(\beta_i) = \ln c_0 + S_0 - f_{D,i}' - \ln \sigma_D - \ln V_2^{(1)} + 2f_{M,i}'. \tag{10}$$

If $\ln K_{MD} > 0$ the dimer is prevalent.

RESULTS

The method described in the previous section has been tested on the GCN4 leucine zipper, a 31-residue segment with the characteristic heptad repeat sequence of leucine zippers comprising residues 3–33 of 1zta plus the N- and C-terminus caps [the monomer sequence considered is RMKQLEDKVEELLSKNYHLENEVARLKKLVG]. The oligomerization equilibrium of the wild type has been addressed both experimentally (O'Shea et al., 1991; d'Avignon et al., 1998) and computationally (Mohanty et al., 1999), as have several of its mutant forms (Harbury et al., 1993; Vieth et al., 1995). Due to the short length of the sequence, and the simplicity of its secondary structure, numerous computational studies have addressed various aspects of the oligomerization process in GCN4, including dimer and multi-mer equilibria (Vieth et al., 1995), the stability of several of its subdomains (Vieth et al., 1996), oligomeric equilibrium of several of its mutant forms (Skolnick et al., 1995), and other parameters of the coiled coil such as the helical content as a function of temperature and a van't Hoff enthalpy analysis to reveal the adequacy of a two-state assumption for the dimerization process (Mohanty et al., 1999).

We have extended the analysis of Mohanty et al. (1999) in two directions. First, we use a Replica Exchange Monte Carlo method instead of the Entropy Sampling Monte Carlo method of that reference, as the former provides a faster rate of convergence to the equilibrium distribution of the dimer form. Second, we extend the method of calculation of the various entropy losses upon dimerization, and show their strong dependence on the energy of the configuration, a dependence that was not taken into account in previous studies.

The results shown are based on two long runs for the monomer and dimer forms, respectively. Initial configurations were chosen close to the native state, but first equilibrated at constant temperature. Several runs with different initial conditions yielded essentially identical results for the various thermodynamic quantities presented, although none of the dimer simulations involved an initial condition in a manifestly anti-parallel configuration. The monomer runs involved 2×10^6 independent configurations or steps after equilibration, with one replica exchange attempted every 500 steps. Quantities for analysis were collected every 250 steps. The dimer simulation comprises two identical, and initially parallel, chains with at least one contact between them. [Two residues are considered to be in contact if the distance between the two corresponding particles is smaller than a predefined cutoff. The value of this cutoff distance depends on the pair of amino acids involved, and has been determined simultaneously with the other parameters that define the interaction potentials. Distances range between $\sim$3 and 5 in units of the lattice spacing.] The simulation in this case is conducted by rejecting all bond moves that would result in no contacts between the chains. The run for the dimer involved 1.3×10^6 configurations, with the same frequency of analysis and of replica exchange. In both cases, 20 independent replicas were run in parallel at dimensionless temperatures in the range $T = 0.5$–1.45 in increments of 0.05. In the low temperature range, the root-mean-squared deviation (RMSD) between the estimated location of the α-carbons in the model and the native configuration is of the order of 3 Å (see Fig. 1). We note

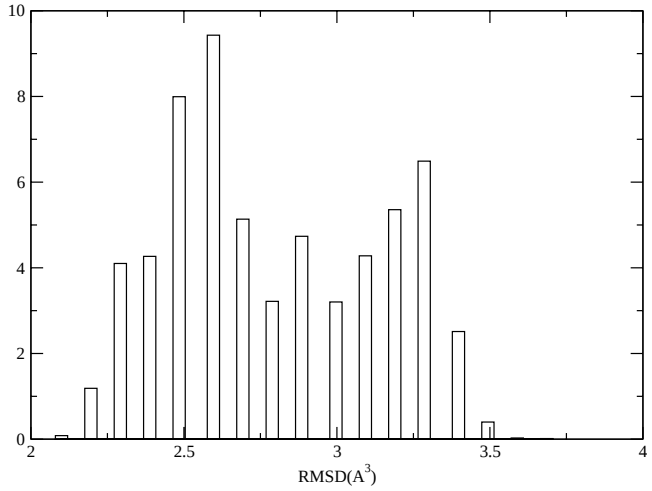

FIGURE 1 Histogram of the sampled root-mean-squared deviation from native (RMSD in Å^3 at the dimensionless temperature $T = 0.6$.).

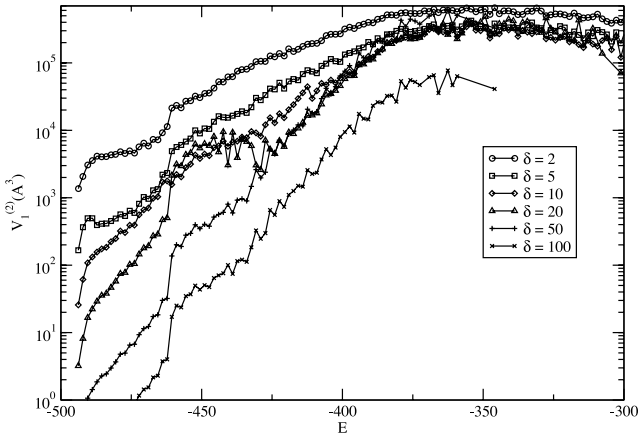

FIGURE 2 Volume of configuration space $V_1^{(2)}$ (in Å^3) that is accessible to rigid translation of chain 2 of the dimer as a function of its energy. Several different choices of the coarse-graining volume Γ_0 are shown in the figure as indicated by the values of δ.

that this RMSD range is a result of the simulation and was not enforced by the use of any restraint.

The analysis presented is based on energy histograms and the subsequent reweighting described in the previous section. The entire range of energies sampled by the monomer and the dimer during the course of the simulation was divided into 100 equal bins, so that in dimensionless units $\Delta E \simeq 0.2926$ for the monomer and $\Delta E \simeq 0.3767$ for the dimer.

In the case of the dimer, the location of all individual particles was also recorded every 250 steps in order to estimate the configuration space volumes $V_1^{(2)}$,φ_α, $\varphi_{\cos\beta}$, and φ_γ. The results presented for $V_1^{(2)}$ are based on the spatial coordinates of particle 15 of chain 2 (the chain that is free to move within the computational volume). Substantially identical results follow from an analysis of any other particles in the chain, except for those in the immediate vicinity of the N- or C-termini.

The configuration space volume φ_α is estimated from the azimuth of the bond between particles 15 and 16 of chain 2, and $\varphi_{\cos\beta}$ follow from cos β, β being the polar angle of this bond. Finally, φ_γ is obtained from the azimuth distribution of the bond between particles 16 and 17. In a freely rotating molecule, α is uniformly distributed in $(-\pi, \pi)$, cos β in $(-1, 1)$, and γ in $(-\pi, \pi)$, resulting in a combined conformational space volume for rigid rotation of $\varphi_\alpha\varphi_{\cos\beta}\varphi_\gamma = 8\pi^2$.

Fig. 2 shows our results for $V_1^{(2)}$ with the same energy bin size ΔE used to construct the histogram. The coarse-graining volume $\Gamma_0 = \Delta x\Delta y\Delta z$ has been obtained by defining $\Delta x = (x_{\max} - x_{\min})/\delta$, and similarly for Δy and Δz. $x_{\min}$ and $x_{\max}$ are the smallest and largest values of x_{15}, the x coordinate of particle 15, for each particular energy bin. We present our results for a range of values of δ in Fig. 2. If δ is too large, the coarse-graining volume is small, and the number of pairs for a given energy $n_t(E)$ is also small. As discussed in the previous section, this leads to underestimating the accessible volume. The value of $V_1^{(2)}$ is seen to increase with decreasing δ, becomes approximately independent of δ in some range, and then further increases with decreasing δ. If δ is too small, the shape of the region being sampled cannot be accurately reproduced with this coarse Γ_0. Note that the values of $V_1^{(2)}$ at low energies are the most difficult to estimate, presumably because the shape of the region in configuration space is not as smooth as that at higher energies. However, as the procedure leading to the computation of the reference entropy relies only on the region of high energies, this inaccuracy does not represent a significant limitation to our results. In what follows, the results that we present correspond to $\delta = 10$.

The behavior just described is qualitatively similar to that shown in Fig. 3 corresponding to the rotation volume

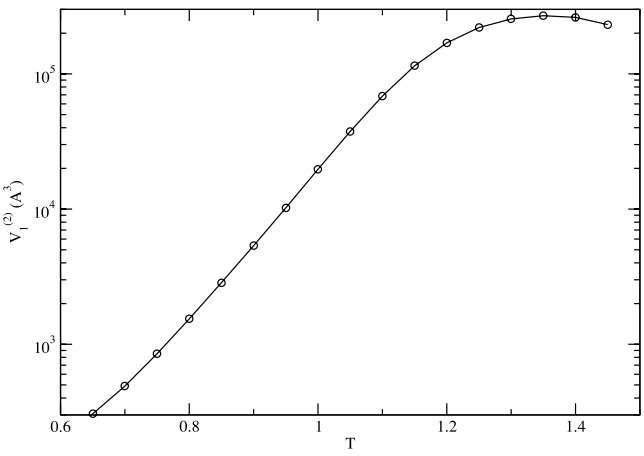

FIGURE 3 Volume of configuration space accessible to rotation of chain 2 of the dimer $\varphi_{\alpha\varphi\cos\beta}\varphi_\gamma$ for four different choices of the coarse-graining volume, as indicated by the values of δ in the figure. The straight line corresponds to the value associated with free rotation, $8\pi^2$.

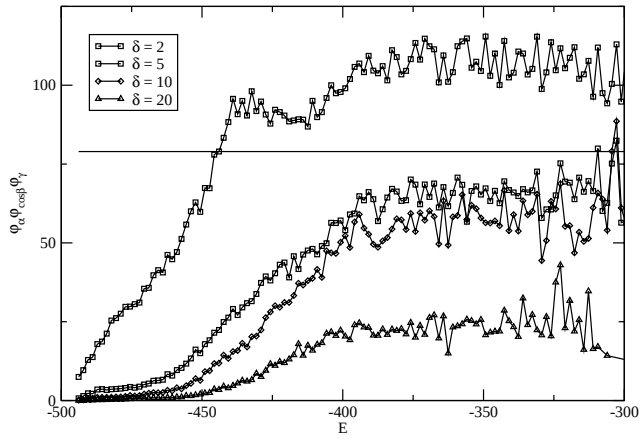

FIGURE 4 Volume of configuration space $V_1^{(2)}$ (in Å^3) that is accessible to rigid translation of chain 2 of the dimer as a function of temperature. The coarse graining volume chosen corresponds to $\delta = 10$ (see Fig. 2).

$\varphi_\alpha\varphi_{\cos\beta}\varphi_\gamma$. In this case we define $\Gamma_0 = \Delta\alpha\Delta(\cos\beta)\Delta\gamma$ with the same definition of $\Delta\alpha$, $\Delta(\cos\beta)$ and $\Delta\gamma$ in terms of the quantity δ. The figure also shows (*solid line*) the value $8\pi^2$ that corresponds to free rotation of chain 2 relative to chain 1. As can be seen from the figure, the values obtained approach this limit at high energies for the intermediate values of δ shown.

The simplest approximation in the literature to estimate the entropy change of dimerization is as minus the entropy of translation and rotation of an ideal gas, under the assumption that as far as the translational and rotational degrees of freedom are concerned, the chain in solution behaves as an ideal gas. The translational entropy of an ideal gas per particle is given by Hill (1986): $S_t^{id} = k_B[5/2 - \ln(\Lambda^3 c_0)]$, where c_0 is the inverse volume per chain, and Λ the thermal wavelength $\Lambda = h/\sqrt{2\pi M k_B T}$. The rotational entropy can also be calculated exactly (Hill, 1986), and is about the same order of magnitude for a typical protein. It is well known that the ideal gas approximation greatly overestimates the entropy loss of dimerization, up to as much as a factor of 10 (Yu et al., 2001). For the case under consideration, the molecular mass of GCN4 is ~100 Da per residue, yielding a thermal wavelength of $\Lambda \sim 0.1$ Å (we use the molecular mass per residue and not per chain in order to compare the ideal gas result with the calculation of the volume loss $V_1^{(2)}$ also defined per residue). For example, we find at $c_0 = 10$ μM that $S_t^{id} \simeq 28\ k_B$ (or an entropy change of dimerization of $-28\ k_B$). Adding a similar contribution for the rotational component would lead to an estimate of the entropy change of dimerization (combined translation and rotation) of $S_{tr} \sim (-50\ k_B) - (-60\ k_B)$. In contrast, the value that we find at the transition temperature for $c_0 = 10$ μM is $S_{tr} \simeq -12\ k_B$ (from Fig. 4, and a loss of rotational volume of $10.6/8\pi^2$). This is approximately a factor of five smaller than the ideal gas result.

In the case of the synthetic peptide, it has been determined experimentally (Yu et al., 2001) that $S_{tr}^o \simeq -5\ k_B$, where S_{tr}^o is the combined entropy change of translation and rotation at the standard concentration $c_0 = 1$ M. The ideal gas prediction is $S_{tr}^{id} \simeq 50\ k_B$, and hence the entropy is overestimated by a factor of ~10. Although both experiments and our numerical estimate show a similar trend, there are still a number of important differences between the two. First, whereas our simulations consider two chains to be a dimer if there is at least one contact between *any* two resides belonging to different chains, the experiments considered explicitly cross-linking the two chains through the formation of disulfide bonds. Therefore, the translation volume available to the dimer in the simulation could be up to $\sim 30^2$ times larger than the disulfide bridged monomer, at least at high energies. Second, solvent effects contributing to the experimental entropy loss cannot be properly accounted for in our calculation. The solvent is only implicitly included in the interaction potentials, which of course only contribute in the simulation when there are residue contacts.

Our limiting results at large energies can be understood by purely geometric considerations. In this limit, the two chains of the dimer are moving almost freely, and our result for $V_1^{(2)}$ simply follows from the volume available to chain 2 relative to chain 1, when the motion of the two is largely uncorrelated except for the constraint that at least one interchain contact be maintained at all times. The length of the helical monomer is of the order of 45 Å, so that the volume swept by the second chain relative to the first is $\sim(45\ \text{Å})^3 \sim 10^5$ Å^3, which is essentially the value of $V_1^{(2)}$ that we find numerically in the limit of high energies. In this limit we find $S_{tr}(10\ \mu\text{M}, E \to \infty) \simeq -6\ k_B$, a value that manifestly does not include any solvent effects. Near the transition temperature, the available volume $V_1^{(2)}$ decreases by about two orders of magnitude as a consequence of chain-chain interactions, but we cannot determine how much of the decrease could be ascribed to chain interactions versus solvent effects. Finally, we mention that although the value of $V_1^{(2)}$ can be determined at all energies (or temperatures), the partition function only factors into relative translation and internal degrees of freedom at high energies (or high temperatures) when $V_1^{(2)}$ becomes independent of energy. Otherwise, contributions to the various thermodynamic quantities cannot be separated into relative translation and rotation degrees of freedom on the one hand, and internal motions within the chains on the other.

We next address the determination of the constant S_0 of Eq. 7 that required to place the monomer and dimer free energies in the same scale. This quantity is obtained directly from Eqs. 5 and 6, as shown in Fig. 5. In this figure we plot $S_{D,int}(E) + S_0$ and $2S_{M,int}(E/2)$ with S_0 adjusted graphically so that the two curves coincide at large E. Note that both curves superimpose to a good accuracy for a range of energies, indicating the consistency of the approach.

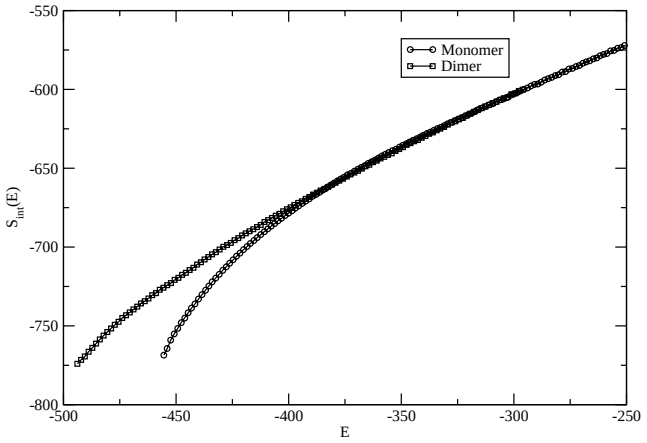

FIGURE 5 Rescaled internal entropy of the dimer $S_{D,int}$ + S0, with $S_0 = 52$ (*squares*), and twice the internal entropy of the monomer $2S_{M,int}(E/2)$ (*circles*). The value of S_0 has been graphically determined to make the two curves coincide in the range of large E. As can be seen from the figure, the two curves superimpose quite accurately over a significant range of energies.

We show next our results for K_{MD} in Fig. 6, with K_{MD} defined in Eq. 9, as a function of the dimensionless temperature. When $K_{KD} = 1$ the fractions of the monomer and dimer forms are equal ($x_M = x_D = 0.5$). At low temperatures $K_{MD} > 1$, indicating a prevalence of the dimer form, and the reverse is true at high temperatures. For the sake of illustration, the figure shows the values of K_{MD} at two different concentrations $c_0 = 10\ \mu$M and $c_0 = 1$ mM. The contribution to $\ln K_{MD}$ from $V_1^{(2)}$ is of the order of $\ln 3 \times 10^5 \simeq 13$ compared with the earlier result of $\ln 67 \simeq 4$. As it can be seen from Fig. 6, this is a substantial contribution. The transition temperatures found are lower than those given by Mohanty et al. (1999). This is due to later refinements of the interaction potentials, and slightly different values of the weights given to the various terms in the energy (see Appendix A). Recall that the interaction parameters of the model have been tuned so that native protein structures correspond to minima of the potential hypersurface. Therefore, one ought to bear in mind that the absolute scales of energy (or temperature) used are, to a certain extent, arbitrary. As a matter of fact, the model could be expected to yield agreement with experiments only when dimensionless ratios are computed, so that the absolute scale of energy (or temperature) drops out.

As a further illustration subject to the caveats just mentioned, we calculate the internal energy of folding per chain $\Delta E = \langle E_D \rangle/2 - \langle E_M \rangle$, where $\langle E_D \rangle$ and $\langle E_M \rangle$ are the ensemble averages of the energy of the dimer and monomer phases, respectively. At the transition temperature (see Fig. 6), we find $\Delta E = -25\ k_B T$, that would correspond to $\Delta E = -15$ kcal/mol at room temperature. Folding enthalpies of the GCN4 leucine zipper have been measured experimentally (see, e.g., Kenar et al. (1995) and Holtzer et al. (2001)), and are in the range of -45 kcal/mol to -60 kcal/mol. The measured enthalpy difference is expected to be approximately equal to the internal energy difference, and hence the value measured is about three time larger than our estimate. This is not an unexpected difference, as the coarse-grained nature of the model leads to less efficient residue packing at the transition temperature than the wild protein. Better agreement with experiments is expected

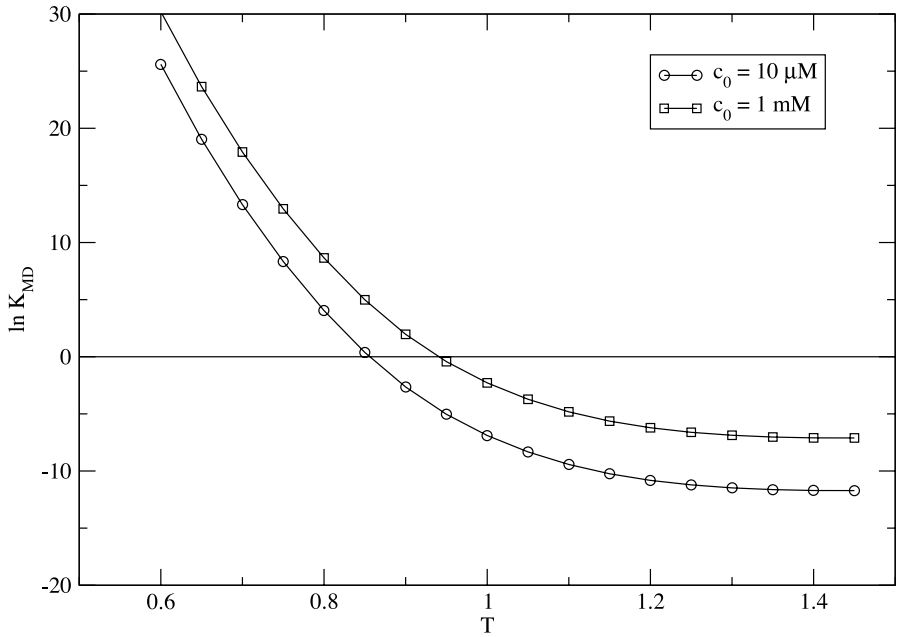

FIGURE 6 Equilibrium reaction constant for monomer-dimer equilibrium as a function of dimensionless temperature and for the two concentrations indicated.

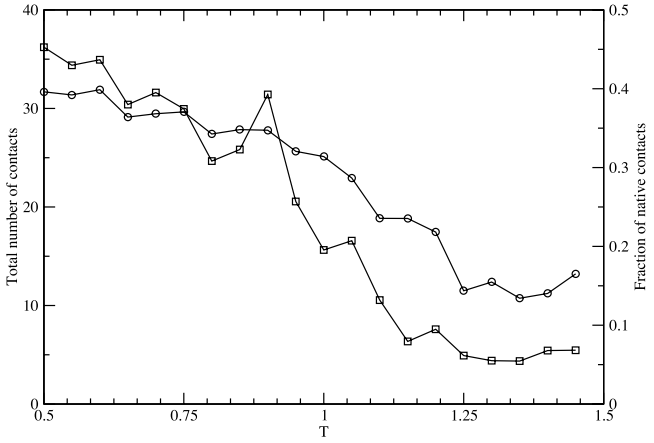

FIGURE 7 Total number of inter-chain contacts (○, *left axis*) as a function of temperature, and (□, *right axis*) the fraction of the total number of inter-chain contacts that are native.

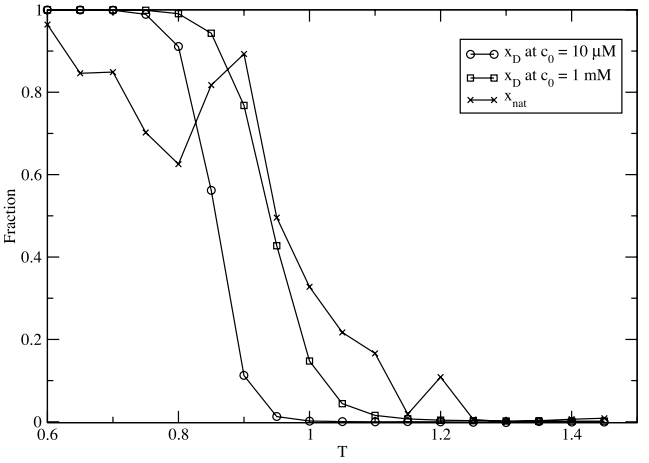

FIGURE 8 Equilibrium fraction of the dimer as a function of temperature. $x_D = (1 + 4K_{MD} - \sqrt{1 + 8K_{MD}})/4K_{MD}$, for the two concentrations indicated. The value of the equilibrium constant K_{MD} is shown in Fig. 6. We also show the fraction of configurations that, at the given temperature, had at least 50% native contacts.

when dimensionless ratios are computed, so that the absolute scale of energy (or temperature) drops out. Examples of such quantities include relative contributions to the total internal energy from the various interaction terms in the potential (Vieth et al., 1996), or their relative change upon mutation (Vieth et al., 1995).

It is also interesting to examine the contact map of the dimer phase as given by the simulation. A contact between residues belonging to different chains is considered native if it appears in the contact map of the native protein in its lattice representation (leading to a total of 27 contacts in our case). As discussed above, the only constraint in the simulation is that there be at least one contact between the two chains. Therefore, the question arises as to whether the dimer retains a significant fraction of native contacts in the vicinity of the transition temperature, or whether there is a significant fraction of out of register dimer configurations that are structurally very different from the native state. Fig. 7 shows the total number of inter-chain contacts as a function of temperature. There is a smooth decline from ~30 contacts at the lowest temperature to ~10 at the highest temperature. Therefore, the constraint used in the simulation leads to a considerable number of inter-chain contacts even at high temperatures. The same figure shows the fraction of the total number of contacts that are native. Because the fraction decays to below 0.1 at high temperatures, and there are ~10 contacts in total, there is on average less than one native contact per configuration at high temperatures. This can be contrasted with ~15 native contacts at low temperature.

An analysis that relies on the total number of contacts is necessarily biased at high temperature because of the constraint introduced to define the dimer phase. Furthermore, given that many residues involved in the inter-chain contact map are involved in two or more contacts, thermal fluctuations can lead to a large reduction in the number of native contacts even though the two monomers are substantially in register with one another. For these reasons, we have also calculated the ensemble average of the fraction of configurations that have *at least* 50% native contacts (although the details of the results are seen not to depend significantly on the choice of the percentage as long as it is not too close to 100%). The results are shown in Fig. 8 as a function of temperature. This fraction approaches one at low temperature, changes quickly around the transition region, and decays to zero at high temperatures. The dip or raise of ~20% around $T = 0.8-0.9$ is a manifestation of the statistical uncertainty in our calculation of the fraction. We also show in this figure the equilibrium fraction of the dimer form for the same two concentrations shown in Fig. 6. From the figure we conclude that in this range of temperature and concentration the decrease in the fraction of native contacts as given by our model can be mainly attributed to the appearance of the monomer form, and not to a significant contribution from out of register dimers.

To conclude, we have presented a numerical method to estimate the entropy loss of dimerization from a Monte Carlo simulation that is free from any assumptions or restrictions to the motion of the individual chains. We have considered the specific example of the GCN4 leucine zipper monomer-dimer equilibrium and shown that, in agreement with previous results (Vieth et al., 1996; Yu et al., 2001), the ideal gas approximation greatly overestimates the entropy loss of dimerization. Previous numerical simulations had also significantly overestimated this quantity. Compare, for example, the value of $V_1^{(2)} = 67.6\text{Å}^3$ given in Table 1 of Vieth et al. (1996), and the values shown in Fig. 2. Our results are also qualitatively consistent with experimental data (Yu et al., 2001), even though our model does not consider any explicit solvent effects, and that our calculations are restricted to very low concentrations so that monomer-dimer and dimer-dimer interactions are negligible.

Therefore, we cannot determine whether the smaller values of the translational and rotational entropy losses obtained in experiments relative to ideal gas statistics are due primarily to solvent effects or to chain interactions.

APPENDIX A

Lattice protein model and interaction force parameters

The model protein used in this work uses a reduced representation of the protein backbone on a regular lattice. The model comprises a sequence of bonds connecting particles located at the center of mass of the corresponding residue and backbone α-carbon. The particles are then placed in a three-dimensional simple cubic lattice with spacing of 1.45 Å. Further details on this model can be found in Kolinski et al. (1999).

A sequence of configurations is generated by a Monte Carlo scheme with Metropolis updating. The method uses three different types of individual transitions. In the first case, a single particle and its two corresponding bonds are selected for an attempted update. The second type of transition involves three consecutive bonds and the corresponding two adjacent particles. The third involves a rigid translation of a small fragment of the chain comprising three particles, and the ensuing rearrangement of the end bonds. These transitions are attempted sequentially for all the bonds in the chain. Two separate transitions are also included to adjust the position of the N- and C-termini particles. The set of all these attempted transitions constitutes a Monte Carlo step (MCS). Further details about the transitions used in the Monte Carlo updating have been given by Kolinski and Skolnick (1998).

Interaction potentials can be grouped into generic and sequence-specific. The former are sequence-independent and lead to protein-like structures, whereas the latter are derived from a statistical analysis of the protein database, and explicitly depend on the identity of the amino acids involved. We next list the values of the various parameters used in our calculations. Sequence-dependent short-range interactions are defined by Eq. 1 of Kolinski and Skolnick (1998). We use a common multiplicative factor in our calculations $\varepsilon_{\text{short}} = 0.325$ (this factor is explicitly shown in Eq. 12 of Kolinski et al. (1999) with a value of 0.75 instead). A three-body potential that is sequence-specific is also used, with an amplitude $\varepsilon_{3b} = 0.25$. The generic, short-range conformational biases of Eqs. 3–5 of Kolinski et al. (1999) involve $\varepsilon_{\text{gen}} = 1.25$. Hydrogen-bonding energies within the main chain are also included, with an amplitude $\varepsilon_{\text{H-bond}} = 0.325$ in Eq. 7 of Kolinski et al. (1999). Long-range and sequence-dependent interactions are modeled by a set of square well potentials as described in Eq. 9 of Kolinski et al. (1999). We have chosen $E^{\text{rep}} = 4$ in Eq. 8 of that reference, and a common multiplicative factor $\varepsilon_{\text{pair}} = 2.0$ (to be compared with the factor of 1.25 in Eq. 12 of Kolinski et al. (1999). Two additional multibody potentials are introduced to include hydrophobic effects and preferences for parallel or anti-parallel packing among the residues. We define as the scale of Eq. 10 in Kolinski et al. (1999) $\varepsilon_{\text{surface}} = 0.75$ (instead of the value 0.5 shown in Eq. 12 of that reference). Finally, we have used a factor $\varepsilon_{\text{multi}} = 0.75$ in Eq. 11 of Kolinski et al. (1999) (instead of the value 0.5 in Eq. 12).

This research was supported by National Science Foundation Grant 9986019. J.V. is also supported by National Institutes of General Medical Sciences Grant GM64150-01.

REFERENCES

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535.

Crick, F. H. C. 1953. The packing of α-helices: simple coiled-coils. *Acta Crystallogr.* 6:689–697.

d'Avignon, D. A., G. L. Bretthorst, M. E. Holtzer, and A. Holtzer. 1998. Site-specific thermodynamics and kinetics of a coiled-coil transition by spin inversion transfer NMR. *Biophys. J.* 74:3190–3197.

Ferrenberg, A. M., and R. H. Swendsen. 1989a. Optimized Monte Carlo data analysis. *Computers in Physics.* September/October:101–104.

Ferrenberg, A. M., and R. H. Swendsen. 1989b. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* 63:1195–1198.

Geyer, C. J. 1992. Practical Markov chain Monte Carlo. *Stat. Sci.* 7:437–483.

Hao, M.-H., and H. A. Scheraga. 1994. Monte Carlo simulation of a first order transition for protein folding. *J. Chem. Phys.* 98:4940–4948.

Harbury, P. B., T. Zhang, P. S. Kim, and T. Alber. 1993. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science.* 262:1401–1407.

Hill, T. L. 1986. An Introduction to Statistical Thermodynamics. Dover, New York.

Holtzer, M. E., G. L. Bretthorst, D. A. d'Avignon, R. H. Angeletti, L. Mints, and A. Holtzer. 2001. Temperature dependence of the folding and unfolding kinetics of the GCN4 leucine zipper via C-13(α)-NMR. *Biophys. J.* 80:939–951.

Hukushima, K., and K. Nemoto. 1996. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.* 65:1604–1608.

Kenar, K. T., B. García Moreno, and E. Freire. 1995. A calorimetric characterization of the salt dependence of the stability of the GCN4 leucine zipper. *Protein Sci.* 4:1934–1938.

Kihara, D., A. Kolinski, and J. Skolnick. 2001. TOUCHSTONE: An ab initio protein structure prediction method that uses threading-based tertiary restraints. *Proc. Natl. Acad. Sci.* 98:10125–10130.

Kolinski, A., P. Rotkiewicz, B. Ilkowski, and J. Skolnick. 1999. A method for the improvement of threading-based protein models. *Proteins: Struct., Funct., Genet.* 37:592–610.

Kolinski, A., P. Rotkiewicz, B. Ilkowski, and J. Skolnick. 2000. Protein folding: flexible lattice models. *Prog. Theor. Phys. Suppl.* 138:292–300.

Kolinski, A., and J. Skolnick. 1994. Monte Carlo simulations of protein folding. 1. Lattice model and interaction scheme. *Proteins: Struct., Funct., Genet.* 18:338–352.

Kolinski, A., and J. Skolnick. 1996. Lattice models of protein folding, dynamics and thermodynamics. R. G. Landes, Austin.

Kolinski, A., and J. Skolnick. 1998. Assembly of protein structure from sparse experimental data: an efficient Monte Carlo model. *Proteins: Struct., Funct., Genet.* 32:475–494.

Ma, S.-K. 1985. Statistical Mechanics. World Scientific, Singapore.

Mayer, J. E., and M. G. Mayer. 1963. Statistical Mechanics. John Wiley and Sons, New York.

Mohanty, D., A. Kolinski, and J. Skolnick. 1999. De novo simulations of the folding thermodynamics of the GCN4 leucine zipper. *Biophys. J.* 77:54–69.

O'Shea, E. K., J. D. Klemm, P. S. Kim, and T. Alber. 1991. X-ray structure of GCN4 leucine zipper, a two stranded, parallel coiled coil. *Science.* 254:539–544.

Skolnick, J., M. Vieth, A. Kolinski, and C. L. Brooks III. 1995. De novo simulations of the folding of GCN4 and its mutants. *In* Modeling of Biomolecular Structures and Mechanisms. A. Pullman, J. Jortner, and B. Pullman, editors. Kluwer Academic, The Netherlands.

Swendsen, R. H., and J.-S. Wang. 1986. Replica Monte Carlo simulation of spin glasses. *Phys. Rev. Lett.* 57:2607–2610.

Vieth, M., A. Kolinski, C. L. Brooks III, and J. Skolnick. 1995. Prediction of quaternary structure of coiled coils. Application to mutants of the GCN4 leucine zipper. *J. Mol. Biol.* 251:448–467.

Vieth, M., A. Kolinski, and J. Skolnick. 1996. Method for predicting the state of association of discretized protein models. Application to leucine zippers. *Biochemistry.* 35:955–967.

Yu, B. Y., P. L. Privalov, and R. S. Hodges. 2001. Contribution of translational and rotational motions to molecular association in aqueous solution. *Biophys. J.* 81:1632–1642.