De Novo Simulations of the Folding Thermodynamics of the GCN4 Leucine Zipper

Debasisa Mohanty,* Andrzej Kolinski,*# and Jeffrey Skolnick*

*Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037, USA, and #Department of Chemistry, University of Warsaw, 02-093 Warsaw, Poland

ABSTRACT Entropy Sampling Monte Carlo (ESMC) simulations were carried out to study the thermodynamics of the folding transition in the GCN4 leucine zipper (GCN4-lz) in the context of a reduced model. Using the calculated partition functions for the monomer and dimer, and taking into account the equilibrium between the monomer and dimer, the average helix content of the GCN4-lz was computed over a range of temperatures and chain concentrations. The predicted helix contents for the native and denatured states of GCN4-lz agree with the experimental values. Similar to experimental results, our helix content versus temperature curves show a small linear decline in helix content with an increase in temperature in the native region. This is followed by a sharp transition to the denatured state. van't Hoff analysis of the helix content versus temperature curves indicates that the folding transition can be described using a two-state model. This indicates that knowledge-based potentials can be used to describe the properties of the folded and unfolded states of proteins.

INTRODUCTION

Leucine zippers belong to the general class of protein structural motifs known as coiled coils (Crick, 1953; Cohen and Pary, 1990; O'Shea et al., 1991). Coiled coils occur in a diverse class of proteins ranging from fibrous proteins such as myosin, tropomyosin, and keratin (Johnson and Smillie, 1975; Phillips et al., 1986; Fraser and MacRae, 1971) to transcriptional activators (Landshultz et al., 1988; Harrison, 1991; Smeal et al., 1989) such as GCN4, Fos, and Jun. They are also present in many kinases (DiDonato et al., 1997). The coiled coil motif consists of amphipathic right-handed alpha helices wrapped around each other with a small left-handed superhelical twist (Crick, 1953).

The amino acid sequences of coiled coils consist of a characteristic heptad repeat $(abcdefg)_n$ (Cohen and Pary, 1990; Hodges et al., 1981; McLachlan and Stewart, 1975). Residues a and d are mostly hydrophobic and form a tightly packed core at the interhelical interface. Residues b, c, and f are mostly hydrophilic, while residues e and g are typically charged. It is believed that the identities of residues in the e and g position dictate the relative orientation of the helices in the coiled coil.

Coiled coil motifs having leucines in the d position in the heptad repeat are known as leucine zippers (Landshultz et al., 1988). They are quite short, typically 14 to 45 residues in length, unlike the long coiled coils in tropomyosin (Bairoch, 1990). Leucine zippers can form homo (O'Shea et al., 1991) or heterodimeric (O'Shea et al., 1991; Smeal et al., 1989) coiled coil structures and play an important role in subunit dimerization and subsequent DNA binding of transcription activators, such as GCN4 (Talanian et al., 1990). Apart from their biological importance, leucine zippers are also ideal systems for exploring questions related to protein folding (Alber, 1992; Harbury et al., 1993; O'Shea et al., 1993) and de novo protein design (Hodges et al., 1981; Su and Hodges, 1994; Monera et al., 1996; Lovejoy et al., 1993; O'Neil et al., 1990) because they represent the simplest example of quarternary structures in proteins. Therefore, leucine zipper motifs have been attractive targets for both experimental (O'Shea et al., 1991, 1993; Harbury et al., 1993; Lumb et al., 1994; Kenar et al., 1995; Sosnick et al., 1996; Zitzewitz et al., 1995; Lovett et al., 1996; Holtzer et al., 1997) and theoretical studies (Krystek et al., 1991; Nilges and Brunger, 1993; Zhang and Hermans, 1993; DeLano and Brunger, 1994; Vieth et al., 1994–1996). In this spirit, here we present a computational study of the folding thermodynamics of the GCN4 leucine zipper.

The high-resolution structure of the GCN4 leucine zipper, GCN4-lz, (O'Shea et al., 1991), has been obtained from x-ray crystallography. In the crystal state, GCN4-lz adopts a dimeric, parallel coiled coil structure (O'Shea et al., 1991). The effects of various mutations in the GCN4-lz sequence on the stability and state of oligomerization (Harbury et al., 1993; O'Shea et al., 1993) of the coiled coil structure have been investigated in detail using CD, equilibrium ultracentrifugation, and gel filtration. The thermal stability of various subdomains of GCN4-lz (Lumb et al., 1994) has also been determined. Thermodynamic (Kenar et al., 1995; Sosnick et al., 1996) and kinetic (Zitzewitz et al., 1995) studies that attempt to probe the nature of the conformational transition in the GCN4-lz indicate that the folding/unfolding transition is well described by a simple two-state model (Kenar et al., 1995; Sosnick et al., 1996), i.e., the equilibrium population is essentially comprised of only a native dimer and a completely denatured monomer. However, recent unfolding studies (Holtzer et al., 1997) on a GCN4-like leucine zipper indicate that the equilibrium na-

Received for publication 8 December 1998 and in final form 14 April 1999.

Address reprint requests to Dr. Jeffrey Skolnick, Dept. of Molecular Biology, TPC5, Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla, CA 92037. Tel.: 619-784-8821; Fax: 619-784-8895; E-mail: skolnick@scripps.edu.

0006-3495/99/07/54/16 $2.00

tive population is much richer than that suggested by a simple two-state model.

Surprisingly, compared to the large number of experimental studies, there have been relatively few theoretical studies (Krystek et al., 1991; Nilges and Brunger, 1993; Zhang and Hermans, 1993; DeLano and Brunger, 1994; Vieth et al., 1994–1996) on leucine zippers. Molecular mechanics (Krystek et al., 1991) and free energy perturbation (Zhang and Hermans, 1993) calculations on model leucine zipper structures indicate that leucines in the d position of the heptad make a major contribution to the stability of dimeric coiled coils. In contrast, residues at position a play a relatively minor role. Starting from a pair of parallel, in-register alpha helices, Nilges and Brunger (1991, 1993) used an all-atom force field and a molecular dynamics-based simulated annealing protocol to predict the coiled coil structure of the GCN4-lz. In their simulations (Nilges and Brunger, 1993), they used helical backbone hydrogen bond restraints and interhelical distance restraints. They predicted a structure whose backbone had a coordinate root mean squared deviation (RMSD) of 1.26 Å from the subsequently solved crystal structure (O'Shea et al., 1991) of GCN4-lz. Ab initio folding of the GCN4-lz, starting from a pair of random chains, was first attempted by Vieth and co-workers (Vieth et al., 1994). Using a high coordination lattice representation of the protein and knowledge-based potentials (Kolinski and Skolnick, 1994), they obtained interesting insights into the folding process and mechanism of chain assembly. The predicted native structures had a C_α RMSD from GCN4-lz crystal structure in the range of 2.3 to 3.7 Å. Subsequent refinement of these structures using an all-atom representation and a solvation shell resulted in a family of structures with a backbone heavy atom RMSD of 0.81 Å from the GCN4-lz crystal structure. Vieth and co-workers subsequently developed theoretical methods (Vieth et al., 1995, 1996) that could successfully predict the equilibrium between different oligomeric species of GCN4-lz, its mutants, and various subdomains. Hence, simulations based on a lattice protein representation and knowledge-based potentials have been successful in reproducing a number of experimental observations related to the structure and thermodynamics of the GCN4-lz.

In view of the encouraging results obtained from lattice-based computer simulations (Vieth et al., 1994–1996) on GCN4-lz, it is tempting to ask if such lattice models and knowledge-based potentials can be used for simulations of folding thermodynamics. In such a simulation, one would be interested in predicting the various microscopic conformational states that contribute to an observed macroscopic property at a given temperature. Hence, the result of such a simulation would directly tell us if the molecular population at transition temperature consists of a completely folded native dimer in equilibrium with a structureless denatured state, as suggested by simple two-state models (Kenar et al., 1995; Sosnick et al., 1996), or are more aptly described by a complex array of conformational states, as suggested by Holtzer and co-workers (Holtzer et al., 1997). However, the success of such a statistical thermodynamic calculation is crucially dependent on whether the knowledge-based potential cannot only describe the properties of the native state, but also the unfolded states of GCN4-lz. One minimal requirement is that such a model reproduce the conformational properties of the denatured state, in particular, its helix content. In Vieth's folding simulations (Vieth et al., 1994), the helix content of the isolated chains ranges from 30 to 35% and increases to 90% upon formation of the native dimer. In contrast, experiment (Holtzer et al., 1997) suggests that the helix content of thermally denatured chains is ~10–15%. The intrinsic secondary structure content of the denatured state is related to the relative weight of the local (short-range) versus the nonlocal (long-range) interactions. Therefore, it might be possible to reduce the intrinsic helix content in the denatured state by adjusting the relative strengths of the short- versus long-range interactions. However, in practice, reducing the intrinsic helix content for isolated chains increases the mean folding time. If the intrinsic helix propensity for isolated chains is too small, then the resulting native structure for the dimer may not be helical. Since the principal aim of the work of Vieth and co-workers (Vieth et al., 1994) was the ab initio prediction of native structure, they opted for an energy parametrization that allowed the folding of coiled coil sequences to their native state within a reasonable mean folding time, but the effect of this parametrization on the structure of the denatured state had not been fully explored.

In the present work we use lattice-based computer simulations for a complete thermodynamic characterization of the native and denatured states of the GCN4-lz. The specific questions we attempt to address are the following: Can we predict the experimentally observed conformational properties of both the native and denatured states of the GCN4-lz using the existing knowledge-based parameters? If not, can the agreement with experiment be obtained by a simple reweighting of the relative strengths of short- versus long-range interactions, or are more drastic changes in the energy functions required? If our parameters are capable of predicting the conformational properties of the native and denatured states of the GCN4-lz, we can also answer the following questions related to the thermodynamics of the folding transition. Is the folding transition adequately described by a two-state model, as suggested by experiment? If so, what is the origin of this behavior?

For a complete thermodynamic analysis of the various monomeric and dimeric states of the GCN4-lz, it is essential to estimate the entropies associated with the various energy states. Recent work (Kolinski et al., 1996) has demonstrated that the Entropy Sampling Monte Carlo (ESMC) technique of Hao and Scheraga (1994) can be used to obtain a reasonably accurate estimate of the entropies for 310 type lattice protein models. For a single chain, the ESMC technique suggested by Hao and Scheraga can be used in a straightforward way to calculate various thermodynamic properties. However, for multimeric systems, one needs to calculate the equilibrium between the monomer and the

multimer (Vieth et al., 1996); hence, a number of modifications to the original ESMC protocol are required. Therefore, to study folding/unfolding thermodynamics of GCN4-lz, we have carried out ESMC simulations with appropriate modifications for the two-chain system.

METHODS

In this section a brief description of the following is given: protein models, interaction scheme, ESMC formalism, and the formalism for obtaining the thermodynamic variables for a two-chain system from ESMC simulations.

Protein model

The (310) lattice model used here has been described in detail elsewhere (Kolinski and Skolnick, 1994). Each amino acid of a given polypeptide chain is represented by a set of two points corresponding to the position of the C_α and the center of mass of the amino acid side chain. The C_α positions are restricted to lattice points on an underlying cubic lattice, with a lattice spacing of 1.22 Å. The virtual bond vector connecting two consecutive C_α positions can take 90 possible orientations defined by the set of vectors {(3,1,1), . . . , (3,1,0), . . . , (3,0,0), . . . , (2,2,1), . . . , (2,2,0), . . . , . . .}. However, to exclude nonphysical conformations, there are further restrictions that limit two consecutive virtual bonds to an angle in the range of 72.5 to 154°. Using this high-resolution lattice representation, it is possible to build lattice models of proteins with an average C_α RMSD of 0.6 to 0.7 Å (Godzik et al., 1993) from the actual structure. The spheres representing the side chains are not confined to lattice points, and there are multiple rotamers for all amino acids except Gly, Ala, and Pro. The spatial resolution of the rotamers in the model are such that the center of mass of the side chain in a given rotameric state in a real protein is no farther than 1 Å from another member of the rotamer library.

Interaction scheme

The knowledge-based potential has the same functional form and parameters as in the work of Vieth and co-workers (Vieth et al., 1994, 1995). The total energy (E_{tot}) of the system consists of energy contributions from hydrogen bonds (E_{hb}), an effective Ramachandran potential (E_{R14}), a local side chain orientational coupling term (E_{beta}), a rotamer energy (E_{rot}), a burial energy (E_{one}), a pair potential (E_{pair}), and a cooperative pair potential (E_{temp}). The reader is referred to the work of Vieth and coworkers (Vieth et al., 1995) for a detailed description of these terms. The various interactions in the model can be divided into short- and long-range terms. The short-range interactions are E_{hb}, E_{hcoop}, E_{R14}, E_{beta}, and E_{rot} as they describe the local conformational preference in a polypeptide chain. Since we are considering helical structures, we classify E_{hb} as being short-range, but in β-proteins it may be long-range as well. The long-range interactions are E_{one}, E_{pair}, and E_{temp} that result from interactions between residues separated by at least three amino acids along a chain, but which are close in space. The structure and thermodynamics of a protein are obviously the result of the balance between the short- and long-range interactions. In our model, this balance is maintained by use of appropriate scale factors. Thus, the total energy (E_{tot}) in our model is given by

$$\begin{aligned} E_{tot} &= E_{long} + E_{short} \\ E_{long} &= S_{pair}{*}E_{pair} + S_{one}{*}E_{one} + S_{temp}{*}E_{temp} \\ E_{short} &= S_{hb}{*}E_{hb} + S_{R14}{*}E_{R14} + S_{beta}{*}E_{beta} + S_{rot}{*}E_{rot} \end{aligned} \quad (1)$$

In the work of Vieth et al., the scale factors were $S_{pair} = 5.0$, $S_{one} = 0.5$, $S_{temp} = 4.25$, $S_{hb} = 1.0$, $S_{R14} = 0.25$, $S_{beta} = 1.0$, $S_{rot} = 0.5$. Here, we carried out one set of simulations using these scale factors. This will be referred to as parameter set I. We then carried out another set of simulations using $S_{pair} = 5.0$, $S_{one} = 0.5$, $S_{temp} = 4.25$, $S_{hb} = 0.5$, $S_{R14} = 0.10$, $S_{beta} = 0.5$, $S_{rot} = 0.5$. This will be referred to as parameter set II. In parameter set II, the lower scale factors for hydrogen bond, orientational coupling, and Ramachandran potential reduced the weight of short-range interactions relative to the long-range interactions. As will be discussed later in the Results section, this reduction in the weight of short-range interactions significantly reduced the helix content of the denatured state and made the model in better agreement with experiment.

Conformational sampling scheme and move set

The most widely used method of conformational sampling for lattice models is the asymmetric Metropolis Monte Carlo (MMC) scheme (Metropolis et al., 1953). Here, the transition probability from a conformation with energy E_i to that with energy E_j is given by

$$P_{ij}^{MMC} = \min\{1, \exp(-\Delta E_{ij}/kT)\} \quad (2)$$

Here, $\Delta E_{ij} = E_j - E_i$, k is Boltzmann's constant, and T is the simulation temperature. Since the transition probability is decided by the energy difference between the two conformations, this method is sensitive to the presence of energy barriers. Hence, to achieve an adequate sampling of representative structures at all energy states, which is essential for any statistical thermodynamic calculation, the MMC protocol needs a sufficiently complete move set (Kolinski and Skolnick, 1994) and the simulation must be carried out over a range of temperatures. Even then, there is no obvious way to demonstrate convergence.

As elegantly demonstrated in a series of recent theoretical work (Hao and Scheraga, 1994a,b, 1995; Kolinski et al., 1996), the ESMC scheme not only overcomes these drawbacks of MMC, but offers several other advantages; in particular, it provides an objective measure of convergence. ESMC can be formulated in several different ways (Lee, 1993; Berg and Neuhaus, 1991; Hao and Scheraga, 1994a), some of which are very similar to the multicanonical MC (Hansmann and Okamoto, 1993; Okamoto and Hansmann, 1995). Here, the ESMC scheme as proposed by Hao and Scheraga (1994a,b) was used. The transition probability from a conformation with energy E_i to a conformation with energy E_j is given by

$$P_{ij}^{ESMC} = \min\{1, \exp(-\Delta S_{ij}/k)\} \quad (3)$$

Here, k is the Boltzmann constant, $\Delta S_{ij} = S(E_j) - S(E_i)$, and $S(E_i)$ and $S(E_j)$ are the entropies associated with energy states E_i and E_j, respectively. Thus, ESMC generates an artificial distribution of states controlled by their relative entropies.

At the beginning of the simulation, the entropies of the various states, $S(E)$, are not known. Thus, the simulation is started with an approximate guess for the entropy, $J(E)$, which could be a constant. The sampled conformations are then used to obtain a density of states histogram, $H(E)$. In turn, $H(E)$ is used to get a better approximation for the entropy. If the kth iteration consists of an ESMC run with $S(E)$ approximated by $J_{k-1}(E)$, then the new approximation for $S(E)$, $J_k(E)$ is given by

$$J_k(E) = J_{k-1}(E) + \ln(\max\{1, H_k(E)\}) \quad (4)$$

After a sufficient number of iterations, when convergence is achieved, all energy states are equally sampled and the density of states histogram becomes flat. Thus, one has a means of determining whether or not convergence was achieved. After convergence, the $J(E)$ curve shifts by a constant in successive iterations and the true entropy, $S(E)$, of the system is given by $J(E) + C$, where C is some constant.

If the system gets trapped in low entropy states, it may take a very large number of iterations to achieve convergence. Kolinski et al. (1996) have suggested a means to accelerate the convergence of ESMC simulations by making use of a "conformational pool." The conformational pool is essentially a library of seed structures, which can be obtained from initial ESMC iterations or from MMC folding or unfolding simulations. Then, during the ESMC run, periodically, one randomly selects a new structure from the conformational pool and accepts it subject to Eq. 3. Then, if successful, the

ESMC simulation is continued with the structure picked from the pool. Thus, by picking new starting structures from the conformational pool, one can randomly shift the system between distant energy levels, and yet detailed balance is maintained. This way, the sampling process can surmount possible entropic barriers. The other advantage of using a conformational pool is that the long ESMC iteration essentially becomes a series of short runs, each of which starts with an independent new conformation. Thus, the computational task involved in simulation can be easily parallelized on multiprocessor machines.

In the present work the simulations were carried out using eight processors of a CRAY T3E. At the beginning of the simulation, the processors randomly picked structures from different regions of conformational space. Starting from these structures, each processor executed several cycles of Monte Carlo moves (Vieth et al., 1995) using the transition probability given by Eq. 3. Each cycle consisted of $(N-2)*M$ two-bond moves, $(N-3)*M$ three-bond moves, $2*M$ chain end moves, M shifts of chain pieces, and M rigid body shifts and rotamer rearrangements after each backbone conformational change. Here, M is the number of chains in the system and N is the number of beads in each chain. In our case, N is 35, including two dummy beads at the ends, and M is one or two, depending on whether the simulation is for a monomer or dimer. As mentioned before, the main task in an ESMC simulation is to sample different conformations of the molecule and construct the density of states histogram (Hao and Scheraga, 1994a). Hence, the conformations sampled by each processor were stored at 50-cycle intervals to build the density of states histogram. A bin size of 4 kT was used in the construction of the density of states histogram. The choice of this bin size is based on the work of Kolinski et al., where it had been observed that a coarse-grained histogram accelerates sampling, but there was no substantial loss of conformational resolution (Kolinski et al., 1996).

Starting from the initial structure, each processor executed 25,000 MC cycles and generated 500 conformations for building the density of states histogram. Then, each processor randomly picked a new structure from the conformational pool and executed another 25,000 MC cycles to generate another 500 conformations for the density of states histogram. This process of picking a new initial structure from the pool and collecting conformations for density of states histograms was repeated 10 times until each processor collected a total of 5000 conformations. Then, the individual histograms obtained from all eight processors were combined and the resulting histogram, consisting of $(5000 \times 8 =)$ 40,000 points, was used to update the entropy curve using Eq. 4, and the updated entropy was communicated to the individual processors. This entire process will be referred to as one communication cycle, which essentially consists of 80 short ESMC runs, each of which began from a different starting structure. The next communication cycle was executed on the eight processors of CRAY T3E using the updated entropy obtained from the previous communication cycle. One iteration of an ESMC run consisted of four such communication cycles, requiring ~28 h of CPU time on eight processors of CRAY T3E for one iteration of an ESMC run for the dimer.

The convergence of the simulation was estimated by the following function:

$$\mathrm{Err}^{k}(E_i) = \frac{100*\mathrm{abs}(\Delta J(E_i) - \Delta J_0)}{\Delta J(E_i)} \tag{5a}$$

Here, $\mathrm{Err}^{k}(E_i)$ is the error in convergence in the ith energy bin in the kth iteration.

$$\Delta J(E_i) = J_k(E_i) - J_{k-1}(E_i) \tag{5b}$$

$$\Delta J_0 = \sum \Delta J(E_i)/\mathrm{nbin} \tag{5c}$$

and nbin is the total number of bins over which statistics have been collected in the kth iteration.

The ESMC simulations were considered converged when for successive iterations, $\mathrm{Err}(E_i)$ was <5% in each energy bin. The total computational cost to build up the entropy curve for the dimer from scratch was ~600 CPU hours on eight processors of a CRAY T3E, and it took 20 iterations. For the monomer, it was ~200 CPU hours on four processors of a CRAY T3E.

Generalization of ESMC for a two-chain system and formalism for calculation of thermodynamic parameters

In this section we describe the required modifications to the ESMC protocol for calculating thermodynamic variables for a two- (or multi) chain system.

Let us first consider the monomer. In order to calculate the equilibrium constant, one must separate the translational, rotational, and internal degrees of freedom of the chain (Davidson, 1962; Herschbach, 1959). Thus, we fix the first C_α in space and perform ESMC simulations. The entropy for the monomer, $S_M(E)$, obtained from ESMC simulations contains contributions from the internal conformational entropy ($S_{\text{int-conf,M}}$) and rotational entropy ($S_{\text{rot,M}}$). The rotational entropy of the monomer is related to the number of conformational states sampled for the first two C_α bond vectors, Ω_M, by the equation,

$$S_{\text{rot,M}} = \ln \Omega_M \tag{6}$$

and can be calculated from the manifold of structures obtained from a converged ESMC run. Thus, internal conformational entropy for the monomer, $S_{\text{int-conf,M}}$, can be obtained by the equation,

$$S_{\text{int-conf,M}} = S_M(E) - S_{\text{rot,M}} \tag{7}$$

Hence, the monomer configurational partition function, $Z_{\text{int-conf,M}}$, is given by,

$$Z_{\text{int-conf,M}} = \sum \exp(-E/kT + S_{\text{int-conf,M}}(E)) \tag{8}$$

However, $Z_{\text{int-conf,M}}$ can be constructed only to within a constant because of the arbitrary constant associated with the entropy obtained from ESMC.

ESMC simulations are then carried out for the dimer. The dimer is defined as any configuration of two chains containing at least one interchain side chain contact. Hence, in the dimer simulation, configurations without any interchain contacts are rejected. During the simulations, the first bead of chain 1 in the dimer is kept fixed in space so as to exclude the translational entropy. The entropy of the dimer, $S_D(E)$, obtained from ESMC simulations consists of three classes of terms:

1. The translational entropy of bead 1 in chain 2, which is in fact related to the average volume accessible to this bead, $\langle V \rangle$, by

$$S_{\text{trans,D}} = \ln\langle V \rangle \tag{9}$$

2. The rotational entropy of the dimer, which is related to the number of accessible conformations for the first two C_α bond vectors in chains 1 and 2, Ω_{D1} and Ω_{D2}, by

$$S_{\text{rot,D}} = \ln(\Omega_{D1}*\Omega_{D2}) \tag{10}$$

3. The conformational entropy arising from internal configurations of both chains in the dimer, $S_{\text{int-conf,D}}$.

Therefore, $S_{\text{trans,D}}$ and $S_{\text{rot,D}}$ can be calculated from the manifold of structures obtained from a converged ESMC simulation for the dimer using Eqs. 8 and 9, and the internal conformational entropy for the dimer can be obtained by

$$S_{\text{int-conf,D}}(E) = S_D(E) - S_{\text{trans,D}} - S_{\text{rot,D}} \tag{11}$$

Then, within a constant, the internal configurational partition function of the dimer is given by

$$Z_{\text{int-conf,D}} = \sum \exp(-E/kT + S_{\text{int-conf,D}}(E)) \tag{12}$$

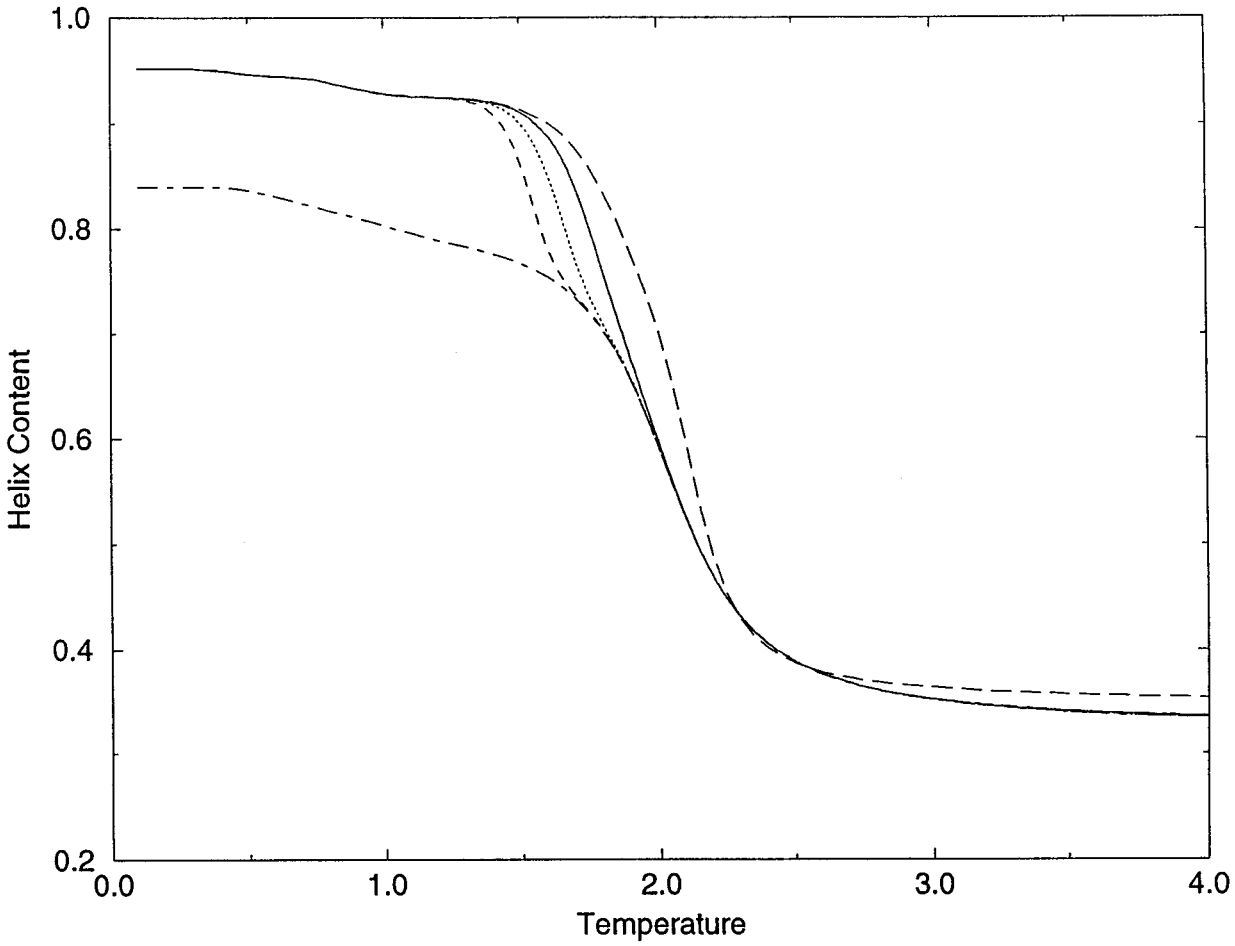

FIGURE 1 The average helix content versus temperature curves for simulations with parameter set I. The dashed and dashed-dotted lines represent the average helix contents of the dimer and monomer, respectively. The dashed, dotted, and solid lines represent the overall helix contents at chain concentrations of 2 μM, 43 μM, and 1 mM, respectively.

Now that we have calculated the total internal partition functions of the monomer and dimer, any internal thermodynamic or conformational property, W, can be obtained from

$$\langle W\rangle = \frac{\sum W(E)\exp(-E/kT + S_{\text{int-conf,L}}(E))}{Z_{\text{int-conf,L}}} \tag{13}$$

Here, $L = D$ or M.

At this point, we have the total internal partition functions of the monomer and dimer to within a constant. The problem is that the constant provided by ESMC is arbitrary; thus, we need a procedure to relate the partition function of the dimer to that of the monomer. This is equivalent to demanding that all species have the same reference state. If so, then the equilibrium constant follows from Herschbach (1959) and Mayer and Mayer (1963)

$$K_{\text{MD}} = \frac{\langle V\rangle \Omega_{\text{D1}}\Omega_{\text{D2}}Z_{\text{int-conf,D}}}{\sigma_D \Omega_M^2 Z_{\text{int-conf,M}}^2} \tag{14}$$

Here, σ_{D} is the symmetry number and equals 2 for a homodimer and 1 for a heterodimer.

Therefore, we have to place the monomer and dimer in the same reference state. We first observe that for internal configurational contributions in the limit of very large energies, we would expect that

$$S_{\text{int-conf,D}}(E) = 2S_{\text{int-conf,M}}(E_{\text{M}}) \tag{15}$$

Here, $S_{\text{int-conf,D}}(E)$ is the internal configurational entropy of the dimer with energy E, $S_{\text{int-conf,M}}(E_{\text{M}})$ is the internal configurational entropy of the monomer having an energy per chain, E_{M}, and $E = 2E_{\text{M}}$. This equation simply says that in very high energy states, the internal configurational entropy accessible to a chain in the dimer should be the same as that in a monomer. By demanding that Eq. 15 hold in the limit of large energies, we can shift the entropy curves of the monomer and dimer systems so that they have the same reference state.

The equilibrium constant K_{MD} can be used to calculate the fraction of all chains, X_{D}, that exist as a dimer at any given chain concentration, C_0, using the relation (McQuarrie, 1976)

$$x_{\text{D}} = \frac{(1 + 4K_{\text{MD}}C_0) - \sqrt{1 + 8K_{\text{MD}}C_0}}{4K_{\text{MD}}C_0} \tag{16}$$

The average overall helix content over the entire range of temperatures can be calculated from the equation

$$\theta(T) = x_{\text{D}}(T)\theta_{\text{D}}(T) + 1 - x_{\text{D}}(T))\theta_{\text{M}}(T) \tag{17}$$

Here, $x_{\text{D}}(T)$ is the mole fraction of dimers, $\theta(T)$ is the overall helix content, and θ_{D} and θ_{M} are the average helix contents for the dimer and monomer, respectively, at the temperature T. They are calculated from the manifold of structures for the dimer and monomer using Eq. 13.

A macroscopic measure of how well a given thermodynamic transition is described by a two-state model can be obtained from a van't Hoff

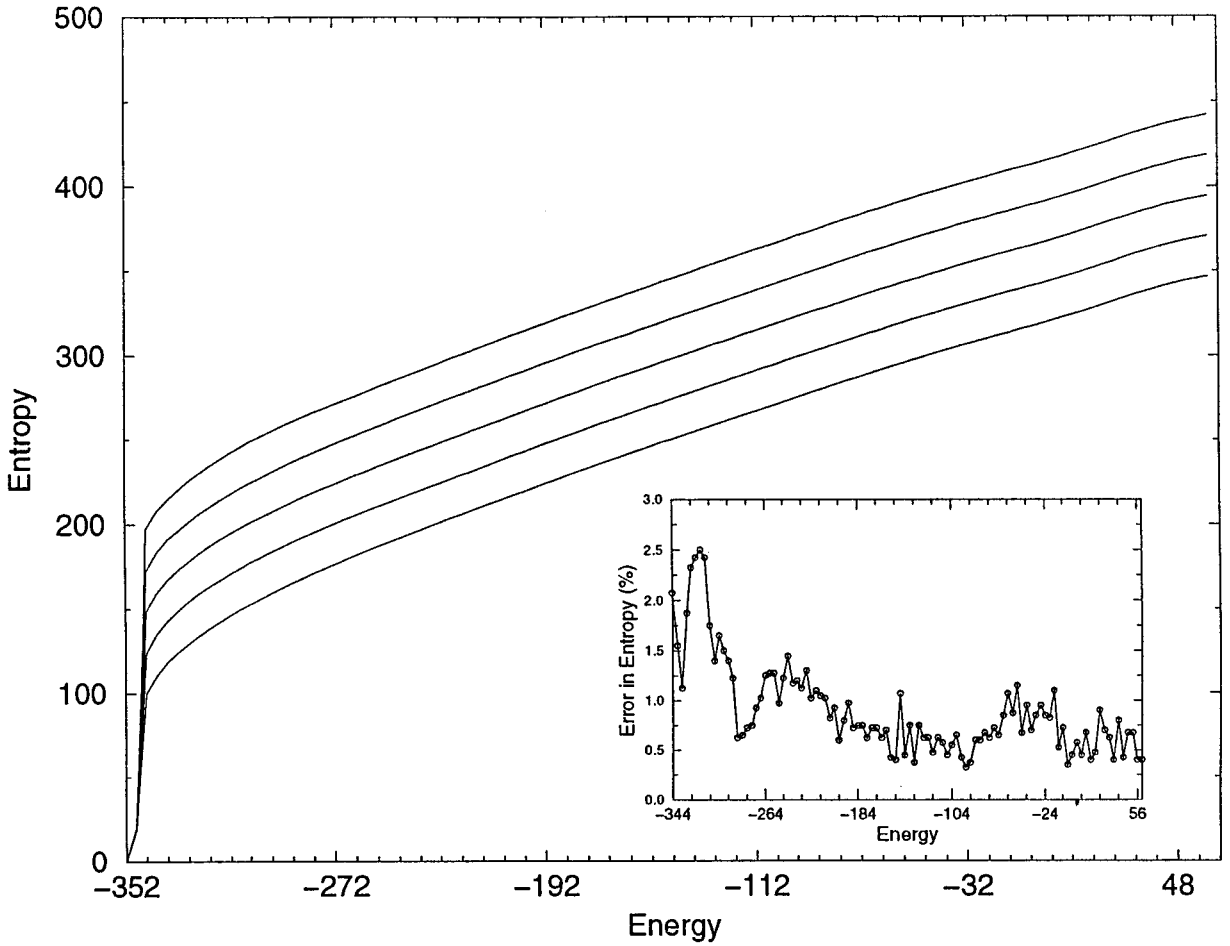

FIGURE 2 Entropy of the dimer of GCN4-lz as a function of its energy. The five different entropy curves have been obtained from last five successive iterations of the ESMC run after convergence of the simulations. The inset shows the error in entropy plotted as a function of the energy.

analysis (Privalov and Gill, 1988; Marky and Breslauer, 1987) of the $\theta(T)$ versus T curve. In the van't Hoff analysis, one computes the apparent fraction of the native dimer, $f_N(T)$ from the $\theta(T)$ versus T curve, assuming that a two-state model holds good for the folding transition. Thus, it is assumed that at any temperature T, the overall helix content, $\theta(T)$ is given by

$$\theta(T) = f_N{*}\theta_N + 1 - f_N){*}\theta_D \tag{18}$$

Hence, $f_N(T)$ can be extracted from the $\theta(T)$ versus T curve using the relation

$$f_N(T) = (\theta(T) - \theta_D)/(\theta_N - \theta_D) \tag{19}$$

Here, θ_N and θ_D are the helix contents of the native and denatured state, respectively, and their values can be obtained from the low and high temperature plateau regions of the $\theta(T)$ versus T curve. However, in many cases, the regions representing the native and denatured states in the $\theta(T)$ versus T curve show a linear decline of helix content with increase in temperature. In such a case, the values of θ_N and θ_D can be chosen in two different ways. One possible way is to assume that θ_N and θ_D change linearly with temperature, and hence their values at any temperature are obtained by fitting straight lines to the native and denatured regions and extrapolating those straight lines to the transition region. This method is known as "baseline fitting" (Marky and Breslauer, 1987). The second possible way is to take the values of $\theta(T)$ at the lowest and highest temperatures as the values for θ_N and θ_D, respectively. This method does not involve any baseline fitting, and it is assumed that θ_N and θ_D do not change with temperature. Hence, for the van't Hoff analysis, the apparent fraction of the native dimer, $f_N(T)$, can be obtained in two different ways, i.e., with and without baseline fitting.

The values for the fraction of native dimer, f_N, can be used to calculate the apparent equilibrium constant, K_{app}, for the transition from the native to the denatured state using the equation

$$K_{app}(T) = f_N/(1 - f_N)^2 \tag{20}$$

The van't Hoff enthalpy can be calculated from the $K_{app}(T)$ versus T curve using

$$\Delta H_{VH} = \frac{d(\ln K_{app})}{dT} kT^2 \tag{21}$$

The calorimetric enthalpy for the transition can be computed using the equation

$$\Delta H_{cal} = \langle E_D \rangle - 2\langle E_M \rangle \tag{22}$$

Here, $\langle E_D \rangle$ and $\langle E_M \rangle$ are the average energies of the dimer and monomer, respectively, at the transition temperature. They can be computed for the monomer and dimer using Eq. 13.

If the transition is in fact two-state, then the van't Hoff enthalpy should be equal to the calorimetric enthalpy obtained directly from Boltzmann averaging of energy over the microscopic states (Privalov and Gill, 1988; Marky and Breslauer, 1987). Hence, ΔH_{VH} can be compared to ΔH_{cal} to

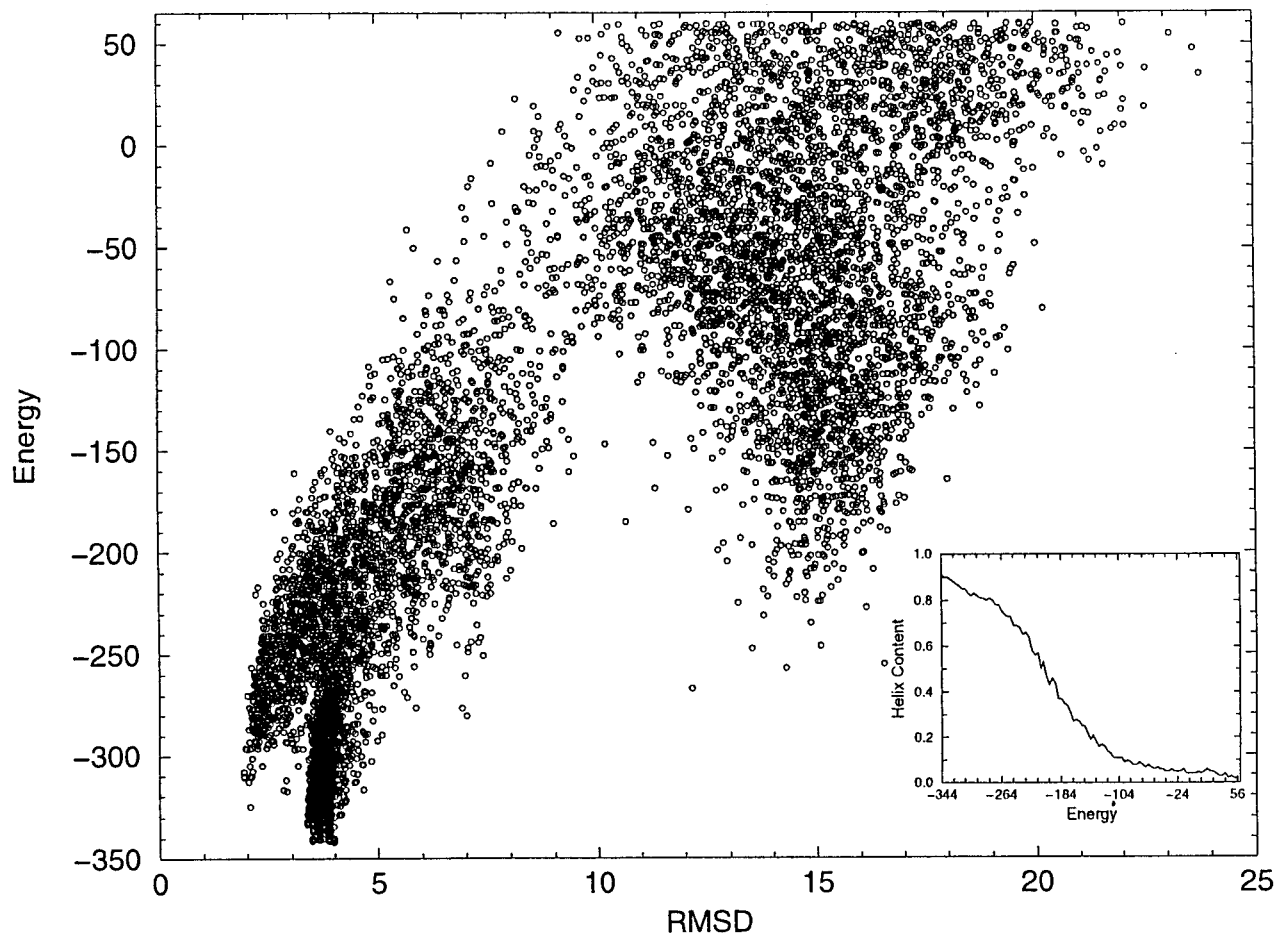

FIGURE 3 Plot of RMSD (from the native GCN4-lz) versus energy for the structures obtained from a converged ESMC run for the dimer. The inset shows the average helix content of these dimer structures as a function of their energy.

estimate how well the two-state model fits the folding transition in our simulation for the GCN4-lz.

RESULTS

As mentioned in the Methods section, we carried out two sets of ESMC simulations for the monomers and dimers of the GCN4-lz. In parameter set I, we used exactly the same scale factors as in the work of Vieth and co-workers (1995). Here, we present the final results for simulations with parameter set I, skipping the details. As described in the Methods section, the entropy versus energy curves for the monomer and dimer of the GCN4-lz were obtained from ESMC simulations, and the entropy curves were used to calculate the partition functions for the monomer and dimer. The equilibrium between monomer and dimer at different temperatures was calculated using the partition functions. The mole fractions of the monomer and dimer at different chain concentrations were calculated from equilibrium constants. The overall helix content for the GCN4-lz was computed using the average helix contents of the dimer, average helix content of the monomer, and mole fractions of the monomer and dimer.

Fig. 1 shows the average helix content for the monomer and dimer of the GCN4-lz at various temperatures and the overall helix content of the GCN4-lz at different chain concentrations. As seen in the figure, the monomer helix content is far too high, being close to 85% at the lowest temperature. At 2 μM chain concentration, after the complete dissociation of two chains, the monomer is 70% helical. This indicates that the GCN4-lz undergoes a folding/unfolding transition with a helical monomer as an intermediate, which is completely at variance with experimental observations (Kenar et al., 1995; Sosnick et al., 1996). Even at 1 mM chain concentration, where dissociation takes place at a substantially higher temperature, this parameter set predicts the existence of a monomer that is $>$50% helical. Thus, it is clear that this parameter set predicts too high a helix content in the denatured state.

Therefore, we decided to explore whether it is possible to reduce the helix content in the denatured state and at the same time correctly predict the native dimer conformation. Hence, we carried out simulations with parameter set II, where we reduced the weight of its short-range interactions relative to the long-range interactions by a simple adjust-

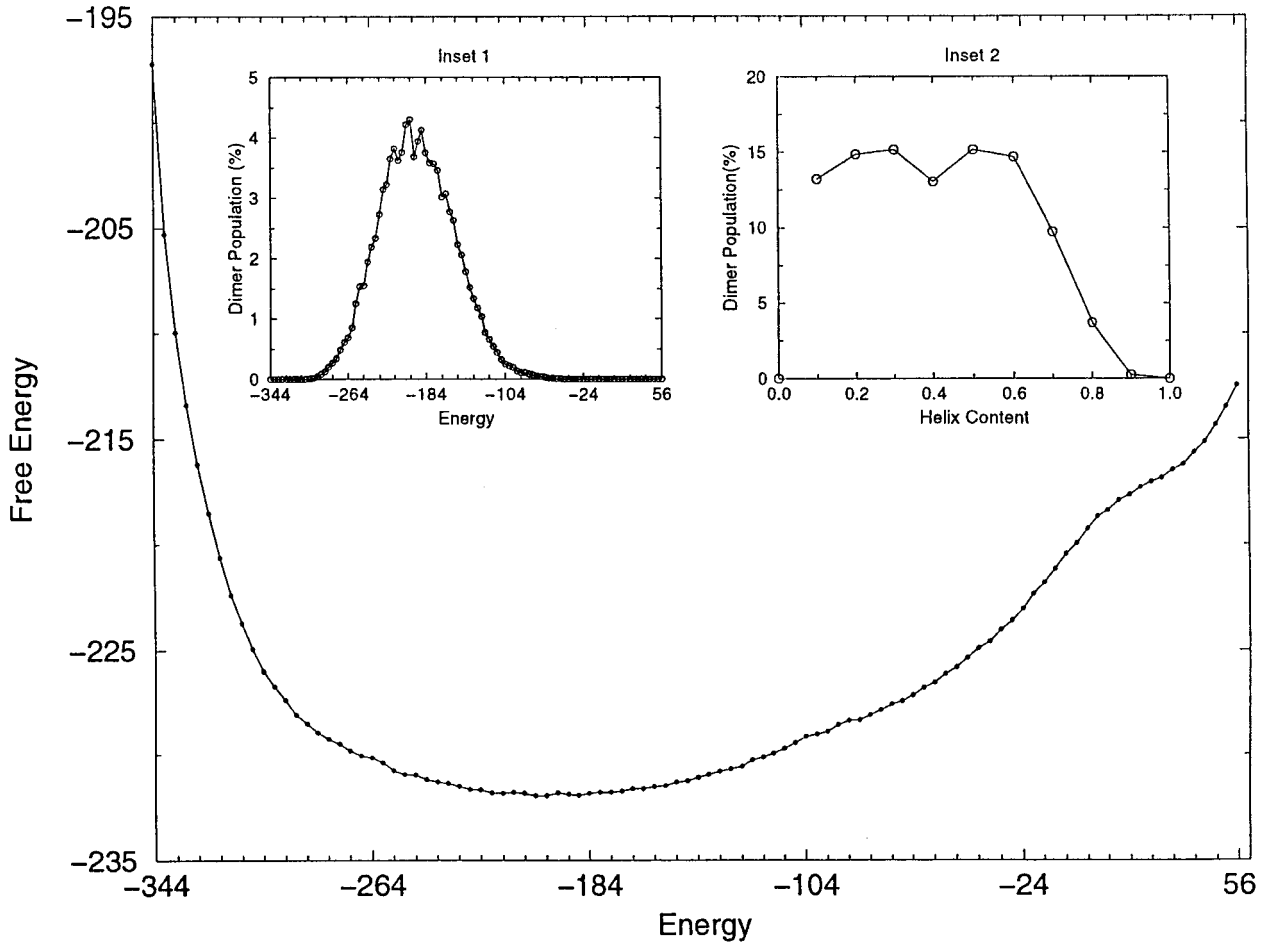

FIGURE 4 Free energy, $F(E)$, of the GCN4-lz dimer as a function of energy, E, at $T = 1.75$, which is the transition temperature for the conformational transition within the dimer. The insets show the population of GCN4-lz dimers as a function of energy and helix content at the transition temperature.

ment of the scale factors of certain energy terms in our potential. Below, we discuss in detail the results of simulations with parameter set II.

Fig. 2 shows the relative entropy versus energy curves for the dimer obtained from a series of converged ESMC runs carried out with parameter set II. The manifold of structures obtained from these converged ESMC runs span the energy range -348 to 60 kT, with the structures in the lowest energy bin rarely sampled. As seen in Fig. 2, the entropy curves have shifted by a constant amount in successive iterations in all bins except for the lowest energy bin at -348 kT. This indicates that we have achieved convergence everywhere except in the lowest energy bin. The inset to Fig. 2 shows the average error in the entropy in each bin. For each converged iteration, the error in a given energy bin was computed using Eq. 5a, and these were averaged over the last five iterations shown in Fig. 2 to obtain the average error in each of the energy bins. As can be seen in Fig. 2, the errors are larger toward the lowest energy bins, but the maximum error is 2.5%, with a $<1\%$ average overall error. This indicates good convergence of the ESMC simulation for the dimer. The error values shown in the figure were expressed as a percentage of the shift in entropy in successive iterations. Since the entropy curves shown in Fig. 2 shift by ~ 23 k in successive iterations, the maximum error in entropy is 0.57 k, and the average error in entropy is <0.23 k.

In Fig. 3 we show a plot of the energy versus RMSD (from native GCN4) for the structures obtained from a typical converged ESMC simulation for the dimer. As can be seen, the lowest energy structures typically have an RMSD ranging from 3.5 to 4 Å from the native GCN4, and the lowest energy misfolded structures are at least 80 kT above the lowest energy minimum. These lowest energy misfolded structures consist of two helical hairpins with very few contacts between the two chains. There is a large population of structures in the range of 2–3 Å from native GCN4, but these are at least 20 kT in energy above the lowest energy structure. Overall, there is a significant correlation between RMSD and energy. The clustering of points at lower energies is a reflection of the lowering of the entropy as one goes to lower and lower energy states. It is important to note that the lowest energy structures obtained from the simulation are close to the native structure for GCN4-lz. Thus, the model with diminished short-range interaction contributions also recovers the native state. This

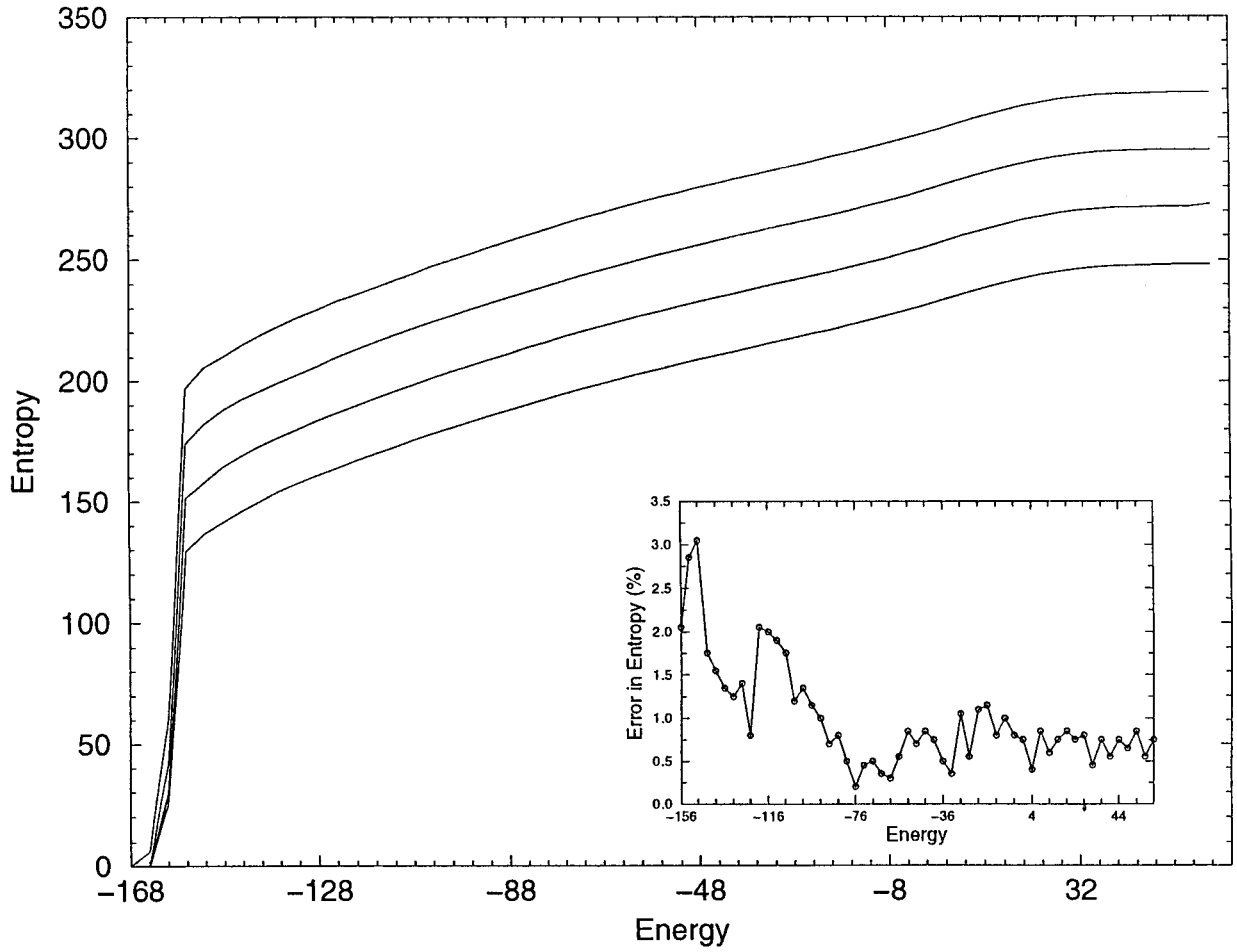

FIGURE 5 Entropy of the monomer of GCN4-lz as a function of its energy. The four different entropy curves have been obtained from the last four successive ESMC iterations after convergence has been achieved. The inset shows the error in entropy as a function of energy.

also can be seen in the inset to Fig. 3, which shows the average helix content of the structures in various energy states, with the lowest energy state having a helicity of 91% and the very high energy states close to 2%.

At any temperature T, the microscopic free energy, $F(E, T)$, for the dimer can be computed from the $S(E)$ curve for the dimer, using the expression $E - TS(E)$ (Hao and Scheraga, 1994a,b; Kolinski et al., 1996). At the transition temperature, the low and high energy states must have equal free energy. Thus, the transition temperature for the conformational transition within the dimer was obtained by demanding that low and high energy states of the dimer have equal free energy. Fig. 4 shows the free energy $F(E, T)$ versus energy curve for the GCN4-lz dimer at $T = 1.75$, which is the transition temperature for the conformational transition within the dimer. It must be noted that the absolute value of the free energies is arbitrary because the entropy has an arbitrary constant. But the most important feature that can be seen in Fig. 4 is that there is no free energy barrier separating the native state and the high energy states with much lower helix content. This indicates that the conformational transition within the dimer is continuous, and that a large number of conformational states will be populated at the transition temperature. This can be seen from the insets to Fig. 4. Inset 1 shows the population at various energy states of the dimer at the transition temperature for the conformational transition within the dimer. Inset 2 shows the population of the dimer as a function of helix content. As can be seen from inset 2, at the transition temperature a large number of states with helix contents varying from 10% to 70% will be significantly populated.

Even though the free energy profile of the dimer indicates the presence of a large number of conformational states, it is important to note that we defined the dimer as any configuration of two chains with at least one interchain side chain contact, which is rather arbitrary. However, at any given temperature, whether or not these dimeric states will be populated will depend on their relative weight compared to monomeric states. It is possible that a dissociation of chains takes place at a temperature lower than 1.75, possibly rendering the distributions shown in Fig. 4 not relevant for the GCN4-lz system. The true distribution of conformational population at any temperature can be determined only by calculating the equilibrium between monomer and dimer.

Fig. 5 shows the relative entropy versus energy for the

monomer, obtained from an ESMC simulation using parameter set II. The various monomer states obtained from the ESMC simulation span an energy range of -164 to 60 kT, but convergence could be achieved only in energy bins up to -156 kT. The results shown here are from the last four converged iterations. In the inset to Fig. 5 we show the average errors in monomer entropy for each of the energy bins. The errors were calculated in the same way as for the dimer and finally averaged over the last four converged iterations. As can be seen from Fig. 5, the errors are larger toward the lowest energy bins, but the maximum error is $\sim$3%, and the average overall error is $<$1%. This indicates that we have also achieved good convergence for the monomer simulation.

An analysis of the monomer structures sampled in a converged ESMC run indicates that the lowest energy structure for the monomer is only 40% helical, and structures having energy higher than -148 kT have an average helix content of $<$20%, dropping to 2% in the highest energy states. Hence, in general, the GCN4 monomer does not seem to favor a highly helical structure even at the lowest temperature. The isolated monomer is likely to have an average helix content of $<$40%. However, at any temperature, whether or not the monomer will be present in the overall population of GCN4-lz can be known only from the calculation of equilibrium constant.

As discussed in detail in the Methods section, to calculate the equilibrium between monomer and dimer we must shift the entropies of the monomer and dimer to the same reference state. Fig. 6 shows the entropies of the monomer and dimer shifted to the same reference state. The solid line is the internal entropy (after subtraction of translation and rotation components), $S_{\text{int-conf,D}}(E)$, versus energy (E) of the dimer. The dashed lines indicate twice the internal entropy for the monomer, i.e., $2S_{\text{int-conf,M}}(E_{\text{M}})$, plotted as a function of E^1, with $E^1 = 2E_{\text{M}}$ and E_{M} being the energy of the monomer. The dashed line essentially represents the internal entropy versus energy curve for a hypothetical dimer consisting of two independent monomers. E^1 varies from -312 to 120 kT, while E varies from -344 to 60 kT. The dimer curve has been shifted by a constant so that at $E = 60$ kT, both curves match. As seen in Fig. 6, from $+60$ to $+16$ kT, both curves almost overlap, indicating that in those energy regions, the dimer essentially behaves as two inde-

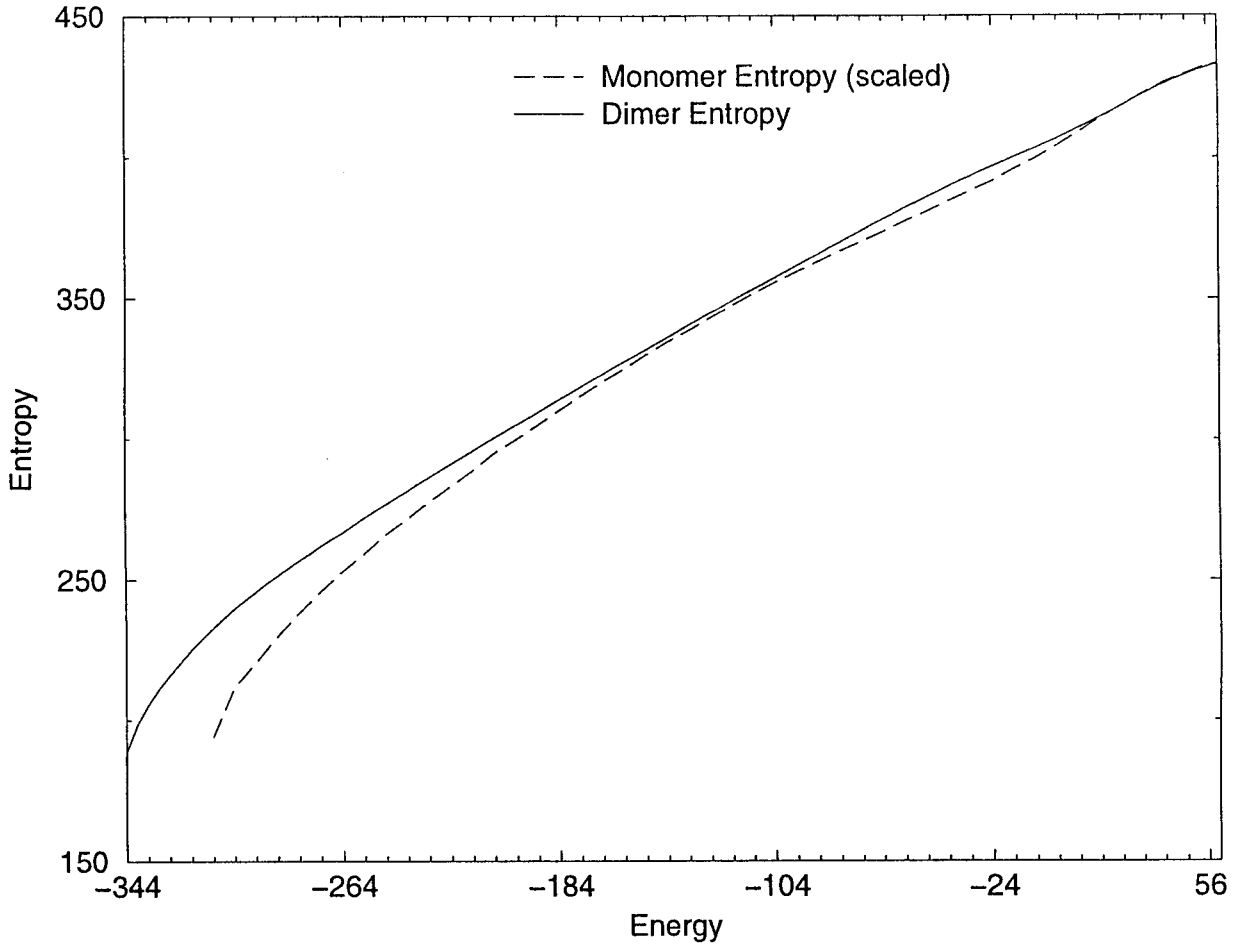

FIGURE 6 The solid line represents the internal entropy of the dimer, $S_{\text{int-conf,D}}(E)$, plotted as a function of its energy, E. The dashed line represents twice the internal entropy of the monomer, $2S_{\text{int-conf,M}}(E_{\text{M}})$, plotted as a function of E^1, with $E^1 = 2E_{\text{M}}$ and E_{M} being the energy of the monomer. The curves have been shifted so that in the limit of very large energies we have $S_{\text{int-conf,D}}(E) = 2S_{\text{int-conf,M}}(E_{\text{M}})$ and $E = 2E_{\text{M}}$.

pendent chains, while the deviation between the two entropy curves increases as the energy becomes lower and lower.

After shifting the internal entropies of the monomer and dimer to have the same reference state, the internal partition functions for the monomer and dimer were computed at various temperatures. Hence, for the monomer-dimer equilibrium, the dominant species at a given temperature was determined from the ratio of their respective partition functions as described in detail in the Methods section. Fig. 7 shows the variation of dimer mole fraction, x_D, with temperature for chain concentrations of 2, 43, and 300 μM. As can be seen, at a 2 μM chain concentration, while the population is entirely dimeric below $T = 1.2$, it becomes essentially monomeric above $T = 1.6$. $T = 1.45$ is the transition midpoint where both monomer and dimer are equally populated. With increasing chain concentration, the transition region shifts to higher temperatures. The midpoints for the transition at 43 and 300 μM chain concentration are at $T = 1.53$ and 1.59, respectively. In general, the x_D versus T curves indicate that the GCN4-lz undergoes a sharp dissociating transition.

In Fig. 8 we show the average overall helix content versus temperature for chain concentrations of 2, 43, and 300 μM, as well as the helix contents for dimer and monomer at various temperatures. Fig. 8 also shows that at a 2 μM chain concentration, the overall helix content for the GCN4-lz declines slowly from a value of 91% at $T = 0.1$ to 83% at $T = 1.2$, and then exhibits a sharp transition to a value close to 10% at $T = 1.6$, with the transition midpoint being at $T = 1.45$. In general, for all chain concentrations, first there is a linear decline in helix content with increase in temperature and then there is a sharp transition with the transition temperature increasing with an increase in chain concentration. This result is consistent with the experimental results from melting studies on the GCN4-lz. Of course, in our theoretical curve, we also see another broad transition that occurs at a still higher temperature and the monomer helix content drops from 10% to ~2%. It is possible that this might occur outside the experimental temperature range. Comparison of the overall helix content curve to the helix content curves for monomer and dimer gives us a qualitative picture of the folding transition in the GCN4-lz. At the lowest temperature, the molecule is a dimer that is close to

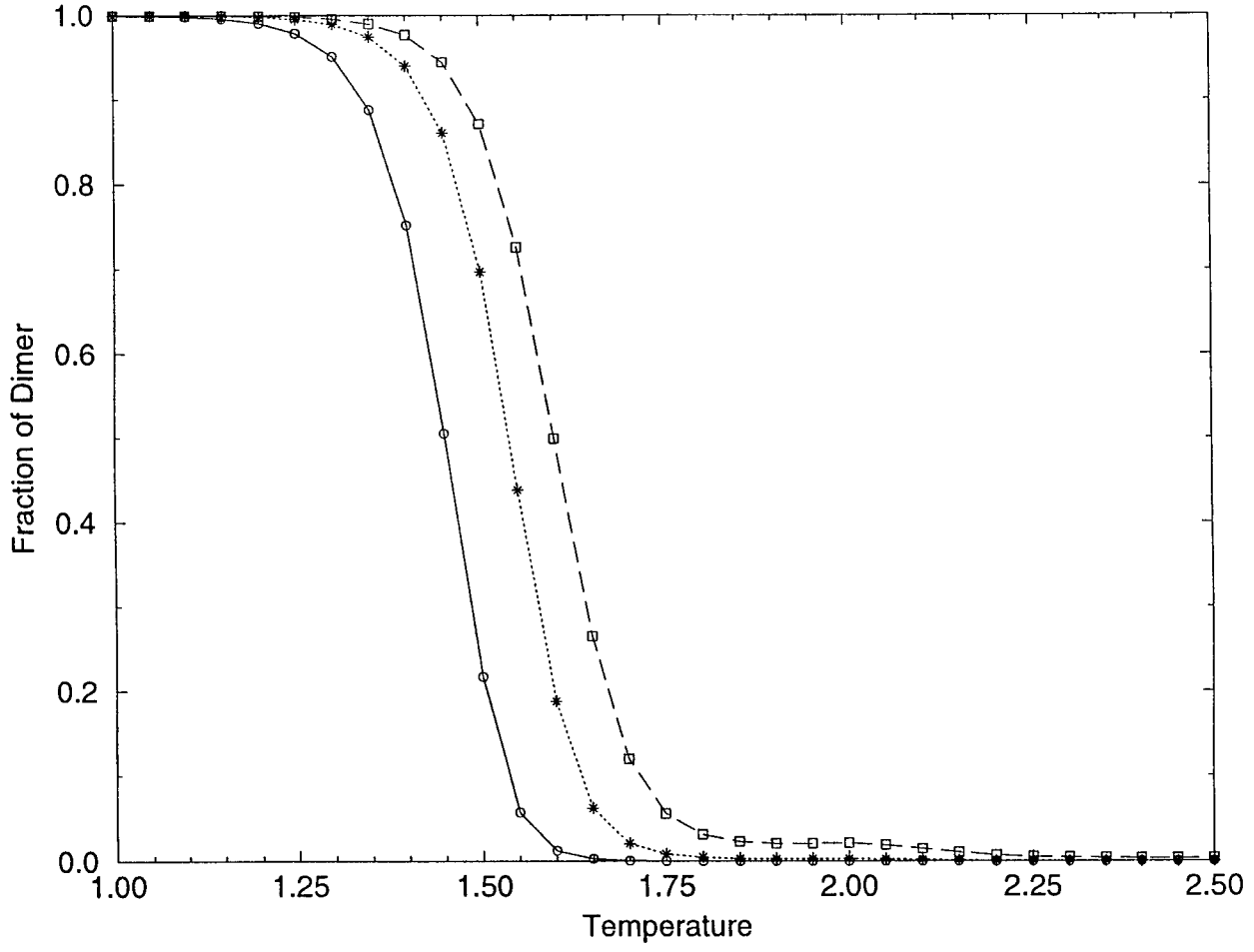

FIGURE 7 The mole fraction of dimers as a function of temperature, at chain concentrations of 2 μM *(solid line and circle)*, 43 μM *(dotted line and star)*, and 300 μM *(dashed line and square)*.

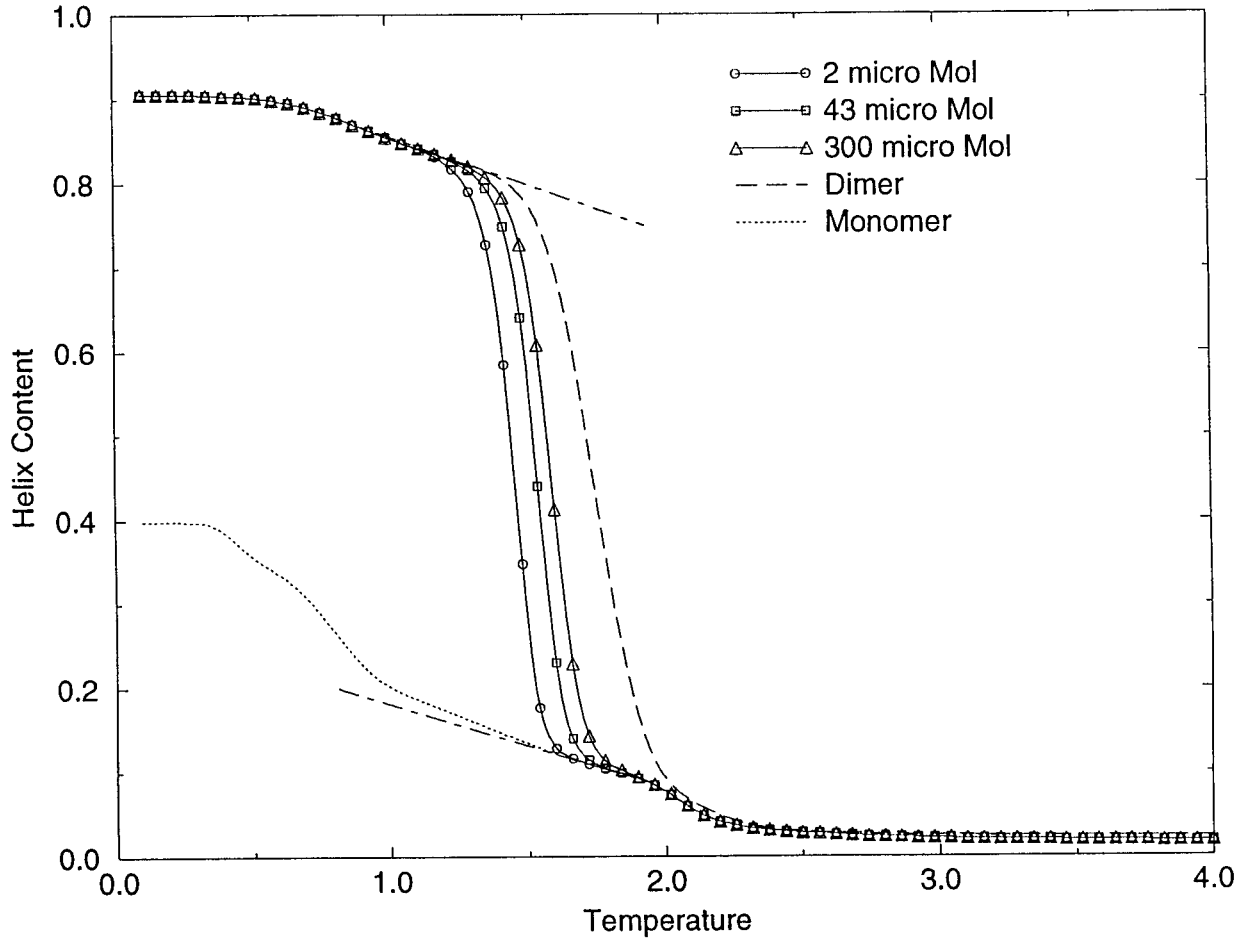

FIGURE 8 The average helix content of the monomer *(dotted line)* and dimer *(dashed line)* as a function of the temperature. The solid lines represent the average overall helix content versus temperature curves at chain concentrations of 2 μM *(circles)*, 43 μM *(squares)*, and 300 μM *(triangles)*. The dashed-dotted lines represent the baselines fitted to the native and denatured regions that will be used for van't Hoff analysis.

91% helical, and with an increase in temperature there is a continuous unfolding in the dimer. However, by the time the molecule unfolds to a state with 80% helix content, the chains fall apart and the monomeric chains have a helix content of ~10%. The second transition we see corresponds to full melting of the residual helix in the monomer.

Hence, our simulation results suggest that the equilibrium population in the GCN4-lz consists of folded dimers and essentially unfolded single chains. This can be seen in Fig. 9, which shows the population of various microscopic states at $T = 1.45$, i.e., transition midpoint for the chain concentration of 2 μM. At this temperature, the dimer and monomer populations show sharp peaks in two distinct energy regions. At the same temperature, the population of the monomer and dimer as a function of the helix content is shown in the inset to Fig. 9. The population versus helix content curves for the monomer and dimer indicate that the two distinct peaks correspond to folded dimer and unfolded single chains. A similar distribution of monomer and dimer populations is also seen at reduced temperatures of 1.53 and 1.59, which correspond to the transition midpoints for chain concentrations of 43 and 300 μM.

We have also done a van't Hoff analysis (Privalov and Gill, 1988; Marky and Breslauer, 1987) of the overall helix content curve to check whether the two-state model indeed holds. For the van't Hoff analysis, only the portion of the helix content curve up to $T = 1.9$ was used and the higher temperature regions of the curve, which represent a transition within the monomer, were excluded. As described in the Methods section, the results of the van't Hoff analysis would depend on whether or not baseline fitting to the $\theta(T)$ curve was done. We first discuss the results of the van't Hoff analysis obtained by baseline fitting. In Fig. 10 *A* we show the apparent fraction of native state, $f_N(T)$, obtained from the helix content curve by the method of baseline fitting. The f_N values at different temperatures were used to calculate the apparent equilibrium constant (K_{app}) using Eq. 20. Using the temperature derivative of ln K_{app}, the van't Hoff enthalpy was calculated using Eq. 21. The calorimetric enthalpies at the corresponding temperatures were calculated directly from the average energy of the dimer and monomer using Eq. 22. Fig. 10 *B* shows the ratio of the van't Hoff to calorimetric enthalpies, $\Delta H_{VH}/\Delta H_{cal}$, in the transition region for the three different chain concentrations.

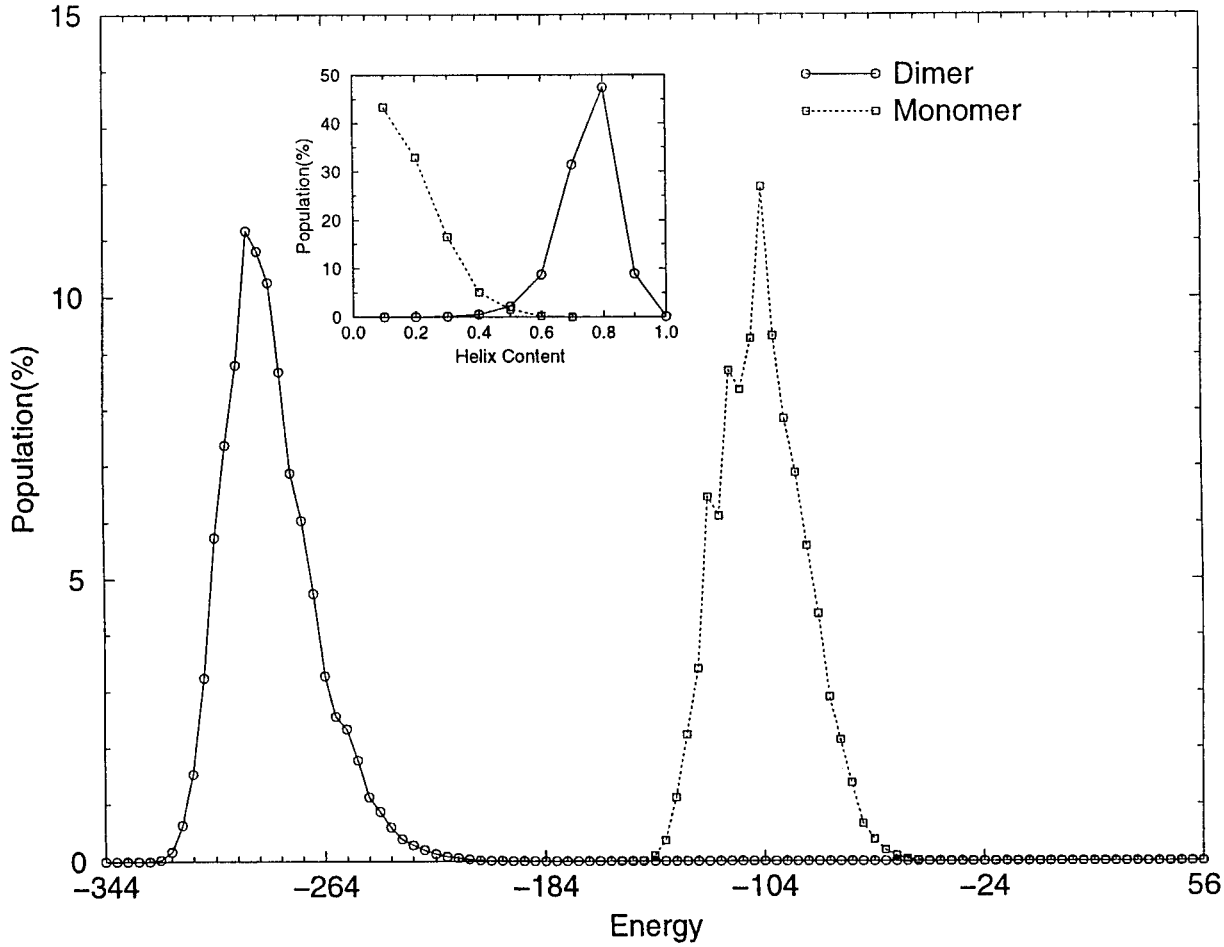

FIGURE 9 The populations of the dimer *(solid line and circle)* and monomer *(dotted line and square)* in various energy states at $T = 1.45$, which is the transition temperature for chain concentration of 2 μM. The inset shows the population of the dimer *(solid line and circle)* and monomer *(dotted line and square)* as a function of helix content at $T = 1.45$.

The transition region has been defined as the temperature range in which $f_N(T)$ varies from 95% to 5%. As shown, $\Delta H_{VH}/\Delta H_{cal}$ is 1.002, 0.997, and 0.999 at reduced temperatures of 1.45, 1.53, and 1.59, corresponding to the midpoints of the transition for the three different chain concentrations considered here. This ratio is also close to 1.0 over most of the transition range. This indicates that the folding transition shown in Fig. 8 can be well-represented by a two-state model as defined by baseline fitting. The experimental studies also use a similar baseline fitting to deduce the helix content from the CD data and conclude that GCN4-lz exhibits a two-state folding transition. Hence, the results of our statistical thermodynamics calculation reproduce the experimentally observed two-state transition (Kenar et al., 1995; Sosnick et al., 1996) in the GCN4-lz.

However, if we do the van't Hoff analysis without baseline fitting, the $\Delta H_{VH}/\Delta H_{cal}$ at the transition temperature is close to 0.8 for all three chain concentrations, thus indicating that there are intermediates present and transition cannot be represented by a two-state model. It is important to understand why the result of the van't Hoff analysis with the baseline fitting differs from that without baseline fitting. In the method of baseline fitting, we used a relaxed definition for the native state. Structures with reduced helix content were considered native, and hence a two-state model was applicable. However, without baseline fitting, we used a strict definition of native state and all structures with minor unfolding were considered transition intermediates. Hence, the conclusion about whether or not the folding/unfolding transition is two-state would depend on whether or not baseline fitting was done. Even though this has been realized before in the experimental analysis of melting curves, different workers have given different justifications for use of baseline. Some (Hvidt et al., 1985; Lehrer and Stafford, 1991) deny conformational significance to the linear region of the melting curve. They attribute it to the temperature dependence of the chiro-optical properties of the native coiled coil, and hence justify the use of a baseline. However, others (Holtzer et al., 1983, 1990; Holtzer and Holtzer, 1992) ascribe a conformational significance to the linear region of the melting curve and attribute it to slow unfolding below room temperature. They have given independent experimental evidence for their interpretation by carrying out site-specific unfolding studies (Holtzer et al., 1997) on

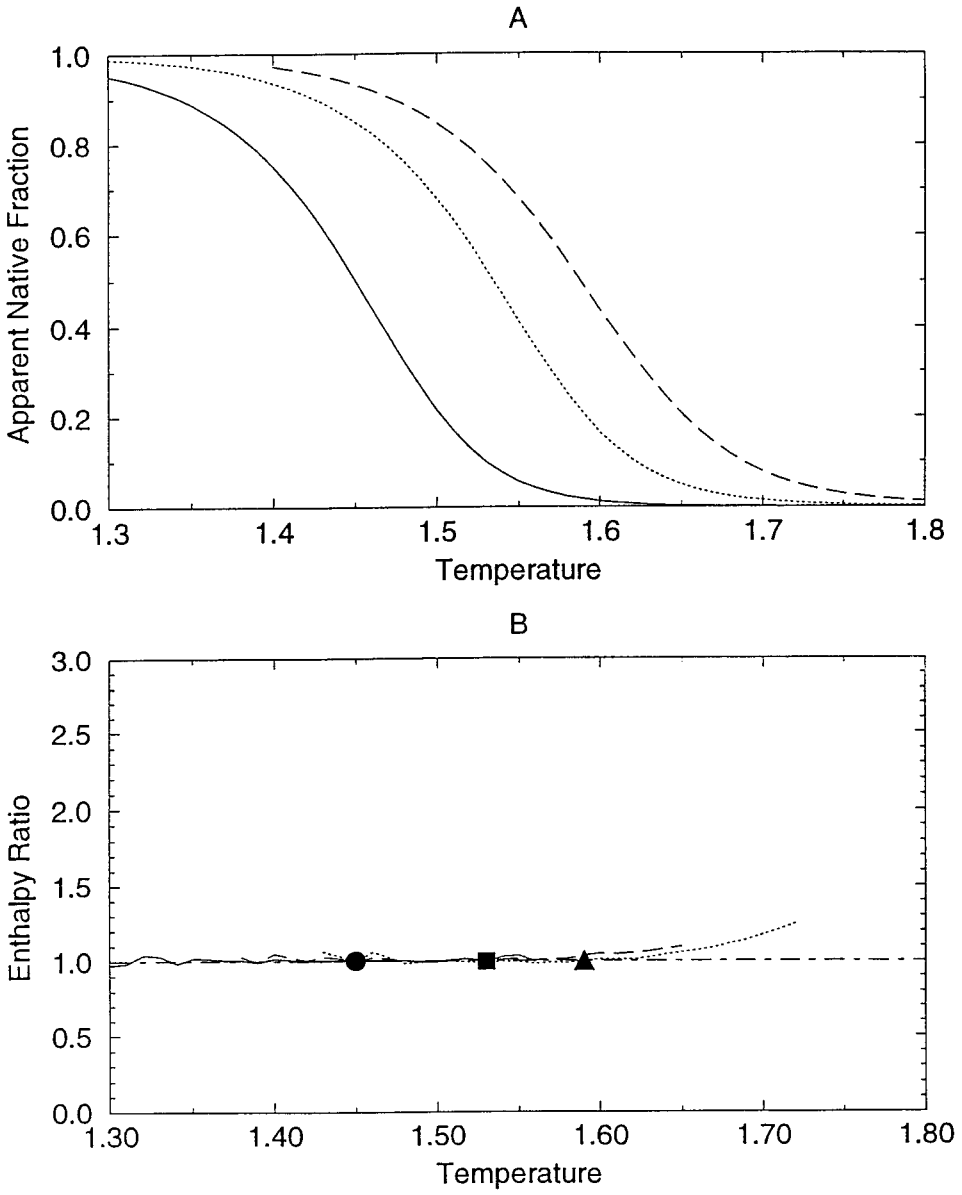

FIGURE 10 *(A)* The apparent fraction of native GCN4-lz, f_N, as a function of temperature at chain concentrations of 2 μM *(solid line)*, 43 μM *(dotted line)*, and 300 μM *(dashed line)*. The f_N values have been extracted from the overall helix content versus temperature curves assuming a two-state model for the folding/unfolding transition. *(B)* The ratio of the van't Hoff to the calorimetric enthalpy, $\Delta H_{VH}/\Delta H_{cal}$, at different temperatures in the transition region for chain chain concentrations of 2 μM *(solid line)*, 43 μM *(dashed line)*, and 300 μM *(dotted line)*. The filled circle, square, and triangle indicate the values of $\Delta H_{VH}/\Delta H_{cal}$ at $T = 1.45$, 1.53, and 1.59, respectively, corresponding to the transition midpoints for these chain concentrations, respectively. The dashed-dotted line represents the $\Delta H_{VH}/\Delta H_{cal}$ value of unity expected for a two-state transition.

a GCN4-like leucine zipper. The results from our simulations agree with the interpretation of Holtzer and co-workers.

Thus, our simulations with knowledge-based potentials (Vieth et al., 1995) clearly indicate that the statistical potentials derived from a database of native protein structures can indeed describe the conformational properties of the native as well as the nonnative states of proteins, and can also reproduce many features of their folding thermodynamics. Since reduced models of proteins often employ such potentials, this observation is of considerable significance, in view of the recent questions (Thomas and Dill, 1996; Ben-Naim, 1997) raised in the literature about whether or not such knowledge-based potentials have physical meaning. While these objections are, in principle, valid in a series of model studies, it has been demonstrated that such potentials of mean force can be extracted from a structural database comprised of native protein structures and are quite close to the "true" potentials (Zhang and Skolnick, 1998). The present work further suggests that these knowledge-based potentials, derived from native structures of proteins, are applicable to nonnative states and that general questions related to thermodynamics can be addressed. This observation is also supported by the results (Mohanty et al., 1998) obtained from a detailed comparison of knowledge-based potentials with detailed atomic potentials for various folded and unfolded conformations of GCN4-lz. The two potentials show a good correlation, which extends from the folded to the unfolded region. Since there is a wide body of evidence that detailed atomic potentials do quite a good job of describing folded as well as unfolded states of proteins (Brooks, 1993, 1995; Brooks et al., 1988; Hirst and Brooks, 1995), a good correlation between the two potentials indicates that knowledge-based potentials might also describe the properties of native and nonnative states. Finally,

knowledge-based potentials are designed to be applied to reduced models to capture some proteinlike features not readily encoded in a molecular mechanics potential as applied to a simplified protein representation.

CONCLUSIONS

In this work we carried out ESMC simulations using lattice model and knowledge-based potentials (Kolinski and Skolnick, 1994; Vieth et al., 1995) to study the thermodynamics of the folding/unfolding transition in the GCN4 leucine zipper. Even though ESMC simulations have been used before for studying the thermodynamics of folding in single chain proteins (Hao and Scheraga, 1994a,b; Kolinski et al., 1996), here we described a method for generalizing the ESMC approach to multichain systems. By using this method we computed the average helix content for the two-chain GCN4-lz system over a range of temperatures and chain concentrations, taking into account the dissociating transition. Simulations with exactly the same parameters as in the earlier folding simulation (Vieth et al., 1994) of Vieth and co-workers predicted a denatured state helix content that is too high. However, by a simple reduction of the scale factors for the short-range interactions, it was possible to diminish the helix content of the denatured state to a value that was in agreement with experiment, while at the same time keeping the native state significantly helical. These results indicate that it is possible to predict the experimentally observed conformational properties of the native and denatured states of the GCN4-lz using this lattice protein model and knowledge-based potentials.

The van't Hoff analysis of the helix content versus temperature curves, with baseline fitting, indicates that the folding transition can be described by a two-state model at chain concentrations ranging from 2 to 300 μM. Hence, our simulations clearly reproduce the two-state folding transition of the GCN4-lz observed experimentally (Kenar et al., 1995; Sosnick et al., 1996). Our results also suggest that the physical origin of this two-state folding transition in the GCN4-lz is very different from that observed for globular proteins. The conformational transition within the dimer of the GCN4-lz is continuous, but the chains fall apart after very little unfolding in the dimer. The balance of short- and long-range interactions in the GCN4-lz is such that the isolated chains have very little helix content, but in the dimer, quarternary interactions stabilize the helical structure. Hence, after dissociation, the molecule essentially loses all its helix content and, at transition temperature, one sees only the folded dimer and unfolded single chains. However, this folded dimer population at the transition temperature does not represent completely folded native states, rather it is comprised of structures with slight unfolding from the native state. Hence, only if a relaxed definition is used for the native state by "baseline fitting," the transition can be described as two-state. If a strict definition is used where only the structures at the lowest temperature are native, then the transition cannot be described by a two-state model; rather, there is an initial loss of helix content within the dimer, which is subsequently followed by chain dissociation.

Even though recent theoretical work on the thermodynamics of folding transitions in idealized protein models (Dill et al., 1995) have given us interesting insights into the origin of all-or-none folding transitions observed for proteins, there have been relatively few simulations incorporating sufficient details in the model for meaningful comparison with experiment. The results presented here clearly demonstrate that using a lattice protein model, knowledge-based potentials, and powerful techniques such as ESMC, it is not only possible to predict the native structure, but also features related to their folding thermodynamics, which can be directly compared with experiment.

The authors thank Dr. C. L. Brooks III for sharing unpublished results. D.M. acknowledges valuable discussions with Drs. M. Vieth and L. Zhang. A.K. is an International Scholar of the Howard Hughes Medical Institute.

This work was supported in part by National Institutes of Health Research Resource Grant P41 RR12255.

REFERENCES

Alber, T. 1992. Structure of the leucine zipper. *Curr. Opin. Genet. & Dev.* 2:205–209.

Ben-Naim, A. 1997. Statistical potentials extracted from protein structures: are these meaningful potentials? *J. Chem. Phys.* 9:3698–3706.

Berg, B. A., and T. Neuhaus. 1991. Multicanonical ensemble: a new approach to simulate first-order phase transitions. *Phys. Rev. Lett.* 68: 9–12.

Brooks, C. L. III. 1993. Molecular simulations of peptide and protein unfolding: in quest of a molten globule. *Curr. Opin. Struct. Biol.* 3:92–98.

Brooks, C. L. III. 1995. Methodological advances in molecular dynamics simulations of biological systems. *Curr. Opin. Struct. Biol.* 5:211–215.

Brooks, C. L. III, M. Karplus, B. M. Pettit. 1988. Proteins: a theoretical prospective of dynamics, structure and thermodynamics. *Adv. Chem. Phys.* 71:1–249.

Cohen, C., and A. H. Parry. 1990. α-Helical coiled coils and bundles: how to design an α-helical protein. *Proteins: Struct., Funct., Genet.* 7:1–15.

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

DeLano, W. L., and A. T. Brunger. 1994. Helix packing in proteins: Prediction and energetic analysis of dimeric, trimeric, and tetrameric GCN4 coiled coil structures. *Proteins: Struct., Funct., Genet.* 20: 105–123.

DiDonato, J. A., M. Hayakawa, D. M. Rothwarf, E. Zandi, and M. Karin. 1997. A cytokine-responsive IκB kinase that activates the transcription factor NF-κB. *Nature.* 388:548–554.

Dill, K. A., S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan. 1995. Principles of protein folding: a perspective from simple exact models. *Protein Sci.* 4:561–602.

Fraser, R. D. B., and T. P. MacRae. 1971. Structure of α-keratin. *Nature.* 233:138–140.

Godzik, A., A. Kolinski, and J. Skolnick. 1993. Lattice representation of globular proteins: how good are they? *J. Comp. Chem.* 14:1194–1202.

Hansmann, U. H. E., and Y. Okamoto. 1993. Prediction of peptide conformation by a multicanonical algorithm: new approach to the multiple minima problem. *J. Comp. Chem.* 14:1333–1338.

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

Hao, M.-H., and H. A. Scheraga. 1994b. Statistical thermodynamics of protein folding: sequence dependence. *J. Phys. Chem.* 98:9882–9893.

Hao, M.-H., and H. A. Scheraga. 1995. Statistical thermodynamics of protein folding: comparison of mean-field theory with Monte Carlo simulations. *J. Chem. Phys.* 102:1334–1348.

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.

Harrison, S. 1991. A structural taxonomy of DNA-binding domains. *Nature.* 353:715–719.

Herschbach, D. R. 1959. Molecular partition function in terms of local properties. *J. Chem. Phys.* 31:1652–1661.

Hirst, J. D., and C. L. Brooks III. 1995. Molecular dynamics studies of isolated helices of myoglobin. *J. Mol. Biol.* 34:7614–7621.

Hodges, R. S., A. S. Saund, P. C. S. Chong, S. A. St-Pierre, and R. E. Reid. 1981. Synthetic model for two-stranded, α-helical coiled coils. *J. Biol. Chem.* 256:1214–1224.

Holtzer, M. E., and A. Holtzer. 1992. α-Helix to random coil transitions: interpretation of CD in the region of linear temperature dependence. *Biopolymers.* 32:1589–1591.

Holtzer, M. E., A. Holtzer, and J. Skolnick. 1983. The α-helix-to-random-coil transition of two-chain, coiled coils. Theory and experiments for thermal denaturation of α-tropomyosin. *Macromolecules.* 16:173–180.

Holtzer, A., M. E. Holtzer, and J. Skolnick. 1990. Does the unfolding transition of two-chain, coiled-coil proteins involve a continuum of intermediates? *In* Protein-Folding. L. M. Gierasch and J. King, editors. AAAS Books, Washington, DC. 177–190.

Holtzer, M. E., E. G. Lovett, D. A. d'Avignon, and A. Holtzer. 1997. Thermal unfolding in a GCN4-like leucine zipper: $^{13}C\alpha$-NMR chemical shifts and local unfolding curves. *Biophys. J.* 73:1031–1041.

Hvidt, S., M. E. Rogers, and W. F. Harrington. 1985. Temperature-dependent optical rotatory dispersion properties of helical muscle proteins and homopolymers. *Biopolymers.* 24:1647–1662.

Johnson, P., and L. B. Smillie. L.B. 1975. Rabbit skeletal α-tropomyosin chains are in register. *Biochem. Biophys. Res. Commun.* 64:1316–1322.

Kenar, K. T., B. Garcia-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.

Kolinski, A., W. Galazka, and J. Skolnick. 1996. On the origin of the cooperativity of protein folding: implications from model simulations. *Proteins: Struct., Funct., Genet.* 26:271–287.

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

Krystek, S. R., Jr., R. E. Bruccoleri, and J. Novotny. 1991. Stabilities of leucine zipper dimers estimated by an empirical free energy method. *Int. J. Pept. Protein Res.* 38:229–236.

Landshultz, W. H., P. F. Johnson, and S. L. McKnight. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science.* 240:1759–1764.

Lee, J. 1993. New Monte Carlo algorithm: entropic sampling. *Phys. Rev. Lett.* 71:211–214.

Lehrer, S. S., W. F. Stafford III. 1991. Preferential assembly of the tropomyosin heterodimer: equilibrium studies. *Biochemistry.* 30:5682–5688.

Lovejoy, B., C. Seunghyon, D. Cascio, D. K. McRorie, W. F. DeGrado, and D. Eisenberg. 1993. Crystal structure of a synthetic triple-stranded α-helical bundle. *Science.* 259:1288–1293.

Lovett, E. G., D. A. d'Avignon, M. E. Holtzer, E. H. Braswell, D. Zhu, and A. Holtzer. 1996. Observation via one-dimensional $^{13}C_\alpha$ NMR of local conformational substates in thermal unfolding equilibria of a synthetic analog of the GCN4 leucine zipper. *Proc. Natl. Acad. Sci. USA.* 93: 1781–1785.

Lumb, K. J., C. M. Carr, and P. S. Kim. 1994. Subdomain folding of the coiled coil leucine zipper from bZIP transcriptional activator GCN4. *Biochemistry.* 33:7361–7367.

Marky, L. A., and K. J. Breslauer. 1987. Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers.* 26:1601–1620.

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

McLachlan, A. D., and M. Stewart, 1975. Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. *J. Mol. Biol.* 98: 293–304.

McQuarrie, D. A. 1976. Statistical Mechanics. Harper's Chemistry Series. Harper and Row, New York.

Metropolis, N. A., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–1092.

Mohanty, D., B. Dominy, A. Kolinski, C. L. Brooks III, and J. Skolnick. 1998. Correlation between knowledge-based and detailed atomic potentials for GCN4-lz unfolding. *Proteins: Struct., Funct., Genet.* in press.

Monera, O. D., N. E. Zhou, P. Lavigne, C. M. Kay, and R. S. Hodges. 1996. Formation of parallel and antiparallel coiled coils controlled by the relative positions of alanine in the hydrophobic core. *J. Biol. Chem.* 271:3995–4001.

Nilges, M., and A. T. Brunger. 1991. Automated modeling of coiled coils: application to the GCN4 dimerization region. *Protein Engn.* 4:649–659.

Nilges, M., and A. T. Brunger. 1993. Successful prediction of the coiled coil geometry of the GCN4 leucine zipper domain by simulated annealing: comparison to the x-ray structure. *Proteins: Struct., Funct., Genet.* 15:133–146.

Okamoto, Y., and U. H. E. Hansmann. 1995. Thermodynamics of helix-coil transitions studied by multicanonical algorithms. *J. Phys. Chem.* 99:11276–11287.

O'Neil, K. T., R. H. Hoess, and W. F. Degrado. 1990. Design of DNA-binding peptides based on the leucine zipper motif. *Science.* 249: 774–778.

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.

O'Shea, E. K., K. J. Lumb, and P. S. Kim. 1993. Peptide "Velcro": design of a heterodimeric coiled coil. *Curr. Biol.* 3:658–667.

Phillips, G. N., Jr., J. P. Fillers, and C. Cohen. 1986. Tropomyosin crystal structure and muscle regulation. *J. Mol. Biol.* 192:111–131.

Privalov, P. L., and S. J. Gill. 1988. Stability of protein structure and hydrophobic interaction. *Adv. Protein Chem.* 39:191–235.

Smeal, T., P. Angel, J. Meek, and M. Karin. 1989. Different requirements for formation of jun:jun and jun:fos complexes. *Genes Dev.* 3:2091–2100.

Sosnick, T. R., S. Jackson, R. R. Wilk, S. W. Englander, and W. F. DeGrado. 1996. *Proteins: Struct., Funct., Genet. 24:427–432.*

Su, J. Y., and R. S. Hodges. 1994. Effect of chain length on the formation and stability of synthetic α-helical coiled coils. *Biochemistry.* 33: 15501–15510.

Talanian, R. V., C. J. Mcknight, and P. S. Kim. 1990. Sequence-specific DNA binding by a short peptide dimer. *Science.* 249:769–770.

Thomas, P., and K. A. Dill. 1996. Statistical potentials extracted from protein structure: how accurate are they? *J. Mol. Biol.* 257:457–469.

Vieth, M., A. Kolinski, C. L. Brooks III, and J. Skolnick. 1994. Prediction of the folding pathways and structure of the GCN4 leucine zipper. *J. Mol. Biol.* 237:361–367.

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

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

Zhang, L., and J. Hermans. 1993. Molecular dynamics study of structure and stability of a model coiled coil. *Proteins: Struct., Funct., Genet.* 16:384–392.

Zhang, L., and J. Skolnick. 1998. How do potentials derived from structural databases relate to "true" potentials? *Protein Sci.* 7:112–122.

Zitzewitz, J. A., O. Bilsel, J. Luo, B. E. Jones, and C. R. Matthews. 1995. Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. *Biochemistry.* 34:12812–12819.