Stabilization of Native and Non-native Structures by Salt Bridges in a Lattice Model of the GCN4 Leucine Dimer

Yanxin Liu, Prem P. Chapagain,† and Bernard S. Gerstman*

Department of Physics, Florida International University, University Park, Miami, Florida 33199

Received: October 14, 2009; Revised Manuscript Received: December 4, 2009

We use computer simulations to investigate the influence of salt bridges on the dimerization of the GCN4 leucine zipper. Use of a lattice model allows the dimerization process to be followed for time scales long enough to investigate large-scale structural changes traversing large distances in configuration space. We calculated the rate, efficiency, and stability of dimerization and free-energy landscapes. We varied the strength of the ionic interactions and find that there is an optimal, intermediate salt bridge strength at which the dimerization process proceeds at the maximum rate. Though especially strong salt bridge strength beneficially stabilizes native dimers if they form, it can also stabilize non-native configurations that can hinder the dimerization process. We give examples of stable, non-native structures that might arise. These structures may be relevant to interesting recent experiments that have provided evidence of unusual forms of the leucine zipper GCN4-p1 called the LZ_{GCN4} x-forms, which dominate the population in certain biochemical conditions.

Introduction

The "protein folding problem" of how a protein can fold into its proper three-dimensional, functional, native state in a biologically relevant time has been studied for decades.[1–6] The complexity of the problem is due to the flexibility of a peptide chain and the large number of interactions experienced by amino acids. The hydrophobic interaction is important in the large-scale collapse of proteins toward compact shapes.[7,8] Other interactions, such as hydrogen bonds, dipole interactions, electrostatic interactions, and van der Waals interactions, are necessary to produce the intricate secondary and tertiary structure of individual proteins,[9–11] as well as the proper interchain alignment in dimers.[12] The role of salt bridges formed by electrostatic, ionic, interactions is especially interesting.[13–18] Intrachain and interchain salt bridges can form between the charged side chains of amino acids, such as lysine and glutamic acid. Experimental results report that salt bridges can stabilize or destabilize native-state structures,[12,13,19–23] as well as affect other aspects of the folding process, such as the folding rate.[23–25]

Coiled-coil dimers have been studied extensively because they are a commonly found dimer structure, but the role of salt bridges in the folding and dimerization of coiled coils is not fully understood.[26–30] The simplicity of the coiled-coil structure arises from the fact that it may contain as few as two α-helices that interact with each other to form a simple protein complex. A well-studied example of a coiled coil is the GCN4-p1 leucine zipper from the yeast transcription factor GCN4.[23,31–36] In the wild-type GCN4-p1 leucine zipper, each of the two chains is a 33-residue peptide. Experiments performed on this peptide have resulted in uncertainty about the role that salt bridges play in stabilizing the native structure,[12,22,23,37,38] as well as how the salt bridges affect the speed and efficiency of the dimerization process.[23]

To address these issues at a molecular level, we performed extensive computer simulations to determine the effects of the strength of salt bridges as a function of temperature on the thermodynamics and kinetics of dimerization. We employed a three-dimensional lattice model incorporating a Monte Carlo−Metropolis algorithm to investigate the effect of salt bridge strength on the rate, efficiency, and stability of the dimerization of a model system with many of the properties of the leucine zipper. Use of a lattice model allows the dimerization process to be followed for time scales long enough to investigate large-scale structural changes traversing large distances in configuration space.

The amino acid sequence used in the computer simulations is described in detail in ref 36. Our investigations of pathways involving non-native molecular configurations provide an explanation for the experimental findings that, though stronger ionic interactions result in more stable dimers, the efficiency of the dimerization process is not helped by very strong salt bridges. This understanding presents the possibility to make targeted amino acid substitutions to improve dimerization.

In agreement with experimental results,[23] we find dimerization rates in our model system to be almost independent of salt bridge strength at low temperatures. At high temperatures, we find that the strength of the salt bridges has a noticeable effect on the dimerization rates. Interestingly, our simulations show that increasing the strength of the ionic interactions does not lead to a monotonic increase in the speed of the dimerization process at high temperatures. Instead, we find that the dimerization process proceeds at the maximum rate at an intermediate salt bridge strength. Though stronger salt bridges beneficially stabilize the native dimer once it forms, stronger salt bridges also stabilize non-native structures that act as kinetic traps and hinder the dimerization process. Forming these non-native salt bridges requires significant bending of peptide chains and is therefore more likely to play a role at higher temperatures when the chains undergo large random thermal structural fluctuations. In order to understand the temperature dependence of dimerization on salt bridge strength, we present quantitative results of dimerization rates, heat capacity, and free energy landscapes as a function of the ionic interaction strength for our model of

* To whom correspondence should be addressed. E-mail: gerstman@fiu.edu. Tel: 305-348-3115. Fax: 305-348-6700.
† E-mail: chapagap@fiu.edu.

10.1021/jp909872a
Published on Web 12/28/2009

the GCN4 leucine zipper (LZ_{GCN4}), and display the molecular structure of the non-native configuration that acts as the kinetic trap.

Interesting recent experiments[39] have investigated an unusual form of the leucine zipper GCN4-p1 called the LZ_{GCN4} x-form. The probability of occurrence of the x-form is low at pH 7.1, but under different conditions the x-form dominates, with 83% of the population at pH 3.2 and a low peptide concentration of 15 μM. Although these conditions are non-natural, Nikolaev and Pervushin propose that conformational variants such as the x-form may be biologically relevant. Though we do not know if the non-native conformation that we describe here is the x-form, our results show with molecular level details how various conformations of the leucine zipper can appear in the folding process.

Methods

Three-Dimensional Lattice Model. The computer simulations, described in detail in ref 36, were performed by employing a Monte Carlo algorithm which simulates the dynamics of the system by changing the internal configuration of each chain, as well as the relative separation and orientation between the two chains. A Metropolis test is incorporated to ensure that the various configurations appear with the correct thermodynamic Boltzmann probability.[40] The model uses an underlying cubic lattice, and includes separate degrees of freedom for an amino acid's backbone and side chain. The lattice representation of the peptide bond allows protein secondary, tertiary, and quaternary structures to be effectively represented.[36,41–44] In the simulations, time is counted in Monte Carlo (MC) steps. Each MC step equates to approximately 1 ns and includes a variety of moves involving individual amino acids, groups of amino acids, or an entire chain.

In this paper, we are focusing on the behavior of the chains once they are close enough to interact. For these investigations, the size of the box is not important as long as it is large enough for the two chains to have the freedom to fluctuate once they have approached each other. We also wanted the concentration to be close to experimental values. We chose a volume of the lattice simulation box containing the two chains that is equivalent to a concentration of 2470 μM, which is close to, but higher than, the experimental concentration used by Privalov[45] of 880 μM. This simulation box is large enough so that it is possible for the two chains to be separated far enough so that they do not interact. This volume played two roles in our simulations. Simulations were initialized so that the chains were separated by enough distance so that they were not interacting at the start of the simulations. In addition, during a simulation, if the two chains approached each other, they could later separate if their attractive interactions were not sufficiently strong. The volume of the simulation box was also big enough to allow the chains the freedom to flip, bend, or rotate, either individually or as a dimer. It is possible to increase the volume of the simulation box to decrease the effective concentration to match the experimental conditions. We did not do this because it would not add to the information that we are seeking which is specifically concerned with inter and intrachain interactions. Instead, increasing the size of the box would have the detrimental effect of making the simulations more computationally costly as the chains searched a larger volume before finding each other.

Interaction Hamiltonian. For any configuration of the two chains that occurs in a simulation, the energy of the system is calculated using the following Hamiltonian.

$$H = \sum_i \Big(\sum_{j>i} \left(a_{ij}^{\mathrm{hp}} E_{ij}^{\mathrm{hp}} + a_{ij}^{\mathrm{es}} E_{ij}^{\mathrm{es}} + a_{ij}^{\mathrm{vdw}} E^{\mathrm{vdw}} + a_{ij}^{\mathrm{diphb}} E^{\mathrm{diphb}} + a_{ij}^{\mathrm{rep}} E^{\mathrm{rep}} \right) + a_i^{\mathrm{l}} E_i^{\mathrm{l}} + a_i^{\mathrm{m}} E_i^{\mathrm{m}} \Big) \quad (1)$$

where i and j are amino acid residue numbers in the primary sequence of the two chains. The terms in eq 1 represent different interactions that occur in proteins: hp represents side-chain−side-chain interactions and can be energy lowering or energy raising depending on whether the interacting side chains are hydrophobic or hydrophilic; es represents electrostatic salt bridges and this is the interaction that we focus on in this paper, vdw is a weak van der Waals attraction between backbones; diphb represent the dipole interaction and hydrogen bond interaction that is especially important in stabilizing the long-range organization of an α-helix; rep is a soft-core repulsion if amino acids approach too close; l represents a structural propensity that enhances the probability for a single amino acid to be in its own local α-helix configuration; and m is a structural propensity that enhances the probability for two adjacent amino acids to cooperate to form a midsize α-helix configuration.

The values of the interactions used in the simulations in this paper are the following: E^{hp} depends on the identity of the two side chains: hydrophobic−hydrophobic = −1.7 kcal/mol, hydrophobic−hydrophilic = +1.0 kcal/mol, hydrophilic−hydrophilic = +0.1 kcal/mol. The strengths of the other interactions are as follows: $E^{\mathrm{vdw}} = -0.1$ kcal/mol, $E^{\mathrm{diphb}} = -0.7$ kcal/mol, $E^{\mathrm{rep}} =$ 2.5 kcal/mol, and $E^{\mathrm{l}} = E^{\mathrm{m}} = -0.5$ kcal/mol. In this paper we are investigating the effects of stronger or weaker electrostatic salt bridge interaction, E^{es}. For each chain configuration in a simulation, the interactions that contribute are determined based upon distance and other criteria that are expressed through the a_{ij} and explained in ref 36.

The electrostatic interaction is a long-range interaction. In order to allow all native salt bridges to be represented, charged residues interact in our model if they are located within a separation of $\sqrt{33}$ lattice spacings. Because of the strength and specificity of the salt bridges, charged amino acids must be placed in the biologically appropriate positions in the primary sequence of each chain. The specific locations of the charged amino acids along the chains are given in the following section describing the leucine zipper structure.

GCN4 Leucine Zipper with Trigger Sequence in Lattice Model. The 33-residue GCN4-p1 takes on an α-helical structure of nine helical turns and occurs at the C-terminus end of the larger yeast transcription factor GCN4 protein containing 281 residues.[32] Each chain in the dimer has nine residues that are on the interfacial surface between the chains. Eight of these residues are hydrophobic. The remaining interfacial residue is hydrophilic and occurs in the middle of each chain at position 16. Residues 17−29 have stronger propensities for α-helix formation and act to trigger dimerization.[30,46,47]

By careful assignment of the primary sequence of amino acids in our computer lattice model, we are able to simulate the important interactions in the leucine zipper. Amino acids are assigned one of four characteristics: hydrophilic, hydrophobic, charged residues that make salt bridges in the dimer, and inert. As with the real GCN4-p1, in our model if the chains dimerize into the native configuration there will be nine residues on each chain that are on the interfacial surface and the interfacial residue that is in the middle of the chain is hydrophilic. The other eight interfacial residues (four on either side) are hydrophobic. The interfacial hydrophilic residue on each chain assists correct dimerization by destabilizing misaligned dimers by placing a

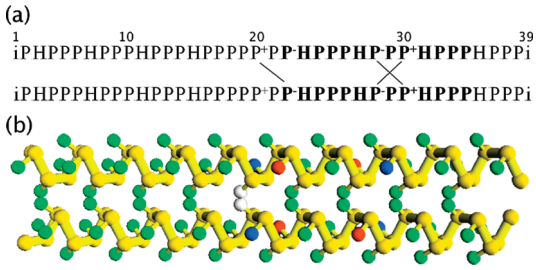

Figure 1. Computer simulation model of the GCN4-p1 leucine zipper. Both native and non-native salt bridges can occur during the simulations. (a) The sequence of amino acids used in each chain: P-hydrophilic, H-hydrophobic, i-inert. Native salt bridges are illustrated by lines. The trigger sequence on each chain is shown in boldface. (b) The native state of the dimer. Blue side chains are positively charged, and red side chains are negatively charged.

hydrophilic residue from one chain close to a hydrophobic side chain from the other peptide[48] if the chains are misaligned.

The trigger sequence hypothesis for dimerization proposed by Kammerer et al. is generally accepted.[30,49] Enhanced α-helical propensity of the 13 amino acids in the trigger sequence is included in our lattice model by assigning a stronger helical propensity $E^{\mathrm{l}} = E^{\mathrm{m}} = E^{*}_{\mathrm{L,M}}$ for the equivalent residues in our model compared to the α-helical propensity of $E^{\mathrm{l}} = E^{\mathrm{m}} = E_{\mathrm{L,M}} = -0.5$ kcal/mol for amino acids outside the trigger segment. In this paper, we use $E^{*}_{\mathrm{L,M}} = -0.8$ kcal/mol, which gives a ratio of $E^{*}_{\mathrm{L,M}}/E_{\mathrm{L,M}} = 1.6$ in agreement with a statistical analysis of secondary structure preference from the protein data bank.[50,51]

In order to allow the formation of the native salt bridges, in our model we assign side chains of the appropriately situated residues to be negatively charged, representing glutamic acid, and the side chains of other residues as positively charged, representing lysine. These charged side chains form three interchain salt bridges in the native dimer structure and are able to do so in our model as shown in Figure 1a. Of significant importance and the focus of our investigations, in our computational model they are also permitted to experience non-native interchain electrostatic attractions and repulsions. The non-native repulsions are important in destabilizing misaligned configurations.[52,53] Intrachain attractive salt bridges can form both in the native dimer as well as in non-native structures. All salt bridges that form in our model are assigned equal strength, including the non-native intrachain salt bridges that stabilize kinetic trap structures. The detrimental effects of non-native salt bridges are found to be especially important at high temperatures when the chains are very flexible and have a higher probability of forming non-native, intrachain salt bridges.

The computer simulation model of the GCN4-p1 leucine zipper is displayed in Figure 1. Part a shows the sequence of amino acids used in each chain. Native salt bridges are illustrated by lines. The trigger sequence on each chain is shown in boldface. Figure 1b shows the native state of the dimer. Blue side chains are positively charged and red side chains are negatively charged.

Structural Parameters. Our lattice model requires four amino acids to construct each of the nine helical turns in the native state. To better model the GCN4-p1, we add an additional residue at one end, as described in detail in ref 36 for a total of 37 residues in each chain that can be in α-helical secondary structure. There are a variety of structural parameters that provide information on the configuration of the system. Secondary structure is monitored through the parameter q that supplies information on how many amino acids are in α-helical secondary structure. The maximum value of q is 74 (37 from each chain), with 13 coming from the amino acids in the trigger sequence of each chain. In the figures, q is represented as a fraction of this maximum of 74.

Native interchain interactions along the interface between the chains are monitored by Q, with $Q = 9$ as the maximum that occurs in the fully formed native dimer. If only the trigger sequences of the two chains are properly aligned, the value will be $Q = 3$. An additional native interchain contact near the C-terminus just outside the trigger sequence usually follows quickly, giving $Q = 4$ as a result of proper alignment of the two trigger sequences. If the two chains interact, but in a non-native misaligned structure, the value can be $Q = 0$. In the figures, we plot Q as a fraction of the maximum native contacts of nine.

The separation between the two chains is measured by d_{cm}, which is the distance between the centers of mass of the two chains. There are many non-native configurations of the chain that can have the same d_{cm}, and therefore d_{cm} is most valuable when used in conjunction with other structural parameters and the chains' energy. The results from using a combination of parameters to investigate the dimerization process are presented.

Results

A variety of kinetic and thermodynamic characteristics and free energy landscapes were calculated. Both the dimer formation and dissociation processes were investigated under various conditions of salt bridge strength and temperature. The strength of the electrostatic interaction in eq 1 is varied from $E^{\mathrm{es}} = 0$ kcal/mol, which is equivalent to a mutant that has no salt bridges, to $E^{\mathrm{es}} = \pm 3.0$ kcal/mol (attraction (−), repulsion (+)). The change in the strength of E^{es} employed in our computer simulations can be achieved in experiments by changing the salt concentration of the solvent.[54–56] We go up to the relatively high value of $E^{\mathrm{es}} = 3.0$ kcal/mol strength in order to show that non-native structural kinetic traps can be created as an unanticipated byproduct of attempts to stabilize the native state.

To obtain a good statistical sample for analyses, we ran 100 independent simulations at each value of E^{es}. Each folding simulation started with the chains in the same random coil configuration but separated far enough so that they do not interact. The only difference between each simulation is the random number seed, which leads to different sequences of random numbers throughout the simulations and produces independent paths through configuration space. We stop a simulation when the chains have correctly dimerized, or if the simulation reaches 20×10^6 MC steps. This is long enough to capture most dimerization events, as shown in Figure 10 of ref 36. Unfolding simulations start with the chains in the native dimer configuration, and stop when the chains have fully dissociated or after 20×10^6 MC steps.

Dimerization Probability: Stronger E^{es} Is Not Always Better. The probability for the two chains to fold and dimerize is shown in Figure 2a as a function of temperature for different strengths of E^{es}. The dimerization probability is the percentage of independent simulations, all starting in the random coil configuration, that dimerize into the native configuration within 20×10^6 MC steps. When $E^{\mathrm{es}} = 0$ kcal/mol, there are no amino acids capable of forming salt bridges, and Figure 2a shows that the system is not able to dimerize with 100% effectiveness at any temperature. For $E^{\mathrm{es}} = 0$, the maximum dimerization probability is 90% which occurs only at $T = 45$ °C and the dimerization probability drops rapidly at both lower and higher temperatures. This shows that the absence of interchain salt

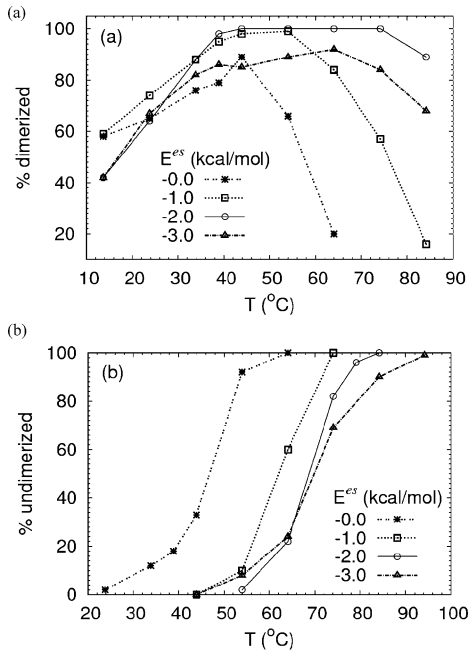

Figure 2. (a) Probability for the chains to dimerize as a function of the strength of the ionic salt bridge interaction, E^{es}, and as a function of temperature. The dimerization probability is the percentage of independent simulations at a given T and E^{es} that dimerize into the native configuration within 20×10^6 MC steps. (b) The probability for a dimer to unfold within 20×10^6 MC steps as a function of T and E^{es}.

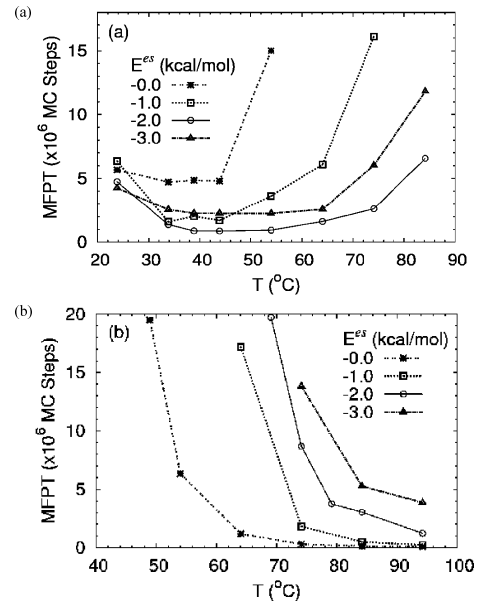

Figure 3. (a) Temperature dependence of the median first passage time (MFPT) for dimerization, as a function of the strength of the ionic interaction strength E^{es}. (b) Temperature dependence of the MFPT for undimerization as a function of E^{es}.

bridges results in unreliable dimerization and supports the experimental work of Meier et al.[57] who found that removing the interchain salt bridge from a de novo designed coiled-coil peptide results in an octameric, globular helical assembly without coiled-coil structure. The result that $E^{es} = 0$ does not allow effective dimerization is easily understandable since an absence of salt bridges leaves short-range hydrophobic side chain interactions as the dominant interaction between chains.

Figure 2a displays the interesting result that if E^{es} is too strong, dimerization is also hindered. Figure 2a shows that dimerization is optimized around $E^{es} = -2.0$ kcal/mol. This is consistent with experimental work[12] in which amino acid substitutions were used to create various salt bridge patterns. In all cases in ref. 12 individual salt bridges strengths were reported to be below 2.0 kcal/mol. In our simulations, we found that for $E^{es} = -2.0$ kcal/mol, dimerization will occur with almost 100% efficiency for a large temperature range of 35 to 75 °C. As shown in Figure 2a, a stronger value of $E^{es} = -3.0$ kcal/mol discourages dimerization and the peptide will not be able to dimerize with 100% probability at any temperature. These results show that the ionic interaction plays a complicated role that does not produce a simple linear, monotonically enhancing effect on the dimerization process. In order to understand the molecular basis of the existence of an optimal salt bridge strength at intermediate values, we carried out further investigations.

Dimer Stability. We investigated the importance of the electrostatic salt bridge on dimer stability by varying E^{es} while keeping the strengths of other interactions constant. The results are displayed in Figure 2b. We found that the strength of the salt bridges has an unexpected effect on the stability of the dimer: stronger E^{es} increases the stability of a dimer, but with a limit.

We started a large group of simulations in the native dimer configuration, defined by the parameter values of $Q = 9$ native interchain side chain contacts with eight of them stabilizing hydrophobic−hydrophobic contacts, $q = 74$ amino acids (37 on each chain) in α-helical configuration, and three pairs of native interchain salt bridges. We defined the dissociated (undimerized) state as having no native interchain contacts, $Q = 0$, and no interchain salt bridges. In Figure 2b we present the results for the probability to dissociate as a function of T and E^{es}. For $E^{es} = 0.0$ kcal/mol (no charged residues), the hydrophobic interchain interactions maintain good stability of the dimer up to $T \sim 30$ °C. For $T > 40$ °C, a dimer is very unstable and likely to unfold. For $E^{es} = -1.0$ kcal/mol, dimers are very stable up to 50 °C. For $E^{es} = -2.0$ kcal/mol, dimers remain stable and do not unfold unless $T > 60$ °C. However, this is the limit of the ability of salt bridge strength to increase dimer stability; if it is further increased in strength to $E^{es} = -3.0$ kcal/mol, there is no appreciable increase in stability and above 60 °C dimers start to dissociate as with $E^{es} = -2.0$ kcal/mol.

The simulation results show the stabilization effect of interchain salt bridges on the leucine zipper dimer as observed in experiments.[12,23,38] Our curves in Figure 2b display cooperative undimerization with similar melting temperatures to those found experimentally in Figure 2 of ref 12. Also in common with Figure 2 of ref 12, our results confirm that stronger salt bridges increase dimer stability. Additionally, our results show that, above a certain strength (2.0 kcal/mol in our simulations), there is little additional benefit to dimer stability gained from stronger salt bridges. Combined with our results in Figure 2a that show that salt bridges that are too strong may hinder the dimerization process, 2.0 kcal/mol may be the optimal value for dimerization speed and stability for the GCN4-p1 system. We investigated the intricacies of the effects of the salt bridges further by looking at the kinetics of dimerization.

Dimerization and Dissociation Kinetics. The temperature dependencies of the dimerization time and dissociation time provide further insight. The time given in Figures 3 is the median first passage time (MFPT). The dimerizing MFPT of Figure 3a is the time at which half of the simulations have dimerized. The trends of the MFPT are similar for all E^{es}. At low temperature, the dimerizing MFPT is high because the chains

have low thermal energy and explore configuration space very slowly in small steps, and also have trouble escaping from shallow configurational traps. At high temperatures, the dimerizing MFPT is also high because the chains have high thermal energy and easily fluctuate throughout all of configuration space. Under these conditions, low energy intermediate configurations that lead to the native state are not especially helpful because the chain can easily jump to numerous different configurations, many of which do not lead toward the native state. The dimerization process is quickest, with the lowest MFPT, at intermediate temperatures. At intermediate temperatures the thermal energy is low enough so that the low energy of the native state is an important advantage, but the temperature is high enough so that non-native configurations can be escaped and do not act as long-lived kinetic traps.

Comparing the dimerizing MFPT for different E^{es} at the same temperature enables us to tell which salt bridge strength most accelerates the dimerization. Ibarra-Molero et al.[23] have reported that salt bridges do not accelerate the folding of the GCN4-p1 leucine zipper at $T = 15$ °C. Our low temperature, 25 °C, results in Figure 3a show a similar lack of dependence of MFPT on the strength of E^{es}. At intermediate temperatures, our results show that dimerization occurs faster than at low temperatures and the dimerization time for all strengths of E^{es} are similar, except for $E^{es} = 0$, which is significantly slower. The absence of interchain salt bridges makes it harder for chains to find each other in the proper arrangement. In addition, the repulsion between non-native same-charged residues in the coiled-coil structure has been found experimentally to be important for the correct alignment of the chains in the dimer.[45,52,58] When the two chains approach with $E^{es} = 0$, the absence of salt bridges removes an important barrier against misalignment of the two chains into structures that form many non-native hydrophobic interactions and act as kinetic traps.

In contrast to low and intermediate temperatures, at high temperatures (>55 °C) the strength of the salt bridge plays a significant role in the kinetics. As with the probability for dimerization shown in Figure 2a, the time for dimerization is also optimized at the value $E^{es} = -2.0$ kcal/mol, as shown in Figure 3a. The dimerization process is slower at both $E^{es} = -1.0$ kcal/mol and $E^{es} = -3.0$ kcal/mol, with the effect becoming larger at higher temperatures.

The MFPT for dissociation is defined as the time at which half of the simulations starting in the same native dimer configuration have undimerized. We present the results for the dissociation MFPT in Figure 3b. Unlike the dimerizing MFPT of Figure 3a, in Figure 3b there is a monotonic dependence on E^{es}, with stronger E^{es} having longer dissociation MFPT. This is consistent with Figure 2b, showing that stronger salt bridges result in more stable dimers once they have formed.

Structural Intermediates. Reliable dimerization requires an efficient kinetic process to find the native state quickly as well as subsequent thermodynamic stability of the product. Figures 2b and 3b show that stronger salt bridges enhance the stability of dimers once they have formed. However, Figures 2a and 3a show that, at higher temperatures, very high salt bridge strength hinders the dimerization process. At high temperatures, intermediate salt bridge strength is optimal for efficient dimerization; both weaker and stronger E^{es} strength slow the dimerization process and make it less probable. In order to understand the molecular basis of the existence of an optimal salt bridge strength at intermediate values, we carried out investigations of the structural intermediates.

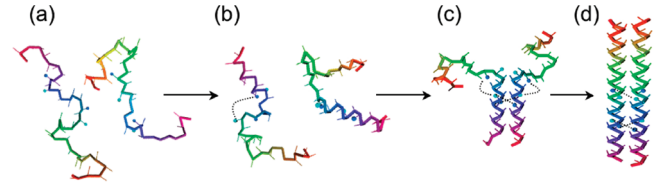

Figure 4. Schematic view of the different dimerizing stages showing the effect of salt bridges. (a) The initial random coil configuration of the separated chains at the beginning of the simulations without any interchain or intrachain salt bridges. (b) Formation of an α-helix in the trigger sequence of each chain along with a non-native intrachain salt bridge. (c) Formation of two native interchain salt bridges along with non-native intrachain salt bridges. (d) Formation of the full dimer with the three native interchain salt bridges.

Experiments have found that the folding and unfolding of the GCN4 leucine zipper is not a two-state process.[35,45] In Figure 4, we present the major structural steps along the dimerization pathway. Starting from separated random coils (Figure 4a), each trigger sequence first folds separately into a partial α-helix (Figure 4b). Then, the α-helices of the two trigger sequences interact through hydrophobic side chains and salt bridges and assume the correct interchain alignment, Figure 4c, to form a partial dimer. In the final step, each chain finishes forming its full α-helix and additional interchain contacts form the entire, long hydrophobic core and native dimer (Figure 4d).

The multistate dimerization process includes the important step of the formation of the partial dimer of Figure 4c that acts as an anchor for the formation of the rest of the dimer. This partial dimer requires amino acids in the trigger sequence to have enhanced α-helical propensity. In addition, salt bridges are important in stabilizing the intermediate structure of Figure 4c. In the configuration of Figure 4c, two of the three native interchain salt bridges have formed and involve amino acids within the trigger sequences. The interactions between the charged residues ensure the correct interchain alignment of the trigger segments by utilizing the electrostatic repulsion from same-charged residues to discourage non-native structures and salt bridges between oppositely charged residues to stabilize the proper configuration. The third native salt bridge, which has not yet formed in Figure 4c, involves amino acids at the edge of the trigger sequence.

Examining many different simulations, we found the molecular structure of an intermediate configuration that is especially detrimental to dimerization at higher temperatures for large E^{es}. The molecular configuration is stabilized by non-native salt bridges that produce kinetic traps. Usually, kinetic traps are less important at high temperatures when the system has high thermal energy and can easily escape. In the dimerization process investigated here, the interesting situation occurs in which an especially deep kinetic trap, involving non-native intrachain salt bridges, is only likely to occur at high temperature and for strong salt bridge interaction strength E^{es}. The non-native, unfavorable, intrachain salt bridges are shown in Figure 4c. They occur within the same chain between residue 20(+) and residue 28(−) or between residue 22(−) and residue 30(+). These non-native salt bridges require that a chain folds back on itself in a contorted shape. This requires large excursions in configuration space and is more probable at high temperatures. The depth of this kinetic trap, and difficulty in escaping, is enhanced by stronger E^{es}. This explains why dimerization is not monotonically enhanced for increasing E^{es} but instead is most efficient for intermediate $E^{es} = -2.0$ kcal/mol as shown in Figures 2a and 3a. Stronger salt bridges stabilize the native state, but also stabilize the kinetic trap. The highly contorted structure of this kinetic trap also

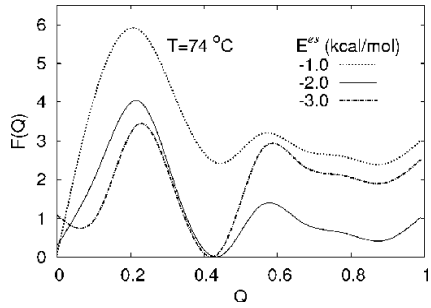

Figure 5. Free energy landscapes as a function of Q (presented as a fraction of the maximum of nine native contacts) for different values of E^{es} computed at the same simulation temperature $T_s = 74$ °C. The weak $E^{es} = -1.0$ kcal/mol curve shows the undimerized state ($Q \sim 0$) to be more stable than the intermediate partially dimerized state ($Q \sim 0.4$) and fully dimerized native state ($Q \sim 1$). There is a huge free energy barrier between the undimerized state and the intermediate state. Intermediate $E^{es} = -2.0$ has lower free energy barriers and the intermediate state and native state are more stable. For strong $E^{es} = -3.0$, the intermediate state is more stable than the native state, and there is a big free energy barrier between them.

explains why it does not produce large anomalies during the dissociation process that is investigated in Figures 2b and 3b. For those simulations, the system was initialized to start in the dimer configuration and it is unlikely to find the contorted kinetic trap structure during the process of dissociation.

Our work shows that it is possible to form stable, non-native structures at different biochemical conditions, as found in the recent experiments described in ref 39 in which the new LZ_{GCN4} x-form was discovered. The non-native structure that we describe has similarities with the x-form. Both are less structured than the native coiled coil and both retain a structured α-helical core. Both our non-native configuration and the x-form can interchange configuration and be in dynamic equilibrium with the native coiled coil. The structure we display in Figure 4c is dimeric and there is some evidence that the x-form is dimeric, but it is not yet certain. This may not be an issue because it is possible that the non-native salt bridges in Figure 4c may also produce stable monomeric structures. An important difference may be that our non-native intermediate structure contains α-helical structure that spans the trigger sequence whereas the x-form appears to have two smaller helical segments; one within the trigger sequence but the other outside. The monomeric versus dimeric question, as well as the location of the helical segments, show that more experimental and computational work is required to elucidate the details of the non-native intermediates.

Free Energy Landscape. We further quantified the dimerization process and the effect of structural kinetic traps by calculating the free energy landscape as a function of various structural parameters. The free energy as a function of a structural parameter x is calculated using $F(x_i) = -kT \ln P(x_i) + c$, where $P(x_i)$ is the probability for a specific value of x to occur and c is an arbitrary constant. The probability $P(x_i)$ is determined by counting the number of MC steps during a simulation in which the system is in a configuration with the specific value x_i. Different structural parameters can be used to elucidate the free energy landscape of such multichain systems.[67] These include helicity q, native interchain contacts Q, the distance between the center of masses of the two chains d_{cm}, and the total energy of the two-chain system E. We found that the native interchain contact parameter Q is especially valuable for detailed investigation of the free energy landscape.

Figure 5 shows free energy profiles for different values of E^{es} computed at the same simulation temperature $T_s = 74$ °C. This temperature is chosen because it is the only temperature at which we can extensively explore both the unfolded and folded dimer configurational space for all values of E^{es}. At higher T, we cannot explore the free energy landscape for the dimer for weaker values of E^{es} (−1.0 kcal/mol) because almost no simulations fold, as seen in Figure 2a. On the other hand, at lower temperatures, all the simulations for strong E^{es} (−3.0 kcal/mol) dimerize quickly and remain folded with high probability. This prevents investigation of the unfolded configuration space.

Figure 5 elucidates the configurations participating in the multistate nature of the GCN4 leucine zipper. There are three states involved: dissociated state with $Q \sim 0$, partially dimerized intermediate state with $Q \sim 4/9$ (0.44), and the native dimer state with $Q \sim 9/9$ (1.0). Higher values of Q imply more organized structure with lower entropy S. Free energy is a balance between entropy and enthalpy: $F = E - TS$. Figure 5 shows that for weak E^{es} (−1.0 kcal/mol) the dissociated, high S, state is more stable with lower F than the intermediate state and low S native state. This is because for low E^{es}, the loss of S upon dimerization is not compensated for by a sufficient decrease in E. The free energy barrier at $Q = 2/9$ (0.22) is especially high because the formation of two interchain salt bridges greatly decreases S, but results in only a small decrease in E. The rate-limiting step that makes dimerization slow and unlikely for weak E^{es} is the crossing of the barrier from $Q = 0$ to $Q = 4/9$.

Dimerization for strong E^{es} (−3.0 kcal/mol) is slow because it requires crossing two free energy barriers. The first barrier again occurs in crossing from $Q = 0$ to $Q = 4/9$. This barrier is not as high as for weak $E^{es} = -1.0$ kcal/mol because the decrease in S is better compensated when the salt bridges are stronger and each releases $E^{es} = -3.0$ kcal/mol. This lowers the free energy barrier at $Q = 2/9$ and allows quick, partial dimerization to the $Q = 4/9$ intermediate of Figure 4c. However, once the system attains the intermediate configuration, it faces a significant barrier that hinders the dimerization to the native configuration at $Q = 9/9$. This is because the structure with $Q = 4/9$ permits the formation of non-native salt bridges at high T. Strong E^{es} makes these non-native salt bridges especially difficult to break and results in the $Q = 4/9$ kinetic trap structure described above and displayed in Figure 4c. To progress to $Q = 5/9$ requires breaking these non-native salt bridges. The free energy barrier hindering the formation of the remaining native interchain contacts occurs in Figure 5 in the region at $Q = 5/9$.

Figure 5 also explains why $E^{es} = -2.0$ kcal/mol is optimal for dimerization. The free energy barrier for partial dimerization to go from $Q = 0$ to $Q = 4/9$ is significantly lower than for weak $E^{es} = -1.0$ kcal/mol, and almost as low and easy to cross as for strong $E^{es} = -3.0$ kcal/mol. However, unlike for $E^{es} = -3.0$ kcal/mol, the $Q = 4/9$ intermediate partial dimer does not become a deep kinetic trap because non-native salt bridges that might form are not strong. Therefore, there is not a large free energy barrier to cross to continue dimerization from $Q = 4/9$ to the native state $Q = 9/9$. $E^{es} = -2.0$ kcal/mol is optimal because it faces one midsize free energy barrier to initiate dimerization, and a small second barrier to complete the process, whereas $E^{es} = -1.0$ kcal/mol faces a very large initial barrier and $E^{es} = -3.0$ kcal/mol faces two midsize barriers.

We further examined the free energy landscape through three-dimensional contour plots in which the free energy is plotted as a function of two structural parameters, $F(x1,x2)$. In all panels of Figure 6, red color denotes stable, low free energy configurations, and blue denotes high free energy. Figure 6, a−c, shows the contour plots of the free energy, $F(Q,q)$, for different values

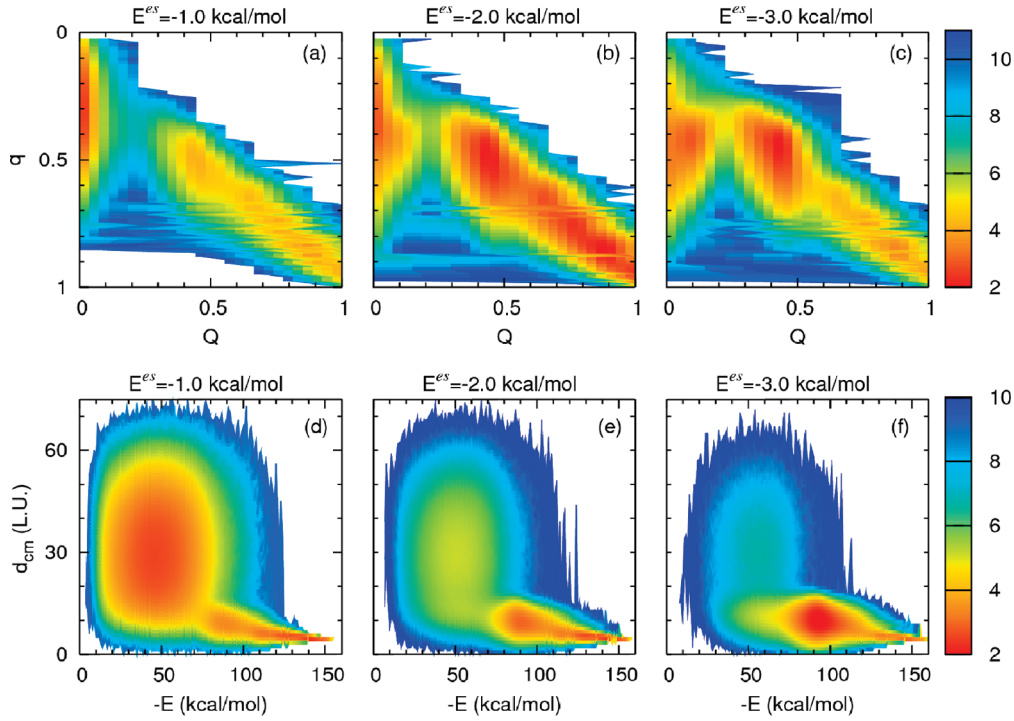

Figure 6. Free energy contour plots as a function of two structural parameters for different values of E^{es} at the same simulation temperature T_s = 74 °C: (a) $F(Q,q)$ as a function of native interchain contacts Q and helicity q with weak $E^{es} = -1.0$ kcal/mol; ((b) $F(Q,q)$ for intermediate $E^{es} = -2.0$ kcal/mol; (c) $F(Q,q)$ for strong $E^{es} = -3.0$ kcal/mol; (d) $F(-E,d_{cm})$ as a function of system energy E and distance between the centers of mass of each chain d_{cm} for weak $E^{es} = -1.0$ kcal/mol; (e) $F(-E,d_{cm})$ for intermediate $E^{es} = -2.0$ kcal/mol; and (f) $F(-E,d_{cm})$ for strong $E^{es} = -3.0$ kcal/mol.

of E^{es} at the same temperature $T = 74$ °C. Figure 6a displays the results for $E^{es} = -1.0$ kcal/mol. The most stable configuration with the lowest F (deepest red color) is the dissociated state with $Q \sim 0$ and $q \sim 0.3$. This is the configuration shown in Figure 4b, where the trigger sequences on each chain independently form partial α-helices, but the two chains are not interacting. The free energy barrier to get to the $Q = 4/9$ intermediate state with higher q (∼0.5) is clearly visible in blue in Figure 6a.

Figure 6b is for $E^{es} = -2.0$ kcal/mol. The native state is the most stable with the lowest free energy, as displayed by the red color. The intermediate configuration with $Q = 4/9$ is almost as stable. Unlike the blue barrier in Figure 6a, there are no strong free energy barriers preventing the system from traveling through configuration space from the dissociated $Q = 0$ and $q = 0.3$ configuration to the native state with $Q = 9/9$ and $q = 74/74$. There is a midsize (yellow-green) barrier in the region around $Q = 2/9$ and $q \sim 0.4$, but almost no barrier between the intermediate $Q = 4/9$ configuration and the native state. This allows dimerization to occur quickly and efficiently.

Figure 6c is for $E^{es} = -3.0$ kcal/mol. The intermediate configuration at $Q = 4/9$ (red) is the most stable. The barrier between the dissociated $Q = q = 0$ configuration and the intermediate $Q = 4/9$ is smaller than for $E^{es} = -2.0$ kcal/mol, but still present (yellow). However, unlike for $E^{es} = -2.0$ kcal/mol, Figure 6c, shows that $E^{es} = -3.0$ kcal/mol must surmount a second midsize barrier to go from the intermediate configuration to the native state.

The free energy, $F(-E,d_{cm})$, for different values of E^{es} at the same $T = 74$ °C is shown in Figure 6d,e,f for values of $E^{es} = -1.0$ kcal/mol, -2.0 kcal/mol, -3.0 kcal/mol, respectively. We see that, in Figure 6d, there is a wide and deep free energy basin at $d_{cm} = 30$. This distance of separation is large and means that the chains are not interacting. The width of this free energy basin makes it hard to escape. In Figure 6f, with $E^{es} = -3.0$ kcal/mol, a deep kinetics trap occurs at small d_{cm} and intermediate E. The small value of d_{cm} means that the chains are close together, but the intermediate value of E means that the system is only partially dimerized. This kinetic traps corresponds to the intermediate $Q = 4/9$ configuration with the non-native intrachain salt bridges of Figure 4c. This region is difficult to escape not because it is wide but because it is very deep. For intermediate $E^{es} = -2.0$ kcal/mol displayed in Figure 6e, the dissociated configuration of large d_{cm} is neither a wide nor deep basin, and the intermediate configuration does not act as a kinetic trap because there is a clear and easy low free energy path to the native state.

Discussion and Conclusions

We have found that the existence of optimal salt bridge strength at intermediate values is due to the balance between two opposing effects. Stronger salt bridges are beneficial because they stabilize the native dimer and guide the dimerization process. However, we also found that stronger salt bridges are detrimental because they stabilize an intermediate configuration that becomes a deep kinetic trap that hinders the system's ability to continue the dimerization process. We found the specific molecular structure that acts as the most important kinetic trap. This intermediate configuration involves non-native intrachain salt bridges that only occur if a chain assumes a contorted shape. If the non-native salt bridges are very strong, the kinetic trap structure is extremely difficult to escape. At low temperature, a chain undergoes small structural fluctuations while dimerizing and the contorted kinetic trap structure is less likely to occur. Therefore, at low temperatures, strong non-native salt bridges are a smaller problem for the dimerization process than at high temperatures.

The non-native kinetic trap structure is especially important at high temperatures. This is because the contorted non-native structure of Figure 4c occupies a small region of the configuration space available to the partially dimerized system. In order for the contorted configuration to occur, before the system correctly dimerizes it must be able to sample large distances in configuration space. This occurs more quickly at high temperatures. It is possible that the same contorted shape may appear at physiological temperatures if the system is observed for a much longer time. Similarly, an important structural transition from α-helical to β-structure has been observed experimentally by Kammerer[59] et al. at high temperatures in the important group of ccβ peptides. It is possible that this structural transition might occur at physiological temperatures, but only after time scales that are much longer than those observed in that work. The time scales for structural transitions depend on both distances and barriers that must be traversed in configuration space. The dependence of the time scale for structural transitions as a function of temperature might be predictable from the free-energy landscapes[48] calculated here, and more fundamentally from energy−entropy landscapes if they can be calculated.[60]

Our results show that stronger interactions that stabilize the native state once it forms are not always better for the efficiency of the folding process. These results are based upon investigations of the GCN4 leucine zipper dimer model, but may be applicable to other systems in regard to the balance between entropy and enthalpy that is important in living systems. Stronger bonds are helpful in protein systems to counteract the loss of entropy when forming organized structures. If the molecular chains are flexible enough to assume a wide range of configurations, stronger interactions that thermodynamically stabilize the native state may also stabilize non-native structures to the extent that they form detrimental kinetic traps that decrease the speed and probability for reaching the native state. Determining the molecular configuration of the important kinetic trap structures can provide guidance for making amino acid substitutions to reduce the importance of the trap.

References and Notes

(1) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10.
(2) Shakhnovich, E. I. *Curr. Opin. Struct. Biol.* **1997**, *7*, 29.
(3) Onuchic, J. N.; Nymeyer, H.; Garcia, A. E.; Chahine, J.; Socci, N. D. *Adv. Protein Chem.* **2000**, *53*, 87.
(4) Oliveberg, M.; Wolynes, P. G. *Q. Rev. Biophys.* **2005**, *38*, 245.
(5) Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3555.
(6) Fersht, A. R.; Daggett, V. *Cell* **2002**, *108*, 573.
(7) Zhou, R.; Huang, X.; Margulis, C. J.; Berne, B. J. *Science* **2004**, *305*, 1605.
(8) MacCallum, J. L.; Moghaddam, M. S.; Chan, H. S.; Tieleman, D. P. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 6206.
(9) Onuchic, J. N.; LutheySchulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545.
(10) Dill, K. A.; Bromberg, S.; Yue, K.; Fiebig, K. M.; Yee, D. P.; Thomas, P. D.; Chan, H. S. *Protein Sci.* **1995**, *4*, 561.
(11) Kolinski, A.; Skolnick, J. *Proteins* **1994**, *18*, 353.
(12) Spek, E. J.; Bui, A. H.; Lu, M.; Kallenbach, N. R. *Protein Sci.* **1998**, *7*, 2431.
(13) Hendsch, Z. S.; Tidor, B. *Protein Sci.* **1994**, *3*, 211.
(14) Waldburger, C. D.; Schildbach, J. F.; Sauer, R. T. *Nat. Struct. Biol.* **1995**, *2*, 122.
(15) Kumar, S.; Nussinov, R. *Proteins: Struct. Funct. Genet.* **2000**, *41*, 485.
(16) Makhatadze, G. I.; Loladze, V. V.; Ermolenko, D. N.; Chen, X. F.; Thomas, S. T. *J. Mol. Biol.* **2003**, *327*, 1135.
(17) Vijayakumar, M.; Zhou, H.-X. *J. Phys. Chem. B* **2001**, *105*, 7334.
(18) Dadarlat, V. M.; Post, C. B. *J. Phys. Chem. B* **2008**, *112*, 6159.
(19) Bosshard, H. R.; Marti, D. N.; Jelesarov, I. *J. Mol. Recognit.* **2004**, *17*, 1.
(20) Kohn, W. D.; Kay, C. M.; Hodges, R. S. *Protein Sci.* **1995**, *4*, 237.
(21) Kammerer, R. A.; Jaravine, V. A.; Frank, S.; Schulthess, T.; Landwehr, R.; Lustig, A.; Garcia-Echeverria, C.; Alexandrescu, A. T.; Engel, J.; Steinmetz, M. O. *J. Biol. Chem.* **2001**, *276*, 13685.
(22) Phelan, P.; Gorfe, A. A.; Jelesarov, I.; Marti, D. N.; Warwicker, J.; Bosshard, H. R. *Biochemistry* **2002**, *41*, 2998.
(23) Ibarra-Molero, B.; Zitzewitz, J. A.; Matthews, C. R. *J. Mol. Biol.* **2004**, *336*, 989.
(24) Paoli, B.; Seeber, M.; Backus, E. H.; Ihalainen, J. A.; Hamm, P.; Caflisch, A. *J. Phys. Chem. B* **2009**, *113*, 4435.
(25) Bjelic, S.; Wieninger, S.; Jelesarov, I.; Karshikoff, A. *Proteins* **2008**, *70*, 810.
(26) Burkhard, P.; Stetefeld, J.; Strelkov, S. V. *Trends Cell Biol.* **2001**, *11*, 82.
(27) Yu, Y. B. *Adv. Drug Delivery Rev.* **2002**, *54*, 1113.
(28) Mason, J. M.; Arndt, K. M. *Chembiochem* **2004**, *5*, 170.
(29) Lupas, A. N.; Gruber, M. *Adv. Protein Chem.* **2005**, *70*, 37.
(30) Steinmetz, M. O.; Jelesarov, I.; Matousek, W. M.; Honnappa, S.; Jahnke, W.; Missimer, J. H.; Frank, S.; Alexandrescu, A. T.; Kammerer, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 7062.
(31) Su, L.; Cukier, R. I. *J Phys Chem B* **2009**, *113*, 9595.
(32) O'Shea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T. *Science* **1991**, *254*, 539.
(33) Vinals, J.; Kolinski, A.; Skolnick, J. *Biophys. J.* **2002**, *83*, 2801.
(34) Missimer, J. H.; Steinmetz, M. O.; Jahnke, W.; Winkler, F. K.; van Gunsteren, W. F.; Daura, X. *Chem. Biodiversity* **2005**, *2*, 1086.
(35) Wang, T.; Lau, W. L.; DeGrado, W. F.; Gai, F. *Biophys. J.* **2005**, *89*, 4180.
(36) Liu, Y.; Chapagain, P. P.; Parra, J. L.; Gerstman, B. S. *J. Chem. Phys.* **2008**, *128*, 045106.
(37) Lumb, K. J.; Kim, P. S. *Science* **1995**, *268*, 436.
(38) Matousek, W. M.; Ciani, B.; Fitch, C. A.; Garcia-Moreno, B.; Kammerer, R. A.; Alexandrescu, A. T. *J. Mol. Biol.* **2007**, *374*, 206.
(39) Nikolaev, Y.; Pervushin, K. *J. Am. Chem. Soc.* **2007**, *129*, 6461.
(40) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
(41) Kolinski, A.; Skolnick, J. *Proteins* **1994**, *18*, 338.
(42) Kolinski, A.; Milik, M.; Skolnick, J. *J. Chem. Phys.* **1991**, *94*, 3978.
(43) Chapagain, P. P.; Gerstman, B. S. *J. Chem. Phys.* **2004**, *120*, 2475.
(44) Gerstman, B. S.; Chapagain, P. P. *J. Chem. Phys.* **2005**, *123*, 054901.
(45) Dragan, A. I.; Privalov, P. L. *J. Mol. Biol.* **2002**, *321*, 891.
(46) Lee, D. L.; Ivaninskii, S.; Burkhard, P.; Hodges, R. S. *Protein Sci.* **2003**, *12*, 1395.
(47) Chapagain, P. P.; Liu, Y.; Gerstman, B. S. *J. Chem. Phys.* **2008**, *129*, 175103.
(48) Chapagain, P. P.; Gerstman, B. S. *Biopolymers* **2006**, *81*, 167.
(49) Kammerer, R. A.; Schulthess, T.; Landwehr, R.; Lustig, A.; Engel, J.; Aebi, U.; Steinmetz, M. O. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13419.
(50) Pace, C. N.; Scholtz, J. M. *Biophys. J.* **1998**, *75*, 422.
(51) Wang, J.; Feng, J. A. *Protein Eng.* **2003**, *16*, 799.
(52) Kohn, W. D.; Monera, O. D.; Kay, C. M.; Hodges, R. S. *J. Biol. Chem.* **1995**, *270*, 25495.
(53) Lavigne, P.; Sonnichsen, F. D.; Kay, C. M.; Hodges, R. S. *Science* **1996**, *271*, 1136.
(54) Wendt, H.; Leder, L.; Harma, H.; Jelesarov, I.; Baici, A.; Bosshard, H. R. *Biochemistry* **1997**, *36*, 204.
(55) Kenar, K. T.; Garcia-Moreno, B.; Freire, E. *Protein Sci.* **1995**, *4*, 1934.
(56) Jelesarov, I.; Durr, E.; Thomas, R. M.; Bosshard, H. R. *Biochemistry* **1998**, *37*, 7539.
(57) Meier, M.; Lustig, A.; Aebi, U.; Burkhard, P. *J. Struct. Biol.* **2002**, *137*, 65.
(58) Marti, D. N.; Bosshard, H. R. *Biochemistry* **2004**, *43*, 12436.
(59) Kammerer, R. A.; Kostrewa, D.; Zurdo, J.; Detken, A.; Garcia-Echeverria, C.; Green, J. D.; Muller, S. A.; Meier, B. H.; Winkler, F. K.; Dobson, C. M.; Steinmetz, M. O. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 4435.
(60) Chapagain, P. P.; Parra, J. L.; Gerstman, B. S.; Liu, Y. *J. Chem. Phys.* **2007**, *127*, 075103.

JP909872A