Temperature Dependence of the Folding and Unfolding Kinetics of the GCN4 Leucine Zipper via $^{13}C^{\alpha}$-NMR

Marilyn Emerson Holtzer,* G. Larry Bretthorst,* D. André d'Avignon,* Ruth Hogue Angeletti,† Lisa Mints,† and Alfred Holtzer*

*Department of Chemistry, Washington University, St. Louis, Missouri 63130-4899; and †Laboratory for Macromolecular Analysis and Proteomics, Albert Einstein College of Medicine, Bronx, New York 10461 USA

ABSTRACT Studies by one-dimensional NMR are reported on the interconversion of folded and unfolded forms of the GCN4 leucine zipper in neutral saline buffer. The peptide bears 99% $^{13}C^{\alpha}$ labels at three sites: V9, L12, and G31. Time-domain $^{13}C^{\alpha}$-NMR spectra are interpreted by global Bayesian lineshape analysis to extract the rate constants for both unfolding and folding as functions of temperature in the range 47–71°C. The data are well fit by the assumption that the same rate constants apply at each labeled site, confirming that only two conformational states need be considered. Results show that 1) both processes require a free energy of activation; 2) unfolding is kinetically enthalpy-opposed and entropy-driven, while folding is the opposite; and 3) the transition state dimer ensemble averages ~40% helical. The activation parameters for unfolding, derived from NMR data at the elevated temperatures where both conformations are populated, lead to estimates of the rate constant at low temperatures (5–15°C) that agree with extant values determined by stopped-flow CD via dilution from denaturing media. However, the corresponding estimated values for the folding rate constant are larger by two to three orders of magnitude than those obtained by stopped flow. We propose that this apparent disagreement is caused by the necessity, in the stopped-flow experiment, for initiation of new helices as the highly denaturant-unfolded molecule adjusts to the newly created benign solvent conditions. This must reduce the success rate of collisions in producing the folded molecule. In the NMR determinations, however, the unfolded chains always have a small, but essential, helix content that makes such initiation unnecessary. Support for this hypothesis is adduced from recent extant experiments on the helix-coil transition in single-chain helical peptides and from demonstration that the folding rate constants for coiled coils, as obtained by stopped flow, are influenced by the nature of the denaturant used.

INTRODUCTION

The leucine zipper of the GCN4 transcription factor (called here GCN4-lz) has become one of the most often used models for examining questions concerning the thermodynamics and kinetics of protein folding. Its popularity stems both from its structural simplicity and from our relatively reliable understanding of the relationship between amino acid sequence and structure in proteins of its type (Crick, 1953; McLachlan and Stewart, 1975).

Structurally, GCN4-lz is a coiled coil in benign media; it consists of two 33-residue right-handed α-helical chains, arranged side by side in parallel and register, and with a slight negative supertwist (O'Shea et al., 1989, 1991). Such structures are known to result from sequences with a pseudo-heptad repeat, the amino acids of which are designated *abcdefg,* and in which residues *a* and *d* are hydrophobic and *e* and *g* oppositely charged (McLachlan and Stewart, 1975; Lupas, 1996). The resulting amphipathicity of the helices leads to helix-helix interactions that direct parallel, registered dimerization. GCN4-lz is also a leucine zipper, a type of coiled coil in which the *d* residue positions are predominantly occupied by leucines.

Received for publication 10 July 2000 and in final form 27 November 2000.

Address reprint requests to Dr. Alfred Holtzer, Dept. of Chemistry, Washington University, One Brookings Dr., Campus Box 1134, St. Louis, MO 63130-4899. Tel.: 314-935-6572; Fax: 314-935-4481; E-mail: holtzer@wuchem.wustl.edu.

0006-3495/01/02/939/13 $2.00

The structurally simplest models for protein folding are the synthetic polypeptides, many of which form single α-helices in aqueous or organic media and which undergo a helix-random coil transition under appropriate conditions (Poland and Scheraga, 1970). However, the population of conformational states in equilibrium in such transitions is actually far more complex than in most proteins. The conformational population in small single-chain globular protein unfolding transitions, for example, is often well described by a two-state model. That is, the molecular population consists of a mixture of native molecules and a vast array of rapidly interconverting, highly disordered conformations approximating the random coil of polymer chemistry. In the case of synthetic polypeptides, however, the equilibrium population comprises a vast array of partially folded, partially disordered chains.

Coiled coils are structurally intermediate in complexity between the topologically complex globular proteins and single chain helices. One simply has two helices side by side. Is the population of states in coiled-coil unfolding equilibria more like globular proteins or more like single-chain peptide helices? That is among the questions we address here.

Studies of the thermodynamics and kinetics of unfolding equilibria in GCN4-like peptides already exist. Some studies, using stopped-flow methods, provide evidence that the transition is two-state, i.e., that the kinetics of both unfolding and folding are satisfactorily represented by the respective elementary uni- and bi-molecular forward and back-

ward processes: F ⇄ 2U, wherein F signifies an intact coiled-coil dimer and U an unfolded monomeric chain (Zitzewitz et al., 1995, 2000; Sosnick et al., 1996; Moran et al., 1999). However, studies by $^{13}C^{\alpha}$-NMR of a GCN4-like peptide (called here GCN4-lzK) show plainly that thermal unfolding curves differ from site to site on the molecule and that more than one *folded* form exists at most sites, findings incompatible with a two-state population (Lovett et al., 1996; Holtzer et al., 1997; d'Avignon et al., 1998, 1999).

These various findings raise questions as to whether observed differences between laboratories are caused by variations in the precise sequences and/or conditions used in these various studies or by something deeper. Various laboratories have referred to somewhat different peptides by the same name, i.e., "GCN4 leucine zipper." Originally, the peptide sequence Ac-RMKQLEDKVEELLSKNYHLENEVARLKKLVGER-OH was referred to verbally by that name and symbolically as GCN4-p1 (O'Shea et al., 1989, 1991). Other authors have used mutants that feature different end-capping (note that GCN4-p1 is capped at the N- but not the C-terminus) or contain other mutated sites for special purposes, e.g., a Y17W mutation to enhance fluorescence (Sosnick et al., 1996; Moran et al., 1999). The $^{13}C^{\alpha}$-NMR studies referred to above use a GCN4-like sequence (GCN4-lzK) that not only includes a C-terminal amide, but also the conservative mutations R1K, H18K, R25K, and R33K. These lysine substitutions result in reduction of the melting temperature of the coiled coil, a feature that lessens sample damage during accumulation of NMR data in thermal unfolding experiments.

The experiments reported on in this paper were all performed on peptide with the original sequence and capping of GCN4-p1, as given explicitly above, except that it contains 99% $^{13}C^{\alpha}$ at sites V9(*a*), L12(*d*), and G31(*b*). This peptide will be referred to here as GCN4-lz, notation that facilitates comparison with our earlier work on its relative GCN4-lzK.

In addition to capping and sequence variation, literature reports vary in the solvent media used for folding studies. Despite the known and powerful influence of solvent on folding, studies very often vary the medium drastically; often, one has to read the paper in considerable detail to find out that a study has been conducted in a medium far from physiological, e.g., at pH 5.5 or lower (Sosnick et al., 1996; Moran et al., 1999). All the experiments we report on here were in the medium $(NaCl)_{100}(NaPi)_{500}(D_2O)_{5514}(7.4)$, wherein we designate complex aqueous media using the formula for each additive with its millimolarity as subscript, followed by the pH in parentheses. The 10% D_2O included in the medium for NMR purposes does not alter the transition (Lovett et al., 1996; Holtzer et al., 1997).

If an unfolding process is strictly two-state, any structured site on the chain will report the same unfolding process. If the conformational population is more complex, one expects different sites to report differently. Insight into the population of states can therefore be gained by using methods that report on folding from a local point of view. Unfortunately, such local probes are not common in native proteins. Aromatic residues are few, and only tryptophan fluoresces strongly. Attached spin or fluorescent labels are invariably large hydrophobes and may report only on what they themselves have caused. Substitution of ^{13}C for ^{12}C, however, is conformationally neutral, so selective substitution provides unperturbed site-specific information.

NMR has other advantages as well. Whether one uses saturation-transfer, spin-inversion transfer, or line-shape analysis to extract dynamic information, only spins are manipulated. The rate information that results therefore refers to the undisturbed chemically equilibrated system. Such information therefore complements studies of folding that involve starting from a denaturant-unfolded state. The NMR studies therefore do not involve an initial collapse of the unfolded form from a thoroughly unfolded, denaturant-swollen state to one characteristic of the normal aqueous environment.

Finally, we emphasize that many studies of the folding kinetics of proteins, including coiled coils, have drawn conclusions about transition states from experiments at a single temperature. However, a crucial test of the entire transition-state idea is in the temperature dependence of rate constants. Without a knowledge of the activation parameters, conclusions on transition states must remain uncertain. However, one must also recognize that studies of strictly thermal unfolding may bring the system into nonphysiological regions of temperature, emphasizing the complementarity of the various approaches.

With these objectives in mind, we report here on studies of the temperature dependence of the rate constants for folding and unfolding of GCN4-lz in near-neutral saline buffer via analysis of the $^{13}C^{\alpha}$-NMR line shapes at sites V9(*a*), L12(*d*), and G31(*b*).

MATERIALS AND METHODS

Synthesis, purification, and characterization of GCN4-lz labeled at V9(*a*), L12(*d*), and G31(*b*)

Peptides were prepared by solid-phase synthesis on an Applied Biosystems 433A peptide synthesizer using N^{α}-9-fluorenylmethyloxycarbonyl (Fmoc) strategies on Wang resins. All reagents were purchased from Perkin-Elmer (Foster City, CA). Syntheses were on 0.1 mmol of resin, but using coupling protocols standard for 0.25 mmol synthesis. N-terminal acetylation was carried out before deprotection and cleavage by reacting with acetic anhydride in diethylamine and *N*-hydroxybenzotriazole dissolved in *N*-methylpyrrolidone. Peptides were deprotected and cleaved from the resins by incubation for 2 h in trifluoracetic acid/thioanisole/ethanedithiol/phenol/water::83.3:4.2:2.1:6.2:4.2. After precipitation and washing with methyl-*tert*-butyl ether at 4°C, the peptides were lyophilized before further purification. Each peptide was characterized by amino acid analysis, analytical HPLC, and electrospray ionization and matrix-assisted laser desorption mass spectrometry.

The $^{13}C^{\alpha}$-labeled F moc-amino acids were prepared as previously described (Lovett et al., 1996). Protocols for synthesizing the labeled

peptides were the same as described above, except that the labeled amino acids were dissolved in 2 ml of *N*-methylpyrrolidone before adding to the amino acid cartridges, and the activation reagents, for those coupling cycles only, were reduced fourfold in concentration.

Our peptides were brought to at least 99% purity by reversed-phase HPLC using a Beckman system equipped with a Vydac preparative (C18, 22 × 250 mm) column for polypeptides. The purity of individual fractions was assessed by analytical reversed-phase HPLC (Vydac C18, 4 × 250 mm), UV spectrophotometry, and mass spectrometry. The molecular masses obtained were within 1 Da of the calculated value, 4042 Da, for peptides bearing three $^{13}C^{\alpha}$ labels.

For both the CD and NMR experiments described below, analytical HPLC was routinely used to check for sample damage after experiments that exposed solutions to temperatures greater than ambient for extended periods. Samples kept at temperatures above 50°C for several hours are particularly vulnerable, and such checks, and checks for reversibility, are essential to assure that the measurements are done on intact samples.

Peptide concentration was assessed by measurement of absorbance at 275 nm, using an extinction coefficient for the tyrosine in GCN4-lz in benign media of 1.40 mM^{-1} cm^{-1}. This value was obtained by alkaline titration of tyrosine in denaturing media by the Edelhoch method (Edelhoch, 1967). Because GCN4-lz has one tyrosine per chain, tyrosine and chain molar-basis concentrations are the same.

CD

Digitized CD data were obtained using a Jasco (Easton, MD) J500A spectropolarimeter equipped with a Jasco IF-500 interface. Procedures for CD data acquisition and analysis have been described earlier (Holtzer et al., 1995). CD data for natural abundance GCN4-lz and for labeled samples are indistinguishable.

$^{13}C^{\alpha}$-NMR

All ^{13}C-NMR data were recorded at 11.4 Tesla (^{13}C resonant frequency: 125.703 MHz), using a Varian (Palo Alto, CA) Unity-Plus 500 spectrometer equipped with a Nalorac (Martinez, CA) 10 mm probe and an Oxford (Oxford, UK) temperature controller. Before measurements, 1.2 ml GCN4-lz in $(NaCl)_{100}(NaPi)_{50}(D_2O)_{5514}$(7.4) was placed in a 10 mm matched susceptibility Shigemi (Allison Park, PA) tube and adjusted to a sample height of no less than 16 mm. The sample was allowed to equilibrate to the desired temperature for at least 20 min before recording NMR data. Collection conditions included 6400–8300 transients at each temperature except those ≥70°C (4000 transients); 17.5 μs pulse length (90° flip angle); spectral width 12,686 Hz; 1.5 s preacquisition delay time; and broad-band (GARP) proton decoupling. During a multi-temperature NMR run, the sample was periodically equilibrated at 25°C and the NMR spectrum recorded. This allows assessment of the sample integrity to ensure that damage had not been sustained by prolonged exposure to elevated temperature. The gas bubble generated in the Shigemi tube in going from ambient conditions to elevated temperature was purged before data collection.

Bayesian data analysis

Basic equations

For an elementary reaction of the form $f \rightleftarrows u$, after a 90° pulse, the Bloch equations for transverse magnetizations contributed by nuclei in each of the two environments, f or u, are (McConnell, 1958; Carrington and McLachlan, 1967):

$$d\hat{\mathbf{M}}'_f/dt = -[i(\hat{\omega}_f - \omega) + k_{fu}]\hat{\mathbf{M}}'_f + k_{uf}\hat{\mathbf{M}}'_u \qquad (1)$$

$$d\hat{\mathbf{M}}'_u/dt = -[i(\hat{\omega}_u - \omega) + k_{uf}]\hat{\mathbf{M}}'_u + k_{fu}\hat{\mathbf{M}}'_u$$

wherein $\hat{M}'_j$ is the complex (signified by the circumflex) transverse magnetization in the rotating frame (signified by the prime) due to nuclei in conformational environment j; $\hat{\omega}_j$ is the complex Larmor angular velocity of nuclei in environment j, given by $\hat{\omega}_j = \omega_j - ir_{2j}$, in which ω_j is the Larmor angular velocity and r_{2j} the transverse relaxation rate constant (reciprocal of the T_2 relaxation time) at j; ω is the angular velocity of the rotating frame (i.e., the spectrometer angular velocity); and k_{fu} and k_{uf} are the rate constants for the forward and reverse reactions, respectively. These Bloch equations can be put into more elegant form (à la Rudin and Sauter, 1992) by introducing the definitions:

$$\hat{S}_f = i(\hat{\omega}_f - \omega) + k_{fu} \qquad (2)$$

$$\hat{S}_u = i(\hat{\omega}_u - \omega) + k_{uf}$$

and the matrix:

$$\hat{\mathbf{M}}' = \begin{pmatrix} \hat{M}'_f \\ \hat{M}'_u \end{pmatrix} \qquad (3)$$

into Eq. 1, which then becomes:

$$d\hat{\mathbf{M}}'/dt = \hat{\mathbf{X}}\hat{\mathbf{M}}' \qquad (4)$$

wherein $\hat{\mathbf{X}}$, the kinetic matrix, is:

$$\hat{\mathbf{X}} = \begin{pmatrix} -\hat{S}_f & k_{uf} \\ k_{fu} & -\hat{S}_u \end{pmatrix}. \qquad (5)$$

The form of differential equation (4) is identical to one encountered in earlier work from this laboratory involving spin inversion transfer experiments (d'Avignon et al., 1998, 1999). However, in that earlier context, all of the quantities involved are real, whereas in the present case, several are complex. Nevertheless, the solutions to Eq. 4 are formally identical to those found earlier:

$$\hat{\mathbf{M}}' = \hat{\mathbf{V}}^{-1}\begin{pmatrix} e^{\hat{a}_1 t} & 0 \\ 0 & e^{\hat{a}_2 t} \end{pmatrix}\hat{\mathbf{V}}\hat{\mathbf{M}}'(0), \qquad (6)$$

in which $\hat{\mathbf{V}}$ and $\hat{\mathbf{V}}^{-1}$ are matrices capable of diagonalizing the kinetic matrix, i.e.,

$$\hat{\mathbf{V}}\hat{\mathbf{X}}\hat{\mathbf{V}}^{-1} = \begin{pmatrix} \hat{a}_1 & 0 \\ 0 & \hat{a}_2 \end{pmatrix} \qquad (7)$$

with:

$$\hat{\mathbf{V}} = \frac{1}{\hat{a}_1 - \hat{a}_2}\begin{pmatrix} -(\hat{S}_f + \hat{a}_2) & k_{uf} \\ \hat{S}_f + \hat{a}_1 & -k_{uf} \end{pmatrix} \qquad (8)$$

and:

$$\hat{\mathbf{V}}^{-1} = \frac{1}{k_{uf}}\begin{pmatrix} k_{uf} & k_{uf} \\ \hat{S}_f + \hat{a}_1 & \hat{S}_f + \hat{a}_2 \end{pmatrix} \qquad (9)$$

and $\hat{a}_1$ and $\hat{a}_2$ are the eigenvalues of the kinetic matrix, obtained from solutions of the usual quadratic equation:

$$\hat{a}_{1,2} = \frac{-(\hat{S}_f + \hat{S}_u)}{2} \pm \frac{[(\hat{S}_f - \hat{S}_u)^2 + 4k_{fu}k_{uf}]^{1/2}}{2} \qquad (10)$$

The reader is reminded that, since the coefficients in the quadratic are not all real, the two solutions in Eq. 10 are not conjugate pairs.

The relationship of these equations to our system must next be clarified, because, in the case of two-stranded coiled coils, the equation expressing the stoichiometry is of the form:

$$F \rightleftarrows 2U \tag{11}$$

wherein F represents one mole of two-stranded coiled coils and U one mole of unfolded, dissociated chains. In chemical kinetics, the equations for an elementary reaction of this form are:

$$d[F]/dt = -k_{\mathrm{unf}}[F] + k_{\mathrm{fld}}[U]^2 \tag{12a}$$

$$d[U]/dt = 2k_{\mathrm{unf}}[F] - 2k_{\mathrm{fld}}[U]^2 \tag{12b}$$

wherein brackets around a chemical species symbol indicate molar concentration of that species.

The NMR experiment measures magnetizations, not concentrations. Because NMR is concerned only with spins, which are conserved in conformational exchange while molecules are not, the reaction appears to be unimolecular both ways. We therefore need to make the correspondence of the rate constants in Eqs. 12 to those in Eq. 1. The change in the magnetization, per second per liter, of nuclei in the folded environment due to the forward reaction is given by the change in their molar concentration multiplied by the average magnetization those folded species have; therefore:

$$(d\hat{\mathbf{M}}'_{\mathrm{f}}/dt)_{\mathrm{unf}} = (d[F]/dt)_{\mathrm{unf}}(\hat{\mathbf{M}}'_{\mathrm{f}}/[F]) = -k_{\mathrm{unf}}\hat{\mathbf{M}}'_{\mathrm{f}} \tag{13}$$

Likewise, the reverse reaction produces a change in the magnetization of the folded form that is given by the change in the concentration of U molecules responsible for the back reaction, multiplied by their mean magnetization:

$$\begin{aligned}(d\hat{\mathbf{M}}'_{\mathrm{f}}/dt)_{\mathrm{fld}} &= (d[F]/dt)_{\mathrm{fld}}(2\hat{\mathbf{M}}'_{\mathrm{u}}/[U]) \\ &= (k_{\mathrm{fld}}[U]^2)(2\hat{\mathbf{M}}'_{\mathrm{u}}/[U]) \\ &= 2k_{\mathrm{fld}}[U]\hat{\mathbf{M}}'_{\mathrm{u}}\end{aligned} \tag{14}$$

Addition of Eqs. 13 and 14 gives the total rate of change of the magnetization of the folded form caused by conformational exchange. Comparison of these results with the Bloch equations given in Eq. 1 indicates that we have only to make the correspondences: $k_{\mathrm{fu}} = k_{\mathrm{unf}}$ and $k_{\mathrm{uf}} = 2k_{\mathrm{fld}}[U]$, for all the equations governing the simpler stoichiometry to hold exactly. Since the total concentration of conformational form U is obtainable from the rate constants themselves ($k_{\mathrm{unf}}/k_{\mathrm{fld}}$ being just the equilibrium constant) and the total peptide concentration, the conformational rate constants (k_{unf}, k_{fld}) follow from determination of the NMR rate constants (k_{fu}, k_{uf}); only the latter are, or need be, directly determined from the time-domain data.

Because we adopt a global Bayesian analysis, in which time-domain NMR data at all temperatures and concentrations are combined into a single set, it is necessary to provide a relationship giving the form of the temperature dependence of the rate constants for the conformational change (Eq. 11). We use the usual form:

$$\begin{aligned}k_{\mathrm{unf,fld}} &= Z\exp(-\Delta G^{\ddagger}_{\mathrm{unf,fld}}/RT) \\ &= Z\exp(\Delta S^{\ddagger}_{\mathrm{unf,fld}}/R)\exp(-\Delta H^{\ddagger}_{\mathrm{unf,fld}}/RT)\end{aligned} \tag{15}$$

wherein $\Delta X^{\ddagger}_{\mathrm{unf,fld}}$ represents the change in any standard thermodynamic quantity X (G, S, or H) in going to the transition state in the unfolding or folding direction, respectively. The preexponential factor Z is generally accepted to be $k\mathrm{T}/h$ (k is Boltzmann's and h Planck's constant) for chemical reactions involving covalent bond breakage or formation. Its value is near 10^{12} s^{-1}. However, the value appropriate to reactions in which only conformational changes occur is controversial, values ranging from 10^7 to 10^{12} having precedents (Bilsel and Matthews, 2000). Here, we use 10^{10}, which is not only a compromise value, but also the approximate value of the diffusion-controlled rate constant for association of two random chains of the size of GCN4-lz in the temperature range of interest here. This issue is further discussed below, where it will be seen that the chosen value of Z does not affect any of our conclusions, and in any case, has no effect on the activation enthalpies.

Parameter estimation

The parameters in the time-domain, two-state exchange lineshape model were estimated globally using Bayesian probability theory (Bretthorst, 1990a,b,c, Bretthorst, 1990c, 1997). The marginal posterior probability for each parameter was computed using a Markov chain Monte Carlo (MCMC) simulation with the joint posterior probability as its target distribution (Gilks et al., 1996). The MCMC simulation was implemented using thermodynamic integration to ensure that it located the global maximum of the joint posterior probability.

The physics of the model is that of two-state exchange as discussed above. In this model there are two basic functions of time at each labeled site: the transverse magnetization from spins in the folded environment and the corresponding magnetization from spins in the unfolded environment. We assume that the conformations at each labeled molecular site are interconverting at the same rate, which means that these rates are functions of four quantities: $\Delta H^{\ddagger}_{\mathrm{unf,fld}}$ and $\Delta S^{\ddagger}_{\mathrm{unf,fld}}$, the standard enthalpies and entropies of activation for the reaction $F \rightleftarrows 2U$. The purpose of the Bayesian analysis is to determine the values of these quantities that best mimic the data (fids). At first, the assumption of like rates at the three sites was considered to be provisional only. The subsequent finding (see below) that the analysis, on that basis, mimics the data well, justifies invocation of Occam's razor and thereby acceptance of the two-state idea, at least for the present.

Table 1 summarizes the 148 parameters involved. In order to determine these parameters from Bayesian theory, one must relate them to the available data, in this case the 11 free induction decays (fid's) obtained from 11 NMR runs. We next sketch how this was done and the origin and nature of the 148 parameters. The two unknown enthalpies and two unknown entropies, along with the estimated temperature, are needed to compute the NMR exchange rates for a given fid. We have a measured value of the temperature in each case, but we do not know it exactly. Consequently, each temperature was allowed to vary, but only within the uncertainty in the measured value (~1°C). There are 11 fids, so there are 11 unknown temperatures. From the computed NMR exchange rates, the weight fraction of each conformation was computed. These fractions and the total magnetization at time zero were then used to compute the total magnetization at each site for each fid. There are 11 fids and 3 sites, making 33 total magnetization parameters. Moreover, each resonance for each conformation for each fid has an unknown Larmor frequency associated with it, yielding (3)(2)(11) = 66 frequencies. Also, each resonance for each conformation has an inherent transverse relaxation rate constant (r_2) associated with it. Unlike frequencies, relaxation rate constants are not very precisely determined by NMR, so small errors in relaxation times do

TABLE 1 Summary of model parameters

Parameter Description	Number of Parameters
$\Delta H^{\ddagger}_{\mathrm{unf}}$, $\Delta H^{\ddagger}_{\mathrm{fld}}$, $\Delta S^{\ddagger}_{\mathrm{unf}}$, $\Delta S^{\ddagger}_{\mathrm{fld}}$	4
1 Temperature per fid	11
1 Total longitudinal magnetization per site per fid	33
2 Frequencies per site per fid	66
4 Longitudinal relaxation parameters per site	12
2 Phase parameters per fid	22
Total parameters in the model	148

fid, Free induction decay.

not make much difference in how well the model fits the data. Consequently, we imposed a linear temperature dependence on the relaxation times. With 2 parameters to determine a line and 6 relaxation rate constants, 12 parameters are needed to describe the relaxation times of the resonances. As seen in Table 1, the grand total is 148 parameters.

The MCMC simulation generates a parameter set and then uses Bayesian probability theory to compute the joint posterior probability for all 148 parameters. Two critical quantities are needed within the Bayesian calculations: the prior probabilities for the parameters (discussed below) and the direct probability for each data set. The latter is essentially a likelihood function. A Gaussian prior probability for the noise was used in the calculation, so the likelihood function is closely related to the mean square residual, i.e., measures the difference between the data and the model. For a given fid, the model of the data is the sum of six magnetizations (three sites, two conformations), each computed as the solution to the two-state exchange differential equations (Eq. 6 above).

Ideally, we would use uniform prior probabilities for many of the parameters. However, the functional form of this model precludes that strategy. If the enthalpies and entropies are not carefully constrained to the physically meaningful region of the parameter space, the computed exchange rates will be zero or infinity. Consequently, one is forced to specify this meaningful region. Each of these parameters was therefore assigned a bounded Gaussian prior probability. In the absence of specific prior information, we initially set the bounds, means, and standard deviations of the priors to our best guess of the physically meaningful region. For example, the relaxation times were constrained between 0.0001 and 1 s, with a mean of 0.03 and a standard deviation of 0.03 s. These bounds are so wide that they serve as little more than a rough guide to the MCMC simulation as to where to look for the parameter. Similarly, the frequencies were specified with a mean equal to the location of the resonance in the discrete Fourier transform with a standard deviation of 10 Hz. Given that the simulation determined these parameters to a few tenths of a hertz, these means and bounds on the prior values simply confine the parameters to frequencies where we know the resonances to be.

After setting the priors, we ran the simulations and examined the results. If any of the marginal probability density for the parameters butted up against the prior bounds, the prior bounds and means were shifted appropriately and the simulation re-run. When all the prior bounds covered the estimated parameters and all the bounds were wide compared to the estimated parameters, we set all the means to be approximately as estimated by the simulation and set the standard deviations on the priors to be very wide compared with how accurately the parameters were determined by the simulation. As a result, the simulation is essentially computing the marginal posterior probability for the parameters using uniform priors. The priors serve only to keep the MCMC from wandering into non-physical regions of the parameter space.

In implementing an MCMC simulation, it is usually better to run many short simulations than one long one. Consequently, our program runs 150 simulations concurrently. For a fixed annealing parameter, the simulations were run through 10 steps for each parameter. We then removed the 5 lowest probability simulations and replaced them by copies of higher probability simulations. We then increased the annealing parameter and repeated the process until the annealing parameter reached unity. At that time the simulations were saved, evolved for a number of additional steps, more samples saved, etc. After gathering a total of 200×150 simulations, we then used the distribution of samples to compute the marginal posterior probability of each parameter.

Because 148 parameters seems to be a dauntingly high number, it may help to put it in perspective. If one were to model each resonance as an exponentially decaying sinusoid, three parameters per resonance plus a zero and first-order phase correction would be required for each NMR run, i.e., each observed fid. With three labeled sites and two conformations each, this totals $(3)(6) + 2 = 20$ parameters per fid. With 11 fids the grand total is 220 parameters. Of course, if one were to model the fids run by run, only 20 parameters would need to be estimated at a time. However, the process would have to be repeated 11 times. Using our global procedure requires that 148 parameters be estimated at one time. Therefore, it is clear that 148 parameters is a marked reduction from 220, representing a substantial input of prior information into the model. The software was tested on simulated data. Chosen sets of parameters were used to generate free induction decays, with noise added to mimic that found in the experiments. The actual experiments were analyzed only after verifying that the programs converged to the known parameter values in the simulations. Other, subsequent tests are described in the Discussion section below.

RESULTS

Sample CD spectra are shown in Fig. 1 as a function of temperature over the range 5.4–73.2°C. The decline in helix content with increasing temperature is evident. There is strong evidence of an isodichroic point at 203 nm, where the mean residue ellipticity is -127 deg cm^2 mmol^{-1}. These results are in the range expected for a coiled coil (Holtzer and Holtzer, 1992). However, this only shows that the peptide groups in the system are in one of two possible states, not that there are only two molecular conformational states. Indeed, similar isodichroic points are observed in coiled coils where there is no doubt that the number of molecular conformations is larger than two.

Fig. 2 shows the negative of the mean residue ellipticity at 222 nm as function of temperature for various concentrations in the range 3.4–1410 μM. Here, the typical behavior of thermally unfolding coiled coils is seen: a shallow fall at low T, followed by much more cooperative decline at higher T. The effect of mass action is clearly evident as an

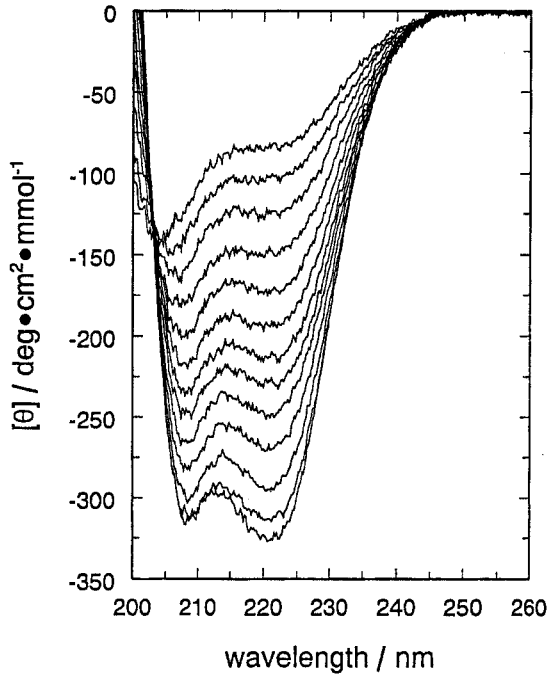

FIGURE 1 CD spectra of 317 μM GCN4-lz in $(NaCl)_{100}(NaPi)_{50}(D_2O)_{5514}$ (7.4) at various temperatures. Reading upward at 222 nm, temperatures are (in °C): 5.4, 21.3, 37.2, 47.7, 52.8, 55.3, 57.9, 60.5, 63.0, 65.6, 68.1, 70.7, 73.2. The isodichroic point is at 203 nm, -127.3 deg cm^2 mmol^{-1}.

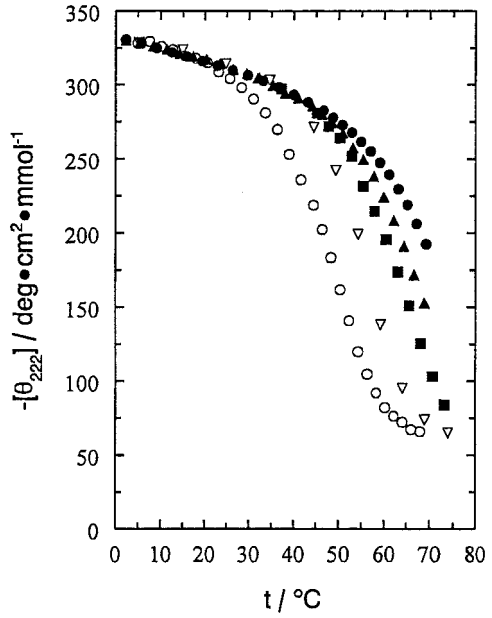

FIGURE 2 Negative of the mean residue ellipticity at 222 nm of GCN4-lz versus temperature at various peptide concentrations. Concentrations are (in μM): 3.41, *open circles*; 39.0, *open dels*; 317, *filled squares*; 528, *filled triangles*; 1410, *filled circles*. Solvent is as in Fig. 1.

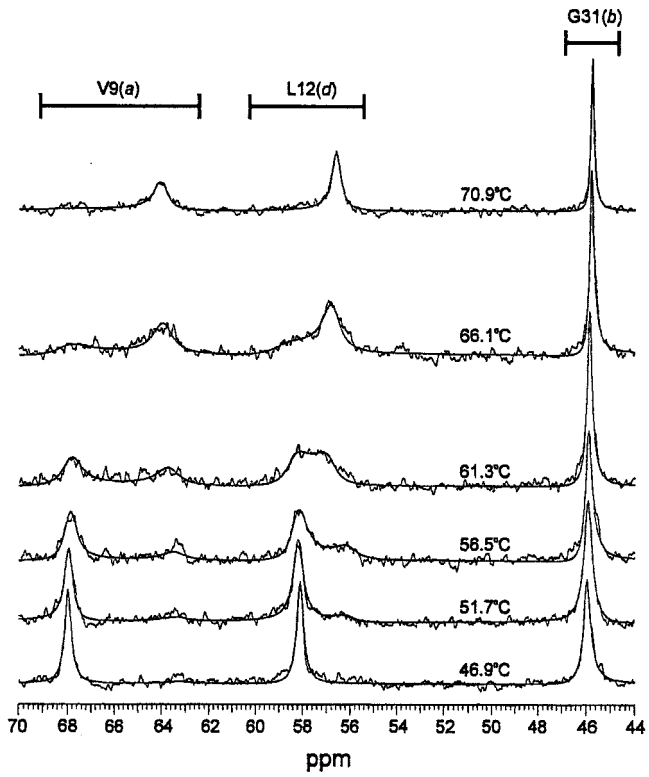

FIGURE 3 Frequency-domain $^{13}C^{\alpha}$-NMR spectra of GCN4-lz at sites V9(*a*), L12(*d*), and G31(*b*) at various temperatures (as marked). Range of chemical shifts for various conformations at each site is also marked. Peptide concentrations (μM) are: 324, bottom two spectra; 340, remaining spectra. Non-noisy lines are results of global Bayesian analysis of all data, including spectra not shown. Chemical shifts are referenced to external 2,2-dimethyl-2-silopentane-5-sulfonic acid (DSS) at 27.8°C. Solvent is as in Fig. 1.

increase in the melting temperature as concentration increases. In general, the melting temperatures of this GCN4-lz peptide are significantly greater than for the GCN4-lzK counterpart studied here earlier.

Fig. 3 displays some of the frequency-domain experimental $^{13}C^{\alpha}$-NMR spectra (noisy curves) at near 335 μM and various temperatures. As is clear from the spectra at the lowest and highest temperatures, the chemical shift difference between folded and unfolded forms is greatest at the V9(*a*) resonances, ~63–68 ppm, and least at the G31(*b*) resonances, near 46 ppm. The NMR time scale for conformational change is set by the frequency difference between the two forms. If the reaction rate constants are large compared to that frequency, only one broadened resonance will be observed. If those constants are small compared to that frequency, two independent resonances will be observed. This means that a given set of interconformational rate constants will play out differently at our three different labeled sites. A conformational reaction with site-independent rate constants will seem slower to a ^{13}C nucleus at V9 than at L12 or at G31. This effect is seen directly in the spectra. At the valine site, the reaction is very clearly evident, for example at 61°C, as two resonances broadened so as to strongly overlap. The leucine site shows two resonances so strongly overlapping as to almost merge, while at the glycine site they have merged. The strong influence of conformational interconversion on the line shape is evident.

For quantitation purposes, the signal-to-noise (S/N) is not as great as one would like. Nor is it possible to improve the S/N by increasing the concentration, since the dissociating nature of the transition simply then requires higher temperatures to reach the transition region, resulting in sample damage. Nevertheless, the spectra calculated from the global Bayesian analysis (smooth lines) fit the experimental spectra rather well. This is also seen in Fig. 4, wherein are displayed the residuals for spectra at 340 and at 1410 μM. Aside from a small anomaly in the glycine region for the high-concentration spectrum, the fit is satisfactory. This anomaly may be connected to the position of glycine near the C-terminus of the peptide. In the related peptide, GCN4-lzK, the G31(*b*) site is rather atypical (Holtzer et al., 1997). Other glitches in the residuals are readily ascribable to noise in the data. The kinetic constants returned by the Bayesian analysis are compiled in Table 2 and appear as Arrhenius plots in Fig. 5. The corresponding thermodynamic values are given in Table 3 and as a van't Hoff plot in Fig. 6. The linearity of these plots justifies the assumption that the heat capacity changes for these processes are small enough to ignore. As Figs. 5 and 6 show, the data (fids) at more than one concentration are well fit, as they should be, by concentration-independent values for the rate and equilibrium constants.

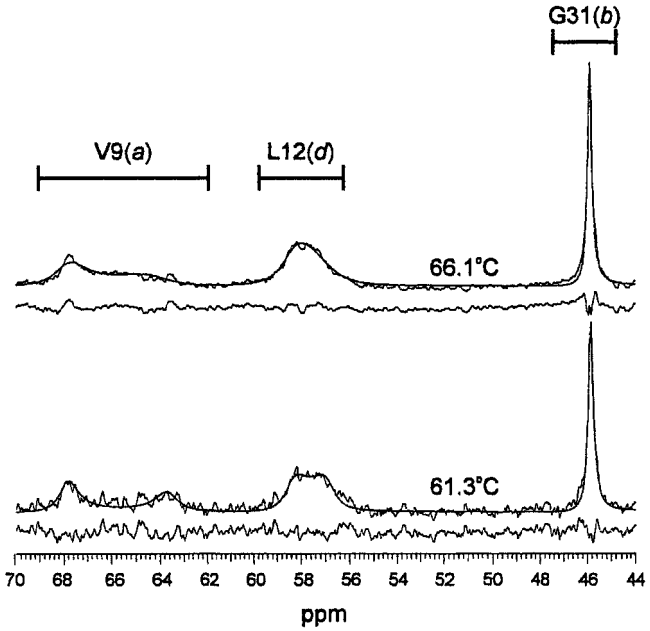

FIGURE 4 Sample frequency-domain $^{13}C^{\alpha}$-NMR spectra of GCN4-lz at sites V9(*a*), L12(*d*), and G31(*b*) with residuals from global Bayesian analysis. *Top*: 1410 μM peptide at 66.1°C. *Bottom*: 340 μM peptide at 61.3°C. Non-noisy lines are from the global Bayesian analysis, and residuals appear below each spectrum. Chemical shift referencing as in Fig. 3. Solvent is as in Fig. 1.

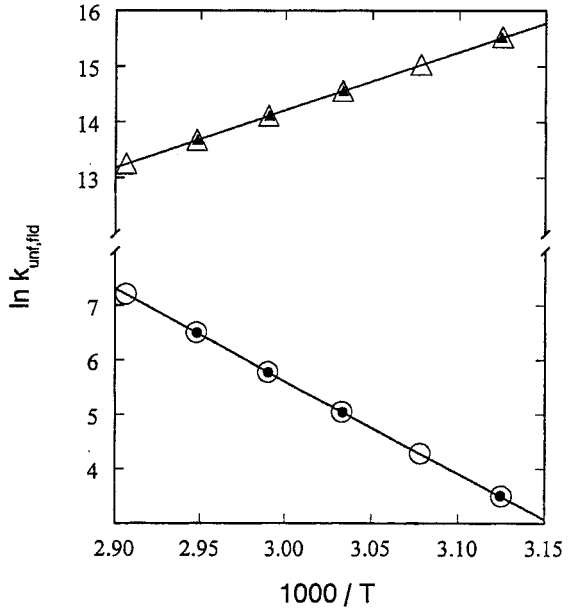

FIGURE 5 Arrhenius plots of kinetic constants for conformational change: F ⇄ 2U of GCN4-lz. *Lower line and circles*, unfolding (forward) direction; *upper line and triangles*, folding (reverse) direction. *Unfilled symbols*, 320–340 μM; *filled symbols*, 1410 μM. *Solid lines* are results of the global Bayesian analysis. Solvent is as in Fig. 1. Note break in ordinate axis.

DISCUSSION

Population of conformational states at equilibrium

The Bayesian analysis of our NMR data assumes that the same unfolding and folding rate constants apply at all three labeled sites. Moreover, these three sites are characterized by chemical shift differences for the folded and unfolded forms that vary significantly from site to site, thus providing three different NMR time scales. The satisfactory agreement between the calculated and measured spectra therefore confirms earlier workers who conclude that the two-state model applies well to GCN4-lz (Zitzewitz et al., 1995, 2000; Sosnick et al., 1996). This result is in marked contrast to the GCN4-lzK peptide, which clearly shows evidence of more than one folded form at the various sites investigated by $^{13}C^{\alpha}$-NMR. This evidence includes the appearance of more than one resonance in the region of the spectrum ascribed to folded forms, and demonstration by magnetization transfer of interconversion among folded forms (Lovett et al., 1996; Holtzer et al., 1997; d'Avignon et al., 1998; d'Avignon et al. 1999).

Nevertheless, the question cannot be considered entirely closed in the case of GCN4-lz. Our studies so far include only the three sites: V9(*a*), L12(*d*), and G31(*b*). Moreover, the first two are in the same heptad. In GCN4-lzK these two sites differ in transition midpoint temperature by 3.5°C, a substantial difference despite their close location (Holtzer et

TABLE 2 Kinetic parameters for GCN4-lz unfolding and folding

	k*	$\Delta H^{\ddagger\dagger}$	$\Delta S^{\ddagger\S}$	$\Delta G^{\ddagger\dagger}$
unf	401 ± 13	33.8 ± 1.9	66.8 ± 5.7	11.4 ± 0.03
fld	(1.20 ± 0.08) × 10^6	−20.6 ± 2.3	−79.3 ± 6.8	6.03 ± 0.04

For $F \rightleftharpoons A^{\ddagger} \rightleftharpoons 2U$ in $(NaCl)_{100}(NaPi)_{50}(7.4)$ using $Z = 10^{10}$. Error ranges are standard deviations from Bayesian analysis.
*s^{-1} for unfolding; $M^{-1} \cdot s^{-1}$ for folding; values apply at 62.7°C, the transition midpoint at 335 μM.
†kcal · su^{-1}; values for $\Delta G^{\ddagger}$ apply at 62.7°C, the transition midpoint at 335 μM.
§cal · K^{-1} · su^{-1}.

TABLE 3 Standard thermodynamic parameters for GCN4-lz

ΔH^{∞}_{unf}*	ΔS^{∞}_{unf} †	ΔG^{∞}_{unf}*	10^6 K_{unf} ‡
54.4 ± 3.0	146 ± 9	5.34 ± 0.05	335 ± 26

For F ⇌ 2U in $(NaCl)_{100}(NaPi)_{50}(7.4)$. Error ranges are standard deviations from Bayesian analysis.
*kcal · su^{-1}; standard state is 1 M, as extrapolated from highly dilute reference state. The value for ΔG^{∞}_{unf} applies at 62.7°C, the transition midpoint at 335 μM.
†cal · K^{-1} · su^{-1}; standard state is 1 M, as extrapolated from highly dilute reference state.
‡Strictly speaking, equilibrium constants have no units; here, the nominal units of K_{unf} are Mol · $liter^{-1}$; the value given is for 62.7°C, the transition midpoint at 335 μM.

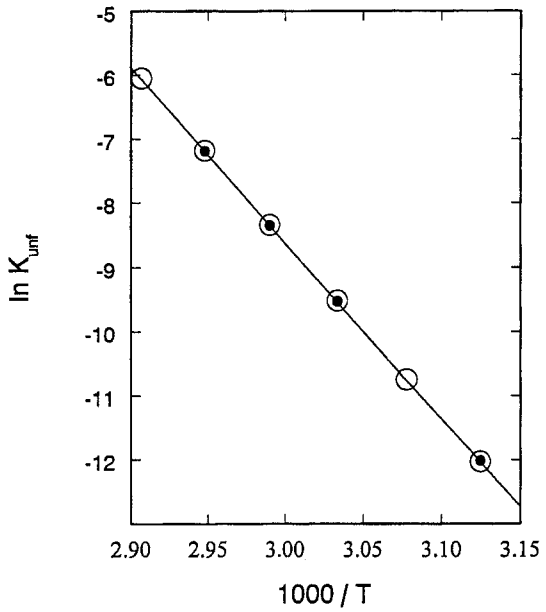

FIGURE 6 van't Hoff plot of equilibrium constants for unfolding reaction F ⇄ 2U of GCN4-lz. *Unfilled circles*, 320–340 μM; *filled circles*, 1410 μM. Solid line is from global Bayesian analysis. Solvent is as in Fig. 1.

al., 1997). In GCN4-lz, the corresponding difference, if any, must be much less. However, it is possible that alternative folded forms could be observed at other sites, even in GCN4-lz. Indeed, there are hints of a shoulder on each of the resonances at lower temperatures and higher concentrations (not shown) in our current labeled GCN4-lz peptide, possibly indicating the presence of such other forms. It may be, therefore, that at the higher temperatures required to unfold GCN4-lz, other folded forms have given way to the one most stable at elevated temperatures. Investigation of other labeled sites in GCN4-lz could answer the question more definitely. Meanwhile, there is no reason to invoke more than two states in examining our present kinetic data.

The GCN4-lz peptide also differs from the GCN4-lzK variant in the time scale for conformational change. GCN4-lzK shows distinct resonances for the folded and unfolded forms, indicating interconversion processes that are slow on the NMR time scale (Lovett et al., 1996). GCN4-lz, however, shows line shapes characteristic of two interconverting sites, evidence of kinetics of interconversion on a time scale comparable to that of the NMR experiment. These pronounced differences between peptides differing by only four conservative changes underscores the difficulty of making predictions about thermodynamics and kinetics for even very minimally mutated peptides.

Thermodynamic parameters for $F \rightleftarrows 2U$

Table 3 indicates an enthalpy of overall unfolding of 54.4 kcal/su. This compares well to the calorimetric value of 45–60 kcal/su (Kenar et al., 1995). However, we find that ΔC_P^{∞} is negligible, whereas the calorimetric value is given as 500 cal K^{-1} su^{-1}. If the standard heat capacity difference were as large as that, our enthalpy value should have changed by ~12 kcal/su (i.e., ~20%) over the temperature range of 25°C spanned by our data. If that were the case, we should have seen it. We have no explanation for this discrepancy between our NMR results and the calorimetric value of the heat capacity change. Our efforts to include such a difference did not improve the agreement between the Bayesian results and our NMR data.

It is of interest to compare our enthalpy value for GCN4-lz with that for tropomyosin, perhaps the prototypical coiled coil that is not a leucine zipper. Since tropomyosin's chain is much longer, the relevant comparison is gram-for-gram. For tropomyosin the value is ~5.2 cal/g (Potekhin and Privalov, 1982), while our value for GCN4-lz is 6.8 cal/g, significantly higher. This indicates that a great deal of the greater stability of these short leucine zipper peptides is energetic in origin.

Kinetics of unfolding

The nature of the transition state

The model used here for analysis of the kinetics is:

$$F \rightleftarrows A^{\ddagger} \rightleftarrows 2U \tag{16}$$

wherein $A^{\ddagger}$ represents the dimeric activated complex (transition state). The values in Table 2 leave little doubt that the process is an activated one. The standard Gibbs energy of activation is substantial, 11.4 kcal/su in the unfolding direction at the transition midpoint (62.7°C at 335 μM). It is also evident that the unfolding reaction, kinetically speaking, is entropy-driven and enthalpy-opposed. The transition state is apparently appreciably unfolded with respect to the folded form. This is, of course, not unexpected, but the values in Tables 2 and 3 provide quantitation. All of these energy relations can be grasped more easily by referring to their graphical representation in Fig. 7.

One of the tenets of transition-state theory is that the concentration of the activated state is low enough to ignore in considering the equilibrium populations. One can verify that this is indeed the case here. For a reaction of the stoichiometry appropriate here, the midpoint occurs where the concentration of the folded form is related to the total formality of peptide (as monomers) by: $[F] = C_0/4 = 83.8$ μM at $C_0 = 335$ μM. The transition state concentration then becomes: $[A^{\ddagger}] = [F] \exp(-\Delta G^{\ddagger}_{unf}/RT_m) = 3.2 \times 10^{-6}$ μM, far lower than for the folded or unfolded form.

Because the value of the activation free energy depends on one's choice of pre-exponential factor, Z, one might inquire as to whether a change in our choice of 10^{10} s^{-1} for this factor affects our conclusions. Within the reasonable limits for Z ($10^7 - 10^{12}$), the value 10^7 provides the smallest

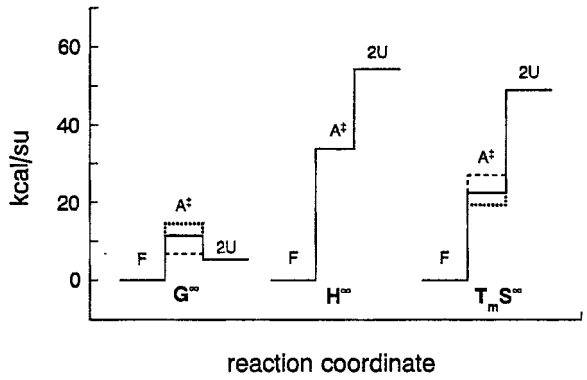

FIGURE 7 Energy level diagram for GCN4-lz from Tables 2 and 3. Standard Gibbs energy, enthalpy, and entropy values for the folded form F are arbitrarily set to zero. Solid lines are for (in s^{-1}): $Z = 10^{10}$; dashed lines for $Z = 10^{7}$; dotted lines for $Z = 10^{12}$.

value for the activation free energy (6.8 kcal/su) and therefore the largest for the transition state concentration, the latter becoming 0.0032 μM, still orders of magnitude lower than that of either of the two principal thermodynamic components. We therefore conclude that the values in Table 2 satisfy this requirement of the theory. Indeed, none of the conclusions drawn here are affected by altering Z within reasonable limits.

The activation enthalpy for unfolding is also high (33.8 kcal/su, 62.1% of the total enthalpy of unfolding) and only partly compensated by the activation entropy. This enthalpy value can be used to learn something about the nature of the transition state, i.e., its "degree of unfoldedness." To do so, we assume that the enthalpy-structure relationship is linear and estimate the enthalpy of the transition state as a weighted sum of those for the unfolded and folded local states:

$$H^{\ddagger} = f_{\mathrm{U}}^{\ddagger}(2H_{\mathrm{U}}^{\circ}) + (1 - f_{\mathrm{U}}^{\ddagger})H_{\mathrm{F}}^{\circ} \tag{16}$$

wherein $f_{\mathrm{U}}{}^{\ddagger}$ is the "fraction unfoldedness" of the transition state. Equation 16 gives:

$$f_{\mathrm{U}}^{\ddagger} = \Delta H_{\mathrm{unf}}^{\ddagger}/\Delta H_{\mathrm{unf}}^{\circ} \tag{17}$$

Implementation of the last equation indicates that the fraction unfoldedness is 62.1%, i.e., the transition state, judging from the enthalpies, is 62.1%, like the unfolded form, and 37.9%, like the folded form.

We can next use this result to estimate the helix content of the transition state. We first need estimates of the helix contents of the folded and unfolded forms at the transition midpoint. These can be obtained from the CD data of Fig. 2. The observed mean residue ellipticity at 222 nm levels off at high temperatures at a value near -60 deg cm^2 $mmol^{-1}$, and CD values for the unfolded forms are but little dependent on T. Using a value of -350 (same units) for the CD of a full helix and zero for the fully random chain, we find a helix content for the unfolded form of $f_{\mathrm{h}}^{\mathrm{U}} = 0.171$, i.e., near 17%, providing one to two helical turns for a 33-residue chain. This is consistent with our data for many types of coiled coils.

To obtain an estimate of the helix content of the folded form we also use the data of Fig. 2. However, the low temperature values indicate a slow linear decline in the negative of the CD with T in the low-T region before the cooperative decline that occurs in the region of higher T. This slow decline at low T almost surely indicates some thermally induced loss of helix content of the dimeric, predominantly folded species (Holtzer et al., 1997). Extrapolating this initial linear decline to the midpoint temperature for solutions near 300 μM gives a value of 284 deg cm^2 $mmol^{-1}$ for the folded form, corresponding to a helix content of 0.811.

Using, then, these measures, we find for the mean helix content of the transition state ensemble: $(0.621)(0.171) + (0.379)(0.811) = 0.414$, i.e., 41.4% helix. This estimate characterizes the structure of the transition state just as the values of Table 2 characterize its various energetic properties. It should be kept in mind that this estimate does not depend on our choice of the preexponential factor Z, since the activation enthalpies are independent of it. The properties of the transition state dictated by our experiments therefore disagree with those found in some extant work that ascribes only marginal helix content to the transition state. We believe our methods to be more direct than the Φ-analysis used in that study (Sosnick et al., 1996).

Comparison with stopped-flow studies at low temperature

Equation 15 and our values of the activation parameters for unfolding (Table 2) allow estimation of the rate constants for temperatures outside the range of our experiments. They therefore make it possible to compare such extrapolated values with extant values obtained by stopped-flow at temperatures sufficiently low so that folding is strongly favored. Such comparison is made in Table 4 with values obtained under solvent conditions similar to ours, but at 5 and 15°C (Zitzewitz et al., 1995, 2000). The agreement of these stopped-flow values of k_{unf} with our NMR-obtained ones is

TABLE 4 Low temperature rate constants from stopped flow and NMR

		$10^{3}\ k_{\mathrm{unf}}$*	$10^{-5}\ k_{\mathrm{fld}}$*
5°C	SF†	3.3	4.2
	NMR‡	10.0	7100
15°C	SF†	18.3	22.0
	NMR‡	88.0	1930

*k_{unf} in s^{-1}; k_{fld} in M^{-1} s^{-1}.
†By extrapolation from GdmCl solution (5°C, Zitzewitz et al., 1995) or urea solution (15°C Zitzewitz et al., 2000).
‡By extrapolation from data at high temperature (47–71°C).

remarkably good, considering that two long extrapolations are involved. Our values are obtained from a data set whose lowest temperature is near 47°C, and must be extrapolated to a region ~35° lower. Moreover, the stopped-flow values for unfolding are obtained at high concentrations of denaturant and must be extrapolated to zero from a lowest value of ~3 M. It follows that these two markedly different techniques, deployed in different laboratories, must be assessing the same unfolding kinetic process.

Kinetics of folding

The mechanism of folding

Table 2 makes clear that folding is also an activated process, as earlier workers have recognized (Zitzewitz et al., 1995). This is directly apparent from the table, which details the temperature dependence through the value for the folding Gibbs energy of activation. However, this was evident earlier from a simple calculation using Smoluchowski's formula for the rate constant of a diffusion-controlled reaction. Tropomyosin was the first coiled coil for which this was done (Mo et al., 1991). Modifying that calculation for the lower molecular mass of the GCN4-lz peptide and accounting for the difference in temperature and aqueous viscosity, we find $\sim 10^{10}$ M^{-1} s^{-1} for the magnitude of the folding rate constant, if it were diffusion-controlled. The observed value is four orders of magnitude lower, indicating that only a fraction of the encounters leads to formation of the intact coiled coil.

As Table 2 also makes clear, the folding is kinetically enthalpy-driven and entropy-opposed. That, however, could have been predicted on physical grounds alone, given the nature of the difference between folded and unfolded coiled coils. These measurements lead to a more detailed model for the process. As noted in the previous section, the enthalpies indicate that the transition state ensemble has appreciable helix content. We believe, therefore, that the folding process must be quite different from that described in some extant work. When two unfolded chains of GCN4-lz meet, they each have (see above) about one or two helical turns. Our experiments suggest that in order for the encounter to be successful, they must form considerably more helix. We therefore suggest that a successful encounter must be characterized by considerable coiled coil formation preliminary to going on to the fully folded native molecule. That this could occur by formation of a registered, partial native structure seems to us unlikely. More likely, the presence of some helix in the first chain induces associative helix formation in the second one, wherever they touch. After all, the heptad pattern of amino acids that characterizes coiled coils exists all along the chain. The result could be formation of a broad ensemble of out-of-register structures (averaging near 40% helix) that then can inch-worm or scissors their way to the registered product coiled coil.

Two extant proposals on this same process are different from one another as well as from our own. One study using ϕ-analysis suggests that the transition state has little helix content (Sosnick et al., 1996). We find that conclusion to be incompatible with our data because of the enthalpies of activation involved. These enthalpies, again, are independent of the arbitrariness in our choice of the preexponential factor Z. We do not see how that finding can be reconciled with an essentially helix-free transition state. In any case, later work from related hands appears to amend that view (Moran et al., 1999). However, one must keep in mind that the ϕ-analysis used data taken in acidic media (pH 4.5–5.5), a very different environment than used here and in the stopped-flow studies alluded to above (Zitzewitz et al., 1995, 2000). The difference that this environmental change produces can be assayed by future experiments.

Another extant view is quite different (Myers and Oas, 1999; Zitzewitz et al., 2000). It avers that the transition state involves a small, locally persistent helix on each chain that has to search for its counterpart on the other chain in order for successful folding to occur. The idea is attractive, and one group even makes an estimate of the rate constant, finding agreement with experiment (Myers and Oas, 1999). Nevertheless, we do not believe their calculation can be squared with our results. There are two grounds for our opinion. First, it is hard to see how their model can yield much temperature dependence for the folding rate constants. The rate limitation effect they propose is exclusively entropic, contrary to what our data indicate. Their model was proposed before the effect of temperature in this system had been investigated. Second, the physical basis of the calculation is questionable. The probability of a collision being reactive is treated as the square of the product of the probability that a chain near the contact point will be helical and the probability that it will present an appropriate contact surface. Each of these probabilities is estimated as crudely obtained for a single chain. Thus, the assumption is made that the two chains in the encounter are independent of one another. However, the transition ensemble is not only unlikely to be indifferent to chain-chain forces, but may be, in part, a creature of their making.

Comparison with stopped-flow studies at low temperature

As seen above, our NMR-derived values for the activation parameters lead to estimates for the unfolding rate constant at low temperatures that agree with those from stopped flow despite the vast differences in the methods and conditions. The corresponding comparison for the folding rate constants given in Table 4 tells a vastly different tale. In the case of folding, the NMR-derived values are two to three orders of magnitude larger than those from stopped flow. In this case, the temperature extrapolation needed is the same as for unfolding, but the extrapolation to zero denaturant is much shorter (from 0.5 M), making those values even less likely

to be in error than those for unfolding. How can such a discrepancy be understood?

In order for the NMR data to yield low-temperature folding rate constants as small as those obtained from stopped flow, our value for the activation enthalpy would have to be almost zero. To test this possibility we constructed two data sets, both very much like the real set; however, one set was made with zero folding activation enthalpy and the other with the same value as determined from the real data set. These two simulated sets were then analyzed using the same Bayesian procedure as used on the real data set. The results left little doubt that the Bayesian analysis easily distinguishes the two, emerging with the correct value in each case.

We next considered the possibility that a change in heat capacity upon activation might alter our temperature extrapolation sufficiently to explain the discrepancy. Such a heat capacity change might not be apparent in the range of T encountered in our experiments, but could be manifest over the long extrapolation. However, calculation soon reveals that the change in activation heat capacity that would be required is appreciably larger than has been estimated for complete folding (Kenar et al., 1995), an extremely unlikely result.

We are therefore driven to accept both the NMR and stopped-flow values as basically correct descriptions of the processes measured, in despite the apparent disagreement. This view is supported by the agreement of the two in the case of the unfolding rate constants. It means that the two values of k_{fld} differ because the processes measured differ. This difference, and how it might lead to very different rate constants, is examined next.

In our NMR determinations, the conformational reaction is never out of equilibrium in the actual solvent milieu of interest. Nevertheless, the equilibrium is dynamic, and NMR is sensitive to the rates. In a stopped-flow experiment, however, the system begins very far from equilibrium in a highly denaturing solvent. Rapid change of the solvent then allows folding to occur. We suggest that such differences lead to the discrepancy found.

The key to the question perhaps lies in a recent finding concerning the rate constants for helix formation (Clarke et al., 1999). It has been a part of the conventional wisdom for almost 35 years that helix formation is accomplished in microseconds. The experimental basis for this view residues in T-jump experiments on synthetic single-chain helix-forming peptides (Schwarz, 1965). A T-jump experiment is somewhat similar to our NMR technique in that the system is at equilibrium to start with. It then is slightly perturbed by the T-jump and the relaxation rate to a new equilibrium is monitored. Recently, however, the rate of helix formation has become accessible by stopped-flow CD, because use of synchrotron radiation yields a 100-fold increase in signal/noise. The startling finding is that, in the stopped flow experiment, helix formation takes milliseconds, not microseconds, a difference of the same order of magnitude as we observe. The investigators making this discovery ascribe the discrepancy to the necessity of nucleating helices in the stopped-flow experiment, whereas in T-jump only helix propagation, a much faster process, is measured (Clarke et al., 1999).

We therefore suggest that in a stopped-flow determination of the folding rate for coiled coils, the highly unfolded initial state is affecting the result. The necessity of initiating helix formation could make each collision much less effective in producing the coiled-coil product. Results for the single-stranded helix-coil transition imply that a rate reduction of three orders of magnitude is not out of line for this effect. In our experiments, the unfolded chains at all times have the conformation dictated by the benign aqueous neutral saline buffer, in which CD results suggest strongly that there is a definite equilibrium presence of helical content. Helix initiation is unnecessary, leading to a different time scale for folding.

It follows that those single-chain globular proteins that are seen in stopped-flow to fold in microseconds must do so by mechanisms not requiring helix initiation. One must also contend with the results for folding the tropomyosin coiled coil (Mo et al., 1991). This protein forms a dimeric intermediate that is 2/3 native with a rate constant that is at least 10^7–10^8 M^{-1} s^{-1} when refolded from 6 M urea, a rate constant one to two orders larger than found for GCN4-lz, but also one to two orders smaller than the tropomyosin collision rate. This must be a result of retention of some helix in urea or the existence of more potential helix-nucleating sites in the very long, 284-residue tropomyosin chain. Because the tropomyosin stopped-flow work had a 40-ms dead time, the dimerization rate constant is only an estimate. Measurements on this system as a function of temperature with better instrumentation may resolve the dimerization step.

In an effort to examine our proposal critically, one might suppose that the extant stopped-flow measurements on GCN4-lz could be used themselves to estimate the activation enthalpy in each direction, since data for two temperatures are available. We next show that these data are inappropriate for the purpose. Using the stopped-flow determined rate constants from Table 4, we find 27.3 kcal/su and 26.3 kcal/su for the enthalpy of activation for unfolding and folding, respectively. These values make no sense. In the first place, they are both positive. Yet, transition from an unfolded state to the activated complex ensemble must surely yield zero or less for the activation enthalpy. Moreover, if one estimates the overall enthalpy of unfolding from these, one finds a mere 1 kcal/su, an absurdly small value. Clearly, something else besides the temperature dependence of the rate constants is contributing to the difference between the stopped-flow values at 5 and at 15°C. We suggest that this difference is chiefly a result of the change in stopped-flow protocol that was made in the two different

determinations. The experiments at 5°C used guanidinium chloride as denaturant (Zitzewitz et al., 1995), whereas the ones at 15°C used urea (Zitzewitz et al., 2000). The more thoroughly unfolded protein from the former may require more time to nucleate the necessary helix content than the latter, contributing to the apparent temperature dependence. Needless to say, appropriate stopped-flow experiments as functions of temperature and denaturant would test this idea.

Further critical scrutiny of our proposal might recognize that it depends on the assumption that the actual collision rate that applies in the stopped-flow experiments exceeds the helix nucleation rate. If the reverse were true, the nucleation rate could not reduce the fractional success of collisions. The following calculation shows that our proposal passes this test also. Use of the Smoluchowski equation as before, with alterations in the temperature and viscosity factors to adapt it to 5°C, yields a diffusion-controlled rate constant near 3×10^9 M^{-1} s^{-1}. The important thing, however, is the rate, not the rate constant. In these stopped-flow experiments, the smallest peptide concentration used was 6 μM, leading to an actual encounter *rate* $\geq(8 \times 10^9)(6 \times 10^{-6})^2 = 0.11$ M s^{-1}. How does this compare with the helix initiation rate? Clarke et al. (1999) give the latter rate constant as 16 s^{-1} near 5°C, giving a helix-forming rate of $(16)(6 \times 10^{-6}) = 1 \times 10^{-4}$ M s^{-1}. Therefore, the encounter rate in the stopped-flow experiments is at least 1000 times the helix-forming rate, indicating that many collisions must occur in the time it takes to initiate helices. The proposal therefore survives this test, at least.

CONCLUSIONS

1. Interconversion of folded dimers and unfolded monomers of the GCN4 leucine zipper in neutral saline buffer is on a time scale appropriate for study by one-dimensional $^{13}C^{\alpha}$-NMR lineshape analysis as a function of temperature;
2. Global Bayesian techniques allow extraction of the activation enthalpies and entropies for both unfolding and folding directions from the time-domain NMR observations;
3. The analysis of data for three labeled sites [V9(*a*), L12(*d*), and G31(*b*)] provides no evidence that different kinetic parameters apply at the different sites, confirming that a two-state model is adequate;
4. The activation parameters obtained suggest that unfolding is kinetically enthalpy-opposed and entropy-driven, while folding is the opposite, and that the transition state ensemble is ~38% folded and 41% helical;
5. When the activation parameters for thermal unfolding, obtained from NMR experiments in the range 47–71°C, are used to calculate the unfolding rate constant at lower temperatures (5–15°C), the results agree well with extant values obtained via stopped flow;
6. When the corresponding comparison is made for the folding rate constants, the values calculated from NMR are two to three orders of magnitude larger than obtained from stopped flow;
7. The apparent discrepancy in the case of the folding rate constants is ascribed to the slowness of helix initiation in chains originally resident in the highly denaturing media necessarily used in the stopped-flow technique, as suggested by recent experiments on the single-stranded helix-coil transition.

Development of the Bayesian techniques was supported by the National Institutes of Health Grant NS-35912 and by a license agreement with Varian Associates. Mass spectrometry was provided by the Washington University Mass Spectrometry Resource, a National Institutes of Health Research Resource (Grant P41RR0954). Peptide syntheses were performed at the Albert Einstein College of Medicine and were supported in part by a grant from the Mathers Foundation. One of us (A.H.) acknowledges the continuing support of the Luftmensch Society.

REFERENCES

Bilsel, O., and C. R. Matthews. 2000. Barriers in protein folding reactions. *Adv. Protein Sci.* 53:153–207.

Bretthorst, G. L. 1990a. Bayesian analysis. I. Parameter estimation using quadrature NMR models. *J. Magn. Reson.* 88:533–551.

Bretthorst, G. L. 1990b. Bayesian analysis. II. Model selection. *J. Magn. Reson.* 88:552–570.

Bretthorst, G. L. 1990c. Bayesian analysis. III. Applications to NMR signal detection, model selection, and parameter estimation. *J. Magn. Reson.* 88:571–595.

Bretthorst, G. L. 1997. Bayesian Analysis Software Package User Guide. Pub. No. 87-190172-00. Rev. A0197, Varian Associates, Palo Alto, CA.

Carrington, A., and A. D. McLachlan. 1967. Introduction to Magnetic Resonance, Chaps. 11 and 12. Harper and Row, New York.

Clarke, D. T., A. J. Doig, B. J. Stapely, and G. R. Jones. 1999. The α-helix folds on the millisecond time scale. *Proc. Natl. Acad. Sci. USA*. 96: 7232–7237.

Crick, F. 1953. The Fourier transform of a coiled coil. *Acta Crystallogr.* 6:689–697.

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

d'Avignon, D. A., G. L. Bretthorst, M. E. Holtzer, and A. Holtzer. 1999. Thermodynamics and kinetics of a folded-folded′ transition at valine 9 of a GCN4-like leucine zipper. *Biophys. J.* 76:2752–2759.

Edelhoch, H. 1967. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry.* 6:1948–1954.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter. 1996. Markov Chain Monte Carlo in Practice. Chapman & Hall, London.

Holtzer, M. E., D. L. Crimmins, and A. Holtzer. 1995. Structural stability of short subsequences of the tropomyosin chain. *Biopolymers.* 35: 125–136.

Holtzer, M. E., and A. Holtzer. 1992. α-Helix to random coil transitions: determination of peptide concentration from the CD at the isodichroic point. *Biopolymers.* 32:1675–1677.

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 equilibria. *Biophys. J.* 73:1031–1041.

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.

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.

Lupas, A. 1996. Coiled coils: new structures and new functions. *TIBS* 21:375–382.

McConnell, H. M. 1958. Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.* 28:430–431.

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

Mo, J., M. E. Holtzer, and A. Holtzer. 1991. Kinetics of self-assembly of $\alpha\alpha$-tropomyosin coiled coils from unfolded chains. *Proc. Natl. Acad. Sci. USA.* 88:916–920.

Moran, L. B., J. P. Schneider, A. Kentsis, G. A. Reddy, and T. R. Sosnick. 1999. Transition state heterogeneity in GCN4 coiled coil folding studied by using multisite mutations and crosslinking. *Proc. Natl. Acad. Sci. USA.* 96:10699–10704.

Myers, J. K., and T. G. Oas. 1999. Reinterpretation of GCN4–p1 folding kinetics: partial helix formation precedes dimerization in coiled coil folding. *J. Mol. Biol.* 289:205–209.

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

O'Shea, E. K., R. Rutkowski, and P. S. Kim. 1989. Evidence that the leucine zipper is a coiled coil. *Science.* 243:538–542.

Poland, D., and H. A. Scheraga. 1970. Theory of the Helix-Coil Transition in Biopolymers. Academic Press, New York.

Potekhin, S. A., and P. L. Privalov. 1982. Co-operative blocks in tropomyosin. *J. Mol. Biol.* 159:519–535.

Rudin, M., and A. Sauter. 1992. Measurement of reaction rates in vivo using magnetization transfer techniques. *In* NMR Basic Principles and Progress. P. Diehl, E. Fluck, H. Günther, R. Kosfeld, and J. Seelig, editors. Springer Verlag, Berlin.

Schwarz, G. 1965. On the kinetics of the helix-coil transition of polypeptides in solution. *J. Mol. Biol.* 11:64–77.

Sosnick, T. R., S. Jackson, R. R. Wilk, W. Englander, and W. F. DeGrado. 1996. The role of helix formation in the folding of a fully α-helical coiled coil. *Proteins: Struct., Funct., Genet.* 24:427–432.

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.

Zitzewitz, J. A., B. Ibarra-Molero, D. R. Fishel, K. L. Terry, and C. R. Matthews. 2000. Preformed secondary structure drives the association reaction of GCN4–p1, a model coiled-coil system. *J. Mol. Biol.* 296: 1105–1116.