Conformational Diffusion and Helix Formation Kinetics

Gerhard Hummer,[1,*] Angel E. García,[2] and Shekhar Garde[3]

[1]*Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520*

[2]*Theoretical Biology and Biophysics Group T-10, MS K710, Los Alamos National Laboratory, Los Alamos, New Mexico 87545*

[3]*Department of Chemical Engineering, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180*

(Received 21 April 2000)

The time, temperature, and sequence dependences of helix formation kinetics of fully atomistic peptide models in explicit solvent are described quantitatively by a diffusive search within the coil state with barrierless transitions into the helical state. Conformational diffusion leads to nonexponential kinetics and jump-width dependences in temperature jump experiments.

PACS numbers: 87.15.Cc, 87.10.+e

Fast, time-resolved spectroscopy probing the earliest, submicrosecond events in protein folding [1–4] showed that α-helix formation occurs on a nanosecond time scale [5,6], considerably faster than β-turn formation [7], loop closure [8], or the overall folding of proteins [1,2,9]. Helix formation kinetics is thus getting within reach of molecular simulations [10,11]. The energy landscapes of small peptides [12] are greatly simplified compared to those of proteins [13–19]. Nevertheless, even for peptides of only a few amino acids, the search for a "folded structure" in a high-dimensional conformation space is far from trivial. Small helical peptides can thus provide us with a revealing view of the molecular processes underlying protein folding at greatly reduced computational cost.

The "coil-to-helix" transition [20–22] is the simplest element of protein secondary structure formation and folding. We quantify helix formation kinetics within a conformational diffusion model, based on extensive molecular dynamics (MD) simulations of all-atom peptide models in explicit solvent. The helix formation kinetics is tied to thermodynamic driving forces: free energies, enthalpies, and entropies. This analysis reveals surprising similarities to the more complex folding of proteins.

Two blocked alanine (Ala) and glycine (Gly) based pentapeptides are studied. Both Ac-Ala_5-NHMe (A_5) and Ac-Ala_2-Gly-Ala_2-NHMe (A_2GA_2) can form 1.5 turns of α helix stabilized by three $(i, i + 4)$ backbone hydrogen bonds. These two peptides were each simulated at temperatures T ranging from 250 to 400 K in explicit solvent of ca. 500 water molecules for about 10 ns per run. Simulation details can be found in Ref. [23]. From a principal-component analysis [24] of configuration space, we identify a helical reference structure with backbone dihedral angles $(\varphi, \psi) = (-57.4, -45.3)$, $(-64.0, -40.9)$, $(-63.9, -41.9)$, $(-63.4, -37.8)$, and $(-70.9, -23.8)$ in units of degrees. Peptides with mean-square-distance (MSD) [25] of less than 0.36 Å^2 from that structure are classified as helical, corresponding to the free energy minimum along the MSD reaction coordinate (Fig. 1). Structures with MSD > 0.36 Å^2 belong to the coil ensemble. With these definitions, equilibrium MD trajectories are analyzed to extract thermodynamic and kinetic parameters of the interconversion between helical and coil states.

Folding is driven by thermodynamics, and we expect that the kinetics of the helix-to-coil transition is tied to thermodynamic functions. Figure 1 shows the free energy profiles for the A_5 and A_2GA_2 peptides along the MSD reaction coordinate. For A_5, we find a minimum corresponding to the helical state with MSD < 0.36 Å^2. Separated by a small barrier ($<k_BT$), a second minimum develops at higher temperatures corresponding to the coil

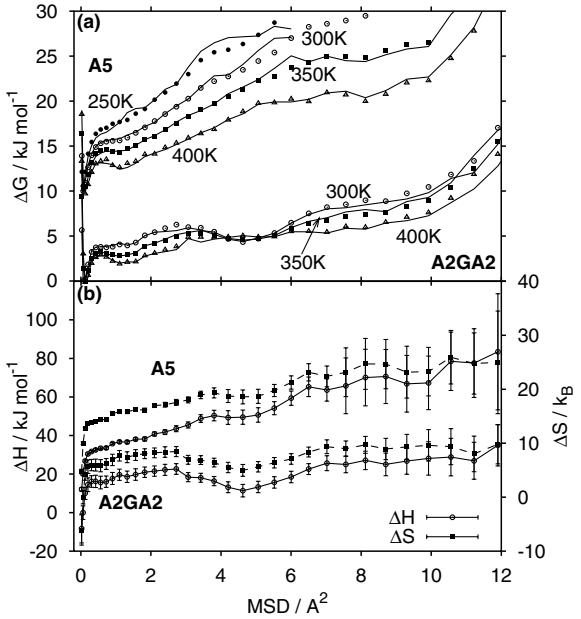

FIG. 1. (a) Free energy along the MSD reaction coordinate for A_5 (top curves) and A_2GA_2 (bottom curves). The lines are obtained from the MD trajectories. The symbols are a fit to a Gibbs free energy $\Delta G = \Delta H - T\Delta S$ with T-independent enthalpies and entropies. (b) Enthalpies ΔH and entropies ΔS along the MSD reaction coordinate for A_5 (top curves) and A_2GA_2 (bottom curves). Lines are guides to the eye. Note that ΔG, ΔH, and ΔS contain arbitrary constants. No correction was made for the "Jacobian" in MSD space, such that $\Delta G \to \infty$ as MSD $\to 0$ (i.e., observing the *exact* reference structure has zero probability).

state. For A_2GA_2, the free energy profile in the coil region is more structured, reflecting the presence of distinct structure classes within the coil ensemble. In particular, the minimum between MSDs of 3 and 6 Å^2 corresponds to turnlike structures [23] that are stereochemically feasible because of the flexibility added by the Gly residue at the center of the peptide. As temperature increases, the free energies of the coil ensemble become more favorable relative to the helical state. At any given MSD, we find that the Gibbs free energy $\Delta G(\mathrm{MSD})$ can be represented by a sum of T-independent enthalpies, ΔH, and entropies, ΔS: $\Delta G(\mathrm{MSD}) = \Delta H(\mathrm{MSD}) - T\Delta S(\mathrm{MSD})$. The agreement of this simple thermodynamic decomposition with the actual data is illustrated in Fig. 1. Note that $\Delta S = -\partial \Delta G/\partial T$ is the thermodynamic entropy incorporating both peptide and water contributions.

As shown in Fig. 1, at MSD > 0.36 Å^2 both ΔH and ΔS of A_5 decrease approximately linearly as the helical state is approached. Below MSDs of about 0.36 Å^2 this decrease accelerates, reflecting enthalpy gains from backbone hydrogen bonds that are traded off by entropy losses from a reduction in the number of possible peptide conformations. For A_2GA_2 compared to A_5, both entropy and enthalpy of helix formation have smaller slopes with respect to MSD. The minima in ΔH and ΔS near an MSD of 4.5 Å^2 reflect the enthalpic competition between hydrogen-bonded turnlike structures and the α helix.

Figure 2 shows the first passage time (FPT) distributions for helix formation and helix-to-coil transitions calculated

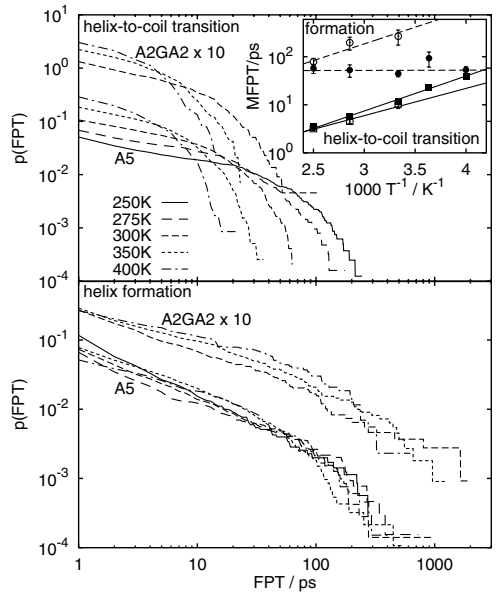

FIG. 2. FPT distributions for helix-to-coil transitions (top) and helix formation (bottom) for peptides A_5 and A_2GA_2 (scaled by ten). The inset in the top panel shows the MFPT as a function of T^{-1} for helix formation (circles, dashed lines) and helix-to-coil transitions (squares, solid lines) of A_5 (filled symbols) and A_2GA_2 (open symbols). Lines are fitted to an Arrhenius form, MFPT $\sim \exp(\Delta H^{\ddagger}/k_B T)$.

from equilibrium MD trajectories. FPTs are defined as the time to reach the dividing surface (MSD $= 0.36$ Å^2) between helix and coil states from an equilibrium distribution in the helix and coil ensembles, respectively. Also shown are the mean first passage times (MFPT) for helix formation and helix-to-coil transitions as a function of T^{-1}. We find that the helix-to-coil transition MFPTs of both peptides follow an Arrhenius T dependence, MFPT $\sim \exp(\Delta H^{\ddagger}/k_B T)$, with similar apparent activation enthalpies of $\Delta H^{\ddagger} = 13.9 \pm 1$ kJ mol^{-1} (A_5) and 10.9 ± 2 kJ mol^{-1} (A_2GA_2). FPT distributions for helix-to-coil transitions are exponential. MFPTs for helix formation are T independent between 250 and 400 K for A_5, and follow an Arrhenius law for A_5GA_5. FPTs for helix formation are strongly nonexponential, with a power-law decay dominating the initial times.

These thermodynamic and kinetic results indicate that helix unfolding is a thermally activated process. On the other hand, the T independence of the MFPTs for helix formation together with the nonexponential FPT distributions suggest that helix formation of A_5 is dominated by conformational diffusion in the coil ensemble with barrierless transitions into the helical state. A diffusive search explains naturally the nonexponential distribution of FPTs [26]: Significantly populated coil conformations close to the helical state lead to an initial nonexponential relaxation, while the long-time behavior is dominated by the slow relaxation of coil conformations distant from the helical state. In addition, the T-independent MFPTs for A_5 helix formation follow from a compensation between temperature-induced increases in the configuration-space diffusivity and the extent of populated coil configuration space. At elevated temperatures, this results in accelerated diffusion over larger average distances.

To quantify this conformational-diffusion model of helix formation kinetics, we numerically solve [27] Smoluchowski's equation for one-dimensional diffusion [13,28],

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial x}\left\{ D e^{-\beta \Delta G(x)} \frac{\partial}{\partial x}\left[e^{\beta \Delta G(x)} \rho \right] \right\} - k(x)\rho \,, \quad (1)$$

where $\beta^{-1} = k_B T$. Implicit is the assumption that the reaction coordinate x changes continuously and in small steps, as helix formation progresses. For x we empirically choose the MSD from the helical reference structure. The corresponding temperature-dependent free energy surfaces $\Delta G(x = \mathrm{MSD})$ are obtained from counting statistics of MD trajectories (see Fig. 1). D is the diffusion coefficient, assumed to be position independent. We use a strongly absorbing sink term $k(x) = 10^3$ ps^{-1} for $x < 0.36$ Å^2 and zero otherwise. $\rho(x, t)$ is the probability distribution of x at time t. An equilibrium distribution, $\rho(x, t = 0) \propto \exp[-\beta \Delta G(x)]$, in the coil region $x \geq 0.36$ Å^2 is used as the $t = 0$ initial distribution, with the helical region $x < 0.36$ Å^2 masked out.

For comparison, we calculate survival probabilities $S(t)$ of the coil state from the MD trajectories and the diffusion

model. $S(t) = \int_{\text{coil}} \rho(x,t)\,dx$ is the probability that the peptide is in a coil state at all times between 0 and t, given that it was in a coil state at time 0. $S(t)$ is related to the FPT distribution $p(t)$ through $S(t) = \int_t^\infty p(t')\,dt'$. The position-independent diffusion coefficients D are adjusted to reproduce the MFPT for helix formation at all temperatures except at 300 K, where D was estimated through matching of $S(t)$ over the first 200 ps.

Figure 3 compares the survival probabilities $S(t)$ for A_5 in the coil state calculated from MD simulations and Smoluchowski theory. We find good agreement between the diffusive-search model and the simulation data at all temperatures. After an initial nonexponential relaxation extending to between 50 and 100 ps, the $S(t)$ curves exhibit an approximately exponential decay. The inset in Fig. 3 shows the T dependence of the estimated diffusion coefficient D. With the exception of the least extensively sampled $T = 250$ K state with sluggish relaxation kinetics reflecting the supercooled solvent conditions, D exhibits Arrhenius behavior, $D \sim \exp(-\Delta H/k_B T)$ with $\Delta H = 13.8\ \text{kJ mol}^{-1}$. Also shown is an estimate of the diffusion coefficient for a harmonic model [27,28], $D_0 \approx \text{var(MSD)}/\tau_{\text{corr}}$, based on the ratio of the variance and the correlation time τ_{corr} of the MSD, both calculated over the whole trajectory. τ_{corr} is estimated by integrating the autocorrelation function of the MSD from time zero to the first crossing of the zero axis. We find good agreement between the two diffusion coefficients, with $D \approx D_0$ at 400 K and D_0 smaller by about a factor of 2 at 250 K. Thus, one-dimensional diffusion along the MSD reaction coordinate

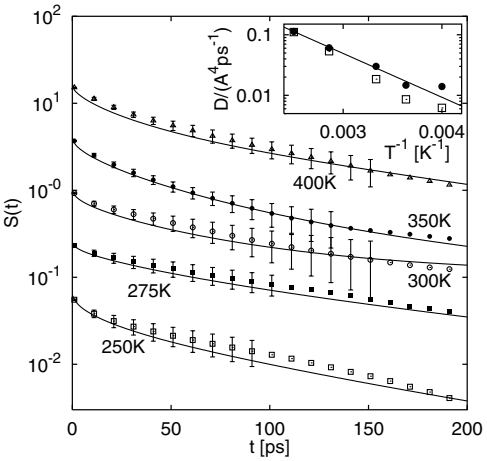

FIG. 3. Survival probability $S(t)$ in the coil state as a function of time. The simulation data are shown as symbols, with errors estimated from block averages for times t with five or more helix formation events contributing to $S(t)$. Calculations based on Smoluchowski's equation for a MSD reaction coordinate are shown as lines. The $S(t)$ curves for different temperatures are shifted vertically. The inset shows the diffusion coefficient D (filled circles) as a function of T^{-1}. The line in the inset is a fit to an Arrhenius law, $D \sim \exp(-\Delta H/k_B T)$. Open squares correspond to estimates of D from the variance of the MSD and the MSD-correlation time [27,28], $D_0 \approx \text{var(MSD)}/\tau_{\text{corr}}$.

quantitatively reproduces the helix formation kinetics of A_5 with diffusion coefficients as expected from equilibrium relaxation dynamics.

The diffusive helix formation model presented here can be tested experimentally. We find that the helix formation kinetics is nonexponential at short times (less than 10 to 100 ps). While this is below the time resolution of the reported experimental helix-coil transition studies, it should be within reach of current experimental technology, and would offer a means of directly comparing molecular simulations to experiments in the time domain. Recent experiments showed nonexponential relaxation kinetics for the T-jump induced re-folding of cold denatured proteins [9], interpreted similarly in terms of "downhill folding" [14,29]. Nonexponential helix-coil relaxation kinetics and weakly T-dependent folding rates were also observed for a synthetic homopolymer forming a helix stabilized by π-π stacking interactions [30]. Here, we accurately reproduce the helix formation kinetics by a diffusive search in the coil state, leading to nonexponential relaxation at the shortest times that turns into exponential relaxation at longer times.

Another experimentally testable prediction of a diffusive folding model is that in T-jump experiments [5,6], the relaxation will depend on the magnitude of the T jump. For conformational diffusion, the relaxation depends on the initial distribution in conformation space which is determined by the temperature *before* the T jump. Jump-width dependence together with nonexponential relaxation was indeed observed for the folding kinetics of cold shock protein A [31].

Moreover, the power-law decay of helix-formation times suggests that repeated single-molecule experiments [32] will lead to a wide distribution of folding times. This in turn directly affects the kinetic interpretation of folding simulations based on a small number of successful folding events [33,34]. On the other hand, repeated helix-to-coil transition simulations [35–37] will produce a rapidly decaying exponential distribution of *unfolding* times that cannot easily be inverted into *folding* times using two-state kinetics.

An analysis of the sequence dependence of helix formation reveals surprising similarities to protein folding models. Within the energy-landscape description of protein folding, the all-Ala peptide can be classified as a "fast-folding" peptide, where the helix formation kinetics is determined by downhill diffusion in conformation space without a significant free energy barrier to the helical state [14,29]. Substituting the central Ala by Gly introduces enthalpically stabilized "traps" into the coil state. Helix formation in A_2GA_2 thus requires a combination of activated escape from these enthalpically trapped conformations and subsequent conformational diffusion. This is in agreement with a theoretical study by Hardin *et al.* [29] who find, based on an effective energy function, that small helical proteins should fold without a barrier. Only as the

roughness of the energy landscape is increased, activated T dependence emerges.

It is remarkable that the "folding" kinetics of a small peptide into an α helix can be reproduced quantitatively by diffusion along a one-dimensional reaction coordinate, with T-dependent diffusion coefficients in agreement with those estimated from equilibrium fluctuations. For larger peptides and proteins, the identification of suitable reaction coordinates will be considerably more difficult, though at least in the case of lattice models of protein folding, diffusive models with carefully chosen reaction coordinates were shown to give accurate descriptions of the kinetics [28]. Thus, while the projection of conformation space onto a one-dimensional coordinate is an approximation with obvious limitations, it leads to a surprisingly simple and quantitative description of complex kinetics that makes the connection of folding to the thermodynamic driving forces—entropies, enthalpies, and free energies—transparent.

We thank W. A. Eaton, A. Szabo, R. B. Dyer, J. Hofrichter, and J. N. Onuchic for many stimulating discussions. This work was supported in part by the U.S. Department of Energy. Computer access to the Los Alamos ASCI Nirvana supercomputer through an IHPCI grant is gratefully acknowledged.

*Email address: hummer@helix.nih.gov

[1] R. B. Dyer, F. Gai, and W. H. Woodruff, Acc. Chem. Res. **31**, 709 (1998).
[2] W. A. Eaton, V. Muñoz, P. A. Thompson, E. R. Henry, and J. Hofrichter, Acc. Chem. Res. **31**, 745 (1998).
[3] M. Gruebele, J. Sabelko, R. Ballew, and J. Ervin, Acc. Chem. Res. **31**, 699 (1998).
[4] I. K. Lednev, A. S. Karnoup, M. C. Sparrow, and S. A. Asher, J. Am. Chem. Soc. **121**, 8074 (1999).
[5] S. Williams, T. P. Causgrove, R. Gilmanshin, K. S. Fang, R. H. Callender, W. H. Woodruff, and R. B. Dyer, Biochemistry **35**, 691 (1996).
[6] P. A. Thompson, W. A. Eaton, and J. Hofrichter, Biochemistry **36**, 9200 (1997).
[7] V. Muñoz, P. A. Thompson, J. Hofrichter, and W. A. Eaton, Nature (London) **390**, 196 (1997).
[8] C. M. Jones, E. R. Henry, Y. Hu, C. K. Chan, S. D. Luck, A. Bhuyan, H. Roder, J. Hofrichter, and W. A. Eaton, Proc. Natl. Acad. Sci. U.S.A. **90**, 11 860 (1993).
[9] J. Sabelko, J. Ervin, and M. Gruebele, Proc. Natl. Acad. Sci. U.S.A. **96**, 6031 (1999).
[10] Y. Duan and P. A. Kollman, Science **282**, 740 (1998).
[11] X. Daura, W. F. van Gunsteren, and A. E. Mark, Proteins Struct. Funct. Genet. **34**, 269 (1999).
[12] C. L. Brooks III and D. A. Case, Chem. Rev. **93**, 2487 (1993).
[13] J. D. Bryngelson and P. G. Wolynes, J. Phys. Chem. **93**, 6902 (1989).
[14] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).
[15] A. Šali, E. Shakhnovich, and M. Karplus, Nature (London) **369**, 248 (1994).
[16] K. A. Dill and H. S. Chan, Nature Struct. Biol. **4**, 10 (1997).
[17] Z. Y. Guo and D. Thirumalai, Biopolymers **36**, 83 (1995).
[18] H. A. Scheraga and M. H. Hao, Adv. Chem. Phys. **105**, 243 (1999).
[19] D. K. Klimov and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **97**, 2544 (2000).
[20] B. H. Zimm and J. K. Bragg, J. Chem. Phys. **31**, 526 (1959).
[21] S. Lifson and A. Roig, J. Chem. Phys. **34**, 1963 (1961).
[22] D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers* (Academic Press, New York, 1970).
[23] G. Hummer, A. E. García, and S. Garde, Proteins Struct. Funct. Genet. (to be published).
[24] A. E. García, Phys. Rev. Lett. **68**, 2696 (1992).
[25] The MSD between two structures $\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ and $\mathbf{y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$ is defined as the minimum average square distance of the N nonhydrogen backbone atoms at positions $\mathbf{x}_i$ and $\mathbf{y}_i$: $\mathrm{MSD} = N^{-1} \min \sum_i |\mathbf{x}_i - \mathbf{y}_i|^2$, where the minimum is taken with respect to all rigid body rotations and translations of conformation $\mathbf{x}$.
[26] D. J. Bicout and A. Szabo, Protein Sci. **9**, 452 (2000).
[27] D. J. Bicout and A. Szabo, J. Chem. Phys. **109**, 2325 (1998).
[28] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, J. Chem. Phys. **104**, 5860 (1996).
[29] C. Hardin, Z. Luthey-Schulten, and P. G. Wolynes, Proteins Struct. Funct. Genet. **34**, 281 (1999).
[30] W. Y. Yang, R. B. Prince, J. Sabelko, J. S. Moore, and M. Gruebele, J. Am. Chem. Soc. **122**, 3248 (2000).
[31] D. Thorn-Leeson, F. Gai, H. M. Rodriguez, L. M. Gregoret, and R. B. Dyer, Proc. Natl. Acad. Sci. U.S.A. **97**, 2527 (2000).
[32] Y. W. Jia, D. S. Talaga, W. L. Lau, H. S. M. Lu, W. F. DeGrado, and R. M. Hochstrasser, Chem. Phys. **247**, 69 (1999).
[33] S. H. Huo and J. E. Straub, Proteins Struct. Funct. Genet. **36**, 249 (1999).
[34] S. S. Sung and X. W. Wu, Proteins Struct. Funct. Genet. **25**, 202 (1996).
[35] K. V. Soman, A. Karimi, and D. A. Case, Biopolymers **31**, 1351 (1991).
[36] J. Tirado-Rives and W. L. Jorgensen, Biochemistry **30**, 3864 (1991).
[37] V. Daggett and M. Levitt, J. Mol. Biol. **223**, 1121 (1992).