A New Twist on the Helix-Coil Transition: A Non-Biological Helix with Protein-Like Intermediates and Traps

Sidney Elmer and Vijay S. Pande*

Department of Chemistry, Stanford University, Stanford, California 94305-5080

Received: May 31, 2000; In Final Form: September 26, 2000

All-atom computer simulations for the folding of a nonbiological helix, a 12-mer poly-phenylacetylene (pPA), are presented. We find that, unlike the classical helix-coil picture, pPA folds via on-pathway intermediate states and can also get trapped in misfolded states, much like that which one finds in simple models for proteins. Quantitative characterization of our pPA folding simulations finds a marked deviation from exponential kinetics, in agreement with experiment. Finally, we suggest how the helix-coil picture can be generalized to include these effects.

Helices are a very common structural motif in biological polymers such as proteins, DNA, and RNA. Also, helical motifs are a natural building block for nonbiological nanoscale structures due to their simplicity and fast folding properties. Moreover, the use of nonbiological protein-like units (such as helices and hairpins) in nanodesign naturally connects protein design techniques to the nonbiological realm.

Helix-coil theory[1,2] has stood as a simple means to understand helical self-assembly for many decades. However, there is growing experimental evidence that helix formation is not necessarily so simple. For example, Clarke et al.[3] have evidence that protein helices may form on millisecond time scales (comparable to time scales for the folding of small proteins). Moreover, Gruebele and Moore and co-workers[4] have direct evidence that poly-phenylacetylene (pPA), a nonbiological helix shown to fold to a helical state,[5] folds with nonexponential kinetics.

Here, we study thousands of molecular dynamics (MD) folding simulations of pPA in full atomic detail,[6] using a simple implicit solvent model[7] with langevin dynamics[8] and the MD code NAMD2.0.[9] These simulations quantitatively reproduce experimentally measured kinetics (see Figure 1) and allow us to theoretically examine how these nonbiological helices form and explain why their kinetics are so rich (nonexponential).

What could lead to nonexponential kinetics? One possibility is the existence of folding intermediates. This would be a marked departure from helix-coil theory, which states that once the critical nucleus (typically considered to be a single turn or interaction) is formed, folding should be rapid (which leads to a single-exponential kinetics). There are three principal possibilities for deviations from single-exponential kinetics: (1) intermediate states, (2) parallel pathways, and/or (3) off-pathway traps. We will see that all of these aspects are present in pPA folding, which is thus quite rich and reminiscent of theoretical models of protein folding.[10−13]

However, unlike proteins, which fold too slowly (tens of microseconds) for atomic-level simulations, pPA folding is considerably faster (tens of nanoseconds), allowing us to directly simulate thousands of folding trajectories and examine why pPA folding is so complex. Below, we first describe one simulation trajectory in detail and then discuss folding properties of our ensemble of simulated trajectories.

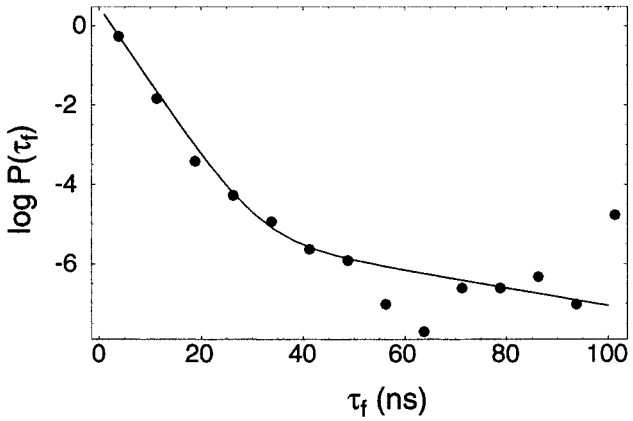

Figure 1. Are pPA folding times exponentially distributed? We plot the log of the folding time distribution log $P(\tau_f)$ for 2228 simulations with optimum interaction strength ϵ (see Figure 2). Runs that reached 100 ns were terminated; therefore, the data point at $t = 100$ ns represents runs which took *at least* 100 ns to fold. We see clear deviations from exponential kinetics. When we fit to a double exponential, we get characteristic times $\tau_1 = 5.3$ and $\tau_2 = 50.1$ ns. These times are comparable to experiment.[4]

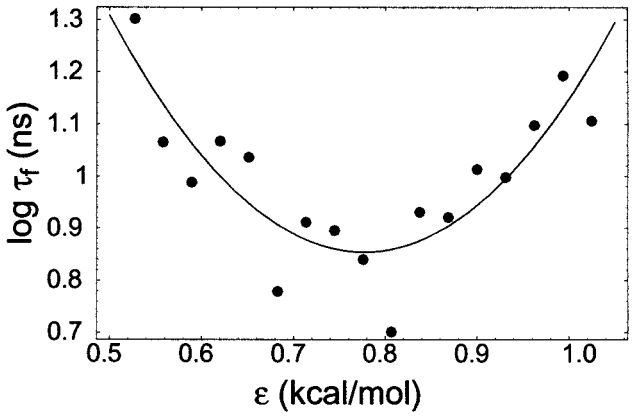

Figure 2. Mean folding time τ_f (dots) vs pPA ring stacking interaction strength ϵ. A parabolic fit is shown to guide the eye. We find that there is an optimum ϵ to minimize the folding time. Similar results have been found in simplified models for protein folding.[10−13] We have also found a linear depedence of τ_f on langevin dynamics viscosity parameter[8] γ (data not shown). Thus, our results can be generalized to dynamics in different solvents by rescaling τ_f for the appropriate viscosity.

* To whom correspondence should be addressed. E-mail: pande@stanford.edu.

10.1021/jp0019761 CCC: $20.00
Published on Web 12/19/2000

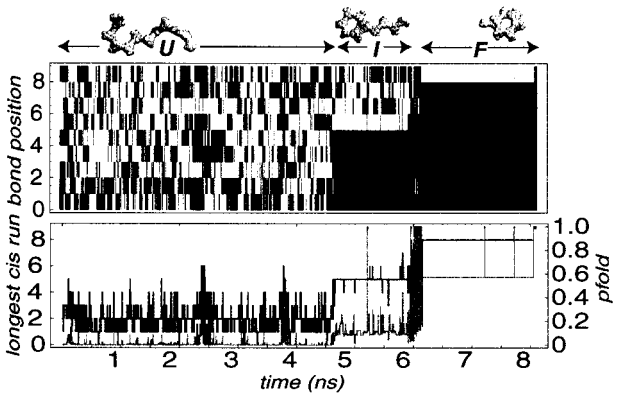

Figure 3. Graphical depiction of a pPA folding trajectory. (a) The 9 pPA dihedral angles (11 minus the two terminal dihedral angles) between phenyl rings (cis as black and trans as white) vs time. The folded state is shown as all-black (because it has all cis isomers). We see that (except for a brief ordering at 2.5ns) no structure forms until about 4.5ns. (b) Length of longest cis stretch vs time. Nucleation is the formation of some large stretch of folded structure (i.e., continuous stretch of cis dihedrals). This plot shows the size of this nucleus vs time. Superimposed on the nucleus size plot is the plot of p_{fold} vs time (gray line). We see that the nucleus size indeed correlates well with the p_{fold}, meaning that the nucleus size is a good reaction coordinate. As the nucleus size sharply jumps, this parameter does a good job of discriminating states U (0−4.5 ns), I (4.5−6 ns), and F (6-ns). Spacefilling snapshots of these states are also shown at the respective simulation time.

Figure 3 shows a single pPA folding trajectory under fastest folding conditions (see Figure 2). We see that the polymer spends some time in the unfolded state, then abruptly folds to and dwells in some partially folded intermediate state, and then finally, it folds. One important aspect of this folding trajectory is the existence of sharp transitions from unfolded state (U) to the intermediate (I) to the folded state (F), which implies that each of these states are free energy minima separated by free energy barriers.

How typical is this folding trajectory? In Figure 4, we show four other folding trajectories under the same conditions (trajectories differed by langevin dynamics random number seeds[9]). We see that although the character of the kinetics is similar (in that one finds U, I, and F states with sharp transitions between them), the precise structure of the intermediate state is not strongly conserved between folding trajectories.

Naturally, the first step toward understanding a thermodynamic system is to find a reaction coordinate, since it is a property of the system which describes the transitions between phases and the overall state of the system. To not bias our result by the choice of a reaction coordinate, we have employed the p_{fold} method[14] to characterize our system: we calculate the probability (p_{fold}) that a given configuration reaches the folded state before unfolding. States in the unfolded and folded basins of attraction will have p_{fold} values of 0 and 1, respectively. Transition states will have $p_{fold} \approx 1/2$. Although p_{fold} can serve as an intrinsic reaction coordinate for all systems, it has no direct geometrical interpretation. The reaction coordinate that provides these details would be a parameter (with either a physical or a geometrical interpretation) that correlates well with the p_{fold} trajectory.

We defined the unfolded state to have a high energy ($E > -2.5$ kcal/mol) and no contacts (two monomers are said to be in contact if the distance between center of the phenyl rings of two monomers are within 4 Å). The folded state was defined

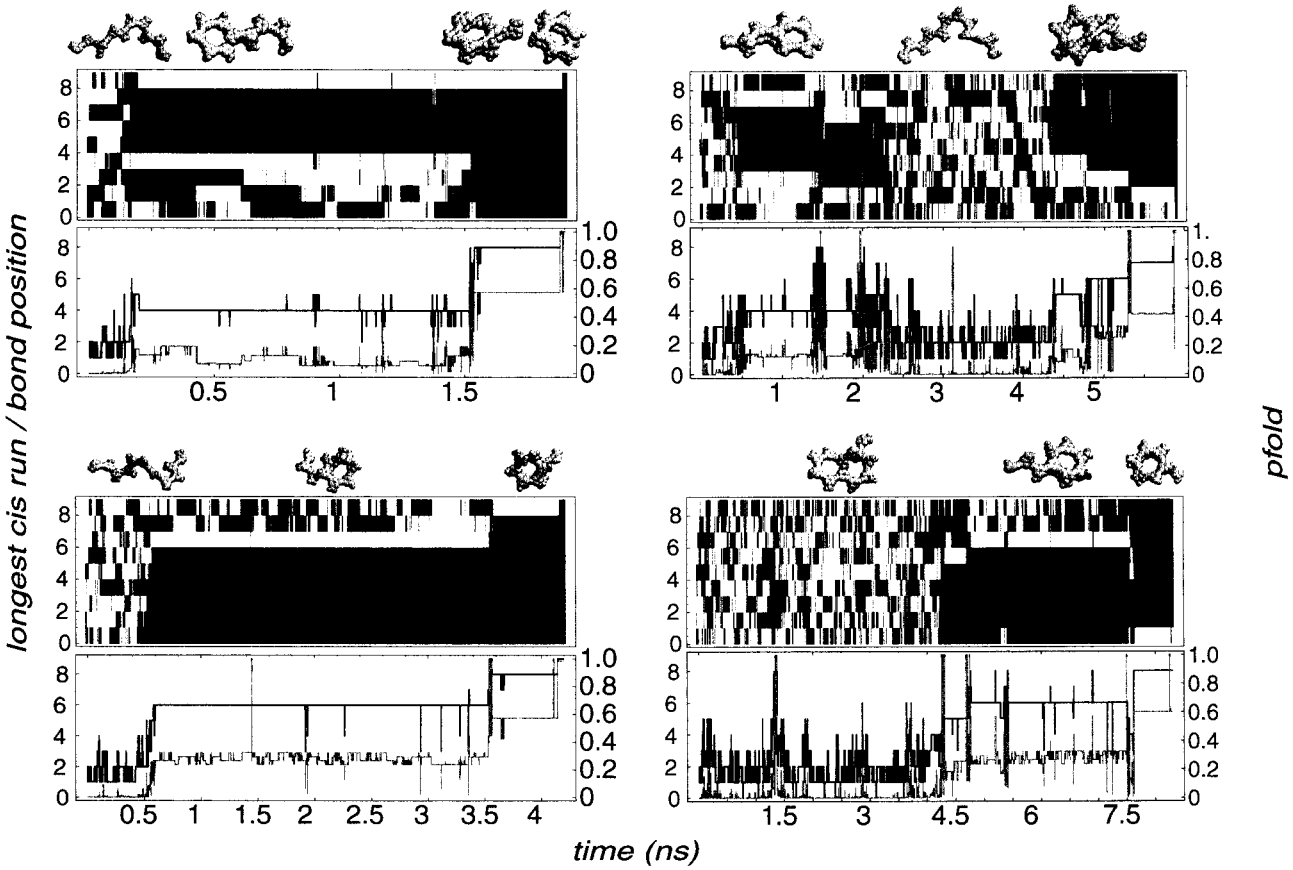

Figure 4. (below) Comparison of folding trajectories. We plot dihedral angles and spacefilling plots vs time (as in Figure 3) for four additional folding trajectories. We see that there is a great deal of variety in the folding pathways. In the upper right frame, we see folding to and unfolding from a trap state.

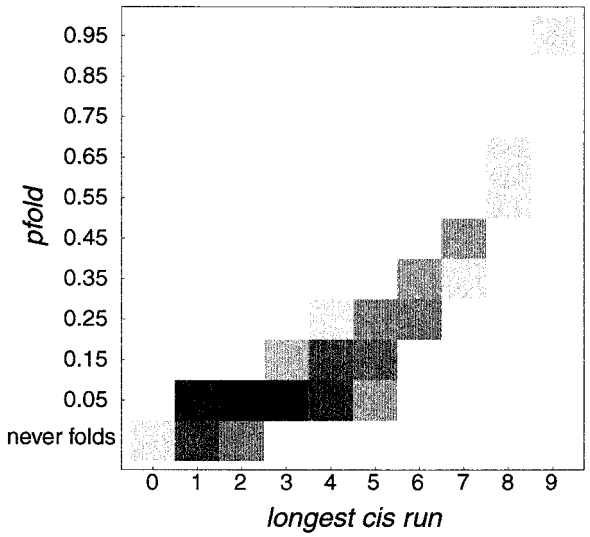

Figure 5. Does the number of consecutive cis bonds (*D*) correlate with p_{fold}? To address this question, we can plot the probability that a conformation has a given value of *D* and p_{fold} (darker shading means greater probability, logarithmically spaced). We see that *D* and p_{fold} correlate extremely strongly, and thus *D* can serve as a good reaction coordinate.

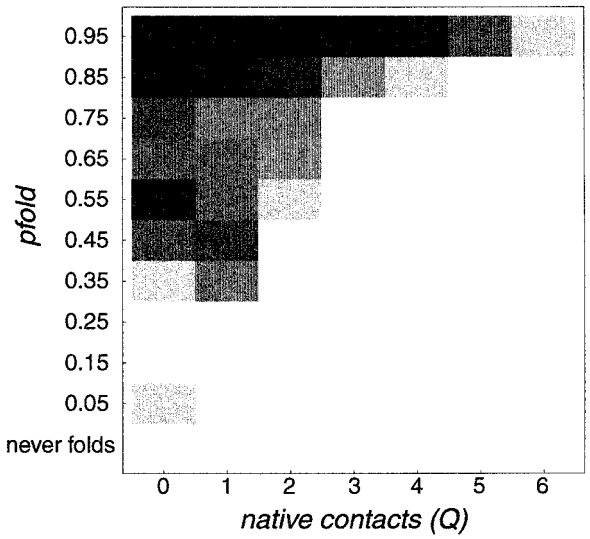

Figure 6. Does the number of native contacts *Q* correlate with p_{fold}? To address this question, we can plot the probability that a conformation has a given value of *Q* and p_{fold} (darker shading means greater probability, logarithmically spaced). We see that *Q* and p_{fold} are virtually uncorrelated and thus, in this system, *Q* is not a good reaction coordinate.

to have six native contacts (a *native* contact is present between two monomers in a given configuration if they are in contact in the given configuration as well as in the native state). On the basis of these definitions of the folded and unfold states, we found that the length of consecutive cis dihedral angles, *D*, does indeed correlate well with the intrinsic p_{fold} reaction coordinate. Figure 5 shows this correlation distribution. The distribution of configurations that appear along the folding trajectory is tightly concentrated around a narrow range of p_{fold} values for a given *D*. And the transition state, $p_{fold} \approx 0.5$, is found to be a critical nucleus of 7 consecutive cis dihedral angles.

This leads to the question, can we use the number of native contacts as a reaction coordinate? Figure 6 shows the correlation distribution for p_{fold} versus *Q*, the number of native contacts, with the same definitions for the folded and unfolded states used for the p_{fold} vs *D* distribution. The distribution is broad and, unlike the number of consecutive cis dihedrals (*D*), *Q* does not correlate well with p_{fold} and thus does not serve as a reaction coordinate for this system.

With p_{fold}, we can also characterize off-path traps as states with some structure (e.g., some contacts) but virtually always completely unfold (p_{fold} very near zero when the unfolded state is defined as $E > -2.5$ kcal/mol and zero contacts) and can never fold without unfolding to some degree ($p_{fold} = 0$ when the unfolded state is defined as $E > -7.5$ kcal/mol and less than 2 contacts). These configurations (e.g., see Figure 4, upper right) involve non-native contacts and form nonlocal structure in the spirit of a protein beta-hairpin.

Thus, we have shown that pPA, a seemingly simple homopolymer, folds into a helix by a rather rich mechanism. This richness appears to come from two aspects of this polymer. First, the monomers are restricted largely to only 180° isomerations,[15] which induces some "steric" frustration and leads to a unique native state. Second, because this chain is a homopolymer, there is a strong possibility for non-native interactions (because *all* interactions are equally preferable in terms of the hydrophobic interactions); this likely helps to stabilize the off-pathway intermediate. This homopolymeric nature also contributes to the large dissimilarity of the structures of on-pathway intermediates because there is no energetic preference for different intermediate states (unlike that which one would expect to find in heteropolymers, such as proteins[13]).

Finally, one could build a simple theoretical model for this system by incorporating on-pathway intermediates into a generalized helix-coil theory: the polymer must form critical nuclei in each step (from $U \rightarrow I$ and $I \rightarrow F$). This inclusion of intermediate states is also interesting in the light of recent theoretical studies of beta-hairpins, dominant building blocks of proteins, which also fold via intermediate states.[16,17] Considering the complexity of the folding of these simple elements of structure, combinations of these elements in proteins and nonbiological self-assembling structures could easily have even richer and more fascinating folding properties.

References and Notes

(1) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526−535.

(2) Lifson, S.; Roig, A. *J. Chem. Phys.* **1961**, *34*, 1963−1974.

(3) Clarke, D. T.; Doig, A. J.; Stapley, B. J.; Jones, G. R. *Proc. Natl. Acad. Sci.* **1999**, pp 7232−7237.

(4) Yang, W. Y.; Prince, R. B.; Sabelko, J.; Moore, J. S.; Gruebele, M. *J. Am. Chem. Soc.* **2000**, *122*, 3248−3249.

(5) Nelson, J. C.; Saven, J. G.; Moore, J. S.; Wolynes, P. G. *Science* **1997**, *277*, 1793−1796.

(6) We built our parameter set based upon the potential equation and values from CHARMM19.[18] We added values for the dihedral angle potential between rings (important for the different isomers of this polymer). For this angle dihredral potential, we used values from ref 15.

(7) Although proteins interact with water both via hydrogen bonding and the hydrophobic effect, pPA only interacts via hydrophobicity. This allows us to greatly simplify our model for the solvent. To mimic the attraction between phenyl rings caused by the hydrophobic effect, we have included a Lenard−Jones interaction between ring carbon atoms $V_{LJ} = \epsilon[(\sigma/r)^{12}-(\sigma/r)^6]$ where we chose $\sigma = 2.8$ Å and varied the well depth ϵ to correspond to different polar solvents. This model is very computationally tractable: on a 400 MHz Pentium II processor, we can simulate approximately 15 ns (twice the folding time) in a CPU day. Moreover, its tractability renders pPA a new type of "simplified" model to study issues related to protein folding.[10−13]

(8) We have used Langevin dynamics, which include a drag term into the force $F_{drag} = -\gamma v$, where v is the velocity. We have set $\gamma = 1$/ps for the results reported in the paper. We also find a linear dependence of folding time on viscosity (data not shown), which is reasonable because small changes in γ likely changes the microscopic time at which dynamics occurs, rather than the underlying mechanism.

(9) Kal, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283−312.

(10) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins: Struct. Funct. Genet.* **1995**, *21*, 167−195.

(11) Chan, H. S.; Dill, K. A. *Nat. Struct. Biol.* **1997**, *4*, 10−19.

(12) Shakhnovich, E. I. *Curr. Opin. Struct. Biol.* **1997**, *7*, 29−40.

(13) Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Rokhsar, D. S. *Curr. Opin. Struct. Biol.* **1998**, *8*, 68−79.

(14) Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. I. *J. Chem. Phys.* **1998**, *108*, 334−350.

(15) Okuyama, K.; Hasegawa, T.; Ito, M.; Mikami, N. *J. Chem. Phys.* **1984**, *88*, 1711−1716.

(16) Pande, V. S.; Rokhsar, D. S. *Proc. Natl. Acad. Sci.* **1999**, *96*(16).

(17) Dinner, A. R.; Lazaridis, T.; Karplus, M. *Proc. Natl. Acad. Sci.* **1999**, *96*, 9068−9073.

(18) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(19) We thank Jeff Moore and Martin Gruebele for their insightful comments and review of our manuscript. Some calculations were run on the T3E at the National Energy Resources Supercomputer Center (NERSC).