Foldamer dynamics expressed via Markov state models. I. Explicit solvent molecular-dynamics simulations in acetonitrile, chloroform, methanol, and water

Sidney P. Elmer, Sanghyun Park, and Vijay S. Pande

Citation: The Journal of Chemical Physics **123**, 114902 (2005); doi: 10.1063/1.2001648
View online: http://dx.doi.org/10.1063/1.2001648
View Table of Contents: http://aip.scitation.org/toc/jcp/123/11
Published by the American Institute of Physics

Foldamer dynamics expressed via Markov state models. I. Explicit solvent molecular-dynamics simulations in acetonitrile, chloroform, methanol, and water

Sidney P. Elmer, Sanghyun Park, and Vijay S. Pande[a)]
Department of Chemistry, Stanford University, Stanford, California 94305-5080

(Received 1 June 2005; accepted 24 June 2005; published online 16 September 2005)

In this article, we analyze the folding dynamics of an all-atom model of a polyphenylacetylene (pPA) 12-mer in explicit solvent for four common organic and aqueous solvents: acetonitrile, chloroform, methanol, and water. The solvent quality has a dramatic effect on the time scales in which pPA 12-mers fold. Acetonitrile was found to manifest ideal folding conditions as suggested by optimal folding times on the order of ~100–200 ns, depending on temperature. In contrast, chloroform and water were observed to hinder the folding of the pPA 12-mer due to extreme solvation conditions relative to acetonitrile; chloroform denatures the oligomer, whereas water promotes aggregation and traps. The pPA 12-mer in a pure methanol solution folded in ~400 ns at 300 K, compared relative to the experimental 12-mer folding time of ~160 ns measured in a 1:1 v/v THF/methanol solution. Requisite in drawing the aforementioned conclusions, analysis techniques based on Markov state models are applied to multiple short independent trajectories to extrapolate the long-time scale dynamics of the 12-mer in each respective solvent. We review the theory of Markov chains and derive a method to impose detailed balance on a transition-probability matrix computed from simulation data. © *2005 American Institute of Physics*. [DOI: 10.1063/1.2001648]

I. INTRODUCTION

It may be argued that the most important substance to living organisms is water.[1] Water's ability to solvate proteins, nucleic acids, and saccharides, in addition to small polar and ionic organic molecules and inorganic ions, allows biological components to be present in a fluid, dynamic medium. Some chemical properties, in particular, which promote the solvation of this broad spectrum of compounds include the ability for water to be both a hydrogen bonding donor and acceptor, its highly polar and high dielectric bulk electrostatic properties, as well as acid-base properties which allow for a very broad range of *p*H environments.

Just as the aqueous environment in biological systems plays a fundamental role in biological function, similarly the solvent medium is a central component to be considered in the design of novel self-assembling nanodevices and functional nanomaterials. In the past few years, polyphenylacetylene[2–4] (pPA), also known as phenylene ethynylene oligomers,[5–8] has gained significant attention in the field of foldamer polymer chemistry. While there already exist demonstrations of biomimetic functionality,[9–16] the potential exists for even more significant and complex applications.[17] In an effort to establish a foundation for understanding the role of the solvent in nonbiological self-assembly processes, Hill and Moore[18] have studied the experimental folding propensities of a pPA 12-mer in a wide variety of organic solvents. They observed that most solvents promote self-assembly to the helical folded state, suggesting that pPA oligomers are inherently biased toward folding.

Our objective for the present study is to build upon the experimental work of Hill *et al.* by using computational methods to gain a greater understanding of the role of the solvent in the folding process of pPA oligomers. We address fundamental questions regarding the optimal solvation environment necessary to design and engineer functional biomimetic nanomaterials without being limited to aqueous solutions. For instance, which solvents promote optimal conditions for rapid self-assembly? Furthermore, how does temperature affect the folding kinetics of the pPA 12-mer? In order to address these issues, we model and characterize the folding dynamics of the pPA 12-mer in explicit solvent. We choose four of the more common solvents reported by Hill and Moore for which explicit solvation models are readily available: water, chloroform, acetonitrile, and methanol.

Each of these four solvents has unique experimental properties that we would like to investigate *in silico*. For instance, chloroform is the solvent of choice to unfold or denature pPA oligomers; on the other hand, acetonitrile has been predominantly used to promote folding.[6] Solvent mixtures of chloroform and acetonitrile have been used to detect the cooperative folding behavior of various lengths of pPA oligomer chains,[6] in addition to monitoring the folding-induced twist-sense biasing of the helix toward a desired chirality when chiral side chains are incorporated into the oligomer backbone.[12] Consequently, chloroform is used to characterize the polymer dynamics in the good solvent regime, while acetonitrile is assumed to be close to the theta solvent regime expressed by optimal folding conditions. Water and methanol, on the other hand, are much more polar solvents;[18,19] in general, pPA oligomers are not soluble in pure water, and only marginally soluble in pure methanol.

[a)]Electronic mail: pande@stanford.edu

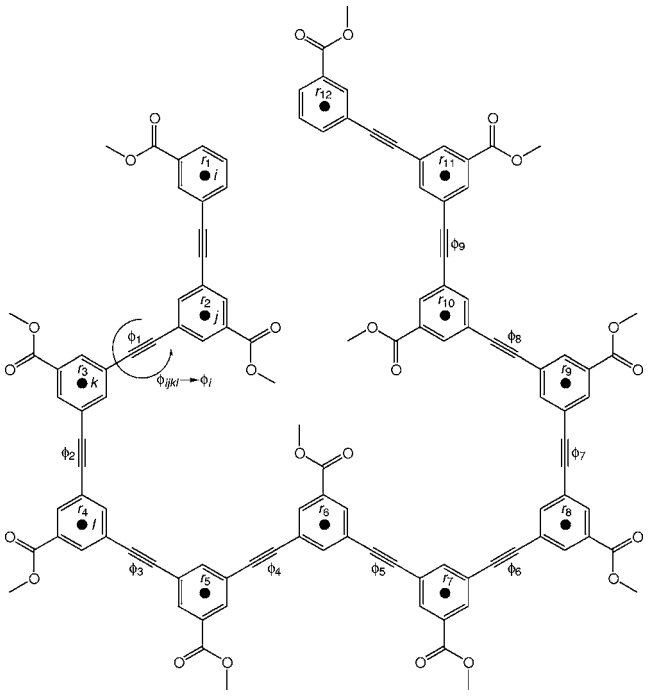

FIG. 1. Chemical structure of a pPA 12-mer. The hydrogen atoms are omitted in this representation, but are present in the all-atom computational model. The side chains are carboxymethyl groups. The methyl group is added to help reduce the charges in the system by capping the carboxylate side chains. The folded state of pPA oligomers consists of a compact helical topology as the oligomer adopts all *cisoid* dihedral angles between adjacent phenyl rings. The dihedral mapping relationship shown for dihedral angle 1 $\phi_{ijkl} \rightarrow \phi_i$ is a general relationship for all dihedral angles.

Very polar side chains may be designed into the pPA backbone to increase the solubility in aqueous solutions, but even then, pPA oligomers aggregate at relatively high water concentrations.[11,19,20] Fortunately, solubility issues are not a concern *in silico* (since we are only simulating a single chain and therefore aggregation is not an issue). Consequently, the water and methanol explicit solvation models are used to characterize the polymer dynamics in the poor solvent regime.

The primary force driving pPA oligomers to fold is solvophobicity.[5] In a poor solvent environment, the oligomer interacts weakly with the solvent (hence the term solvophobicity); therefore, the oligomer strongly attracts itself causing the extended chain to collapse into compact states. The fully helical folded state is one of many such compact states. It is stabilized over other compact states by two important structural contributions. First, the backbone has a rigid, shape-persistent structure[21] due to electronic conjugation over the extended π system of the molecule (see Fig. 1) which preferentially induces the chain to adopt coplanar configurations. Secondly, the π-π aromatic stacking interactions between monomers separated along the sequence by six monomer units $(i,i+6)$ provide energetic stabilization to adopt and remain in a helical topology with regularly repeating structural features, in contrast with random compact states typical of organic polymers.

Lee and Saven have developed a computational model of pPA oligomers in order to study the dynamics of the helical folded state of a pPA 18-mer in TIP3P water.[7] They observed large fluctuations in the radius of the helix in the folded state that were highly correlated with the evacuation of water from the cavity of the helix. In addition, the helical structure was very tolerant to large fluctuations in the dihedral angles along the backbone without disrupting the global folded helical structure. A suggestion was proffered that these highly flexible features of the folded state of pPA oligomers could be exploited to host a wide variety of guest molecules in biomimetic molecular recognition applications.

We have previously characterized the folding kinetics of both pPA 12-mer[2,3] and 20-mer[4] chains in implicit solvent via molecular-dynamics computer simulations. We used a very simple, generic short-range implicit solvent model in which the solvophobic attraction driving the chain to fold is modeled as a Lennard-Jones (LJ) 6–12 potential between distant monomers in the chain: $V_{\mathrm{LJ}}(r)=4\varepsilon[(\sigma/r)^{12}-(\sigma/r)^{6}]$. Modulating the well depth ε was the primary means to modify the solvent quality. From this simplified model, we were able to reproduce the nonexponential cumulative folding time distribution observed experimentally for the 12-mers.[8] Moreover, the folding time extracted from the 12-mer simulations[3] was in quantitative agreement with the experimental folding rate measured from a fit to a two-state model using laser *T*-jump fluorescence spectroscopy.[8]

It is clear that a great deal of research has already been performed, both experimentally and computationally, on pPA structure, dynamics, and potential function. This current work is intended to apply state-of-the-art computational methods to further our understanding of pPA oligomers: to model the dynamics (e.g., folding mechanism) and kinetics (e.g., folding rate) in a range of explicit solvents spanning the extremes of solvent quality.

An essential element of our work concerns the methods used to analyze the simulation data. As more and more research groups gain access to more powerful computational resources, improved methodologies are being developed to analyze the dynamics of molecular trajectories at the ensemble level; meaning that hundreds or even thousands of independent trajectories are generated in order to obtain a representative statistical sample of conformation space. The folding process is a diffusive stochastic process, which requires adequate sampling to characterize the folding process. On the order of hundreds of trajectories are needed to obtain good estimates for the statistical distribution of states and the relative weights of folding pathways.[22] Therefore, the methods used to analyze the simulation data sets must be capable of dealing with multiple trajectories.

Markov state models (MSMs) allow one to easily incorporate multiple trajectories into a simple description of the folding dynamics. Essentially, each molecular trajectory is treated as a Markov chain from which a stochastic transition-probability matrix may be computed between the states defined in the system; the state space decomposition itself is such a crucial component of Markov state models that we wrote a separate companion paper[23] dedicated primarily to this topic. The transition matrix offers an effective way to combine the dynamics of each of the individual trajectories into one mathematical object representing an approximation to the exact dynamics of the molecular system. The longer

the simulation times and the more trajectories that are incorporated into the transition matrix, the better the sampling of conformation space, and the closer the computed transition matrix approaches the exact dynamics of the molecular system defined by a particular state space decomposition.

We present an overview of the theory of Markov chains in Sec. III and apply these Markov state model techniques to the simulation data sets in Sec. IV. Also in Sec. IV, we discuss how the pPA 12-mer folding dynamics are dominated by the solvent quality and check the validity of the Markov state models in predicting the polymer dynamics from multiple short independent trajectories.

II. SIMULATION DETAILS

The simulation strategy which we adopt in this study is to use simple models to generate starting structures for more complex models. Multiple short independent trajectories using a complex model are initiated from these starting structures obtained from the simple model. The transition-probability matrix describing the dynamics of the complex model system is subsequently computed from these multiple short independent trajectories.

Starting structures were taken from a previous study of pPA 12-mers using a generic short-range implicit solvent model under optimal folding conditions (the well depth ε was tuned to minimize the mean first passage time).[2] Using the *K*-mean clustering algorithm, the conformations from an ensemble of over 2000 folding trajectories had been classified into eight distinct clusters.[3] The molecular structure closest to the mean conformation in a given cluster represents the structure of that cluster as a whole; therefore, the eight structures representing the eight clusters in the simplified model were subsequently used as a set of initial coordinates in the current study. These eight structures spanned conformation space fairly evenly, ranging from an unfolded structure, to several helical intermediates with frayed ends, to the helical folded structure; also, a nonhelical conformation analogous to a β turn in proteins and a knotted structure were included (see Fig. 2). The implicit solvent model is fast yet less accurate than the more complex explicit solvent models. Therefore, the implicit solvent models sample conformation space quickly and give rough insights into the more relevant states in the system. Multiple trajectories in explicit solvent are initiated from a single initial state, each with different initial velocities. These trajectories quickly diverge from each other and from the initial state to sample conformation space, making transitions between the states defined by the state space decomposition.

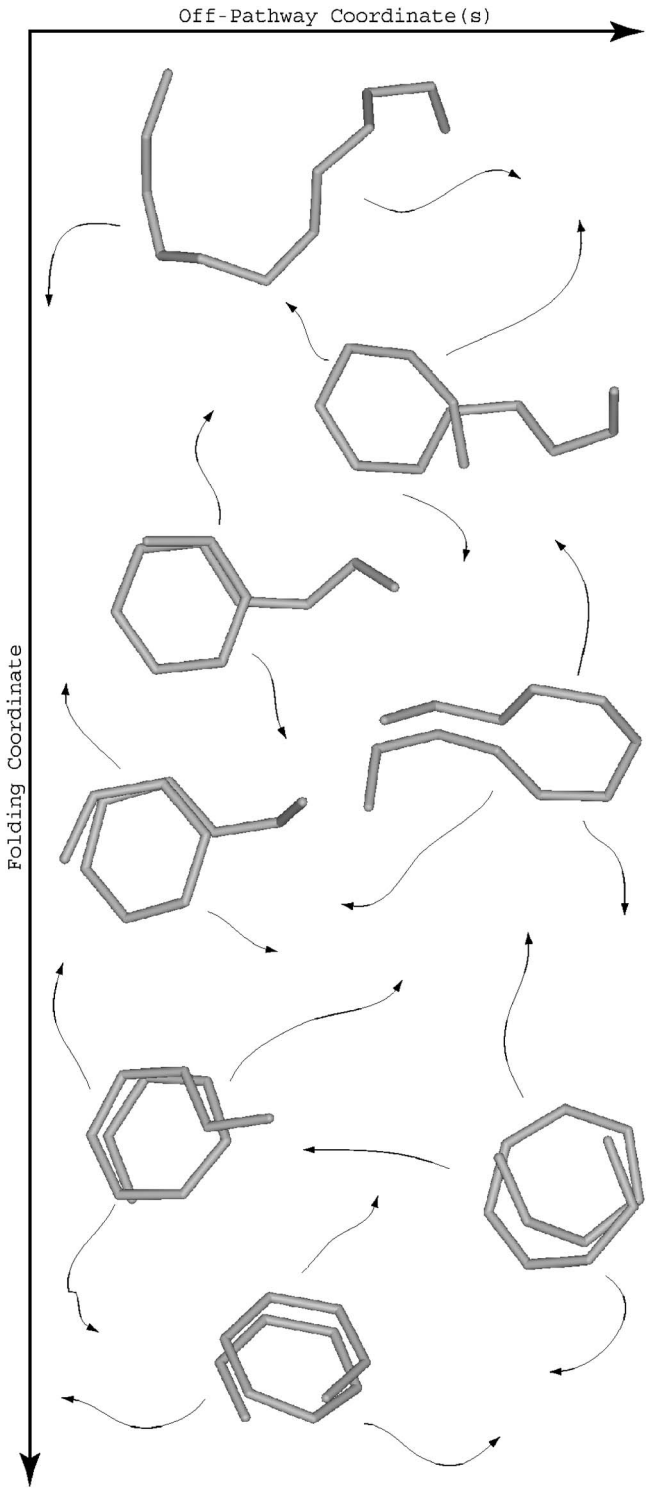

FIG. 2. Illustration of the approach taken to efficiently sample conformation space using multiple short independent trajectories. The starting structures (represented as stick figures, each vertex representing a phenyl ring with the side-chains removed) were taken from the clustering results of a simple implicit solvent model. From these starting structures, multiple trajectories using a more complex explicit solvent model were calculated using the Folding@Home distributed computing network. Central to this paper is the application of Markov state models, analysis methods based on the Markov chains, which provide a method to extrapolate the long time-scale dynamics of these short independent trajectories.

A. pPA model

The molecular mechanics model of an all-atom pPA 12-mer (see Fig. 1) was created using the AMBER8 molecular-dynamics suite and subsequently ported into the GROMACS (http://www.gromacs.org) molecular-dynamics suite, version 3.1.4 modified for use with Folding@Home, to calculate the dynamics trajectories. We used the generalized amber force field (GAFF99) parameter set to define bonded interactions and nonbonded van der Waals terms. The GAFF99 force field was developed primarily to model small organic molecules in a generic, uniform way. The partial atomic charges were calculated using the semiempirical ANTECHAMBER pro-

gram within AMBER8 using the AM1-BCC option for point charge fitting. We also added proper dihedral interactions of the form

$$V_d(\phi) = k_\phi[1 + \cos(2\phi - 180\,^\circ)], \quad (1)$$

across the acetylene bond between two adjacent phenyl rings to bias the chain toward coplanar conformations with energetic minima at 0° (*cisoid*) and 180° (*transoid*). The barrier height to isomerization $k_\phi/2 = 0.6$ kcal/mol was taken from experimental spectroscopic data.[24]

B. Solvent models

The explicit solvent models for water, chloroform, methanol, and acetonitrile, the four solvents used in this study, are well known, standard models. A rigid TIP3P water model with the bond lengths and bond angle constrained using the SETTLE algorithm within GROMACS was used to study the dynamics of pPA 12-mers in aqueous solution. Explicit solvent models for the other three solvents (acetonitrile, methanol, and chloroform) come prepackaged with the AMBER8 distribution and were also ported into GROMACS. These explicit solvent models were developed based on the AMBER99 parameter set, which is compatible with the GAFF99 parameter set used to model the pPA oligomer. Only the bonds were constrained using the LINCS algorithm in GROMACS; thus, these organic solvent models were not completely rigid, as the angles were allowed to fluctuate in all three models.

C. Simulation parameters

The simulations for each explicit solvent were run at two temperatures, 300 and 330 K, to investigate the temperature effects as well as the effects of solvent quality on the dynamics of pPA 12-mer chains. The boiling point for chloroform is 334.85 K,[25] the lowest of the boiling points for the four solvents; hence 330 K is a reasonable choice for the higher-temperature simulations. Eight sets of initial structures were prepared; one set for each solvent at a given temperature. For a particular set of initial structures, each of the eight initial structures was solvated separately, energy minimized, and equilibrated for 200 ps at the appropriate temperature, keeping the oligomer restrained in its initial conformation. These equilibrated structures were the starting points for production dynamics runs at the appropriate temperature.

The world-wide distributed computing platform Folding@Home (http://folding.stanford.edu) was designed as a resource to distribute dynamics calculations over a very heterogeneous supply of processors all donated voluntarily by users from all over the world. We used Folding@Home to compute many independent trajectories for each of the eight separate data sets. A single data set contained 1600 trajectories; for each of the eight initial structures in the data set, 200 trajectories of varying length were computed (time distributions shown in Fig. 3), each differing in random number seed. The aggregate simulation time for all eight data sets amounts to $\sim 300\ \mu$s.

Additional simulation parameters are as follows. A 2-fs time step was used, constraining all bonds using LINCS.

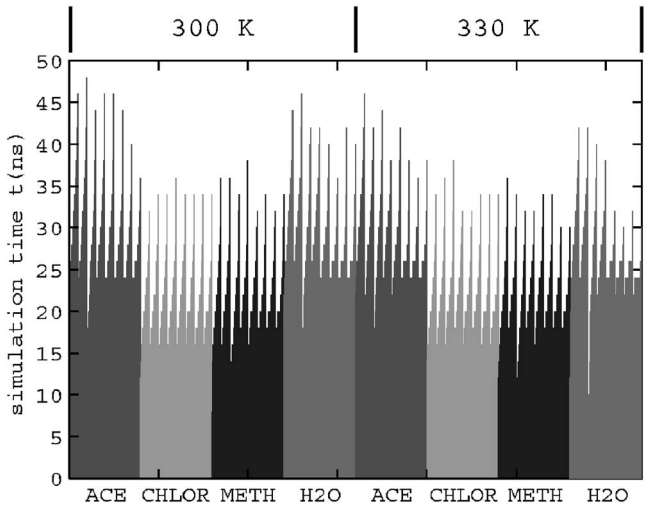

FIG. 3. Simulation time distributions for the eight data sets analyzed in this paper, pPA 12-mers in four solvent systems; acetonitrile, chloroform, methanol, and water; at two temperatures each, 300 and 330 K. The average simulation times taken from these distributions are as follows: at 300 K, the average simulation times for acetonitrile, chloroform, methanol, and water are 26.1, 19.0, 18.9, and 25.6 ns, respectively; at 330 K, the average times are 24.7, 18.9, 19.0, and 23.6 ns, respectively. Contrast these average simulation times with the folding times computed from the Markov models (see Fig. 6).

The particle mesh Ewald (PME) summation algorithm was used to calculate the electrostatic potential energy with a grid spacing of 0.12 nm and a tolerance of 10^{-5}. The van der Waals cutoff distance and the Ewald long-range cutoff separating the direct space from the inverse space were both the same for a given solvent system, although the pair of cutoffs was set to different values for different solvents. The size of the solvent molecule determines the necessary cutoff distances; larger solvent molecules result in larger first and second solvent coordination shells, so the larger solvent molecules require larger cutoffs. The cutoffs were 0.8 nm for water, 0.9 nm for methanol, and 1.0 nm for both acetonitrile and chloroform, respectively, reflecting the ordering of the relative sizes of the solvent molecules: water $<$ methanol $<$ acetonitrile $\approx$ chloroform. The nonbonded van der Waals interactions used a switching function with the switch turned on 0.2 nm closer than the long-range cutoff distance. The neighbor list cutoff was set to be 0.2 nm further away than the cutoff distance. Different size boxes were used for each solvent system; the box size for the acetonitrile simulations was 5 nm, the box size for the chloroform simulations 7 nm, and both water and methanol simulations were run in boxes 6 nm on a side. The simulations were run at constant temperature and pressure using Berendsen temperature and pressure controls.

III. THEORY OF MARKOV CHAINS

The Markov assumption states that the future state of a stochastic system only depends on the current state, or in other words, a Markovian system is senile and has no memory of its history. Stochastic systems which are Markovian are much easier to characterize than systems which strongly depend on their histories. This is because the math-

ematics of Markov processes has been developed thoroughly and is described succinctly by relatively simple principles of linear algebra.[26,27]

Several research groups have developed methodologies for analyzing molecular simulation data using techniques based on Markov chains.[28–31] We build upon the method developed by Swope, Pitera, and Suits to characterize the folding properties of pPA 12-mers in explicit solvent. Markov chains are ideally suited to describe the dynamics of molecular systems due to their mathematical properties encapsulated in the stochastic transition matrix $P_{ij}(\tau)$ over a set of discrete states p_i. We use the row-dominant representation of the transition matrix $P_{ij}(\tau)$, meaning that the rows of the matrix sum to one. The elements of the transition matrix are defined to be the probability that the system is in state p_j at time $t+\tau$, given that the system was initially in state p_i at time t, i.e., $P(p_j, t+\tau | p_i, t) \equiv P_{ij}(\tau)$. The given stochastic process is Markovian when it obeys the following discrete time equation:

$$X_j((m+1)\tau) = \sum_i X_i(m\tau) P_{ij}(\tau), \tag{2}$$

for a probability distribution $X_i(t)$ at a given point in time.

A. Investigating Markov properties

The transition matrix $P_{ij}(\tau)$ may be diagonalized in order to decompose the problem into an eigenspace description of the dynamics.[31] Under the row-dominant convention the first right eigenvector corresponding to the Perron root $\lambda_1=1$ is $e=[1\cdots1]^T$ and the first left eigenvector is the normalized stationary distribution $\pi=[\pi_i\cdots\pi_N]^T$ satisfying $\pi^T e=1$. Assuming the transition matrix satisfies detailed balance $\pi_i P_{ij} = \pi_j P_{ji}$, the eigenvalues λ_i are real and bounded in the range $[-1,1]$, i.e., $|\lambda_i| \leqslant 1$. In addition, the right eigenvectors x_i form an orthonormal basis set with respect to the weighted inner product $\langle x_i, x_j \rangle_\pi = x_i^T \mathcal{D}^2 x_j = \delta_{ij}$, where $\mathcal{D}=\mathrm{diag}(\sqrt{\pi_i})$. The left eigenvectors y_i also form an orthonormal basis set; however, they are orthonormal with respect to the weighted inner product $\langle y_i, y_j \rangle_\pi = y_i^T \mathcal{D}^{-2} y_j = \delta_{ij}$ where the weighting factor $\mathcal{D}^{-2}$ is now the inverse of the diagonal matrix of the normalized stationary distribution. The left and right eigenvectors corresponding to eigenvalue λ_i also have the direct relationship $y_i = \mathcal{D}^2 x_i$.

Returning to the Markov equation [Eq. (2)], the dynamics of a Markovian system may be rewritten in matrix notation by the equation:

$$X((m+1)\tau) = X(m\tau) P(\tau). \tag{3}$$

We emphasize that when the transition matrix $P(\tau)$ satisfies detailed balance, $\pi_i P_{ij} = \pi_j P_{ji}$, the eigenvalues λ_i are real and $|\lambda_i| \leqslant 1$. Furthermore, exploiting the definition of a Markov process (the future state of the system only depends on the current state) this equation may also be written in terms of longer time scale resolutions or lag times:

$$X((m+n)\tau) = X(m\tau)[P(\tau)]^n = X(m\tau) S(n\tau), \tag{4}$$

where $S(n\tau)=[P(\tau)]^n$. The eigenvalues $\mu_i(n)$ and eigenvectors of $S(n\tau)$ are related to those of $P(\tau)$: the eigenvectors (both right x_i and left y_i) are the same whereas the eigenvalues are related by $\mu_i(n)=\lambda_i^n$. From this relationship, a plot of $-n\tau/\ln \mu_i(n)$ is constant over all lag times for a Markovian system,[28] the constant value being the characteristic time τ_i^* of the exponential decay of the ith eigenmode:

$$\tau_i^* \equiv -n\tau/\ln \mu_i(n) = -\tau/\ln \lambda_i. \tag{5}$$

This relationship provides a means to assess whether or not a system is Markovian (or the time scales on which it is Markovian) by computing the transition matrices from the simulation data at increasing lag times and noting the range of lag times where the curves for $-n\tau/\ln \mu_i(n)$ are (approximately) constant.[32] The transition matrices in this Markovian regime may subsequently be used to iterate forward in time in order to characterize the long-time scale dynamics of the system, even when the longest simulation trajectory is shorter than the characteristic time scale on which molecular events occur.

B. Computing transition matrices from simulation data

The state space decomposition $\mathcal{S}$ for the pPA 12-mers studied in this project is described in detail in our companion paper;[23] however, we summarize for the benefit of the reader. For a pPA oligomer of length L monomers, we define the state decomposition in terms of 2^{L-2} helical states s_j. Each state is represented as an $L-2$ bit binary number $\tilde{\delta}(s_j)$ according to the sequence of discretized dihedral angles along the backbone of the chain. For a pPA L-mer, there are $L-3$ dihedral angles ϕ_i; these continuous dihedral angles are discretized according to the dihedral potential [see Eq. (1)] between adjacent phenyl rings along the backbone:

$$\delta(\phi) = [\delta_1 \cdots \delta_{L-3}] \text{ with}$$
$$\delta_i = \begin{cases} 0 & \text{if } -90^\circ < \phi_i < 90^\circ \\ 1 & \text{if } 90^\circ < \phi_i < 270^\circ . \end{cases} \tag{6}$$

The chirality, whether the helical nucleus is left or right handed, is specified by the δ_{L-2} bit. Knotted states $s_{j,k}^*$ are derived from compact helical states s_j according to the position k along the sequence where the chain crosses the ideal helix path, perturbing the helical structure.

A pPA 12-mer has $2^{L-2}=2^{10}=1024$ helical states s_j. In addition, ten knotted states were derived from the folded states s_0 and s_1 each, and additional eight knotted states were derived from states s_2, s_3, s_{512}, and s_{513} each, for a total of 52 knotted states $s_{j,k}^*$. The state space decompostion for a pPA 12-mer therefore consists of a total of 1076 states s_j and $s_{j,k}^*$. Using equivalency relationships based on chiral, knot, and dihedral angle sequence symmetries for helical states and knotted states, the 1076 states are reduced to 285 parent states p_i and $p_{i,k}^*$ by applying the symmetry functions $p_i = \mathcal{O}(s_j)$ and $p_{i,k}^* = \mathcal{O}(s_{j,k}^*)$, respectively (see companion paper[23]), which represent the final-state space decomposition $\mathcal{S}=\{p_i, p_{i,k}^*\}$ of the pPA 12-mer.

From the simulation trajectories, the process of assigning a given molecular conformation ϕ to a state s_j ideally consists of discretizing the dihedral angle sequence ϕ_i into the

$L-2$ bit binary number $\delta(\phi)$ (accounting for chiral symmetry in the δ_{L-2} bit). However, including knotted states in the state space decomposition prevents the use of this simple approach due to the fact that the knotted states have the same sequence of discretized dihedral angles as the helical states from which they are derived. In order to distinguish the knotted states from the helical states when assigning a conformation to a state s_j or $s^*_{j,k}$, a more sophisticated process is required.

A heuristic approach to assigning conformations to states is adopted to overcome this challenge of distinguishing helical states s_j from knotted states $s^*_{j,k}$. We create reference structures $r'(s_j)$ for each of the 1076 states,[33] then assign the conformation to the closest state according to a predefined metric. The root-mean-square deviation (RMSD) is the metric of choice, since it is capable of distinguishing chiral symmetry, as well as distinguishing knots from helices:

$$\mathrm{RMSD}(r,r'(s_j)) = \left[\sum_{i=1}^{L} \|r_i - r'_i(s_j)\|^2\right]^{1/2}. \tag{7}$$

RMSD is a pairwise structural parameter which measures the distance between two molecular conformations, now parametrized in terms of monomer coordinates r rather than dihedral angles ϕ, after an optimal alignment. The monomer coordinates are defined to be the mass-weighted center of mass of the phenyl ring atomic coordinates (see Fig. 1); a pPA L-mer conformation r is a $3L$-component vector. Therefore, given a reference structure $r'(s_j)$ for each of the 1076 helical and knotted states, a given molecular conformation may be assigned to the state which it is closest to in terms of RMSD:

$$s_j = A_{\mathrm{RMSD}}(r): \min_k[\mathrm{RMSD}(r,r'(s_k))] = \mathrm{RMSD}(r,r'(s_j)), \tag{8}$$

where in this case the states s_j include the knotted states $s^*_{j,k}$ to simplify the expression.

The procedure for constructing the transition matrices from the simulation trajectories is described by Swope *et al.*[29] We summarize the process which consists of computing an indicator function $\Omega^{(p_i)}$ and a correlation function $C_{ij}(\tau)$ which are used to compute the transition matrix $P_{ij}(\tau)$ for $\tau > 0$. A single data set consists of a set of M-independent simulation trajectories $r^{(m)}(t)$ for $m=1,\ldots,M$ in which the molecular conformations are parametrized in terms of monomer coordinates r and each trajectory runs for a period of time from $t=0$ to $t=T_m$ (see Fig. 3 for the time distributions). The indicator function

$$\Omega^{(p_i)}(r(t)) = \begin{cases} 1 & \text{if } \mathcal{O}[A_{\mathrm{RMSD}}(r(t))] = p_i \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

represents a succinct approach to group the microstates of molecular conformations $r(t)$ from the simulation data into macrostates p_i. The process of assigning the conformation $r(t)$ to the state s_j or $s^*_{j,k}$ through $A_{\mathrm{RMSD}}(r)$ is a crucial step that may be easily overlooked in the equation for the indicator function [Eq. (9)]; once the assignment has been made to state $s_j = A_{\mathrm{RMSD}}(r)$, the symmetry function $\mathcal{O}(s_j)$ (see companion paper) is applied to assign conformation s_j to its parent state p_i. The conformation may not be assigned directly to the parent state without first being assigned to state s_j, since RMSD is sensitive to chirality and sequence reversal symmetry.

The indicator function $\Omega^{(p_i)}$ is used to approximate the correlation function $C_{ij}(\tau)$ [see Eqs. (2)–(4) of Ref. 29]:

$$C_{ij}(\tau) = \frac{\int dr(0) e^{-\beta H(r(0))} \Omega^{(p_j)}(r(\tau)) \Omega^{(p_i)}(r(0))}{\int dr e^{-\beta H(r)}} \cong \frac{1}{M}\sum_{m=1}^{M} \frac{1}{T_m - \tau} \sum_{t=0}^{T_m-\tau} \Omega^{(p_j)}(r^{(m)}(t+\tau)) \Omega^{(p_i)}(r^{(m)}(t)). \tag{10}$$

The (unnormalized) stationary distribution $\pi_i(\tau)$ may also be computed from the M trajectories [see Eqs. (5) and (6) of Ref. 29]:

$$\pi_i(\tau) = \frac{\int dr e^{-\beta H(r)} \Omega^{(p_i)}(r)}{\int dr e^{-\beta H(r)}} \cong \frac{1}{M}\sum_{m=1}^{M} \frac{1}{T_m - \tau} \sum_{t=0}^{T_m-\tau} \Omega^{(p_i)}(r^{(m)}(t)). \tag{11}$$

The transition function $P_{ij}(\tau)$ is finally calculated using the following Bayes' relationship:[29]

$$P_{ij}(\tau) = \frac{C_{ij}(\tau)}{\pi_i(\tau)}. \tag{12}$$

When computing a transition-probability matrix $P = P_{ij}(\tau)$ directly from the data, invariably detailed balance will not be satisfied due to random noise in the simulation data, which leads to imaginary components in the eigenvalues and associated eigenvectors. We propose a procedure, outlined below, to impose detailed balance on the transition matrix P through a symmetrization. This procedure uses a relationship between P and P^{sym} which is easily invertible, i.e., $P^{\mathrm{sym}}_{ij} = \pi_i P_{ij} \Leftrightarrow P_{ij} = (P^{\mathrm{sym}}_{ij}/\pi_i)$, where π_i is the *normalized* stationary distribution. The stationary distribution π (properly normalized) may be computed from either P or P^{sym}, as shown in steps 1 and 3, respectively, of the following procedure:

$$P, \pi \xrightarrow{1} P^{\mathrm{sym}} \xrightarrow{2} \hat{P}^{\mathrm{sym}} \xrightarrow{3} \hat{P}, \hat{\pi}. \tag{13}$$

(1) Diagonalize P to obtain π (first left eigenvector); then, $P^{\mathrm{sym}}_{ij} = \pi_i P_{ij}$.
(2) Impose symmetry to obtain $\hat{P}^{\mathrm{sym}}$ using the equation: $\hat{P}^{\mathrm{sym}} = \frac{1}{2}[P^{\mathrm{sym}} + (P^{\mathrm{sym}})^T]$.
(3) Calculate the updated stationary distribution $\hat{\pi}$ from the symmetrized matrix: $\hat{\pi}_i = \Sigma_l \hat{P}^{\mathrm{sym}}_{il}$; then use the updated stationary distribution $\hat{\pi}$ to invert from the symmetrized matrix $\hat{P}^{\mathrm{sym}}$ to the stochastic transition matrix $\hat{P}$ which now satisfies detailed balance: $\hat{P}_{ij} = \hat{P}^{\mathrm{sym}}_{ij}/\hat{\pi}_i$.

Performing the prescribed transformation in step 1 should in theory produce a symmetric matrix P^{sym} when P satisfies detailed balance. However, since P does not satisfy detailed balance due to random noise, P^{sym} will not be ex-

actly symmetric. If, on the off-chance P does already satisfy detailed balance before this symmetrization procedure, of course, this procedure may be skipped.

Also of note in step 1, the eigenvalues and eigenvectors from the diagonalization of P will have imaginary components since P does not satisfy detailed balance. This should not be a major concern since the stationary distribution π obtained from the first left eigenvector is always real, and the other imaginary eigenvectors are not needed to perform the symmetrization. Therefore, P^{sym} will be a real matrix as well.

In step 3 it is necessary to use the updated normalized stationary distribution $\hat{\pi}$ computed from the symmetrized matrix $\hat{P}^{\mathrm{sym}}$ in order to obtain a proper stochastic matrix $\hat{P}$ with rows that sum to one. Using the original stationary distribution π will not give back a stochastic matrix satisfying detailed balance due to the symmetrization in step 2.

IV. RESULTS AND DISCUSSION

The dynamic properties of the pPA 12-mer are characterized at both 300 and 330 K for each of the four solvents; thus, eight data sets are simultaneously analyzed and characterized using the Markov state models described in Sec. III. Each data set consists of $M=1600$ independent trajectories initiated evenly between eight starting structures (200 trajectories per starting structure, differing only in initial velocities). From these multiple independent trajectories, the transition matrix[34] $P_{ij}(n\tau_{\mathrm{samp}})$ over the set of 285 parent states p_i and $p^*_{i,k}$ constructed from the state space decomposition $\mathcal{S}=\{p_i,p^*_{i,k}\}$, was computed for the range of lag times $n\tau_{\mathrm{samp}}=0.5-20.0$ ns in increments of $\tau_{\mathrm{samp}}=0.5$ ns. We enforce detailed balance on each transition matrix $P_{ij}(n\tau_{\mathrm{samp}})$ using the symmetrization procedure described in Eq. (13) to address the noise in the data as a precursor to the subsequent kinetic analysis, which highlights the primary role that the solvent quality plays in the folding kinetics of the pPA 12-mer.

A. Demonstration of Markovian behavior

Now that the transition matrices $P_{ij}(n\tau_{\mathrm{samp}})$ satisfy detailed balance, we diagonalize $P_{ij}(n\tau_{\mathrm{samp}})$ to express the dynamics in terms of eigenmodes which are orthonormal with respect to the weighted inner product $\langle x_i,x_j\rangle_\pi$, as outlined in Sec. III. Figure 4 illustrates the nearly exponential decay of the first 20 eigenvalues $\mu_i(n)$ of $P_{ij}(n\tau_{\mathrm{samp}})$ as a function of the lag time $n\tau_{\mathrm{samp}}$ for each of the four solvent systems: (i) acetonitrile, (ii) chloroform, (iii) methanol, and (iv) water; at both temperatures: (a) 300 K and (b) 330 K. The first eigenvalue $\mu_1(n)\equiv 1$ is identically equal to 1 over all lag times as it corresponds to the Perron root. We observe that the second eigenvalue $\mu_2(n)$ has only decayed a minimal amount after 20 ns in all eight data sets (nonaqueous solvents at the higher temperature of 330 K have the highest levels of decay). This suggests that the slowest time scales in the pPA 12-mer folding mechanism are much longer than the average trajectory which is ~20 ns in all eight data sets (see Fig. 3).

In order to determine the time scales at which a system exhibits Markovian behavior, the curves for the characteristic decay time $\tau^*_i=-n\tau_{\mathrm{samp}}/\ln\mu_i(n)$ of the same 20 eigenvalues of $P_{ij}(n\tau_{\mathrm{samp}})$ are plotted in Fig. 5 as a function of the lag time $n\tau_{\mathrm{samp}}$. For a given system to have Markovian properties, the curves for these characteristic times must be constant over an extended range of lag times, which will hereafter be referred to as the Markovian regime of lag times. This is a necessary but not sufficient condition for Markovian behavior.

Inspection of the characteristic time curves in Fig. 5 suggests that acetonitrile (i) exhibits Markovian behavior over a broad range of lag times ranging from ~5 to 15 ns at both temperatures of 300 and 330 K. This observed Markovian behavior is a key component for further investigations of the folding dynamics of the pPA 12-mer in acetonitrile at 300 K, to be given shortly hereafter in Sec. IV C. Moreover, both chloroform (ii) and methanol (iii) appear to be Markovian at the higher temperature of 330 K over the range of lag times from ~8 to 20 ns. On the other hand, it is difficult to discern whether or not chloroform and methanol are Markovian at the lower temperature of 300 K. The characteristic time curves nearly level off in the range ~10–15 ns, which may or may not be a sufficient demonstration of Markovian behavior within that time range. It is easy to conclude that water (iv) is not Markovian at either temperature, as the characteristic decay time curves for the eigenmodes continue to increase nearly linearly over all lag times between 0 and 20 ns, rather than level off to any near constant value. In general, the dynamics of pPA 12-mers are more likely to be Markovian at higher temperatures suggesting that sampling of conformation space is more efficient at higher temperatures. An increase in the amount of thermal energy assists the pPA 12-mer in making transitions between macrostates.

Two features are common to each of the characteristic decay time curves in Fig. 5. First, the curves increase rapidly at short lag times $n\tau_{\mathrm{samp}}=0-5$ ns before they level off. Second, at long lag times $n\tau_{\mathrm{samp}}>15$ ns, the curves increase once again. Swope *et al.*[29] suggest that the lack of Markovian behavior at short lag times may be attributed to the system retaining a memory of its location in conformation space at short times. The system needs adequate time to decorrelate after entering a free-energy basin. The time scales on which the system exhibits Markovian behavior are very dependent upon this decorrelation time. The gentle rise in the characteristic time curves at higher lag times ($n\tau_{\mathrm{samp}}>15$ ns) has also been noted by Swope *et al.*[29]; due to the limited lengths of the trajectories, inadequate sampling at these long time scales purportedly leads to non-Markovian behavior.

The characteristic time curves of pPA 12-mers in explicit water have very interesting non-Markovian features (see Fig. 5). Why does not the pPA 12-mer exhibit Markovian behavior in water? There are several possibilities. First of all, the solvent water may be directly involved in the pPA dynamics. Since we do not account for the water degrees of freedom in the state space decomposition of the oligomer, we may have removed information needed to observe Markovian behavior. Another possibility may be that the oligomer has not adequately sampled conformation space or the transitions between macrostates. Closely related to this sampling issue, the dynamics may be greatly hindered due to the very strong aggregation potential in aqueous solutions, and hence may be Markovian only at lag times greater than 20 ns.

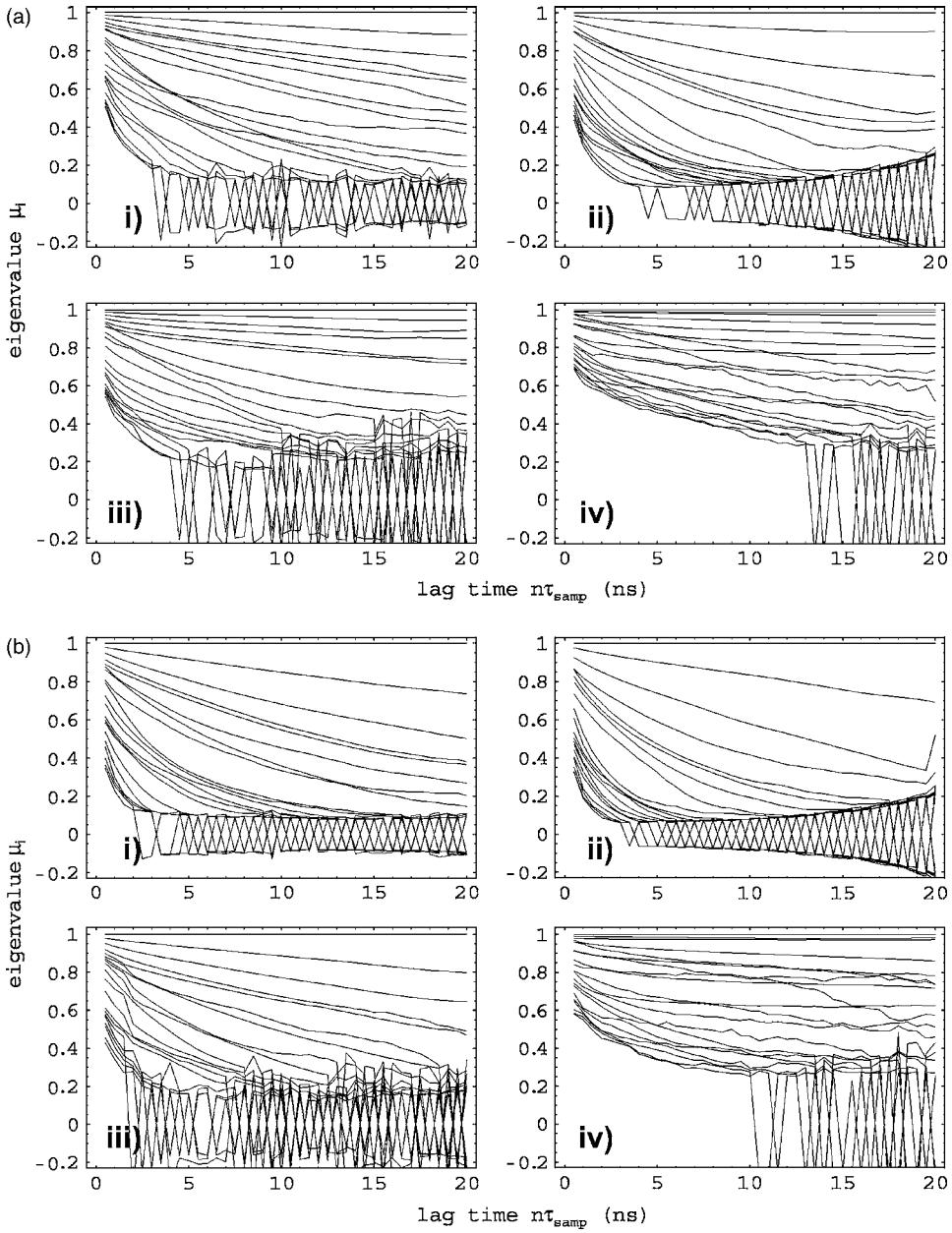

FIG. 4. Eigenvalue spectra for the first 20 eigenvalues $\mu_i(n)$ of the transition matrices $P_{ij}(n\tau_{\text{samp}})$ as a function of the lag time $n\tau_{\text{samp}}$ at different temperatures: (a) 300 K and (b) 330 K; and in different explicit solvents: (i) acetonitrile, (ii) chloroform, (iii) methanol, and (iv) water. The exponential decay of the eigenvalues is apparent, except for the Perron root $\mu_1(n)=1$ which is constant for all $n\tau_{\text{samp}}$. Due to the fact that the eigenvalues are not strictly positive, but can take on negative values, the eigenvalues in the lower portion of the spectra oscillate dramatically.

B. Comparison of folding times in organic solvents

How does the solvent quality inherent to a given solvent affect the pPA 12-mer folding kinetics? Moreover, how does the temperature affect the kinetics? These are important questions that will be addressed in this section by computing the folding time, expressed by the mean first passage time (MFPT), directly from the transition matrices $P_{ij}(n\tau_{\text{samp}})$. The mean first passage time[30] for state p_i is a kinetic measure of the average time taken to reach the folded state from state p_i before unfolding. From the transition matrix at a given lag time $P_{ij}(n\tau_{\text{samp}})$, the MFPT may be computed self-consistently from the following system of equations:

$$\text{MFPT}(p_i,n\tau_{\text{samp}}) = \sum_j P_{ij}(n\tau_{\text{samp}}) \times [n\tau_{\text{samp}} + \text{MFPT}(p_j,n\tau_{\text{samp}})]. \quad (14)$$

A single boundary condition is defined so that the mean first passage time from the folded state p_0 to itself is zero, i.e., $\text{MFPT}(p_0)=0$. Initially, the MFPTs for all states are set to infinity, $\text{MFPT}(p_i)=\infty$, except for the folded state p_0 which again is initially set to zero. This system of equations is solved iteratively until a self-consistent solution is obtained for each of the 285 parent states p_i, at a prespecified tolerance.

The mean first passage time may also be used as an indicator of Markovian behavior. Just as the characteristic decay times τ_i^* must be constant in the Markovian regime, the mean first passage times must also be constant in the Markovian regime of lag times. Again, this is a necessary but not sufficient condition for Markovian behavior.

The majority of the states p_i are unstructured. $D(p_i)$ is a structural parameter which describes the size of the helical nucleus or the number of consecutive helical monomers in state p_i. Since it is explained in greater detail in the companion paper,[23] let it suffice to declare that any state p_i with $D(p_i)\leqslant 3$ is unstructured and is a member of the unfolded

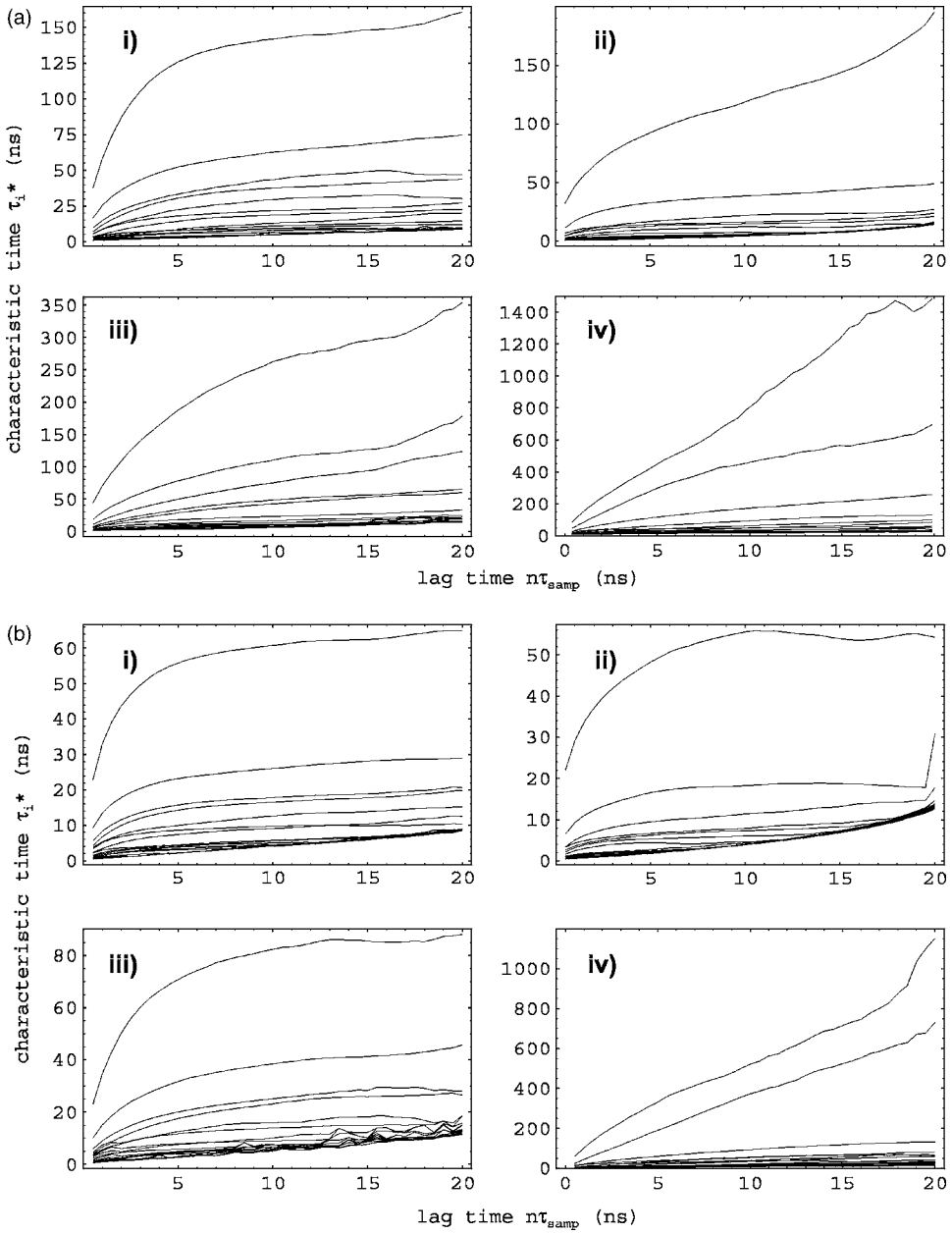

FIG. 5. Characteristic time scales τ_i^* for the first 20 eigenmodes as a function of the lag time $n\tau_{\text{samp}}$ at different temperatures: (a) 300 K and (b) 330 K; and in different explicit solvents: (i) acetonitrile, (ii) chloroform, (iii) methanol, and (iv) water. The range of lag times at which the curves are level or constant reveal the time scales in which the given system is Markovian.

ensemble U, i.e., $U=\{p_i|D(p_i)\leqslant 3\}$. To determine the typical time to fold from the unfolded state, the average of all mean first passage times for the set of unfolded states U is computed:

$$\text{MFPT}(U,n\tau_{\text{samp}})=\frac{1}{N_U}\sum_{p_i\in U}\text{MFPT}(p_i,n\tau_{\text{samp}}), \tag{15}$$

where N_U is the number of states in U. This technique is used to compute the mean first passage time from the unfolded state U to the folded state p_0 for the transition matrices $P_{ij}(n\tau_{\text{samp}})$ over the range of lag times $n\tau_{\text{samp}}=0.5-20.0$ ns. This measure of the folding time, along with the standard deviation among the unfolded states, is shown in Fig. 6(a) for each of the eight solvent/temperature simulation data sets. A complementary view of the temperature dependence and the solvent quality dependence on the folding time scales for pPA 12-mers is illustrated in Fig. 6(b). This bar chart is taken from a time slice at a lag time of 10 ns from the data shown in Fig. 6(a). The solvent system which promotes the fastest folding at a lag time of 10 ns (which was in the Markovian regime for most data sets) was acetonitrile at 330 K in which the characteristic folding time for the pPA 12-mer was ~100 ns.

Several interesting solvent properties may be noted from the folding data shown in Fig. 6. Most importantly, the solvent quality is the primary determinant of the folding kinetics of pPA 12-mers. At a lag time of 10 ns, the range of folding times differ by an order of magnitude, purely due to solvent quality. Temperature also affects the kinetics, particularly for methanol, although not as strongly as the solvent quality. The pPA 12-mers take longer to fold in both the extreme solvent quality regimes, as may be seen from the folding times in chloroform and water, which represent the good and poor solvent regimes, respectively. However, in the theta solvent regime which is characterized by optimal fold-

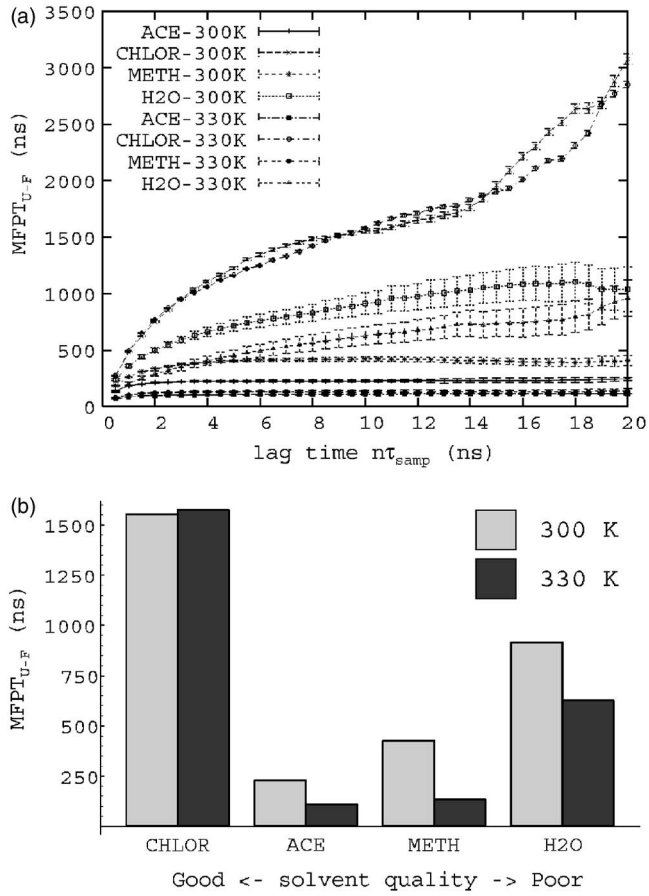

FIG. 6. (a) Mean first passage times from the unfolded state $\{p_i|D(p_i) \leqslant 3\}$ to the folded state p_0 calculated from the Markov models for the eight solvent/temperature systems as a function of lag time $n\tau_{\text{samp}}$. (b) Time slice at $n\tau_{\text{samp}}=10$ ns. The qualitative features of these folding times agree with experiments which suggest chloroform is a good solvent, acetonitrile and methanol (330 K) are in the theta solvent regime, and water is a poor solvent.

ing conditions, the folding times are minimized for both acetonitrile and methanol (especially at higher temperatures of 330 K).

Chloroform has a very high folding time as expected because it is a denaturing solvent, therefore pPA oligomers take a long time to fold if they fold at all in chloroform. The folding time curves for chloroform at both temperatures appear close to leveling off around 10–15-ns lag times. Just as it is difficult to discern the Markovian behavior of pPA 12-mers in chloroform at the lower temperature of 300 K from the characteristic time curves [see Fig. 5(aii)], it is also difficult to discern the Markovian behavior in chloroform using MFPT. Since Markovian behavior is not readily obvious in chloroform, the estimates for the folding times in chloroform shown in Fig. 6 may be inaccurate. However, the trend in the folding times relative to the other solvents is preserved despite any inaccuracies in the estimates, and it is clear that chloroform is a denaturing solvent due to the increased folding time relative to acetonitrile.

Water, on the other hand, has an increased folding time due to the strong aggregation potential as a consequence of the strong attraction of the chain with itself. Therefore, it gets trapped in intermediate states and takes longer to reach the folded state on average relative to a theta solvent such as acetonitrile. Water has already been shown according to the characteristic time curves in Fig. 5 not to be Markovian at any lag time between 0 and 20 ns. Therefore, it is not surprising that the standard deviations for the MFPT$(U, n\tau_{\text{samp}})$ in water are relatively large. In addition, the mean first passage times for water (shown in Fig. 6) are at best rough estimates of the correct folding times in aqueous solutions, but are nonetheless helpful to observe the general trend in the folding times.

Both acetonitrile and methanol have constant curves for the folding times over a broad range of lag times as measured by MFPT [see Fig. 6(a)]. This further supports the evidence gained from the curves for the characteristic decay times shown in Fig. 5 that acetonitrile and methanol have Markovian properties over a broad range of lag times. Not only are the dynamics Markovian, but the folding times appear to be minimized in these solvents [see Fig. 6(b)], especially at the higher temperature of 330 K, supporting the initial assumption that these solvents are in the theta solvent regime, particulary acetonitrile which has low MFPT at both temperatures. Methanol appears to be in the theta solvent regime at the higher temperature; however, the solvent quality for methanol decreases as the temperature decreases as evidenced by the significantly higher mean first passage time at 300 K.

It is interesting to note that more than one measure is needed to draw a convincing argument for Markovian behavior, a natural consequence of the level curves for the characteristic decay times and the MFPT being necessary but not sufficient conditions for the Markovian behavior. This is indeed the case with chloroform at 330 K. The curves for the characteristic decay times appear to exhibit Markovian behavior [see Fig. 5(bii)], while the MFPT curves shown in Fig. 6(a) only provide a tenuous argument for Markovian behavior.

How do the characteristic decay time curves in Fig. 5 compare physically to the folding times measured by the MFPT in Fig. 6? The fundamental difference between the characteristic decay times τ_i^* and the MFPT depends on the final state of the dynamic system. For instance, by definition, the MFPT measures the average time to reach the folded state from state p_i before unfolding. The final state of the system is clearly defined to be the folded state and is the same in all cases. However, the characteristic decay times τ_i^* measure the relaxation times for the individual eigenmodes to reach the *stationary distribution*, which varies from one solvent system to the next. As an example, the stationary distribution of pPA 12-mer states in acetonitrile is dominated by the folded state and other highly compact states. Chloroform being a denaturing solvent, on the other hand, has a stationary distribution of oligomer states consisting primarily of unstructured, unfolded states. Therefore, the final state of the system is different from one system to another when measuring the characteristic times from the eigenmodes, and one needs to be very careful when making a direct physical comparison to the computed MFPT.

In the case of acetonitrile and methanol, the comparison between the MFPT and the characteristic times τ_i^* of the

TABLE I. Interaction enthalpies between the pPA 12-mer and the respective solvent. The solvent quality increases (from poor to good solvent quality) the stronger the chain interacts with the solvent. The expected ordering of the solvents in terms of solvent quality is observed, except for acetonitrile, which is expected to have a solvent quality intermediate between chloroform and methanol.

$\Delta H(T)$ (kJ/mol)	300 K	330 K
$\Delta H_{\mathrm{CHLOR}}(T)$	−610±190	−600±190
$\Delta H_{\mathrm{METH}}(T)$	−570±120	−500±120
$\Delta H_{\mathrm{H2O}}(T)$	−495±95	−440±93
$\Delta H_{\mathrm{ACE}}(T)$	−430±120	−370±120

eigenmodes is valid, since they both essentially measure the same process of folding to the folded state. This may be verified in the case of acetonitrile by looking at the folding time measured by MFPT in Fig. 6 (~100 ns at 330 K) and comparing directly with the sum of the characteristic times of the eigenmodes shown in Fig. 5 (slowest eigenmode ~60 ns at 330 K). It is clear, however, that the characteristic decay times in chloroform do not measure the folding process. The MFPT of the pPA 12-mer in chloroform at both temperatures is well over 1000 ns, while the characteristic decay times essentially measure the unfolding time in chloroform, the slowest eigenmode of which is ~50 ns at 330 K.

It is desirable to be able to quantitatively compare the solvent quality of the four solvents currently being studied. Unfortunately, the abscissa in Fig. 6(b) does not have a distance associated with it. Rather, it is strictly the expected ordering of solvent quality gained from experimental results[18] that drives the ordering of the solvent systems displayed in the chart. Ideally, the solvation free energy[35,36] or the Flory-χ parameter[37] would give a quantitative measure of solvent quality sufficient to generate a distance along the x-axis for each of the eight solvent/temperature systems. However, it is very difficult, if not impossible, to extract solvation free energies from the simulation data according to the methods used in this study.

The best measure of solvent quality we are able to calculate from the simulation data is the interaction enthalpy between the oligomer chain and the given solvent: $\Delta H = \langle H \rangle_{MS} - \langle H \rangle_{MM}$. The total interaction enthalpy is defined to be the sum of the van der Waals and Coulomb nonbonded enthalpies $H = H_{\mathrm{vdw}} + H_{\mathrm{coul}}$. The first term $\langle H \rangle_{MS}$ is the absolute enthalpy of the monomer (M)-solvent (S) interactions; whereas, the second term $\langle H \rangle_{MM}$ is due to the absolute enthalpy of the intramolecular monomer (M)-monomer (M) interactions. The relative interaction enthalpy ΔH measures the strength of attraction between the oligomer and a given solvent. Therefore, a quantitative comparison between the relative interaction enthalpy ΔH of the pPA 12-mer in several solvent systems serves as a direct measure of solvent quality.

The relative interaction enthalpy ΔH for each solvent/temperature system is tabulated in Table I. Chloroform has the strongest interaction enthalpy at both temperatures as expected; the stronger the interaction, the stronger the oligomer attracts the solvent which then screens the oligomer from distant portions of the chain. Thus, the weak intrachain interactions in a good solvent environment cause the oligomer to remain unstructured and extended. Methanol and water both have the expected ordering with respect to solvent quality relative to chloroform and to each other. In this case, the interaction enthalpy between the oligomer (M) and the solvent (S) is weaker than in chloroform; therefore, the oligomer attracts itself more strongly in the poor solvents, water and methanol, causing the chain to collapse.

Unfortunately, the interaction enthalpy $\Delta H_{\mathrm{ACE}}(T)$ between the pPA oligomer and acetonitrile is calculated to be even less than that of water $\Delta H_{\mathrm{H_2O}}(T)$ (see Table I), which would suggest that acetonitrile is a very poor solvent, inconsistent with the folding times shown in Fig. 6. The most glaring deficiency of the interaction enthalpy ΔH as a measure of solvent quality is the failure to obtain the expected ordering of acetonitrile relative to the other three solvents. Nevertheless, this unfortunate result is not surprising given that the primary force driving pPA oligomers to fold is solvophobicity, which is purely an entropic effect. Therefore, the interaction enthalpies may be expected to fail to give the correct ordering, and are therefore a poor quantitative measure of solvent quality. More sophisticated free-energy methods are needed in order to quantitatively compare the solvent quality of these solvents with respect to pPA oligomers.

As a consequence of this failure to obtain quantitative measurements of solvent quality, we chose to order the solvents in Fig. 6(b) in a qualitative sense according to the expected ordering from experimental data. Despite being limited to a qualitative assessment of the solvent quality, we are still able to draw some very important and interesting conclusions from the folding kinetics measured by MFPT. In particular, acetonitrile is the solvent which offers the greatest potential for rapid self-assembly into functional nanostructures (within the set of solvents and temperatures considered in this study). Accordingly, we further investigate the folding dynamics of the pPA 12-mer in acetonitrile in Sec. IV C.

C. pPA 12-mer folding kinetics in acetonitrile at 300 K

As a consequence of running multiple short independent trajectories to sample conformation space, not a single folding event is observed in which the oligomer is initially extended and ultimately folds in the same trajectory. In light of this fact, it is reasonable to question the validity of the folding times (MFPT) computed from the transition matrices. In other words, how well do the Markov state models for the dynamics of pPA 12-mers reproduce the exact dynamics of the all-atom computational model? In order to address this question, we compute an additional data set of folding trajectories which are initiated from an extended conformation and allowed to diffuse unbiased toward the folded state, running as long as necessary to observe molecular folding events. From these folding trajectories, an estimate for the short-time dynamics of the computational model of the pPA 12-mers may be obtained. Then, a direct comparison between the dynamics of these folding trajectories and the dynamics computed from the Markov state models derived

from the multiple independent trajectories will serve as a means to quantitatively validate the predictive capabilities of the Markov state models.

For this additional folding data set, we used the same procedure and parameters outlined in Sec. II to solvate and equilibrate an extended conformation of a pPA 12-mer in each of the four solvents at 300 K (we did not compute folding trajectories at 330 K). These folding trajectories were run on a local LINUX cluster of dual Xeon processor machines using a parallel version of the GROMACS3.1.4 molecular-dynamics program running on two processors on a single node to decrease the wall clock time. For each solvent system 25 trajectories were generated, differing only in random number seed used to compute the initial velocities. The longest trajectory ran for 90 ns; however, not all trajectories ran for the same length of simulation time.

We did not observe any folding events in chloroform, which is not surprising since it is a denaturing solvent, therefore the folding rate is very slow. Also, we did not observe any folding events in water, which is also not surprising. Due to the strong intramolecular attraction of the chain to itself in water, it gets trapped in intermediate conformations and ultimately cannot fold within the time scales accessible to these simulations. We did, however, observe one folding event in methanol. While this is a fantastic observation, one data point does not provide enough information to extract any meaningful conclusions relating to the folding kinetics of pPA oligomers in methanol. Acetonitrile, on the other hand, produced eight folding events! This is a sufficient amount of information to make a direct comparison against the folding dynamics predicted from the Markov state models of the pPA 12-mer in acetonitrile at 300 K.

The fraction of the trajectories which have folded as a function of time is shown in Fig. 7. Utilizing the fact that folding is a diffusive process with exponential folding time distributions, we expect to observe some fast folding events at very short-time periods. In the case of a two-state system, we would expect a linear increase in the fraction folded at very short-time periods, a fit to a straight line providing an estimate for the folding rate of this two-state folding system.[38] The dashed line in Fig. 7 shows the fraction of the folding trajectories initiated from an extended state which reach the folded state by time t.

We see that the pPA 12-mers do not follow the kinetics of a two-state folder since there is an observable curvature corresponding to a quadratic rise of the fraction folded at short times ($t<100$ ns). This is, however, consistent with the experimental kinetic measurements which show significant nonexponential folding time distributions at 300 K,[8] and is also consistent with previous computer simulations using a simple generic implicit solvent model in which a Taylor-series expansion around $t=0$ also revealed a quadratic rise in the fraction folded at very short times.[2,3] The observable curvature in the dashed curve for the fraction folded in Fig. 7 is also consistent with the eigenvalue spectra and the characteristic decay time curves shown in Figs. 4 and 5, respectively, for acetonitrile at 300 K. If the folding kinetics were strictly two state, the second eigenvalue $\mu_2(n)$ would dominate the eigenvalues from the lower portion of the eigenvalue spectrum and there would be a single curve for the characteristic decay time associated with the two-state process. However, there are several eigenvalues which contribute significantly to the overall folding dynamics, lending support to the idea that the pPA 12-mer folding in acetonitrile is more complex than a simple two-state mechanism.

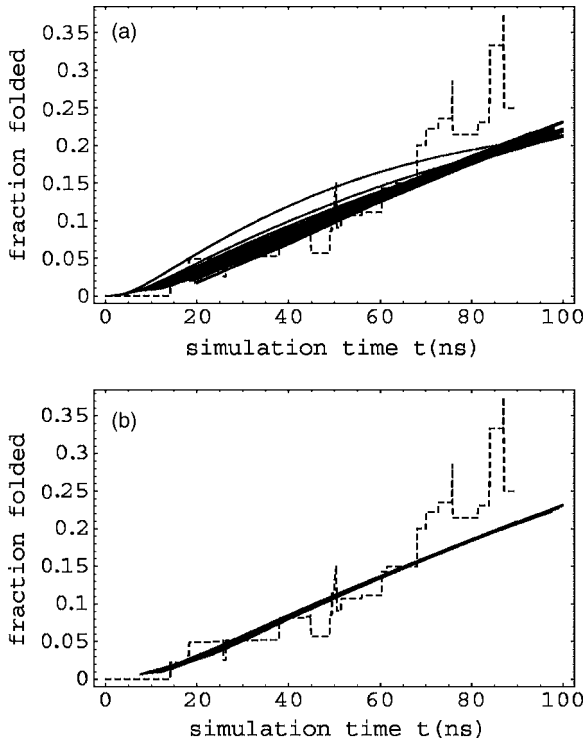

FIG. 7. Fraction of trajectories which have folded as a function of simulation time t. The dashed line is computed from the folding trajectories which were initiated in an extended conformation and allowed to diffuse unbiased to the folded state; therefore, this curve represents an estimation of the full dynamics of the molecular mechanics model of a pPA 12-mer in explicit acetonitrile at 300 K. The solid curves show the accumulation of folded molecules from an initial distribution of purely unfolded states $U=\{p_i|D(p_i)\leqslant 3\}$ computed from the Markov models over a range of lag times: (a) 0.5–20.0 ns and (b) 7.5–15.0 ns (Markovian regime).

The solid curves shown in Fig. 7 are the predicted curves for the fraction folded computed from the Markov state models. From the transition-probability matrices computed from the simulation data at a given lag time $P_{ij}(n\tau_{\text{samp}})$, we iterate forward in time [see Eqs. (3) and (4)] from an initial probability distribution consisting purely of unfolded states[39] and monitor the concentration of the folded state as a function of time t. We note from Fig. 7(a) that the predicted curves are not consistent over the full range of lag times, 0–20 ns in 0.5-ns increments. This is expected since the dynamics are not Markovian over this whole range of lag times. However, restricting the lag times to a narrower range corresponding to the Markovian regime, i.e., 7.5–15.0 ns, reveals that the short-time dynamics predicted by the Markov state models are very consistent, both amongst themselves and also in comparison with the full dynamics for the computational model computed from the folding trajectories. The divergence between the Markov state models and the full dynamics of the computational model above 70 ns may be attrib-

uted to fewer folding trajectories reaching 70 ns; therefore, less data above 70 ns cause significant noise in the dashed curve for the full dynamics.

Now that the Markov state models have been shown to adequately reproduce the dynamics of the computational model for the folding of pPA 12-mers in acetonitrile at 300 K, we can use the transition-probability matrices in the Markovian regime to characterize the kinetic properties of this model system with more confidence. From the Markov model validation process illustrated in Fig. 7(b), and given the satisfactory agreement between the Markov models and the folding data, we conclude that the folding time for the computational model of pPA 12-mers in acetonitrile at 300 K is ~200 ns in terms of MFPT, as computed previously (see Fig. 6).

We are not aware of any reports in the literature of a folding time measured for pPA 12-mers in acetonitrile. However, Yang *et al.*[8] have studied the folding kinetics of a pPA 12-mer in a 1:1 (v/v) tetrahydrofuran (THF)/methanol solution. Since it was not a pure methanol solution, we are not capable of making a direct quantitative comparison to our simulations, as desired. However, a qualitative comparison between the folding time of the computational model of the pPA 12-mer in pure methanol and the experimentally measured folding time in the 1:1 THF/methanol solution is possible. How the THF affects the folding times, whether it speeds up or slows down the dynamics, is unknown. Assuming THF speeds up the pPA dynamics, the results of our simulations in pure methanol will be in good agreement with the experimental folding time measured in the THF/methanol solution, as we now demonstrate.

The experimental folding time of pPA 12-mers in 1:1 THF/methanol solution at 300 K was measured to be ~160 ns, assuming a fit to a two-state model.[8] The folding time, represented as the MFPT, of the computational model for pPA 12-mers in pure methanol at 300 K was calculated to be ~400 ns. Given the experimental uncertainties introduced by fitting the data to a two-state model when the kinetics are multistate and assuming that adding THF to the computational model of the pPA 12-mer in pure methanol would speed up the dynamics, the calculated folding time for the pPA 12-mers in pure methanol, ~400 ns, is in reasonably good agreement with the experimentally measured folding time of ~160 ns in the 1:1 (v/v) THF/methanol solution. Even more encouraging is the fact that the folding times agree reasonably well without having to scale the times, as was needed when comparing the experimental folding time with the folding time for the implicit solvent model to correct for a low friction coefficient in the implicit solvent simulations.[3]

V. SUMMARY

The solvent environment plays a fundamental role in the folding kinetics of self-assembling molecular systems. A major challenge in the field of molecular design entails understanding the properties of the solvent environment which would promote optimal conditions for dynamic processes to occur in a functional, designed system.

Toward this goal, we have studied the solvation contributions to the folding properties of a potential self-assembling functional nanomaterial, polyphenylacetylene (pPA) 12-mer. An ensemble of short, independent molecular-dynamics trajectories in full atomistic detail for this pPA 12-mer in explicit solvent were calculated in each of the solvents, acetonitrile, chloroform, methanol, and water, at two temperatures, 300 and 330 K. Of the four solvents investigated, acetonitrile was found to promote optimal folding conditions at both temperatures, demonstrated by a minimal folding time, measured by the mean first passage time (MFPT). Based on the information gained from the current study, future biomimetic applications of pPA oligomers will ideally be mediated in acetonitrile solutions.

ACKNOWLEDGMENTS

We greatly appreciate the many users of Folding@Home who voluntarily donated time on their PCs to support this project. This work was also supported by the Center on Polymer Interfaces and Macromolecular Assemblies (CPIMA) as part of the NSF Materials Science and Engineering Center program under Grant No. DMR 9808677.

[1] E. Drexler and R. Smalley, Chem. Eng. News **81**, 37 (2003).
[2] S. Elmer and V. Pande, J. Phys. Chem. B **105**, 482 (2001).
[3] S. P. Elmer and V. S. Pande, J. Chem. Phys. **121**, 12760 (2004).
[4] S. P. Elmer and V. S. Pande, J. Chem. Phys. **122**, 124908 (2005).
[5] J. Nelson, J. Saven, J. Moore, and P. Wolynes, Science **277**, 1793 (1997).
[6] R. Prince, J. Saven, P. Wolynes, and J. Moore, J. Am. Chem. Soc. **121**, 3114 (1999).
[7] O. Lee and J. Saven, J. Phys. Chem. B **108**, 11988 (2004).
[8] W. Y. Yang, R. B. Prince, J. Sabelko, J. S. Moore, and M. Gruebele, J. Am. Chem. Soc. **122**, 3248 (2000).
[9] R. Prince, S. Barnes, and J. Moore, J. Am. Chem. Soc. **122**, 2758 (2000).
[10] M. S. Gin and J. S. Moore, Org. Lett. **2**, 135 (2000).
[11] M. J. Mio, R. B. Prince, J. S. Moore, C. Kuebel, and D. C. Martin, J. Am. Chem. Soc. **122**, 6134 (2000).
[12] R. B. Prince, L. Brunsveld, E. W. Meijer, and J. S. Moore, Angew. Chem., Int. Ed. **39**, 228 (2000).
[13] A. Tanatani, M. J. Mio, and J. S. Moore, J. Am. Chem. Soc. **123**, 1792 (2001).
[14] K. Oh, K. S. Jeong, and J. S. Moore, Nature (London) **414**, 889 (2001).
[15] A. Tanatani, T. S. Hughes, and J. S. Moore, Angew. Chem., Int. Ed. **41**, 325 (2001).
[16] D. H. Zhao and J. S. Moore, J. Am. Chem. Soc. **124**, 9996 (2002).
[17] D. J. Hill, M. J. Mio, R. B. Prince, T. S. Hughes, and J. S. Moore, Chem. Rev. (Washington, D.C.) **101**, 3893 (2001).
[18] D. Hill and J. Moore, Proc. Natl. Acad. Sci. U.S.A. **99**, 5053 (2002).
[19] L. Brunsveld, E. W. Meijer, R. B. Prince, and J. S. Moore, J. Am. Chem. Soc. **123**, 7978 (2001).
[20] M. T. Stone and J. S. Moore, Org. Lett. **6**, 469 (2004).
[21] J. S. Moore, Acc. Chem. Res. **30**, 402 (1997).
[22] E. J. Sorin, B. J. Nakatani, Y. M. Rhee, G. Jayachandran, V. Vishal, and V. S. Pande, J. Mol. Biol. **337**, 789 (2004).
[23] S. Elmer, S. Park, and V. Pande, J. Chem. Phys. **123**, 114903 (2005).
[24] K. Okuyama, T. Hasegawa, M. Ito, and N. Mikami, J. Phys. Chem. **88**, 1711 (1984).
[25] CRC Handbook of Chemistry and Physics (CRC, Cleveland, Ohio, 1977), p. v.
[26] N. G. V. Kampen, *Stochastic Processes in Physics and Chemistry*, rev. and enl. ed. (North-Holland, Amsterdam, 1992).
[27] I. Oppenheim, K. E. Shuler, and G. H. Weiss, *Stochastic Processes in Chemical Physics: The Master Equation* (MIT, Cambridge, Mass., 1977).
[28] W. C. Swope, J. W. Pitera, and F. Suits, J. Phys. Chem. B **108**, 6571 (2004).
[29] W. C. Swope, J. W. Pitera, F. Suits *et al.*, J. Phys. Chem. B **108**, 6582 (2004).

[30] N. Singhal, C. D. Snow, and V. S. Pande, J. Chem. Phys. **121**, 415 (2004).
[31] P. Deuflhard, W. Huisinga, A. Fischer, and C. Schutte, Linear Algebr. Appl. **315**, 39 (2000).
[32] The level curves obtained from Eq. (5) are a necessary but not sufficient condition to prove Markov behavior; therefore, this procedure only gives a qualitative demonstration of a system being Markovian.
[33] Idealized reference structures (see companion paper) were constructed for each of the 1076 states. First, the coordinates of a given monomer in the chain were defined to be the centroid of the phenyl ring. The dihedral angles between adjacent monomers were assigned ideal angles to generate left- and right-handed helices: all *cis*-dihedral angles were +5° for the left-handed helices, and −5° for the right-handed helices; all *trans*-dihedral angles were −175° for the left-handed partially helical structures, and +175° for the right-handed partially helical structures. The bond lengths and bond angles were all given the ideal values corresponding to the average values taken from the data: 6.811-nm bond lengths, and 120° for bond angles. The knotted reference structures were constructed in a similar manner except the dihedral angles were perturbed in such a way that the chain crossed back over the ideal helix path at the appropriate position k along the helix path.
[34] For the remainder of the analysis, $P_{ij}(n\tau_{\text{samp}})$ will now be used to represent $S_{ij}(n\tau)$ from Eq. (4).
[35] M. R. Shirts, J. W. Pitera, W. C. Swope, and V. S. Pande, J. Chem. Phys. **119**, 5740 (2003).
[36] M. R. Shirts and V. S. Pande, J. Chem. Phys. **122**, 134508 (2005).
[37] A. Ë. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (AIP, New York, 1994).
[38] A. R. Fersht, Proc. Natl. Acad. Sci. U.S.A. **99**, 14122 (2002).
[39] Each of the N_U states in $U=\{p_i|D(p_i)\leqslant 3\}$ is populated from a uniform distribution. Each state is initially assigned a probability of $X_{p_i\in U}=1/N_U$, while the structured states are initially unpopulated, i.e., $X_{p_i\notin U}=0$.