Foldamer dynamics expressed via Markov state models. II. State space decomposition

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

(Received 1 June 2005; accepted 26 June 2005; published online 16 September 2005)

The structural landscape of poly-phenylacetylene (pPA), otherwise known as *m*-phenylene ethynylene oligomers, has been shown to consist of a very diverse set of conformations, including helices, turns, and knots. Defining a state space decomposition to classify these conformations into easily identifiable states is an important step in understanding the dynamics in relation to Markov state models. We define the state decomposition of pPA oligomers in terms of the sequence of discretized dihedral angles between adjacent phenyl rings along the oligomer backbone. Furthermore, we derive in mathematical detail an approach to further reduce the number of states by grouping symmetrically equivalent states into a single parent state. A more challenging problem requires a formal definition for knotted states in the structural landscape. Assuming that the oligomer chain can only cross the ideal helix path once, we propose a technique to define a knotted state derived from a helical state determined by the position along the helical nucleus where the chain crosses the ideal helix path. Several examples of helical states and knotted states from the pPA 12-mer illustrate the principles outlined in this article. © *2005 American Institute of Physics*. [DOI: 10.1063/1.2008230]

I. INTRODUCTION

As computational resources become ever more available through inexpensive and increasingly powerful hardware, more sophisticated software, and parallelization algorithms, the ability to calculate large molecular simulation data sets at the ensemble level[1,2] is becoming available to more and more research groups interested in studying protein folding[3] and molecular self-assembly dynamics. Consequently, more sophisticated methodologies are being developed to analyze the multitude of molecular trajectories needed to adequately sample conformation space. Analysis techniques[4–7] based on Markov chains have proven to be effective in reducing the high dimensional conformation space into a simpler, more manageable framework to better understand the dynamics of the self-assembly process.

The state space decomposition is an essential component of these methods which employ Markov chains to describe the dynamics. However, defining the state space decomposition is not always a well-defined problem, and varies from one molecular system to another. Moreover, the quality of the analysis results depends significantly on the state definitions. Why is this so? In order for a stochastic process[8,9] (such as self-assembly or protein folding) to be Markovian, the future state of the system must depend only on the current state; in other words, the system must not maintain a memory of its past. Therefore, the state definitions are fundamentally coupled with the Markovian description of the dynamics; a poor state space decomposition may cause the Markov method to fail.

In a pair of seminal papers, Swope *et al.* use the theory[5] of Markov chains in order to develop a method to qualitatively assess whether a system exhibits Markovian behavior, and then apply[6] the method to several model molecular systems. In a simple numerical example of the method, they show that a nine-state Markovian system ultimately may not be Markovian at short lag time scales if the nine states are lumped together to form only three states. At long lag times the dynamics of the three-state system approaches the Markovian dynamics of the nine-state system. This simple exercise illustrates that for systems in which the exact Markovian dynamics is not known, the choice of state definitions may prove to be the difference between success or failure of the method. Applying the method to molecular systems, they observed Markovian behavior in alanine dipeptide simulations, but were unable to observe Markovian behavior in simulations of the *C*-terminal protein G β hairpin. They concluded that the failure of the method applied to the β hairpin resulted from a poor state space decomposition.

In the current article, we are interested in defining the state space decomposition for poly-phenylacetylene (pPA) oligomers, otherwise known as *m*-phenylene ethynylene oligomers,[10,11] for a subsequent analysis of the folding dynamics using Markov state models[4] (MSM) which we report in a companion paper.[12] These molecules are a relatively new class of nonbiological self-assembling oligomers, or "foldamers,"[13,14] which fold into a helical conformation in a poor solvent environment. They have the potential for very interesting applications in molecular recognition[15–17] as functional nanomaterials.[18] Understanding the dynamics of the self-assembly process of pPA oligomers is an important step in designing sequences that can fold on reasonable time scales to a stable functional state.[19–22] As has been expressed previously, the state space decomposition is an integral com-

a)Electronic mail: pande@stanford.edu

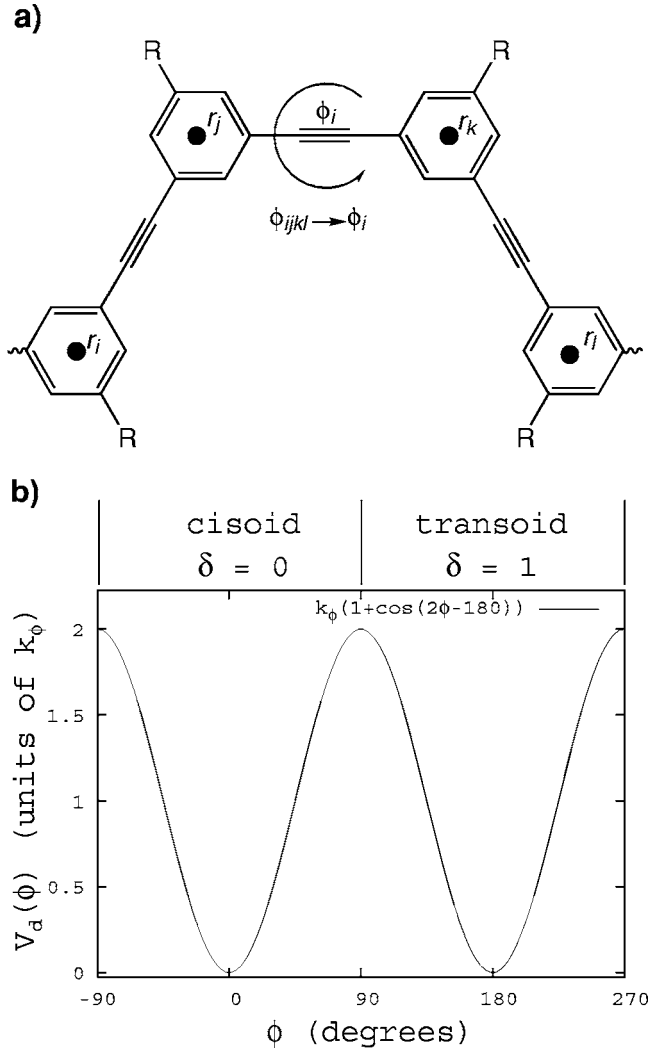

FIG. 1. The barrier height to isomerization in pPA oligomers is fairly high (Ref. 28) $V_d(\phi)/2 \approx 0.6$ kcal/mol which is approximately kT. Therefore, dihedral isomerizations dominate the dynamics of pPA oligomers. As a consequence, molecular conformations are parametrized in terms of these dihedral angles between adjacent phenyl rings. (a) Chemical structure of a pPA oligomer. 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 $\phi_{ijkl} \rightarrow \phi_i$ shown is a general relationship for all dihedral angles. (b) Proper dihedral angle potential across the acetylene bond between two adjacent phenyl rings present in molecular mechanics force field models of pPA oligomers. The energy barriers at ±90° provide a natural framework to discretize state space into a binary representation of the sequence of discretized dihedral angles.

ponent of the analysis methodology based on Markov state models to simplify the description of the dynamics of an ensemble of molecular simulation trajectories.

II. STATE DECOMPOSITION

The dynamics of pPA oligomers are dominated by dihedral isomerizations around the acetylene bond between adjacent phenyl rings [see chemical structure in Fig. 1(a)]. When all of the dihedral angles are in *cisoid* configurations, the oligomer chain will have a helical conformation which is the folded state, stabilized due to a maximal number of contacts. The folding process consists of an initially random extended chain collapsing into compact states, such as partially folded helical structures with a helical nucleus and frayed ends, in addition to knotted structures which may ultimately be trapped states.[20,21] For those trajectories which fold[19–21] through partially folded helical states, the helical nucleus grows as dihedral isomerizations from *transoid* to *cisoid* configurations at the boundaries of the helical nucleus extend the helical content of the helical nucleus. In the right solvent conditions, eventually all dihedral angles will be *cisoid*, completing the folding process to the fully helical folded state. The fact that the dynamics are dominated by the dihedral isomerizations provides a natural connection to define the states of the oligomer in terms of the sequence of dihedral angles along the chain.

The energy function for the molecular mechanics model of pPA oligomers[19–21] includes terms for the proper dihedral angle potential across the acetylene linkage between phenyl rings which are cosine functions with a multiplicity of 2 and phase shift of 180°, [see Fig. 1(b)]:

$$V_d(\phi) = k_\phi[1 + \cos(2\phi - 180\,^\circ)]. \tag{1}$$

The energetic barriers at ±90° inherent to this potential-energy function lend natural boundaries for the discretization of state space. A pPA oligomer with L monomers has L-3 dihedral angles [see Fig. 1(a)]; therefore, given that the dynamics of pPA oligomers are dominated by dihedral isomerizations, the molecular conformations are parametrized in terms of dihedral angles $\phi=[\phi_1 \cdots \phi_{L\text{-}3}]$ rather than the $3N$ atomic coordinates r_i. The sequence of continuous dihedral angles ϕ_i for a given molecular conformation may be discretized into a sequence of binary numbers δ_i determined by whether dihedral angle ϕ_i is either *cisoid* ($\delta_i=0$) or *transoid* ($\delta_i=1$). The discrete representation of a pPA L-mer conformation may thus be parametrized as an L-3 bit binary number $\delta(\phi)$:

$$\delta(\phi) = [\delta_1 \cdots \delta_{L\text{-}3}] \text{ with } \delta_i = \begin{cases} 0 & \text{if } -90\,^\circ < \phi_i < 90\,^\circ \\ 1 & \text{if } 90\,^\circ < \phi_i < 270\,^\circ . \end{cases} \tag{2}$$

In addition, pPA oligomers are achiral, leading in theory to equal populations of left- and right-handed helices. In order to distinguish left- and right-handed helical conformations, an additional bit $\delta_{L\text{-}2}$ is added at the 2^0 position to the dihedral number $\delta(\phi)$; the left-handed helical conformations get a 0 bit, whereas the right-handed helical conformations get a 1 bit. This L-2-bit binary number $\delta(\phi)$ is subsequently converted to a decimal number, $\delta(\phi) \rightarrow j$, which represents a discretized state s_j; as a side note, the left-handed states have even state indices and the right-handed states have odd state indices as a result of the convention for the $\delta_{L\text{-}2}$ bit. The set of $2^{L\text{-}2}$ states $\mathcal{S}=\{s_0, s_1, \ldots, s_{2^{L\text{-}2}-1}\}$ based on this discretization of dihedral angle space represents the state decomposition of conformation space for all realizations of pPA L-mer conformations. Moreover, given a state s_j, the binary number $\tilde{\delta}(s_j)=[\delta_1 \cdots \delta_{L\text{-}2}]$ representing the sequence of discretized dihedral angles δ_i is easily obtained from the state index j by converting the decimal representation of the state index into a binary number. For example, the left-handed folded state of a pPA 12-mer is represented by $\tilde{\delta}(s_0)=[0000000000]$, while

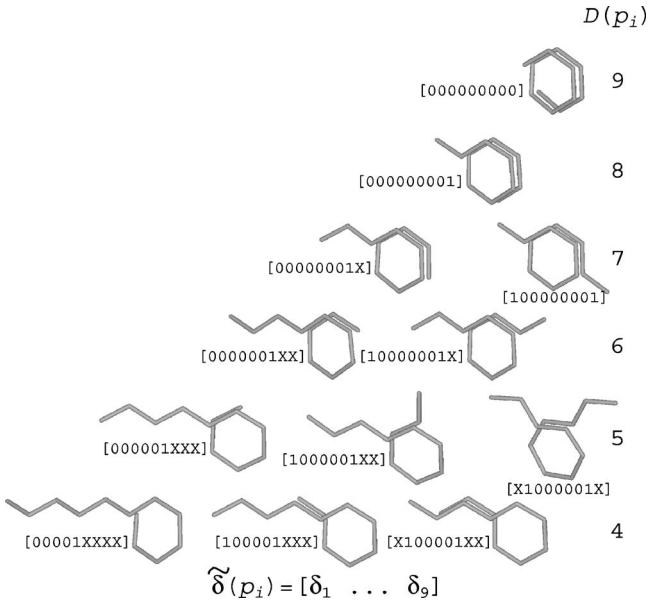

FIG. 2. Images for a select group of idealized reference structures for the helical parent states p_i of the pPA 12-mer. Also shown is the 9-bit binary number $\tilde{\delta}(p_i)$ (the tenth bit for the chirality is not shown). Dihedral angles of unstructured regions of the chain are denoted with an *X*, meaning that the dihedral angle may be *cisoid* ($X=0$) or *transoid* ($X=1$). The structures shown all have *transoid* dihedral angles in the unstructured regions.

the right-handed folded state is $\tilde{\delta}(s_1)=[0000000001]$. The fully extended state is represented by $\tilde{\delta}(s_{1023})=[1111111111]$, and a left-handed helical structure with a single frayed end is $\tilde{\delta}(s_{512})=[1000000000]$. Further examples are shown in Fig. 2.

The states s_j may be characterized structurally based upon the size of the helical nucleus $D(s_j)$, which is defined to be $D(s_j)=$maximum number of consecutive 0 s in $\tilde{\delta}(s_j)$ not including chirality bit $\delta_{L\text{-}2}$. As may be seen from the previous examples, both the left- and right-handed folded states, s_0 and s_1, respectively, have a helical "nucleus" of nine dihedral angles: $D(s_0)=D(s_1)=9$. State s_{512} has a structural parameter $D(s_{512})=8$. Illustrations of states s_j with a given size of helical nucleus $D(s_j)$ are shown in Fig. 2 for a pPA 12-mer.

A. Homopolymers

We have already introduced the notion of conformational symmetry with the concept of chirality leading to left- and right-handed helical conformations and an additional bit in the binary numbers $\delta(\phi)$. An additional symmetry is present in the state space $\mathcal{S}$ of homopolymer pPA oligomers with respect to the reversal of the sequence of discretized dihedral angles $\tilde{\delta}(s_j)$. Reversing the dihedral angle sequence or the chirality of a conformation should not change the dynamics since the forces governing the dynamics will be invariant under the symmetry operations. In view of the invariance of dynamics of symmetrically equivalent states, we propose a method to reduce the state space decomposition in terms of what we refer to as parent states. A general summary of the method dictates that equivalent states from different symmetry classes will have equivalent dynamics and may therefore be grouped into parent states which are sets of equivalent states. The equivalency relationship requires a formal definition in order to reduce the state space decomposition from states s_j to parent states p_i. In order to define the equivalency relationship based on chiral and sequence symmetries, symmetry operations also need to be defined. The next few paragraphs formalize these definitions needed to decompose the state space in terms of the parent states p_i based on the equivalency of states s_j.

We define a partitioning of the state space $\mathcal{S}$ into symmetry classes (disjoint subsets of $\mathcal{S}$) based on sequence and chiral symmetries. $\mathcal{S}=\mathcal{S}_L^L\cup\mathcal{S}_L^R\cup\mathcal{S}_R^L\cup\mathcal{S}_R^R\cup\mathcal{S}_L\cup\mathcal{S}_R$. The subscript represents the chirality whereas the superscript represents the sequence symmetry. The states in either of the two symmetry classes $\mathcal{S}_L$ or $\mathcal{S}_R$ have reversal-symmetric sequences and therefore have neither right nor left sequence symmetry as shown by the absence of the superscript in the symmetry class. The sequence symmetry for state s_j is determined by the position of the helical nucleus (largest run of consecutive 0 bit, representing *cisoid* dihedral angles) in the sequence $\tilde{\delta}(s_j)$. This is more easily explained using an example to illustrate the method to determine the sequence symmetry. The states s_2, s_3, s_{512}, and s_{513} for a pPA 12-mer are symmetry-related states which each have a single frayed end relative to the fully helical folded state [see Fig. 2, $D(p_i)=8$]. The terminus (left or right) at which the frayed end appears, along with the chirality of the helical nucleus determines the respective symmetry class of the states. For instance, state s_2 has L symmetry for both the chirality and the sequence symmetry (0 in the δ_{10} bit, while the helical nucleus is positioned on the left of the sequence, respectively), i.e, $\tilde{\delta}(s_2)=[0000000010]\Rightarrow s_2\in\mathcal{S}_L^L$. Similary, state s_3 is an element of symmetry class $\mathcal{S}_R^L$: $\tilde{\delta}(s_3)=[0000000011]\Rightarrow s_3\in\mathcal{S}_R^L$. In addition, states s_{512} and s_{513} are elements of symmetry classes $\mathcal{S}_L^R$ and $\mathcal{S}_R^R$, respectively: $\tilde{\delta}(s_{512})=[1000000000]\Rightarrow s_{512}\in\mathcal{S}_L^R$ and $\tilde{\delta}(s_{513})=[1000000001]\Rightarrow s_{513}\in\mathcal{S}_R^R$. The symmetry classes will prove to be valuable in determining the symmetry operations needed to establish the equivalency relationship of states s_j.

The set of symmetry operations which act upon a state s_j and return a new state s_i reverses the sequence of discretized dihedral angles, $\mathcal{O}_{\text{rev}}$, mirrors the chirality of the helical nucleus, $\mathcal{O}_{\text{mir}}$, or both, $\mathcal{O}_{\text{mirrev}}$. Moreover, the identity operation $\mathcal{O}_{\text{ident}}$ returns the same state. These symmetry operations are formally defined to be

(1) $s_i=\mathcal{O}_{\text{ident}}(s_j):s_i=s_j$ (identity operation),

(2) $s_i=\mathcal{O}_{\text{mir}}(s_j):\tilde{\delta}_{L\text{-}2}(s_i)=\text{not}\,\tilde{\delta}_{L\text{-}2}(s_j)$, $\tilde{\delta}_k(s_i)=\tilde{\delta}_k(s_j)$ for $k\neq L$-2 (bitwise *not* chirality bit, other L-3 bits preserved),

(3) $s_i=\mathcal{O}_{\text{rev}}(s_j):\tilde{\delta}_k(s_i)=\tilde{\delta}_{L\text{-}2\text{-}k}(s_j)$ for $k=1,\ldots,L$-3, $\tilde{\delta}_{L\text{-}2}(s_i)=\tilde{\delta}_{L\text{-}2}(s_j)$ (reverse ordering of first L-3 bits, preserve chirality bit),

(4) $s_i=O_{\text{mirrev}}(s_j):s_i=\mathcal{O}_{\text{mir}}[\mathcal{O}_{\text{rev}}(s_j)]$ (sequentially apply $\mathcal{O}_{\text{rev}}$ and $\mathcal{O}_{\text{mir}}$ operations).

Furthermore, a single symmetry function $\mathcal{O}(s_j)$ is defined such that it selects the correct symmetry operation to perform

on state s_j according to the symmetry class to which state s_j belongs

$$\mathcal{O}(s_j) = \begin{cases} \mathcal{O}_{\text{ident}}(s_j) & \text{if } s_j \in \mathcal{S}_L^L \text{ or } s_j \in \mathcal{S}_L \\ \mathcal{O}_{\text{rev}}(s_j) & \text{if } s_j \in \mathcal{S}_L^R \\ \mathcal{O}_{\text{mir}}(s_j) & \text{if } s_j \in \mathcal{S}_R^L \text{ or } s_j \in \mathcal{S}_R \\ \mathcal{O}_{\text{mirrev}}(s_j) & \text{if } s_j \in \mathcal{S}_R^R. \end{cases} \tag{3}$$

Essentially, $\mathcal{O}(s_j)$ maps a state s_j from any symmetry class into the $\mathcal{S}_L^L$ or $\mathcal{S}_L$ symmetry classes: $\mathcal{O}(s_j) \rightarrow s_i$, $s_i \in \mathcal{S}_L^L$ or $\mathcal{S}_L$ (depending on whether or not the state is sequence reversal symmetric). With these definitions in place, the equivalency relationship may now be defined; two states s_i and s_j are equivalent if the symmetry function $\mathcal{O}$ transforms each state into the same state: $s_j \sim s_i$ if $\mathcal{O}(s_j) = \mathcal{O}(s_i)$. From the previous examples using states s_2, s_3, s_{512}, and s_{513}, it is apparent that these four states are all equivalent, $s_2 \sim s_3 \sim s_{512} \sim s_{513}$, since the symmetry function $\mathcal{O}$ transforms each respective state into state s_2, $\mathcal{O}(s_2) = \mathcal{O}(s_3) = \mathcal{O}(s_{512}) = \mathcal{O}(s_{513}) = s_2 \in \mathcal{S}_L^L$.

Now that the equivalency relationship has been formally defined between the two states s_i and s_j, the necessary tools are in place to combine multiple equivalent states s_j into a single parent state p_i: $p_i = \{s_j | s_j \sim s_i\}$. The index of the parent state could be given by any of the indices of the states s_j in the set p_i; however, we choose the parent state index to be the state index of the state in symmetry class $\mathcal{S}_L^L$ or $\mathcal{S}_L$. For example, from the example above $s_2 \in \mathcal{S}_L^L$, therefore the parent state index is 2: $p_2 = \{s_2, s_3, s_{512}, s_{513}\}$. Essentially, the states s_i and s_j have equivalent dynamics due to symmetry invariance and therefore may be treated as a single entity p_i. This demonstrates a practical way to apply symmetry in order to reduce the number of states in the state space decomposition $\mathcal{S}$ by redefining the states in terms of parent states, i.e., $\{s_j\} \rightarrow \{p_i\}$ (where the indices i are a subset of the indices j). This is a reasonable approach since excluded volume prohibits the spontaneous reversal of chirality and the chain cannot spontaneously reverse its sequence. Defining the state space decomposition in terms of parent states p_i effectively reduces the number of states from 2^{L-2} states (s_j) to $2^{L-4} + 2^{\text{int}[(L-4)/2]}$ parent states (p_i), which is nearly a reduction by a factor of 4.[23]

At this point, the astute reader may be questioning the need to distinguish the chirality with the δ_{L-2} bit, when the end goal is to reduce the number of states s_j via symmetry to parent states p_i. The need to distinguish the chirality becomes more apparent in the practical application of this method, when assigning conformations observed in the simulation data to states s_j and especially when attempting to assign conformations to trapped states $s_{j,k}^*$ (see Sec. II C). For example, the sequence of discretized dihedral angles $\delta(\phi)$ for conformation ϕ is unable to distinguish a compact knotted state $s_{0,k}^*$ or $s_{1,k}^*$ from the folded state s_0 or s_1, respectively.

Therefore, an improved metric is required which is capable of distinguishing knotted states from helical states. As outlined in our companion paper, we use the pairwise root-mean-square displacement (RMSD) to assign conformations to states since it is capable of distinguishing knots from helical states. However, RMSD also distinguishes chirality and sequence symmetry, as a consequence of an optimal alignment of the two structures to be compared, such that each of the states s_j in the parent state $p_i = \{s_j | s_j \sim s_i\}$ are distinguishable by RMSD. Therefore, in order to assign a given conformation to a particular state s_j or $s_{j,k}^*$, the RMSD is computed between the given conformation ϕ and a reference structure for each of the candidate states s_j or $s_{j,k}^*$ in turn. The state which minimizes the RMSD is the state to which the given conformation is assigned. Subsequently, the equivalency relationship based on symmetry is used to reduce the states s_j and $s_{j,k}^*$ to the parent states p_i and $p_{j,k}^*$ as outlined in this section.

B. Heteropolymers

For heteropolymer pPA oligomers, the sequence of discretized dihedral angles no longer contains symmetry since the sequence of monomers in general is not symmetric.[24] Therefore, only two symmetry classes based on chirality $\mathcal{S} = \mathcal{S}_L \cup \mathcal{S}_R$ will exist for heteropolymers, with only two corresponding symmetry operations, $\mathcal{O}_{\text{ident}}$ and $\mathcal{O}_{\text{mir}}$. Therefore, the symmetry function

$$\mathcal{O}(s_j) = \begin{cases} \mathcal{O}_{\text{ident}}(s_j) & \text{if } s_j \in \mathcal{S}_L \\ \mathcal{O}_{\text{mir}}(s_j) & \text{if } s_j \in \mathcal{S}_R \end{cases} \tag{4}$$

takes a much simpler form for heteropolymers. In this case, the 2^{L-2} state space decomposition may be segregated into only two disjoint subsets of left- or right-handed helical states: $\mathcal{S} = \mathcal{S}_L \cup \mathcal{S}_R$ where

$$\mathcal{S}_L = \{s_0, s_2, \ldots, s_{2^{L-2}-2}\},$$

$$\mathcal{S}_R = \{s_1, s_3, \ldots, s_{2^{L-2}-1}\}.$$

With chirality specified in the δ_{L-2} bit in the 2^0 position in the binary number $\tilde{\delta}(s_j)$, the states s_j and s_{j+1} for any even j are equivalent $s_{j+1} \sim s_j$, i.e., $\mathcal{O}(s_{j+1}) = \mathcal{O}(s_j) = s_i$. Therefore, $p_i = \{s_i, s_{i+1}\}$ and $\mathcal{S} = \{p_0, p_2, \ldots, p_{2^{L-2}-2}\}$, reducing the state decomposition $\mathcal{S}$ from 2^{L-2} states (s_j) to 2^{L-3} parent states (p_i); the number of states is reduced by a factor of 2. The symmetry operations for a homopolymer are obviously much more complicated but allow a state decomposition into much fewer parent states p_i.

C. Knots and traps

We have previously demonstrated the existence of knotted conformations in the conformational state space of pPA oligomers, in addition to the important role they play in the folding dynamics.[20] These knotted conformations contribute to trapped kinetic macrostates in pPA 20-mers. In pPA 12-mers, however, these knotted conformations are not entangled enough to trap the chains in kinetically trapped macrostates, but they do contribute to nonexponential folding kinetics. Given the remarkable kinetic properties of pPA oligomers due primarily to trapped states with knotted topologies, the need for a state space decomposition that includes knotted states becomes apparent.

In general, defining a state decomposition which includes knotted conformations is very difficult as there are

many ways knots may occur. As discussed previously, the folded state of pPA oligomers consists of an ideal helix with all dihedral angles in *cisoid* configurations. A knotted state is defined to be a state in which the chain crosses over itself along the path of the ideal helix in the fully helical folded state or in states with a partially folded helical nucleus. If we make an approximation that the chain can only cross the path of the ideal helix once, which should be valid for short chains, knotted states $s^*_{j,k}$ may be defined and characterized as derivatives of helical states s_j. The additional subscript k demarcates the position within the helical nucleus where the chain crosses itself perturbing the ideal helical structure: $\tilde{\delta}(s^*_{j,k})=\lfloor\delta_1 \ldots \hat{\delta}_k \cdots \delta_{L\text{-}2}\rfloor$ where the bit $\hat{\delta}_k$ at position k is specially marked to specify the position within the helical nucleus where the chain crosses itself along the ideal helix path.[25] The set of knotted states derived from state s_j is defined to be $K(s_j)=\{s^*_{j,k}|k \text{ in helical nucleus}\}$. As a consequence, the state space decomposition which includes knotted states then becomes $\mathcal{S}=\{s_0,s_1,\ldots,s_{2^{L\text{-}2}-1}\}\cup\cup_{i=0}^{2^{L-2}-1}K(s_j)$; for those states s_j which do not lead to knots, $K(s_j)$ will simply be empty.

Just as the number of states in the state space decomposition for helical states s_j is reduced from 2^{L-2} states (s_j) to $\sim 2^{L-4}$ parent states (p_i) by establishing an equivalency relationship between states based on chiral and dihedral angle sequence symmetries, the number of knotted states $s^*_{j,k}$ in the state decomposition may also be reduced by establishing a similar equivalency relationship between the knotted states based on symmetry operations. The symmetry operations and symmetry function defined in Eq. (3) may be extended to operate on knotted states as well. However, adding the symbols $\hat{0}$ and $\hat{1}$ to the knotted state definitions $s^*_{j,k}$ introduces two technical issues: first, $\tilde{\delta}(s^*_{j,k})$ is no longer a pure binary number; second, in reversal-symmetric sequences, a new element of symmetry has been introduced based on whether the symbol $\hat{\delta}_k$ appears on the left or right side of the helical nucleus.

The first issue is relatively simple to address. Fortunately, the fact that $\tilde{\delta}(s^*_{j,k})$ is no longer binary is not problematic. The only symmetry operations that utilize a bitwise binary operator are $\mathcal{O}_{\text{mir}}$ and $\mathcal{O}_{\text{mirrev}}$ (by extension), in which case the chirality bit $\delta_{L\text{-}2}$ is negated to invert the chirality. Since $\delta_{L\text{-}2}$ can never be $\hat{0}$ or $\hat{1}$, the exactness of the operation is maintained. Moreover, reversing the sequence of discretized dihedral numbers is not problematic for the symbol $\hat{0}$ or $\hat{1}$. Therefore, the symmetry operators $\mathcal{O}_{\text{ident}}$, $\mathcal{O}_{\text{mir}}$, $\mathcal{O}_{\text{rev}}$, and $\mathcal{O}_{\text{mirrev}}$ acting on helical states s_j are equally valid and applicable to operate on knotted states $s^*_{j,k}$ as well.

The second issue concerning knotted symmetry is more complicated. The formal definition for the symmetry function acting on a helical state $\mathcal{O}(s_j)$ [see Eq. (3)] needs to be modified to include new symmetry classes resulting from knot symmetry in the knotted states $s^*_{j,k}$. With the new symbol $\hat{\delta}_k$ comes a new symmetry class for knotted states $\mathcal{S}^{*K}_J$ in which the superscript K distinguishes knotted states which cross the ideal helix path on the right side of the sequence from those knotted states which cross on the left side of the sequence, and the asterisk denotes a symmetry class which contains knot symmetry. The symmetry function modified to act on knotted states is therefore [compare Eq. (3)]

$$\mathcal{O}(s^*_{j,k})=\begin{cases}\mathcal{O}_{\text{ident}}(s^*_{j,k}) \text{ if } s^*_{j,k}\in\mathcal{S}^L_L \text{ or } s^*_{j,k}\in\mathcal{S}_L \text{ or } s^*_{j,k}\in\mathcal{S}^{*L}_L\\ \mathcal{O}_{\text{rev}}(s^*_{j,k}) \text{ if } s^*_{j,k}\in\mathcal{S}^R_L \text{ or } s^*_{j,k}\in\mathcal{S}^{*R}_L\\ \mathcal{O}_{\text{mir}}(s^*_{j,k}) \text{ if } s^*_{j,k}\in\mathcal{S}^L_R \text{ or } s^*_{j,k}\in\mathcal{S}_R \text{ or } s^*_{j,k}\in\mathcal{S}^{*L}_R\\ \mathcal{O}_{\text{mirrev}}(s^*_{j,k}) \text{ if } s^*_{j,k}\in\mathcal{S}^R_R \text{ or } s^*_{j,k}\in\mathcal{S}^{*R}_R.\end{cases} \tag{5}$$

Since the individual symmetry operations are essentially unchanged, the only modification to $\mathcal{O}(s^*_{j,k})$ includes additional knotted symmetry classes for selecting the symmetry operation to perform on knotted state $s^*_{j,k}$.

For those knotted states derived from helical states which have reversal-symmetric sequences, the position where the chain crosses the ideal helix path $\hat{\delta}_k$ introduces a new element of knot symmetry. Fortunately, since the superscript is absent in the reversal-symmetric symmetry classes $\mathcal{S}_J$, the superscript may be used to designate the knot symmetry of these sequence reversal-symmetric knotted states $s_j\in\mathcal{S}_J\rightarrow s^*_{j,k}\in\mathcal{S}^{*K}_J$. Helical states which are not reversal symmetric remain in the same symmetry class upon conversion to knotted states since these states do not have knotted symmetry: i.e., $s_j\in\mathcal{S}^K_J\rightarrow s^*_{j,k}\in\mathcal{S}^K_J$. The precedence rule for assigning knotted states $s^*_{j,k}$ to symmetry classes is established so that sequence reversal symmetry takes precedence over knotted symmetry. There is no overlap between these symmetry classes, so there should be no ambiguities in the symmetry class assignments for the knotted states and a precedence rule is really unnecessary, but is only included for completeness. For instance, reversal-symmetric sequences have knot symmetry, and sequences that are not reversal symmetric do not have knot symmetry. With the symmetry operators and symmetry function modified to act on knotted states, an equivalency relationship may be defined, i.e., $s^*_{j,k}\sim s^*_{i,k'}$ if $\mathcal{O}(s^*_{j,k})=\mathcal{O}(s^*_{i,k'})$ by which knotted states $s^*_{j,k}$ may be grouped together into a set of knotted parent states $p^*_{i,k}=\{s^*_{j,k}|s^*_{j,k}\sim s^*_{i,k'}\}$. Thus, the state space decomposition $\mathcal{S}=\{p_i,p^*_{i,k}\}$ based on parent states has been formally defined to include knotted states.

The best way to illustrate the principle of knotted state symmetry is with examples, which we take from the real structures encountered in the analysis of the pPA 12-mer. From the formal state space decomposition for helical states, a pPA 12-mer has $2^{10}=1024$ helical states $\mathcal{S}=\{s_0,s_1,\ldots,s_{1023}\}$. The relevant compact states which may produce knotted conformations in pPA 12-mers are states s_0 and s_1 which are equivalent states in parent state p_0, and states s_2, s_3, s_{512}, and s_{513} which are equivalent states in parent state p_2. These six helical states are further decomposed into six subsets of knotted states $K(s_0)$, $K(s_1)$, $K(s_2)$, $K(s_3)$, $K(s_{512})$, and $K(s_{513})$ (see also Fig. 3). Applying the equivalency relationship, these knotted states $s^*_{j,k}$ may be assigned to parent knotted states $p^*_{i,k}$ to reduce the number of states in the state decomposition. For example, $\tilde{\delta}(s^*_{0,3})=[00\hat{0}\hat{0}0000000]\Rightarrow s^*_{0,3}\in\mathcal{S}^{*L}_L$, $\tilde{\delta}(s^*_{0,7})=[0000000\hat{0}\hat{0}00]\Rightarrow s^*_{0,7}$

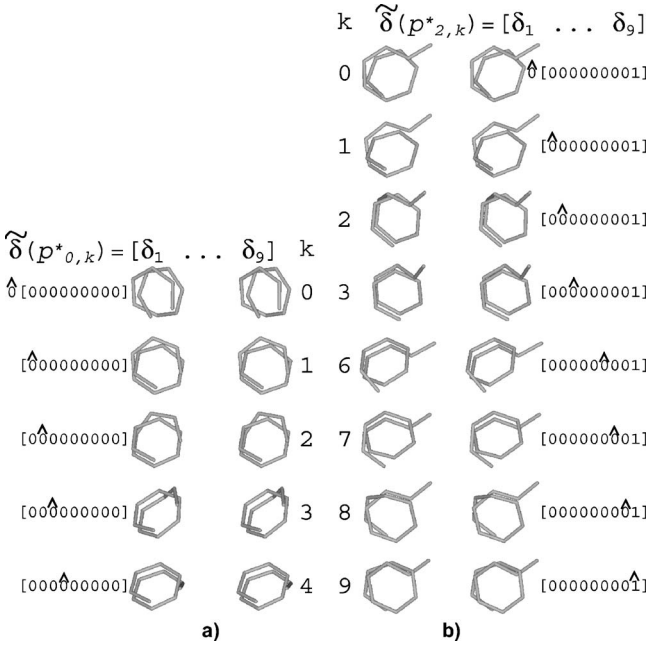

FIG. 3. Relaxed eye stereo images of the 13 parent knotted states $p^*_{i,k}$ of the pPA 12-mer. The 9-bit binary number $\tilde{\delta}(p^*_{i,k})$ (the tenth bit for the chirality is not shown) is shown with the caret (^) to represent where the chain crosses the path of the ideal helix. (a) Knotted states derived from the folded state p_0. (b) Knotted states derived from state p_2 which has one frayed dihedral angle from the fully helical folded state. Since state p_0 is sequence reversal-symmetric with respect to the discretized dihedral angles, knot symmetry ($\hat{\delta}_k$ bit on left or right of sequence) reduces the number of knotted states $s^*_{0,k}$ from ten to five parent knotted states $p^*_{0,k}$.

$\in \mathcal{S}^{*R}_L$, $\tilde{\delta}(s^*_{1,3}) = [00\hat{0}0000001] \Rightarrow s^*_{1,3} \in \mathcal{S}^{*L}_R$, and $\tilde{\delta}(s^*_{1,7}) = [0000000\hat{0}01] \Rightarrow s^*_{1,7} \in \mathcal{S}^{*R}_R$ are the symmetry classes for the equivalent knotted states derived from the fully helical left- and right- handed states s_0 and s_1, respectively, in which the chain crosses the ideal helix path at the third (or seventh) position in the sequence of discretized dihedral angles. The parent knotted state for these equivalent states is $p^*_{0,3} = \{s^*_{0,3}, s^*_{1,3}, s^*_{0,7}, s^*_{1,7}\}$.

The symmetry classes for the knotted states derived from the equivalent helical states s_2, s_3, s_{512}, and s_{513} are given as further examples (see Fig. 3): i.e., $\tilde{\delta}(s^*_{2,3}) = [00\hat{0}0000010] \Rightarrow s^*_{2,3} \in \mathcal{S}^L_L$, $\tilde{\delta}(s^*_{3,3}) = [00\hat{0}0000011] \Rightarrow s^*_{3,3} \in \mathcal{S}^L_R$, $\tilde{\delta}(s^*_{512,6}) = [1000000\hat{0}00] \Rightarrow s^*_{512,6} \in \mathcal{S}^R_L$, and $\tilde{\delta}(s^*_{513,6}) = [1000000\hat{0}01] \Rightarrow s^*_{513,6} \in \mathcal{S}^R_R$. Thus, these four knotted states are all equivalent, $\mathcal{O}(s^*_{2,3}) = \mathcal{O}(s^*_{3,3}) = \mathcal{O}(s^*_{512,6}) = \mathcal{O}(s^*_{513,6}) = s^*_{2,3} \in \mathcal{S}^L_L$, and may be combined into a parent knotted state, $p^*_{2,3} = \{s^*_{2,3}, s^*_{3,3}, s^*_{512,6}, s^*_{513,6}\}$.

While the incorporation of knotted states in the state space decomposition adds complexity to the symmetry classes and equivalency relationships, it is necessary for understanding kinetic traps. Once the trapped states are defined, Markovian state models constructed from an ensemble of trajectories from the simulation data provide a way to measure the mean lifetimes[5,7] of states to quantitatively assess whether a knotted state $p^*_{i,k}$ is kinetically trapped relative to other states p_i.

III. DISCUSSION

From the examples of knotted states $p^*_{i,k}$ in Fig. 3(a), note that the knotted state $p^*_{0,5}$ for a pPA 12-mer does not exist since the chain is not long enough to cross the ideal helix path at position $k=5$. For similar reasons, the states $p^*_{2,4}$ and $p^*_{2,5}$ derived from state p_2 do not exist as implied from the absence of these states from the examples shown in Fig. 3(b). One additional subtle issue arises with knotted chains. The states $p^*_{0,0}$ and $p^*_{2,0}$ are defined in the state decomposition although the position $k=0$ is not in the helical nucleus in a strict sense.[26] Most likely, these states are not traps even in longer chains; nevertheless, they are significantly populated in the simulation data.

The structural perturbations from an ideal helix necessary to observe knotted states are so drastic that one would think the underlying discretized dihedral number would be greatly affected when comparing states s_j and $s^*_{j,k}$. On the contrary, the discretized dihedral number does not change upon converting a helical state to a knotted state. This is explained by the fact that the boundaries for the range of *cisoid* dihedral angles $-90° < \phi < 90°$ in the definition of a *cisoid* discretized dihedral angle [see Eq. (2)] offers a wide range of fluctuations around the idealized helical path as discussed by Lee and Saven,[27] and further discussed in our companion paper, such that the knotted states still have all continuous dihedral angles within this range $-90° < \phi < 90°$ in the helical nucleus. The continuous dihedral angles at the position the chain crosses the ideal helix path are far from the ideal values expected for ideal helices, yet still within the range defined for *cisoid* dihedral angles.

How does the chain become trapped in a knotted state? During the initial stages of the folding process, the rapid collapse from extended states to compact states occurs nonspecifically. As a result of this nonspecific collapse to compact states, the chain may easily end up in a knotted compact state. Once the chain is caught in a knotted state, the dynamics consist of a sterically constrained search of compact conformation space[21] for the folded state or other partially folded intermediate states.

In previous studies conducted on the folding dynamics of pPA oligomers,[20,21] we observed topologically diverse intermediate structures, such as partially folded helical structures and structures analogous to β turns in proteins, in addition to knotted structures. Initially, we were fascinated with this observation since a significant population of β structures was observed without any prior intention to design the β structures into the structural landscape. Defining the state space decomposition of these pPA oligomers in terms of discretized dihedral angles helps to give further insight to understanding the nature of these observed β structures. Close inspection of the structures illustrated in Fig. 2 reveals an inherent tendency for pPA oligomers to form β structures, even though the helical folded state is the most stable structure accessible to the oligomer. Based purely on the combinatorial sequences of 1s and 0s for the 2^{L-3} discrete states s_j (ignoring chirality and traps), a specific β structure shown in Fig. 2 for $D(p_i)=4$ would be equally likely as the folded state s_0, i.e., $P(s_j)=P(s_0)=2^{-(L-3)}$. In the presence of interactions

between monomers, the probabilities of all states s_j will no longer be equal, but will depend on the stabilization energy provided by intermonomer contacts. For any length of pPA oligomer, the helical state s_0 maximizes the number of contacts and hence is the most stable state, which is obviously designated as the folded state. Since the β structures have nonnative contacts in the strand region of the structure, they are marginally stable relative to the folded state. This simple explanation accounts for the significant population of β structures observed in our simulations of pPA 12-mers and 20-mers.[20,21]

Since all of our simulations to date have been studies of homopolymers, there is no preference for native contacts over nonnative contacts. Therefore, we have been primarily concerned with understanding the folding dynamics relative to the helical folded state. Directions for future research on pPA foldamers include incorporating heteropolymer sequences onto the pPA backbone in order to design folded β structures which are more stable than the helical structure. Since trapped states are a major concern in the folding dynamics of pPA oligomers, we would also like to be able to design heteropolymer pPA sequences to fold, whether to the helical structure or the β structure, while avoiding trapped states. The simple combinatoric principles revealed in this article through the state space decomposition based on the sequence of discretized dihedral angles in the pPA backbone will be the foundation upon which our future work in the design of functional nanomaterials will commence.

IV. CONCLUSION

An accurate description of the stochastic folding dynamics of molecular systems requires an ensemble of trajectories to adequately sample conformation space. Markov chains provide a natural foundation for developing methodologies to analyze molecular folding trajectories at the ensemble level. Defining a state space decomposition for the molecular system is a fundamental component of Markovian state models and is crucial for successfully simplifying the description of the folding dynamics. We have derived a method to define the states of a pPA oligomer in terms of discretized dihedral angles between adjacent phenyl rings along the oligomer chain. Moreover, the knotted states were defined to further distinguish the structural features of the diverse conformations present in the structural landscape of pPA oligomers. The state space decomposition presented in this article exemplifies a strategy to simplify the high dimensional space of molecular systems for input into Markovian state models.

ACKNOWLEDGMENT

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. J. Sorin, B. J. Nakatani, Y. M. Rhee, G. Jayachandran, V. Vishal, and V. S. Pande, J. Mol. Biol. **337**, 789 (2004).
[2] E. J. Sorin and V. Pande, Biophys. J. **88**, 2472 (2005).
[3] V. Pande, I. Baker, J. Chapman *et al.*, Biopolymers **68**, 91 (2003).
[4] N. Singhal, C. D. Snow, and V. S. Pande, J. Chem. Phys. **121**, 415 (2004).
[5] W. C. Swope, J. W. Pitera, and F. Suits, J. Phys. Chem. B **108**, 6571 (2004).
[6] W. C. Swope, J. W. Pitera, F. Suits *et al.*, J. Phys. Chem. B **108**, 6582 (2004).
[7] P. Deuflhard, W. Huisinga, A. Fischer, and C. Schutte, Linear Algebr. Appl. **315**, 39 (2000).
[8] N. G. V. Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1992).
[9] I. Oppenheim, K. E. Shuler, and G. H. Weiss, *Stochastic Processes in Chemical Physics: The Master Equation* (MIT, Cambridge, 1977).
[10] J. Nelson, J. Saven, J. Moore, and P. Wolynes, Science **277**, 1793 (1997).
[11] R. Prince, J. Saven, P. Wolynes, and J. Moore, J. Am. Chem. Soc. **121**, 3114 (1999).
[12] S. Elmer, S. Park, and V. Pande, J. Chem. Phys. **123**, 114902 (2005).
[13] S. Gellman, Acc. Chem. Res. **31**, 173 (1998).
[14] D. J. Hill, M. J. Mio, R. B. Prince, T. S. Hughes, and J. S. Moore, Chem. Rev. (Washington, D.C.) **101**, 3893 (2001).
[15] R. Prince, S. Barnes, and J. Moore, J. Am. Chem. Soc. **122**, 2758 (2000).
[16] A. Tanatani, M. J. Mio, and J. S. Moore, J. Am. Chem. Soc. **123**, 1792 (2001).
[17] A. Tanatani, T. S. Hughes, and J. S. Moore, Angew. Chem., Int. Ed. **41**, 325 (2001).
[18] A. Barron and R. Zuckermann, Curr. Opin. Chem. Biol. **3**, 681 (1999).
[19] S. Elmer and V. Pande, J. Phys. Chem. B **105**, 482 (2001).
[20] S. Elmer and V. Pande, J. Chem. Phys. **122**, 124908 (2005).
[21] S. Elmer and V. Pande, J. Chem. Phys. **121**, 12760 (2004).
[22] W. Y. Yang, R. B. Prince, J. Sabelko, J. S. Moore, and M. Gruebele, J. Am. Chem. Soc. **122**, 3248 (2000).
[23] The second term in the number of parent states is a consequence of reversal-symmetric sequences in symmetry classes S_L and S_R; since these states have only chiral symmetry, there are only two equivalent states s_j in the parent states p_i rather than four. The function int(x) truncates a decimal number to its integer value.
[24] For heteropolymer monomer sequences which are sequence reversal symmetric, the definitions for dihedral angle sequence symmetry in homopolymers apply.
[25] For clarity in defining knotted states, the *transoid* dihedral angle that marks the end of the helical nucleus is included as a member of the helical nucleus since the chain may cross the path of the ideal helix at this position as well (see Fig. 3).
[26] Monomers are labeled starting from monomer 1, ending at monomer L. Therefore, the dihedral angles are labeled relative to the monomer labels in the following way: the dihedral angle at position 1 includes monomers 1, 2, 3, and 4, with position 1 occurring between monomers 2 and 3. The remaining dihedral angles are labeled by continuing along the chain in a similar manner. It is noted, therefore, that the position between monomers 1 and 2 is labeled position 0, even though there is no dihedral angle at that position (only three monomers 1, 2, and 3 contribute to position 0). Since there is no dihedral angle at position $k=0$, that position cannot be in the helical nucleus in a strict sense.
[27] O. Lee and J. Saven, J. Phys. Chem. B **108**, 11988 (2004).
[28] K. Okuyama, T. Hasegawa, M. Ito, and N. Mikami, J. Phys. Chem. **88**, 1711 (1984).