Catch bond-like kinetics of helix cracking: Network analysis by molecular dynamics and Milestoning

Steven M. Kreuzer¹ and Ron Elber²

¹*Department of Mechanical Engineering, University of Texas at Austin, Austin, Texas 78712, USA*

²*Institute for Computational Engineering and Sciences, Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, Texas 78712, USA*

(Received 4 April 2013; accepted 29 May 2013; published online 1 July 2013)

The first events of unfolding of secondary structure under load are considered with Molecular Dynamics simulations and Milestoning analysis of a long helix (126 amino acids). The Mean First Passage Time is a non-monotonic function of the applied load with a maximum of 3.6 ns at about 20 pN. Network analysis of the reaction space illustrates the opening and closing of an off-pathway trap that slows unfolding at intermediate load levels. It is illustrated that the nature of the reaction networks changes as a function of load, demonstrating that the process is far from one-dimensional.

© 2013 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4811366>]

I. INTRODUCTION

Alpha helices are one of the primary secondary structures found as short segments in globular proteins and as longer segments in some filamentous proteins. They represent the simplest secondary structure, in which the conformation is stabilized by hydrogen bonds between backbone donors and acceptors of residues that are 4 spaces apart on the peptide backbone. Due to this simplicity, a sizeable literature has focused on understanding the folding and unfolding of helices as a model system. Here we consider a long helix of 126 amino acids that unfolds in the presence of load. Long helices are frequently found in molecular machines such as kinesin and myosin, and their stability under load is therefore of considerable biophysical interest. Prior theoretical investigations of helix unfolding were done without load and for shorter helices (5–20 amino acids^{1–14}).

The present study is a follow up on an earlier Milestoning investigation on the initiation of unfolding for the long helix under load.¹⁵ That study hinted that the Mean First Passage Time (MFPT) of the initiation of unfolding was a non-monotonic function of the applied load magnitude. This non-monotonic behavior is reminiscent of the “catch-bond” phenomenon observed in some inter-molecular structures in which intermediate loads induce unfolding more slowly than either low or high loads.^{16–26} The MFPT calculated at 10 pN was longer than that calculated at both 0 pN and 100 pN. Pathway calculations indicated that this non-monotonic behavior appeared to be related to the formation of non-native (i.e., non α -helical) structures, namely π helices, more prominently at this intermediate load level.

The goal of the present paper is to provide a broad picture of the catch-bond phenomenon in a long helix. In the analysis, we use the Milestoning theory in a network formulation to obtain a comprehensive view of the kinetics and thermodynamics of the unfolding process. Compared to our earlier study,¹⁵ we consider a larger number of load levels to better resolve the molecular mechanism (ten load levels instead of three). The network model was adjusted with respect to the

earlier study to make it a directed graph, which is more detailed and accurate. We use significantly larger statistics to reduce the error bars of our analysis and employ a more strict definition of a “cracking” event. Here a “cracking” event is defined as the first transition of an amino acid to a Ψ dihedral angle $>90^\circ$, representative of an extended conformation; a more complete description of the definition of cracking can be found in Sec. II D, “Coarse graining approach.” We examined the Milestoning networks at greater detail, assessing their numerical accuracy, overall statistical stability, and robustness upon small variation in edge properties. The results add significant new evidence supporting the existence of the catch-bond phenomenon and insights into its origin.

Inter-molecular catch-bonds were first observed for the interactions between rolling immune cells (leukocytes) and the blood vessel wall such that increased blood shear flow rates would lead to longer-lived bonds.¹⁶ A selection of inter-molecular interactions exhibiting catch-bond behavior includes FimH/mannose,^{17–20} L-selectin/endoglycan,^{21,22} P-selectin/PSGL-1,²³ and myosin/actin.²⁴ Explanations for this behavior include a so-called “allosteric” regulation mechanism and a sliding-rebinding mechanism. The allosteric regulation mechanism is based on the idea that intermediate forces induce a primarily rigid-body conformational change in the receptor molecule, such as disruption to inter-domain interactions, leading to structural rearrangements in the binding site that allow stronger interactions with the ligand. The sliding-rebinding mechanism posits that loads applied to the ligand induce a rigid body motion in such a way that the load direction is parallel to the binding surface, resulting in a “sliding” of the ligand along this binding surface. In this mechanism, the parallel orientation of the applied load relative to the binding surface is in contrast to the perpendicular orientation present at low loads. Thus, the intermediate loads are capable of changing the loading orientation due to rotation of the receptor, but these load levels are not large enough to prevent rebinding of individual functional groups of the ligand (e.g., hydrogen bond donors and acceptors) to the receptor. Thus, the two major explanations for the catch-bond

behavior of inter-molecular systems include (a) allosteric rearrangements permitting additional interactions and/or (b) changes to the alignment of the binding surface relative to the load direction allowing the transient formation of non-native interactions that delay unbinding.

The unfolding mechanisms of protein structures under load have been studied using theoretical,^{27–31} experimental (atomic force microscopy³²) and computational (steered molecular dynamics³³) means, typically through the application of forces from either a true or virtual cantilever. In such investigations, the magnitude of the applied force is a function of the instantaneous reference position of the cantilever base and the transient nature of the applied force is due to both enforced displacement of the cantilever base (the “loading rate”) and the extension of the protein such that forces applied to an aligned but folded protein rapidly increase prior to an unfolding transition. Bell’s approach²⁸ is a simple model to describe the unfolding kinetics of such experiments via two parameters: (i) the height of the free energy barrier and (ii) the position of a “transition state” along a one-dimensional reaction coordinate. This model has been used to characterize the dependence of the measured unfolding rate on the experimental loading rate.

Previous simulations of the strength of isolated helical proteins have employed the SMD approach³⁴ to identify unfolding forces and to determine the Bell model parameters under load. These simulations observed unfolding forces on the order of ~ 200 – 300 pN for the slowest simulated loading rates. Experimental determinations of helical protein strength via atomic force microscopy have placed the unfolding force in the 20–40 pN range.^{25,26} The use in simulations of significantly higher loading rates than in the experiment was attributed to time-scale limits of straightforward simulations. Simulations of unfolding under load are restricted to microseconds at most, far shorter than typical experimental times of milliseconds and longer. Forcing unfolding to occur over fast time periods is likely to induce short and direct pathways that are not followed by the system at lower, more typical loads. Hence the mechanism of unfolding may change as a function of loading magnitude. We illustrate this phenomenon in the present paper.

A recent look at unfolding time versus load for a simple analytical energy landscape with a single energy barrier²⁵ illustrated that a “rollover” in lifetime can appear when a degree of freedom other than the molecular extension (i.e., load direction) is essential for the unfolding kinetics. This study demonstrated that this “rollover” can occur for a single unfolding pathway, and therefore may not require multiple distinct pathways such as a “fast” and “slow” pathway whose relative traffic switch due to load magnitude. A study of the unfolding mechanism of a coarse-grained model of protein G²⁶ showed that the unfolding mechanism exhibited a discrete switch between competing unfolding pathways at low and high force. This switch, recognized through a clustering analysis of the unfolding structures, was most pronounced for “stiff” load directions and resulted in a “catch-bond”-like kinetic turnover. In contrast, the unfolding mechanisms for “weak” loading directions exhibited a more gradual shift than that seen in the “stiff” loading directions. Thus, while a simple

analytical potential is capable of generating kinetic turnover without a discrete switch in mechanism, previous results from more detailed energy landscapes with more explicit characteristics of proteins demonstrate that this non-monotonic unfolding behavior can also be a function of traffic along discrete pathways. The α helical system studied here provides an ideal test system for evaluating these competing visions as it has a rough energy landscape relative to analytical models yet it is simple enough to provide comprehensive analysis of unfolding mechanisms, which is harder to obtain for larger systems.

II. METHODS

A. Computational protocol

The system studied is the 126 residue long chain A of the 2fxm.pdb beta-myosin S2 coiled coil, which forms a single ~ 16 nm long α helix. Its sequence is:

GSSPLLKSAEREKEMASMKEEFTRLKEALEKSEARRK
ELEEKMSVSLQLQEKNDLQLQVQAEQDNLADAEERCDQ
LIKNIQLEAKVKEMNERLEDEEEMNAELTAKKRKLE
DECSELKRDIDDLELTLAK.

Simulation conditions were the same as the previous study. Simulations are performed with the GROMACS³⁵ 4.3.5 distribution using the GROMOS96 “53a6” united-atom force field with a 2 fs time step. The original structure was placed in an explicit (SPC model) water-filled box sufficiently large to ensure 1.2 nm of solvation surrounding the helix in the x, y, and z directions, resulting in a $7.4 \times 8.2 \times 25.6$ nm rectangular box (Fig. 1). While different force fields could have been used to describe the solvated helix, the choice here ensures efficiency and consistency with our earlier study.¹⁵ System charge neutrality was enforced by adding 102 Na⁺ and 91 Cl[−] counter-ions at a concentration of 0.1 mole/liter, resulting in a full system containing 1290 protein atoms and 49 605 water molecules, in addition to these counter-ions. Periodic boundary conditions were used throughout all phases of the

FIG. 1. (Top) Helical structure (2fxm.pdb chain A) within simulation box filled with explicit water. (Bottom) Helical structure illustrating two groups (red and blue) comprised of the 18 residues on either end. Constant force loads are applied to these two groups in a constant direction (dashed black line) aligned with the long helical axis.

TABLE I. Number of simulations for each load level.

Load (pN)	Simulations
0	75
5	28
10	28
15	50
20	60
25	83
30	28
40	24
70	30
100	20

MD simulations as was the parallel processing version of the LINCS constraint on fast bond dynamics.^{36,37}

Following energy minimization, the initial structure was heated to 310 K in three 20 ps steps; temperature stabilization was monitored before each temperature increase. Pressurization to 1 atm was performed over 750 ps on the heated system using a Berendsen thermostat to maintain temperature. Final production runs used a Nose-Hoover thermostat with a relaxation constant of 0.1 ps to maintain 310 K and a Parrinello-Rahman barostat with a coupling constant of 1.0 ps. The last algorithms provide the rigorous NPT ensemble.^{38,39} The Coulomb and van der Waals cutoff radii were set to 1.4 nm; Particle Mesh Ewald is used to handle long-range electrostatics. A base set of ten conformations (seeds) obtained from sampling an initial 1 ns equilibrium simulation was used for the production simulations. The conformations of these 10 seeds were used with different sets of randomly sampled velocities to initiate production simulations at different load levels. The set of 10 conformations used as seeds was repeated with new velocities to achieve the total simulation quantity as shown in Table I (note that simulation quantities may not be multiples of 10 in cases where individual simulations did not experience a “cracking” event).

All simulations were run on the Texas Advanced Computing Center’s (TACC) Sun Constellation Linux Cluster, Ranger (<http://www.tacc.utexas.edu>).

Constant force loads were applied to the helix’s C- and N-terminal ends, with the pull groups being the 18 residues on either end. Load magnitudes tested included 0, 5, 10, 15, 20, 25, 30, 40, 70, and 100 pN. The load direction, fixed throughout the simulations, was set in the [0 0 1] direction corresponding to the helix’s nominal axis as defined by the vector between the center of mass of the load end groups. Loads, applied via the “pull” command within GROMACS, were evenly distributed over all atoms of the pull groups, explaining the selection of a relatively large number of “pull group” residues, namely 18, to minimize end effects. The loaded residues were not analyzed in order to minimize any possible complications due to direct load application. Furthermore, it was necessary to exclude two residues on either end of the 86-residue analysis section in a concerted effort to minimize the effects of applied load on the hydrogen bonds spanning the first and last analyzed residues.

In contrast to our previous work, only the first residue to “crack” (defined by a ψ dihedral angle $> 90^\circ$) within a given simulation is considered for analysis; all other residues present in the simulation were ignored during analysis. The decision to consider only the first residue to initiate unfolding was made in an effort to more clearly isolate the initial unfolding event from the influences of neighboring residues that may have already experienced unfolding, such as would occur when an unfolding event is propagating through the helix.

The only simulation time included in the analysis was the time prior to the initiation of unfolding of the first residue; any remaining simulation time was discarded for the purposes of the present study. At least 30 simulations were run for each load level except 100 pN, which was run 20 times. Unfolding with 100 pN is rapid and efficient. The 20 trajectories at this load unfold in a similar way and provided adequate statistics. Additional simulations were run for the 0 pN and 25 pN load levels to provide better statistical confidence at these load levels.

B. Milestoning

1. Overview

The “Milestoning” approach was used to analyze the kinetics and process of the initiation of unfolding. While most implementations of Milestoning focus on a large number of short simulations initiated at intermediate stages of an activated process, the present implementation utilizes Milestoning as an analysis tool for post-processing relatively long “initial-value” simulations. In the current implementation, the raw, full atomic trajectories generated by GROMACS are converted to trajectories of coarse-grained states. The atomic positions at each saved image define the instantaneous coarse-grained state of the residue of interest (see definition of Milestoning states below). The Milestoning approach is then used to analyze the transitions between the coarse-grained states. The interested reader can find a full derivation of the Milestoning equations in Ref. 40; a brief description follows.

2. Milestones

Milestoning is concerned with fluxes between molecular states that are defined by partitions in phase space; we call these partitions “milestones.” The partitions are defined in a coarse space that is a reduction of the full atomic space of the MD simulations. For example, here we use the hydrogen bonds and the backbone torsions of the peptide chain to define the state of a helix. Atomically detailed trajectories in the full phase space of the protein, aqueous solution, and counter ions are mapped into a state trajectory defined by the last milestone or interface that is passed.

Milestones are defined with the help of anchors. Anchors are configurations determined with coarse variables that provide a sample of the important regions of the reaction space (see Fig. 2). For example, consider a system of two anchors. In one of the anchors the system is in a hydrogen-bonded

FIG. 2. Schematic illustration of a set of anchors that are coarse grained descriptions of atomic space (represented here by the $\alpha 3$, $\alpha 1$, $[none]$, $\Psi > 90$, and $\pi 34$ anchors) and a set of milestones that are the interfaces between these anchors (represented here by the $\alpha 3:\alpha 1$, $\alpha 1:\alpha 3$, $\alpha 1:[none]$, and $[none]:\Psi > 90$ designations). For more details, see Sec. II D, “Coarse graining approach,” below. Briefly, the “ α ” designation in the anchor names indicates the existence of α -helical hydrogen bonds while the numeral indicates the number of α -helical bonds currently formed (similar for the $\pi 34$ designation in which 4 π -helical bonds are shown). The $[none]$ and $\Psi > 90$ anchors lack intact hydrogen bonds and have Ψ dihedral angles < 0 and > 90 , respectively. Note the use of directional milestones as illustrated by the two distinct milestones separating the $\alpha 3$ and $\alpha 1$ anchors, each with a unique directionality relative to the two anchors.

state and in the second anchor it is not. A milestone representing the interface between these two anchors is defined as the distance at which the hydrogen bond changes its state (it is turned on (or off)). The milestone is used to map the atomically detailed trajectory $X(t)$ to a trajectory of hydrogen bond (hydrogen bonded q_1 , no hydrogen bond q_0) states. For example, if the system forms a hydrogen bond at time t and dissociates at time t' , the trajectory in milestone space takes the value of q_0 between time 0 and t (the last milestone that it passes (or initiated on) is q_0), the value of q_1 between t and t' and the value of q_0 afterwards. This construction builds on the Markovian Milestoning with Voronoi Tessellation of Vanden Eijnden and Venturoli⁴¹ in which a trajectory confined to a cell generates the necessary first-hitting statistics. Our Milestoning formulation⁴² is different and emphasizes short trajectories between interfaces. In the present study we consider another variant of Milestoning in which long trajectories cross the interfaces and Milestoning is used for analysis (and is in principle exact).

In contrast to our previous study, the milestones considered here are directional. In the earlier definition of the milestones⁴³ it was sufficient to record the passage of a milestone with no reference to the direction of crossing. For example, the milestone would be the same for a system that crosses from a hydrogen bonded to a dissociated state, as in the opposite scenario when the system crosses from a dissociated to a hydrogen bonded state. This mapping is expected to be valid when the dynamics are stochastic and overdamped as is typical for thermal motions of macromolecules. However, Directional Milestoning, which was introduced recently (DiM⁴²) offers more flexibility. It retains the direction of the crossing,

and allows (for example) the forward and backward direction to differ in their geometric definition.

3. Milestoning equations

The Milestoning trajectories that are the outcome of the coarse-grain projection discussed above are used to generate the fundamental quantities of the Milestoning theory. A critical function to Milestoning is $K_{ij}(t)$. The transition kernel $K_{ij}(t)$ is an “ $n \times n$ ” matrix (where “ n ” is the number of milestones) that contains the probability of transitioning from milestone “ i ” (given that we start in “ i ”) to milestone “ j ” and is estimated by $n_{ij}(t)/n_i$, in which $n_{ij}(t)$ is the number of trajectory fragments transitioning from milestone “ i ” to milestone “ j ” (passing milestone “ j ” for the first time while the previous milestone passed is “ i ”) at time “ t ” and n_i is the number of trajectories initiated at milestone i .

If we focus on equilibrium properties and on the Mean First Passage Time (the mean of the times that it takes the system to reach a pre-defined target for the first time: MFPT), then only the zero and the first moments of $K_{ij}(t)$ as a function of time are required. The zero moment, which we call the *transition matrix*, is $K_{ij} = \int_0^\infty K_{ij}(t) dt$ and the first moment is $\langle t_{ij} \rangle = \int_0^\infty t K_{ij}(t) dt$. We use below the average lifetime of a milestone, $\langle \tau \rangle_i$. It is the time that the trajectory spends after hitting milestone “ i ” before crossing (any) other milestone. It is defined by $\langle \tau \rangle_i = \sum_j \langle t_{ij} \rangle = \sum_j [\int_0^\infty t \cdot K_{ij}(t) dt]$. Thus, the system is reduced from an atomic description to a lower dimensionality mesh in coarse-grained space considering only the initiating milestone, terminating milestone, and time.

These two fundamental quantities are then used to calculate the Mean First Passage Time (MFPT or $\langle \tau \rangle_{if}$) from milestone i to the final (assumed absorbing) milestone f by

$$\langle \tau \rangle_{if} = p_i \sum_j [\mathbf{I} - \mathbf{K}]_{ij}^{-1} \langle \tau \rangle_j. \quad (1)$$

The use of the index “f” may be confusing since it appears explicitly only on the left side of the equation. However, the final state “f” is special since it is absorbing and the transition matrix was modified accordingly (i.e., $(\mathbf{K})_{ff} = 0 \forall j$). For more details on the construction of the boundary conditions of Eq. (1) see Ref. 42. The initial milestone distribution p_i is the initial ($t = 0$) probability density, $\mathbf{I}$ is the identity matrix and K_{ij} is an element of the transition matrix ($(\mathbf{K})_{ij} = \int_0^\infty K_{ij}(t) dt$). In order to model an absorbing boundary condition at f we set $K_{fi} = 0 \forall i$. Hence, trajectories that arrive to the final state vanish. We note the important difference between $\langle \tau \rangle_j$ and $\langle \tau \rangle_{if}$: the former is the average lifetime of individual milestones (j) and the latter is the Mean First Passage Time between the initiating milestones (i) and absorbing milestones (f), respectively.

To obtain stationary flux we set the elements of the transition matrix to $K_{fi} = 0 \quad i \neq 1$ and $K_{f1} = 1$. This choice creates a cycle in which trajectories that make it to the absorbing boundary at f are immediately moved back to the origin. The transition matrix and average milestone lifetimes can also be used to calculate the stationary probability distribution $p_{i,stat}$ of being in a state i (the last milestone that was passed is i) as well as the stationary reactive flux vector ($\mathbf{q}_{stat}$). The stationary reactive flux vector is calculated directly from the linear equation below or alternatively as the eigenvector with zero eigenvalue of the matrix $\mathbf{I} - \mathbf{K}$.

$$\mathbf{q}_{stat} (\mathbf{I} - \mathbf{K}) = 0. \quad (2)$$

The stationary probability distribution is a function of both the transition matrix—via the stationary reactive flux vector—and the average milestone lifetime:

$$p_{i,stat} = (\mathbf{q}_{stat})_i \cdot \langle \tau \rangle_i, \quad (3)$$

where the multiplication is of the i th elements and is not a vector dot product. Note that the stationary probability distribution is not normalized, as written. This is because it is determined from the flux vector (a solution of Eq. (2)) which is known up to a positive scaling factor. The normalized probability distribution is simply

$$p_{i,stat} = \frac{(\mathbf{q}_{stat})_i \cdot \langle \tau \rangle_i}{\sum_i (\mathbf{q}_{stat})_i \cdot \langle \tau \rangle_i}. \quad (4)$$

C. Statistical analysis of the Milestoning network

At the core of the Milestoning calculations one finds the matrix K_{ij} . The overall accuracy of the calculation is determined by the computational precision of this matrix. Here we propose a statistical model to assess this uncertainty. The matrix element K_{ij} is a random variable that is obtained as a sum of other random variables. The elementary event that is used

to compute K_{ij} is the outcome of a trajectory. A trajectory initiated at milestone i may (or may not) transition to milestone j . Hence, for this single event K_{ij} can be zero or one. As more events are collected we improve our estimate of K_{ij} as an average of zeroes and ones. We write

$$\langle \mathbf{K}_{ij} \rangle \cong \frac{1}{n_i} \sum_{l=1, \dots, n_i} \delta_{k(l),j} = \langle \delta_{k(l),j} \rangle, \quad (5)$$

where n_i is the number of trajectories we have initiated at milestone i , l is the trajectory index, $k(l)$ is the outcome, terminating milestone of trajectory fragment l , and $\delta_{k(l),j}$ is the Kronecker delta function which is zero if the terminating milestone, k , of the l th trajectory fragment is different from j . It is one otherwise.

As is typical for statistical mechanics calculations, the trajectories are assumed to be independent and are sampled from the correct distribution. The typical number of trajectories fragments used to generate the Milestoning quantities for each milestone is between several tens and several hundreds. Following the Central Limit Theorem, we can therefore assume that K_{ij} is sampled from a normal distribution for a sufficiently large number of initial trajectories. We estimate the parameters of the normal distribution (mean, μ and variance, σ^2) from the trajectory data.

$$\mu_{ij} = \mathbf{K}_{ij}, \quad (6)$$

$$\sigma_{ij}^2 = \frac{1}{n_i} \sum_l \delta_{k(l),j}^2 - \left(\frac{1}{n_i} \sum_l \delta_{k(l),j} \right)^2 = \mathbf{K}_{ij} - \mathbf{K}_{ij}^2.$$

The variance of the average $\langle K_{ij} \rangle$ is σ_{ij}^2/n_i . Of course, we must also satisfy the intrinsic constraint that the probability of the trajectory to go somewhere is one: $\sum_j \mathbf{K}_{ij} = 1$. We impose this constraint by sampling the individual elements of the K_{ij} completely independently from the normal distribution for each individual element $N(\mu, \sigma^2/N)$ and we normalize each of the elements by the sum of the sampled elements of a given row of the matrix (i.e., for each milestone). Hence from a complete row of M milestones only $M - 1$ elements are independent. A similar procedure is used for the milestone lifetime $\langle \tau \rangle_i$ where all the elements of the life time vector are assumed independent and normalization is not required. This sampling procedure was repeated 1000 times for each load level, providing a distribution of 1000 MFPT values and associated quantities derived from the transition matrix and milestone lifetime vector (such as the milestone probability, $p_{i,stat}$, and stationary flux, $\mathbf{q}_{stat}$).

D. Coarse graining approach

Milestoning requires the generation of trajectories in which an atomically detailed trajectory (as generated by the simulation software) is mapped to the coarse-grained space. The mapping implemented here is based on the “coarse anchor” definitions used previously. Briefly, the anchors defined here on a single-residue basis are a function of (a) the Ψ dihedral angle value and (b) the quantity and type of hydrogen bonds spanning across the residue. These degrees of freedom were previously shown to provide a useful proxy for the

system structure; as a result, other metrics (such as the Φ dihedral angle) are ignored in the present analysis. We do not differentiate between different residues along the helix. Instead we pick the first residue that cracks.

The hydrogen bonds considered to “span” the residue of interest (where “spanning the residue” means that the carbonyl-group donor precedes and the amide-group acceptor succeeds the residue of interest) can take one of three forms: 3_{10} , α -, and π -helical in which the carbonyl group from residue “ i ” (which can be any of a number of residues that precede the residue of interest) forms a hydrogen bond with the amide group of residue “ $i+3$,” “ $i+4$,” or “ $i+5$,” respectively (which could be any of a number of residues that succeed the residue of interest). The hydrogen bonds are considered in addition to the Ψ dihedral in order to incorporate the requirement that the hydrogen bonds spanning a residue must, in the vast majority of cases, be broken to permit the flexibility in the Ψ angle to allow the initiation of unfolding. Note that the “spanning” condition allows a number of possible bonds to span a residue at any point in time; specifically, there is a maximum of 2, 3, or 4 hydrogen bonds spanning a residue if the bonds at a specific time are 3_{10} , α -, and π -helical in nature, respectively. Note that for the purposes of this analysis a hydrogen bond is considered to exist if the N–O distance is <0.35 nm and the angle between the HN and NO vectors is $>150^\circ$.

The full anchor definition (Table II) requires considering both the value of the Ψ dihedral angle as well as the type and quantity of hydrogen bonds spanning the residue at a given point in time. Thus, there can be “pure” 3_{10} , α -, and

TABLE II. Anchor definitions as a function of the quantity and type of hydrogen bond spanning the residue of interest and the Ψ dihedral angle. The keyword “any” indicates that any quantity of a given type of hydrogen bond may be present for a given anchor; for example, the 3_{10} state can have “any” number of 3_{10} -helical bonds (for this anchor, that would be either 1 or 2 bonds). Note that certain anchors allow a variable quantity of hydrogen bonds; for example, the $\pi 34$ anchor can have either 3 or 4 π -helical hydrogen bonds. These variable-bond-quantity anchors are not to be confused with milestones, which are defined to be the interface between domains of any two anchors listed here. The full atomically detailed trajectories computed with GROMACS addressed in Sec. II A are mapped to Milestoning trajectories and are used to compute K_{ij} and $\langle \tau \rangle_i$.

Anchor designation	Quantity of hydrogen bonds			Ψ angle values (deg)
	3_{10}	α	π	
$\Psi > 90$	Any	Any	Any	$>90, <-150$
$90 > \Psi > 0$	Any	Any	Any	$0-90$
none	0	0	0	<0
3_{10}	Any	0	0	<0
$3_{10}/\alpha 2$	Any	2	0	<0
$3_{10}/\alpha 1$	Any	1	0	<0
$\alpha 3$	0	3	0	<0
$\alpha 2$	0	2	0	<0
$\alpha 1$	0	1	0	<0
$\pi/\alpha 3$	0	3	Any	<0
$\pi/\alpha 2$	0	2	Any	<0
$\pi/\alpha 1$	0	1	Any	<0
$\pi 12$	0	0	1 or 2	<0
$\pi 34$	0	0	3 or 4	<0

π -helical states in which only one type—but possibly different quantities—of bonds are formed and there can be “mixed” states in which multiple types and quantities of bonds are formed, such as the $\pi/\alpha 1$ state in which one α -helical bond and at least one π -helical bond span the residue of interest. Additionally, there is a “none” state in which no hydrogen bonds are formed but the Ψ angle does not exceed 0° . The value of the Ψ angle is used to define when a residue has initiated unfolding—and is therefore at the final state for the present analysis—by reaching a value greater than 90° ; an intermediate unfolded state is defined by $0^\circ < \Psi < 90^\circ$. See Table II and Ref. 15 for a full list of the anchors and their definitions.

E. Pathway calculations

Given the K_{ij} matrix and the Milestoning solution for the kinetics and thermodynamics we can represent the process as a network in which each edge on the network connects two anchors and the weight of the edge is the net flux. The flux is the number of trajectories that passes through that milestone in unit time. This network can be analyzed to determine pathways of maximum flux to obtain qualitative insights into mechanisms. The continuous MaxFlux theory⁴⁴ and algorithm^{45,46} were previously discussed. We used an exact and discrete MaxFlux approach⁴⁰ to determine the maximum flux pathways from the folded ($\alpha 3$) to unfolded ($\Psi > 90$) states on the network.¹⁵ Recall that the transition matrix allows the calculation of the stationary reactive flux vector (q_{stat}). When using directional milestones there are two milestones ($\alpha\beta$ and $\beta\alpha$) between anchors “ α ” and “ β ,” each with their own stationary reactive fluxes: $q_{\alpha\beta}$ and $q_{\beta\alpha}$. The difference between these two fluxes is the net flux and the weight along the edge connecting α and β : $w_{\alpha\beta} = q_{\alpha\beta} - q_{\beta\alpha}$. Negative $w_{\alpha\beta}$ corresponds to a net flux in which trajectories transition from β to α more frequently than from α to β . A weight matrix w of maximal dimensionality $n_{anchors}^2$ can be determined from q_{stat} ; it is this weight matrix that is used for the pathway calculations. Note that in this paper the stationary reactive fluxes are defined in milestone space while the weight matrix is in anchor space. Elements from this weight matrix can be referred to as $w_{\alpha\beta}$ or w_i since milestone “ i ” is composed of the interface between anchors “ α ” and “ β .”

1. The MaxFlux algorithm

The MaxFlux algorithm,⁴⁵ as implemented for a Milestoning graph,⁴⁰ seeks a path that connects the reactant and product states and that carries maximum flux. A related algorithm⁴⁷ was proposed in the context of the Markov State Model.^{48–50} As used in the current study, the following algorithm finds the MaxFlux path:

- Step 1. Find the link e_{min} with the smallest weight $w_{min} = \min_i(w_i)$ on the graph G (corresponding to the edge/milestone with the smallest flux) and mark it for elimination. Note that the weight w_i consists of connections between two anchors, e.g., k and l . If no edges on graph G remain that have not been marked in Step 2 as part of the path, then **stop**.

Step 2. Check if, after removing $e_{\min}$, it is still possible to find a sequence of edges leading from the reactants to the products. If yes, then remove $e_{\min}$ from the graph and return to Step 1. If no, then store $e_{\min}$ as an edge on the desired path and remove it from consideration of the minimum weight edge (Step 1). Proceed to Step 1.

The collection of edges that remain after all non-marked edges are removed is the “MaxFlux path,” and is guaranteed to provide the maximum flux path in the discrete space generated in the Milestoning calculations. Finally, a “transition state” can be chosen as the edge along a MaxFlux path with the smallest weight (net flux). Additional paths, e.g., second best, third best, etc., may be found by eliminating the “transition state” in the first-order (i.e., MaxFlux) path from consideration and repeating the algorithm.

F. K_{ij} correlation calculation

In contrast to pathway calculations, which generate connections between edges to define the course that a reaction will take, the K_{ij} correlation calculation illustrates the significance of individual elements of the transition matrix, and therefore individual edges of the graph. Recall that the MFPT calculation scheme involves a sampling of the individual elements of the transition matrix (K_{ij}) and milestone lifetime vector ($\langle\tau\rangle_i$) to calculate a single MFPT. This sampling procedure is repeated “n” times for each load level (where “n” is 1000 for the present study), providing a set of MFPT values ($\langle\tau\rangle_{if}^{(n)}$) corresponding to a set of transition matrices ($K_{ij}^{(n)}$) and a set of milestone lifetime vectors ($\langle\tau\rangle_i^{(n)}$). The correlation is calculated between the set of MFPT values and the set of values for a given transition matrix element. In this way it is possible to determine whether high values for a given transition, such as that corresponding to the rotation of the Ψ angle beyond 90° , correlate to high or low MFPT values. Note that this calculation does not prove that high values of a given transition matrix element are *responsible* for the high or low MFPT value. Instead, this correlation shows that the MFPT is sensitive to the topology of edges. For example, sensitivity to specific K_{ij} may indicate the opening (or closing) of an off pathway intermediate, which influences the overall rate (but is not on the MaxFlux pathway).

G. K_{ij} sensitivity calculation

In Sec. II F, we explored stochastic changes for the whole matrix, which can impact simultaneously a large number of elements. It is also of interest to examine the impact of small local variation of the transition matrix K_{ij} . In order to determine the effect of changing individual K_{ij} elements on the calculated MFPT, the sensitivity of the MFPT to K_{ij} was determined:

$$\Omega_{ij} = \frac{1}{\langle\tau\rangle_{if}} \delta \langle\tau\rangle_{if} / \delta K_{ij}. \quad (7)$$

As explained in Eq. (1) we are primarily interested in the overall mean first passage time from the initial milestone “i” to the

special absorbing state “f.” The absorbing process is defined in the transition matrix $\mathbf{K}$.⁴² Here we examine the relative sensitivity of the mean first passage time to values of the $\mathbf{K}$ matrix. This partial derivative was numerically calculated by adjusting the transition matrix element of interest while keeping all other transition matrix values fixed and computing the resulting MFPT. The transition matrix element was adjusted as a percentage of its average value (75%, 80%, 85%, 90%, 95%, 100%, 105%, 110%, 115%, 120%, and 125%) with MFPTs computed for each adjustment. Normalization of the rows of the K_{ij} matrix was maintained by renormalizing the other entries in the i th row. A linear regression was done to the resulting MFPT and K_{ij} element pairs, with the slope of this regression providing Ω_{ij} . Note that this calculation is in contrast to the correlation analysis which allows every element of the transition matrix to change.

III. RESULTS

A. MFPT of the initiation of helix unfolding

The initiation of unfolding as a function of applied load magnitude followed a non-monotonic curve in which the initiation of unfolding was longest (~ 3.5 ns) for loads in the 15–30 pN range (Fig. 3). The unfolding time for loads up to 10 pN was approximately 1 ns less (~ 2.5 ns) than this peak value. The MFPT followed a monotonic profile for loads above 40 pN, decreasing with increasing load for both 70 pN (~ 2.0 ns) and 100 pN (~ 1.0 ns). Critically, this non-monotonic MFPT vs. load curve follows a profile that is similar in nature to the lifetime of inter-molecular catch bonds. The statistical sampling approach provides an estimation of confidence in the MFPT calculated for each load level, clearly indicating that the MFPT results presented are statistically ($p < 0.005$) different between four load levels (0 pN, 25 pN, 70 pN, and 100 pN) that will be the focus of the following analysis.

The calculated MFPT values using Milestoning favorably compare to the estimate of the MFPT based on observed times from the straightforward molecular dynamics calculations. Note that the present implementation of Milestoning differs from earlier implementations of Milestoning that focused on more efficient calculations by using trajectory fragments.⁵¹ Here, an entire, initial-condition trajectory is post-processed according to the Milestoning methodology. Thus, the assumption of de-correlation of trajectory fragments is not necessary. The agreement in Fig. 3 demonstrates the fidelity of the quantities used to calculate the MFPT—namely the transition matrix and average milestone lifetime—and increases confidence in the accuracy of the results presented below that rely on analysis of these fundamental quantities.

B. Milestone probability distributions

The milestone stationary probabilities were calculated for each MFPT calculation using Eq. (3). The averages of these probabilities over all MFPT samples for a given load level were plotted to illustrate the multi-dimensional nature of the energy landscape explored during the initiation of unfolding. The edge-specific differences in the average probabilities between load levels provide an indication of the shift in

FIG. 3. Simulation time of first observed initiation of helical unfolding (“MD MFPT”) in straightforward MD simulations and MFPT of helical unfolding initiation (“Milestoning MFPT”) calculated via Milestoning analysis of the simulation trajectories. Observed simulation times are computed directly from multiple molecular dynamic trajectories at a given load by recording the first observation of $\Psi > 90^\circ$; data points are the averages and error bars are the standard deviations of these observations. Data points and error bars for the Milestoning MFPT are the averages and standard deviations from the Milestoning sampling protocol.

system behavior as a function of load. Note that the stationary probabilities underlying these calculations are a function of both the milestone connectivity (as captured by the transition matrix) as well as the local milestone kinetics (as captured by the average milestone lifetime vector).

While the native, α -helical states are the most probable ($p_{\alpha 3:\alpha 2} = 0.63$ at 0 pN), both the π -helical ($p_{\pi 12:\pi 34} = 0.03$ at 0 pN) and 3_{10} -helical ($p_{\alpha 1:3_{10}/\alpha 1} = 0.01$ at 0 pN) states are also well visited. Thus, prior to the initiation of unfolding, the system explores roughly the entire energy landscape of helical states. While the relative distribution of this exploration is somewhat load dependent (see below), this wide exploration is seen for all load levels below 100 pN. This exploration is indicative of the highly diffusive nature of the state dynamics prior to unfolding, even in the presence of relatively large (≤ 70 pN) loads.

Differences in the probability networks illustrate the shift in the multi-dimensional energy landscape with different magnitudes of applied load. The comparison between the 0 pN and 25 pN distributions illustrates that the system explores the π -helical states more at 25 pN than in the absence of applied load. As a result, there is a lower probability for the states nominally close to unfolding, including the connections between the 3_{10} , $\alpha 1$, and *none* states. In contrast, the 70 pN load level has a much lower probability for the π -helical states than does the 25 pN. Instead, the system under 70 pN has a much higher probability of staying in the native state or exploring the 3_{10} -helical states. Thus, the 0 pN–25 pN and 25 pN–70 pN comparisons illustrate that the 25 pN load level has a higher tendency to visit the π -helical states than either lower (0 pN) or higher (70 pN) loads. Interestingly, the lack of a distinct division between regions of the graphs in the 70 pN–100 pN comparison shows the relative similarity between these two load levels, with the 100 pN providing a still-higher probability of remaining in the native α -helical states

prior to the initiation of unfolding than even the 70 pN. Perhaps not surprisingly, the diffusive behavior is reduced at the highest load level as the reaction space is more focused in the neighborhood of direct paths.

C. Net flux networks

Net flux networks are calculated for each load level by taking the difference between the fluxes of the two directional milestones between each anchor pair (see Fig. 5). These networks indicate any bias in the diffusive exploration of the accessible states of the system given the absorbing condition at milestone “f.”

As with the probability and total flux networks, the net flux networks illustrate the diffusive nature of the system evolution at low loads (0 pN and 25 pN), with more directed pathways present at the higher loads (70 pN and 100 pN). Notably—and as was seen with the total fluxes—the net fluxes of the π -helical states have a higher net flux for the intermediate 25 pN load than is seen in the 0 pN load. Indeed, a circulation appears in the 25 pN net flux network between the $\pi 12$, $\pi 34$, and $\pi/\alpha 1$ anchors that may contribute to the higher milestone probabilities seen for the milestones connecting these anchors.

The connections to and from the $\alpha 1$ state provides a notable comparison between the net flux networks of the different load levels. In the absence of load (0 pN), the $\alpha 1$ anchor has a tendency to form a companion 3_{10} -helical bond or experience a rotation of the Ψ angle beyond 0° . Additionally, the connection between the [*none*] and $\alpha 1$ anchors is actually in favor of the re-formation of the α -helical bond. The 25 pN net fluxes from the $\alpha 1$ anchor indicate that the lone hydrogen bond in this state is expected to break, leading to a relatively strong connection to the [*none*] anchor in which there are no intact hydrogen bonds. At the higher load levels, the evolution of the system from the $\alpha 1$ state is to simultaneously break the α -helical bond and form a 3_{10} -helical bond. Note that the 3_{10} -helical hydrogen bond is formed between a donor and acceptor that are “closer” along the peptide backbone (between residues i and $i+4$ for the α -helical bond while the 3_{10} -helical bond is formed between residues i and $i+3$). Additionally, our previous study of the Ψ dihedral angle distributions as a function of anchor indicate that the Ψ dihedral angle minimum for a 3_{10} -helical structure is less negative—and therefore closer to unfolding—than the Ψ dihedral angle minimum for the α -helical structure. Thus, the connections from the $\alpha 1$ anchor indicate that at relatively large loads (70 pN and 100 pN), the progression to unfolding is one in which work is done against the rotational degrees of freedom along the backbone of the helix in order to effectively “stretch” the helix, resulting in 3_{10} -helical hydrogen bond donors and acceptors that are in closer proximity than at intermediate load levels (25 pN), where the applied load acts to break the remaining α -helical hydrogen bond.

D. Average milestone lifetime networks

Networks of the average lifetimes of individual milestones were graphed to provide an indication of the

FIG. 4. Milestoning probabilities of networks of different loads are provided. The plot (a) includes the network of probabilities of different edges for zero load, which is also used as the reference distribution. The rest of the networks are for differences in probabilities of edges of networks of a high and low (reference) load network. Positive values (red) mean that the probability is increased for the higher load level. Only probabilities above 0.001 are presented; others are omitted for clarity. Line thickness for (a) is determined by the $\log_{10}$ of probabilities; line thickness for (b)–(d) is determined by the absolute difference in probability without any logarithmic adjustment. Each edge is the sum of the probabilities of the two directional milestones connecting the two relevant anchors. Note that as the load increase from zero to 25 pN the contribution of the π helix to the network is increasing, while it is reduced when the load grows to 70 pN. Turning on and off the contribution of the off the pathway π helix explains the non-monotonic behavior of the MFPT. The absence of connections to the $\Psi > 90^\circ$ anchor reflects the fact that the analysis of the simulations only considered the trajectory up to the first transition to this anchor. Therefore, the milestones connecting to this final anchor have an undefined lifetime, represented here as a lifetime of zero, (see Fig. 6). The milestone probability is also zero despite the presence of significant fluxes through these milestones (see Fig. 5).

relationship between the kinetics of local transitions between anchors (see Fig. 6). These average lifetimes are one of the fundamental quantities used in the calculation of the MFPT that are directly calculated from the simulation. Differences in the average lifetimes are used to illustrate the effect of load increases in speeding or slowing the local dynamics.

The average milestone lifetimes are three orders of magnitude less than the overall MFPT of unfolding initiation. For instance, the average lifetime of the $\alpha 3:\alpha 2$ milestones are ~ 6 ps (in contrast to the ~ 1 – 3 ns of the MFPT). The differences in these time scales underscore the diffusive nature of the process such that no one transition dominates the overall process. It also illustrates the promise of Milestoning in computing long time behaviors from short-time easy-to-compute trajectory fragments.

Differences in the average milestone lifetimes reveal the extent to which these local kinetics change with higher load levels. While 25 pN has increased lifetimes of some milestones relative to the unloaded scenario, increasing load levels beyond 25 pN clearly decrease the majority of milestone lifetimes. Two key milestones with increased lifetimes for 25 pN relative to 0 pN are the $\pi 12:90^\circ > \Psi > 0^\circ$ and $3_{10}/\alpha 1:90^\circ > \Psi > 0^\circ$ milestones. While these particular milestones are certainly near the unfolded state, the increase in time for these milestones from 0 pN to 25 pN (the change in lifetime from 0 pN to 25 pN: $\pi 12:90^\circ > \Psi > 0^\circ = 6.4$ ps, $3_{10}/\alpha 1:90^\circ > \Psi > 0^\circ = 1.9$ ps) is still three orders of magnitude less than the increase in overall MFPT between these load levels (1.2 ns), indicating that changes in the local lifetimes do not directly account for the observed differences in MFPT. Only the complete description of the network including diffusive

FIG. 5. Net flux networks as a function of applied load magnitude. The net fluxes are calculated as the difference between the directional milestone fluxes for a given pair of anchors. The directional milestone fluxes are calculated via solution of $\mathbf{q}_{stat}(\mathbf{I} - \mathbf{K}) = 0$. The edge thickness corresponds to the magnitude of the net flux, with the edges of each network normalized by the highest net flux for a given load level. The direction of the edges corresponds to the directionality of the net flux. Edges with normalized values less than 10% of the maximum net flux are omitted for clarity.

dynamics along many edges can describe the MFPT shifts quantitatively.

E. K_{ij} sensitivity analysis

Correlations between the sampled K_{ij} element values and the calculated MFPT values were determined for each load level (see Fig. 7). Positive correlations indicate that increases in the K_{ij} element value were correlated with increased MFPT, and therefore the increase in K_{ij} slowed initiation of unfolding; vice versa for the negative correlations (Fig. 8). Plots of these correlations illustrate individual transitions that have a retarding influence on the initiation of unfolding. Note that a K_{ij} element links two milestones (i and j), and therefore three anchors ($\alpha:\beta:\gamma$ where milestone i is the interface between anchors α and β and milestone j is the interface between β and γ and anchor β is the shared connection between the two milestones); these plots illustrate the correlation for a milestone transition by showing an edge between the first and third anchor.

The highest correlations between K_{ij} element value and MFPT for all the load levels occur for the K_{ij} elements connecting the $\Psi > 90^\circ$ state with the 310 , $\alpha 1$, and $[none]$ states.

Each of these transitions pass through the $90^\circ > \Psi > 0^\circ$ intermediate state on their way to the unfolded state. These three transitions represent all the final transitions. Interestingly, the values of these transition elements monotonically increase with applied load magnitude (see Fig. 9). Thus, while the correlation analysis indicates that the final transitions are correlated with the calculated MFPT for a given load level, these final transitions do not solely determine the MFPT. If the final transition was a true rate-limiting step then the calculated MFPT with respect to load would be expected to monotonically follow the transition probabilities of the final transitions, which is clearly not the case.

In addition to the correlations for the final transitions, the overall connectivity of the correlation graphs is noteworthy, particularly for the 25 pN. The large quantity of transition elements whose value correlates with MFPT underscores the highly connected nature of the transition matrix in that many transitions—rather than a few, isolated, well-connected pathways—influence the MFPT calculation. Further, the highly correlated nature of the 25 pN graph illustrates the wide variety of states visited at this intermediate load level relative to the lower and higher loads. Additionally, the modest—yet still significant—values of the correlations further emphasizes the multi-dimensionality of the

FIG. 6. The average milestone lifetimes for the 0 pN (a) network as well as comparisons between 0 pN–25 pN (b), 25 pN–70 pN (c), and 70 pN–100 pN (d). The comparisons made in (b)–(d) are the arithmetic differences between the milestone lifetimes of the two load levels. For example, in (b) the edge values plotted are calculated as $\Delta \langle \tau \rangle_i^{0\text{pN}-25\text{pN}} = \langle \tau \rangle_i^{0\text{pN}} - \langle \tau \rangle_i^{25\text{pN}}$. The average milestone lifetime values graphed here are the vector elements $\langle \tau \rangle_i$. For clarity, only milestone lifetimes above 0.5 ps are presented. Note that the lifetimes of the milestones are presented as the edges between anchors, which serve as the nodes of the network. Line thickness corresponds to the magnitude of the milestone lifetime; where relevant, the color corresponds to increasing or decreasing lifetime with respect to load. Dashed lines are renderings of relatively thin, and therefore short-lived, edges. The absence of connections to the $\Psi > 90^\circ$ anchor reflects the fact that the analysis of the simulations only considered the trajectory up to the first transition to this anchor. Therefore, any milestones representing connections to this final milestone have an undefined lifetime, represented here as a zero lifetime.

transition network as very high correlations would imply that a few, well-connected pathways dominated.

In addition to the negatively correlated K_{ij} elements that are associated with speeding the initiation of unfolding, a set of positively correlated connections represent the milestone transitions associated with slowing the initiation of unfolding. It is noteworthy that a large proportion of these retarding transitions represent bond re-formation. The relative importance of bond re-formation transitions further indicates that, even in the presence of large forces, the unfolding of the helical system is diffusive, rather than highly directed process.

The MFPT sensitivity calculation displays a strongly negative correlation to the transition matrix K_{ij} elements with the strongest negative correlation to the MFPT (i.e., elements connecting the $\Psi > 90^\circ$ state with the 310 , $\alpha1$, and $[none]$ states) (see Fig. 11 for select sensitivity values). Noticeably, this sensitivity is highest for the zero load case (0 pN) in which the unfolding event is expected to be dependent on

thermal motion to significantly perturb the helix. Conversely, the sensitivity to individual elements of the transition matrix is lowest for the highest load level (100 pN), an observation that is supported by the lack of significant correlation between these K_{ij} elements and the MFPT (Fig. 8). This difference between the sensitivity for the zero and high loads suggests that the highest load level has multiple parallel pathways available, and therefore a perturbation to one pathway is compensated by the other pathways. While the zero load case may have additional pathways available, the values of the K_{ij} matrix elements are 3–5 $\times$ lower than those in the highest load.

The sensitivity of individual K_{ij} element values associated with breaking or forming native α -helical hydrogen bonds reveals an asymmetry in which the MFPT calculation is 3–5 $\times$ more sensitive to perturbations in the K_{ij} elements representing bond breakage. This asymmetry may suggest that bond re-formation can occur through a variety of means, such as through non-native intermediates. In contrast to the final transitions (Fig. 10), the transitions through the

FIG. 7. The correlations of individual transition element values with the calculated MFPT. Correlations are calculated between the set of values for a given transition element and the set of calculated MFPTs during the sampling procedure used for the MFPT calculation at each load level. The correlations are displayed as the outer anchor connection between two milestones; the transition anchor that links the two milestones together is omitted for clarity. Only correlations that are statistically significant ($p < 0.05$) are displayed. The thickness of the lines corresponds to the correlation coefficient.

different quantities of native α -helical bonds are less sensitive to load. Note the relatively extreme insensitivity of the MFPT calculation to the formation of π -helical bonds, even for the 25 pN intermediate load level. We emphasize that the sensitivity analysis is local and is in the neighborhood of a particular value. If the transition probability of a particular edge is very high then the MFPT may be insensitive to changes in K_{ij} (that channel is already open) while still making significant contribution to the MFPT.

F. MaxFlux pathway probability networks

MaxFlux pathways were calculated to determine the most likely unfolding pathways for each load level. For each load level and each sampled K_{ij} (and therefore each MFPT calculation) a total of three MaxFlux pathways were determined, providing a set of MaxFlux pathway triplets for each load (see Fig. 7). The probabilities of a given edge appearing on one of these three MaxFlux pathways was then determined for each load level; an edge appearing in more than one pathway for a given sampling of K_{ij} was only counted once. These percentages indicate the most likely MaxFlux pathways

for a given load level. The density of the graph thus indicates the multitude of parallel unfolding pathways possible for each load level.

Multiple unfolding pathways were present for all load levels, indicating that the unfolding process is a competition between parallel unfolding pathways. The overwhelming majority of unfolding pathways pass through the $90^\circ > \Psi > 0^\circ$ state before the final transition to the $\Psi > 90^\circ$ state. However, there are multiple approaches to this penultimate $90^\circ > \Psi > 0^\circ$ state, with highly probable transitions from the 310 , $\alpha1$, and $[none]$ states. Note that this is the same triplet of connections whose K_{ij} elements are highly correlated with the MFPT as discussed above. The parallel nature of the pathways is further reinforced in the connections to this triplet of states that connect with the $90^\circ > \Psi > 0^\circ$ state.

A pathway “transition edge” is defined as the edge with the lowest weight (net flux) on a given pathway. The average weights of the transition edges were calculated for the three pathways of each load level. These transition edge weights were strongly anti-correlated with the overall MFPTs with calculated correlation coefficients of -0.91 , -0.88 , and -0.86 for Paths 1–3, respectively. Thus, the magnitudes of the

FIG. 8. Correlation between the sampled K_{ij} element value and calculated MFPT for the transition from the $310:90^\circ > \Psi > 0^\circ$ milestone to the $90^\circ > \Psi > 0^\circ: \Psi > 90^\circ$ milestone at 0 pN. The correlation value for this set of data was -0.41 .

rate limiting transitions of the MaxFlux pathways are strongly related to the overall MFPT calculation. Note that these transition weights consider the stationary flux as calculated from the whole transition matrix. Thus, in contrast to the transition probabilities K_{ij} for the highly correlated final transitions discussed above (Fig. 13), the transition weight elements incorporate effects from the whole network.

Interestingly, the 25 pN (and 0 pN to a lesser extent) pathways have noticeable, if not dominant, passages through the π -helical states. These π -helical transitions are entirely absent from the 70 pN and 100 pN MaxFlux pathways. Critically, the edge weights to the $90^\circ > \Psi > 0^\circ$ are the lowest from the $\pi 12$ state relative to the other major connections to the $90^\circ > \Psi > 0^\circ$ state (Fig. 14). Thus, the 25 pN MaxFlux pathways traverse through the states with the lowest net flux

FIG. 9. Transition matrix elements with strong negative correlations with the calculated MFPT; these elements represent the final transition to the unfolded state. Values are the averages determined from the post-processing analysis of the simulations and represent the mean values used in the MFPT calculation sampling procedure for each load level; error bars represent the standard deviation of these transition values throughout the sampling procedure. Note that this plot does not explain the non-monotonic behavior of the MFPT.

FIG. 10. Sensitivity (Ω_{ij}) of the MFPT calculation to changes in the transition matrix elements with strong negative correlations to the MFPT. The sensitivity is numerically calculated as the slope of the linear regression between MFPT and K_{ij} element values: $\Omega_{ij} = \delta \langle \tau \rangle_{if} / \delta K_{ij}$. Negative sensitivities indicate that the MFPT decreases as the K_{ij} element increases.

connections to the unfolded states, thereby slowing the average process as reflected in the calculated MFPT.

As noted above for the net flux networks (Fig. 5), the MaxFlux pathways for the higher load levels (70 pN and 100 pN) have a high probability of passing through the 310 anchor indicative of a “stretch” of the helix that positions 310 -helical hydrogen bond donor and acceptors in close proximity. In the 100 pN case, the MaxFlux pathways through the 310 -helical states are much more probable than those passing directly from the $\alpha 1$ anchor to the $[none]$ anchor (the former pathways accounting for 80%–100% of the MaxFlux pathways while the latter pathways accounting for 0%–20% of the MaxFlux pathways).

The critical edges to the $90^\circ > \Psi > 0^\circ$ state have notable geometric properties related to their Ψ dihedral angle (see Fig. 15). The most probable Ψ angle—and therefore the

FIG. 11. Sensitivity (Ω_{ij}) of the MFPT calculation to changes in transition matrix elements associated with breaking α -helical hydrogen bonds. The sensitivity is numerically calculated as the slope of the linear regression between MFPT and K_{ij} element values: $\Omega_{ij} = \delta \langle \tau \rangle_{if} / \delta K_{ij}$. Negative sensitivities indicate that the MFPT decreases as the K_{ij} element increases. Note that the filled symbols represent successive bond breakage while the open symbols represent hydrogen bond formation.

FIG. 12. MaxFlux pathway distributions representing the probability of a given edge being part of one of three MaxFlux pathways calculated for each sampled K_{ij} . Color represents the probability of an edge being part of a MaxFlux pathway; thickness is the average weight of that edge. The results explored 1000 sampled networks at each load. See text on the sensitivity analysis of the networks for more details.

minimum of the energy landscape—for the $\pi 12$ anchor was found previously to be -80° . Conversely, the most probable Ψ angle for the $\alpha 1$ anchor was found at -47° while that of the 310 anchor was -36° . Notably, the edge weights for the connections between these anchors and the $90^\circ > \Psi > 0^\circ$ state for the non-zero load levels followed the same qualitative order ($310 > \alpha 1 > \pi 12$). Therefore, a transition from the $\pi 12$ state to the $90^\circ > \Psi > 0^\circ$ state requires traversing through, on average, a much larger Ψ dihedral conformational space than is required to transition from the 310 anchor.

Interestingly, the edge weight of the $[none]:90^\circ > \Psi > 0^\circ$ transition is intermediate between the edge weight of the $\alpha 1$ and 310 transitions to the $90^\circ > \Psi > 0^\circ$ state. Not surprisingly given the unstabilized nature of the $[none]$ anchor, the Ψ angle distribution for this anchor was previously found to be wider than the distribution for any of the single-bond states, with significant overlap between with the $\pi 12$ Ψ angle distribution. Thus, while systems in the $[none]$ state do not need to break a hydrogen bond, they are free to explore a wider conformational space that includes more negative Ψ angles more reminiscent of the $\pi 12$ states. This suggests that the hydrogen bonds present in the 310 state may actually serve to

stabilize a near-unfolded state in which the Ψ dihedral angle is maintained nearer to the critical transition beyond 0° .

IV. DISCUSSION

Perhaps the most significant observation of the present simulation studies is that the force-induced “cracking” of secondary structure elements (helices) occurs via dominant reaction pathways that change with respect to load and that the corresponding MFPTs is a non-monotonic function of load. This result is supported by large statistics and many more load levels compared to an earlier study.¹⁵ By “crack” we mean the first or a few amino acids that adopt extended chain conformations in which the Ψ dihedral angle is $>90^\circ$. The most direct and experimentally testable prediction is presented in Fig. 3 in which the MFPT has a peak at 20 pN. Hence, our simulations suggest that the MFPT in the range of 0 to 20 pN is first *increasing* as a function of load and only as the load increases further the MFPT drops to lower values. We have conducted the MFPT calculations using two different simulation technologies: straightforward MD simulations and Milestoning providing further support to the convergence of

FIG. 13. Average weight (net flux) of the transition edge for the three pathways calculated for each iteration of the MFPT calculation at a given load level. The transition edge is defined as the edge along a given pathway with the lowest net flux, and therefore represents the rate-limiting transition for that pathway. Data points are the averages over all 1st, 2nd, or 3rd pathways calculated during each iteration of the MFPT calculation sampling procedure for a given load level; error bars are the standard deviation of the distribution of these weights.

the atomically detailed simulations. However, we comment that the calculations were done with only one empirical force field (GROMACS 53a6), leaving open the possibility that inaccuracies in the force field impact the observations.

While the MFPT calculations of this non-equilibrium system are most relevant to experiment it is also of considerable interest to investigate the mechanisms of force-induced cracks. For that purpose we have employed the network picture of Directional Milestoning.^{40,42} With the network of fluxes at hand we illustrate that there is no single reaction pathway, although the number of dominant paths is not large. We showed that at intermediate load levels (25 pN) (and in accord with the non-monotonic behavior of the MFPT) a new flux-channel is open allowing the formation of a π helix

FIG. 14. Edge weights (net fluxes) of four key transitions to the $90^\circ > \Psi > 0^\circ$ state. Data points are the averages of all samples of the K_{ij} matrix for the MFPT calculation; error bars are the standard deviation of the MFPT computed from the samples.

FIG. 15. Ψ angle distributions for four critical anchors whose interfaces with the $90^\circ > \Psi > 0^\circ$ state form the majority of transition edges for the MaxFlux pathways. Energies are calculated from the Boltzmann inversion of the Ψ dihedral angle distribution for each anchor ($E_{\text{anchor}}(\psi) = -\ln[p(\psi)]$). Distributions of the Ψ angle were determined in the zero load case; our previous work demonstrated that the Ψ angle distributions for these anchors were not qualitatively different at higher loads. Distributions are extracted from our previous simulations.¹⁵

(Fig. 4). At low load levels the barrier for the formation of the π helix is high and the system remains more focused on the direct path. As the load increases to 25 pN the path to the π helix is less activated (i.e., lower energy) and the system spends time at this intermediate state (Fig. 4 where red edges indicate higher probabilities). As the load increases further to 70 pN the off-the-pathway intermediate is eliminated and the networks become more direct, as expected at large stretching forces. Typical experimental (Atomic Force Microscopy) unfolding forces are 20–40 pN which is the intermediate load level in our simulations. This observation suggests that for helix cracking the mechanism uncovered by the highest load is different from experiment, which in turn is also different from the cracking mechanism at 0 pN.

Another summary of our results is provided in Fig. 12 in which we computed average MaxFlux pathways from an ensemble of networks. The networks are sampled according to the known statistical uncertainties of K_{ij} . For each sampled network we compute three MaxFlux pathways and we present the distribution of pathways for all samples at a given load. The diversity of pathways is most significant at 25 pN in which MaxFlux pathways have significant likelihood of passing through the π helix. This observation illustrates again (with a different measure) the rise and fall of the off-the-pathway intermediate as a function of the external load.

On a more technical note, we provide a number of different views of the cracking network. These different views lead to similar mechanistic conclusions and seem to support the overall argument. However, they do differ in details and therefore offer complementary views. For example, in Fig. 4 we show the probability, p_i , that the last milestone that was passed is i . For example, at no load the trajectories most strongly occupy the 3/2/1 helical hydrogen bonds. This does not mean however that these states are important for the reaction. More relevant to the reaction is the net flux as illustrated by

Vanden-Eijnden and E in their Transition Path Theory.⁵² The flux, q , (not necessarily the net flux) is given by $p/\langle\tau\rangle = q$ where $\langle\tau\rangle$ is the lifetime of the milestone. These net fluxes are presented in Fig. 6. We observe in Fig. 4 that the probability of being at the π helix increases when the load increases from 0 pN to 25 pN and the probabilities of being on the way to the crack decrease. Under the same conditions, the lifetime of the π helix increases (Fig. 6) and the lifetimes of the edges leading to unfolding decreases. Hence the net reactive-flux increases with the two measures, probability and the lifetime of the milestone, and these observations are consistent with the overall MFPT.

Another measure that we use to examine the network is sensitivity. We consider the major experimental outcome from the network (the MFPT) and its sensitivity to individual elements of the transition matrix K_{ij} . For that purpose we examined multiple networks generated according to the estimated distributions of the K_{ij} (Fig. 7). This analysis examines network stability and potentially identifies bottlenecks. For example, if the MFPT is strongly correlated with a particular K_{ij} element we may assume that this transition is critical for the overall rate and consider the corresponding edge to be a transition state. Interestingly, not a single element has a correlation coefficient larger than the absolute value of 0.5. Hence, according to this measure we expect the network to be reasonably stable with respect to perturbations and not to have a clear transition state.

A more significant correlation is obtained in the analysis of the MaxFlux pathways. Each of the sampled networks used in the calculation of the MFPT was used to determine a corresponding set of 3 MaxFlux pathways. On each of these pathways we determine a “transition state” which is the edge of the minimal flux. The net fluxes (weights) of these edges have correlation coefficients greater than 0.8 with the MFPT. This observation suggests that for the calculations of the network, the best representation is the net-flux level and that the MaxFlux pathways indeed capture the essential features of the transition. The different contributions of the “transition states” as a function of load (Fig. 13) illustrate again the “anomaly” at intermediate load levels in which the π helix makes significant contribution to the reactive flux.

Our network emphasizes fluxes and as such it is different from other networks that emphasize meta-stable states and probabilities of these states.^{48,50} Besides the above argument suggesting the flux as a natural way of describing complex kinetics, there are additional technical points of why we use a flux-based method versus equilibrium and meta-stability picture. Metastability arguments that are used in the deriving the Markov State Model are different from Milestoning networks in a number of ways. A metastable state is assumed to be in quasi-equilibrium. This requires a waiting time in the meta-stable state to allow for local equilibration. Milestoning does not require local equilibration but rather de-correlation of fluxes. Milestoning can therefore use a more fine partition of the states and shorter trajectory fragments in the analysis. This fine picture can capture mechanisms even at short times. Transition times can be estimated at temporal ranges for which metastability and local equilibrium have not been reached (yet). We have illustrated that de-correlation times

can be as short as a few hundred femtoseconds, still providing accurate overall kinetics.⁴² Hence Milestoning makes it possible to obtain accurate kinetics starting from picoseconds and going forward to exceptionally long time scales (e.g., hours.⁵³). An advantage of the Markov State Model is that the book keeping is simpler and the number of states is smaller compared to the Milestoning method. An interesting recent development is the merging of the two technologies.⁴⁹

V. CONCLUSIONS AND FINAL REMARKS

We investigated, using straightforward Molecular Dynamic simulations and Milestoning analyses, the cracking of a long helix that is used in molecular machines. The cracking under load is expected to be relevant to biophysical studies of folding and unfolding of secondary structure elements. We observed non-monotonic behavior of the Mean First Passage Time, which is a prediction of the present study that can be tested experimentally and may be related to the catch-bond phenomenon of biophysics. Using a network description of the kinetics we examined the mechanism of crack formation and identify an off-the-pathway state (the π helix) that is active at intermediate load levels and causes the non-monotonic behavior of the MFPT. Network analyses were illustrated to be useful tools in studying complex dynamics of biological molecules.

ACKNOWLEDGMENTS

R.E. acknowledges the grant support of NIH grant GM05976, NSF grant CCF-0833162, and Welch grant F-1783. S.M.K. acknowledges the support of NSF grant CBET1133351. Use of Texas Advance Computing Center (TACC) computer resources is acknowledged.

- ¹R. Czerminski and R. Elber, *J. Chem. Phys.* **92**(9), 5580–5601 (1990).
- ²G. S. Jas, W. A. Hegefeld, P. Majek, K. Kuczera, and R. Elber, *J. Phys. Chem. B* **116**(23), 6598–6610 (2012).
- ³K. Kuczera, G. S. Jas, and R. Elber, *J. Phys. Chem. A* **113**(26), 7461–7473 (2009).
- ⁴C. L. Brooks, *Acc. Chem. Res.* **35**(6), 447–454 (2002).
- ⁵C. L. Brooks, *Curr. Opin. Struct. Biol.* **3**(1), 92–98 (1993).
- ⁶C. L. Brooks and D. A. Case, *Chem. Rev.* **93**(7), 2487–2502 (1993).
- ⁷D. J. Tobias, J. E. Mertz, and C. L. Brooks, *Biochemistry* **30**(24), 6054–6058 (1991).
- ⁸E. J. Sorin and V. S. Pande, *Biophys. J.* **88**(4), 2472–2493 (2005).
- ⁹N. V. Buchete and G. Hummer, *J. Phys. Chem. B* **112**(19), 6057–6069 (2008).
- ¹⁰N. V. Buchete and G. Hummer, *Phys. Rev. E* **77**(3), 030902 (2008).
- ¹¹G. Hummer, A. E. Garcia, and S. Garde, *Proteins: Struct., Funct., Genet.* **42**(1), 77–84 (2001).
- ¹²G. Hummer, A. E. Garcia, and S. Garde, *Phys. Rev. Lett.* **85**(12), 2637–2640 (2000).
- ¹³D. De Sancho and R. B. Best, *J. Am. Chem. Soc.* **133**(17), 6809–6816 (2011).
- ¹⁴W. A. Hegefeld, S. E. Chen, K. Y. DeLeon, K. Kuczera, and G. S. Jas, *J. Phys. Chem. A* **114**(47), 12391–12402 (2010).
- ¹⁵S. M. Kreuzer, R. Elber, and T. J. Moon, *J. Phys. Chem. B* **116**(29), 8662–8691 (2012).
- ¹⁶E. V. Sokurenko, V. Vogel, and W. E. Thomas, *Cell Host Microbe* **4**(4), 314–323 (2008).
- ¹⁷R. R. Isberg and P. Barnes, *Cell* **110**(1), 1–4 (2002).
- ¹⁸W. E. Thomas, L. M. Nilsson, M. Forero, E. V. Sokurenko, and V. Vogel, *Mol. Microbiol.* **53**(5), 1545–1557 (2004).
- ¹⁹O. Yakovenko, S. Sharma, M. Forero, V. Tchesnokova, P. Aprikian, B. Kidd, A. Mach, V. Vogel, E. Sokurenko, and W. E. Thomas, *J. Biol. Chem.* **283**(17), 11596–11605 (2008).

- ²⁰M. Forero, W. E. Thomas, C. Bland, L. M. Nilsson, E. V. Sokurenko, and V. Vogel, *Nano Lett.* **4**(9), 1593–1597 (2004).
- ²¹K. K. Sarangapani, T. Yago, A. G. Klopocki, M. B. Lawrence, C. B. Fieger, S. D. Rosen, R. P. McEver, and C. Zhu, *J. Biol. Chem.* **279**(3), 2291–2298 (2004).
- ²²T. Yago, J. H. Wu, C. D. Wey, A. G. Klopocki, C. Zhu, and R. P. McEver, *J. Cell Biol.* **166**(6), 913–923 (2004).
- ²³K. Konstantopoulos, W. D. Hanley, and D. Wirtz, *Curr. Biol.* **13**(15), R611–R613 (2003).
- ²⁴B. Guo and W. H. Guilford, *Proc. Natl. Acad. Sci. U.S.A.* **103**(26), 9844–9849 (2006).
- ²⁵F. Berkemeier, M. Bertz, S. B. Xiao, N. Pinotsis, M. Wilmanns, F. Grater, and M. Rief, *Proc. Natl. Acad. Sci. U.S.A.* **108**(34), 14139–14144 (2011).
- ²⁶I. Schwaiger, C. Sattler, D. R. Hostetter, and M. Rief, *Nature Mater.* **1**(4), 232–235 (2002).
- ²⁷G. Hummer and A. Szabo, *Acc. Chem. Res.* **38**(7), 504–513 (2005).
- ²⁸G. I. Bell, *Science* **200**(4342), 618–627 (1978).
- ²⁹C. Hyeon and D. Thirumalai, *J. Chem. Phys.* **137**(5), 055103 (2012).
- ³⁰O. K. Dudko, T. G. W. Graham, and R. B. Best, *Phys. Rev. Lett.* **107**(20), 208301 (2011).
- ³¹H. Lannon, E. Vanden-Eijnden, and J. Brujic, *Biophys. J.* **103**(10), 2215–2222 (2012).
- ³²B. Jagannathan, P. J. Elms, C. Bustamante, and S. Marqusee, *Proc. Natl. Acad. Sci. U.S.A.* **109**(44), 17820–17825 (2012).
- ³³J. Hsin, J. Strumpfer, E. H. Lee, and K. Schulten, in *Annual Review of Biophysics*, edited by D. C. Rees, K. A. Dill, and J. R. Williamson (Annual Reviews, Palo Alto, 2011), Vol. 40, pp. 187–203.
- ³⁴T. Ackbarow, X. Chen, S. Ketten, and M. J. Buehler, *Proc. Natl. Acad. Sci. U.S.A.* **104**(42), 16410–16415 (2007).
- ³⁵B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, *J. Chem. Theory Comput.* **4**(3), 435–447 (2008).
- ³⁶B. Hess, H. Bekker, H. J. C. Berendsen, and J. Fraaije, *J. Comput. Chem.* **18**(12), 1463–1472 (1997).
- ³⁷B. Hess, *J. Chem. Theory Comput.* **4**(1), 116–122 (2008).
- ³⁸W. G. Hoover, *Phys. Rev. A* **31**(3), 1695–1697 (1985).
- ³⁹M. Parrinello and A. Rahman, *J. Appl. Phys.* **52**(12), 7182–7190 (1981).
- ⁴⁰S. Kirmizialtin and R. Elber, *J. Phys. Chem. A* **115**(23), 6137–6148 (2011).
- ⁴¹E. Vanden-Eijnden and M. Venturoli, *J. Chem. Phys.* **130**(19), 194103 (2009).
- ⁴²P. Majek and R. Elber, *J. Chem. Theory Comput.* **6**(6), 1805–1817 (2010).
- ⁴³A. M. A. West, R. Elber, and D. Shalloway, *J. Chem. Phys.* **126**(14), 145104 (2007).
- ⁴⁴M. Berkowitz, J. D. Morgan, J. A. McCammon, and S. H. Northrup, *J. Chem. Phys.* **79**(11), 5563–5565 (1983).
- ⁴⁵S. H. Huo and J. E. Straub, *J. Chem. Phys.* **107**(13), 5000–5006 (1997).
- ⁴⁶R. J. Zhao, J. F. Shen, and R. D. Skeel, *J. Chem. Theory Comput.* **6**(8), 2411–2423 (2010).
- ⁴⁷F. Noe, C. Schutte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, *Proc. Natl. Acad. Sci. U.S.A.* **106**(45), 19011–19016 (2009).
- ⁴⁸A. Berezhkovskii, G. Hummer, and A. Szabo, *J. Chem. Phys.* **130**(20), 205102 (2009).
- ⁴⁹C. Schutte, F. Noe, J. F. Lu, M. Sarich, and E. Vanden-Eijnden, *J. Chem. Phys.* **134**(20), 204105 (2011).
- ⁵⁰J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, *J. Chem. Phys.* **126**(15), 155101 (2007).
- ⁵¹A. Cardenas and R. Elber, in *Innovation in Biomolecular Modeling and Simulation*, edited by T. Schlick (Royal Society of Chemistry, London, 2012).
- ⁵²W. E and E. Vanden-Eijnden, in *Annual Review of Physical Chemistry*, (2010), Vol. 61, pp. 391–420.
- ⁵³A. E. Cardenas, G. S. Jas, K. Y. DeLeon, W. A. Hegefeld, K. Kuczera, and R. Elber, *J. Phys. Chem. B* **116**(9), 2739–2750 (2012).