

Early Events in Helix Unfolding under External Forces: A Milestoning Analysis

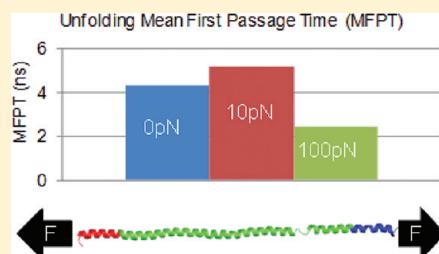
Steven M. Kreuzer,[†] Ron Elber,^{*,‡,§} and Tess J. Moon^{*,†,‡}

[†]Department of Mechanical Engineering, University of Texas at Austin, Austin, Texas 78712, United States

[‡]Institute for Computational Engineering and Sciences (ICES), University of Texas at Austin, Austin, Texas 78712, United States

[§]Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, Texas 78712, United States

ABSTRACT: Initial events of helix breakage as a function of load are considered using molecular dynamics simulations and milestoning analysis. A helix length of ~100 amino acids is considered as a model for typical helices found in molecular machines and as a model that minimizes end effects for early events of unfolding. Transitions of individual amino acids (averaged over the helix's interior residues) are examined and its surrounding hydrogen bonds are considered. Dense kinetic networks are constructed that, with milestoning analysis, provide the overall kinetics of early breakage events. Network analysis and selection of MaxFlux pathways illustrate that load impacts unfolding mechanisms in addition to time scales. At relatively high (100 pN) load levels, the principal intermediate is the β_{10} -helix, while at relatively low (10 pN) levels the α -helix is significantly populated, albeit not as an unfolding intermediate. Coarse variables are examined at different levels of resolution; the rate of unfolding illustrates remarkable stability under changes in the coarsening. Consistent prediction of about ~5 ns for the time of a single amino-acid unfolding event are obtained. Hydrogen bonds are much faster coarse variables (by about 2 orders of magnitude) compared to backbone torsional transition, which gates unfolding and thereby provides the appropriate coarse variable for the initiation of unfolding. Results provide an atomic description of "catch–bond" behavior, based on alternative pathways, in which unfolding of a simple protein structural element occurs over longer timescales for intermediate (10 pN) loads than for zero (0 pN) or large (100 pN) loads.



I. INTRODUCTION

Alpha helices are fundamental building blocks of protein folds. As structure elements frequently found in globular proteins, they continue to attract considerable attention in experiments,^{1–5} theory,^{3,6,7} and simulations.^{1,2,8–12} Typical helices found in proteins *in vivo* are short (~10–20 amino acids), a necessary restriction to support compact three-dimensional shapes of proteins. It is, therefore, not surprising that most of the experiments and simulations on the properties of helices focus on relatively short helical peptides of about 20 amino acids or less.^{1,2,5,9,12,13} However, significantly longer helices are also found *in vivo* that are making important and unique contributions in the field of molecular biophysics. For example, molecular motors—like kinesin^{14–17} and myosin^{18–24}—use order of magnitude longer helices as key components to "walk" along cytoskeletal proteins to transmit energy and to conduct work.

For a number of reasons, it is intriguing to investigate the physical, biological, and mechanical properties of these longer helices, in addition to the well-studied short helices. First, it is difficult to extract from short helices the general and average properties of a single amino acid that are useful in phenomenological models of helix folding. While studies of short helices are strongly influenced by end effects, studies of long helices may enable the definition and calculation of an "average" helical residue that is approximately invariant to translation (in the helical sense) far from the helix edges. Second, the central role that long helices play in the proper functioning of motor pro-

teins renders the study of their mechanical properties to be particularly important. Here, we specifically explore the onset of mechanical instability in long helices in which the first "crack" is observed. Such an initial crack in the basic building block of a motor is critical to the motor's overall stability under load and, thus, likely to be important to their function. Third, the helix's longer length potentially permits a new type of experiment directly probing the molecule's mechanical properties, namely, single-molecule pulling experiments. While pulling experiments^{25–35} have been used for some time to probe mechanical properties of biological macromolecules, here we examine, via computations, the feasibility of using a single-molecule pulling experiment to investigate the early events in the unfolding of a relatively long, helical structure. Fourth, the process of pulling by force is fundamentally different from the unfolding of short helices. Unfolded short helices have vast conformational space from which to sample, much larger than the conformational space of the folded state. Notably, this is not the case for force-induced unfolding in which the extended stretch (and unfolded) configuration is as unique as the folded state, providing yet

Special Issue: Macromolecular Systems Understood through Multi-scale and Enhanced Sampling Techniques

Received: January 24, 2012

Revised: March 29, 2012

Published: April 3, 2012

another motivation to investigate long helix unfolding. The impact of this difference is clearly illustrated in the present manuscript in which unfolding times and mechanisms vary nonmonotonically with the applied pulling force magnitude. This non-monotonic, “catch–bond” behavior is shown to be due to changes in the unfolding pathway with load, thereby providing an atomic description for this phenomena.

We focus on molecular dynamics (MD) simulations of a relatively long, heterogeneous helix embedded in aqueous solution. While these atomically detailed simulations of biological macromolecules hold the promise of first-principle understanding of complex mechanisms, while enabling direct comparison to experimental data, they are unfortunately severely restricted with respect to both temporal and spatial scales. Enhanced sampling approaches are required to investigate the time courses and mechanisms of local crack formations in the helix. Toward these ends, we chose the method of milestones^{9,22,36–41} (for the present task, a similar alternative is the Markov state model^{42,43}) to analyze the MD trajectories and make it possible to construct coarse-grained models over extended temporal and spatial scales.

This paper is organized as follows. In the next Methods section, we discuss the milestoneing method. While the milestoneing algorithm and theory have been discussed extensively elsewhere,^{36–38} its application here, employing relatively long trajectories and corresponding analysis, is different and requires additional considerations. Further, we elaborate on the impact of choosing different sets of coarse variables, as well as models with variable spatial resolutions. We then discuss various findings in the Results section, followed by the Discussion and Conclusions sections.

II. METHODS

II.1. Methods: The Milestoning Algorithm. We consider a local, conformational transition in which an amino acid, initially in a folded, α -helical conformation, transitions to an unfolded, β -strand-like conformation. We conduct atomic-level simulations in which the Cartesian coordinates of all atoms in the peptide and the water molecules are explicitly represented and followed as a function of time. That said, an atomic-level description of the peptide’s state at each “snapshot” in time at this level of spatial resolution is clearly overkill. At any given point in time, the secondary-structure state of each amino acid is uniquely determined by the following degrees of freedom: (i) the hydrogen bonds that are along the peptide’s backbone and surrounding the amino acid and (ii) the peptide’s backbone dihedral angles, primarily ψ .

From an equilibrium perspective, it is suggestive to define a free-energy landscape as a function of these degrees of freedom to obtain needed information on the probability of forming or unforming a helix as a function of system variables, such as the strength of the pulling force. However, for the study of kinetics, the free-energy landscape is neither sufficient nor likely to be the most efficient way to proceed. We are missing complex functions and parameters, such as hydrodynamic interactions, nondiagonal diffusion, and friction memory tensors to correctly capture the kinetics of these highly nonlinear coordinates. It is therefore useful to construct an alternative coarse-graining approach to atomically detailed simulations that will capture, in a single simulation, a complete kinetic and thermodynamic model. Here, we propose to use the theory and algorithm of milestoneing to obtain such a coarse-grained model for the initial events in unfolding of a long helix.

The system’s atomic-level coordinates are coarse grained into (i) the status of these hydrogen bonds and (ii) the value of the dihedral angle ψ . Further, two different levels of coarse graining (“coarse” and “fine”) are used to explore the variation of the results with the degree of coarsening and to discuss mechanisms viewed at the different spatial scales. We denote the current state of the hydrogen bonds and dihedral angle ψ by a state vector Y . A hydrogen bond is assumed “on”, denoted by a 1, if (i) the distance between the nitrogen atom of the amide group to the oxygen atom of the carbonyl group is less than 3.5 Å and (ii) the NH–O angle is greater than 150°. (We also experimented, as discussed in the Results, with a hydrogen bond cutoff distance of 4.0 Å and an NH–O angle of 120°, only to find a small variation in the results.) The space of the dihedral angle ψ is divided into three bins: −150 to 0°; 0 to 90°; and 90 to 180° or −180 to −150°. The state vectors corresponding to the folded, α -helical and the unfolded, β -strand conformations are Y_α and Y_β , respectively. A trajectory in coarse space at any given time t is denoted by $Y(t)$.

The first step of milestoneing is the identification of “anchors”. Anchors are amino-acid conformations represented by the coarse-graining variables discussed above (hydrogen bond status and dihedral angle value); they are collectively referred to as the “amino-acid state” or more briefly “AA-state”. We note that the coarse variables used here have been discretized: (i) the hydrogen bonds either exist (“1”) or do not exist (“0”), according to geometrically defined criteria, and (ii) the dihedral angle ψ is similarly placed into one of three bins (defined above). That said, the anchors used here completely span the space, incorporating all possible values of the coarse variables within the desired scope of coarsening; thus, the mapping from the Cartesian coordinates to the coarse variables is unambiguous. As the number of discrete states is not prohibitively large, and as the transitions between the states are only marginally activated, we are able to obtain meaningful statistics for all possible AA-states using straightforward MD trajectories.

Analysis of full-length MD trajectories is in contrast to the more common implementation of milestoneing in which anchors are defined through a separate calculation, e.g., reaction path calculations or replica exchange simulations. From this perspective, the sampling of configurations is more similar to past Markov state models⁴² than to prior milestoneing calculations. In the spirit of a recent study⁴⁴ that brought milestoneing to the Markov state model, we bring ideas from the Markov state model (use of long trajectory sampling) into milestoneing. In typical milestoneing calculations,^{9,22,36–38} the processes under consideration are activated and/or have very long diffusion times that direct MD simulations are unable to properly sample. As a result, a process for anchor identification must precede the MD simulations, which are then performed only between the domains assigned to anchors.^{9,12,22} While anchor-to-anchor simulation has the advantage that the trajectories can be much shorter and more efficient, it has the disadvantage that the calculations of kinetics assume decorrelation of short trajectories and depend on anchor identification.^{38,41} In the present investigation, the MD simulations are run in a straightforward fashion with no *a priori* identification/selection of anchors. This is clearly an advantage, as different types of anchors can be tested *a posteriori* using the same set of MD trajectories. The limitation, of course, is our ability to adequately sample the space of coarse-grained variables in the straightforward MD simulations between the initial and final states.

Once the anchors are defined, the next step is the identification of “milestones”, defined as the interface between two anchors. As our anchors (or AA-states) are defined by discrete coarse variables—here, the presence/absence of hydrogen bonds spanning a residue and the bin of the dihedral angle ψ —the milestones represent changes in (i) the number of existing hydrogen bonds spanning a residue, (ii) the type of hydrogen bonds spanning a residue, and/or (iii) the bin of the dihedral angle ψ .

In milestoning, we focus on the detailed statistics of transitions between the interfaces (see Figure 1) that are analyzed

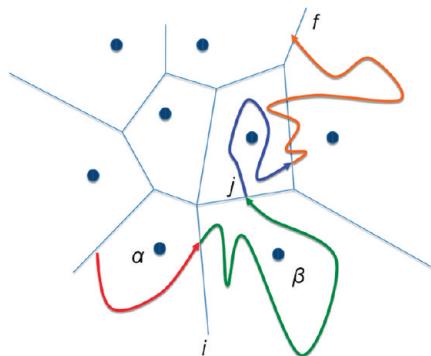


Figure 1. Anchors, milestones, and trajectory schematic. Schematic representation of an exact trajectory (multicolor curve) on a two-dimensional energy surface. Light blue small circles and straight lines denote anchors and milestones, respectively. Milestones divide the complete trajectory into fragments, which represent transitions between milestones. Each distinct color indicates a different trajectory fragment—and a transition. Transitions between milestones are enumerated in order to estimate the flux used in the coarse-grained equation describing the kinetics.

using statistical mechanics tools to obtain complete descriptions of the process’s kinetics and thermodynamics at the level of coarse variables. Consider a long duration, reactive trajectory $Y(t)$. This trajectory is followed as a function of time and further identified by the last milestone that is passed, e.g., $Y_{\alpha:\beta}(t)$, where $\alpha:\beta$ is the last transition interface traversed between two anchors, α and β . Figure 1 illustrates a representative trajectory and its milestones; the different colors correspond to different milestone transitions.

The current milestoning implementation differs from previous uses in that we break few (10–20), relatively long trajectories into smaller fragments between milestones. Moreover, the current implementation benefits from simulations of a naturally occurring sequence/structure in which many different amino acids (86 residues; 14 distinct AA types) are tracked simultaneously. As we seek a general mechanism for early events of helix unfolding, we do not *a priori* differentiate between events that occur at different positions along the helix. This approach helps us define an “average” helical residue, gaining significantly more statistics. That said, even with this relatively long helix, amino-acid type and position potentially influence the kinetic and thermodynamic properties of the process. We therefore examine this possibility in the Results section and conclude that, at least for some important observables, e.g., the mean first passage time for local unfolding, the concept of an “average” amino acid can be useful, yielding accurate results.

The original formulation of milestoning relied, in part, on the assumption of fragment independence, premised on a “first

hitting distribution” (FHD) for the initiation of trajectory fragments.^{37,38} Therefore, trajectories were terminated only when a new milestone was passed and specifically not when recrossing the initiating milestone. In the same spirit, here the trajectory fragments “cut” from the full, long trajectories are defined by the crossing of a new milestone. To illustrate, consider the trajectory fragments in Figure 1: the orange fragment starts at the end of the blue fragment and recrosses its initiating milestone several times before terminating at the next (noninitiating) milestone it traverses. The entire orange trajectory portion, including recrossings, is therefore one fragment. For each fragment, three pieces of information are retained: the fragment’s initiating and terminating milestones and its duration (also known as its lifetime).

II.2. Methods: Milestoning Theory. With milestones defined, we now consider the transition probability density K_{ij} between phase space points X_i and X_j at the interfaces at time t , $K_{ij}(X_i, X_j; t - t')$, also referred to as the “Kernel”. Note we use the vector X_i to denote a full phase space vector constrained to the interface i defined by the set of coarse variables Y_i . The Kernel is the probability that a trajectory starting at phase space point X_i at interface i will make it to a phase space point X_j at milestone j after time t (Figure 1). Let the number of trajectories that pass X_i at time t and reach X_j at time t' be $n_{ij}(X_i, X_j, t - t')$, $n_i(X_i, t)$ be the number of trajectories that pass milestone i at time t , then $K_{ij}(X_i, X_j; t - t') \cong n_{ij}(X_i, X_j; t - t')/n_i(X_i, t)$. We restrict our attention to processes that are stationary at least locally—but not necessarily in equilibrium—in which we can write $g(t, t') = g(t - t') = g(\tau)$, where $\tau = t - t'$.^{37,40}

We further define the following entities: $P(X_j, t)$, the probability that, at time t , the last milestone crossed was j at an interface point X_j ; and $q_i(X_j, t')$, the probability density, also known as the “reaction flux”, that, exactly at time t' , a given trajectory hits a milestone coordinate X_j .

This Kernel is general and can be estimated for different types of dynamics, e.g., Newtonian or Langevin dynamics.³⁸ The reaction flux, $q_i(X_j, t)$, passing through milestone i at exactly X_i at time t , is

$$q_i(X_i, t) = P(X_i, t)\delta(t^+) + \sum_{j \in \bar{i}} \int dX_j \int dt' q_j(X_j, t') K_{ji}(X_i, X_j; t - t') \quad (1.1)$$

Equation 1.1 states that milestone i is crossed at time t by one of two mechanisms: either it is exactly at i at time $t = 0^+$ (first term) or, exactly at time t , it transitions to milestone i from another milestone j (second term). Only transitions between milestones that can be reached without traversing other milestones first are considered; thus, \bar{i} is a subset of milestones that can reach directly milestone i and is known as the *reachable* milestones from milestones i .

While eq 1.1 is exact, it can also be used as a starting point for spatial and temporal coarse graining. The *essential* coarse-graining assumption of the Milestoning theory is partial independence of initial conditions; that is, the interfaces are presumed to be sufficiently separated such that a trajectory initiated at say X_i does not “remember” the precise location of the initiating phase space point within milestone i when it terminates at X_j . Mathematically, this can be written as

$$K_{ij}(X_i, X_j; t - t') \cong K_{ij}(X_j; t - t') \quad (1.2)$$

This Milestoning approximation has been extensively investigated in the past,^{40,41} both theoretically and computationally. These rigorous results, as well as empirical and practical

guidelines, make it possible to efficiently and accurately coarse grain the system, as well as provide consistent results with microscopic data at higher degrees of computational efficiency. The procedure is similar in spirit to other coarse-graining approaches in which one assumes the existence of local equilibrium, and time independence of fast variables that quickly relax to local equilibrium. Examples are the equilibrium calculations of the potential of mean force⁴⁵ or the creation of a master equation in the Markov state model.^{42,46} Here, however, the approximation is different, as we only require decorrelation of the flux and not population. Flux, i.e., the velocity times probability, is likely to decorrelate faster than spatial distribution, as velocities in the condensed phase rapidly decorrelate, having a time constant of about 1 ps.

Once one accepts the essential Milestoning assumption (eq 1.2), it is useful to integrate over the entire phase space X_i and (if applicable) X_j to define the following functions:

$$q_i(t) = \int q_i(X_i, t) dX_i \quad (1.3)$$

$$K_{ij}(t - t') = \int K_{ij}(X_i, X_j; t - t') dX_i dX_j \quad (1.4)$$

$$P_i(t) = \int P_i(X_i, t) dX_i \quad (1.5)$$

Substituting eq 1.2 into eq 1.1 and integrating over phase space X_i , we obtain

$$\begin{aligned} \int q_i(X_i, t) dX_i &= \int P_i(X_i) dX_i \delta(t^+) \\ &+ \sum_{j \in \bar{i}} \iint \int_0^t q_j(X_j, t') K_{ji}(X_i; t - t') dX_i dX_j dt' \end{aligned} \quad (1.6)$$

Equation 1.6, together with eq 1.4, provides the Milestoning formula

$$q_i(t) = P_i(0) \delta(t^+) + \sum_{j \in \bar{i}} \int_0^t q_j(t') K_{ji}(t - t') dt' \quad (1.7)$$

Importantly, the explicit dependence on coordinates is removed in eq 1.7; only the indices of the milestones remain. Further, the Kernel $K_{ij}(t - t')$ is now estimated by $n_{ij}(t - t')/n_i$, where n_i is the total number of trajectory fragments crossing interface i and $n_{ij}(t - t')$ is the number of trajectory fragments that reach milestone j from milestone i after time τ . Thus, the system is reduced to a mesh in coarse-grained space: initiating milestone, terminating milestone, and time.

As eq 1.7 for the q_i components of the reactive flux column vector \mathbf{q} is linear, it can be solved analytically using Laplace transforms and corresponding initial-value and final-value theorems.^{37,47} Via the Laplace transform framework, the stationary reactive flux vector, \mathbf{q}_{stat} , can be determined using

$$(\mathbf{I} - \mathbf{K})\mathbf{q}_{\text{stat}} = 0 \quad (1.8)$$

where \mathbf{I} is the identity matrix, and \mathbf{K} is the *stationary* transition probability matrix with elements $(\mathbf{K})_{ij} = \int_0^\infty K_{ij}(t) dt$. If $(\mathbf{P})_i$ are the elements of probability distribution vector \mathbf{P} , then the $(\mathbf{P}_{\text{stat}})_i$ elements of the stationary probability distribution vector \mathbf{P}_{stat} are given by a scalar multiplication:

$$(\mathbf{P}_{\text{stat}})_i = (\mathbf{q}_{\text{stat}})_i \langle \tau \rangle_i \quad (1.9)$$

where $\langle \tau \rangle_i$ the time-averaged lifetime of milestone i , is defined by $\langle \tau \rangle_i = \sum_i [\int_0^\infty t K_{ii}(t) dt]$. Note, $\langle \tau \rangle_i$ form the elements of

the time-averaged milestone lifetime vector $\langle \tau \rangle$. The overall mean first passage time (MFPT) from milestone i to the final (assumed absorbing) milestone f , $\langle \tau \rangle_{if}$ is given by

$$\langle \tau \rangle_{if} = P_{\text{init}} \sum_j [\mathbf{I} - \mathbf{K}_{\text{final}}]^{-1} \langle \tau \rangle_j \quad (1.10)$$

where the initial probability distribution \mathbf{P}_{init} has element $(\mathbf{P}_{\text{init}})_i$ equal to 1 for initiating milestone i and equal to 0 for all other milestones. Further, the absorbing final transition probability matrix $\mathbf{K}_{\text{final}}$ is a modification of \mathbf{K} where the absorbing condition is enforced by zeroing the column associated with the final milestone. We note the important difference between $\langle \tau \rangle_i$ and $\langle \tau \rangle_{if}$, the former is the average lifetime of individual milestones, and the latter is the mean first passage time between the initiating milestones and absorbing milestones i and f , respectively.

We emphasize the simplicity of eqs 1.8–1.10. In order to obtain the stationary distribution and overall mean first passage time, we only need to solve a linear problem. For example, the reactive flux vector, \mathbf{q}_{stat} , is the zero-eigenvalue eigenvector of the matrix $[\mathbf{I} - \mathbf{K}]$. The dimensionality of $[\mathbf{I} - \mathbf{K}]$ is equal to the number of milestones, which, in the present case, is 182 and 1722 for the coarse- and fine-anchor meshes, respectively. Thus, in estimating the kernels, the linear algebra calculations are trivial as compared to the high costs of MD simulations. To illustrate, it can require months of core time to estimate the kernels but take only a few minutes to solve the linear problem on a reasonable computer.⁹

So what, if anything, is gained by using this non-straightforward calculation of kinetics and thermodynamics? Milestoning offers a mathematically well-defined procedure of mapping thermodynamics and kinetics onto an internally consistent grid of coarse-grain degrees of freedom. The reduced system, assuming that the mapping is correctly done, is much simpler to analyze and is easier to understand and to extend the investigations to long-time behavior. Also, by defining an “average” amino acid and its thermodynamic and kinetic properties, it is possible to extend the investigation in the present system to other long helices of varying lengths.

II.3. Methods: Computational Protocol. In order to reveal the “natural” coarse variables necessary for effective description of helix unfolding, explicit-water molecular dynamics simulations are performed for three axial load conditions: 0, 10, and 100 pN. The model structure is chain A of the human cardiac β -myosin S2 structure (2fxm.pdb),⁴⁸ residues 1–126; throughout this paper, it is referred to as 2fxmA. Its sequence is the following: GSSPLLKSAEREKEMASMKEEFLTRKEALEKSEARRKELEEKMVSLQLQEKNNDLQLQLVQAEQDNLADAEERCDQLIKNKIQLEAKVKEMNERLEDEEEEMNAELTAKKRKLEDECSELKRDIDDLETLAK.

Simulations are performed with the GROMACS 4.0.5⁴⁹ distribution using the GROMOS96 “53a6” united-atom force field⁵⁰ 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 $7.4 \times 8.2 \times 25.6$ nm³ rectangular box. For system charge neutrality, 102 Na⁺ and 91 Cl⁻ counterions were added providing a concentration of 0.1 M, resulting in a full system containing 1290 protein atoms and 49 605 water molecules, in addition to these counterions. Periodic boundary conditions were used throughout all phases of the MD simulations, as was the parallel processing version of the LINCS constraint on fast-bond dynamics.⁵¹

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 thermostat⁵² to maintain temperature. Final production runs used a Nose–Hoover thermostat^{53,54} with a relaxation constant of 0.1 ps to maintain 310 K and a Parrinello–Rahman barostat⁵⁵ with a coupling constant of 1.0 ps. The Coulomb and van der Waals cutoff radii were set to 1.4 nm; PME electrostatics handled long-range electrostatics.⁵⁶ An initial 1 ns equilibrium simulation provided the seeds for subsequent production runs. All simulations were run on the Texas Advanced Computing Center's (TACC) Sun Constellation Linux Cluster, Ranger (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. The load-application 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 helix's interior 86 residues were used in the analysis, purposely excluding the pull group residues 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.

Total simulation time on TACC's Ranger for each load level was 60 ns, albeit it was split between shorter individual simulation runs (see Table 1). For both the 0 and 10 pN load

Table 1. Constant-Force Molecular Dynamics Simulation Summary^a

load level (pN)	number of runs (n)	individual run times (ns)	cumulative run time (ns)
0	10	6	60
10	10	6	60
100	20	2, 3, or 4	60

^aSimulation quantity and durations for the unloaded (0 pN) and constant-force molecular dynamics (CFMD) (10 and 100 pN) loading scenarios. The 0 and 10 pN simulations consisted of 10 independent runs of 6 ns each, while the 100 pN simulations consisted of 20 runs of either 2 ns (*n* = 3), 3 ns (*n* = 10), or 4 ns (*n* = 6), providing a net 60 ns of raw data for analysis. The 100 pN simulations were shorter than either the 0 or 10 pN due to their proclivity to unfold more rapidly and a desire to avoid accumulating unfolded data. For calculation of AA-state probabilities (Figure 7) and AA-state exit probabilities (Figure 10), each run was treated as an individual “block” of data, allowing the calculation of block averages and block standard errors across the runs within each load.

levels, the simulation time was divided across 10 different 6 ns runs; the 100 pN load levels included 20 different runs of either 2, 3, or 4 ns duration. As the 100 pN simulations demonstrated a strong proclivity for unfolding a large portion of the helix's interior, runs were terminated after 2 or 3 ns if more than 20% of the interior residues had unfolded; otherwise, they were run to 4 ns. We note that the relatively large number of residues analyzed within the helix's interior (86) over the full simulation duration (60 ns) provides roughly the same number of

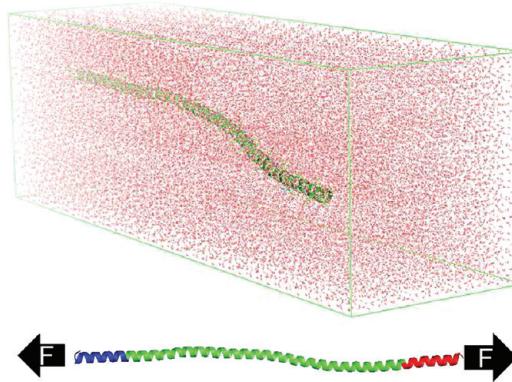


Figure 2. CFMD simulation box and load application. (top) The $7.4 \times 8.2 \times 25.6$ nm rectangular, simulation box, filled with explicit water, used in the constant-force molecular dynamics (CFMD) computations. The simulated helix contained 126 protein residues from chain A of the human cardiac β -myosin S2 domain (2fxm.pdb;⁴⁸ chain A referred to as 2fxmA): 49 605 explicit SPC water molecules and 193 counterions (102 Na⁺ and 91 Cl⁻). (bottom) Illustration of loading groups on the 2fxmA chain. Loading groups were the 18 C- and N-terminal residues; 86 of the 90 residues between these C- and N-terminal loading groups were used for analysis. Two unloaded residues on either side of the 86-residue interior were not used in the analysis to minimize end effects from the applied loads. GROMACS's CFMD loading method partitions the total, user-specified, load (F_{tot}) in a user-prescribed direction over all atoms of the loaded groups on a mass-weighted basis such that the forces on atom *i* are given by $F_i = (m_i/M)F_{\text{tot}}$ where m_i is the mass of the atom *i* and M is the sum of all the masses. The CFMD approach maintains these two quantities (magnitude and direction) constant during the entire simulation. All protein structures were generated with PyMol (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC).

“residue-nanoseconds” (~ 5000) of data as 1 μ s of simulation for a pentapeptide.

II.4. Methods: Coarse-Grained Model System Definitions. As discussed in the Theory section above, the first step in milestone calculations involves the selection of appropriate anchors and coarse variables describing the space of system conformations. We investigate early, local events in helix unfolding in which the helix at large is intact and a few residues participate in an unfolding “seed”. For each residue, the coarse variables are the status of the hydrogen bonds spanning the residue and the bin of the residue's dihedral angle ψ ; the dihedral angle φ is specifically not considered in this analysis, as φ is relatively invariant for the helix and β -strand conformations. Data used in the milestone analysis, in which transitions between states were counted and analyzed, are from all 86-interior residues within the helix's nonloaded interior. In contrast to the previous milestone investigations^{9,22,36,38,41} in which a large number of short trajectories are used, we employ relatively long trajectories. The theory, however, remains the same and the kernel allows for the calculation of kinetics over the appropriate time scale.

Note that the term “AA-state” refers to the state or condition of the amino acid—specifically its ψ dihedral angle as well as the status of hydrogen bonds spanning the amino-acid backbone. As a result, “AA-state” is used synonymously with the term “anchor”. In the present study, we use two sets of state definitions to coarse grain our system: (i) a set of *coarse* AA-states describing the secondary structure of a particular residue and (ii) a set of *fine* AA-states describing the specific state or permutation of hydrogen bonds within the purely α -helical

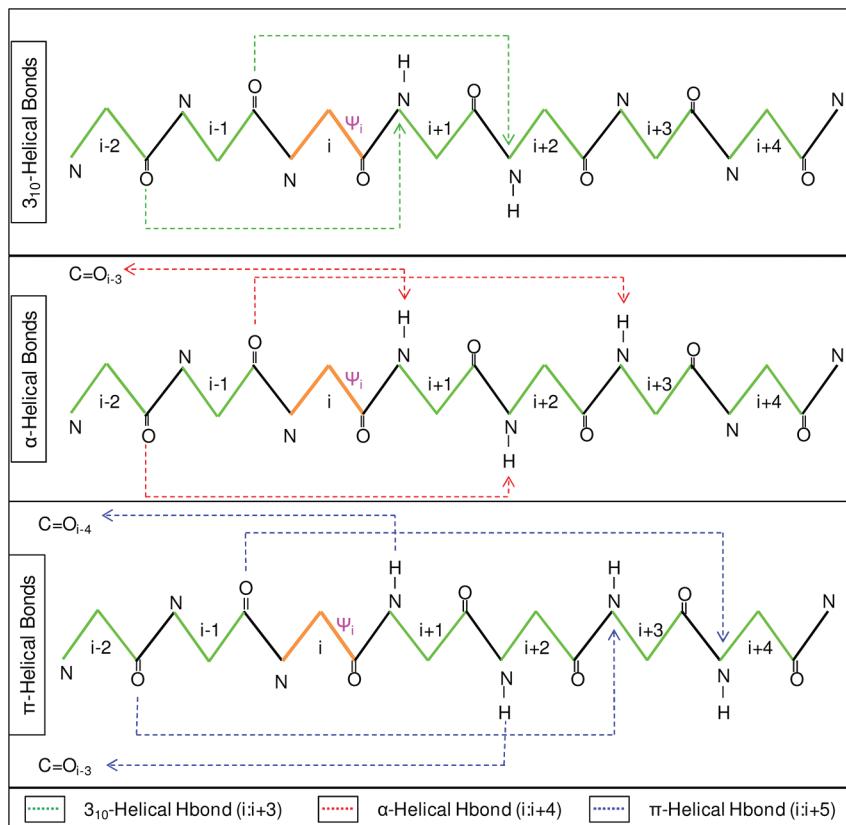


Figure 3. 3₁₀-, α -, and π -helical hydrogen-bond types spanning a residue. Residue status in the coarse system is defined by (i) the set of hydrogen bonds spanning the residue of interest (orange) as well as (ii) the residue's dihedral angle ψ . Hydrogen bonds appear in three varieties depending on the residues whose amide and carbonyl groups are bonded: 3₁₀-helical hydrogen bonds bind an amide group with the carbonyl three residues away ($i:i+3$), while α -helical hydrogen bonds have $i:i+4$ binding and π -helical hydrogen bonds have $i:i+5$ binding. Any given residue may have multiple hydrogen bonds spanning it, as illustrated here. Coarse anchor definition requires evaluating the state of each of these bond types and classifying the residue based on the resulting set of existing bonds. Hydrogen bonds are considered to exist if the N–O distance is <0.35 nm and the NH–O angle is $>150^\circ$.

states (as illustrated in Table 3, for the fine anchors only, residues exhibiting non- α -helical characteristics, like π - or 3₁₀-helical bonds, were “lumped” into the π - or 3₁₀ fine anchor classifications, respectively). These two anchor sets are used separately in two levels of analysis. Importantly, these anchor definitions are assigned *a priori* to reflect expected/possible conformations of the system, while the time series are generated directly with CFMD without a reference to the coarse variables. In both cases, the anchors are defined on the basis of the existence of local hydrogen bonds where a hydrogen bond is formed according to the criteria employed by the GROMACS 4.0.5 “g_hbond” utility: (i) a donor–acceptor (N–O) distance less than 0.35 nm and (ii) an NH–O angle greater than 150°.

The coarse set of AA-states is analyzed first to determine the readily accessible system conformations, including non- α -helical secondary structures. From this coarse analysis, the focus shifts to the finer resolution in which the kinetics of the coarse α -helical states is studied in greater detail. Additionally, the fine AA-states incorporate a wider set of possible hydrogen bonds such that the three α -helical bonds spanning the residue of interest are combined with the two additional α -helical bonds that are formed by the amino and carbonyl groups of the residue itself, bringing the total number of α -helical hydrogen bonds considered in the fine AA-states to five.

In order to explain the difference in scope between the two AA-state sets, the coarse AA-states are predicated on

determining the *quantity* of hydrogen bonds of different types, while the fine AA-states count the *permutation* of existing hydrogen bonds. Due to the wider scope of the coarse AA-states, it is not possible to easily enumerate all bond permutations of all possible secondary structure states as is done with the fine AA-states, thus the desire to focus on a less-specific AA-state definition. Thus, the relative advantages/disadvantages of the two definitions are, whereas the coarse AA-states compute general information about transient secondary structures, the fine AA-states elucidate the necessity of specific bonds and sensitivity to specific bond combinations. It is affirming (see section III.2.5 and Table 9) that mean first passage times (MFPT) of helix breakage are found to be insensitive to the different types of coarse graining mentioned above.

In contrast to the conventional biochemical designations of secondary structure limited to properties (hydrogen bonds or φ , ψ values) confined to each individual amino acid, here we consider all hydrogen bonds that span across the amino acid. The description we adopt is necessary to describe transitions that occur far from the helical ends, since they are strongly coupled to the internal rotations of the particular amino acid at the center. We further comment that the use of torsion values alone is insufficient to capture flips between different types of hydrogen bonds, since structural fluctuations during the simulations make it difficult to obtain one-to-one correspondence between hydrogen bonds and dihedral angles. The added

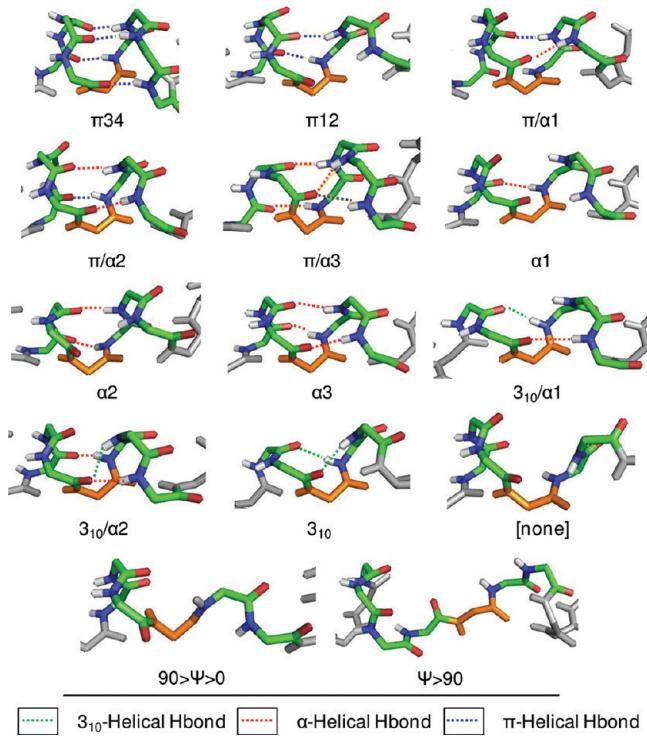


Figure 4. Coarse anchor illustrations. The 14 coarse anchors illustrating the local hydrogen-bond network spanning a given residue (orange). Spanning hydrogen bonds may be one of three varieties: $3_{10}^{\text{-}}$ -helical (residues $i:i+3$), α -helical (residues $i:i+4$), or π -helical (residues $i:i+5$). Combinations of different hydrogen bond types and quantities define the coarse anchors. Hydrogen bonds are considered to exist if the N–O distance is <0.35 nm and the NH–O angle is $>150^\circ$ (side chains are omitted for clarity).

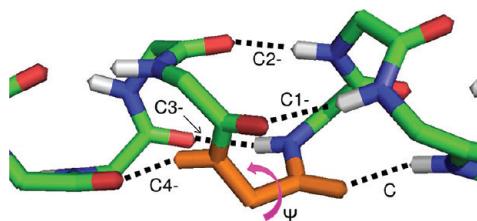


Figure 5. Fine-anchor hydrogen bonds and naming convention. Five hydrogen bonds surrounding a residue are enumerated according to the position of the carbonyl group (red atoms in the above stick representation) relative to the residue of interest (orange). Hydrogen bonds are considered to exist if the N–O distance is <0.35 nm and the NH–O angle is $>150^\circ$. In the absence of $3_{10}^{\text{-}}$ or π -helical bonds, the permutation of existing bonds in this five hydrogen bond set and the dihedral angle ψ determine the amino acid's fine anchor or fine AA-state.

insight enabled by this secondary structure classification is key to (i) establishing the degree of helical structure “disruption”, as represented by the quantity of hydrogen bonds, and (ii) avoiding ambiguity in secondary structure assignment due, for example, to overlapping ψ -angle distributions for the different helical conformations (see Figure 8).

II.4.1. Methods: Coarse-Anchor System. Coarse anchors, defining the secondary structure of a particular residue, are based on the status of hydrogen bonds spanning the residue, as well as the residue’s dihedral angle ψ (Table 2). In contrast to selections made solely based on the residue’s dihedral angle ψ ,

this secondary structure description is chosen to reflect the local backbone conformations and hydrogen bond topology surrounding the residue. In particular, a residue’s coarse anchor is calculated by evaluating the existence of three types of hydrogen bonds spanning the residue: $3_{10}^{\text{-}}$, α -, and π -helical bonds in which a carbonyl group from residue “ i ” binds to the amide group of residue “ $i+3$ ”, “ $i+4$ ”, or “ $i+5$ ”, respectively. A preliminary check on the existence of gamma turns—in which a carbonyl group from residue “ i ” binds to the amide group of residue “ $i+2$ ”—revealed that a probability of less than 0.5% for the 0 and 10 pN scenarios (0 pN, 0.37%; 10 pN, 0.38%) and less than 2% for the 100 pN scenario (1.5%). Given these hydrogen bond donor/acceptor pairs, it is possible for a number of hydrogen bonds to span any given residue (where “spanning the residue” means that the carbonyl-group donor precedes and the amide-group acceptor succeeds the residue of interest). Specifically, there are possibly 2, 3, or 4 bonds spanning a residue for the pure $3_{10}^{\text{-}}$, α -, or π -helical bonds, respectively. At each image and for each residue, the existence of a hydrogen bond of each type is calculated for all possible bonds (9 in total) across the residue. The particular combination of bonds formed defines the secondary structure of that residue, thereby incorporating secondary structure information into a residue description. These bond combinations also include transition structures in which hydrogen bonds of two different types span a residue at the same time, e.g., a $\pi/\alpha 1$ state in which one α -helical bond and at least one π -helical bond span a given residue.

Table 2. Coarse Anchor/Amino-Acid State Definitions^a

coarse anchor/AA-state designation	quantity of hydrogen bonds			ψ angle values (deg)
	310	α	π	
$\psi > 90^\circ$	any	any	any	$>90^\circ, <-150^\circ$
$90^\circ > \psi > 0^\circ$	any	any	any	$0^\circ - 90^\circ$
<i>none</i>	0	0	0	$<0^\circ$
310	any	0	0	$<0^\circ$
310/ $\alpha 2$	any	2	0	$<0^\circ$
310/ $\alpha 1$	any	1	0	$<0^\circ$
$\alpha 3$	0	3	0	$<0^\circ$
$\alpha 2$	0	2	0	$<0^\circ$
$\alpha 1$	0	1	0	$<0^\circ$
$\pi/\alpha 3$	0	3	any	$<0^\circ$
$\pi/\alpha 2$	0	2	any	$<0^\circ$
$\pi/\alpha 1$	0	1	any	$<0^\circ$
$\pi 12$	0	0	1 or 2	$<0^\circ$
$\pi 34$	0	0	3 or 4	$<0^\circ$

^aCoarse anchors are based on both the type and quantity of hydrogen bonds spanning a residue, as well as the residue’s dihedral angle ψ , reflecting the fact that multiple types (e.g., $\pi/\alpha 1$, $3_{10}/\alpha 1$) and multiple quantities of bonds (e.g., $\alpha 3$, $\alpha 2$, $\alpha 1$) may span a residue at any given time. The anchor *none* denotes the zero-hydrogen bond condition, while two other anchors reflect two positive dihedral ψ angle ranges (φ angle is not evaluated). Hydrogen bonds are considered to exist if the N–O distance is <0.35 nm and the NH–O angle is $>150^\circ$.

In addition to the hydrogen bonds defining a given residue’s secondary structure, the coarse anchors also include three zero-hydrogen-bond states that represent the three different dihedral angle ψ ranges for this zero-hydrogen condition. They include (i) a state for which the dihedral angle ψ is negative (referred to as *none* due to the absence of hydrogen bonds and the negative

ψ angle), (ii) the dihedral angle ψ is positive but less than 90° ($90^\circ > \psi > 0^\circ$), and (iii) the dihedral angle ψ is greater than 90° or less than -150° ($\psi > 90^\circ$).

II.4.2. Methods: Fine Anchor System. Fine anchors are defined by the (i) hydrogen bond pattern for all α -helical states having negative- ψ angles, (ii) two different, positive- ψ angle states for four rotationally flexible hydrogen bond patterns, and (iii) two “basket” states to reflect residues containing either π - or 3_{10} -helical bonds (Table 3). Specifically, the 42 fine anchors are the direct sum of

- 32 fine anchors that represent the hydrogen bond pattern for all α -helical states having negative- ψ angles. They reflect the 32 ($=2^5$) possible permutations of the five specific hydrogen bonds associated with a given residue: the three spanning across the residue (“C1-”, “C2-”, and “C3-”) as well as the two bonds formed by the residue’s amino (“C4-”) and carbonyl (“C”) groups. We note that, while these designations are labeled with the relative carbon position from the carbonyl group, the hydrogen bond calculation is performed on the basis of the oxygen position.
- 8 fine anchors that represent two different, positive- ψ angle states for the four rotationally flexible, hydrogen bond patterns, namely, patterns 1, 3, 4, and 9 (see Figure 17 for energy landscapes demonstrating this flexibility). Hence, additional positive ψ angle anchors are defined “1+”, “1++”, “3+”, “3++”, “4+”, “4++”, “9+”, and “9++”, where the “+” and “++” denote dihedral angle ranges ($90^\circ > \psi > 0^\circ$) and ($\psi > 90^\circ$ or $\psi < -150^\circ$), respectively.
- 2 fine anchors that represent two “basket” states to reflect residues containing either π - or 3_{10} -helical bonds. Note that, in the fine representation, we do not split the π - or 3_{10} -helices into specific permutations of the possible bonds, as the probability of these anchors in the coarse description is low relative to the α -helical states (see Figure 8 for coarse anchor probabilities).

Note that there is a degree of overlap between the coarse and fine anchors. The fine anchors subsume certain coarse anchors in that the *quantity* of existing “C3-”, “C2-”, and “C1-” bonds is considered in the coarse-anchor definition. The fine anchors add the two hydrogen bonds formed with the residue (“C4-” and “C”), as well as considering not just the quantity of bonds but their precise permutations (i.e., there are five different anchors in which one bond exists). That said, the precise set of existing hydrogen bonds is only considered for residues in the pure α -helical condition, i.e., those containing neither π - nor 3_{10} -helical bonds.

II.4.3. Methods: Data Analysis. After the atomic, full-coordinate, space data from the MD simulations was binned into appropriate AA-states, analysis was performed on the resulting AA-state trajectories. At each time step, the status of the spanning hydrogen bonds and the dihedral angle ψ of each residue are evaluated. The AA-state corresponding to the state of these coarse variables is determined according to the anchor system employed (either coarse or fine). Certain data analysis is based solely on the instantaneous AA-state, e.g., AA-state trajectories and probabilities. Other metrics, e.g., net fluxes between anchors, are transition dependent and are evaluated with consideration to the prior AA-state.

Calculation of net fluxes provides the opportunity to determine a reaction pathway for better understanding of molecular mechanisms. Rather than determine this pathway from the few

long trajectories (e.g., by tracing the progression of individual amino acids throughout the entirety of these long trajectories), the net flux analysis breaks these long trajectories into trajectory fragments or transitions, occurring hundreds or thousands of times more often. Hence, reaction pathways can be more accurately computed from this ensemble of transitions—each computed over an ensemble of trajectories—which provides better insight into the reaction mechanism(s).

Consider a graph G with N nodes (“anchors”) and E edges (“milestones”) connecting these nodes. Analysis of the AA-state trajectories allows the calculation of q_{kb} , the total flux between anchors k and l , by determining the number of transitions from k to l ; similarly, q_{lk} , the total flux between anchors l and k , can be calculated. For processes at equilibrium and/or completely reversible transitions, $q_{kl} = q_{lk}$. However, for processes not at equilibrium, there is no requirement that $q_{kl} = q_{lk}$. For example, under external load, the unfolding pathways may differ from refolding pathways. To wit, a number of unfolding events, which do not refold, are observed. As key transitions may be “directional”, it is expedient to cast the trajectory data in terms of a net flux w_{kl} from k to l , where $w_{kl} = q_{kl} - q_{lk}$ is defined as the positive net flux from k to l . Moreover, on graph G depicting the net flux between nodes k and l , the edge connecting k and l is assigned the corresponding weight w_{kl} .

The MaxFlux algorithm,⁵⁷ as implemented for a milestoning graph,³⁷ seeks a path that connects the reactant and product states and that carries maximum flux. Another algorithm was proposed in the context of the Markov state model.⁵⁸ As used in the current study, the following algorithm finds the MaxFlux path:

Step 1. Find the link e_{\min} with the smallest weight $w_{\min} = \min(w_i)$ 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 nonmarked edges are removed is the “MaxFlux path”; it 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. 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.

Milestoning analysis requires calculation of the last milestone passed; thus, it requires history of the trajectory and not just the current position. The few (10–20) long trajectories at each load level are partitioned into trajectory fragments according to the algorithm described in section II.1. Briefly,

Step 1: Evaluate current anchor according to anchor system definition.

Step 2: Compare current anchor to previous anchor in order to determine if a milestone has been crossed: if yes, then proceed to Step 3; if no, then return to Step 1 at

Table 3. Fine-Anchor Definitions^a

Fine Anchor Designation	Quantity of α Helical Hbonds	Hydrogen Bond Status (1=Intact; 0=Broken)					ψ Angle Values
(# Indicates Hydrogen Bond Pattern)		C4-	C3-	C2-	C-	C	
32	5	1	1	1	1	1	all
31	4	1	0	1	1	1	all
30	4	0	1	1	1	1	all
29	4	1	1	1	1	0	all
28	4	1	1	1	0	1	all
27	4	1	1	0	1	1	all
26	3	0	0	1	1	1	all
25	3	1	0	1	1	0	all
24	3	0	1	1	1	0	all
23	3	1	1	1	0	0	all
22	3	1	0	1	0	1	all
21	3	0	1	1	0	1	all
20	3	1	0	0	1	1	all
19	3	0	1	0	1	1	all
18	3	1	1	0	1	0	all
17	3	1	1	0	0	1	all
16	2	0	0	1	0	1	all
15	2	0	0	1	1	0	all
14	2	1	0	1	0	0	all
13	2	0	1	1	0	0	all
12	2	0	0	0	1	1	all
11	2	1	0	0	1	0	all
10	2	0	1	0	1	0	all
9	2	1	0	0	0	1	-150°:0°
9+	2	1	0	0	0	1	0°:90°
9++	2	1	0	0	0	1	90°:180° or -180°:-150°
8	2	0	1	0	0	1	all
7	2	1	1	0	0	0	all
6	1	0	0	0	1	0	all
5	1	0	0	1	0	0	all
4	1	0	0	0	0	1	-150°:0°
4+	1	0	0	0	0	1	0°:90°
4++	1	0	0	0	0	1	90°:180° or -180°:-150°
3	1	1	0	0	0	0	-150°:0°
3+	1	1	0	0	0	0	0°:90°
3++	1	1	0	0	0	0	90°:180° or -180°:-150°
2	1	0	1	0	0	0	all
1	0	0	0	0	0	0	-150°:0°
1+	0	0	0	0	0	0	0°:90°
1++	0	0	0	0	0	0	90°:180° or -180°:-150°
π	NOTE: this fine anchor incorporates any residues with a π -helical hydrogen bond (and therefore not in a pure α -helical state)						
310	NOTE: this fine anchor incorporates any residues with a 310-helical hydrogen bond (and therefore not in a pure α -helical state)						

^aFine-anchors definitions based on the hydrogen bond pattern of five hydrogen bonds spanning the residue; dihedral angle ψ conformation; and π - or 3₁₀-helical bonds, if any. Additional positive- ψ angle anchors are defined for patterns 1, 3, 4, and 9, where the “+” and “++” denote angles ($90^\circ > \psi > 0^\circ$) and ($\psi > 90^\circ$ or $\psi < -150^\circ$), respectively, as CFMD simulations indicated that these patterns exhibited significant rotational flexibility. Fine anchors are also included for a residue that is not in the pure α -helical state, due to the presence of either 3₁₀- or π -helical bond states; in the CFMD simulations, these anchors were infrequently visited as compared to the α -helices. Hydrogen bonds are considered to exist if the N–O distance is <0.35 nm and the NH–O angle is >150°.

next trajectory frame after incrementing milestone lifetime counter.

Step 3: Compare milestone crossed to previous milestone crossed; if the milestones are different, then proceed to

Step 4, if not, then return to Step 1 at next trajectory frame after incrementing milestone lifetime counter.

Step 4: Increment transition counter n_{ij} where j is the new milestone. Store milestone i lifetime for calculation

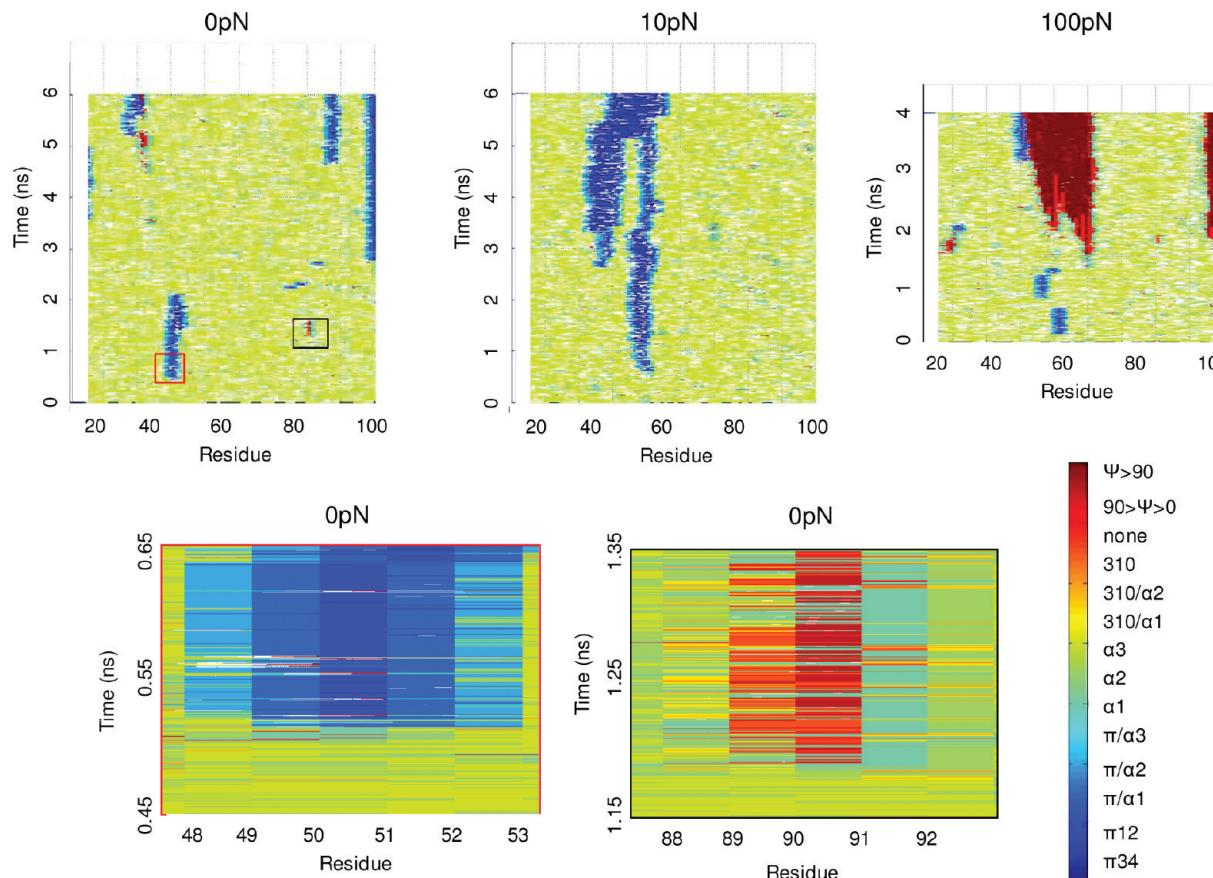


Figure 6. Sample amino acid-state trajectories. (top) Sample AA-state trajectories (time increasing from bottom to top) displaying creation/destruction of non- α -helical secondary structures for the three examined load levels (n = number of runs; Table 1): 0 pN ($n = 10$), 10 pN ($n = 10$), and 100 pN ($n = 20$). Amino-acid states are defined by the quantity and type of hydrogen bonds spanning a given residue as well as the residue's ψ angle. For these illustrations, hydrogen bonds are considered to exist if the N–O distance is <0.35 nm and the NH–O angle is $>150^\circ$. (bottom) Expanded view revealing transitory creation/destruction of non- α -helical secondary structures (left: π helix; right: unfolded domain) over a time scale of 10s of picoseconds and spanning about five sequential residues. Many transitions between AA-states reflect that these states are well mixed.

of $\langle \tau \rangle_i$. Increment milestone trajectory counter for new milestone, n_j . Return to Step 1.

This milestones analysis provides the necessary components ($\langle \tau \rangle$ and K) for the calculation of the mean first passage time $\langle \tau \rangle_{if}$ and the coarse-grained transition graph between states.

III. RESULTS

Results regarding the early events of unfolding and the extent to which they are affected by applied load level and amino-acid type and location are presented using two different coarse-grain state definitions: (i) *coarse anchors*, describing each residue's secondary structure, and (ii) *fine anchors*, describing the specific states of hydrogen bonds within the α -helical states.

III.1. Coarse Anchors. **III.1.1. Results: Coarse Anchors: Representative Amino-Acid State Trajectories.** Amino-acid state trajectory visualizations illustrate the helix's spatial and temporal structural evolution, as well as the relative variability between and within the three applied load levels. For each run, amino-acid states (or AA-states) are calculated for the helix's central 86-residue analysis section for each image saved at picosecond intervals for each run of the different three load levels. Notably, multiple, distinct transitory (lasting tens to hundreds of picoseconds) secondary structures may appear, evolve, and dissolve over time within a given simulation. Moreover, the states and transitions are well visited, demonstrating

that the data is sufficiently rich for milestones analysis. Nevertheless, some long-lived species (on the simulation time scale) are also observed. Hence, the present simulation is not sufficient to describe the entire unfolding process and we focus instead on early events in which a single amino acid loses its α -helical structure.

Amino-acid state trajectories illustrate the formation of persistent (longer than 1 ns) non- α -helical secondary structures, namely, π helices and their hybrids. While these non- α secondary structures form at the three load levels, their formation is strongly dependent upon load level. As illustrated in Figure 6, these secondary structures, in some cases, form at individual residues, propagate to neighboring residues, and potentially merge with like neighbors, while, in other cases, they are arrested and do not propagate along the helix. Critically, the data reveals that these non- α -helical structures are not necessarily intermediates to unfolding, and, in most cases, do not precede unfolding events.

In addition to the formation of non- α -helical structures, amino-acid state trajectories also illustrate unfolding events in which an individual residue's ψ angle rotates, in the positive direction, beyond 0° . As with the formation of π helices, unfolding events may initiate at individual residues and spread to neighboring residues. On the other hand, these unfolding events can also be arrested with helix refolding into either an α - or π -helical secondary structure. While π -helices are relatively

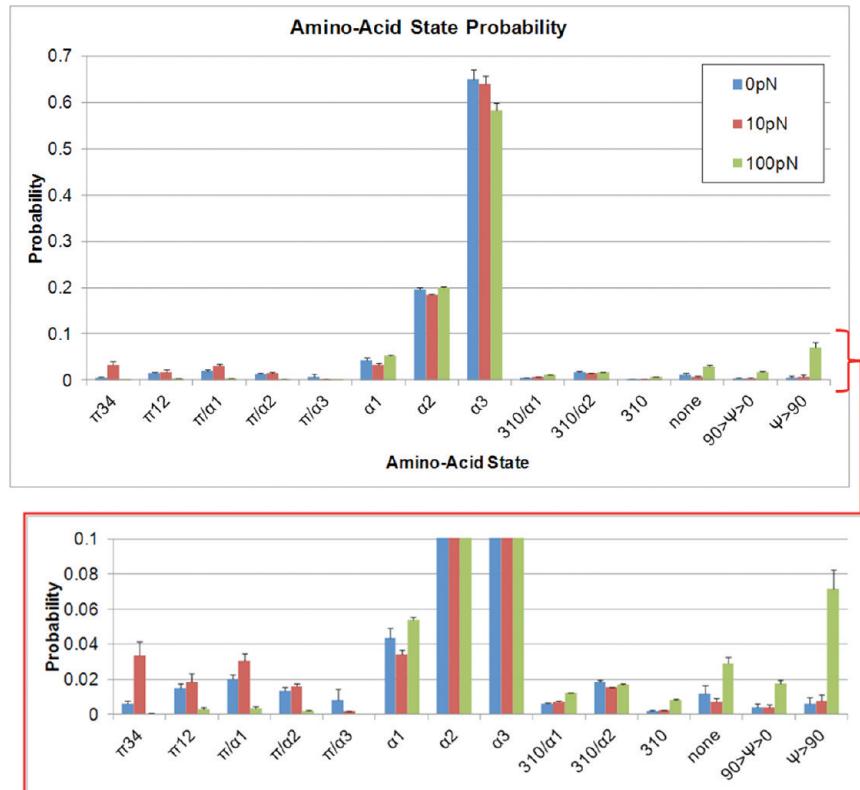


Figure 7. Amino-acid state probabilities as a function of load. (top) Average amino-acid state probabilities as a function of load level, based on all simulation times for each load level. Amino-acid states are defined by the quantity and type of hydrogen bonds spanning a given residue, as well as the residue's ψ angle. Values are the block averages and block standard errors where each block is an individual trajectory within each load level (see Table 1). (bottom) Expanded view of the lowest range of amino-acid state probabilities to illustrate in greater detail the non- α -helical secondary structure populations.

long-lived structures, it is clear from the amino-acid state trajectories that the 3_{10} -helices are relatively short-lived structures.

Interestingly, it is also possible for multiple non- α structures (either π helices or unfolded residues) to simultaneously exist along the helix but at different—and, perhaps surprisingly, sometimes adjacent—residues, within a particular helix, suggesting that helices generally have a heterogeneous AA state composition. While there are instances of structural transitions occurring from the helix's loaded, terminal ends, non- α structures predominantly form in the helix's interior.

While the evolution and propagation of persistent non- α -helical structures is clear from the amino-acid state trajectories, closer examination of the initiation of these structures reveals rapid transitions (on the order of a few picoseconds) between amino-acid states. These relatively rapid transitions mark the breaking and/or formation of individual hydrogen bonds. Critically, while large, macroscopic structural transitions (π helix formation and/or multiple-residue unfolding) occur only a handful of times within each simulation run, closer examination of all the data—but particularly at transitions—indicates that on a single-residue basis the chosen system states are well-visited and therefore statistically appropriate for milestoneing.⁴⁰

III.1.2. Results: Coarse Anchors: Amino Acid-State Probability. In addition to the AA-state trajectories, amino-acid state probabilities quantify each helix's secondary structure population at each load level. While trajectories of each simulation run qualitatively illustrate the system's evolution, AA-state probabilities provide a quantitative description of the time-averaged,

helix structural state, as well as indicating load-dependent, secondary structure propensities (Figure 7).

Consistent with the sample amino-acid state trajectories, the time-averaged amino-acid state probabilities reveal the proclivity of the 2fxmA sequence to retain its α -helical conformation (sum of α_3 , α_2 , α_1 probabilities $\sim 85\text{--}90\%$)—even under a constant external stretching force (≤ 100 pN) on the time scale of the simulation. Moreover, as expected, increasing load monotonically decreased the α_3 state population ($66\% \rightarrow 58\%$), as the applied axial load favors breaking at least some of these bonds. This trend is not apparent, however, for the far less prevalent α_2 ($\sim 19\text{--}20\%$) or α_1 ($\sim 3\text{--}5\%$) states that are similarly prevalent at the three load levels.

While increasing load diminishes the α_3 state population, it clearly plays a role in populating non- α -helical secondary structures. In particular, the 10 pN—and 0 pN to a somewhat lesser degree—load favors the formation of π helices, while the 100 pN load helix is virtually devoid of π helical structures. Notably, each π structural variant population is $<1\%$ for the 100 pN load helix. On the other hand, the 100 pN load favors the formation of 3_{10} -helices and demonstrates a much greater tendency to unfold the helix. The 100 pN load helix is significantly more unfolded, $\psi > 90^\circ$ state population $>7\%$ as compared to $<1\%$ for either the 0 or 10 pN load levels. Stated another way, the Pearson correlation coefficient of π -helical—including π_{34} , π_{12} , π/α_1 , π/α_2 , and π/α_3 —states' occupancy with load magnitude is -0.86 , while the Pearson correlation coefficient of 3_{10} -helical—including 3_{10} , $3_{10}/\alpha_1$, and $3_{10}/\alpha_2$ —states' occupancy with load magnitude is 0.99 . Increasing load

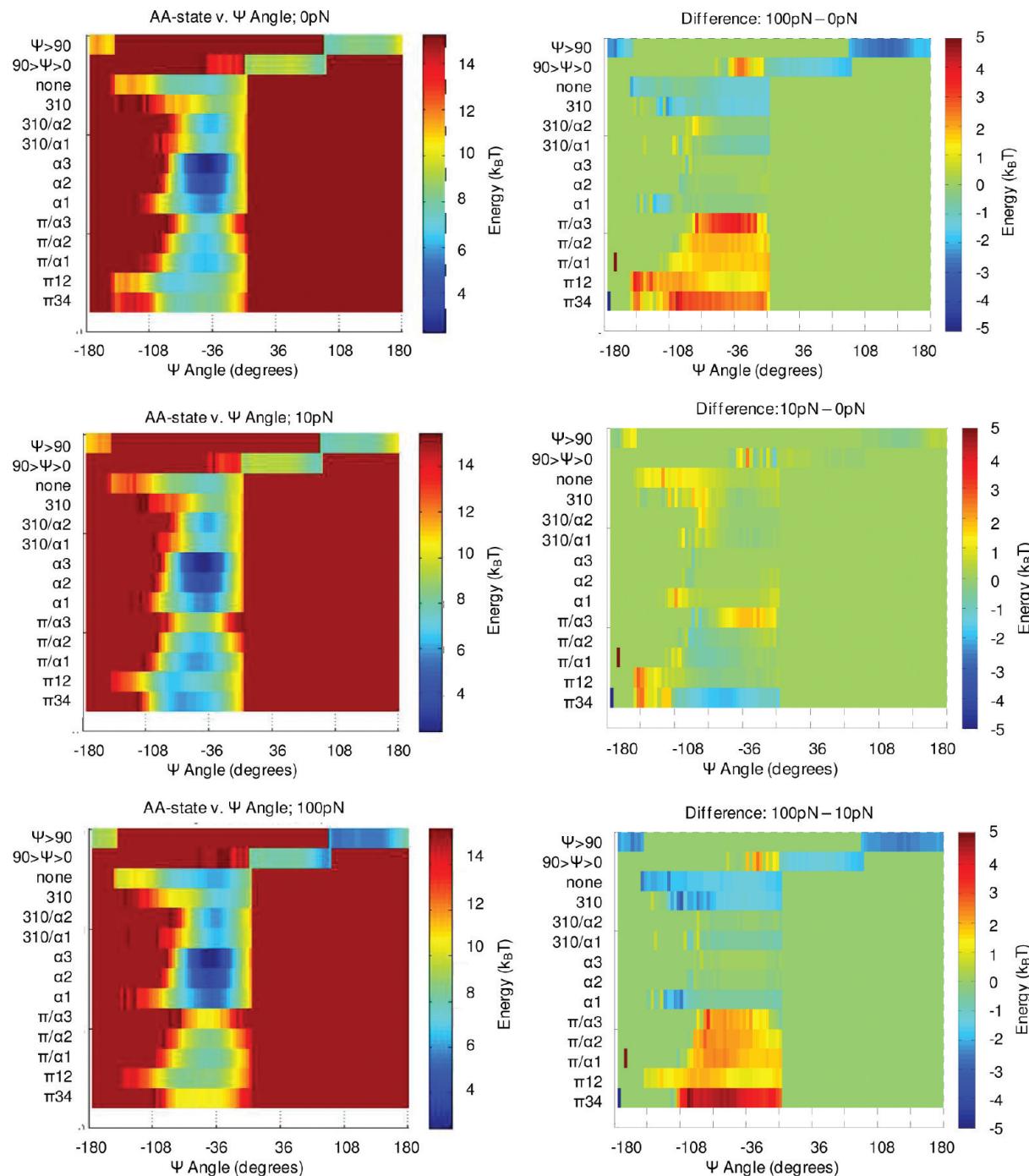


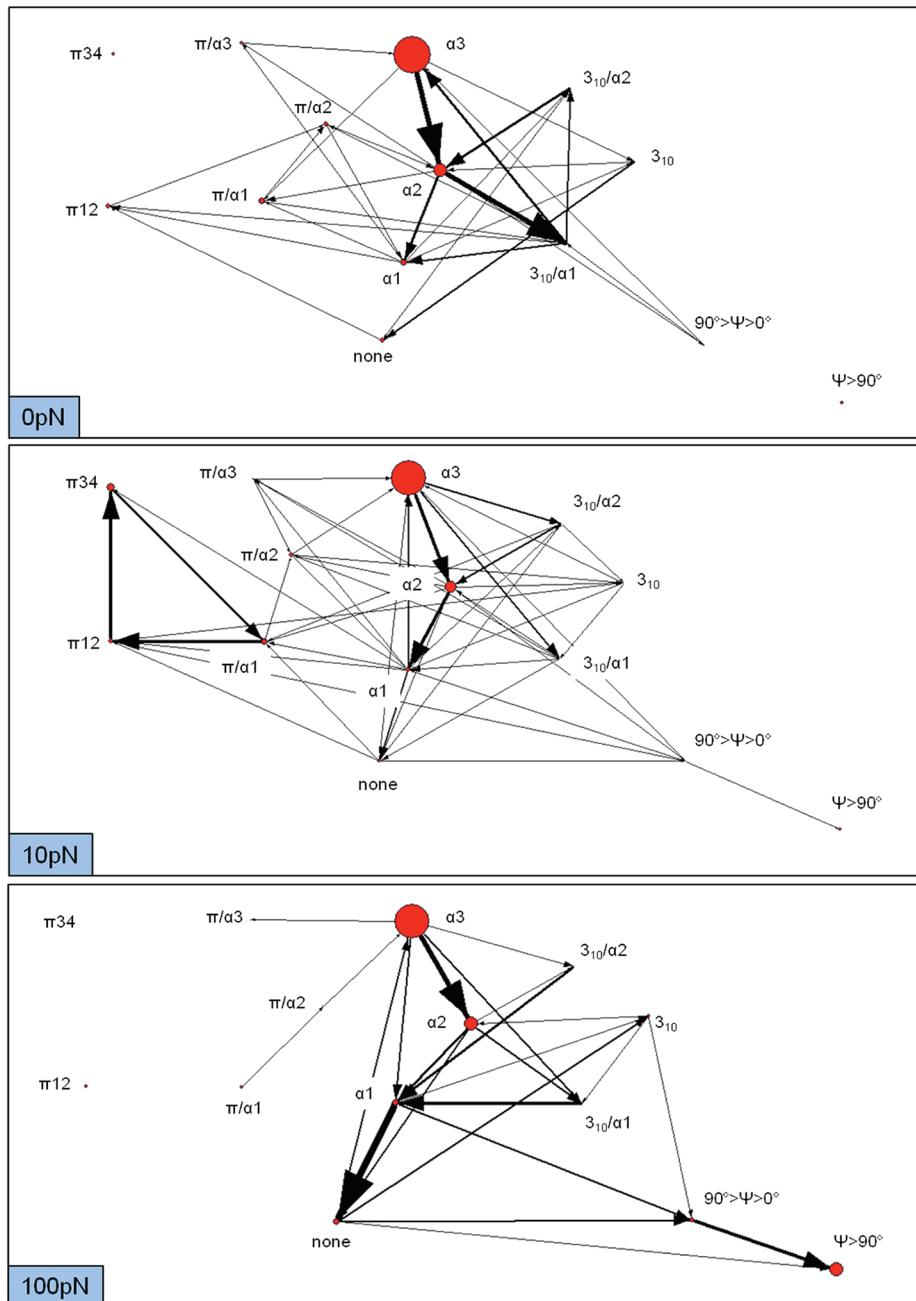
Figure 8. ψ angle distribution versus amino-acid state for each load. (left) One-dimensional, dihedral angle ψ potentials of mean force for each amino-acid state and load. Potentials of mean force calculated via Boltzmann inversion, where $E_{\text{AA-state}}(\psi) = -\ln[p_{\text{AA-state}}(\psi)]$, of dihedral angle ψ probability distributions for each amino-acid state. Probability distributions calculated over all simulation time for each load level. Simulations were run at 310 K; energy is in units of $k_B T$. (right) Differences in dihedral angle ψ potentials of mean force between load levels. Negative energy populations (blue regions) correspond to a relatively higher population—and subsequently lower energy—of the larger force.

primes the helix for unfolding by favoring the “longer” (relative to α -helical) 3_{10} -helical states, while the “shorter” π -helical states are eliminated at higher loads.

As was noted with the amino-acid state trajectories, the population of the 3_{10} -helix state is virtually nonexistent at lower loads ($\sim 0.2\%$ for both 0 and 10 pN), reflecting the lack of persistent structures of this type. Even with the highest load, the pure 3_{10} -helix state has a population of $< 1\%$ (0.8%). Conversely, the population of combined, pure π -helical states

(π_{34} , π_{12}) at 10 pN is $> 5\%$. Interestingly enough, the structural hybrid or transition states— π/α_1 , π/α_2 , $3_{10}/\alpha_3$, and $3_{10}/\alpha_2$ —are each visited as often as the pure non- α -helical states— π_{34} , π_{12} , and 3_{10} —reflecting the highly connected, frequently visited, state network responsible for the diffusive nature of the unfolding initiation process.

III.1.3. Results: Coarse Anchors: ψ Angle Distributions by Coarse Anchor. As alternative depictions of the coarse anchors, the ψ angle distributions for each AA-state demonstrate the



*Line width normalized within each load.

Figure 9. Flux networks as a function of load. Net fluxes between amino-acid states, illustrating channels between helical states, including non- α -helical secondary structures. Net flux between two states i and j , defined as $\Delta q_{ij} = q_{ij} - q_{ji}$, is calculated using all simulation time within each load level; arrows point in the direction of positive flux. State circle diameter is proportional to amino-acid state probability, while network edge thicknesses scale with the magnitude of net flux, normalized within each load level. (For visual clarity, fluxes below 5% of the maximum net flux for a given load level are omitted.) Possible unfolding pathways for the 0 pN scenario include $\alpha_3 \rightarrow \alpha_2 \rightarrow 3_{10}/\alpha_1 \rightarrow \text{none}$; this pathway does not connect to the unfolded, $\psi > 90^\circ$ state, perhaps due to the lack of applied load. A high flux 10 pN pathway is $\alpha_3 \rightarrow \alpha_2 \rightarrow \alpha_1 \rightarrow \text{none} \rightarrow 90^\circ > \psi > 0^\circ \rightarrow \psi > 90^\circ$, where the last two fluxes are comparatively small relative to the earlier net fluxes. A high flux 100 pN pathway is the same as the 10 pN example but has higher net fluxes for each graph edge.

relative “rotational flexibility” of a residue within its secondary structure (Figure 8). Given that unfolding is generally defined as a positive rotation of ψ beyond 0° , it is potentially insightful to examine the proclivity of each secondary structure to approach this threshold. Additionally, as others have hypothesized that these various secondary structures are (necessary) intermediates of unfolding, it is useful to understand how their

relative rotational flexibility—or lack thereof—may contribute to intermediate behavior.

While the α -helical states are noticeably the most visited—and, therefore, the lowest free energy—states, the ψ -angle distribution widths reflect the differing flexibilities of the corresponding secondary structures. In particular, there is a small but distinct increase in the ψ angle corresponding to the

energy minimum of the 3_{10} , α -, and π -helical states, consistent with expected ψ values based on literature values.

While there are notable differences between the distributions for different secondary structure states, it is clear from these ψ angle distributions that any existing hydrogen bonds constrain the residue so that unfolding is only possible when all hydrogen bonds are broken. Thus, it can be inferred that breaking the hydrogen bonds spanning a particular residue is necessary—but not, as will be seen later, sufficient—for unfolding. Moreover, it is clear from comparison of different hydrogen bond quantities within a specific type of secondary structure (i.e., comparing π/α_3 , π/α_2 , and π/α_1 or alternatively $\pi34$ and $\pi12$) that additional hydrogen bonds further constrain the rotational flexibility of the ψ angle. Note the distribution for the $\pi34$ and $\pi12$ AA-states in which the lower bond number state ($\pi12$) has noticeably greater flexibility—and therefore entropy—than the higher bond number state ($\pi34$); the same holds true for the pure α states in which α_1 has greater entropy than α_3 . Thus, as more bonds are broken, the residue has greater conformational flexibility. As will be seen later (Figure 10), hydrogen bond reformation is a critical step in the overall unfolding process; therefore, this flexibility may play a critical role in unfolding initiation.

Finally, these distributions confirm the efficacy of assigning secondary structure based on hydrogen-bond connectivity rather than the ψ angle value. Specifically, the overlap in amino-acid state ψ angle distributions clearly illustrates that, if the secondary structure designation was solely based on ψ angle, there would be multiple degeneracies corresponding to ψ angles between roughly -20 and -80° , due to the omission of information regarding the hydrogen bond states. Further, the 3_{10} structure can (and does) serve as a ψ -angle intermediate between the α -helical and unfolded states, while the π -helical structures, with their decreased ψ angles, are inconsistent with “primed to unfold” structure that would be expected in an unfolding intermediate.

III.1.4. Results: Coarse Anchors: Flux Networks. Net fluxes between two amino-acid states i and j , defined as $\Delta q_{ij} = q_{ij} - q_{ji}$, reveal the connectivity between states, as well as the directional tendencies of amino-acid state transitions (Figure 9). Arrows point in the direction of positive flux. Flux networks depict the “flow” of amino-acid states, illustrating the reaction pathway from the initial, native α_3 state, through various secondary structures, and ultimately to the completely unfolded $\psi>90^\circ$ state. Qualitative comparison between graphs for different load levels reveals the role of load in altering the reaction pathway through various helical secondary structures, particularly in determining whether certain non- α secondary structures are true unfolding intermediates or “off-the-pathway” intermediates to unfolding.

Here, the flux networks clearly illustrate that the unfolding pathway depends on the magnitude of applied load. The 0 pN net fluxes show connectivity between all states, but there is relatively little net flux into the nonbonded (*none*) and unfolded ($90^\circ>\psi>0^\circ$, $\psi>90^\circ$) states. While the 10 pN case shows a similarly small net flux to unfolding, the most glaring feature at this intermediate load level is the tendency to form relatively long-lived, π helices; the reasons for this are examined further in the next section. This tendency to form π -helices is consistent with the amino-acid state populations noted above. Further and consistent with the ψ angle distributions, the net flux pathways reveal that the π -helix formation at this load level

is not a structural intermediate to unfolding but rather a separate, “off-the-pathway”, diversionary state.

In stark contrast, the 100 pN flux networks clearly reveal the proclivity to unfold via breaking bonds in the α -helical state, possibly passing rapidly through the 3_{10} unfolding intermediate, followed by an accelerated rate of ψ angle rotation, consistent with the zero-hydrogen bond ψ angle distribution. Thus, these net flux reaction pathways illustrate that load magnitude affects not only the rate of unfolding but, critically, also the pathway—either by directly, albeit progressively, through unfolding intermediates (100 pN) or by being diverted into another, relatively stable but nonunfolding intermediate state (10 pN). Interestingly, the 10 and 100 pN graphs share a common high-flux unfolding pathway, in spite of the fact that the 10 pN has a proclivity to be diverted due to hydrogen bond reformation into the formation of π -helical structures.

The five highest, net fluxes at each load level provide a qualitative comparison between the loads (see Table 4). The states connected by these largest net fluxes indicate the helix’s behavior for each load. The five largest net fluxes for the 0 pN load are for transitions between the α - and 3_{10} -helical states and their structural intermediates. On the other hand, for 10 pN, three of its five largest net fluxes involve the formation of π helices, while the largest net fluxes of the 100 pN involve the later steps for unfolding, primarily the α_1 states. Interestingly, the magnitude of the largest net fluxes is roughly the same ($\sim 3\text{--}4$ events/ns), notable exceptions being the two highest for both the 0 and 100 pN.

While the individual load levels illustrate several heavily trafficked pathways toward unfolding, it is vital to note that the different amino-acid states are multiply connected via many parallel, competing pathways. Given that there is no distinct reaction pathway between the native (α_1 , α_2 , α_3) and completely unfolded ($\psi>90^\circ$) state, any methodology investigating the kinetics of the unfolding process must account for this high degree of multiplicity and connectivity. As discussed in the Methods section, the milestones methodology implicitly incorporates multiple, parallel channels in its formulation and is thus ideally suited for capturing the system’s dynamics.

III.1.5. Results: Coarse Anchors: Amino-Acid State Exit Probabilities. State exit probabilities clearly illustrate the overriding tendency to form—rather than to break—hydrogen bonds (Figure 10). This is particularly obvious for the α_2 state that has an 85% probability of forming a third bond, while a less than 10% likelihood of breaking another bond.

Hydrogen bond reformation tends to stabilize the secondary structure—even in the no-bond state. Bond reformation (in 3_{10} , α , or π secondary structure in the no-bond state) occurs >95% of the time for the low loads (0 and 10 pN) and >90% of the time for the highest load (100 pN). Notably, this effect disappears with rotation of the ψ dihedral to a positive angle; in such cases, the system—regardless of load—is >50% likely to continue rotating to the $\psi>90^\circ$ state. Thus, based on these state exit probabilities, it can be inferred that bond reformation tends to forestall the initiation of unfolding. Furthermore, while bond failure is necessary to unfold, these results further emphasize that even the breakage of *all* hydrogen bonds spanning a residue is insufficient to initiate unfolding. The hydrogen bonds rapidly reform unless significant positive ψ rotation prevents their reformation.

An interesting side effect of the α -helix’s tendency to reform hydrogen bonds is the load-dependent formation of

Table 4. Five Highest Net Fluxes between Coarse Anchors^a

0 pN			10 pN			100 pN		
net flux per ns	origin AA-state	destination AA-state	net flux per ns	origin AA-state	destination AA-state	net flux per ns	origin AA-state	destination AA-state
5.48	α_3	α_2	3.77	π_{12}	π_{34}	6.39	α_1	none
5.40	α_2	$3_{10}/\alpha_1$	3.63	α_2	α_1	5.06	α_3	α_2
2.61	$3_{10}/\alpha_1$	α_3	3.48	π/α_1	π_{12}	3.55	$90^\circ > \psi > 0^\circ$	$\psi > 90^\circ$
2.44	$3_{10}/\alpha_2$	α_2	3.38	α_3	α_2	3.25	$3_{10}/\alpha_1$	α_1
2.25	α_2	α_1	2.51	π_{34}	π/α_1	2.59	$3_{10}/\alpha_2$	α_1

^aFive largest net fluxes (per nanosecond, per trajectory) and their corresponding amino-acid states for each load level. Net fluxes are (i) calculated using all simulation data for each load level and (ii) normalized by the total simulation time within each load level.

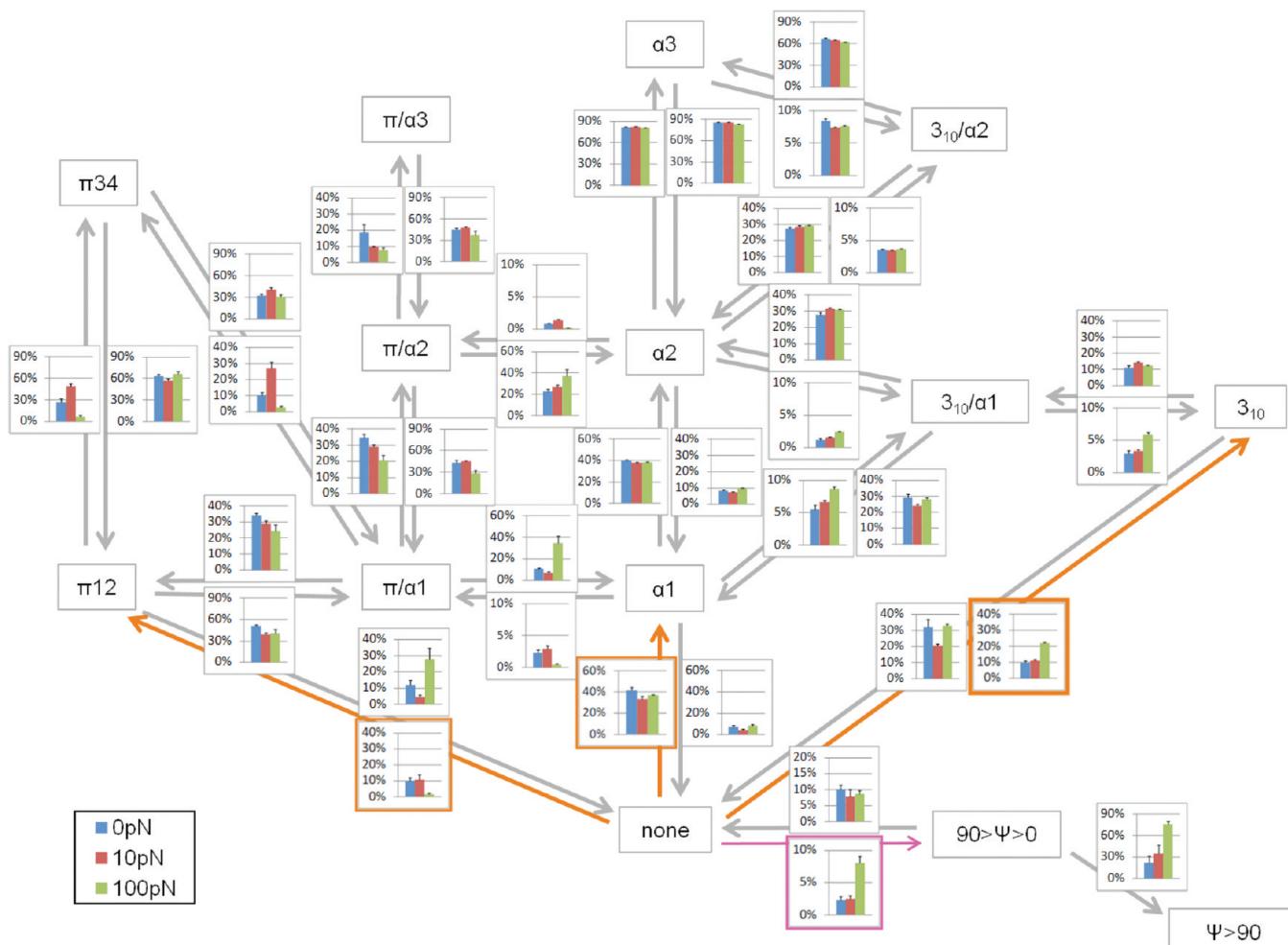


Figure 10. Amino-acid state exit probabilities as a function of load. Exit probabilities from each amino-acid state, calculated as the likelihood of transitioning from a given state to another connected state. Exit probabilities from state i to j are calculated as $p_{\text{exit},ij} = n_{ij}/\sum_k n_{ik}$, where the normalization is over all exits from state i to any state k . The sum of all exit probabilities for a given state equals 1. Probabilities are calculated over all residues and all simulation time within each load level; values are the block averages and block standard errors where each block is an individual trajectory within each load level (see Table 1). Certain, low-probability connectivities (e.g., α_3/α_1) are omitted for clarity. Exit probabilities (orange) illustrate the role of hydrogen bond reformation in the creation of non- α -helical structures. Meanwhile, exit probabilities (magenta) between *none* and $90^\circ > \psi > 0^\circ$ illustrate the role of load magnitude in the rotation of the dihedral angle ψ .

non- α -helical structures that favors π helices at 10 pN and 3_{10} helices at 100 pN. It is insightful to examine the formation of bonds from the no-bond state to the π_{12} , α_1 or 3_{10} states (Figure 10 in orange). While all three load levels are most likely to form an α -helical bond (with >35% likelihood), the 10 pN load level has a >10% likelihood to form either one or two π -helical bonds, while the 100 pN shows a >20% likelihood to form a 3_{10} -helical bond. A similar phenomenon is seen with the

α_1 state and its transitions to the π/α_1 , α_2 , or $3_{10}/\alpha_1$ states. Thus, examining states via their exit probabilities reveals the subtle connectivity differences underlying helix-level observations, such as a tendency to form π helices under a 10 pN axial load. Furthermore, these exit probabilities illustrate that non- α secondary structure formation is most likely the result of bond reformation—not breakage—and thus should not necessarily be thought of as a structural intermediate on the

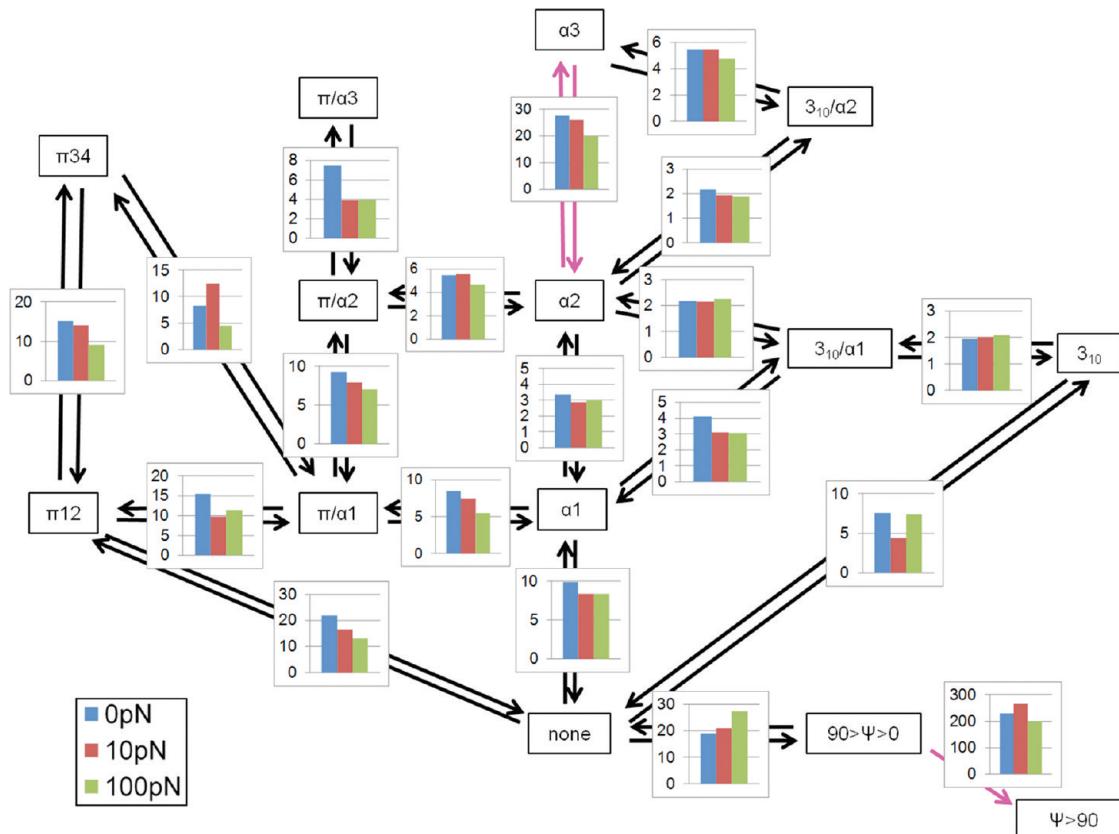


Figure 11. Average milestone lifetimes ($\langle \tau \rangle$) as a function of load. Average local lifetimes ($\langle \tau \rangle_i$) for select milestones (others omitted for clarity), as a function of applied load and calculated over all simulation time for each load level. Lifetimes calculated according to $\langle \tau \rangle_i = \sum_j [\int_0^\infty t \cdot K_{ij}(t) dt]$, averaging milestone transition times over all unique milestone visits (i.e., recrossings of the same milestone do not terminate the trajectory or the visit lifetime). Pink arrows indicate the initiating and terminating milestones in the calculation of mean first passage time for the coarse anchor system. Note that this figure illustrates time scales for the three essential processes within the initiation of the unfolding mechanism: (i) fast (~single ps) hydrogen bond formation/dissociation, (ii) intermediate (~tens of ps) milestone lifetimes, and (iii) slow (~hundreds of ps) dihedral angle rotations.

way to unfolding, but rather a structure—at least temporarily—“rescued” by hydrogen bond reformation.

III.1.6. Results: Coarse Anchors: Average milestone Lifetimes. In order to assess the role of load in accelerating individual amino-acid state transitions—in addition to governing the pathway of state transitions—average milestone lifetimes ($\langle \tau \rangle$) are calculated (Figure 11). As milestones occupy the boundaries between two states, they accommodate repeated transitions between these neighboring states. In contrast to transition-state theory, recrossing a surface in milestones is accounted for exactly (Figure 12). Hence, milestone lifetimes provide metrics to describe the kinetics of meaningful individual transitions and quantify their time scales.

A key observation of the average milestone lifetimes is that these individual structural transitions occur rapidly, on the order of one to tens of picoseconds. The one notable exception is the terminal, “completely” unfolded milestone ($90^\circ > \psi > 0^\circ$: $\psi > 90^\circ$), as it is assumed to be an unrecoverable transition (i.e., an absorbing boundary). Consistent with the representative amino-acid state trajectories and the absence of persistent 3_{10} -related structures, the average milestone lifetimes of these structures are typically less than 5 ps. On the other hand, the π -related structures are noticeably longer lasting, on the order of 10 ps or more.

The effect of increasing load on these structural transitions is typically to reduce their average milestone lifetime with few

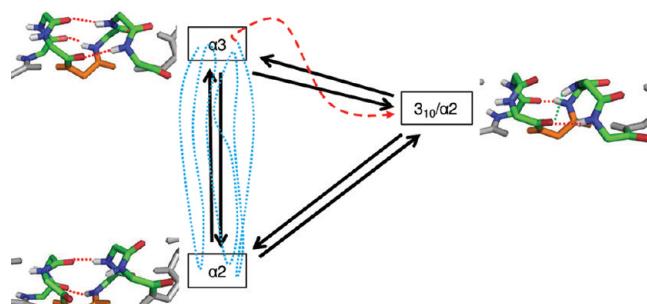


Figure 12. Local mean first passage time, $\langle \tau \rangle_i$: calculation schematic. $\langle \tau \rangle_i$ is the average lifetime of a milestone, not an anchor. As a milestone (α_3/α_2) is defined to be the interface between two anchors, this average lifetime can include many transitions between the same two anchors (α_3 and α_2). Calculation of the lifetime of a given visit to a milestone is initiated when first crossing the milestone from another milestone and is terminated as soon as the system transitions to a third anchor ($3_{10}/\alpha_2$), thereby crossing another milestone ($\alpha_3:3_{10}/\alpha_2$).

exceptions. While there is a qualitative role for load in average milestone lifetime, the effect of increasing load is typically only on the order of a few picoseconds. That said, an interesting feature of load in stabilizing the unfolded state can be seen with the monotonically increasing lifetime of the $none:90^\circ > \psi > 0^\circ$ milestone with increasing load, which is consistent with the notion that applied loads work to

prevent a return of the ψ angle to a negative value. Higher loads are presumably more effective in preventing this return.

III.1.7. Results: Coarse Anchors: Mean First Passage Time of Unfolding Initiation. To estimate the kinetics of unfolding initiation for both the unloaded (equilibrium) and loaded scenarios, the mean first passage time (MFPT) was calculated from a native α -helix ($\alpha_3:\alpha_2$) to a completely rotated ψ angle ($90^\circ > \psi > 0^\circ : \psi > 90^\circ$) milestone (Figure 13). The MFPT calculation combines the average lifetimes of individual structural transitions/milestones with the multiply connected, parallel pathways of unfolding as represented by the transition probabilities, to provide a single MFPT to the fully unfolding state. These estimates also provide insight into the role of load in unfolding. For the coarse anchor system, the calculated MFPTs are 4.19 ns (0 pN), 5.16 ns (10 pN), and 2.40 ns (100 pN), markedly nonmonotonic.

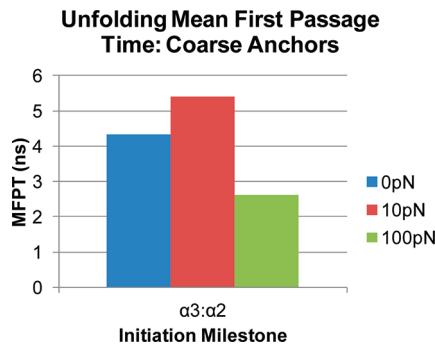


Figure 13. Mean first passage times for unfolding a single residue as a function of load. Mean first passage time from the initiation milestone ($\alpha_3 : \alpha_2$) to the destination milestone ($90^\circ > \psi > 0^\circ : \psi > 90^\circ$) calculated for each load level via $\langle \tau \rangle_{if} = P_{init} [I - K_{final}]^{-1} \langle \tau \rangle$. Calculation of MFPTs is based on all simulation data for a given load level.

The MFPT between the native, initial milestone ($\alpha_3:\alpha_2$) and the fully unfolded ($90^\circ > \psi > 0^\circ : \psi > 90^\circ$) milestone displays a “kinetic turnover” in which the MFPT for the intermediate load level (10 pN) is higher than either the unloaded (0 pN) or high load level (100 pN). Interestingly, this nonmonotonic MFPT is in marked contrast to the monotonically lower average lifetimes of individual milestones for different loads, indicating that unfolding MFPT is not a serial, direct sum of individual structural transitions but rather a change in the path.

Furthermore, the estimated MFPTs for unfolding initiation are on the order of a few nanoseconds, although individual milestone lifetimes are at most ~ 30 ps—with the one exception of the final, fully unfolded ($90^\circ > \psi > 0^\circ : \psi > 90^\circ$) milestone lifetime (~ 200 ps). The several orders of magnitude difference between individual amino-state transitions and the overall unfolding initiation process reflects the diffusive, highly connected, frequently traversed trajectory (shown in Figure 9) in which many states are visited before randomly “stumbling” on the final state. Remarkably, even a relatively large, constant axial load (100 pN), is insufficient to “pull” the system directly through its intermediates; here again, hydrogen bond reformation rescues the vulnerable, no hydrogen-bond secondary structure before chance structural disturbances have the opportunity to rotate the dihedral angle to a positive value.

III.1.8. Results: Coarse Anchors: Amino-Acid State Dependence on Amino-Acid Type. The dependence of amino-acid state on amino-acid type is examined to assess the proclivity, if

any, of specific amino acids and/or types (electrostatic, polar, hydrophobic, etc.) to transition to a non- α -helical secondary structures (Table 5). Pearson correlation coefficients of amino-acid state occupancies with increasing load magnitude were calculated for each amino acid type. Likewise, Pearson correlation coefficients are calculated between the amino-acid state occupancy percentage and increasing load magnitude for each amino-acid type and occupancy state combination. Given the helix’s amphipathic nature, including a number of alternating charged and hydrophobic residues consistent with a helix found in a coiled-coil, at the outset it was not obvious whether amino-acid type plays a role in helix unfolding. For example, it is possible that hydrophobic residues, which are normally in contact with the complementary hydrophobic side of the helix coiled coil partner, may become unstable when isolating the individual helix as has been done in the current study.

Table 5. Correlation of AA-State Occupancy with Load Level as a Function of Amino-Acid Type^a

Residue Type	Presence in 2fxmA Analysis Section		Pearson Correlation of Load with AA-state Occupancy					
	Quantity	%	π , π/α	310, 310/ α	α	none	$90^\circ > \psi > 0^\circ : \psi > 90^\circ$	
Charged	Arg	5	6%	-0.88	0.99	-1.00	1.00	0.98
	Lys	12	14%	-0.89	0.99	-0.91	0.91	1.00
	Asp	6	7%	-0.99	1.00	-0.99	0.98	0.99
	Glu	19	22%	-0.95	0.99	-0.94	1.00	0.99
Polar	Ser	2	2%	-0.93	0.93	0.92	0.67	0.94
	Thr	1	1%	1.00	1.00	-1.00	1.00	1.00
	Asn	5	6%	-0.64	0.99	-0.75	0.84	0.98
	Gln	7	8%	-0.65	0.99	-0.41	0.98	1.00
Hydrophobic	Cys	1	1%	-0.50	0.98	-0.43	0.91	1.00
	Ala	8	9%	-0.61	1.00	-0.86	0.90	0.99
	Ile	2	2%	-0.48	0.87	0.18	0.94	0.89
	Leu	12	14%	-0.96	1.00	-0.79	0.95	1.00
Hydrophobic	Met	3	3%	-0.94	0.85	-0.96	1.00	0.99
	Val	3	3%	-0.79	0.99	-0.86	0.91	1.00

^aPearson correlation coefficients of applied load with amino-acid state occupancies (Figure 7). Composition of the 2fxmA helix’s central, 86-residue, analysis section, grouped by amino acid type. For simplicity, correlations are calculated with state types (e.g., α) such that coarse anchors with different bond quantities (e.g., α_1 , α_2 , α_3) are summed prior to correlation calculation. Positive correlations indicate an increase in state occupancy with load. Correlations are calculated between the AA-state occupancy percentage and load magnitude for each amino-acid type and state occupancy combination. AA-state occupancy calculated as the average over all instances of each residue type within the analysis section for all simulation times within each load level. Note that not all amino acid types are present in the studied helix. Residues absent from the interior analysis section of 2fxmA are Gly, Pro, Tyr, Phe, Trp, and His.

Examination of the data reveals relatively consistent amino-acid state distributions for each amino acid within individual load levels (Figure 14). There are noticeable differences in amino-acid state populations across load levels, reflecting the overall state population differences with load. In particular, the 10 pN state distributions reflect a greater tendency to form π -helical structures, while the 100 pN distributions show a greater tendency to break bonds and unfold. These observations are supported by the Pearson correlation calculations in which the π - and α -helical states show a strong negative correlation with load magnitude, indicating that increased loads lead to lower occupancies of these states. Meanwhile,

3_{10} -helical, no-bond (*none*), and positive dihedral angle ψ states have strong positive correlations with load magnitude, indicating that they are increasingly populated with increased load.

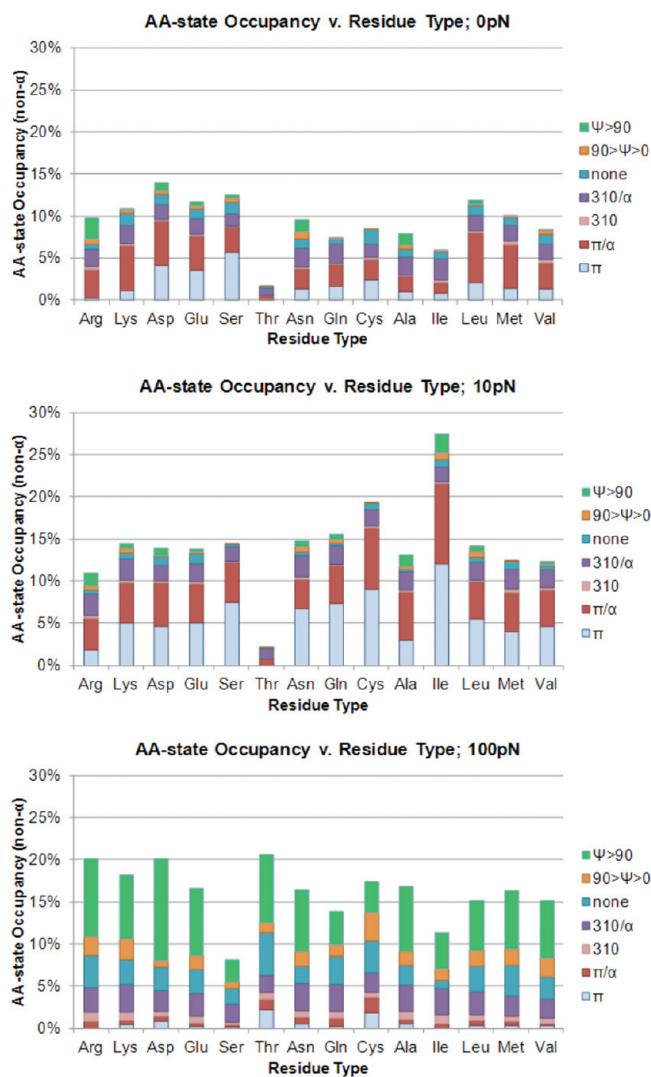


Figure 14. Amino-acid type dependence of AA-state occupancy as a function of load. Simulation time (%) spent in non- α -helical amino-acid states for each of the 14 amino-acid types found in the helix's central, 86-residue, section as a function of applied load level (for clarity, time spent in α states omitted). Percentages calculated for each amino acid type are average values over all instances of each amino acid within the analysis section over all data for each load but presented as a per-amino-acid basis. For simplicity, occupancies are calculated with state types (e.g., π) such that coarse anchors with different bond quantities (e.g., $\pi12$, $\pi34$) are summed.

The two major outliers, threonine and isoleucine, are relatively rare within the helix's studied section, appearing only once and twice, respectively. While there are other amino acids appearing in similar quantities without having outlier state populations, such as cysteine and serine, there is no obvious dependence of amino-acid state population of a non-helical structure on specific amino acid.

There also appears to be no major dependence on amino-acid type. The four electrically charged amino acids present in the helix's 86-residue interior (Arg, Lys, Asp, Glu) have similar distributions to those of the nonoutlier hydrophobic amino

acids (Ala, Leu, Met, Val). Thus, there does not appear to be a dependence of amino-acid state on amino-acid type, justifying the subsequent analysis in which the state populations and kinetics for all residues are lumped together. Consistent with the qualitative observations shown in Figure 14, correlations (not shown) calculated between amino-acid type (electrostatic, noncharged polar, and hydrophobic) and state occupancy revealed all correlations to be less than 0.38, with 14 of the 24 having correlations less than 0.10.

III.1.9. Results: Coarse Anchors: Amino-Acid State Occupancy with Helix Position. The dependence of amino-acid state on residue position was calculated for each load to reveal any possible location dependence of unfolding (Figure 15). Average amino-acid state occupancy was calculated for each amino acid position in the interior analysis section, averaged over all simulation time for a given load level. For comparison purposes, the sequence of the analysis section is repeated and coded for amino acid type, indicating a relatively even distribution of the three major amino acid types (electrostatic, hydrophobic, and polar uncharged) along the helix length.

In general, the amino-acid state occupancies versus residue number indicate no distinguishable dependence of amino-acid state on position along the helix for the 0 and 10 pN load magnitudes: the Pearson correlation coefficient of the non- α -helical state population (corresponding to the total height of each state occupancy presented in Figure 15) with respect to residue number is -0.00 for 0 pN and -0.11 for 10 pN (Table 6). Interestingly, though, the amino-acid state occupancies of these lower load levels show a significant population of π -helical states, noted previously, distributed along the helix in a seemingly random fashion.

In contrast, the Pearson correlation coefficient of non- α -helical states with residue position is 0.80 for the 100 pN case. Indeed, the 100 pN load level state occupancies show a marked unfolding tendency toward the C-terminal. As both the N- and C-terminals of the helix have a group of 18 residues that are loaded with equal but opposite forces, this C-terminal tendency cannot be solely attributed to their proximity to loaded residues. The analysis section of the sequence near the C-terminal provides no clear explanation for the increase in unfolding in this region. That said, the tendency of the C-terminal to (un)fold first has been previously reported; in particular, it was observed computationally, for example, in refs 59 and 60.

The relatively even distribution of amino acid type along the helix length—particularly the lack of distinct groupings—and the resulting lack of correlation for state occupancy with residue position (with the exception of 100 pN which has a proclivity to unfold from the C-terminal end) underscores the invariance of unfolding with amino acid type, as illustrated in Figure 14. If amino acid type is a strong determinant of state occupancy, then it would be expected that state occupancies would be non-normal at each position of the amino acid. In contrast, regions of high or low state occupancies appear along longer, multiresidue lengths of helix.

Mean first passage times for the initiation of helix unfolding, which were calculated for three different residue (sub)sections within the helix, reveal a small dependence on residue location (Table 7). We note that this location dependence is roughly independent of load, as the trend in MFPTs is essentially the same for each (sub)section. There is a roughly 0.5 ns (or $\sim 10\text{--}20\%$) difference in MFPTs for the 10-residue samples relative to the full analysis section for each load level. Moreover, the

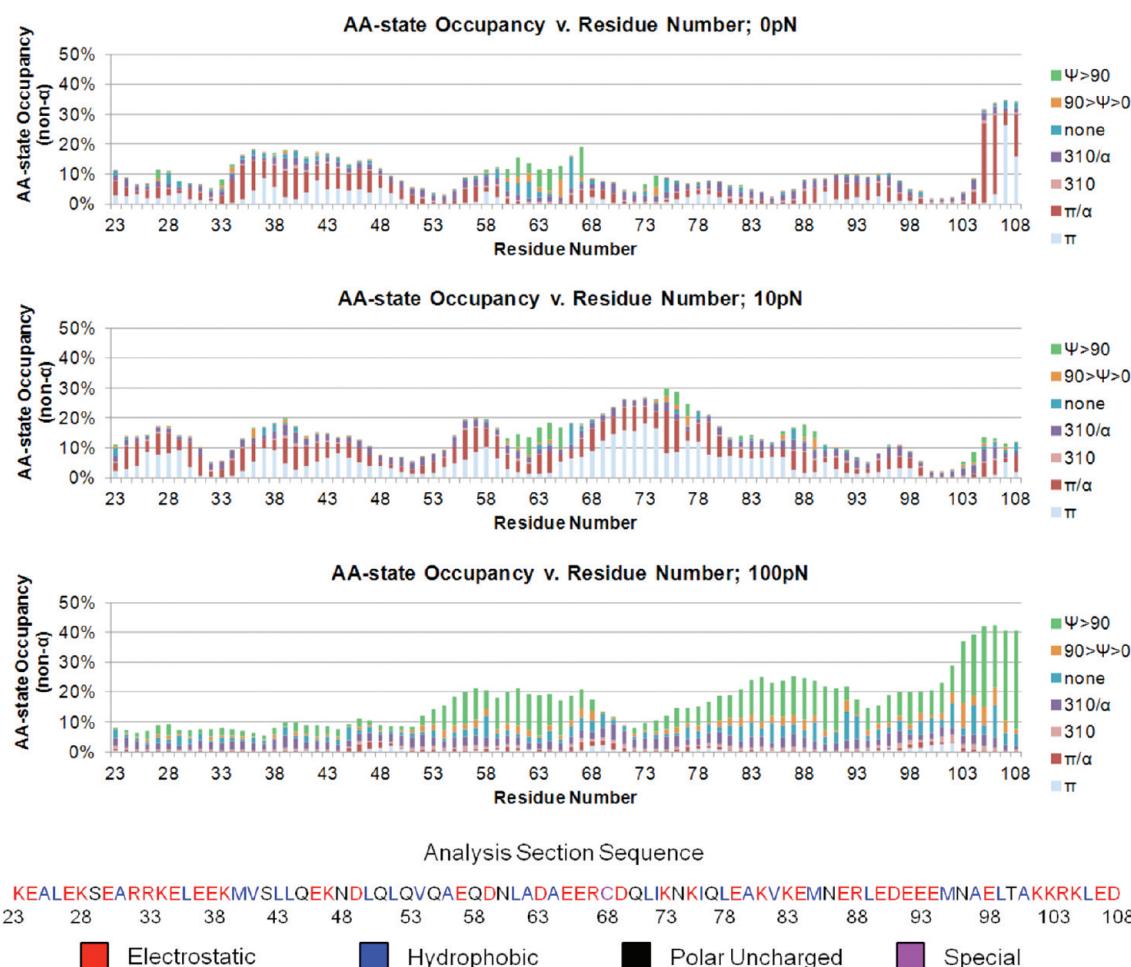


Figure 15. AA-state occupancy dependence on helix position as a function of load. Average amino-acid state occupancy versus residue number for the three load levels. Percentages are calculated with respect to all simulation time for a given load level at the residue position indicated. Pure α -helical states ($\alpha_1, \alpha_2, \alpha_3$) are omitted for clarity. For simplicity, occupancies are calculated with state types (e.g., π) such that coarse anchors with different bond quantities (e.g., π_{12}, π_{34}) are summed. The sequence of the analysis section is repeated for comparison and highlighted to reflect amino acid type.

Table 6. Correlation of AA-State Occupancy with Residue Position^a

load level (pN)	Pearson correlation of AA-state occupancy with residue position	
	non- α -helical state occupancy	non- α -helical + non- $\psi > 90^\circ$ state occupancy
0	0.00	0.02
10	-0.11	-0.16
100	0.80	0.71

^aPearson correlation coefficients for amino-acid state occupancy with residue position for all non- α -helical AA-states and non- α -helical + non- $\psi > 90^\circ$ AA-states for each load level. Correlation coefficients were calculated after summing the AA-state occupancies of each residue position across all simulation data for each load level. The non- α correlations indicate the extent of helix perturbation into non- α -helical structures of any kind; the non- α -helical + non- $\psi > 90^\circ$ correlations are included to reduce any influence of long-term unfolding on correlations.

N-terminal unfolding times are longer than the full analysis section—for each load level. Meanwhile, the MFPT for the C-terminal shows a faster unfolding time for the 100 pN load level yet a slower one for the 10 pN load level. In either case, the times, with respect to load level, follow the same

Table 7. Mean First Passage Times for Unfolding Initiation as a Function of Helix Position^a

load (pN)	full analysis section	N-terminal 10 residues w/ in analysis section	C-terminal 10 residues w/ in analysis section
	MFPT (ns)	MFPT (ns)	MFPT (ns)
0	4.2	4.6	n/a
10	5.2	5.8	5.5
100	2.4	2.9	2.0

^aMean first passage times of helix unfolding initiation (between the $\alpha_3:\alpha_2$ and $90^\circ > \psi > 0^\circ: \psi > 90^\circ$ milestones for each load level) for three different residue (sub)sections within the helix's analysis section. The "full analysis section" results reflect all 86-interior residues. Results from the N- and C-terminal 10 residues within the analysis section are calculated using K and $\langle \tau \rangle$ data from only these sections. In all cases, all simulation time within each load level was used. Note that the C-terminal MFPT for 0 pN is not applicable, as none of these residues unfolded during the simulations.

nonmonotonic trend in which the 10 pN has the highest MFPT regardless of residue position. These results underscore the benefit of simulating a long helix in which any end-effects can be examined, as well as the concept of an "average" residue behavior tested.

III.1.10. Results: Coarse Anchors: MFPT Dependence on Hydrogen Bond Definition. MFPTs for each load level were calculated to determine the sensitivity of the findings to the precise definition of the hydrogen bond (Table 8). Notably, changing either limit—the N–O distance (from 0.35 to 0.40 nm) or the NH–O angle (from 150 to 120°)—resulted in a less than 2% difference in the calculated MFPT for any load level. While all MFPTs calculated with altered hydrogen bond limits were within 1% of the standard definition, we note that changing the NH–O angle limit resulted in a slightly greater change in MFPT for 10 pN (0.74% vs. 0.37%) and a slightly smaller change in MFPT with 100 pN (0.76% vs. 1.14%).

Table 8. MFPT Dependence on Hydrogen Bond Definition^a

load (pN)	Hbond limits: distance = 0.35 nm, angle = 150°	Hbond limits: distance = 0.40 nm, angle = 150°	Hbond limits: distance = 0.35 nm, angle = 120°
	MFPT (ns)	MFPT (ns)	MFPT (ns)
0	4.43	4.42	4.45
10	5.40	5.42	5.44
100	2.63	2.66	2.65

^aMean first passage time (MFPT) of the initiation of helix unfolding from the α_1 , $\alpha_3:\alpha_2$ milestone to the absorbing $90^\circ > \psi > 0^\circ : \psi > 90^\circ$ milestone using data from the coarse anchor system. The geometrical limits on hydrogen bond existence used throughout the current work are based on two criteria: N–O distance < 0.35 nm and HN–O angle $> 150^\circ$. These results—in which each of these two criteria are independently changed—represent a test on the sensitivity of the MFPT calculation. Quantities used in the calculation of MFPT are based on all simulation data within each load level according to the listed hydrogen bond criteria.

Considering the extremely small difference in calculated MFPT, these results indicate that the reported unfolding initiation times are relatively insensitive to the precise hydrogen bond definition. Furthermore, as changes to the hydrogen bond limits are expected to change the relative population of various anchors as bonds “exist” over a greater conformational space, the relative indifference of MFPT to hydrogen bond limits provides further support for the robustness of the MFPT calculation to anchor choice. This robustness is further evidenced by the comparison between MFPT calculated with coarse and fine anchor sets (see Table 9).

We note that the difference in calculated MFPT with altered hydrogen bond limits (less than 2%) is much smaller than the difference with residue location (less than 20%) (see Table 8). This comparison illustrates the greater dependence of the MFPT calculation on the underlying simulation data, via different residue sets, than on the partitioning of the same data within different anchors. The importance of MFPT calculation on underlying simulation data underscores the value of the residue property averaging done here as individual residues are subjected to a variety of “boundary conditions” based on the condition and type of neighboring residues. Results based on shorter helices in which end effects are expected to dominate may not be applicable to more constrained helices (and vice versa).

III.2. Fine Anchors. In addition to revealing the mechanisms and kinetics of unfolding, as it relates to the formation of non- α -helical secondary structures, via the coarse anchor set, the milestones methodology can be applied to the same simulations to further examine the progression of unfolding through the predominantly α -helical states at a finer resolution. The fine

resolution analysis can demonstrate whether the same mechanisms identified with coarser resolution remain the same on the finer scale. In particular, this finer-resolution anchor set allows an investigation of the sensitivity of the milestones calculation to choice of anchors and coarse variables.

III.2.1. Results: Fine Anchors: ψ Angle Distribution v. Fine Anchor. In order to determine the relative rotational flexibilities of residues within each fine anchor, potentials of mean force (PMF) are calculated across all residues in the helix’s interior over all simulation time within a given load (60 ns) (Figure 17). These residue rotational flexibilities indicate (i) the rotational constraints placed on the residue by the existing hydrogen bonds and perhaps more subtly (ii) the criticality of the specific hydrogen bonds surrounding a given residue. The relative ability to explore positive dihedral angles ψ for each fine anchor illustrates the pattern’s propensity to unfold given its set of hydrogen bonds.

Comparisons across load levels reveal similarities in residue rotational flexibilities that are independent of load magnitude. For all load levels, positive dihedral angle ψ rotation beyond 0° is possible in only 4 of the 32 distinct permutations of the five residue-specific hydrogen bonds, namely, patterns 9 (C4- and C hydrogen bonds), 4 (C hydrogen bond), 3 (C4- hydrogen bond), and 1 (no hydrogen bonds) (Figure 16). Notably, the dihedral angle ψ rotations of all remaining patterns are restricted to negative values. Given that these four patterns show flexibility in the dihedral angle ψ with these bonds existing, it is not necessary to break these hydrogen bonds in order to rotate the dihedral angle ψ to initiate unfolding. Furthermore, of the four patterns demonstrating dihedral angle ψ flexibility, the fine anchor in which only the C bond exists (fine anchor 4) shows a lower population—and therefore higher energy—in the $\psi > 0$ range than the C4- hydrogen bond fine anchor 3, thereby indicating that the C4- hydrogen bond is a stronger constraint on the residue flexibility than the C hydrogen bond. These observations indicate that at least three of the five local hydrogen bonds (all except the C4- and C- hydrogen bonds) break prior to dihedral angle ψ rotation, providing support for the choice of coarse anchors in which the bonds spanning a residue in helical conformation (the C3-, C2-, and C1- hydrogen bonds in the fine anchors) define the presence of the α -helical state.

In addition to observations of the most flexible states, it is clear from the PMFs within each load level that fine anchor population and flexibility are not only dependent on the quantity of existing bonds but also highly dependent on the precise combination of bonds. While the 5- and 4-hydrogen bond patterns (32 and 27–31, respectively) show negligible difference in either free-energy-well depth or flexibility, the patterns within the 3-, 2-, and 1-hydrogen bond groups (patterns 17–26, 7–16, and 2–6, respectively) have clear and important differences. It is instructive to compare, for instance, both the well depth and spread for fine anchor 23 (C2-, C4-, and C3- hydrogen bonds) with that of fine anchor 24 (C2-, C-, and C3- hydrogen bonds). While both patterns have three hydrogen bonds existing, their properties—including both the spread and depth of their respective distributions—are markedly different. These differences may indicate that either these patterns do not appear with great frequency or, when they do appear, they quickly transition to another fine anchor through either hydrogen bond formation or failure.

Load dependencies are most obvious in the population of individual patterns as represented by the depth of the energy

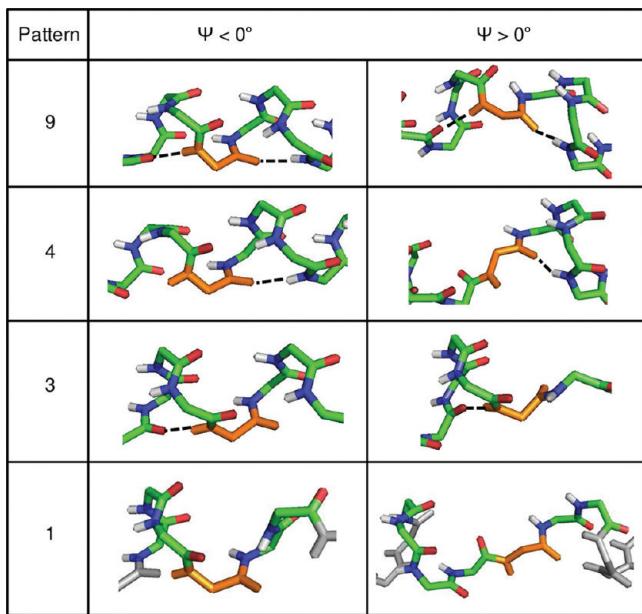


Figure 16. Hydrogen-bond patterns with marked $\psi > 0^\circ$ flexibility. Four patterns had marked positive dihedral angle ψ distributions: pattern 9 (10001), 4 (00001), 3 (10000), and 1 (00000). They represent permutations of the intact C4- and C bond pair; these two hydrogen bonds are those formed by—and thus not spanning—the residue of interest (orange). Notably, no pattern with a hydrogen bond spanning the residue has a noticeable positive dihedral angle ψ distribution. Further, we note that these two bonds are thus not included in the coarse anchors or AA-states.

well. In particular, it is clear that the 100 pN load causes a higher rate of unfolding, since the flexible patterns (1, 3, 4, and 9) have deeper energy wells in both the positive and negative ψ regions relative to both lower load levels. In comparison, the 10 pN PMFs have remarkably shallow energy wells in these flexible patterns even in comparison to the 0 pN load level, indicating that unfolding occurs less frequently at this intermediate load level.

While the depths of the various lower hydrogen-bond quantity states are noticeably different between loads, pattern flexibilities are remarkably similar across load levels. Given these similarities in flexibility, it appears that applied loads are more likely to affect the transitions between patterns rather than the properties of the patterns themselves. Load's effect on pattern transitions is in contrast to the picture of the coarse anchors in which the flexibility of AA-states was partially dependent on load magnitude (see Figure 8). This difference between coarse and fine anchors may be a function of the choice of fine anchors as permutations of the purely α -helical states only.

III.2.2. Results: Fine Anchors: Net Flux Networks. Graphs of net flux, i.e., flux networks, between the fine anchors illustrate the density of connections between individual patterns of the purely α -helical system (Figure 18). Reaction pathway calculations through these flux networks, via the MaxFlux algorithm, are presented in Figure 19. As with the coarse anchor system, large net fluxes indicate a likely connection along an unfolding pathway. The graphical depiction of reaction pathways of unfolding through the α -helical, fine anchor system indicates the highly parallel nature of unfolding pathways. While certain pathways may be emphasized, as we do in the MaxFlux analysis, it is apparent that unfolding can occur in

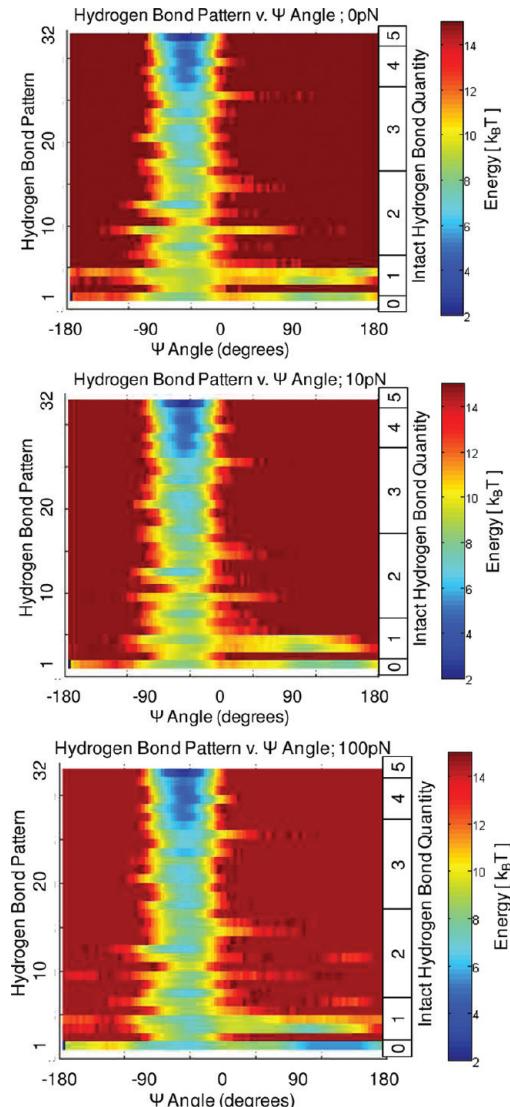


Figure 17. Hydrogen-bond pattern dihedral angle ψ potentials of mean force (PMF) of the dihedral angle ψ for the 32 hydrogen-bond patterns as a function of load. Hydrogen bonds associated with a given α -helical residue are used to determine an instantaneous fine anchor associated with the permutation of existing bonds. In addition to three spanning the residue, these hydrogen bonds include two bound to the residue's amide and carbonyl groups. The dihedral angle ψ distribution for each fine anchor within a load is calculated across all 90 interior residues of the helix and normalized with respect to total simulation time for that load. The PMF is calculated via Boltzmann inversion of the distribution for each fine anchor: $E_{\text{HBond Pattern}}(\psi) = -\ln[p_{\text{HBond Pattern}}(\psi)]$. Simulations were run at 310 K, energy is in units of $k_B T$.

many different ways at each load level. The highly random, parallel nature of the unfolding pathways underscores the need for a comprehensive kinetic theory accounting for reaction pathway redundancy; again, the milestone theory provides an ideal tool for analyzing such systems.

While the nearly equally weighted density of the net flux graphs reveals the parallel nature of unfolding, certain pathways experience noticeably higher traffic at the highest load level, 100 pN. As evidenced by the absence of thick, directed connections between anchors, both the 0 and 10 pN graphs are devoid of obvious and specific unfolding pathways. Conversely, the 100 pN graph indicates at least one major unfolding

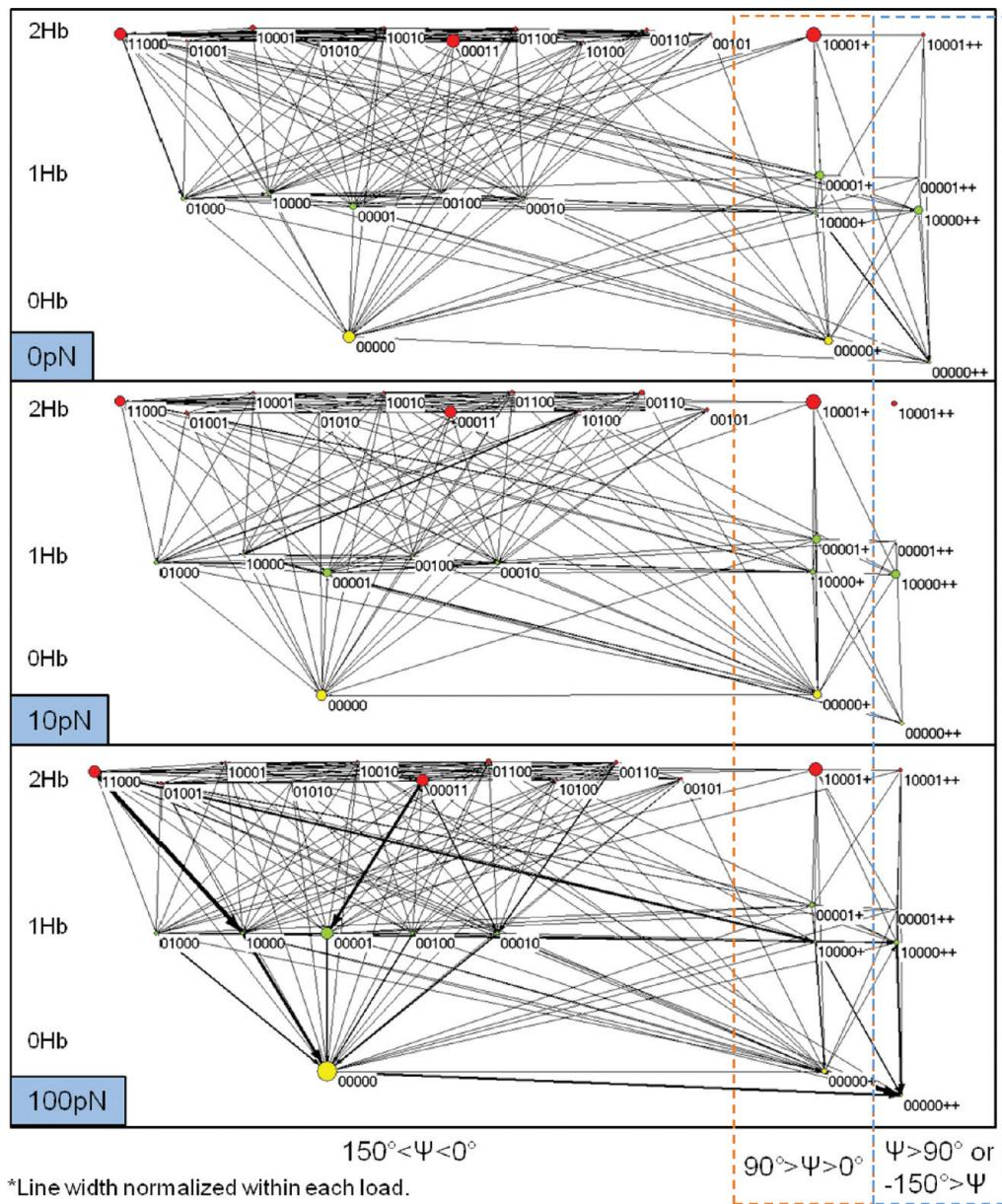


Figure 18. Flux networks through fine anchors as a function of load. Graphs of net flux between the different fine anchors of the α -helical system, in which fine anchors are defined by all the possible permutations of the five rotationally constraining hydrogen bonds and ψ -angle magnitude, as a function of applied axial load. Only net fluxes between fine anchors with two or fewer existing bonds are shown; fluxes involving patterns with greater numbers of bonds are omitted for clarity. Patterns, with dihedral angle ψ distributions demonstrating positive angles, are further divided into two groups to distinguish the severity of rotation. The patterns with a “+” and “++” designation have dihedral angle ψ between 0 and 90° and greater than 90° , respectively. The size of the anchor circle corresponds to the total flux into a fine anchor. The width of connections between fine anchors corresponds to the magnitude of net flux, while the arrows indicate the direction of that net flux. Net flux between two states i and j is calculated as $\Delta q_{ij} = q_{ij} - q_{ji}$; arrows represent the direction of positive flux such that a negative value for Δq_{ij} is represented as an arrow from j to i . Net fluxes are calculated using all simulation time within each load level. (Net fluxes to the π - and 3_{10} -helical anchors are omitted for clarity.)

pathway sequentially connecting patterns 11000, 10000, 00000, 00000+, and 00000++; indeed, this pathway is the first calculated MaxFlux channel (see section III.2.3 and Figure 19). This pathway (preceded by failure of the C2-, C-, and C hydrogen bonds) involves first breaking the C3- hydrogen bond, then the C4- hydrogen bond, then a rotation of the dihedral angle ψ from negative to positive and eventually past 90° . While there are many additional pathways with less net flux, this particular reaction pathway demonstrates the sequential nature of helix unfolding, in which individual hydrogen bonds fail prior to dihedral angle ψ rotation. Critically,

this particular pathway is associated with helix unfolding from the C-terminal and in isolated helices loaded through their backbone via connections to other secondary structures. Interestingly, the equivalent N-terminal unfolding pathway (00011, 00001, 00000, 00000+, 00000++) also appears with an elevated net flux relative to other pathways, the principle difference being that, unlike the C-terminal pathway, there is a low net flux between the one-hydrogen bond and zero-hydrogen bond states.

In addition to identifiable, sequential patterns of unfolding, the connectivity of the positive dihedral angle ψ , fine anchors

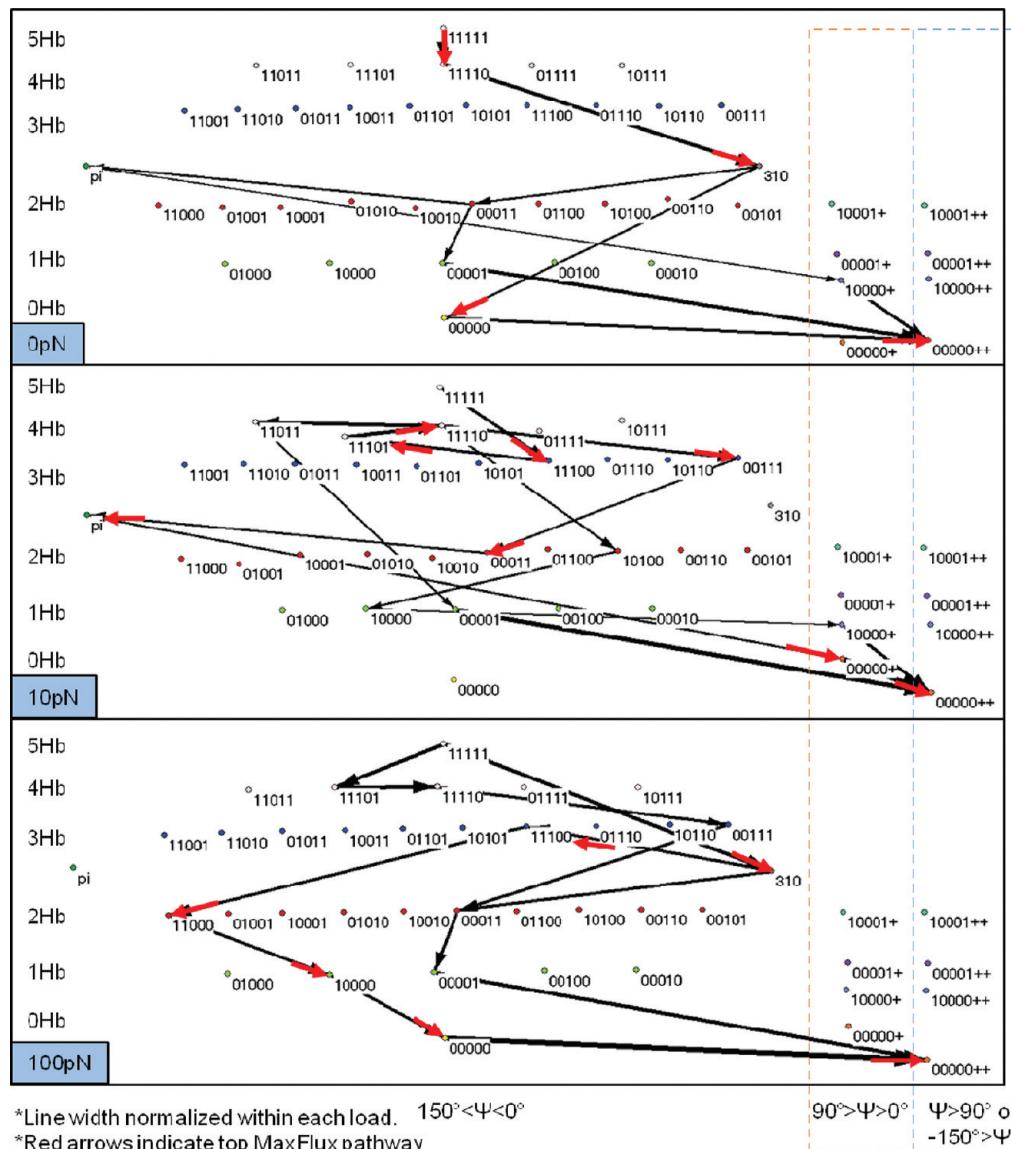


Figure 19. MaxFlux pathways through fine anchors. MaxFlux pathways through the fine anchor system from the native α -helical state with five existing hydrogen bonds (11111) to the initial, unfolded state with a ψ dihedral angle greater than 90° (00000++). Three MaxFlux pathways with greatest flux are shown for each load level: the red arrows indicate the top MaxFlux pathway. The patterns with a “+” and “++” designation have ψ angles between 0 and 90° and greater than 90° , respectively. The width of connections between fine anchors corresponds to the magnitude of net flux, while the arrows indicate the direction of that net flux. Net flux between two states i and j is calculated as $\Delta q_{ij} = q_{ij} - q_{ji}$ except for fluxes involving the “product” state (00000++) in which the net flux is set as the incoming flux consistent with a study of the initiation—and not propagation—of unfolding.

illustrates an additional role for load magnitude in determining the unfolding pathway. In particular, the 0 and 100 pN load levels demonstrate an elevated ability to rotate the dihedral angle ψ to positive values prior to bond breaking; this is in marked contrast with the paucity of connections to the “+” and “++” anchors on the 10 pN graphs. Additionally, the 100 pN load shows higher net fluxes to both these “+” and “++” anchors with hydrogen bonds, as well as between the “+” and “++” anchors, since hydrogen bonds are broken while the dihedral angle ψ is rotated. Thus, the unfolding pathway to a no-hydrogen bond, $\psi > 90^\circ$ state (00000++) at higher loads (100 pN) can proceed two ways: either via breaking of all hydrogen bonds, followed by a dihedral angle ψ rotation or a dihedral angle ψ rotation, followed by failure of the last existing hydrogen bonds.

III.2.3. Results: Fine Anchors: MaxFlux Pathways. Calculated MaxFlux pathways indicate the sequential nature of the initiation of helix unfolding as a process in which individual—or sometimes pairs of—bonds break in multiple steps prior to a ψ angle rotation (Figure 19). Additionally, while α -helical bonds may break during the unfolding process, there are also occurrences in which bond reformation occurs. One such pathway is the top MaxFlux pathway for 10 pN in which the first four states are 11111, 11110, 11101, and 11111 (numbering convention is from C4- to C). While the first transition involves breaking the “C” and “C” hydrogen bonds, the second transition is the creation of the “C” bond. What is clear from each of the trajectories, however, is that helix unfolding does not proceed with one or two large transitions.

Differences between MaxFlux reaction pathways through the fine anchor system calculated for the three load levels reveal clear load magnitude dependence (Figure 21). The unloaded, 0 pN MaxFlux pathways show the formation of the 3_{10} -helical state as an intermediate prior to unfolding in each of the three calculated pathways. Interestingly, one of the 0 pN MaxFlux pathways transitions from the 3_{10} -helical (3_{10}) to α -helical (00011) to π -helical (pi), prior to an initial rotation of the dihedral angle ψ . This pathway volatility in which hydrogen bonds of different helical character are formed—and broken—within a single pathway prior to unfolding underscores the variability of the unloaded, unfolding mechanism.

Additionally, the 0 pN MaxFlux pathways demonstrate that rotation of the dihedral angle ψ can occur prior to breaking of each of the five α -helical, hydrogen bonds evaluated in the fine anchor system. This property is seen in the pathways visiting the 10000+ and 00001 states as the penultimate state prior to unfolding. Interestingly, the third MaxFlux pathway follows a more predictable route in which all five α -helical hydrogen bonds are broken prior to full dihedral angle ψ rotation. Thus, unfolding can occur via multiple different mechanisms, even within the same load (or unloaded) conditions.

The “transition state” of the MaxFlux pathway is defined as the pathway edge having the smallest net flux. Eliminating this edge from consideration allows the calculation of successive pathways of lower net flux across each respective “transition state”. For each of the three MaxFlux pathways at 0 pN, this “transition state” was the second-to-last transition. We note that, due to the focus on the initiation of unfolding, there was no consideration of a “return” flux from the final state and thus the total flux into the final state is equal to the net flux. For the 0 pN scenario, these pathway edges were: (i) 3_{10} to 00000 (net flux of 34), (ii) 00011 to 00001 (net flux of 27), and (iii) π to 10000+ (net flux of 16). While these net fluxes were on the order of tens of transitions, the total fluxes for these states were on the order of hundreds of transitions, indicating that the data represent a well-sampled system.

Consistent with previous observations, the 10 pN MaxFlux pathways show a proclivity to visit the π -helical (pi) state as an intermediate to unfolding in one of the pathways. However, while previous results with the coarse anchor system indicate that π -helical states in the 10 pN load scenario have a relatively high AA-state occupancy, only one—albeit the highest flux—of the top three MaxFlux pathways visit this state. Thus, these π -helical visits are not required for unfolding but rather represent a possible mechanism.

Lastly, the 100 pN MaxFlux pathways visit the 3_{10} -helical state as an intermediate in the two highest flux pathways. Interestingly, these 3_{10} -helical visits do not lead directly to an unfolded state, but rather each of these two pathways return to a purely α -helical state prior to unfolding. Thus, the final transitions to unfolding for the MaxFlux pathways of the 100 pN simulations do not require transitions through non- α -helical intermediate states. Additionally, the three MaxFlux pathways of the 100 pN simulations pass through α -helical states representing either the N- (11000) or C-terminal (00011) structures. The highest flux pathway is through the N-terminal state and has a transition state net flux of 90 transitions between the 10000 and 00000 states. The second and third highest flux pathways go through the C-terminal state and have a transition state net flux of 84 and 75, respectively, both pathways passing between the 00011 and 00001 states.

The magnitude of the transition edge fluxes for each load level indicates that, while the 100 pN load dramatically accelerates unfolding relative to 0 pN, the 10 pN transition fluxes are roughly equal to the 0 pN values (Figure 20). We note that these transition edge fluxes represent kinetics through specific, individual unfolding pathways. This single-pathway representation of kinetics is in contrast to the concept of the MFPT, which incorporates not only parallel mechanisms but also behaviors such as hydrogen bond reformation that can serve to slow overall processes. While the transition edge fluxes for 0 and 10 pN are roughly equal, the calculated MFPT values indicate that the 10 pN unfolding is a slower process than the 0 pN by roughly 20%. This difference in the effect of load on kinetics between the two metrics demonstrates that the 10 pN simulations are slowed not by the rates of unfolding transitions but rather by a tendency to be diverted from the unfolding pathway, as supported by the differences in exit probabilities seen in Figure 10.

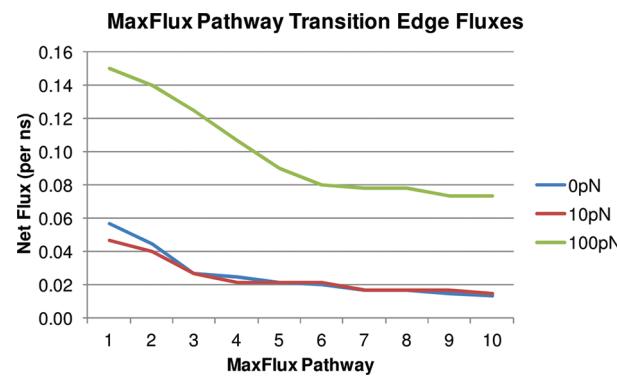


Figure 20. Net fluxes of transition edges for the top 10 MaxFlux pathways. Net fluxes for the transition edges (defined as the pathway edge with smallest net flux) of the 10 largest MaxFlux pathways. Net fluxes are normalized by total simulation time and thus presented on a per nanosecond basis. Successive, multiple pathways for each load are calculated by eliminating the transition edge from the previous pathway. All simulation time for each load is used in the calculation of net fluxes. Note the relatively slow decay of the flux carried by the different MaxFlux pathways. This result underlines the large number of channels contributing to the total flux.

Net fluxes through the transition edge of the 10 largest MaxFlux pathways for each load level indicate that the dominant pathway has a 2× or greater transition flux than all but the five highest fluxes (Figure 20). The drop-off in transition edge flux with successive pathways provides an indication of the relative dominance of the top MaxFlux channels. However, despite the presence of these dominant pathways, the plateau in transition edge fluxes demonstrates the “parallelness” of the unfolding mechanism and further underscores the value of a theoretical approach, such as milestoneing, that calculates kinetics over multiply connected, parallel pathways.

III.2.4. Results: Fine Anchors: Hydrogen Bond Energy Landscapes. The energy landscape for the angle and distance of the two hydrogen bonds forming the 11000 fine anchor are calculated to better understand a specific unfolding pathway through the fine anchor system (identified in the 100 pN fine anchor net flux graph) and the role of load in accelerating this pathway (Figure 22). Net flux calculations identified the 11000, 10000, 00000, 00000+, and 00000++ fine anchor pathway

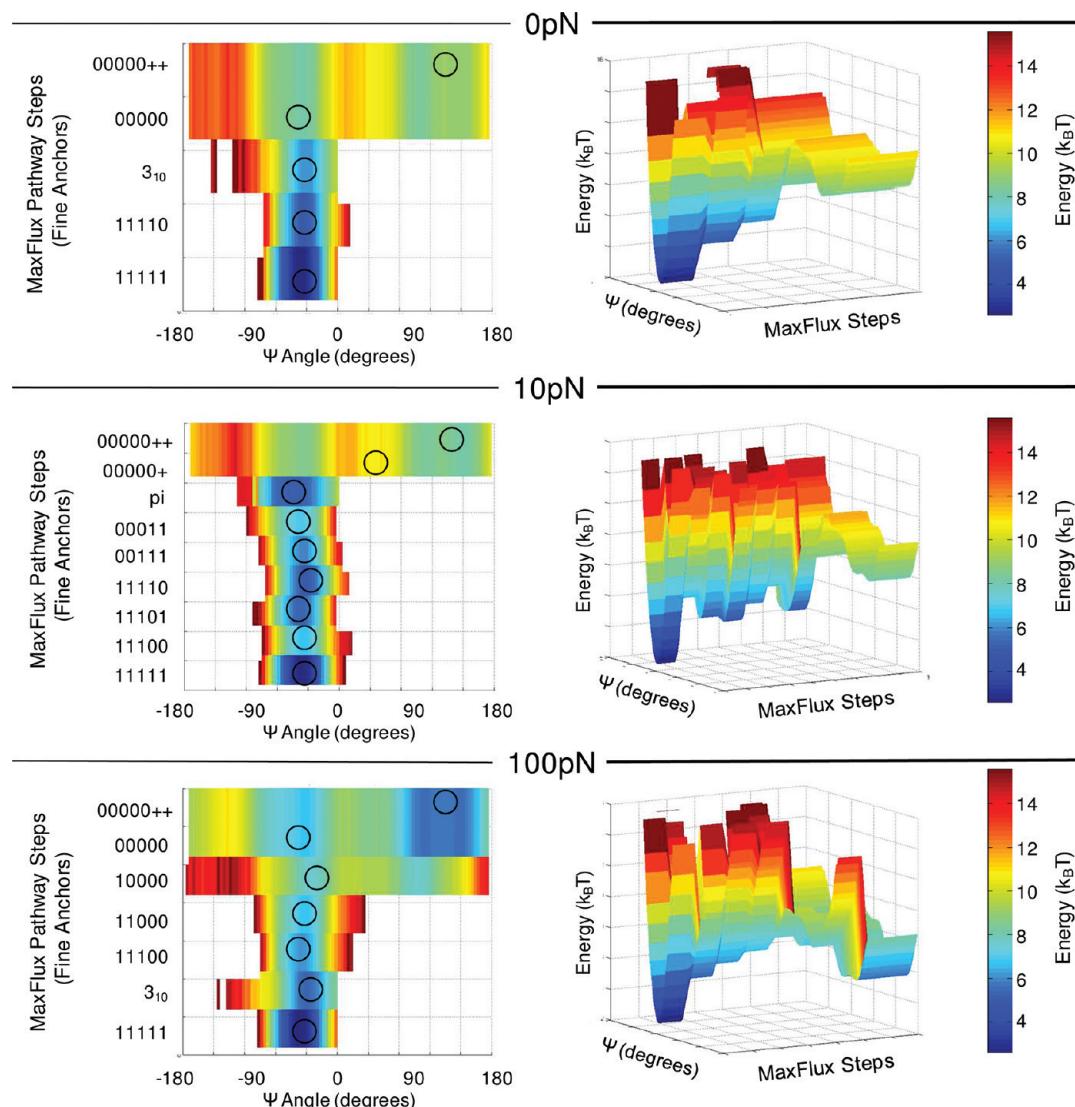


Figure 21. MaxFlux energy landscapes and schematic trajectory. (left) ψ angle energy landscape of each pattern in a vertical progression defining the top MaxFlux pathway for each load level (Figure 19). Dots roughly define the average position of the fine anchor of each step within the pattern landscape; note that some fine anchors (such as 00000 and 00000++) share a pattern (00000) and thus the energy landscape may be repeated. The dihedral angle ψ distribution for each pattern along the pathway is calculated across all 90 interior residues of the helix and normalized with respect to total simulation time for that load. The PMF is calculated via Boltzmann inversion of the distribution for each fine anchor: $E_{\text{HBond Pattern}}(\psi) = -\ln[p_{\text{HBond Pattern}}(\psi)]$. Simulations were run at 310 K, energy is in units of $k_B T$. (right) Alternative perspective of pathway energy landscapes (which are shown on the left), illustrating the depth of the energy wells of each pattern.

(corresponding to a sequential failure of the C3- then C2-hydrogen bonds prior to dihedral angle ψ rotation) as a major coordinate of unfolding at the highest load level. Notably, this pathway was heavily favored over the 11000, 01000, 00000, 00000+, and 00000++ pathway in which the C4- hydrogen bond breaks before the C3- hydrogen bond.

Comparisons of the two-dimensional energy landscapes of the C4- and C3- hydrogen bonds in the 11000 fine anchor reveal a shifting of the population density toward longer distances in the C3- hydrogen bond. Given the classical limit of 0.35 nm for hydrogen bond distance, this shift of the C3- landscape confirms the observation that this bond is more likely than not to break on the way to unfolding. This explanation for the observed pathway specificity suggests a role for protein structure in determining unfolding behavior. Specifically, the hydrogen bond topology—and not simply the quantity of bonds—constraining a system's rotational flexibility (here the

dihedral angle ψ) can be pivotal in determining the pathway of unfolding.

The load dependence of hydrogen bond distance supports the well-established, Bell-model-like notion that applied loads “tilt” underlying energy landscapes in the direction of load application, which, in the current pattern, is roughly in the direction of the hydrogen-bond distance. In particular, the hydrogen bond PMFs illustrate that the 10 pN load has a negligible effect on the hydrogen bond characteristics as compared to the unloaded (0 pN) case. This is manifested in the net flux unfolding pathways in which the net fluxes between the 11000 and 10000 anchors are roughly equal for the 0 and 10 pN loading cases. In marked contrast, the higher load magnitude produces a noticeable shift in hydrogen bond distribution, moving both the energy minima and decreasing the energy barrier to unfolding by roughly 0.5 $k_B T$ relative to that of the lower loads. This effect is noticeably present in the

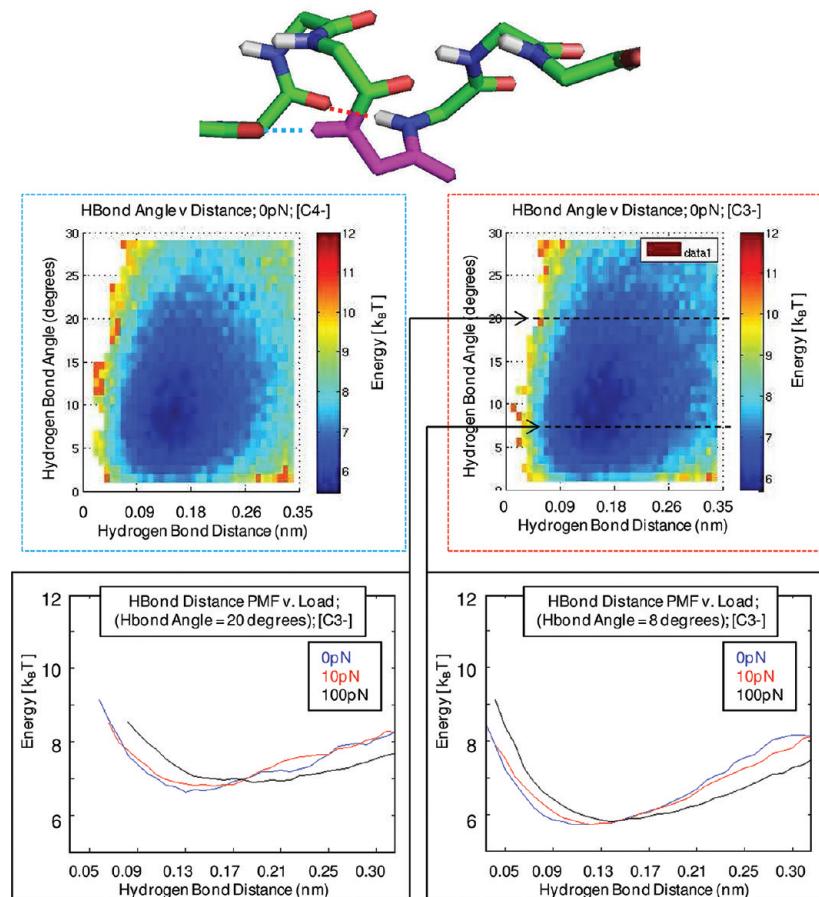


Figure 22. Fine anchor, hydrogen-bond energy landscapes. (top) The 11000 fine anchor showing the existing C4- (blue) and C3- (red) hydrogen bonds. (middle) Hydrogen-bond angle versus distance energy landscape for the bonds of the 11000 fine anchor (C4- and C3-hydrogen bonds). Two-dimensional energy landscapes are generated via Boltzmann inversion of a two-dimensional probability density for each hydrogen bond (where d is the hydrogen bond distance and θ is the hydrogen bond angle): $E_{\text{HBond}}(d, \theta) = -\ln[p_{\text{HBond}}(d, \theta)]$. Underlying probability density is normalized to the total time spent in the 11000 anchor during all 0 pN simulations, regardless of dihedral angle ψ . Simulations were run at 310 K, energy is in units of $k_B T$. (bottom) One-dimensional potentials of mean force (PMF) along the hydrogen bond distance coordinate for the C3-bond at two sample angles (left, 20°; right, 8°) in order to demonstrate the effect of load on shifting the local energy minima and the tilting of the energy landscape. PMFs are slices of the two-dimensional energy landscape. Probability distribution calculated over all instances of the 11000 anchor during the full simulation time within each load.

reaction pathway graphs in which the net flux between the 11000 and 10000 anchors is markedly higher for 100 pN than it is for either of the lower load levels. Thus, the hydrogen bond energy landscapes illustrate a role for large loads (100 pN) in affecting energy landscapes—and therefore stability—of individual hydrogen bonds.

III.2.5. Results: Fine Anchors: Mean First Passage Time of Unfolding Initiation. The mean first passage time (MFPT) for the unfolding process through the fine anchor system illustrates both the flexibility of milestone choices as well as the role of load in the unfolding process (Figure 23). Calculated MFPTs for full unfolding for the fine anchor set are 4.32 ns (0 pN), 5.18 ns (10 pN), and 2.45 ns (100 pN). Presented data is the average MFPTs initiating at the five milestones between fine anchors with 5- and 4-hydrogen bonds; the standard deviations of MFPT for these five different initiation milestones is <0.1% for each load level. The (extremely) close comparison between the MFPT calculated using the coarse and fine anchor sets demonstrates that milestone calculations can be performed with a variety of anchor definitions yet still yield consistent estimates for process kinetics: the differences in unfolding times between the two anchor sets is less than 3% for each load level (Table 9).

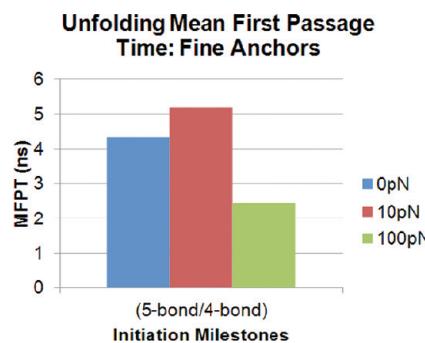


Figure 23. Mean first passage time from 5/4 hydrogen bond milestones to $90^\circ > \psi > 0^\circ : \psi > 90^\circ$ milestone as a function of load. Mean first passage time (MFPT) of unfolding from the initial milestones formed between pattern 32 (having 5 hydrogen bonds) and patterns 27, 28, 29, 30, and 31 (having 4 hydrogen bonds) to the final milestone between anchors 1+ and 1++. The data presented is the average of MFPTs from the five milestones between the 5- and 4-hydrogen bond fine anchors. Quantities used in the calculation of MFPT are based on all simulation data within each load level.

It is important to note that the fine anchor set included anchors incorporating only the permutations of the specific

Table 9. Mean First Passage Times for Coarse and Fine Anchors^a

load level (pN)	mean first passage time (ns)		
	coarse anchors (n = 1)	fine anchors (n = 5)	percent difference
0	4.3	4.3	0%
10	5.4	5.2	4%
100	2.6	2.5	7%

^aCalculated mean first passage times from the fully α -helical state to the unfolded $\psi>90^\circ$ state for each of the two levels of coarse-graining (coarse and fine anchors) as well as the percent difference between the two values for each load level. The initiation milestone of the coarse anchors was $\alpha_3:\alpha_2$, and the termination milestone was $90^\circ>\psi>0^\circ:\psi>90^\circ$. Given the higher resolution of the fine anchors, there are five initiation milestones corresponding to the fully α -helical state. These milestones are formed by the interface between patterns 27, 28, 29, 30, and 31 (the four-hydrogen bond states) and pattern 32 (the five-hydrogen bond state). The average of the MFPT calculated for each of these five initiation milestones is presented; the standard deviation of these five MFPTs was less than 1 ps.

α -helical hydrogen bonds surrounding a residue, and thus did not explicitly include π - or 3_{10} -helical bond permutations. In lieu of explicitly including all of these non- α -helical patterns in the fine anchor set, two broad anchors incorporating all bond permutations within these coarse secondary structures are defined: one for any residue with a π -helical bond and another for any residue with a 3_{10} -helical bond. Therefore, the noted tendency of the 10 pN load level to create π -helical structures is implicitly included via these additional anchors. Indeed, the fine anchor set MFPT shows the same “kinetic turnover” as the coarse anchor set in which the 10 pN unfolding process is slower than either the 0 or 100 pN.

IV. DISCUSSION

IV.1. Unfolding Initiation Summary. Detailed analysis of the CFMD trajectories of the long, heterogeneous-sequence, α -helix—stretched by constant (0, 10, 100 pN) axial force—yields some interesting, perhaps even surprising, results regarding the early events of unfolding and how they are affected, if at all, by applied load level and amino-acid type and location.

Initiation of unfolding must follow a sequence of hydrogen bond dissociation or flips to other types of hydrogen bonds. Unfolding requires that “primed unfolding vulnerabilities” (states with zero or at most one intact hydrogen bond spanning the residue) meet “chance structural disturbances” (transient fluctuations/rotations in the adjacent residues sufficiently large to positively rotate the residue’s dihedral angle). That said, “primed unfolding vulnerabilities” far outweigh “chance structural disturbances” due to the 2 orders of magnitude difference in the respective time scales for hydrogen bond loss/formation and backbone structural fluctuations.

Equally interesting, the CFMD results reveal that the early events in unfolding are driven not by hydrogen bond breakage but instead by hydrogen-bond reformation, often to “non-native” binding partners, into non- α -helical, intermediate structures having populations and lifetimes that are (in some cases, strongly) influenced by the applied load level: 0 pN, both 3_{10} and π helices; 10 pN, primarily persistent π helices; 100 pN, most often transient 3_{10} helices. Moreover, these intermediate structures and their hybrids form a highly connected, frequently visited, state network, resulting in a highly diffusive unfolding initiation process.

Milestoning analysis on flux networks demonstrates that hydrogen-bond reformation to “non-native” binding partners affect both the load-influenced unfolding pathways, as well as their corresponding mean first passage times. Remarkably, a modest stretching load (10 pN) actually forestalls the initiation of unfolding—relative to the 0 and 100 pN loads—due to hydrogen-bond reformation into relatively long-lived, π helices. The unfolding pathway and MFPT predictions are demonstrably robust in that they are invariant to anchor choice, amino acid type, and location along the helix, as well as the specific hydrogen bond definition.

IV.2. Unfolding Kinetics. Rate calculations performed via milestoning reveal a mean first passage time (MFPT) on the order of 2–5 ns for the initiation of unfolding, depending on load magnitude (Table 9). Our MFPTs compare favorably to published MFPTs of the elementary step of unfolding from a different helical system (21-residue blocked alanine peptide) acquired via milestoning and experiment,^{1,61} which are on the order of 0.5–9 ns.^{1,9} These time scales for unfolding are also comparable to computationally derived helical pentapeptide folding times of 0.2–7.2 ns.^{1,12} Interestingly, these kinetic results are from other, much smaller helical systems (5 or 21 residues) than the 2fxmA helix studied here (126 total residues).

We note that, to our knowledge, this is the first time that the fundamental mechanism—including kinetics—of the initiation of helical unfolding of a long, heterogeneous helix under external loads has been studied at the resolution of individual hydrogen bonds. Interestingly, despite the size difference in the studied systems, i.e., 5, 21, and 126 residues, the marked similarity in results, bridging very different helical systems, bolsters our confidence that the reported mechanism is fundamental to the process of helical unfolding. This is in spite of the fact that residues within these smaller helical systems are subject to considerable end-effects, due to the capacity of the unconstrained peptides to explore far greater conformational space than is available for residues within the interior of a much longer helix.

While the current investigation focuses only on the initiation of helix unfolding, it is possible that the effects of the additional conformational-space constraints are felt most prominently in the propagation of an initial unfolding event in such a way that they are responsible for keeping the initial “crack” vibrating within a narrow conformational space, allowing the reformation of bonds and subsequent “healing” of the crack. Indeed, the similarity between unfolding initiation times of 0 pN (4.3 ns) and 100 pN (2.5 ns) is in marked contrast to (i) qualitative observations from the AA-state trajectories (Figure 6), (ii) populations of unfolded AA-states (0 pN, <0.5%; 100 pN, >7%; Figure 7), and (iii) final transition probabilities ($90^\circ>\psi>0^\circ \rightarrow \psi>90^\circ$; 0 pN, 22%; 100 pN, 76%; Figure 10). Taken together, these suggest that the propagation of unfolding requires a step beyond simple initiation. The relative paucity of propagation under 0 and 10 pN load may speak to the dominance of these constraining effects in the absence of relatively large (100 pN) applied loads. This reliance on propagation for full unfolding has potential implications in the development of α -helical “strength” models.

Our estimates of the MFPT of the initiation of unfolding suggest that, in order to experimentally characterize the mechanical properties of an α -helix (with all its hydrogen bonds nominally intact), one needs to apply a quasi-steady load (\leq 100 pN) and perform observations on time scales significantly shorter than \sim 2–5 ns. For experiments performed

outside these parameters, our results suggest a strong likelihood that the observations are dominated by the rotation of the ψ dihedral angle and thus the propagation of unfolding. While not the focus of this paper, this study has yielded data to suggest that the behavior of hydrogen bonds during unfolding propagation is different than that prior to (and during) unfolding initiation.

IV.3. Degree of Coarse Graining. While calculated MFPTs agree with computational^{1,9,12} and experimental^{1,12} results of other helical systems, the calculated MFPTs of unfolding initiation, further, are *independent* of the choice of coarse grain anchor system (Table 9). This agreement between calculated MFPTs provides evidence of the accuracy of the milestoning approach in capturing simulation data. Moreover, the current analysis approach, in which coarse graining levels are applied *a posteriori*, enables elucidation of a number of different scale-dependent observations—notably the full enumeration of possible AA-states. Using the full set of viable AA-states that define the anchor system, analysis of the simulation data determined properties of the AA-states (e.g., Figures 7, 8, 14, and 15) as well as identified—in a tractable manner—information about specific transitions between AA-states (Figures 9–11), particularly for non- α -helical structures. On the other hand, the fine anchors identify individual hydrogen bonds constraining the ψ dihedral angle (Figure 17), as well as pathways through coarse anchors (Figures 18–20) that together can be used to explore the effects of applied loads on the behavior of these individual bonds (Figure 22). Thus, while both *coarse* and *fine* anchor sets provide similar observations of overall behavior (i.e., MFPT), the use of milestoning *a posteriori* permits targeted analysis of specific phenomena and scale-dependent observations.

In order to make these scale-dependent observations, the milestoning analysis methodology applied here does not evaluate the kinetics through a set of predefined structures along a reaction channel. Prior implementations of both milestoning and similar kinetic rate-calculation methodologies have relied on finding a reaction pathway between product and reactant states, identifying anchors along this pathway via geometric clustering using a single scalar, such as fraction of native contacts, and finally running a large quantity of simulations between each of these milestones.^{9,12,37} In contrast, this work prescribes anchors *a posteriori*—and therefore milestones—as the possible states of the system in terms of hydrogen bond states and backbone dihedral angle. This is viable within the residue-level, helical system considered here due to the relatively easily identifiable anchors available and our restricted study of only early events of unfolding. Furthermore, the use of CFMD limits complications associated with the rate dependence of the mechanical response of proteins and permits a quasi-steady-state picture of helix unfolding initiation at different load magnitudes.

IV.4. Mechanism Time Scales. The capacity of the current analysis tools to identify and track individual hydrogen bond activity enables the revelation of two distinct time scales of α -helical unfolding initiation: (i) a fast, hydrogen bond transition time scale (5–10 ps; Figure 11) and (ii) a considerably slower MFPT associated with full ψ dihedral rotation (2–5 ns). Fast time scales of hydrogen-bond formation and breaking have been previously noted in simulations of pentapeptides^{1,62} with values consistent with those reported here (≤ 15 ps). Thus, our observations match simulations at both—hydrogen bond breakage/reformation and full ψ dihedral rotation—time

scales. Together, these strongly suggest that hydrogen bond failure is not a good measure for helix unfolding, a role better suited to ψ dihedral rotation. Indeed, the difference in reported time scales between hydrogen bond failure and the initiation of unfolding emphasizes the role of hydrogen bond reformation. AA-state exit probabilities (Figure 10) illustrate the strong preference of intermediate states with free carbonyl/amide groups to reform bonds rather than break existing bonds, likely due to the fact these free hydrogen bond donors/acceptors remain in the bonding vicinity of possible partners until ψ dihedral rotation. Therefore, it is likely that this difference in time scales plays a key role in any robustness of the unfolding mechanism conferred from bond reformation.

IV.5. Mechanism Load Dependence. Taken together, the use of CFMD and the ability to use a natural basis for anchor definitions permit direct observation, for the first time, of a load dependence of unfolding mechanism in which hydrogen bond reformation plays a critical role. While the largest load level studied (100 pN) has an unfolding process that rarely visits diversionary states such as the π -helical states, the low load (10 pN) has a demonstrated proclivity to form persistent, π -helical structures (Figure 6). Exit probabilities indicate that these diversionary states form as a result of “incorrect” bond reformation, such that systems with available amide and/or carbonyl groups have a greater tendency to form π -helical bonds under 10 pN than 100 pN (Figure 10). In contrast, if the larger load level visits non- α -helical states, they are overwhelmingly the 3_{10} states that are natural intermediates to unfolding. Thus, while differences in mechanism time scales indicate that all load levels have a tendency to reform bonds, the specific type of helical bond formed (π , α , or 3_{10}) is a key determinant of the specific unfolding mechanism. The ψ -angle energy landscapes for each state demonstrate that the π -helical states, which are more likely with 10 pN, have a more negative ψ angle distribution than do the 3_{10} states (Figure 8). Thus, while both 10 and 100 pN simulations have a tendency to reform bonds and visit non- α -helical states, the 10 pN visits a conformation that is not an unfolding intermediate, in fact that actually opposes unfolding due to an external pulling force.

Finally, the fine-anchor system’s calculated MaxFlux pathways (Figures 19 and 21), as well as the magnitude of net flux of the transition edge of the top 10 pathways (Figure 20), indicate that the basic process of unfolding is largely the same for each load level: sequential failure of α -helical bonds, followed by a final ψ -angle rotation when in a primed unfolding vulnerable state. While specific details of these MaxFlux pathways may vary, similarities exist between the pathways and the magnitude of fluxes; e.g., the 00011 fine anchor appears to be a necessary step along a pathway of each load level. Despite these similarities in reaching primed vulnerabilities, the final step of ψ rotation is notably load dependent such that the probability of the final transition from the *none* AA-state to the $90^\circ > \psi > 0^\circ$ state is $\sim 3\times$ more likely at 100 pN (8%) than at 10 pN (3%) or 0 pN (2%) (Figure 10).

Current results indicate a strong proclivity of certain load levels to form π -helical structures, a conformation that, by at least some accounts,^{63,64} is rare and, in some sense, counterintuitive, as they are further away from unfolding than the α -helical states. However, both structural and experimental studies have observed π helices.^{4,5,65,66} In particular, an analysis of the PDB identified π -bulges—in which one or more π -helical bonds are formed in an otherwise α -helical secondary structure—occurring in 4.7% of studied structures.⁴ Addition-

ally, a recent UV resonance Raman spectroscopy study of helical tendencies in a 21-residue alanine rich peptide showed an ~3% fraction of π -bulges at the temperature of the current simulations (310 K).⁵ The unloaded simulations in the current study show a π -helical fraction of ~6% at the same temperature. Therefore, while the formation of π -helices in the current work may at first seem counterintuitive, other researchers using varied techniques have observed π -helices.^{4,5,65–68} Here, these π -helices appear to form as a result of bond failure and subsequent “non-native” reformation—notably not as unfolding intermediates—thus their existence in the current study may be a property associated with loading. The phenomenon that we observe of non-monotonic behavior with load is consistent with the catch-bond behavior observed in experimental systems. (VENDOR: ADD NEW REF 69)

IV.6. Potential Effect of Force-Field Dependence. We concede that these results may be subject to force-field differences inherent in the use of molecular dynamics.⁷⁰ Numerous studies have attempted to assess the helical tendencies of various force fields with small peptides, typically pentapeptides.^{1,8} These studies have shown that the helical character of the studied peptides is indeed dependent on force field. Additionally, theoretical analysis, via the well-known Lifson–Roig model of helix nucleation and formation,⁶ illustrate a strong dependence of helicity on protein length such that longer sequences (like a 126-residue polypeptide) tend to more reliably form stable helices. Finally, while the current simulations reflect the force field used, there is no generally agreed upon force field for modeling helices, though some comparative simulations using different force fields were conducted.^{1,8,71} That said, these results present a common basis for comparison and are self-consistent within the accuracy of the G53a6 force field.⁵⁰

V. CONCLUSIONS

We considered the atomically detailed mechanisms of early events in helix unfolding. Straightforward molecular dynamics simulations were used to sample these events in a long, heterogeneous, solvated helix with a natural sequence under applied external load. While the MD simulations were straightforward, the analysis of the data combined the residue's hydrogen bond status and dihedral angle values to provide a unique description of helical state. In order to create a simpler-to-understand, coarse-grained model, we used the theory of milestoning with two different coarsening schemes and extracted mechanisms of unfolding and their respective kinetic time scales. Despite these diverse analysis approaches and the heterogeneous nature of the helix's sequence, we found a consistent prediction of ~5 ns for the helix unfolding initiation, which implies constraints on testing parameters for pulling experiments, as well as the interpretation of their results. Moreover, our study revealed a hierarchical sequence of unfolding—including off- and on-pathway states—due to the fact that hydrogen bond dissociation and reformation is 2 orders of magnitude faster than torsional transitions. The initiation of helix unfolding is nonmonotonic under load; different pathways are (sometimes frequently) visited, and the time scale is maximal at an intermediate load. Observation of this non-monotonic behavior provides an atomic-level description for catch–bond behavior in the unfolding of a simple protein segment. The combination of molecular dynamic simulations and milestoning analysis was illustrated to be a useful tool that provides a quantitative, yet elemental, picture of complex processes in molecular biophysics.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: ron@ices.utexas.edu (R.E.); moon@austin.utexas.edu (T.J.M.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Supported in part by National Science Foundation grant CBET1133351 to S.M.K. and T.J.M. Supplemental support of a 2011–2012 Moncrief Grand Challenge Faculty award and the Temple Foundation Faculty Fellowship #1 to T.J.M. is also appreciatively acknowledged. Support of National Institutes of Health, National Institute of General Medical Sciences grant GM59796 to R.E. is gratefully acknowledged. Finally, The University of Texas at Austin's Institute for Computational Engineering and Sciences (ICES) and Texas Advanced Computing Center (TACC) are acknowledged for providing the computational infrastructure and HPC resources that have contributed to the research results reported within this paper.

■ REFERENCES

- (1) Hegefled, W. A.; Chen, S. E.; DeLeon, K. Y.; Kuczera, K.; Jas, G. *S. J. Phys. Chem. A* **2010**, *12391–12402*.
- (2) Jas, G. S.; Kuczera, K. *Biophys. J.* **2004**, *87*, *3786–3798*.
- (3) Scholtz, J. M.; Baldwin, R. L. *Annu. Rev. Biophys. Biom.* **1992**, *21*, *95–118*.
- (4) Cartailler, J.-P.; Luecke, H. *Structure* **2004**, *12*, *133–144*.
- (5) Mikhonin, A. V.; Asher, S. A. *J. Am. Chem. Soc.* **2006**, *128*, *13789–13795*.
- (6) Lifson, S.; Roig, A. *J. Chem. Phys.* **1961**, *34*, *1963–1974*.
- (7) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, *526–535*.
- (8) Best, R. B.; Buchete, N.-V.; Hummer, G. *Biophys. J.* **2008**, *95*, *L07–L09*.
- (9) Kuczera, K.; Jas, G. S.; Elber, R. *J. Phys. Chem. A* **2009**, *113*, *7461–7473*.
- (10) Ackbarow, T.; Chen, X.; Keten, S.; Buehler, M. J. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, *16410*.
- (11) Ackbarow, T.; Keten, S.; Buehler, M. J. *J. Phys.: Condens. Matter* **2009**, *21*, *035111*.
- (12) Jas, G. S.; Hegefled, W. A.; Majek, P.; Kuczera, K.; Elber, R. *J. Phys. Chem. B* [Online early access]. DOI: 10.1021/jp211645s. Published Online Feb 15, 2012.
- (13) Bai, D.; Elber, R. *J. Chem. Theory Comput.* **2006**, *2*, *484–494*.
- (14) Visscher, K.; Schnitzer, M. J.; Block, S. M. *Nature* **1999**, *400*, *184–189*.
- (15) Block, S. M.; Goldstein, L. S. B.; Schnapp, B. J. *Nature* **1990**, *348*, *348–352*.
- (16) Hirokawa, N.; Noda, Y.; Tanaka, Y.; Niwa, S. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, *682–696*.
- (17) Hirokawa, N.; Noda, Y. *Physiol. Rev.* **2008**, *88*, *1089–1118*.
- (18) Al-Khayat, H. A.; Hudson, L.; Reedy, M. K.; Irving, T. C.; Squire, J. M. *Biophys. J.* **2003**, *85*, *1063–1079*.
- (19) Burghardt, T. P.; Hu, J. Y.; Ajtai, K. *Biophys. Chem.* **2007**, *131*, *15–28*.
- (20) Cooke, R.; Franks, K. E. *J. Mol. Biol.* **1978**, *120*, *361–373*.
- (21) De La Cruz, E. M.; Wells, A. L.; Rosenfeld, S. S.; Ostap, E. M.; Sweeney, H. L. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, *13726–13731*.
- (22) Elber, R.; West, A. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, *5001–5005*.
- (23) Finer, J. T.; Simmons, R. M.; Spudich, J. A. *Nature* **1994**, *368*, *113–119*.
- (24) Kee, Y.-S.; Robinson, D. N. *Curr. Biol.* **2008**, *18*, *R860–R862*.
- (25) Isralewitz, B.; Gao, M.; Schulten, K. *Curr. Opin. Struct. Biol.* **2001**, *11*, *224–230*.

- (26) Lu, H.; Isralewitz, B.; Krammer, A.; Vogel, V.; Schulten, K. *Biophys. J.* **1998**, *75*, 662–671.
- (27) Dietz, H.; Berkemeier, F.; Bertz, M.; Rief, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12724–12728.
- (28) Dietz, H.; Rief, M. *Phys. Rev. Lett.* **2008**, *100*, 98101.
- (29) Schwaiger, I.; Schleicher, M.; Noegel, A. A.; Rief, M. *EMBO J.* **2005**, *6*, 46.
- (30) Suzuki, Y.; Dudko, O. K. *Phys. Rev. Lett.* **2010**, *104*, 048101.
- (31) Kirmizialtin, S. *J. Chem. Phys.* **2005**, *122*, 234915.
- (32) Graham, T. G. W.; Best, R. B. *J. Phys. Chem. B* **2011**, *115*, 1546–1561.
- (33) Best, R. B.; Paci, E.; Hummer, G.; Dudko, O. K. *J. Phys. Chem. B* **2008**, *112*, 5968–5976.
- (34) West, D. K.; Olmsted, P. D.; Paci, E. *J. Chem. Phys.* **2006**, *124*, 154909.
- (35) Dudko, O. K.; Hummer, G.; Szabo, A. *Phys. Rev. Lett.* **2006**, *96*, 108101.
- (36) Faradjian, A. K.; Elber, R. *J. Chem. Phys.* **2004**, *120*, 10880–10889.
- (37) Kirmizialtin, S.; Elber, R. *J. Phys. Chem. A* **2011**, *6137–6148*.
- (38) Májek, P.; Elber, R. *J. Chem. Theory Comput.* **2010**, *6*, 1805–1817.
- (39) Ron, E. *Biophys. J.* **2007**, *92*, L85–L87.
- (40) Vanden-Eijnden, E.; Venturoli, M.; Ciccotti, G.; Elber, R. *J. Chem. Phys.* **2008**, *129*, 174102–174113.
- (41) West, A. M. A.; Elber, R.; Shalloway, D. *J. Chem. Phys.* **2007**, *126*, 145104.
- (42) Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C. *J. Chem. Phys.* **2007**, *126*, 155101–155117.
- (43) Sarich, M.; Noé, F.; Schütte, C. *Multiscale Model. Simul.* **2010**, *8*, 1154–1177.
- (44) Schutte, C.; Noe, F.; Lu, J.; Sarich, M.; Vanden-Eijnden, E. *J. Chem. Phys.* **2011**, *134*, 204105–204115.
- (45) Valleau, J. In *Classical and quantum dynamics in condensed phase simulations*; Bruce, J., Berne, G. C., Coker, D. F., Eds.; World Scientific: Singapore, 1998.
- (46) Hummer, G.; Kevrekidis, I. G. *J. Chem. Phys.* **2003**, *118*, 10762–10773.
- (47) Shalloway, D.; Faradjian, A. K. *J. Chem. Phys.* **2006**, *124*, 54112–S4120.
- (48) Blankenfeldt, W.; Thomä, N. H.; Wray, J. S.; Gautel, M.; Schlichting, I. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17713–17717.
- (49) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (50) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.
- (51) Hess, B. *J. Chem. Theory Comput.* **2007**, *4*, 116–122.
- (52) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (53) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (54) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- (55) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (56) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (57) Huo, S.; Straub, J. E. *J. Chem. Phys.* **1997**, *107*, 5000–5006.
- (58) Noe, F.; Ille, F.; Smith, J. C.; Fischer, S. *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 534–544.
- (59) Czerninski, R.; Elber, R. *J. Chem. Phys.* **1990**, *92*, 5580–6502.
- (60) Soman, K. V.; Karimi, A.; Case, D. A. *Biopolymers* **1991**, *31*, 1351–1361.
- (61) Mohammed, O. F.; Jas, G. S.; Lin, M. M.; Zewail, A. H. *Angew. Chem., Int. Ed.* **2009**, *48*, 5628–5632.
- (62) Margulis, C. J.; Stern, H. A.; Berne, B. J. *J. Phys. Chem. B* **2002**, *106*, 10748–10752.
- (63) Bolin, K. A.; Millhauser, G. L. *Acc. Chem. Res.* **1999**, *32*, 1027–1033.
- (64) Rohl, C. A.; Doig, A. J. *Protein Sci.* **1996**, *5*, 1687–1696.
- (65) Weaver, T. M. *Protein Sci.* **2000**, *9*, 201–206.
- (66) Fodje, M. N.; Al-Karadaghi, S. *Protein Eng.* **2002**, *15*, 353–358.
- (67) Armen, R.; Alonso, D. O. V.; Daggett, V. *Protein Sci.* **2003**, *12*, 1145–1157.
- (68) Riek, R. P.; Graham, R. M. *J. Struct. Biol.* **2011**, *173*, 153–160.
- (69) Prezhdo, O. V.; Pereverzev, Y. V. *Accounts Chem. Res.* **2009**, *42*, 693–703.
- (70) Mackerell, A. D. *J. Comput. Chem.* **2004**, *25*, 1584–1604.
- (71) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.