Published in final edited form as:
J Phys Chem B. 2012 June 14; 116(23): 6598–6610. doi:10.1021/jp211645s.

Experiments and comprehensive simulations of the formation of a helical turn

Gouri S. Jas[a], **Wendy Hegefeld**[a], **Peter Májek**[b], **Krzysztof Kuczera**[c], and **Ron Elber**[b,d,*]

[a]Department of Chemistry, Biochemistry, and Institute of Biomedical Studies, Baylor University, Waco, TX 76706

[b]Institute of Computational Engineering and Sciences (ICES), University of Texas at Austin, Austin, TX 78712

[c]Departments of Chemistry and Molecular Biosciences, The University of Kansas, Lawrence, KS 66045

[d]Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, TX 78712

Abstract

We consider the kinetics and thermodynamics of a helical turn formation in the peptide Ac-WAAAH-NH_2. NMR measurements indicate that the peptide has significant tendency to form a structure of a helical turn, while temperature dependent CD establishes the helix fraction at different temperatures. Molecular Dynamics and Milestoning simulations agree with experimental observables and suggests an atomically detailed picture for the turn formation. Using a network representation two alternative mechanisms of folding are identified: (i) a direct cooperative mechanism from the unfolded to the folded state without intermediate formation of hydrogen bonds and (ii) an indirect mechanism with structural intermediates with two residues in a helical conformation. This picture is consistent with kinetic measurements that reveal two experimental time scales of sub nanosecond and several nanoseconds.

Keywords

folding; pentatpeptide; Molecular Dynamics; NMR; Circular Dichroism; Milestoning

I. Introduction

The mechanism in which proteins fold has been a challenge for atomically detailed simulations due to the vast conformational space available to the peptide chain and the surprising efficiency in which proteins fold in nature. The large number of conformations makes exhaustive enumeration impossible in practice. Nevertheless, proteins fold accurately and quickly on time scales as short as microseconds[1]. Therefore concrete mechanisms speeding up the process must be present in nature, and can be used for more efficient computer simulations.

The funnel picture of protein folding[2] is successful in explaining folding efficiency. It is a model in which the free energy landscape of the protein chain is systematically tilted towards to the native structure of the protein. The bias persists even when the protein chain is completely unfolded and no resemblance to the native structure can be found. How does the unfolded chain “know” where to go? In other words what are the fundamental physical

corresponding author: Ron Elber, ron@ices.utexas.edu, phone: 512-232-5415, Fax: 512-471-8694.

interactions within the chain and between the protein and solvent that are able to drive the unfolded chain in the correct direction? For folding to be efficient this drive must be active even when no fingerprints of the correct structure are found in the present conformation.

The phenomenology we have in mind is of the Go model[3]. In Go models for protein folding a bias is introduced to the energy function that rewards native geometries. Go models were successful in explaining numerous folding mechanisms and pathways[4]. Besides folding they were found useful in studies of conformational changes and of assembly formation[5]. There are two general types of Go models. The first type biases contacts, some of which can be far apart along the peptide chain. The second type biases chain conformation, such as a single or a few torsion angles. It is the second type that we consider in this paper. By virtue of the locality of the bias it is clear that we consider short protein fragments (peptides).

Local bias is at the core of simplified models for protein folding. For example in one reference [6] a correct or an incorrect “bond” is introduced between sequential amino acids. A bias toward a correct bond, (in each of the bonds), speeds up folding profoundly. Local bias was also discussed in the context of spin glass models of protein folding. Paramagnetic spins under the influence of an external magnetic field are biased (locally) to orientation preferred by their interactions with the field or towards their native state[7].

The models above, while intuitively attractive and formally promising in the sense that they solve the kinetic problem, do not always provide the concrete type of physical interactions that are responsible for these biases. These biases are enforced as a general principle. Seeking a concrete example for which a physical picture of these interactions can be extracted, we note that a well-known structural fragment with significant bias to fold is a helix. The kinetics and thermodynamics of helix folding were studied extensively experimentally[8] and theoretically[9] motivated at least partially by the above mentioned considerations.

The extensive list of investigations given above (which is not complete) raises the question of whether we really need yet another experimental and simulation study of helix folding?

There are two main contributions in the present investigation that we hope will motivate the reader to read on. The first is the examination of the short peptide WAAAH, or WH5. While most of past investigations focused on longer peptides, like twenty one [10] or thirteen [8a, 11] amino acids, the observation that peptides as short as five amino acids incline to helical structures came as a surprise, and motivated further research. This observation offers a smaller nucleation element for protein folding that can further speed up protein folding in the spirit of reference [6]. A few studies of WH5 and similar peptides have been published [9j, k, 9p, 12] so further justification of the present manuscript is needed. From an experimental perspective the paper provides NMR evidence, which was not available before, that WH5 is indeed a helical turn. CD spectroscopy supplements this information and determines the population of the helical turn. Both measurements provide new quantitative data on the formation of the helix, which are intriguing enough to be examined by detailed simulations.

The second new contribution of the present paper is the use of Milestoning, a recent simulation methodology for kinetics [13]. Straightforward molecular dynamics trajectories were already computed for WH5 and related systems [9b, 9j, 9l]. Replica Exchange simulations were used by reference [9p], and a reaction coordinate was used in reference [9k]. At the least, Milestoning is using a different set of assumptions.

Moreover, statistics from straightforward MD simulations [9b, 9j, 9m] are sufficient to address overall questions on thermodynamics (relative stability of a helix) and kinetics (overall rate

of kinetics) but they are not sufficient to address detailed mechanistic questions (e.g., the identification of individual folding channels and determination of their weights). This can be understood as follows. The problem in direct atomically detailed simulations of folding is of path multiplicity. Observing multiple pathways in a single trajectory requires a sufficiently long trajectory that is going back and forth between the folded and unfolded states and able to explore alternative pathways and their weights. For example, as described in the present manuscript, for WH5 we observe about ten significant pathways leading from later intermediates to the folded state. To sample these pathways independently, which differ significantly in their weights, will require at least one hundred trajectories, each, with a single folding event.

Formally a single reaction coordinate can be defined and used in numerous ways (e.g. the fraction of native contacts, number of hydrogen bonds, or iso-committor[14]). However, the vast hypersurfaces defined as orthogonal to the one-dimensional reaction coordinate can be exceptionally difficult to sample by Molecular Dynamics simulations. The unfolded state includes many rotational conformers of the molecule under consideration that may be classified into a single value of the reaction coordinate and are therefore difficult to sample. In practice, simulations of the unfolded state necessitate the use of multiple pathways, where the neighborhood of individual pathways can be sampled adequately. In the present manuscript we illustrate the problem of using a single reaction coordinate for folding by examining a choice that was used in the past of the radius of gyration.

Anther approach that seeks multiple paths is the Markov State Model (MSM). Indeed it was applied for a related peptide A_5 [15] and to WH5[9p]. MSM implementations are using a different set of approximations than Milestoning (such as the use of Replica Exchange for dynamics, and the assumption of Markovian state and process). The mechanistic picture from Milestoning is therefore expected to be at least complementary to MSM as we elaborate in the Discussions.

II. Experimental studies

The peptide we examine is Ac-WAAAH$^+$-NH_2 (WH5). It is blocked on both ends by uncharged termini to resemble more closely the usual peptide segments of proteins. It has three alanine residues that have strong tendency to form a helix. The histidine and tryptophan are useful for detection and measurement of kinetics, with protonated histidine corresponding to acidic experimental pH conditions.

NMR

NMR data for this peptide are summarized in Figure 1.

The ROE cross-peaks characteristic of helical peptide conformations (i.e., d_{NN} (i, i+1), $d_{\alpha N}$ (i, i+1), and $d_{\beta N}$ (i, i+1)) are readily observed in the spectra (the HN-HN and HA-HN spectral regions are shown in the figure). Further evidence of a helical peptide conformation for WAAAH is provided by the observation of the $d_{\alpha N}$ (i, i+2) cross-peaks involving W1-Hα and A3-Hα (the A2-Hα – A4-HN cross-peak could not be observed due to peak overlap with A3-Hα – A4-HN), and $d_{\alpha N}$ (i, i+3) and $d_{\alpha N}$ (i, i+4) cross-peaks involving W1-Hα. In addition, the Cα chemical shifts for residues A2–A4 are ~1 ppm larger on average with respect to the random coil shift of 52 ppm. This points to a partial helical conformation for the peptide, as a downfield shift of ~ 3 ppm would be expected for alanine residues in a fully helical conformation. Finally, the $^3J_{HNH\alpha}$ coupling constants for residues W1 – A4 are < 6 Hz. The coupling constants of this magnitude are again characteristic of partial helical peptide conformations (upon increasing the temperature to 20°C the couplings for residues

W1 – A3 increases slightly, which indicates a shift to lower fraction of helical peptide conformations at the higher temperature).

Circular Dichroism

Another experimental evidence for significant formation of an alpha helix conformation is provided by ultraviolet circular dichroism spectra (CD) as a function of the temperature. The spectra (Figure 2) were measured in 20 mM acetate buffer at pH 4.8.

The far UV CD at low temperature indicates presence of a much higher population of α-helix conformation. At high temperatures, the CD spectrum is characteristic of a random coil with some residual structure. Singular-value decomposition (SVD) of the temperature dependent CD spectra produced two main components (Fig. 2, B). Component-1 has the features of an α-helix. The component-2 corresponds to a CD spectrum of a coil conformation. Fig. 2C shows the spectrum obtained by subtracting the second component from the first one. This is also a characteristic signature spectrum of a α-helix. Far UV CD spectra were measured at several different pH values to monitor the fractional changes associated with the helical conformation in acidic, neutral, and basic environment. The peptide concentrations for these measurements were 350 μM. There were no observed changes in the measured CD spectrum upon varying peptide concentrations.

In summary, the analysis illustrated that WH5 is forming a helical turn and further suggested that the probability of forming a helix at 300 K is above 20 percent. These observations will be used to verify the simulation results. Beside the experimental evidence given in this paper we also compare our data to kinetic experiments presented in reference [16].

III. Theory and Computational Protocols

III.1 Molecular Dynamics

A microsecond long Molecular Dynamics (MD) trajectory was conducted at constant temperature of 300 K and pressure of 1 bar using a 2 fs for a time step in a cubic box size of 30 Å and 801 TIP3P water molecules[17]. The program GROMACS was used in these simulations[18], and the force field was OPLS-AA[19]. The trajectory was analyzed from equilibrium and kinetic perspectives elucidating the underlying free energy surface and the mechanisms of transitions to and from the folded state. In the present manuscript the 1 μs trajectory is also used to define anchors for the Milestoning calculations and to compare the overall thermodynamic and kinetic description in MD, Milestoning and experiment.

III. 2 Milestoning

III.2.1 Theory and algorithm—To elucidate folding mechanism and concrete kinetics of WH5 we use Milestoning[13a, 20] and more specifically, the recent Directional Milestoning[13b] (DiM). Milestoning is a theory and algorithm that partitions the phase space of the system using interfaces that we call Milestones. The theory of Milestoning has been discussed previously.

We review it here for completeness of the present manuscript, and to better connect the theory to novel algorithmic steps introduced here for the first time. Transitions between Milestones are recorded and used in a non-Markovian probabilistic model that enables the calculation of rate and thermodynamic of the system. DiM was discussed in the past[13b,c]. However, for the completeness of this paper we describe it below. There are four basic steps in Milestoning calculations: (i) determination of anchors and Milestones, (ii) sampling of configurations that are restricted to Milestones, (iii) computing short trajectories between the Milestones to estimate the transition probabilities between them, and (iv) putting it all to

together in non-Markovian modeling of the dynamics to obtain a quantitative description of the kinetics and the thermodynamics.

In the first step of Milestoning a set of anchors is determined. In general anchors are points in phase space which we approximate here as points in configuration space. Anchors provide rough coverage of conformational space. The coverage need not be complete or uniform but is expected to include the major important portions that are visited frequently or that are necessary during the process of transition. They are also expected to be sufficiently dense so that relatively short trajectories will be able to transition from one anchor domain to a domain of another anchor. In Figure 3 we illustrate a set of randomly distributed points on a two-dimensional potential surface of the Mueller potential surface that can serve as anchors.

Milestones are interfaces that separate domains that we assign to anchors. Originally, Milestoning was introduced as a way to compute rate along reaction coordinates[20]. Recently however Vanden Eijnden and Venturoli[21] put forward a clever approach to extend Milestoning to high dimension and assign Voronoi cells to different anchors. Trajectories were initiated in a cell and the number of hits at an interface was estimated. We have followed their idea with one twist. The Milestoning theory, discussed below, requires that the time to reach one Milestone from another will be sufficiently long such that a memory loss requirement will be fullfilled. Voronoi cells have their interfaces cross and the transition time from one Milestone to another can be as short as zero near or at the crossing domains, violating the requirement of the Milestoning theory. We therefore defined anchor interfaces, M_{ij}, from α to β that avoid crossing (Figure 3). The equation for the set of points Y that makes this interface is

$$M_{\alpha\beta} \equiv \left\{ Y | \forall k, d(Y, Y_\gamma) \geq d(Y, Y_\beta) = \sqrt{d(Y, Y_\alpha)^2 - \Delta_i^2} \right\} \quad (1)$$

Y is the set of coarse variables that are used to describe the state of the system, for example the set of all rotatable bonds in a polymer. The shift Δ is empirical. When it is set to zero we recover the Voronoi cell description. Here it is set to the shortest distance between any pair of anchors $\Delta_i = \min_{j \neq i} d(Y_i, Y_j)$ [13b]. In other studies [13c] it was set to the fixed value of 0.1Å, which is the minimal distance that the system needs to travel from the domain of one anchor to the next.

With the definition of the Milestones at hand, we consider the transition probability between phase space points at the interfaces, $K_{\alpha\beta}(X_\alpha, X_\beta; t)$, which we also called the Kernel. Note that we use the vector X_α to denote a full phase space vector constrained to the interface α defined by the set of coarse variables Y_α. It is the probability that a trajectory starting at phase space point X_α at interface α will make it to a phase space point X_β at Milestone β after time t (Figure 3). Since the Milestones are close in space the Kernel can be estimated using short trajectories. Let the total number of trajectories initiated at X_α be $n_\alpha(X_\alpha)$. Let the number of trajectories that were initiated at X_α and reach X_β at time t be $n_{\alpha\beta}(X_\alpha, X_\beta; t)$, then $K_{\alpha\beta}(X_\alpha, X_\beta; t) \cong n_{\alpha\beta}(X_\alpha, X_\beta; t)/n_\alpha(X_\alpha)$. Sampling configurations from the distribution of $n_\alpha(X_\alpha)$ requires careful consideration [13b,22]. This distribution includes end-points of trajectories that hit interface α for the first time. We separate the sampling into two steps. First, we sample configurations constrained or restrained to the hypersurface α according to the canonical ensemble. Hence we assume that the reaction conditions are close to equilibrium. In the second step we compute trajectories starting from the sampled points backward in time until they reach a Milestone (Figure 3.b). If the Milestone reached is the same as the starting Milestone, we remove that point from our set since it is a not a first

hitting point. If the trajectory hits another Milestone we keep that point in the sampled set, integrate the equations of motion forward in time until the trajectory hits another Milestone for the first time, and record the time and the Milestone of termination to estimate $n_{\alpha\beta}(X_\alpha, X_\beta; t)$. This procedure of backward and forward integration from interface to determine a first hitting distribution is similar to the procedure described in Partial Path Transition Interface Sampling[23].

This Kernel is general and can be estimated for different types of dynamics, for example, Newtonian or Langevin dynamics[13b]. The equation below for the reaction flux, $q_\alpha(X_\alpha, t)$ (the number of trajectories that passes through Milestone α at X_α exactly at time t) is

$$q_\alpha(X_\alpha, t) = P_\alpha(X_\alpha)\delta(t^+) + \sum_{\beta\in\alpha}\int dX_\beta \int dt' q_\beta(X_\beta, t')K_{\beta\alpha}(X_\alpha, X_\beta; t - t') \quad (2)$$

Only transitions between Milestones that can be reached without passing other Milestones along the way are considered ($\bar{\alpha}$ is the subset of Milestones that can reach directly Milestone α and are called *reachable* Milestones from α).

Equation (2) is exact. However, it is not useful numerically since we are required to compute trajectories from any phase space point at the Milestones to all other phase space points at the reachable Milestones. The essential assumption of the Milestoning theory is partial independence on initial conditions. We assume that the interfaces are sufficiently separated such that a trajectory initiate at (say) X_α does not "remember" the precise location of the initiating phase space point at Milestone α at the time of termination at X_β. Mathematically, it means

$$K_{\alpha\beta}(X_\alpha, X_\beta; t) \cong K_{\alpha\beta}(X_\beta; t) \quad (3)$$

The Milestoning approximation has been investigated extensively in the past [13a, b, 22], both theoretically and computationally. Rigorous results, as well as empirical and practical guidelines make it possible to conduct the calculations efficiently and accurately. A summary of previous observations follows

In Eq. (3) the trajectory still "remembers" the interface it came from but no longer the precise position at the Milestone. When do we expect this approximation to work? If the Kernel $K_{\alpha\beta}(X_\alpha, X_\beta; t)$ is a slowly varying function of X_α then the Milestoning approximation is expected to be sound.

A limit in which the Milestoning assumption of Equation (3) is satisfied is of memory loss or de-correlation. This is the limit that we use most often and also in the present case. Given the chaotic nature of condensed phase trajectories, the precise initial conditions at interface α are not relevant after a sufficiently long period and the initial and final phase space points de-correlate. How do we know that de-correlation indeed happen? In simple systems we can check that the distribution generated by the terminating trajectories on the Milestone is the same as the distribution created by step (ii) of our procedure [13b, 22]. Comparing distributions in higher dimension can however be complex. In practice we found [13a] that velocity relaxation time is a good indicator for the time required for de-correlation. Other tests that we routinely do include adding or subtracting anchors or Milestones [13a]. Correct calculations of rate and thermodynamics are not dependent on the number of Milestones.

Another interesting limit in which the Milestoning methodology is giving the correct mean first passage time is when the Milestones are iso-committors [22]. Iso-committors are

hypersurfaces that are sets of phase space points with equal probability to reach the product state before the reactant. They were proposed as the ultimate reaction coordinate [14]. However, determining exact iso-committor surfaces is complex if not impossible for molecular systems of moderate sizes. Therefore the iso-committors must be approximated in practice. The net result is that rather than approximating the Kernel we have to approximate the iso-committors. Approximations for iso-committors are available for pathways that are dominated by relatively narrow reaction tubes [14], or pathways based on a small number of coarse variables [24]. Past studies used *posteriori* analyses of reactive trajectories [25] sampled by transition interface sampling or optimization of iso-committors functions modeled as a function of a few coarse variables [24,26]. For the present problem of multiple channels, we will require highly complex, and non-planar iso-committor surfaces. It is therefore easier to consider multiple reaction coordinates in a network framework as discussed below instead of using a single reaction coordinate and the associate complex iso-committor surfaces.

Accepting the Milestoning assumption stated in Eq. 3 it is useful to define the following functions:

$$\begin{aligned} q_\alpha(t) &= \int q_\alpha(X_\alpha, t) dX_\alpha \\ K_{\alpha\beta}(t) &= \int K_{\alpha\beta}(X_\alpha, X_\beta; t) dX_\alpha dX_\beta \\ P_\alpha(t) &= \int P_\alpha(X_\alpha, t) dX_\alpha \end{aligned} \tag{4}$$

We now substitute Eq. (3) into Equation (2) and integrate over X_α to obtain

$$\int q_\alpha(X_\alpha, t) dX_\alpha = \int P_\alpha(X_\alpha) dX_\alpha \delta(t^+) + \sum_{\beta \in \overline{\alpha}} \int \int_0^t q_\beta(X_\beta, t') K_{\beta\alpha}(X_\alpha; t - t') dX_\alpha \, dX_\beta \, dt'$$

The last equation together with Equation (4) provides the Milestoning equation

$$q_\alpha(t) = P_\alpha(0)\delta(t^+) + \sum_{\beta \in \overline{\alpha}} \int_0^t q_\beta(t') K_{\beta\alpha}(t - t') dt' \tag{5}$$

Note that the explicit dependence on coordinates was removed; only the indices of the Milestones remain. Note also that the Kernel is now estimated as $n_{\alpha\beta}(t)/n_\alpha$ where n_α is the number of trajectories initiated at interface α from the first hitting point distribution, and $n_{\alpha\beta}(t)$ is the number of trajectories that made it to Milestone β after time t. Equation (5) for the reactive flux q_α is linear and can be solved analytically with the help of Laplace transforms [13c,27]. The reactive flux, $\mathbf{q}_{stat}$, the stationary distribution, $p_{\alpha,stat}$, and the overall mean first passage time, $\langle \boldsymbol{\tau} \rangle_{\alpha f}$ from Milestone α to the final Milestone f which is absorbing are given by

$$\begin{aligned} \mathbf{q}_{stat}(\mathbf{I} - \mathbf{K}) &= 0 \\ p_{\alpha.stat} &= (\mathbf{q}_{stat})_\alpha \cdot \langle t_\alpha \rangle \\ \langle \tau \rangle_{\alpha f} &= P_\alpha \sum_\beta [\mathbf{I} - \mathbf{K}]^{-1}_{\alpha\beta} \cdot \langle t \rangle_\beta \end{aligned} \tag{6}$$

where the notation **q** is used to denote a vector of fluxes with elements q_α. The matrix **I** is the identity matrix and **K** is a matrix with elements defined by $(\mathbf{K})_{\alpha\beta}=\int_0^\infty K_{\alpha\beta}(t)dt$. The term $\langle t\rangle_\beta$ is the lifetime of Milestone β and is given by $\langle t\rangle_\beta=\sum_\alpha\int_0^\infty t\cdot K_{\beta\alpha}(t)dt$.

We emphasize the simplicity of Equation (6). To obtain the stationary distribution and overall mean first passage time we only need to solve a linear problem. For example, the stationary flux $\mathbf{q}_{\text{stat}}$ is the eigenvector with zero eigenvalue of the matrix **I–K**. The dimensionality of this matrix is the number of Milestones in the system, which in the present case is below 10,000. The linear algebra component of the calculation is therefore trivial compared to the Molecular Dynamics simulations to estimate the Kernels. Estimating the Kernels can take months of core time while solving the linear problem takes only a few minutes on a reasonable computer.

What is the gain in using this non-straightforward calculation of kinetics and thermodynamics? Milestoning is dramatically more efficient than straightforward Molecular Dynamics. There are a number of reasons for this increase in efficiency that have been discussed and illustrated a number of times in the past and are briefly mentioned below.

Consider first a system with a significant barrier ΔU. In the overdamped limit the transition time is proportional to $\tau \propto \exp(\Delta U/k_BT)$ where k_B is the Boltzmann factor and T the absolute temperature. If we divide the barrier between M Milestones then the time for a transition between two milestones will be of order of $\tau_M \propto \exp(\Delta U/MkT)$. Since there are M Milestones the total time required to complete the transition will be of order of $M\cdot\tau_M \propto M\cdot\exp(\Delta U/MkT)$. For substantial barriers of order of (say) 50 kT, the use of 100 Milestones, which is a typical number, essentially eliminates the activated nature of the process. Compare barrier-passage times of $\exp(50) \approx 5.2\cdot10^{21}$ to $100\cdot\exp(0.5) \approx 122$.

The above argument suggests exponential speedup for systems that pass over large barriers. Interestingly a significant speed-up factor is also obtained for free diffusion. Consider two states separated by length L. The time scale to pass a distance L with free diffusion is about $\tau D \propto L^2$ where D is the diffusion constant. Dividing the distance to M segments we have for one segment $\tau_M D \propto (L/M)^2$. Collecting the efforts from all the segments we have $M\cdot\tau_M D \propto L^2/M$. Hence Milestoning calculation is faster by a factor of M which is significant with typical M in the hundreds.

The third factor is of parallelization. In Milestoning we use a very large number of short trajectories. In the present paper we run more than 200,000 trajectories. These independent short trajectories can be trivially distributed on a large number of cores, something that is hard to do on a fewer longer trajectories, providing a speed up factor proportional to the number of cores. In light of the first two factors it is also obvious that the trajectories in Milestoning are very short making them particularly attractive to distributed computing. In Figure 4 we show a distribution of times of all the Milestoning trajectories we conducted in the present study.

It is clear that most of the trajectories are indeed of picosecond time length, even though a long tail of a small number of nanosecond trajectories is also observed. The total number of trajectories is 201,432 (we run about 50 trajectory per interface) and all of the trajectories terminate. The average life-time of a terminating trajectory was 33.8 picoseconds. The accumulated time of all the runs was 11.8 microseconds of which 1 microsecond was used at the beginning to identify contributing interfaces, and 200 nanoseconds for sampling at the interfaces. The longest trajectories sampled in the set of terminating trajectories are of

lengths of a few nanoseconds. These outliers illustrate a few trajectories of lengths of the transition times that if used as single sampling trajectory can lead o significant errors. The outliers provide only one transition between two Milestones while their total length if comparable to the reaction time scale (see sections **IV** and **V**).

III.2.2 Simulations of WH5—The Milestoning simulations were conducted with the program MOIL[28] using the same force field OPLS-AA as in the Molecular Dynamic (MD) simulations mentioned earlier. Since our prime interests in the present study are kinetics and mechanisms, we have used in the Milestoning calculations the NVE ensemble. The same box size and number of particles are used as in the MD simulations. The time step was 1fs and the Particle Meshed Ewald was used with a grid of 32×32×32 to account for long-range electrostatic calculations.

The first step in the Milestoning calculations (as discussed above) is the identification of anchors. Anchors can be chosen in many ways (e.g. reaction path calculations [29], high temperature trajectories, uniform and random sampling[13b, 21]). In the present study we use clustering of the configurations saved every 1ps during the $1\mu s$ Molecular Dynamics trajectory to define anchors. The clustering of the one million structures was conducted in the ten dimensional space of the five (φ, ψ) pairs. Euclidean distance in torsion space (with periodicity) was used for the clustering based on a greedy algorithm. The ideal helical structure was assigned to the center of the first cluster. A structure is assigned to be a center of a new cluster if its distances from all other cluster centers is larger than 3 radians, or an average distance per torsion of about 0.3 radians; 153 cluster centers (and anchors) were obtained. A sample of anchor configurations is shown in Figure 5.

The clustering described above is based on geometry. Ideally, the clustering should have been based on kinetics. The Milestoning assumption is based on time scales and not on geometric proximity. We use geometry for initial clustering since it is straightforward but it is not the best measure. It is possible that a large free energy barrier separates two similar conformations, making them more time-separated than suggested by geometrical distance. Indirect and geometrically longer kinetic pathways through a number of intermediate anchors may be a more probable way to proceed in the last case. Even more problematic from a computational viewpoint is the existence of anchors that transition to others too rapidly, and violate the fundamental assumption of Milestone de-correlation. Filtering such cases as early as possible will save considerable computational resources. Therefore, after the anchors were suggested by the geometrical clustering, short trajectories were initiated from each of the anchors using velocities sampled from the Maxwell distribution at 300K. The trajectories were terminated either after 3ps of trajectory time or if they hit a Milestone. If a terminated trajectory was shorter than 100fs the corresponding anchor was removed from the set. Earlier investigations [13a] suggest that the velocity de-correlation time was of sub-picosecond times and that Milestoning calculations with termination times shorter than the velocity relaxation time are inaccurate. This procedure reduced the number of anchors to 90.

For 90 anchors the maximum number of Milestones can be 90×89=8,010 (the Directional Milestoning calculations are of course asymmetric). This number is an indicator of possible challenges for Milestoning calculation; namely, the number of Milestones can grow exponentially with the dimensionality of the coarse variables. Each of the Milestones requires sampling at the interface and at least a few tens of sampling trajectories to other Milestones. The above estimate is, however, of a worst-case scenario since not all Milestones are reachable from other Milestones. As a reminder, “reachable” Milestones are interfaces that are accessible with trajectories that do not cross another Milestone along the way.

Similarly to importance sampling in equilibrium simulations we expect that the space of reactive trajectories is limited and does not require a full exploration of all Milestone space. In the most attractive circumstances the trajectories are concentrated in the neighborhood of a one-dimensional tunnel or a reaction coordinate in which the number of Milestones is independent of the dimensionality of the system. Even in more complex examples not all Milestone pairs are expected to be reachable. Therefore to save computational resources and in preparation for future studies of even more complex systems we establish the following iterative procedure to sample different Milestones, within the Milestones and between the Milestones.

We divide the trajectory calculations into three phases: (i) search, (ii) sample, and (iii) terminate. In the first phase of (i) trajectories are launched from the anchor coordinates with velocities sampled from the Maxwell distribution at 300K. The trajectories are terminated when they reach a Milestone and the last configuration of the trajectory is kept. Fifty trajectories are used per anchor and the Milestones “discovered” are forward for the next step. The total number of Milestones that we considered (including the refinement of step (iii) to be discussed below) is 6186, which is not too far from the maximum number of 8010 mentioned above. The sampling procedure of Milestones is therefore quite effective even if it did not produced the computational saving we were hoping for in this particular case. The present system is highly reachable.

In step (ii) we sample configurations at the Milestone starting from the terminating configurations of step (i). The sampling is constrained to remain at the Milestones. The constraints are enforced in the simulations either exactly with Lagrangian constraints and SHAKE [13a] or with a restraining potential such as a umbrella biasing potential. The umbrella potential forces the system to remain in the neighborhood of the constraints[13b]. In the present study we used umbrella sampling as described in the Appendix C of reference [13b]. Structures from 1ps intervals of the sampling trajectory were tested for first hitting point distribution by integrating the trajectories backward in time until they hit another milestone (or hit the same Milestone and removed from the set) as described in section *III.2.1 Theory and algorithm*. The sampling provided about 50 initial conditions for the trajectories at the Milestones.

In step (iii) n_α trajectories are computed forward in time until they hit another Milestone and are terminated. We no longer terminate the trajectory if it crosses the Milestone it started from in contrast to the backward trajectories of step (ii). Their hitting time and the identity of the Milestone hit are recorded, and are used to estimate $n_{\alpha\beta}(t)$. Statistically converged $n_{\alpha\beta}(t)$ are used to estimate $K_{\alpha\beta}(t)= n_{\alpha\beta}(t)/n_\alpha$. If the interest focuses only on generating the stationary distribution and the overall first passage time it is sufficient to estimate the zero moment of $K_{\alpha\beta}(t) - (\mathbf{K})_{\alpha\beta} \equiv \int_0^\infty K_{\alpha\beta}(t)dt$ and an averaged first moment $\langle t\rangle_\alpha = \sum_\beta \int_0^\infty t \cdot K_{\alpha\beta}(t)dt$. These moments are used in Equation (6). Estimating the moments is statistically easier than estimating the full distribution as a function of time. Higher order moments of the MFPT require higher moments of the Kernel.

Possible termination of a trajectory on a Milestone is checked against all Milestones and not only on the Milestones discovered in step (i). It is possible that a new Milestone is reached at this stage. A “new” Milestone means a Milestone that is not on the list generated in step (i). In this case we add the new Milestone to the list and return to step (ii) for sampling of the newly discovered interface. This iterative procedure improves the sampling of Milestoning and allows for consistency checks. Of course it is still possible that some hidden contributions from other milestones remain. In the present case, the number of Milestones that we use is close to the maximum possible and therefore we believe that it is close to

complete. In the present investigation we fixed the number of anchors. In principle new anchors can be "discovered" as well as a part of the simulation. For example, if a configuration is found during a trajectory that is within a distance d larger than a predetermined threshold d_t from all other anchors then it makes sense to define the configuration as a new anchor and to add it to the set. This expansion, which is similar to earlier work on Markov State Models [30], will be integrated to our code in the future.

IV. Results from Straightforward Molecular Dynamics

There is more than one way to identify the folded state and to measure helical content. Here we present classification of the structures sampled from MD according to their hydrogen bond content. A summary of the results is in Table I. The normal α-helical H-bond is i to i +4, which makes W1-CO to H5-HN the only normal alpha-helical H-bond. Similar definition of states was used by [9p] in their simulations of WH5 using Replica Exchange and Markov State Model.

A formed hydrogen bond is denoted by "1" and a dissociated bond by "0". With 3 hydrogen bonds, a total of 2^3=8 hydrogen-bond microstates are possible in WH5, as shown in Table I. The largest observed population in the trajectory is of the unfolded state (0.776) with zero hydrogen bonds. We denote this state by 000. The third largest population (0.041) is of the helix with three hydrogen bonds, which we denote by 111. To appreciate the significance of the above probability we consider the hydrogen bond co-operativity. If hydrogen bonds are formed independently of each other, which we call the non-cooperative assumption, we can estimate the statistical weight of each of the 8 states from the populations of the three individual hydrogen bonds. It turns out that the 111 conformation is enriched in the trajectory by a factor of 20 compared to an estimate based on lack of co-operativity. This observation supports our notion of significant tendency to a structure and potential usability of the peptide as a folding nucleus. Also of interest are the slight enrichments of the states 110 and 011 by about a factor of 2 and the mild enrichment of the unfolded state 000 by a factor of 1.2. Other states are significantly depleted in accord with a picture of transitions between only a few states. A looser definition of the folded state is of a helix fraction (the fraction of time a particular amino acid is in a helical conformation). This definition suggests that the peptide spends 11–18% of its time in a helical conformation. This is not far from the 20% of helical conformation estimated from the UV-CD spectra at that temperature.

An alternative view of the structure of the peptide focuses on the dihedrals ϕ and ψ. Rather than counting the hydrogen bonds we probe if the (ϕ,ψ) values of each of the five amino acids is in a helical domain or not. We have five pairs when we include he blocking groups. The helix structure is defined with a narrow radius of 20 degrees around the ideal helix value (−62°,−41°). This picture is complementary to the hydrogen bonding view and is not necessarily identical. It takes into account potentially other forces (e.g. van der Waals, or preferred packing) that can also lead to a helical conformation. It is therefore useful to list the additional analysis below. We use again the 0/1 notation but here a value of 1 denotes a residue in a helical conformation and the value of zero means not a helix. The string representing the peptide state is of length five (in contrast to length of three when considering hydrogen bonds) which eliminates ambiguities. Note the significantly higher co-operativity of the helical state in dihedral space compared to the co-operativity measured from hydrogen bonding.

For the unfolded state the distribution of Trp-His side chain distances is very broad, in the range 3–18 Å. However for states with central hydrogen bonds formed (111, 011, 110 and 010) the distance distributions exhibit maxima at short distances of 3–6 Å. The correlation

coefficient between the time series of the central hydrogen bond length and the Trp-His is between 0.3–0.6 suggesting significant coupling between the two quantities. The distance between the Trp and His is therefore a useful measure of the extent of peptide folding. Indeed measurements of folding and unfolding kinetics are based on T-jump experiments that follow fluorescence relaxation of the Trp-His pair. The relaxation time is related to the folding and unfolding times by $\tau_r^{-1} = \tau_f^{-1} + \tau_u^{-1}$ where τ_r is the relaxation time, τ_f folding time and τ_u unfolding time. The experimental estimate has two exponential decays τ_r= 850±300ps or 5.3±1.9 ns. Computationally, 36 folding and 37 unfolding events were observed in the microsecond trajectory making it possible to estimate the folding and unfolding times as τ_f=23ns and τ_u=4.1ns. The calculated relaxation time is 3.5ns, in reasonable agreement with the long relaxation time observed in the experiment. In turn, the shorter measured relaxation time, 850±300ps, is comparable to the 0.6 ns relaxation time for formation of the central hydrogen bond, HB2, in the MD.

V. Results from Milestoning calculations

While the calculations are (of course) conducted using 90 anchors, it is useful to consider reduced representation for the analysis of the results. We first compare the Milestoning calculations with the Molecular Dynamics simulations at equilibrium. In Figure 4 we compare the hydrogen bonding patterns of the system (as defined in Table 1) extracted from the MD simulations and from Milestoning. While concrete probabilities differ, the probability of being in any hydrogen bonded state versus unfolded state is similar (probability of 0.82/0.78 for the state with zero hydrogen bond for Milestoning/MD respectively). With the very different sampling protocols in the two cases we gain confidence in our following analysis.

The relaxation time was found in Milestoning to be 4.0 ns, computed by the accumulated flow from nearby Milestones to the helix. This is quite close to the MD value of 3.5ns and to the experimental longer time of 5.3ns. We have considerably more statistics in the Milestoning calculations. Further insight into the reaction is obtained from a network analysis.

In Figure 7 we present a network picture for all the transitions between the 32 torsional states. Each pair of (ϕ,ψ) values (total of five pairs) is marked either as helical (1) or as something else (0) according to its position in the Ramachandran plot. This analysis is a better match for comparison with models like Lifson and Roig [9d] for helix folding. For example the state 00000 has no residues in the helix conformation while the state 11111 is a complete helix. A node in the graph corresponds to one secondary structure microstate as defined above and arrows denote directed flux between states. Thicker arrows correspond to larger fluxes. Interestingly the path that carries most of the weight to the folded state is a direct path from the unfolded state to the helical structure. Of course the unfolded state in our notation includes a large number of peptide configurations that differ significantly in their Cartesian coordinates. So the large weight is likely to include an entropic effect. Nevertheless, it is surprising that conformations with partial helical structure do not play a more significant role along the folding pathway.

Another observation is kinetic co-operativity of the folding pathways. Starting from the unfolded state (no residues in the helical conformation) the significantly populated pathways lead to states of two helical residues. States of one amino acid in a helical configuration are not populated significantly, nor do they serve as off-the-pathway traps (an off-the pathway trap is defined as a state with a significant flow of probability into the trap, flow that must be reversed to reach a desired product). This is consistent with the analysis of the 1μs

Molecular Dynamics trajectory in which a state with a single hydrogen bond was illustrated to have negative co-operativity.

We comment that the network presented in figure 7 is somewhat crude since we partition the space into cells defined by boundaries that are somewhat arbitrary. For example the low frequency state 11110 is not included, and the probable states 10000 and 01000 were sufficiently distorted so they were included in the 00000 state. Similarly the slight deviations between the Molecular Dynamics calculations and the Milestoning results shown in figure 6 results from a similar source.

There are ten states with two helical amino acids with detectable population. Eight of the ten states have a direct edge to the folded state. The state 10001 with maximally separated helical residues does not lead directly to the folded state. The state 00110 is not populated at all, supporting the notion of helix initiation from the chain terminals, initiation that progresses in one direction [9m].

In contrast to a sequential mechanism of folding, the states of three residues in a helical state are dead ends or off the pathway intermediates. There is only one pathway with significant flux that leads from states with 3 helical residues to the folded helix. If a molecule is misdirected to a partial helix of 3 amino acids it is likely to aim backward to a partial helix of 2 amino acids before trying again to reach the folded state.

The existence of two folding mechanisms direct from unfolded to the folded state and another from an intermediate with 2 hydrogen bonds to the folded (or to the unfolded state) is consistent with the kinetic partition mechanism of Thirumalai[31] for biological polymers. The kinetic partition mechanism predicts two parallel folding pathways one which is direct from the unfolded state to the folded state and one that involves significant intermediates. The same picture is observed here.

V. Discussions

Do the present results support the concept of a single helical turn as a nucleation site (i.e. as a structural site with a probability to form a particular shape which is significantly larger than random)? Consider first a reference system in which there is no energetic bias toward the helical conformation and only the entropy of the chain requires consideration. This argument is in the spirit of the Lifson and Roig model for helix formation[9d]. The three core pairs of (ϕ,ψ) dihedrals determine the torsion space of the helix. Let the range in the map covered by a single amino acid be u. The probability of finding one amino acid in the unfolded state is therefore ($1-u$). Acceptable conformations of peptides from the steric perspective mostly occupy the range of negative ϕ, and complete range of ψ. For qualitative analysis we set the helix to be in the full range of ϕ and consider a range of 18 degrees for ψ. Hence the probability of a single unfolded residue is $1 - 18/360 = 0.95$. Assuming lack of correlation between the states we have $0.95^3=0.86$ for the probability of the fully unfolded state, which is not far from the ~0.80 we obtained in Milestoning and ~0.78 in the straightforward MD simulations. Hence there is only slight enhancement of residues in helical conformations of WH5 in experiments or in our simulations compared to a random model.

A different analysis is more related to co-operativity of helix formation. Consider conformations that are at least partially a helix. Within this subset, do we observe enrichment of conformations with a larger number of hydrogen bonds compared to what we expect by chance? The probability of finding 3 residues in the helical conformation, or two sequential residues in a helix state is 0.08 from MD and 0.12 from Milestoning. The probability of these states following the random model is $0.05^3+2\times0.05^2=0.05125$ a factor of

about 2 lower than the simulated probabilities. As a nucleation site, the peptide under consideration has significant enrichment of the helical state and in the spirit of the Zwanzig et al. model[6] can serve at early events of protein folding. The present peptide is too small to make a meaningful comparison to the Lifson-Roig model of helix co-operativity[9d].

It is interesting that the anchors one step before folding in our network analysis include two residues in helical conformations. Note that in Milestoning we primarily consider fluxes. The two helical residues carry low probability (but significant flux). Hence, we do not argue that the two residues in a helical conformation are a stable intermediate, just that the reactive flux is passing through them. More studies will be required to identify metastable intermediates (if present).

The present study and the network picture we propose for helix formation in a pentapeptide make it possible to compare experimental observables with simulations and further compare with different computational approaches. We first comment that two relaxation times are observed experimentally. The two distinct folding mechanisms observed in the network, direct folding without formation of intermediate and folding through the earlier formation of a conformer with two helical residues can explain the distinct relaxation time scales. The long time scale of a few nanoseconds corresponds to the direct channel while in the shorter relaxation time trajectories pass through two hydrogen bond intermediates. Alternative explanation for the short time scale is seen in MD. It is the formation of the central hydrogen bond in 0.6ns relaxation time.

Another intriguing observation of the present study that awaits experimental verification is the existence of dead ends; i.e. that three residues in helical conformation are not on the pathway to folding. Rather they form off the pathway intermediates. One of these residues should change its conformation back to the unfolded state (to have two residues in the helical domain) before the molecule can proceed to the folded state.

It is tempting to project the network on a single reaction coordinate, with a simple physical interpretation, like was done in reference [9m] for a computed reaction coordinate and in reference [9k] with a distance measure. Reaction coordinates are useful for thermodynamic analysis. They provide simple picture for the process and the probability of finding the system at a specific position along the reaction coordinate can be computed. However, considerations of kinetics are more subtle. For a successful use of a reaction coordinate or an order parameter in a study of kinetics, separation of time scale is necessary. I.e. in kinetic calculations it is assumed that the motion perpendicular to the reaction coordinate is faster than motion along it. We have chosen the anchors with time scales in mind as we removed anchors that transition too fast and illustrate that most Milestones transition at comparable speed (even if they have the same value of the assumed reaction coordinate). Hence separation of time scales is not likely. From this perspective we illustrate in Figure 8 that neither the radius of gyration nor a distance measure are a good reaction coordinate for kinetics. Both do not show one-to-one correspondence with the anchors. Studies of reaction coordinates (rather than ad-hoc assumptions) can be made. One important choice is the iso-committor surfaces [14,24–25,32], which requires significant additional calculations. It is also possible to start from a network (Fig. 7) and to find a MaxFlux path [33] on this network, like was done in [13c].

If a barrier is found along a pathway segment that keeps a single value of the order parameter then the free energy profile for kinetics is incorrect. During the thermal averaging at a fixed value of the order parameter the barrier will be “washed out” and the kinetics will be predicted to be too fast.

This is not to say that optimal pathway(s) could not be found. It is possible to estimate the pathways of maximum flux on the Milestone network following the method described in references[30b13c]. These pathways can (potentially) lead to better mechanistic understanding. However, the relative simple network of Figure 7 makes such analysis unnecessary in the present case.

Hummer et al. [9j] used a different force field AMBER in straightforward MD and suggests remarkably short time scale (0.1ns) for the folding of the related peptide of A_5. This time is significantly shorter compared to the present simulations and experiments[16]. Detailed comparison between different force fields and experimental results for this peptide was published in reference [9l]. It should be noted that the folding mechanism described in that paper resembles our network pathway with 2 amino acids with helical conformations as intermediates, which indeed corresponds to shorter time observed experimentally. Hummer et al. proposed sequential formations of hydrogen bonds leading to the creation of a helix. In our study the emphasis in the mechanism is on the positions in the (ϕ,ψ) map and not necessarily on hydrogen bond formation.

Another intriguing study of pentapeptides is in reference [9k]. A reaction coordinate based on a distance measure was used and a free energy profile was computed. The resulting free energy profile was found flat and barrier-less. However, as argued above, it is difficult to project the above system onto one dimension free energy surface and a projection of this type may result in kinetics that are too fast. Indeed in a later manuscript [9j] the same authors suggested an enthalpy barrier of about 13.9 kJ/mol.

Perhaps more related to the present study are the two investigations employing the Markov State Model (MSM) to investigate peptide folding [9p, 15]. Similar to Milestoning, MSM employs partition of space and constructs a probabilistic model for the kinetics and thermodynamics. However, Milestoning is different from MSM in a number of ways that impact the results and the efficiency of the calculations. In brief, Equation (5) is using a non-Markovian model, and no meta-stable states are computed or assumed. Milestoning is based on fluxes through interfaces, not on states. An obvious consequence of this difference can be demonstrated in passage over large barriers, which is explained in section III.2.1, where it is possible to speed up barrier passing by placing a number of Milestones as we climb up the barrier. Estimates of the flux can be done as usual. However, intermediate Markovian states along the barrier are hard to imagine and to define. As a result the number of states and anchors that can be use in MSM is usually much smaller than in Milestoning. Indeed 4 states were used in [15] and 32 in [9p] compared the 90 anchors used in Milestoning. Moreover, the MSM calculations are mostly based on analysis of long trajectories (hundreds of nanoseconds) and not on a large number of short trajectories (tens of picoseconds). The MSM is very successful in re-producing the long time behavior of the kinetics and good agreement with the detailed simulations that it is based on is demonstrated. The small number of states also suggests simpler interpretation than in a Milestoning calculation in which tens of thousands of interfaces require tracking. The accumulated computational resources of these different investigations were comparable in this case (a few microseconds) while Milestoning is better suited to run in parallel a large number of cores. The detailed picture with 90 anchors allows us to analyze shorter time behavior not accessible to MSM. Finally, it is interesting to note that a study merging Milestoning principles of fluxes between interfaces with MSM has been published recently[34].

VI. Conclusions

We combined experimental and computational results on a pentapeptide Ac-WAAAH-NH_2 to obtain a deeper understanding of a folding mechanism. The use of Milestoning makes it

possible to compute a comprehensive kinetic and thermodynamic network for this small peptide. The use of NMR and CD helped establish the presence of a helical turn, even if at low probability. In Milestoning we exploit the efficient convergence properties of a large ensemble of short trajectories to estimate a local operator. From biophysical perspective this study establishes unexpected off the pathway intermediates with three residues in the helix state that cannot directly proceed to the folded state. Instead they must dissociate to two residues in a helical state first before reforming the helical turn.

Acknowledgments

Supported in part by NIH grant GM059796 to RE. Supported in part by Baylor URG to GSJ. Supported in part by Big XII fellowship from University of Kansas to KK. GSJ thanks Dr. Ad Bax from LCP, NIDDK, NIH for critical review and help with the NMR measurements. GSJ gratefully acknowledge Dr. William A. Eaton for critical review of the manuscript and helpful discussion. RE thanks Eric Vanden Eijnden for enlightening discussions.

References

1. Yang WY, Gruebele M. Biophysical Journal. 2004; 87(1):596–608. [PubMed: 15240492]
2. Socci ND, Onuchic JN, Wolynes PG. Proteins-Structure Function and Bioinformatics. 1998; 32(2): 136–158.
3. Go N. Annual Review of Biophysics and Bioengineering. 1983; 12:183–210.
4. Hardin C, Luthey-Schulten Z, Wolynes PG. Proteins-Structure Function and Bioinformatics. 1999; 34(3):281–294.
5. Levy Y, Caflisch A, Onuchic JN, Wolynes PG. Journal of Molecular Biology. 2004; 340(1):67–79. [PubMed: 15184023]
6. Zwanzig R, Szabo A, Bagchi B. Proceedings of the National Academy of Sciences of the United States of America. 1992; 89(1):20–22. [PubMed: 1729690]
7. Bryngelson JD, Wolynes PG. Proceedings of the National Academy of Sciences of the United States of America. 1987; 84(21):7524–7528. [PubMed: 3478708]
8. (a) Scholtz JM, Baldwin RL. Annual Review of Biophysics and Biomolecular Structure. 1992; 21:95–118.(b) Thompson PA, Munoz V, Jas GS, Henry ER, Eaton WA, Hofrichter J. Journal of Physical Chemistry B. 2000; 104(2):378–389.(c) Williams S, Causgrove TP, Gilmanshin R, Fang KS, Callender RH, Woodruff WH, Dyer RB. Biochemistry. 1996; 35(3):691–697. [PubMed: 8547249] (d) Clarke DT, Doig AJ, Stapley BJ, Jones GR. Proceedings of the National Academy of Sciences of the United States of America. 1999; 96(13):7232–7237. [PubMed: 10377397] (e) Jas GS, Eaton WA, Hofrichter J. Journal of Physical Chemistry B. 2001; 105(1):261–272.
9. (a) Brooks CL, Case DA. Chemical Reviews. 1993; 93(7):2487–2502.(b) Tobias DJ, Brooks CL. Biochemistry. 1991; 30(24):6059–6070. [PubMed: 2043644] (c) Scheraga HA. Pure and Applied Chemistry. 1978; 50(4):315–324.(d) Lifson S, Roig A. J Chem Phys. 1961; 34:1963–1974.(e) Zimm B, Bragg J. J Chem Phys. 1959; 31:526–535.(f) Brooks CL. Journal of Physical Chemistry. 1996; 100(7):2546–2549.(g) Daggett V, Levitt M. Journal of Molecular Biology. 1992; 223(4): 1121–1138. [PubMed: 1538392] (h) Buchete NV, Straub JE. Journal of Physical Chemistry B. 2001; 105(28):6684–6697.(i) Sorin EJ, Pande VS. Biophysical Journal. 2005; 88(4):2472–2493. [PubMed: 15665128] (j) Hummer G, Garcia AE, Garde S. Proteins-Structure Function and Genetics. 2001; 42(1):77–84.(k) Hummer G, Garcia AE, Garde S. Physical Review Letters. 2000; 85(12):2637–2640. [PubMed: 10978126] (l) Hegefeld WA, Chen SE, DeLeon KY, Kuczera K, Jas GS. J Phys Chem A. 2010; 114(47):12391–12402. [PubMed: 21058639] (m) Kuczera K, Jas GS, Elber R. J Phys Chem A. 2009; 113(26):7461–7473. [PubMed: 19354256] (n) Mahadevan J, Lee KH, Kuczera K. Journal of Physical Chemistry B. 2001; 105(9):1863–1876.(o) Barth E, Kuczera K, Leimkuhler B, Skeel RD. J Comput Chem. 1995; 16(10):1192–1209.(p) De Sancho D, Best RB. Journal of the American Chemical Society. 2011; 133(17):6809–6816. [PubMed: 21480610]
10. Eaton WA, Munoz V, Hagen SJ, Jas GS, Lapidus LJ, Henry ER, Hofrichter J. Annual Review of Biophysics and Biomolecular Structure. 2000; 29:327–359.

11. (a) Elber R, Meller J, Olender R. Journal of Physical Chemistry B. 1999; 103(6):899–911.(b) Bierzynski A, Kim PS, Baldwin RL. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences. 1982; 79(8):2470–2474.
12. Hummer G, Kevrekidis IG. Journal of Chemical Physics. 2003; 118(23):10762–10773.
13. (a) West AMA, Elber R, Shalloway D. Journal of Chemical Physics. 2007; 126(14)(b) Majek P, Elber R. Milestoning without a Reaction Coordinate. Journal of Chemical Theory and Computation. 2010; 6(6):1805–1817. [PubMed: 20596240] (c) Kirmizialtin S, Elber R. J Phys Chem A. 2011; 115(23):6137–6148. [PubMed: 21500798]
14. EWN, Vanden-Eijnden E. Annual Review of Physical Chemistry. 2010; 61:391–420.
15. Buchete NV, Hummer G. Journal of Physical Chemistry B. 2008; 112(19):6057–6069.
16. Mohammed OF, Jas GS, Lin MM, Zewail AH. Angewandte Chemie-International Edition. 2009; 48(31):5628–5632.
17. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Journal of Chemical Physics. 1983; 79(2):926–935.
18. Hess B, Kutzner C, van der Spoel D, Lindahl E. Journal of Chemical Theory and Computation. 2008; 4(3):435–447.
19. Kaminski G, Friesner R, Tirado-Rives J, Jorgensen WL. The Journal of Physical Chemistry B. 2001; 105(28):6474–6487.
20. Faradjian AK, Elber R. Journal of Chemical Physics. 2004; 120(23):10880–10889. [PubMed: 15268118]
21. Vanden-Eijnden E, Venturoli M. Journal of Chemical Physics. 2009; 130(19)
22. Vanden Eijnden E, Venturoli M, Ciccotti G, Elber R. Journal of Chemical Physics. 2008; 129(17): 174102. [PubMed: 19045328]
23. Moroni D, Bolhuis PG, van Erp TS. Journal of Chemical Physics. 2004; 120(9):4055–4065. [PubMed: 15268572]
24. Ma A, Dinner AR. Journal of Physical Chemistry B. 2005; 109(14):6769–6779.
25. Lechner W, Dellago C, Bolhuis PG. Journal of Chemical Physics. 2011; 135(15)
26. Peters B, Beckham GT, Trout BL. Journal of Chemical Physics. 2007; 127(3)
27. Shalloway D, Faradjian AK. Journal of Chemical Physics. 2006; 124(5)
28. Elber R, Roitberg A, Simmerling C, Goldstein R, Li HY, Verkhivker G, Keasar C, Zhang J, Ulitsky A. Computer Physics Communications. 1995; 91(1–3):159–189.
29. (a) Elber R, West A. Proceedings of the National Academy of Sciences USA. 2010; 107:5001–5005.(b) Elber R. Biophysical Journal. 2007; 92(9):L85–L87. [PubMed: 17325010]
30. (a) Chodera JD, Singhal N, Pande VS, Dill KA, Swope WC. Journal of Chemical Physics. 2007; 126(15)(b) Noe F, Schutte C, Vanden-Eijnden E, Reich L, Weikl TR. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106(45):19011–19016. [PubMed: 19887634]
31. Thirumalai D, Klimov DK, Woodson SA. Theoretical Chemistry Accounts. 1997; 96(1):14–22.
32. Peters B, Trout BL. Journal of Chemical Physics. 2006; 125(5)
33. Huo SH, Straub JE. Journal of Chemical Physics. 1997; 107(13):5000–5006.
34. Schutte C, Noe F, Lu JF, Sarich M, Vanden-Eijnden E. Journal of Chemical Physics. 2011; 134(20)

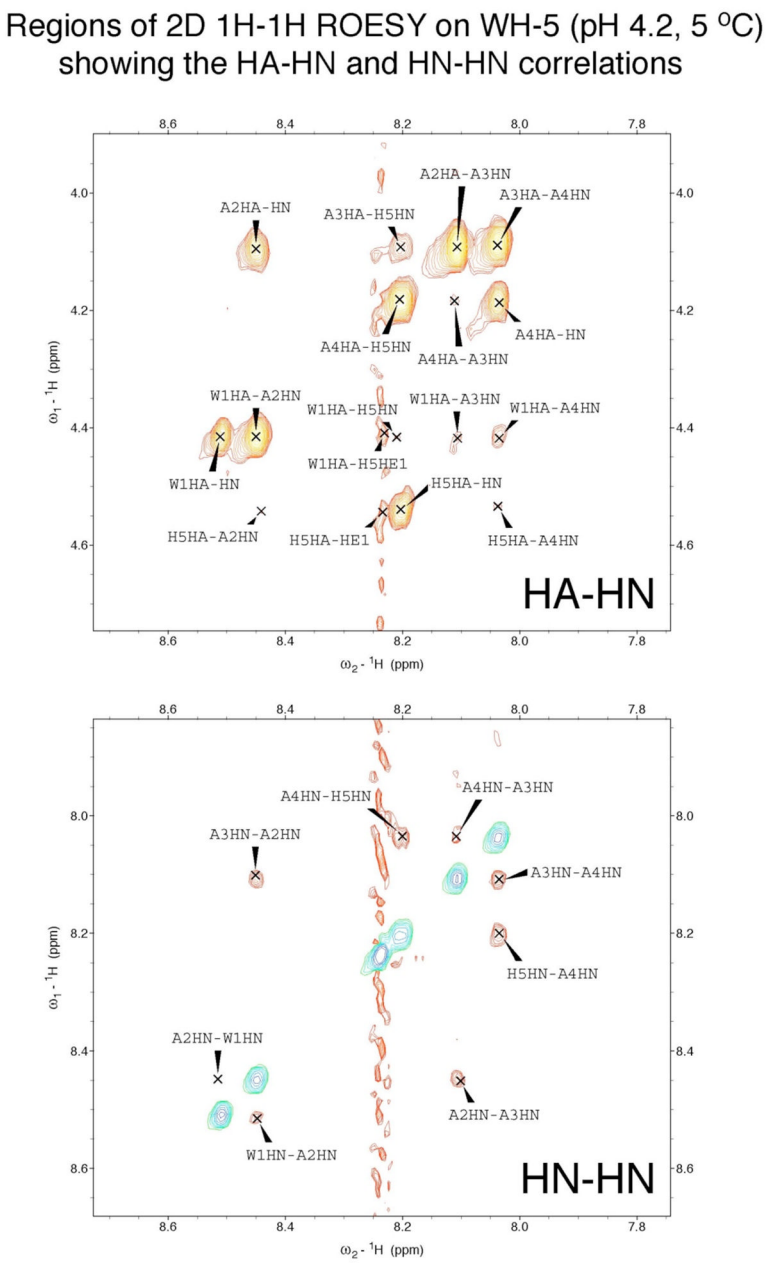

Figure 1.

2D ^{1}H-^{1}H ROESY and 2D ^{1}H-^{13}C HSQC measurements were employed to measure the ^{1}H and ^{13}C chemical shifts and obtain ^{1}H-^{1}H distance information. The $^{3}J_{HNH\alpha}$ coupling constants were also determined. The ROESY and HSQC experiments were performed on a 5 mM sample of the peptide, WAAAH, at pH 4.2 and 5 °C (the $^{3}J_{HNH\alpha}$ constants were also measured at 20 °C).

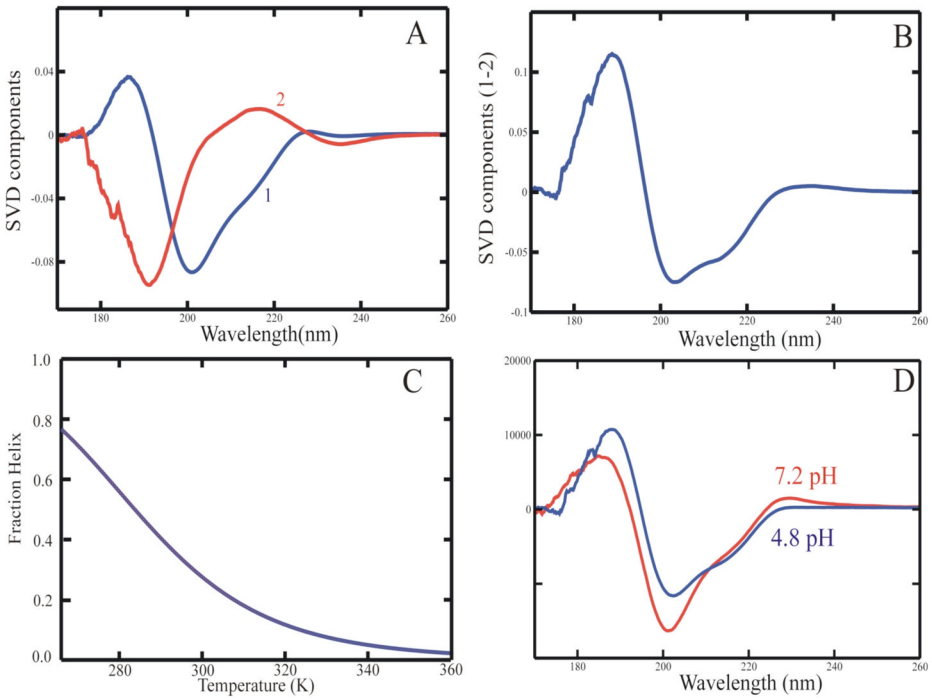

Figure 2.

Far-UV Circular Dichroism (CD) measurements in Wh5. (A) SVD analysis of the measured CD spectra of Wh5 as a function of temperature. Far UV spectra were recorded every 10 K from 266 K to 363 K in 20 mM acetate buffer and at pH 4.8. The first (blue) and second (red) components are presented here, as all other components were noise. (B) Signature helix CD spectrum is obtained by subtracting the second (red) SVD component from the first (blue) SVD component. (C) Fraction helix as a function of temperature obtained from the measured unfolding as a function of temperature. (D) Low temperature (266K) CD spectrum of wh5 in pH 4.8 and pH 7.2.

Figure 3.a

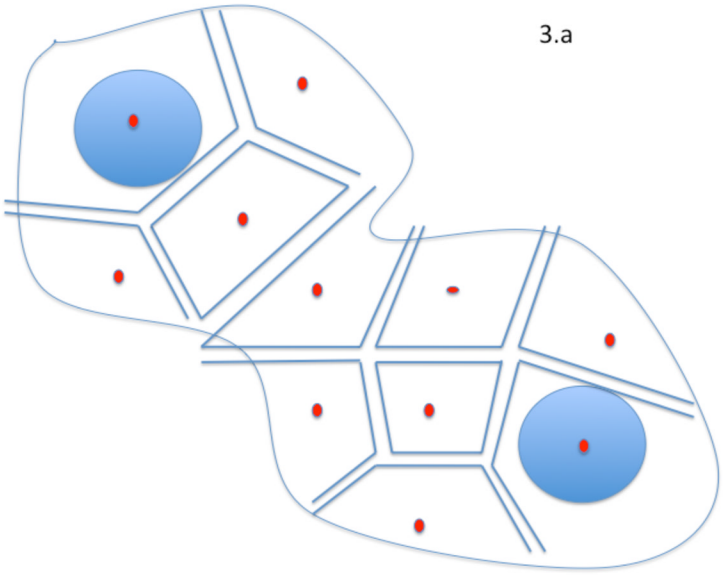

Figure 3.b

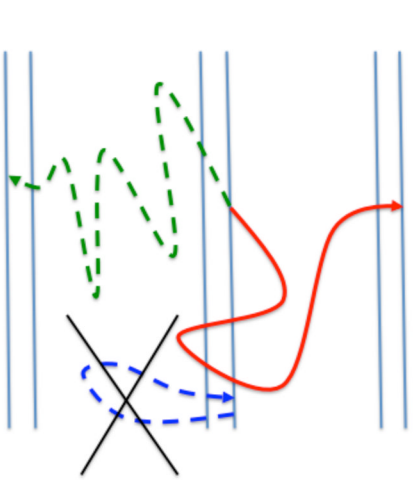

Figure 3.

A schematic representation of anchor distribution and Milestones on a two dimensional energy surface. In 3.a we consider transitions between two states denoted by blue circles. Anchors are red points and vectors in phase space distributed along the potential energy surface (here only the coordinate part of the anchors is used). We separate the anchors by directional Milestones. The Milestone closer to an anchor is of trajectories pointing into the cell of the anchor. Milestones are separated by at least Δ (see text for more details). In 3.b we illustrate trajectories between directional Milestones. We first checked if the trajectories are sampled from first hitting distribution. The dashed lines are backward trajectories initiated from the same Milestone. Backward trajectories that hit first the same Milestone they started from are removed from the set (dashed blue line is removed). Backward trajectories that that hit first another Milestone are kept (green trajectory) and are integrated

forward in time until they hit a new Milestone (red line). The time from initiation to forward termination is recorded, as well as the index of the initiating and terminating Milestones.

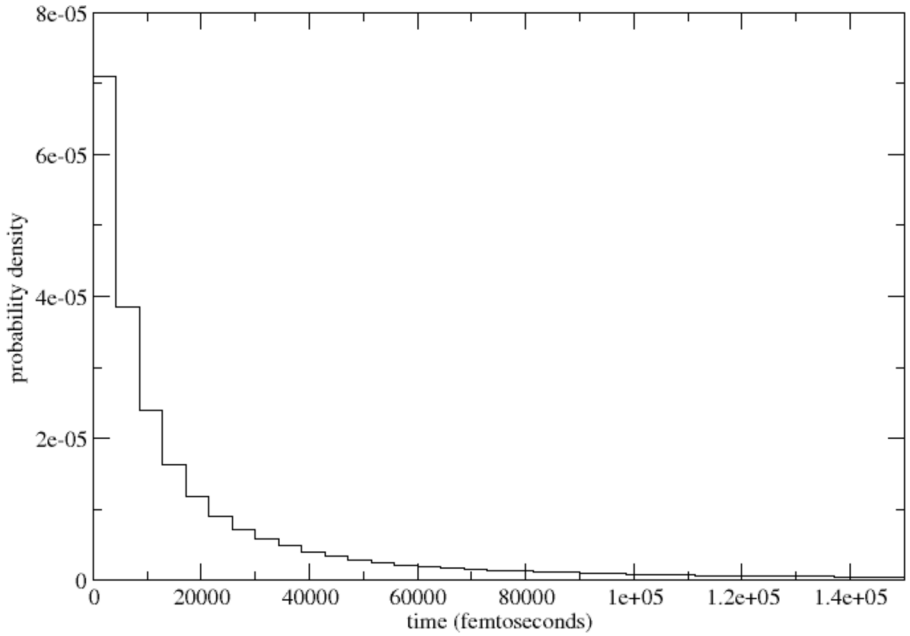

Figure 4.

The distribution of termination times from all the 201,432 Milestoning trajectories. Note the high significant peak at tens of picoseconds and the long tail that continues to a few nanoseconds. The average termination time is 33.8 picoseconds illustrating the efficiency and the highly parallelize algorithm.

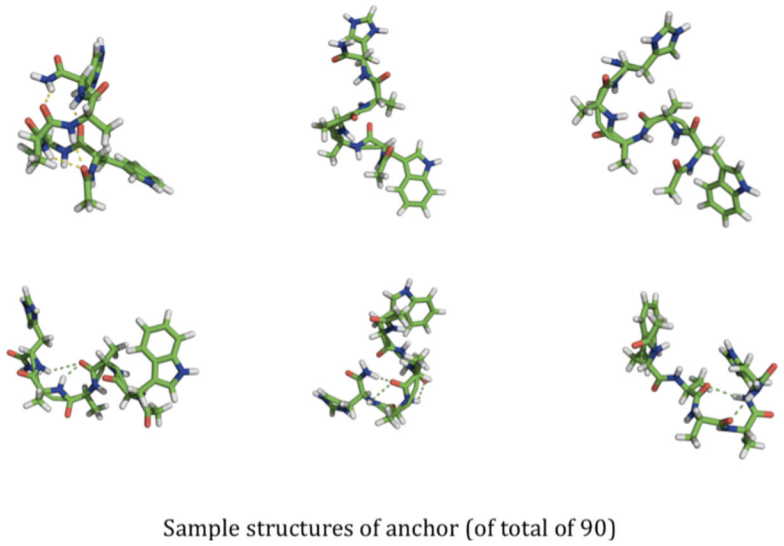

Figure 5.
Sample of peptide conformations that were used as anchors. Anchor is a coarse description of the peptide and is defined by the five (ϕ,ψ) pairs. The distances between the anchors are defined as Euclidean distance (with periodicity) using these coarse variables. The simulations are conducted (of course) with all atoms and in a water box. The anchors are used only to define the boundary between the interfaces.

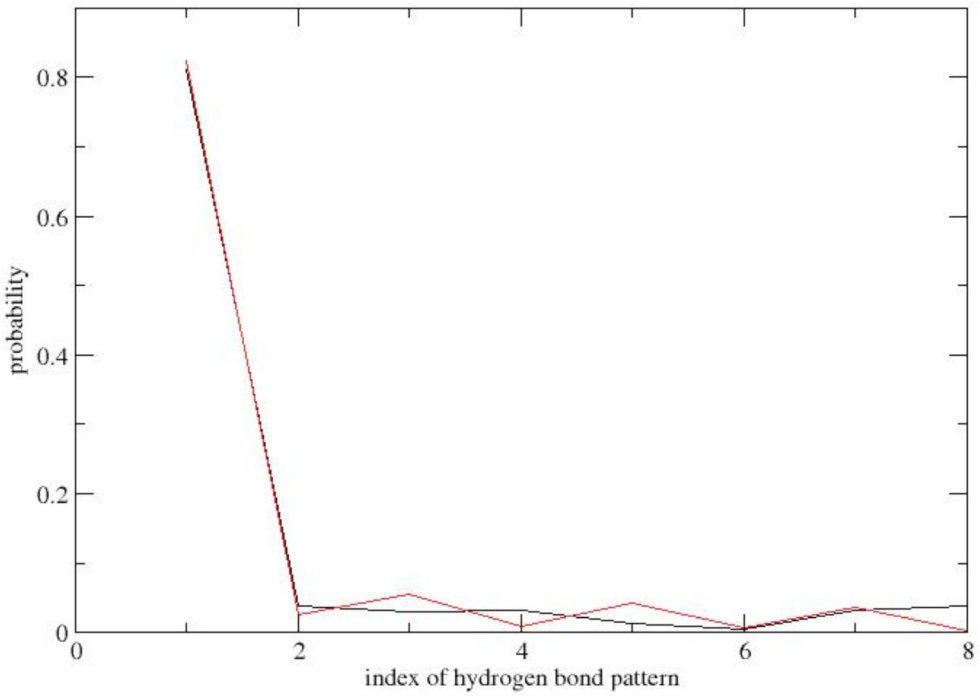

Figure 6.

Comparing hydrogen-bonding patterns (as defined in Table 1) obtained in Molecular dynamics and in Milestoning simulations. The probability of a pattern versus the index of the pattern are shown (see text and Table 1 for more details)

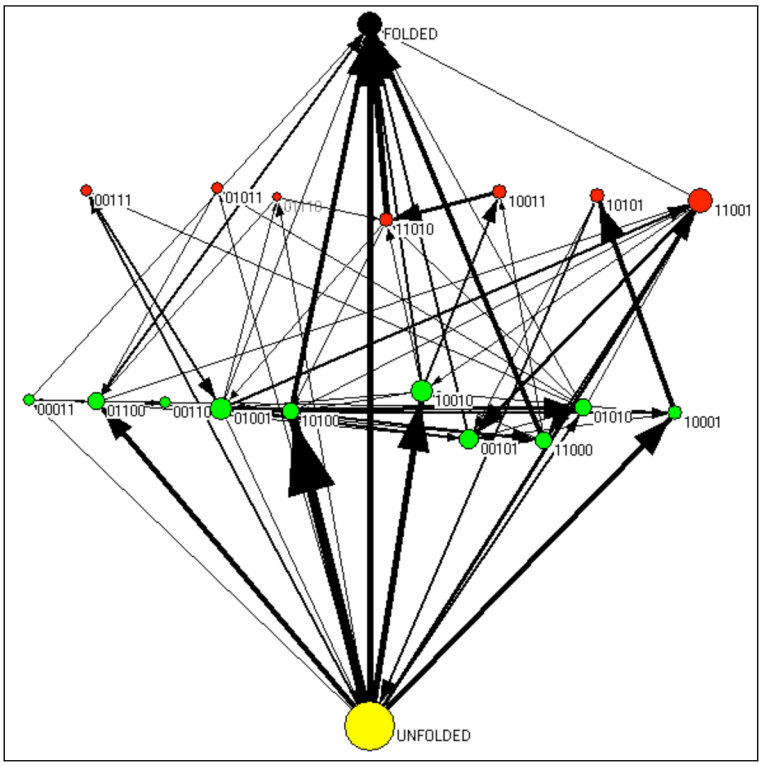

Figure 7.
A network description of secondary structure transitions in the system of WH5. A node is a state with a predefined secondary structure. A node is denoted by a circle and the area of the circle is proportional to the equilibrium probability of a state. We use “1” to denote an amino acid in a helical conformation and “0” for other states. The unfolded state is “00000” and the fully folded state is “11111”. An arrow denotes a possible direct transition between two states. The thicker the arrow the larger is the flux between the two states. Note that a direct transition from the unfolded state to the folded state is the pathway that carries the largest amount of flux.

Figure 8.a

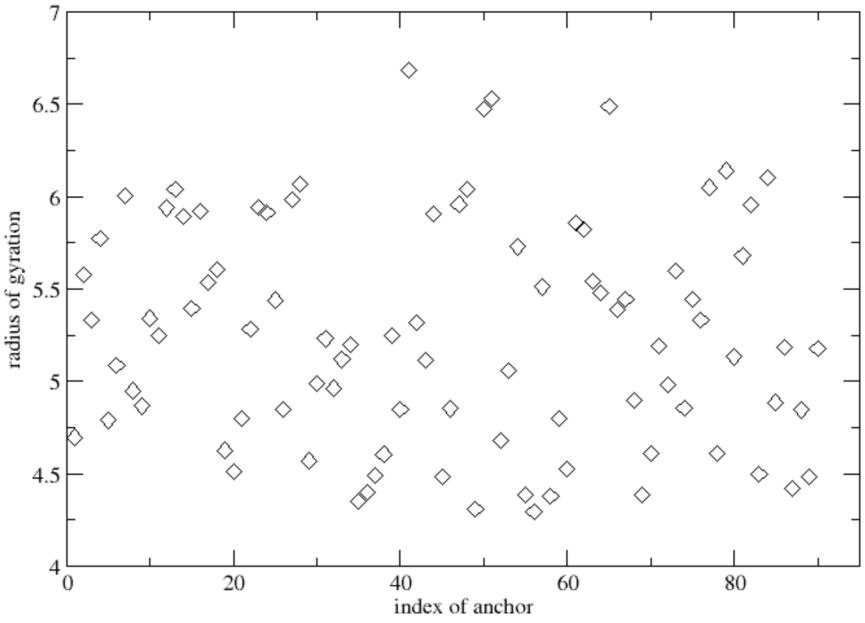

Figure 8.b

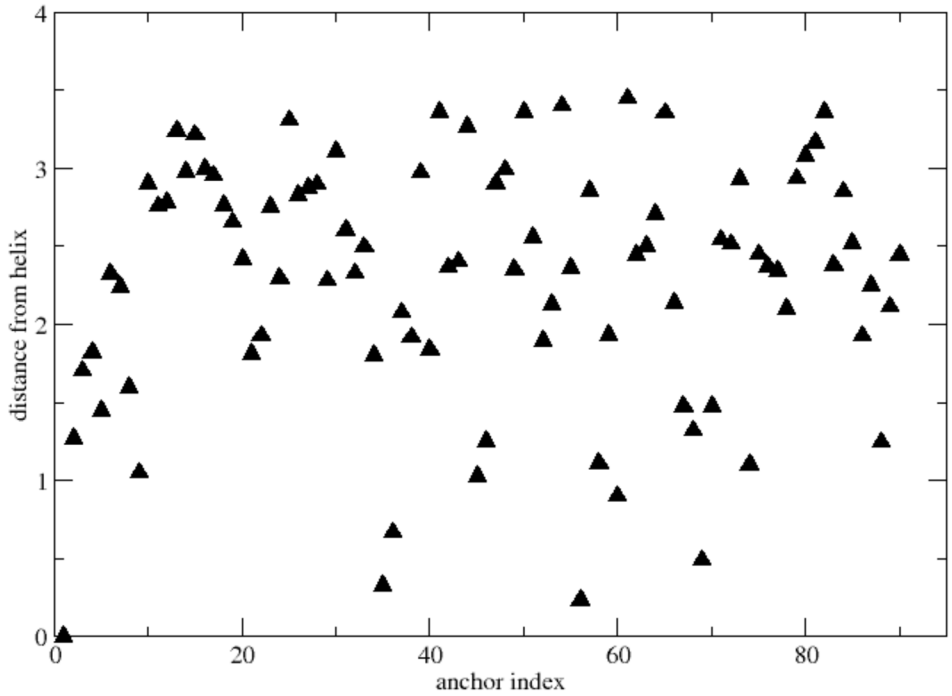

Figure 8.

Projections of the anchor index on two plausible reaction coordinates for peptide folding: (8.a) radius of gyration, and (8.b) distance from the helix. The set of anchors that we used (the five pairs of (ϕ,ψ) dihedrals) is necessary to capture the essential kinetics observed in the Molecular Dynamics simulations and in our Milestoning study. For a reaction coordinate to capture the same space spanned by the anchors, it must provide one-to-one correspondence between the value of the reaction coordinate and the anchor index. It is obvious that the above two are not single value function of the anchor index. Therefore, we do not believe that a one dimensional reaction coordinate with reasonably smooth orthogonal surface (formally iso-committor surface can always be defined) can describe the correct kinetics of the process.

Table I

Hydrogen bond patterns in the MD trajectory - pattern 'ijk' corresponds to states i,j,k of hydrogen bonds HB1, HB2, and HB3, respectively, with i,j,k=1 if the N…O distance is below 3.6 A. The co-operativity is the ratio of the observed population to that predicted by a model of independent hydrogen bond formation, based on the average trajectory populations of 0.136, 0.164 and 0.094 for HB1, HB2 and HB3, respectively. The bonds are defined as HB1: Ac-CO to A4-HN, HB2: W1-CO to H5-HN and HB3: A2-CO to NH2 blocking group.

Structure number	HB pattern	MD population	Co-operativity
1	000	0.776	1.2
2	100	0.036	0.35
3	010	0.039	0.30
4	110	0.054	2.7
5	001	0.017	0.25
6	101	0.006	0.06
7	011	0.031	2.3
8	111	0.041	20.

Table 2

Helical patterns of a pentapeptide analyzed in MD simulations. The analysis of co-operativity is similar to the calculation performed in Table 1.

Index	Pattern	MD population	Co-operativity
1	00000	0.6399	1.1
2	10000	0.0674	0.6
3	01000	0.0617	0.6
4	11000	0.0278	1.5
5	00100	0.0615	0.7
6	10100	0.0165	1.0
7	01100	0.0163	1.1
8	11100	0.0208	7.2
9	00010	0.0320	0.8
10	10010	0.0080	1.1
11	01010	0.0042	0.6
12	11010	0.0041	3.2
13	00110	0.0027	0.4
14	10110	0.0035	3.0
15	01110	0.0033	3.1
16	11110	0.0061	30.1
17	00001	0.0153	1.1
18	10001	0.0014	0.5
19	01001	0.0020	0.8
20	11001	0.0006	1.4
21	00101	0.0016	0.7
22	10101	0.0003	0.8
23	01101	0.0005	1.4
24	11101	0.0004	5.9
25	00011	0.0003	0.3
26	10011	0.0001	0.8
27	01011	0.0002	1.0
28	11011	0.0002	6.1
29	00111	0.0002	1.2
30	10111	0.0002	8.4
31	01111	0.0004	14.3
32	11111	0.0005	126.5