

# **Extending Molecular Dynamics Timescales with Milestoning: Example of Complex Kinetics in A Solvated Peptide**

Anthony M. A. West

Department of Physics

Cornell University

Ithaca NY 14853

Ron Elber<sup>\*</sup>

Department of Computer Science

Upson Hall 4130

Cornell University, Ithaca NY 14853

And

David Shalloway

Department of Molecular Biology and Genetics

Cornell University, Ithaca NY 14853

## **Abstract**

A recently introduced computational algorithm to extend time scales of atomically detailed simulations is illustrated. The algorithm, Milestoning, is based on partitioning the dynamics to a sequence of trajectories between “milestones” and constructing a non-Markovian model for the motion along a reaction coordinate. The kinetics of a conformational transition in a blocked alanine is computed and shown to be accurate, more efficient than straightforward Molecular Dynamics by a factor of about 9, and non-exponential. A general scaling argument predicts linear speed-up with the number of milestones for diffusive processes, and exponential speed-up for transitions over barriers. The algorithm is also trivial to parallelize. As a side result Milestoning also produces the free energy profile along the reaction coordinate, and is able to describe non-equilibrium motions along one (or a few) degrees of freedom.

# I. Introduction

Molecular Dynamics (MD) simulations are widely used in studies of biological macromolecules<sup>1</sup>. The simulations provide a widely tested and consistent tool for the study of equilibrium properties. However, the picture is less clear for the study of kinetics and time scales. Time scales of straightforward molecular dynamics simulations are significantly shorter than the experimental time scales of many biomolecular processes. Even though significant efforts have been made to design special hardware and software systems to extend the scope of MD simulations (e.g. a loose but very large cluster of PCs<sup>2</sup>, or a new type of supercomputer) these are not resources that are widely accessible to users. In practice MD simulations remain confined to times below microseconds. This means that direct integration is not feasible for many kinetic phenomena in biophysics (e.g. fast protein folding events take a millisecond).

In the present manuscript we demonstrate that Milestoning,<sup>3</sup> an approximate simulation method we developed recently, is accurate for kinetic studies and provides significant computational speed-up. The calculations of the non-exponential rate of a conformational transition in alanine dipeptide are faster by a factor of nine (on a serial machine) compared to straightforward MD simulations. General scaling arguments are provided that make it possible to predict speed up in other systems.

We organize the manuscript as follows: In the next section we summarize the main formulas of the milestoning approach. We refer the reader to an earlier manuscript<sup>4</sup> for

more detailed discussions and derivations. We then describe the concrete choice of the system and the implementation in a general molecular dynamics code. We finally present the results, discuss accuracy measures, and compare Milestoning to other approaches.

## II. Theory

Milestoning begins by assuming the availability of a set of non-intersecting *hypersurfaces* that represent the progress of the reaction (figure 1). Specifically, we choose  $M$  hypersurfaces  $H_s$  called milestones that partition the configuration space into  $M+1$  domains. While the hypersurfaces can be determined with the help of a reaction coordinate (see Section **V. Method**), the use of a reaction coordinate is not necessary. Figure 1 illustrates the arrangement:

Place Figure 1 Here

Local kinetics between milestones are determined by preparing a statistical ensemble of configurations in each  $H_s$ . In the canonical ensemble, which we prepared by Molecular Dynamics at constant temperature, configurations are sampled according to the usual weight  $\exp\left(\frac{-U(X)}{kT}\right)$ , subject to the constraint that  $X \in H_s$ , reducing the overall dimensionality of the system by one. Once a sample of configurations is available we integrate each ensemble member until its trajectory, which starts from  $X \in H_s$ , is

terminated when it reaches one of its two neighboring milestones in the forward ( $H_{s+1}$ ) or the backward direction ( $H_{s-1}$ ). The termination times are recorded and collected into two distributions, the distribution of termination times from  $s$  to  $s+1$   $K_s^+(\tau)$ , and the termination times of trajectories going from  $s$  to  $s-1$   $K_s^-(\tau)$ . For simplicity, we denote the two neighboring milestones by  $H_{s\pm 1}$ , and the two distributions of termination times by  $K_s^\pm(\tau)$  (throughout this manuscript the notations  $A_{s\pm 1}$  or  $A^\pm$  mean  $A_{s-1}$  or  $A_{s+1}$ , or  $A^+$  or  $A^-$ , respectively). The time  $\tau$  taken to reach a neighbor is called an *incubation time*, and the  $2M-2$  distributions of incubation times are denoted  $K_s^\pm(\tau)$ ,  $s = 1, 2, \dots, M$ .

We therefore have  $\int_0^\infty K_s(\tau) d\tau = 1$ ,  $s = 1, 2, \dots, M$ , where  $K_s \equiv K_s^+ + K_s^-$ .

These local kinetics are then “glued” together to recover the global kinetics of the process. The approximations involved in this “gluing” procedure will be discussed in the next section; for the moment we focus on the algorithm itself. For each  $s$ , we define  $P_s(t)$  to be the probability that the system’s position lies between  $H_{s-1}$  and  $H_{s+1}$  at time  $t$ , and that the last milestone that was crossed was  $H_s$ . (Note that the probabilities  $P_s(t)$  therefore overlap spatially, between neighboring milestones.) We shall also more loosely say that  $P_s(t)$  is the probability of being “at” milestone  $s$  at time  $t$ , but it should be understood that it is not the probability of being exactly in  $H_s$ . In a previous paper we showed that  $P_s(t)$  can be obtained by solving the integral equations

$$\begin{aligned}
P_s(t) &= \int_0^t \left[ 1 - \int_0^{t-t'} K_s(t-t') \right] Q_s(t') dt' \\
Q_s(t) &= \eta_s \delta(t-0^+) + \int_0^t \left[ K_{s+1}^-(t-t') Q_{s+1}(t') + K_{s-1}^+(t-t') Q_{s-1}(t') \right] dt'
\end{aligned} \tag{1}$$

These equations provide probability balance. Trajectories that make it into milestone  $s$  must come from milestones  $s+1$  or  $s-1$ , and trajectories lost from  $s$  must do so either to milestone  $s+1$  or to milestone  $s-1$ . In the second equation  $\eta_s$  is the initial milestone probability distribution and  $Q_s(t)$  is defined as the probability that the system makes a transition to milestone  $s$  at absolute time  $t$ . The first equation says that the probability of being at milestone  $s$  at time  $t$  equals the probability of having gotten there at some earlier time  $t'$  and having not left yet through the mechanism specified by  $K_s(t-t')$ . The second says that the probability of making a transition to milestone  $s$  (defined by  $Q_s$ ) equals the probability of first making a transition to one of its two neighbors  $s \pm 1$  at some earlier time  $t'$  and then hopping to  $s$  following the  $K_{s \pm 1}$  distribution of incubation times. (At this level of the model, a transition is a “hop.”) The delta function is shifted from  $t=0$  to  $t=0^+$  so that we have  $P_s(0) = 0$  and  $P_s(0^+) = \eta_s$ ,  $s = 1, 2, \dots, M$ , which simplifies the Laplace transform of the time-derivative:

$$\int_0^\infty \frac{dP}{dt} \exp(-\alpha t) dt = \tilde{P}(\alpha) = \alpha \tilde{P}(\alpha) - P(0) = \alpha \tilde{P}(\alpha).$$

The probability distribution  $P_s(t)$  represents the system’s behavior along the entire set of milestones as a function of time (hereafter the enumeration  $s = 1, 2, \dots, M$  is implicit).

For example, the *equilibrium* probability at Milestone  $s$  is

$$P_s^{eq} \equiv \lim_{t \rightarrow \infty} P_s(t) \quad (2)$$

This gives the free energy  $F_s$  of the milestones, via

$$F_s = -k_B T \log P_s^{eq} \quad (3)$$

As another example, suppose we begin at an initial milestone  $i$  so that  $\eta_s = \delta_{si}$  (Kronecker delta), and suppose we make milestone  $f \neq i$  absorbing by setting

$$K_f^\pm = 0 \quad (4)$$

so that all the probability flows from  $i$  to  $f$ . The function  $\varphi(t) \equiv \dot{P}_f(t)$  is called the *first passage time distribution* (FPTD). In particular, its first moment

$$\langle\langle \tau \rangle\rangle \equiv \int_0^\infty \tau \varphi(\tau) d\tau \quad (5)$$

is a measure of the inverse reaction rate, and is called the *mean first passage time*

(MFPT). The second moment of  $\varphi$ ,  $\langle\langle \tau^2 \rangle\rangle \equiv \int_0^\infty \tau^2 \varphi(\tau) d\tau$ , is also interesting since it gives

the width of the FPTD, and the standard deviation  $\sqrt{\langle\langle \tau^2 \rangle\rangle - \langle\langle \tau \rangle\rangle^2}$  measures the

deviation of the reaction from a pure single exponential behavior. Indeed, we can analyze

the reaction probability  $P_f(t) = \int_0^t \varphi(\tau) d\tau$  to determine if the kinetics are characterized

by, say, one or more exponentials. In the case of one exponential  $P_f(t) \sim (1 - e^{-kt})$ , the MFPT is the reciprocal of the reaction rate  $k = \langle\langle\tau\rangle\rangle^{-1}$ . It is important to emphasize that the calculation of MFPT does not imply single exponential kinetics. In fact, and as we demonstrate in this paper, Milestoning describes accurately the non-exponential kinetics observed in the conformational transition of alanine dipeptide.

In this paper, we compute  $P_s(t)$ ,  $F_s$ , and  $\langle\langle\tau\rangle\rangle$  for alanine dipeptide's conformational transition. We do not compute the last two by solving directly the integral equation (1) but instead use a more efficient method given in Appendix A, which was partly discussed in Shalloway and Faradjian<sup>5</sup>.

### III. The computational gain of Milestoning

The Milestoning approach allows for three sources of computational gain compared to straightforward molecular dynamics trajectories.

1. **Parallelization:** The first (which is trivial) is the ease of parallelization of the simulation, assuming a large pool of compute nodes. We need to run a larger number of trajectories in the milestoning approach compared to straightforward molecular dynamics since each milestone requires its own set of independent initial conditions, therefore Milestoning can use more processors. If we sample a total of  $J$  trajectories for different milestones and different initial conditions within a milestone, each of these trajectories

can be computed on a separate CPU for maximum efficiency (the Milestoning trajectories are significantly shorter than the usual MD trajectories, see below).

**2. Diffusive enhancement:** The second gain is for diffusive processes. If the typical distance between the reactant and the product is  $L$  and the system dynamics are diffusive (on a flat free energy surface), then the total time of a straightforward simulation of the entire reaction is proportional to  $L^2$ . If we distribute  $M$  milestones between the reactant and product, however, each milestone simulation (i.e., a single incubating trajectory from the ensembles required for the computation of  $K_s^\pm(\tau)$  for one particular  $s$ ) will take only a time  $\sim (L/M)^2$ . Since there are  $M$  milestones to compute, the calculation time of a complete trajectory is proportional to  $L^2/M$ . This is a factor of  $M$  (in addition to parallelization) that is obtained for a serial calculation.

**3. Exponential bootstrapping:** The third enhancement is associated with a concrete example (similarly to the gain of diffusive processes). Consider a sequence of milestones that follows a broad free energy barrier. Assume that trajectories initiated at a particular milestone  $H_s$  terminate at  $H_{s+1}$  with probability  $q \ll 1$ . Assume further that the probability of reaching  $H_{s+2}$  from  $H_{s+1}$  is  $q' \ll 1$ . In straightforward MD simulations we will need about  $1/(q \cdot q')$  trajectories to sample a transition from  $H_s$  to  $H_{s+2}$ . In the Milestoning approach we will need only  $(1/q) + (1/q')$  trajectories. This substantial savings is exponential in the number of milestones  $M$ .

All three factors suggest that the more milestones we have, the more efficient is the calculation. However, adding milestones without limit does not necessarily work since

the protocol discussed so far makes a crucial assumption about the dynamics of the system that cannot be justified in the general case for  $M \rightarrow \infty$ . In practice the number of milestones is kept reasonably small to ensure the correctness of the application while still taking advantage of the computational gains mentioned above. In the next section we discuss this assumption and how to verify its validity.

## IV. The Milestoning Approximation

Obviously there is no free lunch and the computational gains mentioned above for the Milestoning approach follow from certain assumptions about the reaction. In particular, the process of dividing the kinetics into a sequence of events requires *loss of memory* between contiguous milestones, i.e. the way we initiate a trajectory at milestone  $H_s$  must be independent of the detailed history that brought it to that particular milestone. Algorithmically, trajectories are initiated for a local kinetic study (between milestones) from a statistical ensemble that depends only on the initial coordinates and velocities (e.g. a Boltzmann distribution constrained to the hypersurface  $H_s$ ).

It is important to realize that this choice is different from the assumption of a system in a *global* statistical equilibrium. Since we examine kinetics, it is assumed that during the diffusion time from milestone  $H_s$  to milestone(s)  $H_{s\pm 1}$  the distribution of trajectories has enough time before termination to equilibrate the degrees of freedom within the corresponding milestones. More formally, we write

$$\int_0^{\infty} \rho_{FPT}(H_{s-1} \rightarrow H_s, t) dt \cong \int_0^{\infty} \rho_{FPT}(H_{s+1} \rightarrow H_s, t) dt \cong \rho_{eq}(H_s) \quad (7)$$

where  $\rho_{FPT}(H_s \rightarrow H_{s'}, t)$  is the spatial distribution of terminating trajectories at the neighboring milestones. The equilibrium distribution (as a function of the internal coordinates of the milestone  $H_s$ ) is  $\rho_{eq}(H_s)$  and it should approximate  $\rho_{FPT}(H_{s-1} \rightarrow H_s, t)$  and  $\rho_{FPT}(H_{s+1} \rightarrow H_s, t)$ . We expect that the milestoning approximation will be better when the number of milestones is small, giving more time for the transitional distributions to reach local statistical equilibrium. However, the smaller the number of milestones, the smaller is the computational gain. A compromise is desired based on a measure of the quality of the approximation. One possible measure is to compare the densities directly, however, this is an expensive calculation in high dimensions ( $3N - 1$ ) (where  $N$  is the number of atoms). Therefore in the present manuscript we propose and carry out different (and cheaper) tests that are presented in detail in the **Results** section.

The most telling test that we performed is of the accuracy of the rate calculation as a function of the number of milestones. In principle the rate should be independent of the number of milestones provided that the number is sufficiently small (as we demonstrated in our earlier<sup>3</sup> and present paper). Here we show that accurate results are obtained already for 19 milestoning allowing for significant speedup. While this test provides the answer we need, it requires multiple runs for the same problem. We therefore explore two other possibilities with mixed success. (At present the best test is still to re-compute the rate with different numbers of milestones).

The second test we performed is to compare moments of the distributions in (7). For the example described in this paper this test was insensitive, i.e. the moments were too similar, or too noisy to differentiate between cases that provide accurate rates and cases that do not.

The third test is more limited in scope. Nevertheless and perhaps surprisingly, it correlates well with rate accuracy. We examine the velocity relaxation time along the reaction coordinate and compare it to the incubation time between milestones (defined by

$\tau_s^{\parallel} = \int_0^{\infty} K_s(\tau) d\tau$ ). For the calculations to be accurate the incubation time must be longer

than the velocity relaxation time. This condition is clearly weaker than (7) since Milestoning implies coordinate relaxation in addition to velocity relaxation, and coordinate relaxation is known to be slower. Nevertheless, empirically this test was found to work well (**VI. Results**).

## V. Methods

In figure 2 we show a picture of the blocked alanine that we studied in aqueous solution. Also shown is the  $\psi$  dihedral angle (defined by the four consecutive atoms -  $N - C_{\alpha} - C - N$ ).

Place Figure 2 Here We investigate the kinetics of a transition from an  $\alpha$  helix to a  $\beta$  sheet of this small peptide in aqueous solution. The first task in setting up a Milestoning

calculation is a concrete definition of the hypersurfaces, or milestones. The hypersurfaces should be chosen to support the time scale argument outlined in Section **IV. The Milestoning Approximation**. That is, the relaxation to equilibrium in the milestone must be fast compared to the transition time between milestones. A parameter that we control and can help establish the time scale separation (as we demonstrate in Section **VI. Results**) is the distance between sequential milestones. If the distance is large then the transition time between the milestones will be long without affecting the relaxation time within the milestone. However we do not know of a rigorous and/or efficient method for optimizing such general hypersurfaces, and we therefore use inter-milestone distance as a heuristic. Intuitively we expect that motions in a direction that is not costly in energy (or free energy) relax more quickly to statistical equilibrium; i.e. that the vector  $g_s$  orthogonal to hypersurface  $s$  is parallel to the gradient of the energy (free energy). This intuition suggests that  $g_s$  is in the direction of the minimum energy (free energy) path, an argument which is in accord with some implementations of the transition state theory<sup>6,7</sup>.

In the present study of alanine dipeptide the system is relatively small and it is simple to come up with a good minimum energy path without explicit calculation. To construct the milestones we consider a set of coordinates  $x_s$ . Each of the  $x_s$  is in one of the milestones  $H_s$ . It is well known that the  $\alpha$  to  $\beta$  transition in alanine dipeptide is described primarily by changes in the  $\psi$  dihedral angle while keeping a negative  $\phi$  (see for instance<sup>8</sup>). The changes in the  $\psi$  dihedral angle during the transition allow for only mild variations in the energy or free energy compared to energy changes observed when other internal degrees of freedom are varied. This observation suggests  $\psi$  as a reasonable

choice for a minimum energy path. We therefore construct the initial coordinates  $x_s \in H_s$  using different values of  $\psi$ , and we somewhat arbitrarily set  $\phi = -150^\circ$ . We compute  $x_s$  by specifying values for the two torsion angles  $(\phi, \psi)$  as harmonic constraint inputs to MOIL's conjugate gradient minimizer of the potential energy<sup>9</sup>. To prepare the entire set of initiating structures  $x_s$  for all milestones, we created 144 such structures having  $(\phi_s, \psi_s) = (-150, 175 + 2.5s)^\circ$ ,  $s = 1, 2, \dots, 144$ , so that they span the entire  $360^\circ$  range of  $\psi$ .

The sequence of coordinates  $x_s$  is still insufficient to define the milestones, which are hypersurfaces in  $N - 1$  dimensions. This dilemma is similar to the use of minimum energy and free energy paths for rate calculations. In these calculations hypersurfaces orthogonal to the reaction coordinate are used to define the dividing surface between reactants and products. However, only the reaction coordinate (which is a one dimensional curve) is typically known and not the dividing hypersurface. A useful and widely used approximation<sup>10-12</sup> is the use of a Cartesian hyperplane perpendicular to the reaction coordinate. A Cartesian representation is especially useful in more complex reaction coordinates that cannot be described in terms of one or a few internal coordinates. In anticipation of such cases we developed and tested here the most general approach. A hyperplane is determined by a point in  $H_s$  ( $x_s$ ), and by a vector orthogonal to it, which we already called  $g_s$ , and which we define as

$$g_s \equiv \begin{cases} x_2 - x_1 & s = 1 \\ x_{s+1} - x_{s-1} & 1 < s < 144 \\ x_{144} - x_{143} & s = 144 \end{cases} \quad (8)$$

Thus the  $s$ -th milestone is the hyperplane  $H_s$  perpendicular to  $g_s$  and passing through  $x_s$ .

The resulting structures are then solvated with 248 TIP3 water molecules<sup>13</sup> in a box of volume  $(20 \text{ \AA})^3$ . The peptide structure is then fixed while the waters are allowed to relax for 200 ps. We used SHAKE<sup>14</sup> to fix water bond lengths and bond angles; the peptide bonds were unconstrained. The AMBER/OPLS<sup>15</sup> force field was used with periodic boundary conditions. The Lennard-Jones cutoff was 7 Å, and particle-mesh Ewald summation<sup>16</sup> of electrostatic forces was done. The real space cutoff distance was 8 Å. Initial velocities were sampled at a temperature of 303 K, and coupling to a heat bath of the same temperature was effected by scaling the velocities every 50 MD steps (scaling was performed only if the temperature deviates by more than 15 degrees from 303 K).

We determined  $K_s^\pm(t)$ , the first passage time distributions from milestone  $s$  in the direction from  $s$  to  $s+1$  ( $K_s^+(\tau)$ ), and in the direction  $s$  to  $s-1$  ( $K_s^-(\tau)$ ), by distributing molecular dynamics (MD) simulations across 154 CPUs with the program MOIL<sup>9</sup>. The trajectories were initiated at milestone  $s$  and were integrated until termination at milestone  $s \pm 1$ . The termination times were binned to generate  $K_s^\pm(t)$ .

The simulations are carried out in two phases. In the first phase, called “orthogonal equilibration” (OEQ), we prepare a canonical ensemble in each  $H_s$  by generating a 300-ps Verlet trajectory (0.5fs for a time step). The trajectory starts from  $x_s$  and is subject to

six linear constraints holding the peptide's center of mass at zero and fixing its overall orientation, and a seventh linear constraint confining the peptide to  $H_s$ . The corresponding equations of motion are <sup>17</sup>:

$$\begin{aligned}
 M \frac{d^2 x}{dt^2} &= -\frac{dU}{dx} - \sum_{i=1,\dots,7} \lambda_i \frac{d\sigma_i}{dx} \\
 \sigma_{1-3} &= \sum_n m_n \cdot r_n = 0 \quad r = (r_x, r_y, r_z) \\
 \sigma_{4-6} &= \sum_n m (r_n^0 \times r_n) = 0 \\
 \sigma_7 &= g_s \cdot (x - x_s) = 0
 \end{aligned} \tag{9}$$

The diagonal mass matrix is  $M$ , and  $m_n$  is the mass of particle  $n$ . The time step was  $h = 0.5$  fs and structures were sampled every 200 steps, so each milestone ensemble consisted of 3000 points. Since the constraints are linear, an analytical solution for one integration step of the constrained dynamics can be found, facilitating efficient and stable calculation (**Appendix B**).

Figure 3 shows typical equilibrium samplings taken from the M=11 simulation, and projected on two-dimensional  $\phi, \psi$  space:

Place Figure 3 Here

In the second phase, called “first passage” (FP), we remove the seventh linear constraint confining the peptide to a hyperplane. We integrate each member of the OEQ ensembles (1 fs timestep) until its trajectory reaches one of its two neighboring milestones. We keep the constraints on overall molecular orientations and translations since the milestones are expressed in Cartesian coordinates in an absolute frame, which are sensitive to overall rigid body shifts. The impact of the missing overall rotations on

the kinetics is small if the radius of gyration does not change significantly during the reaction (as discussed in<sup>17</sup>).

To determine when a trajectory reaches a neighbor, we use the following protocol. Let  $x(t)$  be the trajectory's position at time  $t$  after it leaves  $H_s$ , so that  $g_s \cdot (x(0) - x_s) = 0$ . Define the two inner products between the trajectory and its neighboring milestones,  $d^\pm(t) \equiv g_{s\pm 1} \cdot (x(t) - x_{s\pm 1})$ . Then we have

$$d^+(0) < 0 \quad \& \quad d^-(0) > 0 \quad (10)$$

When the trajectory finally reaches a neighbor, the corresponding inner product will change sign. So as soon as one of the two conditions

$$d^\pm(0)d^\pm(t^\pm) \leq 0 \quad (11)$$

holds, we terminate the simulation of that trajectory and take  $t^\pm$  to be its incubation time:  $t^+$  would contribute to the histogram  $K_s^+$ , and  $t^-$  would contribute to  $K_s^-$ . If the milestones are too far from each other additional complications in determining the crossing time may occur and are discussed in Appendix C.

To investigate how milestone predictions depend on  $M$ , we ran nine simulations using different subsets of the above set of 144 milestones. The first simulation, which used all 144 milestones, showed a free energy barrier at  $H_{25}$ , an  $\alpha$  well at  $H_{59}$ , and a relatively flat region corresponding to the  $\beta$  conformation extending up to  $\sim H_{140}$  (see

Figure 9). We decided to use these three hyperplanes in all subsets, so as to study the kinetics of the reaction of the  $\alpha \rightarrow \beta$  reaction in all nine simulations. The following table shows the milestones used:

Place Table 1 Here

The final milestone is not always  $H_{140}$  because we tried to space the milestones equally starting from milestone 1.

Two different boundary conditions are used to calculate the free energy (3), the MFPT (5) and  $P_s(t)$ . To compute the free energy we impose reflecting boundary conditions by ignoring trajectories that transition backwards from the first milestone and trajectories that transition forwards from the last milestone. To compute the rate of the  $\alpha \rightarrow \beta$  reaction (MFPT and  $P_s(t)$ ) we ignore incubation data at milestones 1 through 24, as well as incubation data coming from trajectories that transition from milestone 25 to one of the first 24 milestones. In keeping with (4) we also ignore incubation data at the final milestone. Thus, the  $\alpha \rightarrow \beta$  reaction is defined by the  $M = 3$  simulation: it is the reaction in which alanine dipeptide begins in the  $\alpha$  conformation and transitions to the end of the  $\beta$  conformation by increasing its  $\psi$  torsion, i.e., without crossing the barrier at  $H_{25}$ .

## VI. Results

As mentioned above, we used a per-milestone ensemble size of 3000. Figure 4 shows a sample incubation distribution, from the  $M = 7$  simulation:

Place Figure 4 Here

We can use the convergence of the time-averaged incubation time  $\langle \tau \rangle_s$  at each milestone  $s$  as a measure of the entire simulation's convergence. The following plot shows a sample of these average incubations, taken from the  $M = 11$  simulation:

Place Figure 5 Here.

The remaining plots look similar, and we conclude that roughly 500 points per milestone are required for convergence. Figure 6 shows a plot of the time dependence of the  $\alpha \rightarrow \beta$  reaction probability  $P_M(t)$  for five of the nine simulations, computed with the integral equation (1):

Place Figure 6 Here

Its legend also shows the predicted MFPTs in parentheses. As  $M$  decreases, both the curves and the rates tend to the true ( $M = 3$ ) kinetics.

In Figure 7 we present a semilog plot of the MFPTs:

Place Figure 7 Here

The MFPT agrees with the correct  $M = 3$  value up to  $M=19$ . Therefore we can use up to 19 milestones and save considerable computational resources as illustrated below. How to decide on the correct (and efficient) number of milestone to use? In the example of this manuscript we illustrate (see below) that as long as the typical transition time between milestones is longer than the velocity relaxation time, which is typically less than a picosecond, we obtain accurate estimate of the time evolution of the system. We define the milestone Average Incubation Time (AIT) as

$$\bar{\tau} \equiv \frac{1}{M} \sum_{i=1}^M \int_0^{\infty} \tau K_{s_i}(\tau) d\tau$$

where  $\{s_1, s_2, \dots, s_M\}$  are the milestone numbers used in a simulation. Table 2 summarizes the results of MFPT, and AIT for different number of milestones.

Insert Table 2 Here

The AIT of 19 hyperplanes is 373 femtoseconds (fs). To patch a complete reactive trajectory we need 19 trajectories of similar length:  $19 \times 373 = 7087$  fs. This should be compared with a straightforward molecular dynamics trajectory, which requires 64,000 fs. Since the cost of a trajectory is proportional to its length (or the number of force evaluations), the 19 milestones run is more efficient by a factor of 9. Similarly the 7 milestones require  $7 \times 3581 = 25067$  femtoseconds, a speed-up of 2.5. Interestingly, the 37 milestones run that is not diffusive (see below) still shows significant speedup for a serial calculation. The total time required for 37 milestones is  $37 \times 129 = 4773$  a factor of 13.4 speed-up.

The accuracy deteriorates if we use more than 19 milestones. We now give evidence that this deterioration is due to a violation of the fundamental milestone assumption (7). Consider the auto-correlation function of the time-derivative of the  $\psi$  torsion,  $\langle \dot{\psi}(0)\dot{\psi}(t) \rangle$ . We computed this quantity by running and averaging many short simulations of the peptide. The result, shown in Figure 8, reveals vibrations with periods  $\sim 25$  fs and  $\sim 200$  fs, modulated by an exponential decay with time constant  $\sim 300$  fs:

Insert Figure 8 Here

This exponential damping reflects the system's memory loss due to collisions with the water, so we have the lower limit  $\tau^\perp \geq 300$  fs, where we use  $\tau^\perp$  to denote the relaxation time of all degrees of freedom orthogonal to the reaction coordinate. If we compare this time scale to the AIT, then for 19 milestones we have 373 femtoseconds which is still

larger than the above number. On the other hand the AIT for 37 milestones is significantly shorter (129 femtoseconds) suggesting that the rate calculation will not be accurate since the system did not have sufficient time to lose memory, as is indeed the case. This observation also suggests that different dynamical models with no significant memory of velocity (e.g. Brownian dynamics) should be accurate in the milestoneing model. Nevertheless, Brownian dynamics has the caveat of an ill-defined microscopic parameter for the study of dynamics (the friction), so the rapid memory loss of this model is coming at the cost of uncertainty in using this dynamics for microscopic systems.

Accepting the above explanation for the breakdown of the milestoneing approximation for a large number of milestones, we offer a conjecture for why the inaccurate times are consistently larger than the exact MFPT (figure 7 and table 2). If the velocity is not fully relaxed, (for  $M \geq 37$ ) the exact trajectory will tend to continue in the same direction. Starting at  $H_s$ , to compute approximate fragment of the whole trajectory we randomize the direction of the velocity. The trajectory is as likely to go backward as to go forward. Clearly the more direct (exact) trajectory will be shorter, suggesting why calculations with milestones that are too close are overestimating the MFPT.

We can measure coordinate relaxation and condition (7) also by comparing the moments of the equilibrium and first passage time distributions. However, we found that the moments of the torsions of the generated ensembles are noisy, making it difficult to identify the breakdown of the Milestoning approximation according to this measure. While some deviations from equilibrium (Eq. (7)) were detected, no systematic correlations with the errors of the rate were observed. We suggest, then, that the rate

diverges at large  $M$  not because the peptide's position is insufficiently relaxed, but because the peptide has not "forgotten" its velocity.

Finally, we show the free energy profiles (3) of the nine simulations in Figure 9:

Place Figure 9 Here

As mentioned earlier, we used the  $M = 144$  curve to identify milestones  $H_{25}$ ,  $H_{59}$ , and  $H_{140}$ , for the reaction of interest ( $\alpha \rightarrow \beta$ ) for the remaining eight simulations. Notice that the features of the profile become increasingly less pronounced with fewer milestones. The simulations that violate (7) ( $M > 19$ ) clearly distort the profile, whereas the other simulations give free energies that are more accurate. The essential features of the true free energy are captured even by simulations having so few milestones as to offer only a caricature of the profile. When we shift their minima to zero the five curves nearly overlap. A recent paper outlined the use of a Markov model to compute a free energy surface<sup>18</sup> based on a similar idea, which brings us to the comparison between our non-Markovian and a Markovian approach.

## **VII. Markov model and a single exponential relaxation**

The main result of this work is that Milestoning predicts accurate rates with a reasonably large number of milestones (19 or smaller), which allows for significant



$$\dot{P}_s(t) = \sum_{s'} W_{ss'} P_{s'}(t)$$

The MFPT is obtained from Eq. 5 as explained in **Appendix A** and is identical to the MFPT obtained from the non-Markovian equations. Note however that the Markovian solution does not imply a rate law or single-exponential kinetics. This limit would have been obtained if the smallest nonzero eigenvalue  $\gamma_1$  of  $W$  dominates the relaxation. The third column of Table 2 shows the “exponential” times for all the simulations. Comparing them to the milestoning predictions (second column), it is evident that the single exponent model failed; the process is clearly non exponential.

The deviation of the full Markovian model from the non-Markovian approach can be checked in Table 2 (the standard deviation for the first passage time distribution

$\sqrt{\langle\langle\tau^2\rangle\rangle - \langle\langle\tau\rangle\rangle^2}$  for the Markovian and the Milestoning model). We have also examined complete time courses of the two models. The differences between the Markovian and non-Markovian models are small.

## VIII. Discussion

In this section we compare the Milestoning approach to other related methods and put it in perspective. We comment that milestoning is a technique to accelerate molecular dynamics calculations with a given force field. The quality of the results will depend

on the potential that we use, which in the present case is the combination of AMBER<sup>20</sup> and OPLS<sup>15</sup> as implemented in the molecular dynamics package MOIL<sup>9</sup>.

**Avoiding reaction coordinate:** The study of long time processes in computational modeling has a long history. The broadly used approaches are based on a reaction coordinate (or order parameter) to guide the dynamics. Nevertheless, it would be nice (and less restrictive) if the reaction coordinate was eliminated, and rates could be computed directly. In principle, if MD trajectories could be run for long enough this would have been the most straightforward solution. While a few long time trajectories have been reported this is insufficient to compute kinetics<sup>21,22</sup>. Another interesting technique for exponential processes is the use of short time trajectories to interpolate to longer times<sup>23,24</sup>. Finally, the method of Transition Path Sampling (TPS)<sup>25</sup> is a recent option for calculations of rates without the use of a reaction coordinate. TPS is clearly a promising and interesting method, and advances were made in the efficient sampling of these rare trajectories for alanine dipeptide<sup>26</sup>. However, some restrictions remain: (i) the system is assumed to be in equilibrium, (ii) the trajectory times before converging to reactants or products must be short, and comparable to the times of usual MD, (iii) the kinetics must be exponential. Hence, while these approaches are useful for a number of cases, there remains a significant number of problems that require different tools. Milestoning can extend the time scale of individual trajectories, and compute non-exponential non-equilibrium kinetics (as discussed below) and is therefore complementary to other techniques. On the other hands Milestoning requires a prior

definition of a reaction coordinate or dividing hypersurfaces, which is a disadvantage compared to the above approaches.

**Generating reaction coordinates:** Useful algorithms to compute minimum energy paths<sup>10, 12, 27</sup> and free energy paths<sup>28</sup> are available. Another way of defining reaction coordinates for the Milestoning approach is the use of SDEL trajectories<sup>29</sup>. These trajectories have their high frequency modes filtered. Overall the practical determination of the reaction coordinate appropriate for the calculation of kinetics remains heuristic. In Milestoning we seek rate calculations that are reasonably robust (or insensitive) to a reasonable choice of the hypersurfaces, monitoring the rate as a function of the distance between the milestones is a way of assessing this sensitivity.

It is important to emphasize that the use of a hypersurface for a milestone does not mean complete sampling over the hypersurface's  $3N-1$  dimensions ( $N$  is the number of atoms). Rather, the picture we have in mind is of sampling in the neighborhood of the curvilinear path which is a reaction coordinate or a filtered trajectory. It is difficult to expect that a time scale separation will be found for motions along the reaction coordinate and the rest of the “universe”. It is more reasonable to expect local equilibrium in the neighborhood of the reaction coordinate that we consider.

Despite the difficulties mentioned above reaction coordinates are useful and are a widely used tool for computing long time dynamics. Transition State Theory<sup>7</sup> is a clear success story (for a review see<sup>6</sup>), and other further developments for non-exponential processes broaden the scope of the same approach. Even the recent TPS<sup>25</sup>, designed originally to be “reaction-coordinate free”, was adjusted to produce two related techniques TIS and

PPTIS that require reaction coordinates. PPTIS is similar conceptually to Milestoning and is discussed in details below.

**Computing rates from reaction coordinate.** The estimate of a time scale from a reaction coordinate goes back to the transition state theory for systems with a significant (free) energy barrier. A recent elegant theory builds on a free energy path calculation to compute the rate<sup>28,30</sup>. However, the use of a reaction coordinate is not limited to cases in which it is necessary to identify a transition state. In TIS<sup>31</sup> trajectories are followed in phase space between many interfaces to compute exponential kinetics (high free energy barriers). The reaction coordinate is used to monitor all trajectories and to bootstrap phase space points that look promising. This is similar to the third enhancement factor for milestoning trajectories that we mentioned in section **III. The computational gain of Milestoning** (exponential bootstrapping). The enhancements are for calculations of rate constants at equilibrium and are especially effective for large barriers.

The idea of reducing the dynamics of a complex molecular system to a Markov model with a discrete set of states (potentially along reaction coordinates) was explored by a number of investigators<sup>2,32,33</sup> demonstrating considerable promise. The present paper is similar in concept to the above studies, but different in implementation. From the different techniques mentioned the method of PPTIS is the most similar to Milestoning. The similarities and differences are discussed below.

Another methodology that builds on the transition path sampling concept is the PPTIS<sup>34</sup> in which the memory of the velocity is assumed lost. Again equilibrium and a large barrier are assumed (although the barrier is wide and diffusive). The computational

enhancements in this case are diffusive enhancement and exponential bootstrapping; two of the enhancements for milestones that were discussed in section III. From the accuracy viewpoint the two techniques use different approximations: PPTIS assumes loss of memory in time, and Milestoning loss of spatial memory. Which of the two is more accurate is likely to depend on the system at hand. It is possible to merge the two techniques (e.g. introducing spatial memory to Milestoning, or non-Markovian analysis to PPTIS). There is, however, a price to pay in terms of efficiency and applicability of such a merger.

PPTIS is not as parallelizable as Milestoning since it computes interface ensembles sequentially, which can make it less efficient. PPTIS also aims at computing the kinetics of two-state exponential processes in equilibrium. Milestoning allows for the study of non-exponential kinetics and non-equilibrium progression along the reaction coordinate (we assume equilibrium for all other degrees of freedom). In the present paper we illustrated non-exponential kinetics for a widely used model system in biophysics. The non-Markovian contributions were small however.

**Computing free energies:** While our chief interest is in kinetics, a side product of our analysis is a rapid way of computing free energies. The results and the method proposed in this paper are similar in spirit to a recent PPTIS<sup>18</sup> paper that we became aware of after this study was completed. Similarly to the technique presented in this paper, the equilibrium vector is extracted from the kinetic data. In PPTIS it is done for a Markovian model; we have done it for the non-Markovian version. The two should agree

asymptotically, i.e., in the long-time limit. The spatial distribution in the milestone is in equilibrium at the long time limit making the proposed procedure exact in both PPTIS and milestoning.

**Non-Markovian model:** In milestoning we use a non-Markovian model to compute the kinetics of a reduced one-dimensional model generated from straightforward atomically detailed trajectories. The cost is dominated by the calculation of the first passage time distributions  $K_s^\pm(\tau)$  (as opposed to solving the model). The negligible cost of solving the final model (the integral equations) suggests that we should use the best approach available to solve it since the overall computational cost will not be affected. The non-Markovian model is in principle more accurate than the more widely used Markovian model. In a Markov model the incubation kinetics (the kinetics of hopping from one milestone to the next) is a single exponential. It is hard to predict when this assumption is valid, however, and there is no reason to assume that it must be correct, and as argued above the non-Markovian calculations do not change the overall cost of the study. They are expected to be significant when the incubation process (transition between nearby milestones) is also non-exponential, as in the “mechanicity” scenario that Kolomeisky and Fisher discuss in the context of motor proteins<sup>35</sup>. In the present example this was true only for 3 milestones that had significant waiting time. However, with an eye to more complex systems with significantly larger incubation times we anticipate that such non exponential behavior will be observed in future studies. A related non-Markovian model to examine transition state re-crossing was introduced by Straub and co-workers<sup>36</sup>. In large scale simulations that we are pursuing now (Ron Elber, The R to T transition in

Scapharca Hemoglobin, work in progress) incubation times of individual milestones exceed tens of picoseconds. This suggests that the waiting or diffusion times and non-exponential behavior will be more significant for complex (mechanistic) systems.

**Extending variables of Milestoning** A more detailed description of the system that is likely to add significantly to the efficiency of milestoning is the direction of the velocity. I.e. instead of a state being defined by a single index  $s$ , it is possible to define it as  $s+$  (velocity in forward direction upon entry), or  $s-$  (velocity in the negative direction. This extension may resolve the problem of velocity memory mentioned in the **Results** section since  $s+$  to  $s+$  will be a more favorable transition for close milestones compared to the transitions  $s+$  to  $s-$  (currently the two different types of transitions not differentiable)

**Extending milestoning to more than one dimension:** Another limitation of the present approach is the use of a single reaction coordinate (or a sequence of hypersurfaces leading from reactants to products), which can be interpreted as the existence of a single slow variable in the system. If there are other slow variables that are coupled to the current reaction coordinate and that lead to alternative reaction paths (and therefore a different set of hypersurfaces) then an extension of the formulation is required. There are two parts to this extension. The first part is the replacement of nearest-neighbor milestone connectivity ( $i \leftrightarrow i \pm 1$ ) by a more general transition ( $i \leftrightarrow j$ ). This is straightforward since the formulation remains exactly the same except for the replacement of the tri-diagonal incubation matrix  $K$  by a general matrix, as was pointed out in the paper by Shalloway and Faradjian<sup>5</sup> (see also Appendix A). The second part is more complex and is the reduction scheme from the molecular simulation to the abstract space of milestones. It

is possible to create “boxes” in  $N-k$  dimensions (for  $k$  slow variables) that will be bound by hypersurfaces on which trajectories will be terminated to construct  $n \cdot (l)^k$  first passage distributions (where  $l$  is the number of milestones per slow degrees of freedom, and  $n$  is the number of nearby milestones). While such elaborate models can be built, it is clear that the computational complexity increases exponentially with the number of slow degrees of freedom. In practice 3 degrees of freedom is probably the limit of what we can do with current technology.

## **Acknowledgments**

We thank Eric van den Eijnden for many useful discussions, and John Straub and Eric Darve for pointing us to important references. This research was supported by an NIH grant GM59796 to RE. We acknowledge NIH grant RR020889 for the purchase of a computer cluster on which these calculations were performed.

## Appendix A:

### Efficient Calculation of the overall first passage time and free energy

If the prime focus of the calculation is to get an exact and complete time course of the reaction then solving the integral equation as described in<sup>3</sup> is necessary. However, if we are interested only in the equilibrium distribution,  $P_s^{eq}$ , and in the overall mean first passage time  $\langle\langle\tau\rangle\rangle$ , a shortcut is possible. Let us rewrite the  $2M$  equations (1) as just two equations, by defining the matrix

$$\begin{aligned} K_{s,s\pm 1}(\tau) &\equiv K_{s\pm 1}^{\mp}(\tau) \\ K_{s,s}(\tau) &\equiv -K_s^+(\tau) - K_s^-(\tau) \\ K_{s,s'}(\tau) &\equiv 0, \quad |s - s'| > 1 \end{aligned}$$

$s = 1, 2, \dots, M$ . (This matrix  $K$  is not to be confused with the function  $K_s = K_s^+ + K_s^-$ , which is not used in this Appendix.) Let  $K_D$  be the matrix of diagonal elements of  $K$ , and let  $K_d \equiv K - K_D$  be the matrix of off-diagonal elements. Note the following properties:

$$\vec{1} \cdot K = \vec{0} \quad (\text{zero column-sum}) \quad (\text{A1a})$$

$$\int_0^{\infty} K_D(\tau) d\tau = -I \quad (\text{A1b})$$

where  $\vec{1}$  is the  $M$ -vector of ones,  $\vec{0}$  is the  $M$ -vector of zeros, and  $I$  is the  $M \times M$  identity matrix. For simplicity we use angle-brackets to denote an infinite integration, so (A1b) is written conveniently as  $\langle K_D \rangle = -I$ . With the matrix  $K$  we can write (1) more simply as

$$\begin{aligned}\vec{P}(t) &= \int_0^t \left[ I + \int_0^{t-t'} K_D(\tau) d\tau \right] \vec{Q}(t') dt' \\ \vec{Q}(t) &= \vec{\rho} \delta(t-0^+) + \int_0^t K_a(t-t') \vec{Q}(t') dt'\end{aligned}\tag{A2}$$

where  $P_s$  is the  $s$ -th element of the vector  $\vec{P}$ , etc.

Now because the Laplace transform of a constant  $c$  is  $\tilde{c}(\alpha) = c/\alpha$ , we can express the equilibrium value of any time-dependent quantity  $f$  as

$$f^{eq} \equiv \lim_{t \rightarrow \infty} f(t) = \lim_{\alpha \rightarrow 0} \alpha \tilde{f}(\alpha)$$

Thus to obtain the equilibrium distribution (2), we need the Laplace transform of (A2):

$$\begin{aligned}\tilde{\vec{P}}(\alpha) &= \left[ \frac{I + \tilde{K}_D(\alpha)}{\alpha} \right] \tilde{\vec{Q}}(\alpha) \\ \tilde{\vec{Q}}(\alpha) &= \vec{\rho} + \tilde{K}_a(\alpha) \tilde{\vec{Q}}(\alpha)\end{aligned}\tag{A3}$$

Multiplying the first of these two equations by  $\alpha$ , and using the Laplace expansion for small  $\alpha$

$$\begin{aligned}
\tilde{f}(\alpha) &= \int_0^{\infty} e^{-\alpha t} f(t) dt \\
&= \int_0^{\infty} f(t) dt - \alpha \int_0^{\infty} t f(t) dt + \frac{\alpha^2}{2} \int_0^{\infty} t^2 f(t) dt + O(\alpha^3) \\
&= \langle f \rangle - \alpha \langle t f \rangle + \frac{\alpha^2}{2} \langle t^2 f \rangle + O(\alpha^3)
\end{aligned} \tag{A4}$$

(readers concerned about analyticity at  $\alpha = 0$  are referred to<sup>5</sup>), we have

$$\begin{aligned}
\vec{P}^{eq} &= \left[ \lim_{\alpha \rightarrow 0} \frac{I + \tilde{K}_D(\alpha)}{\alpha} \right] \vec{Q}^{eq} \\
&= \left[ \lim_{\alpha \rightarrow 0} \frac{I + \langle K_D \rangle - \alpha \langle \tau K_D \rangle + O(\alpha^2)}{\alpha} \right] \vec{Q}^{eq} \\
&= -\langle \tau K_D \rangle \vec{Q}^{eq}
\end{aligned} \tag{A5}$$

where we have used (A1b). So the only quantities needed to compute  $P^{eq}$  are the average incubation times, and  $\vec{Q}^{eq}$ . We obtain  $\vec{Q}^{eq}$  by multiplying the second of equations (A3) by  $\alpha$  and taking the limit:

$$\begin{aligned}
\vec{Q}^{eq} &= \lim_{\alpha \rightarrow 0} \alpha \tilde{Q}(\alpha) \\
&= \lim_{\alpha \rightarrow 0} \left[ \alpha \vec{\rho} + \tilde{K}_d(\alpha) \alpha \tilde{Q}(\alpha) \right] \\
&= \langle K_d \rangle \vec{Q}^{eq}
\end{aligned}$$

Subtracting  $\vec{Q}^{eq}$  from both sides of this equation and using (A1b) again, we see that  $\vec{Q}^{eq}$  is just the zero-eigenvector of the integrated transition matrix:

$$\langle K \rangle \vec{Q}^{eq} = \vec{0}$$

So (A5) becomes

$$\langle K \rangle \langle \tau K_D \rangle^{-1} \vec{P}^{eq} = 0 \quad (\text{A5b})$$

For the MFPT, we select an initial milestone  $i$  and set  $\rho = \hat{\varepsilon}_i$  ( $\hat{\varepsilon}_s$  denotes the unit vector at milestone  $s$ ), and we impose the absorbing boundary conditions (4) on the final state  $f$ , so

$$K \cdot \hat{\varepsilon}_f = \vec{0} \quad (\text{A6})$$

This condition modifies (A1b) to

$$\langle K_D \rangle = -I + \hat{\varepsilon}_f \otimes \hat{\varepsilon}_f \quad (\otimes = \text{outer product}) \quad (\text{A7})$$

Now from (A4) we see that

$$\langle \langle \tau \rangle \rangle = -d\tilde{\varphi}/d\alpha|_{\alpha=0} \quad (\text{A8})$$

so let us compute the Laplace transform of  $\varphi(t)$ . The boundary condition (A6) implies that  $\varphi(t) = Q_f(t) = \hat{\varepsilon}_f \cdot \vec{Q}(t)$ . It also implies that the second of equations (A3) can be inverted and solved even at  $\alpha = 0$ :

$$\vec{Q}(\alpha) = [I - \tilde{K}_d(\alpha)]^{-1} \cdot \hat{\varepsilon}_i$$

Thus,

$$\tilde{\varphi}(\alpha) = \hat{\varepsilon}_f \cdot [I - \tilde{K}_d(\alpha)]^{-1} \cdot \hat{\varepsilon}_i \quad (\text{A9})$$

and (A8) gives

$$\langle\langle\tau\rangle\rangle = \hat{\varepsilon}_f \cdot [I - \langle K_d \rangle]^{-1} \langle \tau K_d \rangle [I - \langle K_d \rangle]^{-1} \cdot \hat{\varepsilon}_i \quad (\text{A10})$$

So again the only required quantities are the mean incubations  $\int_0^\infty \tau K_{s\pm 1}^\mp(\tau) d\tau$  and the

splitting probabilities  $\int_0^\infty K_{s\pm 1}^\mp(\tau) d\tau$ . We can simplify (A10) by noting that if we add  $\langle K_d \rangle$

to both sides of (A7) and take the inner product with  $\bar{1}$ , we obtain

$$\bar{1} \cdot [I - \langle K_d \rangle] = \bar{1} \cdot [-\langle K \rangle + \hat{\varepsilon}_f \otimes \hat{\varepsilon}_f] = \hat{\varepsilon}_f, \text{ which implies } \hat{\varepsilon}_f \cdot [I - \langle K_d \rangle]^{-1} = \bar{1}. \text{ Using this}$$

result in (A10), we write

$$\begin{aligned} \langle\langle\tau\rangle\rangle &= \bar{1} \cdot \langle \tau K_d \rangle [I - \langle K_d \rangle]^{-1} \cdot \hat{\varepsilon}_i \\ &= \bar{1} \cdot \langle \tau K_d \rangle [-\langle K \rangle + \hat{\varepsilon}_f \otimes \hat{\varepsilon}_f]^{-1} \cdot \hat{\varepsilon}_i \end{aligned} \quad (\text{A10b})$$

We also note from (A4) that the second moment is  $\langle\langle\tau^2\rangle\rangle = d^2\tilde{\varphi}/d\alpha^2|_{\alpha=0}$ , yielding:

$$\langle\langle\tau^2\rangle\rangle = \bar{1} \cdot \left[ \langle \tau^2 \bar{K}_D \rangle - 2 \langle \tau \bar{K}_D \rangle \langle \bar{K} \rangle^{-1} \langle \tau \bar{K}_d \rangle \right] \cdot \langle \bar{K} \rangle^{-1} \cdot \hat{\varepsilon}_i$$

where  $\bar{X} \equiv X + \hat{\varepsilon}_f \otimes \hat{\varepsilon}_f$ .

Finally, we note that the usual way to a ‘‘Markovization’’ of this non-Markovian model is through the *generalized master equation* (See Reference for 4 and references therein):

$$\frac{d\vec{P}}{dt} = \vec{\rho}\delta(t-0^+) + \int_0^t \Gamma(t-t')\vec{P}(t') dt' \quad (\text{A11})$$

where the rate matrix  $\Gamma$  is related to  $K$  via their Laplace transforms:

$$\tilde{\Gamma}(\alpha) = \alpha\tilde{K}(\alpha)[I + \tilde{K}_D(\alpha)]^{-1} \quad (\text{A12})$$

Markovian dynamics are then recovered simply by defining the Markov rate matrix  $W$  to be an integration over  $\Gamma$ ,

$$W \equiv \langle \Gamma \rangle = \tilde{\Gamma}(0) = -\langle K \rangle \langle \tau K_D \rangle^{-1} \quad (\text{A13})$$

(where again we have used (A4) with (A1b)). Equation (A13) is the matrix equivalent of equation (12) of Section VII. Substituting  $\Gamma(\tau) = W\delta(\tau - 0^+)$  into (A11) then gives the master equation  $d\vec{P}/dt = \vec{\rho}\delta(t - 0^+) + W\vec{P}$ . This approach preserves the equilibrium distribution, i.e., (A5b) is nothing more than the *Markovian* equilibrium condition  $WP^{eq} = 0$ . It can also be shown that the Markovization preserves the MFPT (A10b) too, but does not preserve any higher moments of the FPTD [Shalloway & Faradjian, 2006; Haenggi & Talkner, 1983].

## Appendix B: Dynamics constrained to a milestone

Following equations

$$\begin{aligned}
 M \frac{d^2 x}{dt^2} &= -\frac{dU}{dx} - \sum_{i=1,\dots,7} \lambda_i \frac{d\sigma_i}{dx} \\
 \sigma_{1-3} &= \sum_n m_n \cdot r_n = 0 \quad r = (r_x, r_y, r_z) \\
 \sigma_{3-6} &= \sum_n m(r_n^0 \times r_n) = 0 \\
 \sigma_7 &= g_s \cdot (x - x_s) = 0
 \end{aligned} \tag{9}$$

we write an explicit algorithm for the solution of the linearly constrained equations of motion. The constrained equations of motions are used in the two phases of the Milestoning simulation. In the first (equation (9) – the so called OEQ phase) we sample configurations at the hypersurface  $H_s$ , which we approximate as a hyperplane in Cartesian space ( $\sigma_7$ ), in addition to constraints on overall translation and rotation. Linear constraints are also used in the second part of the calculation (so called FP phase) in which the hypersurface constraint is removed but the constraints on translation and rotations remains for easy evaluation of the trajectory termination. A useful preparatory step is the orthonormalization of the seven constraints. We define

$$\bar{\sigma}_i = \sum_{j=1,\dots,7} a_{ij} \sigma_j \quad i = 1, \dots, 7 \tag{B1}$$

Such that  $\left(\frac{d\bar{\sigma}_i}{dx}\right)^t \frac{\Delta t^2}{2} M^{-1} \left(\frac{d\bar{\sigma}_j}{dx}\right) = \delta_{ij}$  ( $M$  is the diagonal mass matrix, and the

coefficients  $a_{ij}$  are determined by a Gram Smit procedure<sup>37</sup> Note that since the

constraints are linear the orthonormalization needs to be done only once at the beginning of the calculations. We denote  $\frac{d\bar{\sigma}_i}{dx} \equiv \bar{q}_i$ , and the transformed constraints are now written as  $\bar{\sigma}_i = q_i^t \cdot x - c_i$ , where the  $c_i$  are constants. The same space spanned by the original constraints is covered by the new constraints. Since the Lagrange multipliers  $\lambda_i$  are still undetermined, we can redefine them in terms of the new constraints, to have

$$\begin{aligned} M \frac{d^2 x}{dt^2} &= -\frac{dU}{dx} - \sum_{i=1, \dots, 7} \bar{\lambda}_i \bar{q}_i \\ \bar{\sigma}_i &= 0 \quad i=1, \dots, 7 \end{aligned} \quad (\text{B2})$$

The normal Verlet algorithm can be used in conjunction with the explicit solution of the constraints as discussed below ( $v_k$  is the velocity at time slice  $k$ ).

$$\begin{aligned} x_k &= x_{k-1} + v_{k-1} \Delta t - \frac{\Delta t^2}{2M} \left( \frac{dU}{dx_k} + \sum_i \bar{\lambda}_{ik} \bar{q}_i \right) \\ \bar{\sigma}_i &= q_i^t \cdot x_k - c_i = 0 \end{aligned} \quad (\text{B3})$$

Substituting  $x_k$  in the equation for  $\bar{\sigma}_i$  we have (for all  $i$ )

$$\begin{aligned} \bar{\sigma}_i &= q_i^t \cdot \left( x_{k-1} + v_{k-1} \Delta t - \frac{\Delta t^2}{2} M^{-1} \frac{dU}{dx_k} \right) + \frac{\Delta t^2}{2} q_i^t \cdot M^{-1} \sum_j \bar{\lambda}_{jk} q_j - c_i = 0 \\ \bar{\lambda}_{ik} &= \left[ c_i - q_i^t \cdot \left( x_{k-1} + v_{k-1} \Delta t - \frac{\Delta t^2}{2} M^{-1} \frac{dU}{dx_k} \right) \right] \end{aligned} \quad (\text{B4})$$

We have used the orthogonality of the gradients of the constraints. The above equation is an exact analytical solution for the Lagrange multipliers (and the new coordinate set) for a single time step. A similar solution follows for the velocities

$$\begin{aligned}
v_k &= v_{k-1} - \frac{\Delta t^2}{2M} \left( \frac{dU}{dx_{k-1}} + \sum_i \bar{\xi}_{ik-1} \cdot q_i + \frac{dU}{dx_k} + \sum_i \bar{\xi}_{ik} \cdot q_i \right) \\
&= v_{k-1} - \frac{\Delta t^2}{2M} \left( \frac{dU}{dx_{k-1}} + \frac{dU}{dx_k} + \sum_i \bar{\eta}_{ik} \cdot q_i \right)
\end{aligned} \tag{B5}$$

Since both Lagrange multipliers multiply the same constant vector,  $q_i$ , we can merge them into a single unknown parameter. We also have

$$\begin{aligned}
\sigma_{vi} &= q_i^t \cdot v_k = 0 \quad \forall i \\
q_i^t \cdot \left( v_{k-1} - \frac{\Delta t^2}{2} M^{-1} \left( \frac{dU}{dx_{k-1}} + \frac{dU}{dx_k} + \sum_j \bar{\eta}_{jk} \cdot q_j \right) \right) &= 0 \\
\bar{\eta}_{ik} &= \frac{2M}{\Delta t^2} q_i^t \cdot \left( v_{k-1} - \frac{\Delta t^2}{2} M^{-1} \left( \frac{dU}{dx_{k-1}} + \frac{dU}{dx_k} \right) \right)
\end{aligned} \tag{B6}$$

This completes the closed expression for the algorithm.

## Appendix C: Termination of milestone trajectories

Note that if the milestone structures  $x_s$  are spaced so widely apart that the  $g_s$  become bad approximations to the local reaction path gradient, then condition (10) may not hold and the protocol may break down. The following scenarios illustrate the problem. Suppose our system has a two-dimensional configuration space and a free energy that tends to confine it to the unit circle. Suppose we place four milestones on the RC, as in panel (a) of Figure 10, and suppose that a trajectory begun at milestone 2 reaches milestone 3 before it reaches milestone 1:

Place Figure 10 Here

Throughout this time,  $d^+ = g_3 \cdot (x(t) - x_3)$  remains negative and  $d^- = g_1 \cdot (x(t) - x_1)$  remains positive, as expected. However, as is clear from figure 10, when the trajectory finally reaches milestone 3,  $d^+$  and  $d^-$  flip signs *simultaneously*, so the protocol cannot decide which milestone was reached. Similarly, suppose we use three milestones, as in panel (b), and suppose again that a trajectory begins at milestone 2. If it diffuses far enough toward milestone 3, it will cross  $H_1$  first and  $d^-$  will flip sign *before*  $d^+$  does. So here the protocol can decide, but it decides wrongly.

The solution to these problems is to use as many structures  $x_s$  as it takes to define a smooth reaction path gradient  $g_s$ , and to amend the protocol as follows. See panel (c). As in panel (a), we are still interested in using only  $H_1$ ,  $H_3$ ,  $H_5$ , and  $H_7$  as milestones, so we still say  $M = 4$ . But we use (11) to track a trajectory as it crosses the intermediate hyperplanes  $H_2$ ,  $H_4$ ,  $H_6$ , and  $H_8$ . We terminate a trajectory begun at  $H_3$  only when it

reaches  $H_1$  or  $H_5$ , and we do not decide that it has reached  $H_5$  unless it has already reached  $H_4$ ; similarly, we do not decide that it has reached  $H_1$  unless it has already reached  $H_2$ ; and so on. With enough intermediate hyperplanes,  $g_s$  will be smooth, (10) will hold, and (11) will give correct incubation times.

## References

- <sup>1</sup> B. Leimkuhler, C. Chipot, R. Elber, et al., *New Algorithms for Macromolecular Simulation* (Springer, Berlin, 2006).
- <sup>2</sup> G. Jayachandran, V. Vishal, and V. S. Pande, *Journal of Chemical Physics* **124**, 164092 (2006).
- <sup>3</sup> K. Faradjian A. and R. Elber, *J. Chem. Phys.* **120**, 10880 (2004).
- <sup>4</sup> A. K. Faradjian and R. Elber, *J. Chem. Phys.* **120**, 10880 (2004).
- <sup>5</sup> D. Shalloway and A. K. Faradjian, *Journal of Chemical Physics* **124**, 054112 (2006).
- <sup>6</sup> D. G. Truhlar, B. C. Garrett, and S. J. Klippenstein, *Journal of Physical Chemistry* **100**, 12771 (1996).
- <sup>7</sup> H. Eyring, *Journal of chemical physics* **3**, 107 (1935).
- <sup>8</sup> B. Ensing, M. De Vivo, Z. W. Liu, et al., *Accounts of Chemical Research* **39**, 73 (2006).
- <sup>9</sup> R. Elber, A. Roitberg, C. Simmerling, et al., *Computer Physics Communications* **91**, 159 (1995).
- <sup>10</sup> A. Ulitsky and R. Elber, *Journal of Chemical Physics* **92**, 1510 (1990).
- <sup>11</sup> H. Jonsson, G. Mills, and K. W. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne, G. Ciccotti and D. F. Coker (World Scientific, Singapore, 1998), p. 385.
- <sup>12</sup> E. Weinan, R. Weiqing, and E. Vanden-Eijnden, *Physical Review B* **66**, 52301 (2002).

- 13 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, et al., *Journal of Chemical Physics* **79**, 926 (1983).
- 14 J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, *Journal of Computational Physics* **23**, 327 (1977).
- 15 W. L. Jorgensen and J. Tirado-Rives, *Journal of the American Chemical Society* **110**, 1666 (1988).
- 16 T. A. Darden, A. Toukmaji, and L. G. Pedersen, *Journal de Chimie Physique et de Physico-Chimie Biologique* **94**, 1346 (1997).
- 17 R. Elber, *Journal of Chemical Physics* **93**, 4312 (1990).
- 18 D. Moroni, ten Wolde P.R., and P. G. Bolhuis, *Physical Review E* **71**, 056709 (2005).
- 19 P. Haenggi and P. Talkner, *Physical Review Letters* **51**, 2242 (1983).
- 20 S. J. Weiner, P. A. Kollman, D. A. Case, et al., *Journal of the American Chemical Society* **106**, 765 (1984).
- 21 B. L. de Groot and H. Grubmuller, . *Science* **294**, 2353 (2001).
- 22 Y. Duan and P. A. Kollman, *Science* **282**, 740 (1998).
- 23 V. S. Pande, I. Baker, J. Chapman, et al., *Biopolymers* **68**, 91 (2003).
- 24 A. F. Voter, F. Montalenti, and T. C. Germann, *Annual Review of Materials Research* **32**, 321 (2002).
- 25 C. Dellago, P. G. Bolhuis, and P. L. Geissler, *Advances in chemical physics* **123**, 1 (2002).
- 26 A. Ma and A. Dinner, *Journal of Physical Chemistry B* **109**, 6769 (2005).

- 27 H. Jonsson, G. Mills, and K. W. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne, G. Ciccotti and D. F. Coker (World Scientific, Singapore, 1998), p. 385.
- 28 W. Ren, E. Vanden-Eijnden, and P. Maragakis, *Journal of Chemical physics* **123**, 134109 (2005).
- 29 R. Elber, A. Ghosh, and A. Cardenas, *Accounts of Chemical Research* **35**, 396 (2002).
- 30 E. Weinan, W. Ren, and E. Vanden-Eijnden, *Chemical Physics Letters* **413**, 242 (2005).
- 31 T. S. van Erp, D. Moroni, and P. G. Bolhuis, *Journal of Chemical Physics* **118**, 7762 (2003).
- 32 G. Hummer and I. G. Kevrekidis, *Journal of Chemical Physics* **118**, 10762 (2003).
- 33 Huber G.A. and S. Kim, *Biophysical Journal* **70**, 97 (1996).
- 34 D. Moroni, Bolhuis P.G., and T. S. van Erp, *Journal of Chemical Physics* **120**, 4055 (2004).
- 35 A. B. Kolomeisky and M. E. Fisher, *Journal of Chemical Physics* **113**, 10867 (2000).
- 36 J. Straub, D. Hsu, and B. Berne, *Journal of Physical Chemistry* **89**, 5188 (1985).
- 37 G. Arfken, *Mathematical methods for physicists* (Academic press, San Diego, 1985).

# FIGURES

Figure 1

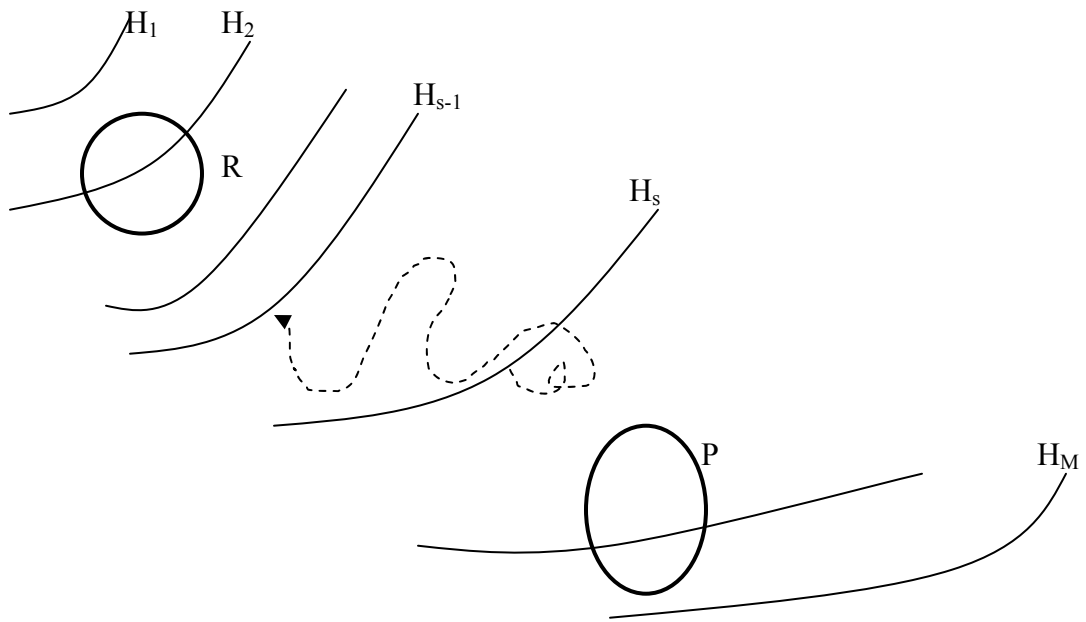
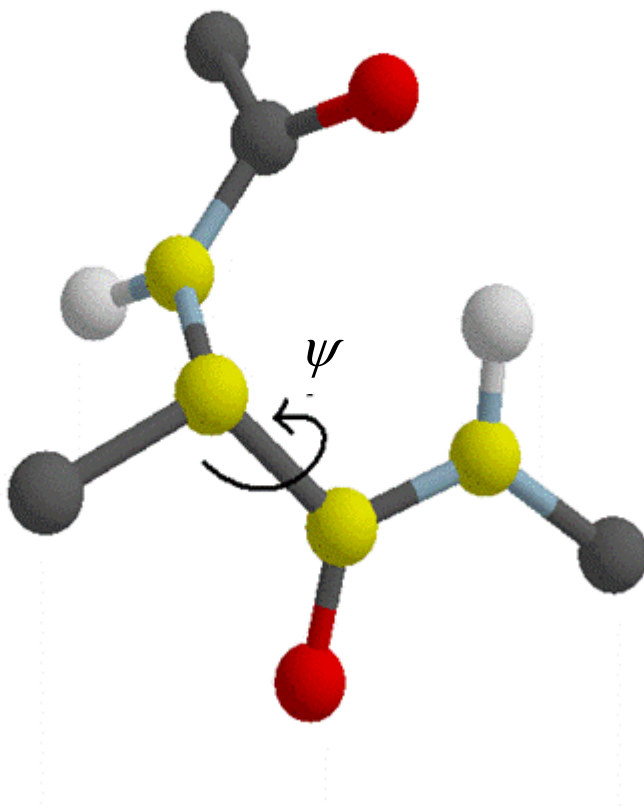


Fig. 2



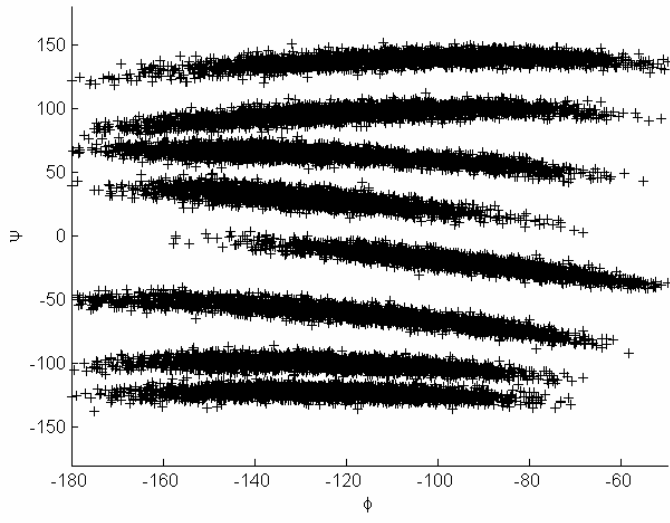


Fig. 3

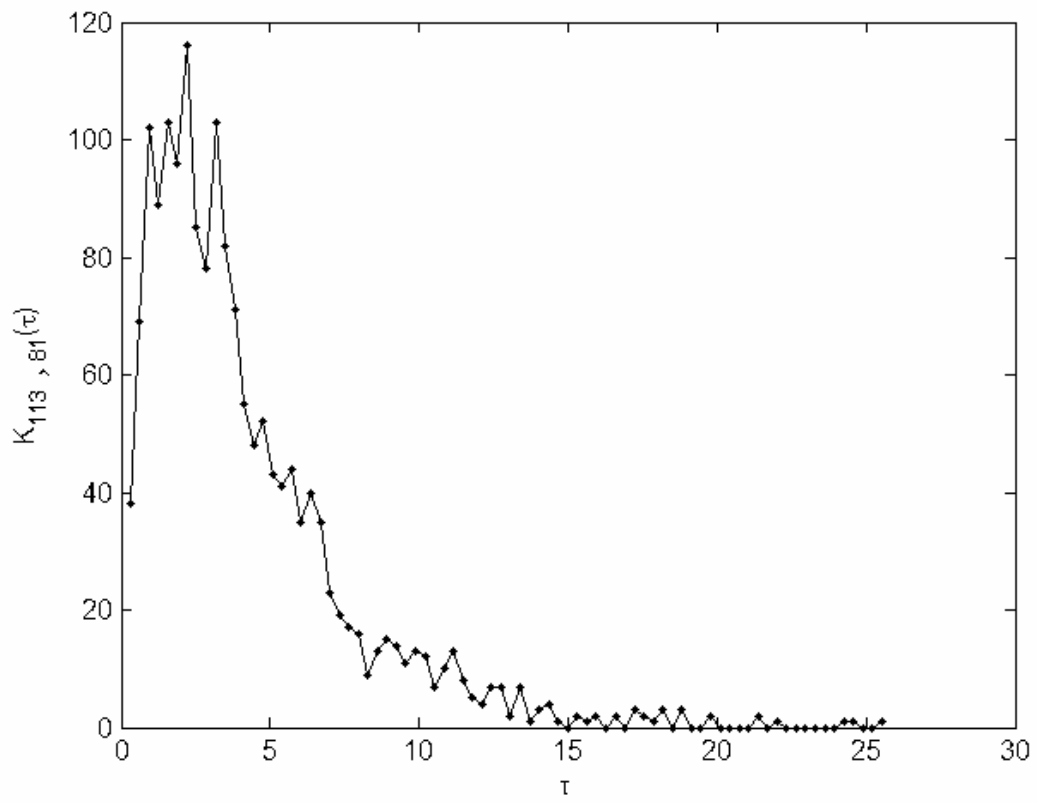


Fig. 4

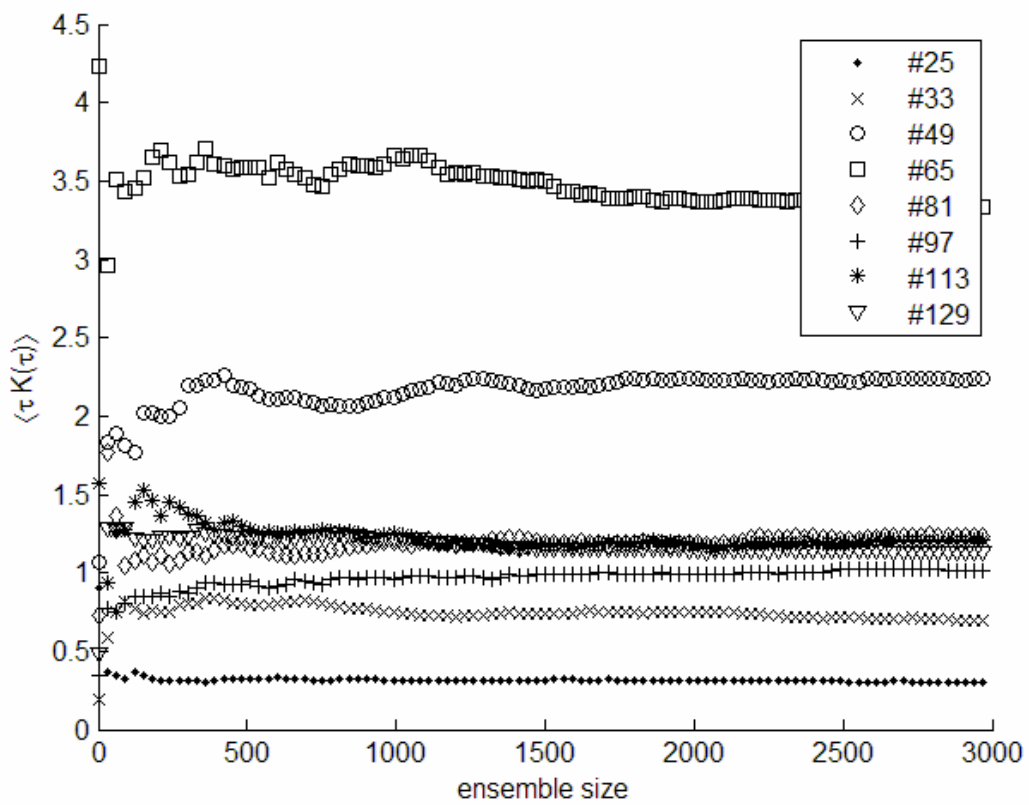


Fig. 5

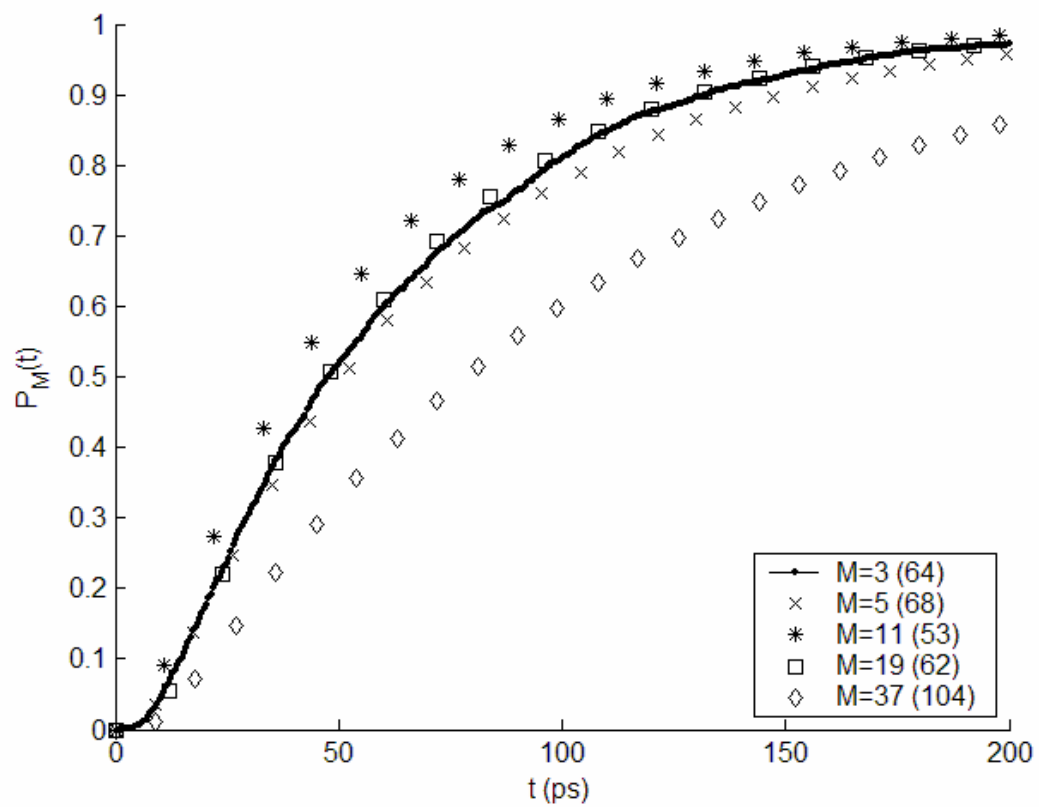


Fig. 6

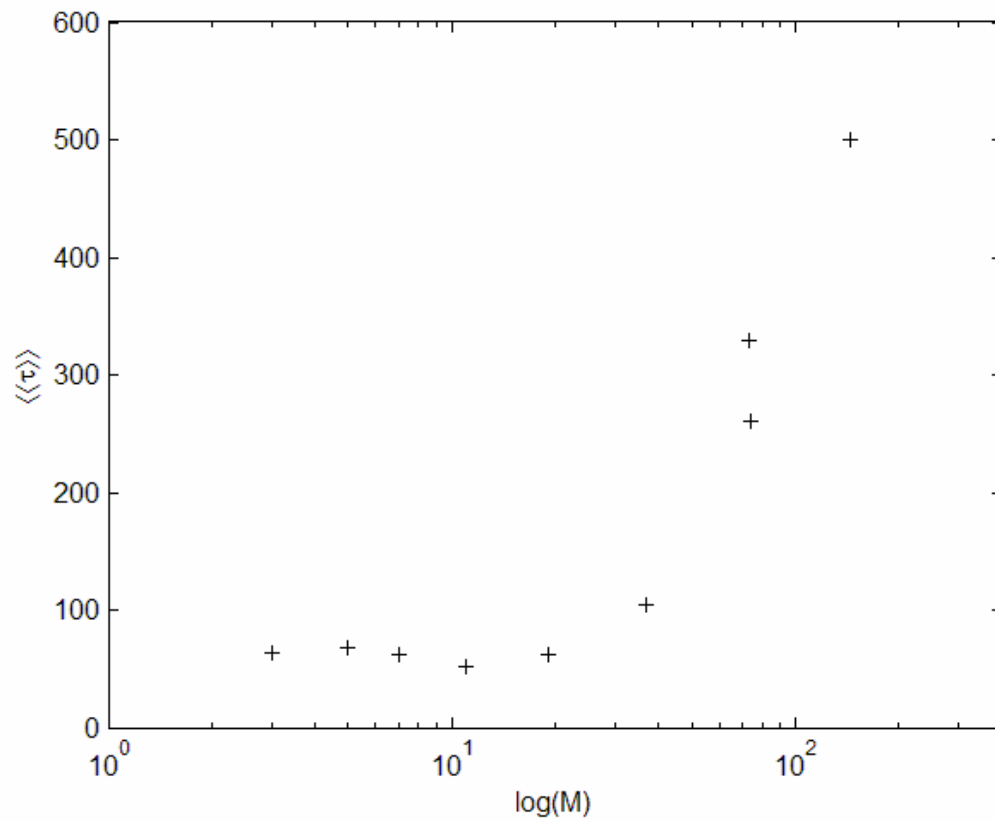


Fig. 7

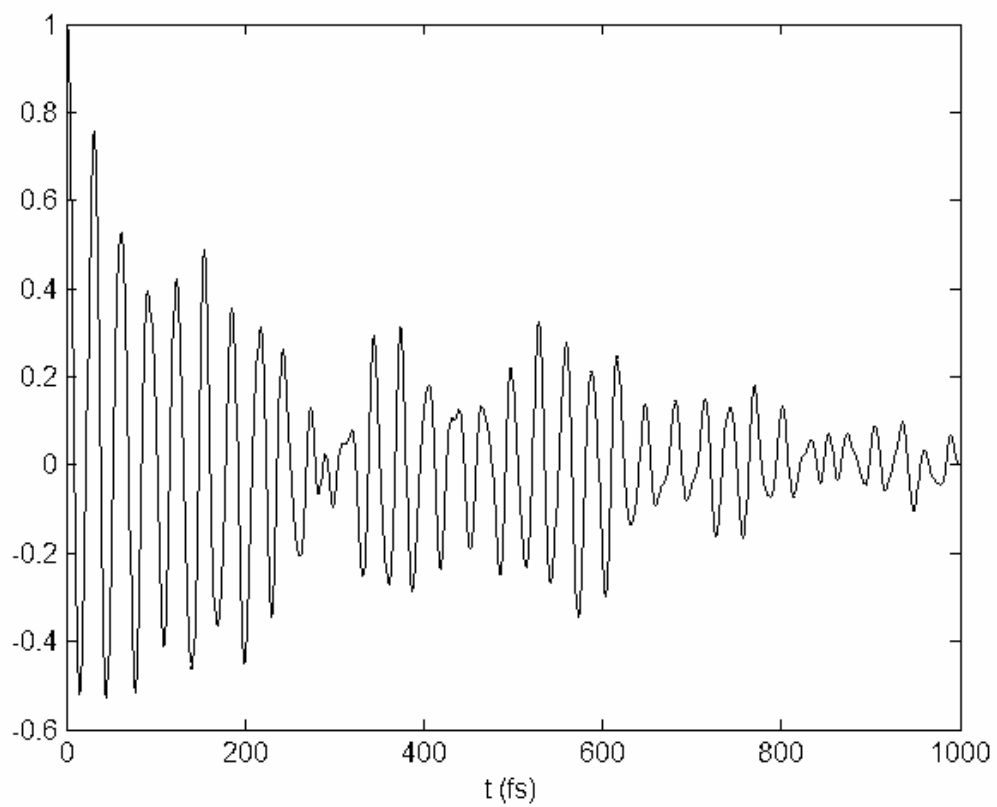


Fig. 8

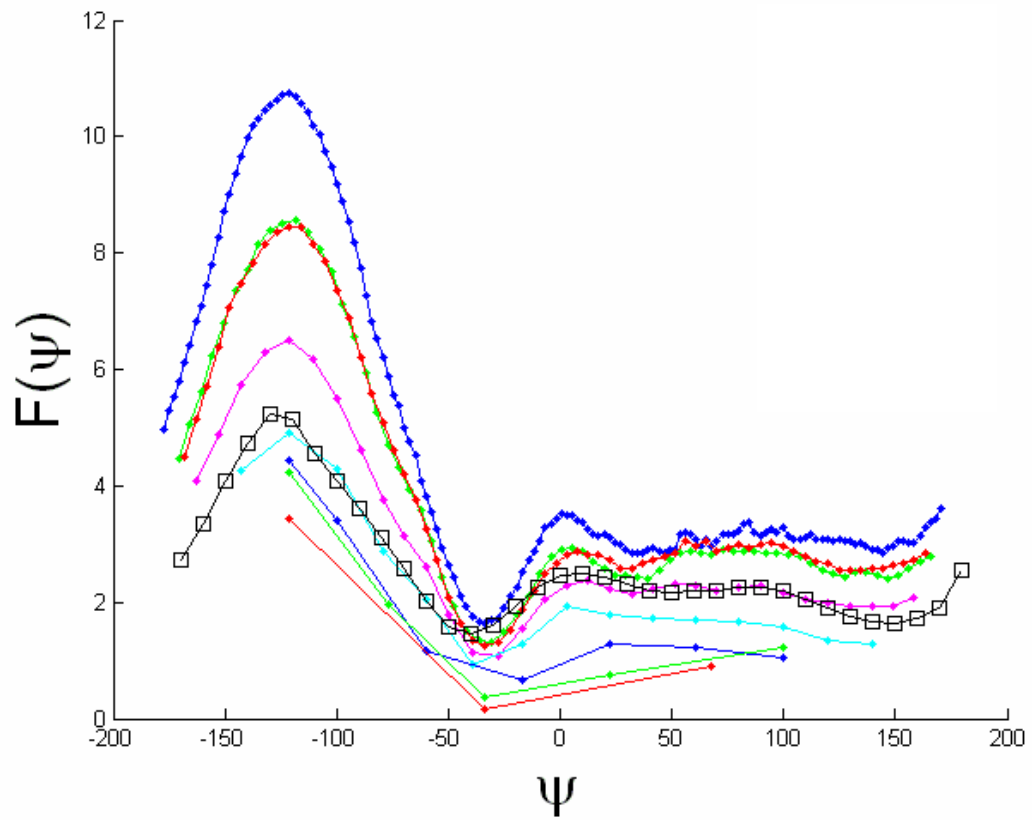


Fig 9a

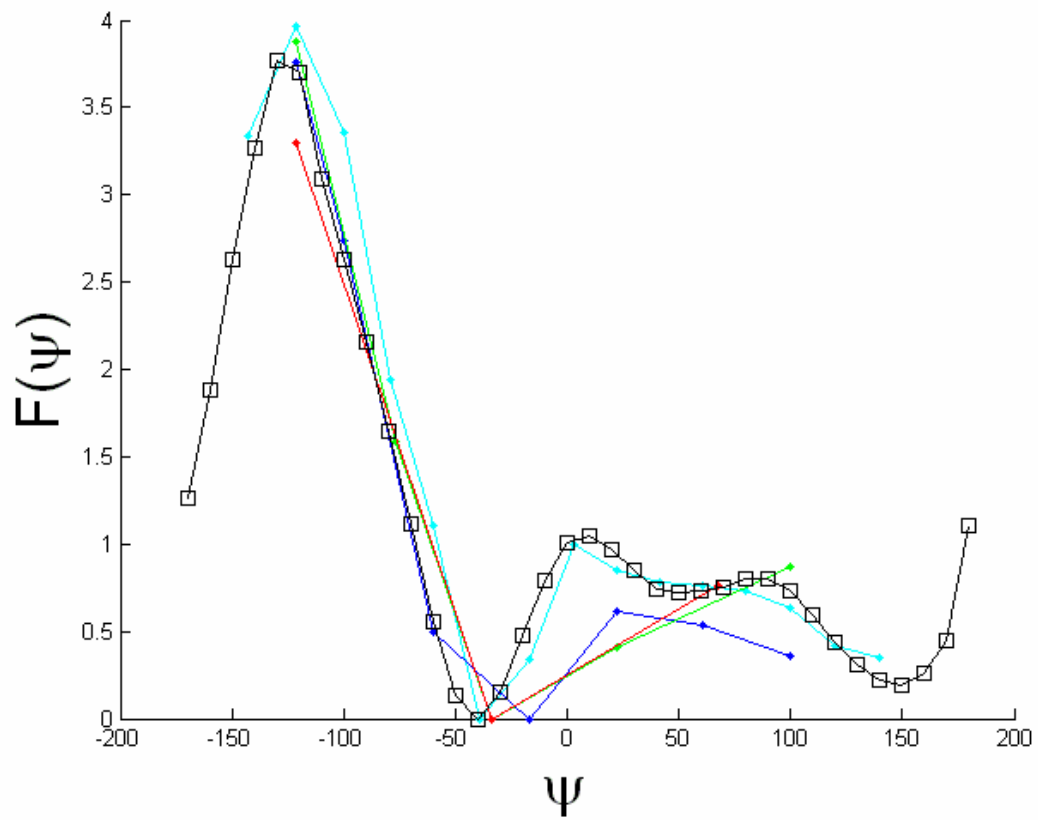


Fig. 9b

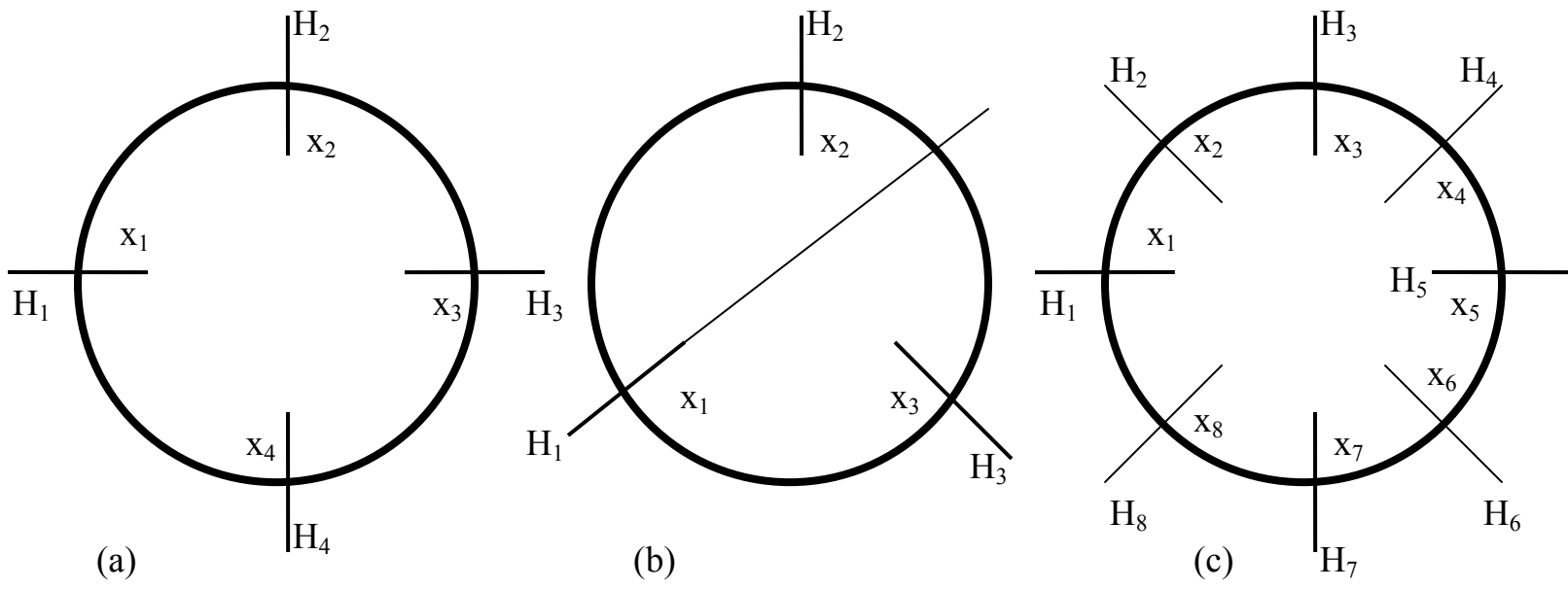


Fig. 10

## FIGURE CAPTIONS

Figure 1. A schematic representation of the milestoning approach. A set of non-intersecting hypersurfaces  $H_1, \dots, H_M$  establishes “milestones” between the reactant (R) and the product (P). The circles indicate the volume in coordinate space that defines the boundaries of each of the final states. Also shown (dashed line) a single trajectory initiated at  $H_s$  and terminated at  $H_{s-1}$ . An ensemble of such trajectories is used to compute the first passage time distributions between the milestones (see text for more details).

Figure 2. A ball-and-stick model of blocked alanine, the molecular system whose transition from an  $\alpha$  helix to a  $\beta$  sheet in aqueous solution is studied in the present paper (water molecules not shown). Also shown is the dihedral angle  $\psi$  that was used as a reaction coordinate. The conformation shown is  $C7_{eq}$ .

Figure 3.  $\phi, \psi$ -projection of the OEQ ensembles of the M=11 simulation, showing hyperplanes 25, 33, 49, 65, 81, 97, 113, 129.. Note that the hyperplane is defined in Cartesian space (including all the peptide degrees of freedom). The points shown satisfy the Cartesian constraints exactly, so their projection onto the (non-Cartesian)  $\phi, \psi$  plane is not one-dimensional.

Figure 4. The local first passage time distribution of the population initiated at milestone 6 to milestone 5 (the trajectories are terminated either at milestone 5 or 7). This distribution is taken from the seven-milestone simulations (see table 1).

Figure 5. Convergence of the local first passage time (between milestones) in picoseconds as a function of the number of trajectories. Although we have used 3,000 trajectories, reasonable convergence can be achieved with just a few hundred.

Figure 6. The rate of product appearance (milestone  $M$  is absorbing) as a function of the number of milestones. The results for 3 milestones are exact and calculations with 19 milestones (or less) are accurate. Corresponding overall first passage times are given in parentheses. Note that the calculation of the time course does not assume exponential kinetics. Indeed the early-time regime shows a significant deviation from exponential behavior. See further discussion on the non-exponential kinetics of the transition in alanine dipeptide in section **VII. Markov model and a single exponential relaxation.**

Figure 7. The dependence of the overall first passage time on the number of milestones. We obviously expect the accuracy of the rate calculations to increase as the number of milestones decreases. However, the plot shows early convergence to the right answer. The calculations are accurate with 19 and fewer milestones. The speed-up obtained with 19 milestones compared to straightforward molecular dynamics trajectories is a factor of 9. See text for the analysis.

Figure 8. Velocity correlation function for the  $\psi$  dihedral angle. A rough estimate for the relaxation time suggests that it is smaller than 300 femtoseconds. Instead of an exact time derivative we have used the finite difference expression  $\dot{\psi} \approx \Delta\psi/\Delta t$

Figure 9. Computation of free energies with Milestoning. Multiple free energy curves as a function of the dihedral angle  $\psi$  are shown for a different number of milestones. Exact results (squares) were obtained by a long Molecular Dynamics trajectory. Figure 9.a includes free energy profiles (going from top to the lower curves) of runs with 144, 74, 73, 37, 19, 11, 7, 5 milestones while the lower figure (figure 9.b) only  $M \leq 19$  ( $M=19, 17, 7, 5$ ). When the Milestoning approximation breaks down the free energy surface is distorted in values, but it still maintains the correct positions of maxima and minima. Note that for a sufficiently small number of milestones (in which the kinetics are described accurately) the coarse free energy surfaces are similar.

Figure 10. A schematic drawing of the termination condition of a Milestoning trajectory that follows a circle. The toy model is an illustration of a potential error in determining termination times (see text for more details).

## Tables

<i>M</i>	<i>milestone numbers</i>
3	25, 59, 140
5	1, 25, 59, 100, 140
7	1, 25, 42, 59, 81, 113, 140
11	1, 17, 25, 33, 49, 65, 81, 97, 113, 129, 140
19	1, 9, 17, 25, ... , 129, 137
37	1, 5, 9, 13, ... , 137, 141
73	odd-numbered milestones
74	even-numbered milestones, and milestone 143
144	all milestones

Table 1: Summary of Milestoning runs.

A detailed list of all the milestoning runs performed in this study. The run with three milestones is exact. Run with 19 milestones or less provide accurate answers.

Table 2

$M$	$\langle\langle\tau\rangle\rangle$ (ps)	$1/ \gamma_1 $ (ps)	$\bar{\tau}$ (fs)	$\sqrt{\langle\langle\tau^2\rangle\rangle - \langle\langle\tau\rangle\rangle^2}$ <i>Markov</i>	$\sqrt{\langle\langle\tau^2\rangle\rangle - \langle\langle\tau\rangle\rangle^2}$ <i>Milestoning</i>
144	500(3.1)	421	31.2		
74	261(1.2)	228	57.7		
73	330(1.6)	284	58.3		
37	104(0.63)	100	129	93.1	93.1
19	62(0.47)	41	373	51.2	51.2
11	53(0.50)	40	1,305	45.5	45.6
7	62(0.73)	32	3,581	55.0	53.8
5	68(0.93)	23	10,902	62.5	60.6
3	64(1.04)	-	-	63.9	53.8

Table 2. The first column is the number of milestones that were used in the rate calculation. The second column shows computed MFPTs of the alanine dipeptide  $\alpha \rightarrow \beta$  ( $H_{59} \rightarrow H_{140}$ ) transition, as predicted by (5). Statistical errors for the first passage time (in brackets) were computed according to the procedure of reference<sup>5</sup>. The results of the last row ( $M=3$ ) are for an exact model. The third column shows the predictions of a single rate constant (exponential rate— see section **VII. Markov model and a single exponential relaxation**). Note the significant deviation of the MFPT from the prediction of a single, exponential reaction, i.e. the  $1/\gamma_1$  prediction. The fourth column shows the

milestone averaged incubation time (AIT). The last two columns are the standard deviations of the first passage time distribution with the Markovian model (not single exponential!) and with Milestoning. The second moment is a measure of the non-exponential behavior of the kinetics. We did not compute the second moments for simulations with  $M > 37$  because even their first moments are highly inaccurate. Overall the differences between the Markovian model and Milestoning are not large. For the non-Markovian contribution to be important, the incubation kinetics must be non-exponential. This is the case for 3 milestones.