9-2007

# Diffusive Hidden Markov Model Characterization of DNA Looping Dynamics in Tethered Particle Experiments

John Beausang
*University of Pennsylvania*

Philip C. Nelson
*University of Pennsylvania*, nelson@physics.upenn.edu

# Diffusive Hidden Markov Model Characterization of DNA Looping Dynamics in Tethered Particle Experiments

## Abstract

In many biochemical processes, proteins bound to DNA at distant sites are brought into close proximity by loops in the underlying DNA. For example, the function of some gene-regulatory proteins depends on such "DNA looping" interactions. We present a new technique for characterizing the kinetics of loop formation in vitro, as observed using the tethered particle method, and apply it to experimental data on looping induced by lambda repressor. Our method uses a modified ("diffusive") hidden Markov analysis that directly incorporates the Brownian motion of the observed tethered bead. We compare looping lifetimes found with our method (which we find are consistent over a range of sampling frequencies) to those obtained via the traditional threshold-crossing analysis (which can vary depending on how the raw data are filtered in the time domain). Our method does not involve any time filtering and can detect sudden changes in looping behavior. For example, we show how our method can identify transitions between long-lived, kinetically distinct states that would otherwise be difficult to discern.

## Disciplines

Physical Sciences and Mathematics | Physics

# Diffusive hidden Markov model characterization of DNA looping dynamics in tethered particle experiments

**John F Beausang and Philip C Nelson**

Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104 USA, fax: (215) 898-2010, phone:(215) 898-7001

E-mail: `nelson@physics.upenn.edu`

**Abstract.**

In many biochemical processes, proteins bound to DNA at distant sites are brought into close proximity by loops in the underlying DNA. For example, the function of some gene-regulatory proteins depends on such "DNA looping" interactions. We present a new technique for characterizing the kinetics of loop formation *in vitro*, as observed using the tethered particle method, and apply it to experimental data on looping induced by *lambda* repressor. Our method uses a modified ("diffusive") hidden Markov analysis that directly incorporates the Brownian motion of the observed tethered bead. We compare looping lifetimes found with our method (which we find are consistent over a range of sampling frequencies) to those obtained via the traditional threshold-crossing analysis (which can vary depending on how the raw data are filtered in the time domain). Our method does not involve any time filtering and can detect sudden changes in looping behavior. For example, we show how our method can identify transitions between long-lived, kinetically distinct states that would otherwise be difficult to discern.

## 1. Introduction and summary

### 1.1. Tethered particle motion

Molecular biophysics relies on an ever-expanding toolkit of methods to "see" processes not visible by traditional light microscopy. The toolkit is vast in part because each method has its own set of strengths and weaknesses, suiting it for a particular class of measurement challenges. For example, some methods may yield excellent spatial resolution but limited information about the dynamics of a process, particularly in a physiologically relevant context.

Recent methods that observe state transitions in single molecules allow us to see differences in behavior between different molecules, as well as making it possible to extract detailed kinetic information. Tethered particle motion (TPM) is one such technique to monitor conformational changes in single molecules of DNA, such as loop formation (figure 1), in real time. Our goal in this paper is to extend the range of applicability of TPM as a tool to study the kinetics of such changes, adding many details to a previous letter [1].

In TPM, a bead is tethered to the surface of a microscope slide by a single DNA molecule and undergoes Brownian motion without any externally applied force (for example, see [2]– [14]). Lateral displacements of the bead are passively observed through an optical microscope. TPM is suited for the study of short molecules (from 100 to several thousand basepairs long). Unlike methods that stretch the DNA, TPM allows the study of proteins acting on molecules not subjected to externally applied tension. Thus TPM offers the possibility to quantify the dynamics of loop formation and breakdown,
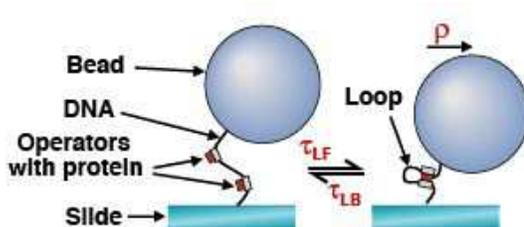
**Figure 1.** A DNA molecule flexibly links a bead to a surface. The motion of the bead's center is observed and tracked, for example as described in Refs. [15, 16. In each video frame, the position vector, usually projected to the $xy$ plane, is found. After drift subtraction, the mean of this position vector defines the anchoring point; the vector **r** discussed in this article is always understood to be the drift-subtracted, projected position, measured relative to this anchoring point. The length of this vector is called $\rho$. Time constants $\tau_{\mathrm{LF}}$ and $\tau_{\mathrm{LB}}$ determine the rates to form and break loops, respectively. The figure is simplified; in the actual experiment leading to the data we analyze, the DNA had two sets of three binding sites ("operators") for cI repressor protein [1], and each operator can at a given moment be occupied or unoccupied.

and then determine the dependence on biophysical parameters such as loop size, tether length, repressor concentration, operator sequence, and interoperator sequence. An additional advantage of TPM is that it can be parallelized: Many tethered particles can be simultaneously tracked in a single run.

Some of the current implementations of TPM allow single-particle tracking of multiple beads with 20–50 ms time and $\sim$10 nm spatial resolution, allowing rather precise determination of effective tether length from data [16–18]. (Other current work observes a time-averaged image instead of using single particle tracking [14].) Thus TPM provides an attractive assay for studying the formation of DNA loops by protein complexes that bind to multiple sites on a molecule, including the kinetics of such loop formation in solution. For example, van den Broek *et al.* have performed such a measurement, using as their model system the restriction enzymes NaeI and NarI [19]. Other single-molecule techniques exist for observing juxtaposition of parts of molecules (e.g., Förster resonance energy transfer [20]). These methods do not require the relatively large reporting particle (the bead) needed in TPM. But fluorescence techniques are subject to photobleaching, which limits their usefulness when studying very long-timescale transitions. In contrast, the TPM method easily allows the observation of a single DNA molecule over periods of 30–60 minutes [15, 16]. Even for short-timescale transitions, the long observations possible with TPM give an advantage, because the hidden Markov analysis developed in this paper can extract reliable information from long time series, even when it has low signal to noise ratio.

Here we present an analysis method for TPM experiments on DNA looping that models the bead motion and looping transitions in terms of quantities determined empirically from data on non-looping DNA, plus two additional rate constants (for loop formation and breakdown). Then we apply a maximum-likelihood approach to determine the numerical values of the rate constants. We illustrate our approach by applying it to looping data from the *lambda* regulatory-protein complex. In this context (and others such as the *lac* complex [4, 14]), the analysis of looping kinetics can answer basic biological questions about gene regulation [21–23]. For example, is repression accomplished via a permanent loop? Or, does the loop open transiently, allowing some subthreshold activity of transcription in the "repressed" state? In the case of lambda phage, the latter scenario could allow faster response to changes in the cell's environment, without an unacceptable rate of "accidental" transitions to the lytic program. Indeed using our method, we found dynamic looping activity at concentrations of cI repressor protein corresponding physiologically to the lysogenic state of lambda [1]. Also, the cooperativity between multiple cI proteins bound to the *lambda* DNA may result in loops of varying stability as operator occupancy changes on a time scale slower than individual looping events.

### 1.2. Hidden Markov modeling

The technical problems we wish to address are easily stated. We wish to deduce the kinetics of an invisible process (DNA loop formation) indirectly, by observing bead motion. As loops form and break, the effective DNA tether length switches between two (or more) distinct values. Simply observing the magnitude of the bead's motion to determine the state of the tether (and thus the rate of loop formation) is not sufficient, because the expected distributions of projected bead position for the looped and unlooped states overlap substantially ( [17]; see figure 4 below). We need a way to avoid misinterpreting these cases as looping events. We could try to minimize this problem by filtering the data over a large enough time window. For example, Refs. [15, 19] used a window of width 4 s, but such filtering degrades our ability to see brief but genuine looping events. Indeed we will argue that in some systems, for example the *lambda* system studied in this paper, the mean state lifetimes are not much longer than the characteristic time scale for the bead to diffuse through its allowed domain. Windowing the data will then not yield unambiguous results; the kinetic rate constants inferred by the method of [19], applied to our system, depend on the arbitrary choice of window size (see [1, 24] and section 3.1 below).

Our difficulty is in some ways similar to that faced in the interpretation of ion-channel data, where again one wishes to infer invisible state changes from their visible, but noisy, consequences. In that context, hidden Markov modeling, combined with the maximum

likelihood method (the "HMM technique" [25]), offers advantages over simpler histogram schemes (see for example [26, 27]). The HMM technique supposes a particular kinetic scheme for the underlying process with unknown rate constants. For any given values of those constants, and any time series of the observable, one then calculates the likelihood of that time series emerging in that model. Substituting actual experimental data for the time series, the inferred values for the unknown time constants are the ones that maximize this likelihood.

The HMM technique does not require time-domain filtering (windowing) to the data (although for some applications it may be desirable, e.g. [28, 29]). Nor does it require the selection of a threshold, across which each data point is labeled as being definitively in the looped or unlooped state. Finally, when the underlying process has more than two states, HMM in principle allows one to extract a more complete set of rate constants than does the method of dwell time histograms [20, 29].

For such reasons, HMM and other maximum-likelihood approaches have recently proven to be a powerful technique for the analysis of a number of single-molecule experiments [20, 30–35]. Much of this paper will be devoted to adapting the HMM method to the context of TPM measurements of DNA looping. To accomplish this goal, we will need to overcome some obstacles, which we now outline.

### 1.3. Need for diffusive HMM

TPM experiments generate data sets consisting of a set of bead images, from which a time series of bead center locations $\mathcal{O} = \{(x_t, y_t)\} = \{\mathbf{r}_t\}$ is determined [16]. Underlying this observed signal is the desired sequence $\{q_t\}$ of the DNA conformational state (e.g., in a 2-state system, $q = 1$ means unlooped, and $q = 2$ looped, at each time point). In principle, the probability distribution $P(\mathbf{r}_t, q_t)$ for the bead position and looping state at time $t$ could depend on the entire prior history of the system, that is, on $\mathbf{r}_{t-1}, q_{t-1}, \mathbf{r}_{t-2}, q_{t-2} \ldots \mathbf{r}_1, q_1$. Standard hidden Markov modeling [25] assumes that the observed signal is uncorrelated, depending only on the current hidden state via a distribution $p_{\text{bead}}(\mathbf{r}_t | q_t)$, and that this hidden state in turn depends only on the previous one, via a "transition matrix" $T(q_t | q_{t-1})$:

$$P_{\text{HMM}}(\mathbf{r}_t, q_t) = p_{\text{bead}}(\mathbf{r}_t | q_t) T(q_t | q_{t-1}) \tag{1}$$

As we will show in section 3.2, however, this assumption does not apply to TPM experiments, in part because of correlations in position due to the diffusive character of the bead's motion. Essentially, our modified DHMM replaces the above assumption with a slightly more general version, in which $P(\mathbf{r}_t, q_t)$ depends on *both* $\mathbf{r}_{t-1}$ and $q_{t-1}$. Finding and justifying a form for this probability function, with just two unknown fit parameters, is the crux of our problem.

### 1.4. Outline of this paper and summary of notation

Table 2 in the Glossary section gives a summary of the notation used in this paper. Readers familiar with TPM experiments and hidden Markov modeling may wish to use this table and skip directly to section 4, where we formulate our approach.

Section 2 outlines the class of experiments we study and shows some typical data. Section 3 outlines methods used in prior work, specifically the windowing/threshold method and standard hidden Markov modeling, along with our critique. Section 5 gives some illustrative results for the experimental dataset studied in Ref. [1].

## 2. Typical experiments

For concreteness, we will analyze a representative TPM experiment [1, 15, 16] in which the projected $(x, y)$ positions of up to 6 beads are simultaneously imaged using differential interference contrast microscopy and recorded along with a time stamp for up to 45 minutes. Images are recorded from alternate rows of pixels every 20 ms from a charge coupled device operating in interlaced mode. Due to the difficulty in obtaining precise alignment of the two rows of pixels, we treat each time series as two separate sets; the resulting 40 ms time resolution is adequate for our purposes. Rejection of anomalous beads (e.g., double tethers, surface adhesion, etc.) and correction for microscope drift are described in Ref. [16]. Simultaneous tracking of multiple beads allows microscope drift to be estimated from the collective motion of all the beads and then subtracted from each bead separately, [16]. We determine the long–time drift by first finding the average $(x, y)$ position of each bead in a 20 s and then a spatial average over all the beads is calculated by summing over each of the 20 s time averages. An interpolating function that passes through these points is fitted and subtracted from each measurement [16]. This second average over the beads results in a smoother interpolation that minimizes any inadvertent removal of true bead motion. Rarely (2 points in figure 2), an anomalous outlying point or a missing frame is replaced with an interpolation of the neighboring points. Whenever we refer to $x$ and $y$ below, we are referring to the drift-corrected time series.

Adding 200 nM of cI repressor protein converts the homogeneous tethered Brownian motion of the particle to a regime characterized by abrupt, dynamic transitions in bead motion (figures 2–3), a fraction of which have long enough dwell times ($\sim$1–100 sec) to be visible by eye in the microscope.

We also use data from two kinds of control experiments: *(1)* tethered beads with no protein and thus no looping, and *(2)* a small subset of beads with cI present that remain in a permanently looped configuration for many minutes (data not shown, see [1]). The first of these controls is obtained with every experiment from a preliminary 20 minute
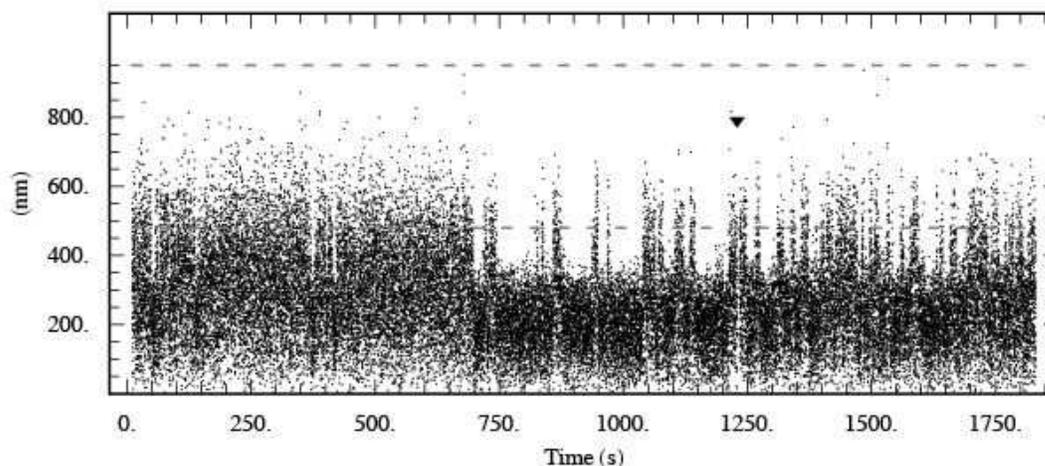
**Figure 2.** Typical time series of bead positions. DNA constructs of total length $3477$ bp were attached at one end to a glass coverslip and at the other to a $480$ nm diameter bead. The vertical axis gives the actual distance of the bead center from its attachment point, after drift subtraction. The trace shown passed the tests discussed in Ref. [16], for example the ones that eliminate doubly-tethered beads. The DNA construct contained two sets of three operator sites. The two sets of operators were separated by $2317$ bp. The system contained cI repressor protein at concentration $200$ nM (where $200$ nM refers to the concentration of cI dimers since monomers do not bind to DNA); repressor proteins bind to the operator sites on the DNA, and to each other, looping the DNA as in figure 1. A sharp transition can be seen from a regime of no loop formation to one of dynamical loop formation at $\sim 650$ s. Later section 5.3 will argue that the latter regime itself consists of two kinetically distinct subregimes. The dashed lines represent the two values of $\rho_{\max}$ corresponding respectively to the looped and unlooped states; control data in these two states was observed never to exceed these values. A brief sticking event, indicated by the inverted triangle, was excised from the data prior to analysis. Section 4.3.1 uses these observed values of $\rho_{\max}$ to create truncated Gaussian step distributions for our models of the looped and unlooped Brownian motion. (Experimental data for this and subsequent figures from [1], kindly supplied by C. Zurla and L. Finzi.)

recording with no repressor present. Figures 4–5 show two important reduced forms of typical experimental data (solid curves), along with simulation results to be discussed in later sections.

## 3. Prior methods

Before constructing the DHMM method, we briefly review two existing methods, to highlight their limitations in the context of tethered particle motion experiments.
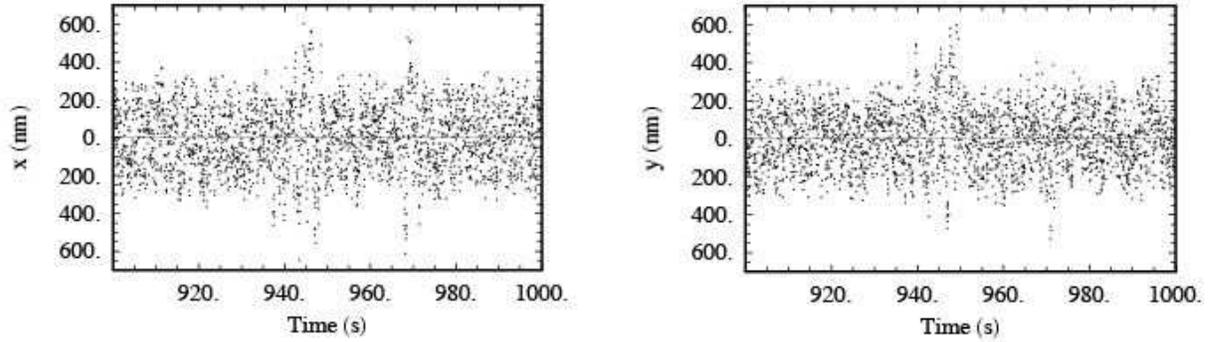
**Figure 3.** Time series for $x$ and $y$ corresponding to figure 2. The graphs show that our drift subtraction scheme leads to visually similar traces for $x$ and $y$.
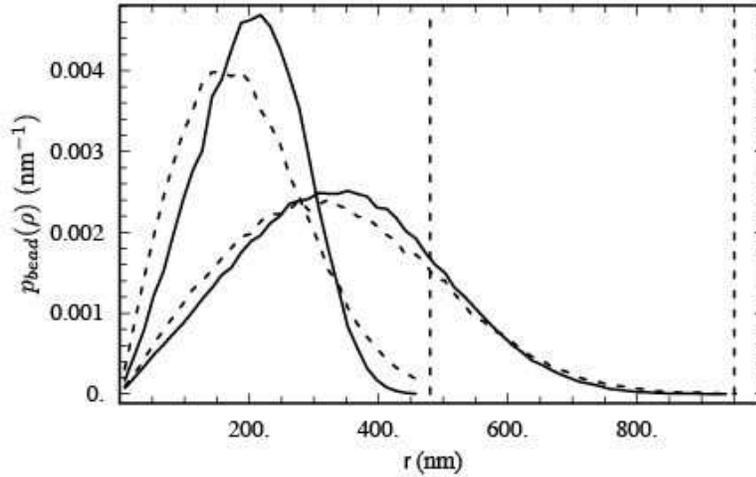


**Figure 4.** *Left solid curve:* Normalized probability distribution function (pdf) of bead center location for experimental control data corresponding to a permanently looped tether (about 65 000 video frames). *Right solid curve:* Corresponding pdf for unlooped control data (about 200 000 video frames). *Dashed curves:* Corresponding pdfs for simulated bead motion, computed using the step distribution functions found in section 4.3.1, with similar numbers of simulated steps. Although the agreement with the experimental data is not perfect, it is quite nontrivial: The step distribution functions were not chosen to make these distributions agree, but rather to match the observed distributions of steps between pairs of adjacent video frames. In contrast to this dynamical model, the equilibrium calculations of Refs. [17 and [16] are less ambitious; they evaluate expectations based on Monte Carlo integration of the Boltzmann distribution. (Data from [1].)
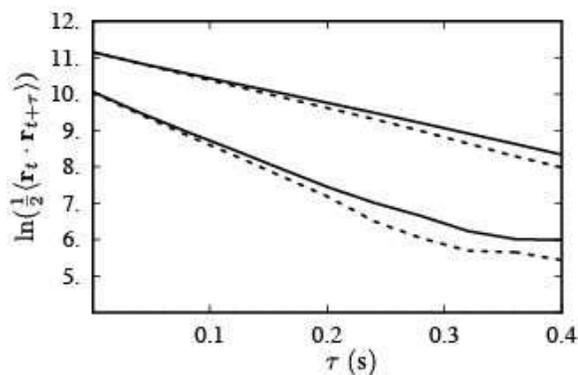
**Figure 5.** *Upper solid curve:* Semilog plot of the time autocorrelation function, $\frac{1}{2}\langle \mathbf{r}_t \cdot \mathbf{r}_{t+\tau} \rangle$ for unlooped experimental control data used in figure 4. *Lower solid curve:* Analogous function for permanently looped control data. *Dashed curves:* Analogous functions for simulated control data, computed using the step distribution functions found in section 4.3.1. As in figure 4, the agreement is nontrivial: The simulation was based on pairs of data points differing by just one video frame, so the approximate agreement supports our assumption that the tethered bead motion is Markovian. (The equilibrium calculations of Refs. [16,17] made no attempt to calculate autocorrelations.)

## 3.1. Windowing/threshold method

A traditional method for determining transition rates from a signal that depends on the hidden, underlying state is to record the signal over many transitions; transition events are defined when the signal crosses some threshold value. The dwell times between threshold crossings of the experimentally distinguishable states are separately histogrammed and then fit to an exponential (or multiexponential) distribution. The parameters of the fits are the state lifetimes; their inverses are the transition rates.

TPM experiments present some difficulties for this straightforward approach, however. First, the observed signal (bead center position) is only indirectly related to the desired DNA looping state. For example, in the unlooped state the bead will spend an appreciable fraction of its time close to the attachment point, mimicking the looped state, see figure 4. Additionally, the height coordinate $z$ of bead position is either not measured (e.g. in [15, 16]) or is measured to less precision than $x, y$ (e.g. in [13]), increasing further the overlap between the pdfs (figure 4). Thus we cannot state with certainty the looping states of a majority of the individual data points in our time series; in our example data, measurements with $\rho_t$ less than $\sim 450$ nm can be in either looped state.

To overcome this difficulty, the raw data are often filtered using a sliding time window of width $W$, whose value is chosen to make clearly visible steps emerge in the data. Often this filtering step is a calculation of the signal's variance in the window, which has the

additional advantage of being insensitive to instrumental drift over time scales slower than $W$. Next, a threshold is chosen that separates the high- and low-variance states, which now appear more clearly in the filtered data. Dwell times are then defined between those threshold crossings of the filtered data whose durations exceed the filter dead time (see [24, 36]). If the window $W$ contains many independent measurements, then we get an accurate measure of the variance, and hence few spurious reported threshold crossings between the high- and low-variance states.

Unfortunately, any filter based on a finite-sample estimate of bead position variance converges slowly as we increase $W$. For a time series of *uncorrelated* samples, the variance of the estimated variance is proportional to $1/$(number of samples in the window $W$) [28, 37]. This observation may make it seem that we need only take data at a high enough frame rate $\gamma$ to make $1/(W\gamma)$ sufficiently small. But successive snapshots of bead position are *not* uncorrelated. The appropriate prefactor is not $1/(W\gamma)$ but rather $\tau_{\mathrm{diff}}/W$, where $\tau_{\mathrm{diff}}$ is the diffusion time for the bead to traverse its range of motion. We estimate $\tau_{\mathrm{diff}} \approx 140\,\mathrm{ms}$ for TPM experiments from the $1/e$ decay of the position autocorrelation function with $\sim\mu$m length lambda DNA fragment and $\sim 0.5\,\mu$m diameter beads.

Hence the fractional statistical noise in our estimate of bead variance, $\sqrt{\tau_{\mathrm{diff}}/W}$, decreases slowly with increasing $W$; $W$ must therefore be taken large to obtain sufficient noise rejection (sufficiently rare spurious threshold crossings). Indeed, we found using control data that taking $W$ less than 1–2 seconds introduces spurious looping events in the unlooped control data where there should be none. On the other hand, $W$ must be taken smaller than the shortest state lifetime ($\sim 1$ second); otherwise we miss too many genuine state transitions. These two constraints set a fundamental limit on the accuracy of the windowing/threshold method, and in fact we found that applying the threshold-crossing method to experimental data of reasonable duration resulted in reported lifetimes that depend on the choice of $W$ (see figure 6) [1].

Techniques do exist to correct for missed events [19, 24, 36] and to improve the reliability of the threshold method. Our DHMM model will bypass this difficulty altogether, by avoiding the filtering step.

### 3.2. Standard HMM analysis

This section briefly reviews the standard hidden Markov method, considers its possible application to tethered particle experiments, and points out a serious limitation in its applicability. The purpose is to motivate our modification of standard HMM in section 4, and to establish some mathematical notation to be used there.

DNA looping experiments are natural candidates for application of hidden Markov methods, because the probability distribution functions (pdfs) for the bead position (called
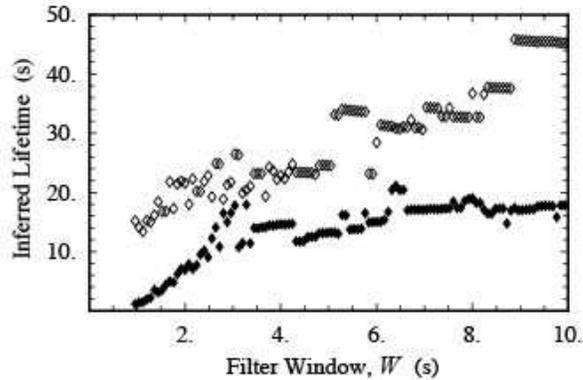
**Figure 6.** *Black dots:* Lifetime results for the windowing/threshold method for DNA loop formation lifetime (points), as a function of filter window size $W$. These results are based on the segment of data between 650–1750 seconds in figure 2 (total of 27 500 video frames). For each value of $W$, dwell times longer than twice the filter dead time (i.e. $W$) were histogrammed and fit to single exponentials to determine $\tau_{\mathrm{LF}}$. No attempt was made to correct for missed events. *Open dots:* Similar results for loop breakdown time scale $\tau_{\mathrm{LB}}$. The dependence of the inferred lifetime on the filter can complicate the "best" choice of $W$. Our DHMM method uses no window.

$p_{\mathrm{bead}}(\mathbf{r}|q)$ in (1)) can be calculated *a priori* [16, 17] or measured directly [15, 16]. To model dynamical looping data, then, we could imagine an underlying (hidden) 2-state Markov process describing loop formation and breakdown, then suppose that in each of the two states the observed quantity (bead position) is drawn from the pdf appropriate to the current looping state. The underlying process has two unknown parameters, the rate constants $1/\tau_{\mathrm{LF}}$ for loop formation and $1/\tau_{\mathrm{LB}}$ for loop breakdown. Using assumed values for these parameters, and the known pdfs for bead position, we can calculate the likelihood that any given time series of bead positions would be observed. (For a general introduction to maximum likelihood methods see Refs. [37, 38.) Substituting an actual observed time series and maximizing the likelihood over $\tau_{\mathrm{LF}}$ and $\tau_{\mathrm{LB}}$ then yields the best estimates for the lifetimes, which are the quantities of interest to us.

Before we critique the approach just outlined, let us make it somewhat more explicit. The standard HMM [25] supposes that an observed signal reflects two processes: An autonomous Markov process (in our case, loop formation and breakdown) generates a time series $\mathcal{Q} = \{q_t\}$ at a discrete set of times $t = 1, \ldots M$, using a $2 \times 2$ matrix of transition probabilities to represent $T(q|q')$ in (1), where the $t$ subscripts are dropped and the prime denotes the preceding measurement. $\mathcal{Q}$ is not directly observable, but it influences an observable signal $\mathcal{O} = \{\rho_t\}$: At each time, $\rho_t$ is drawn from a probability distribution $p_{\mathrm{bead}}(\rho_t|q_t)$, which depends only on the current value of $q_t$. In particular, there

is assumed to be no "back reaction" from the observed value of $\rho$ onto the underlying Markov process, and no "memory" in the process that generates $\mathcal{O} = \{\rho_t\}$.

The overall likelihood of observing a time series $\mathcal{O}$ is then taken to be a sum over all possible hidden trajectories $\mathcal{Q}$:

$$P_{\text{tot,HMM}}(\mathcal{O}) = \sum_{\mathcal{Q}} \left[ \prod_{t=2}^{M} p_{\text{bead}}(\rho_t|q_t) T(q_t|q_{t-1}) \right] \pi(\rho_1, q_1) \qquad (2)$$

where we initialized the product with the probabilities $\pi(\rho_1, q_1)$. In our case, $\rho$ is the radial distance $\rho = \sqrt{x^2 + y^2}$ from the projected bead center to the tether attachment point (because $p_{\text{bead}}$ is circularly symmetric).

The evaluation of (2) may at first seem prohibitively difficult: Typically we may have $M = 70\,000$ video frames, and hence $2^M$ terms in the sum over $\mathcal{Q}$! But closer inspection shows that (2) is the product of $M$ $2 \times 2$ matrices, which can be evaluated in order-$M$ steps. Namely, (2) can be rewritten as

$$P_{\text{tot,HMM}}(\mathcal{O}) = (1,1) \cdot \mathsf{T}_{\text{HMM}}(\rho_M) \cdots \mathsf{T}_{\text{HMM}}(\rho_2) \Pi(\rho_1) \qquad (3)$$

where

$$\mathsf{T}_{\text{HMM}}(\rho) = \begin{pmatrix} 1 - (\Delta t/\tau_{\text{LF}}) & (\Delta t/\tau_{\text{LB}}) \\ (\Delta t/\tau_{\text{LF}}) & 1 - (\Delta t/\tau_{\text{LB}}) \end{pmatrix} \begin{pmatrix} p_{\text{bead}}(\rho|q = 1) & 0 \\ 0 & p_{\text{bead}}(\rho|q = 2) \end{pmatrix} \qquad (4)$$

The row vector $(1, 1)$ is used in (3) to obtain a scalar probability, and $\Pi(\rho)_q = \pi(\rho, q)$ is regarded as a column vector. The first factor in (4) is just $T(q|q')$ regarded as a matrix; $\Delta t$ is a time step that is much smaller than $\tau_{\text{LF}}$ or $\tau_{\text{LB}}$.

With this insight, evaluating the expression and even optimizing it over the parameters $\tau_{\text{LF}}$ and $\tau_{\text{LB}}$ is not so prohibitive after all. A numerical issue does arise when dealing with long time traces, in that the value of $P_{\text{tot}}$ becomes very small. To handle this, we perform a normalization procedure after each time step [25] in the calculation of (3). For example, after the first $t$ multiplications we divide each of the two entries in the matrix by their sum $s_t$, thus keeping all of the terms in the multiplication close to unity. We maintain a running tally $S_t$ of the normalizing factors $\ln s_t$ so that by the end of the calculation of (3), the final value $S_M$ is the logarithm of the desired $P_{\text{tot}}(\mathcal{O})$.

One problem with the procedure outlined above, as mentioned in section 1.3, is that it neglects correlations in observed bead position due to the Brownian character of tethered particle motion. Equation (2) assumes that the "noise" (bead motion) has no memory. Although this may be a valid assumption for the noise in ion channels, it is certainly not the case for tethered particle motion, if the sample times are separated by less than the bead's diffusion time. And indeed, we found that HMM in this form was not able to determine reasonable lifetimes in experimental data (figure A1). Instead, the algorithm reports very short lifetimes, inconsistent with obvious looping events visible in the data. This spurious

short lifetime depends on the sampling rate of the measurement, and is present in data from control experiments with no looping.

Appendix A illustrates an *ad hoc* approach for addressing bead diffusion in an HMM analysis by thinning the data. While this may be acceptable in some circumstances, we found that a better approach is to modify the HMM procedure itself to include bead diffusion directly. Section 4 will show that the required modification is simple, and computationally no more difficult than the calculations sketched in this section.

## 4. DHMM method

### 4.1. Formulation

The previous section argued that the standard HMM approach cannot be applied to tethered particle motion because its starting point, (2), is not valid. First, even if the rate of loop formation is zero (as it is in the absence of any cI protein), there will still be correlations in the observed bead positions that the algorithm interprets as spurious transitions. In fact, the time series of bead positions itself has a Markov character: Each position $\mathbf{r}_t$ is drawn from a probability distribution $T_{\mathrm{unl}}(\mathbf{r}_t|\mathbf{r}_{t-1})$ depending only on $\mathbf{r}_{t-1}$, independent of the positions at prior times [39].

A second violation of (2) is apparent when we note that the HMM assumption that the loop formation/breakdown process is autonomous (independent of the observed bead positions) is also not valid. As an extreme example, we note that if at some time the bead position is observed to be so far from the attachment point that the tether is stretched nearly straight, then it is geometrically impossible for a loop to form and the momentary rate of loop formation must equal zero.

In short, the dynamics of the bead/tether system must be regarded as a single, extended Markov process, with a joint conditional probability $T_{\mathrm{DHMM}}(\mathbf{r}_t|\mathbf{r}_{t-1})_{q_t,q_{t-1}}$; we replace (3) by

$$P_{\mathrm{tot,DHMM}}(\mathcal{O}) = (1,1) \cdot T_{\mathrm{DHMM}}(\mathbf{r}_M|\mathbf{r}_{M-1}) \cdots T_{\mathrm{DHMM}}(\mathbf{r}_2|\mathbf{r}_1)\Pi(\mathbf{r}_1) \qquad (5)$$

Henceforth we abbreviate $P_{\mathrm{tot,DHMM}}$ as $P_{\mathrm{tot}}$.

Note that since we wish to capture the dynamics of the bead motion, we can no longer reduce our description of the bead center from its position vector $\mathbf{r}$ to its length $\rho$, because there are pairs of points with the same $\rho$ that are nevertheless spatially distant. Replacing $\mathbf{r}$ by $\rho$ would therefore discard some valuable, observed, information. We must, however, reduce from the full 3-dimensional position to its 2D projection when dealing with experiments in which $z$ is not observed. In this paper we will always make this reduction, so we simply write $\mathbf{r}$ for the projected position.

What should we now take for our transition matrix $\mathsf{T}_{\mathrm{DHMM}}(\mathbf{r}_t|\mathbf{r}_{t-1})_{q_t,q_{t-1}}$? It may seem that we have lost the most desirable feature of HMM, namely that the transition matrix was fully determined by the two fit parameters $\tau_{\mathrm{LF}}$ and $\tau_{\mathrm{LB}}$, along with functions known *a priori* (the two empirically observed pdfs $p_{\mathrm{bead}}(\rho|q)$). We need a correspondingly simple proposal for the matrix $\mathsf{T}_{\mathrm{DHMM}}$, which depends on *two* continuous quantities $\mathbf{r}, \mathbf{r}'$.

Our answer to this question is a heuristic proposal for $\mathsf{T}_{\mathrm{DHMM}}$. Although not rigorously justified, our proposal retains the properties of depending on empirically determined functions and two fit parameters $\tau_{\mathrm{LF}}, \tau_{\mathrm{LB}}$; moreover, it incorporates the diffusive character of bead motion, which was not included in (2). The empirical functions describing the tethered motion of the bead are $T_{\mathrm{unl}}(\mathbf{r}|\mathbf{r}')$ and $T_{\mathrm{loop}}(\mathbf{r}|\mathbf{r}')$ for the unlooped and permanently looped states, respectively. We first state our proposal, then discuss its meaning. Subsequent sections will show how to extract the empirical functions from control data, then how to apply the model to data on dynamic looping.

We assume that the matrix $\mathsf{T}_{\mathrm{DHMM}}$ takes the form

$$\mathsf{T}_{\mathrm{DHMM}}(\mathbf{r}|\mathbf{r}') = \begin{pmatrix} 1 - (\Delta t/\tau_{\mathrm{LF}})\Theta & (\Delta t/\tau_{\mathrm{LB}}) \\ (\Delta t/\tau_{\mathrm{LF}})\Theta & 1 - (\Delta t/\tau_{\mathrm{LB}}) \end{pmatrix} \begin{pmatrix} T_{\mathrm{unl}}(\mathbf{r}|\mathbf{r}') & 0 \\ 0 & T_{\mathrm{loop}}(\mathbf{r}|\mathbf{r}') \end{pmatrix} \quad (6)$$

In this formula, $\Theta \equiv 1 - \Theta(\rho - \rho_{\mathrm{max}})$ is a step function, equal to 1 if looping is geometrically permitted, and 0 otherwise. The value of $\rho_{\mathrm{max}}$ is to be obtained from experimental data, as described in section 4.3.

When looping is either forbidden or obligatory, (6) reduces to tethered Brownian motion in either the unlooped or looped states respectively (no dynamic looping). In other cases, we can think of the formula as describing an alternating series of transitions. First, the bead takes a diffusive step based on its current looping state (second matrix factor in (6)). Next, the tether updates its looping state, using probabilities that depend in a simple way on its current position (first matrix factor in (6)). Then the process repeats. If $\Delta t$ is much smaller than either the diffusion time or the loop formation/breakdown times, then we make no significant error by decomposing the process in this way. Note that (6) is properly normalized, as we can check by summing it over all final states $\mathbf{r}, q$.

## 4.2. Critique of DHMM approach

Although it is reasonable, our proposal does make some strong and perhaps naive assumptions about the looping process. Before turning to the implementation of the method, we now pause to make some of these assumptions explicit, and indicate how future experiments could help justify them. The material in this section will not be used later; the reader may wish to skip to section 4.3.

DNA loop formation involves the motion of the molecule through a high-dimensional space of shapes, driven by thermal motion, subject to a free energy landscape determined

by the molecule's elasticity. When binding sites on the proteins encounter each other, or the DNA's operator sequences, binding may ensue depending on the precision of their alignment and the relevant binding constants. Phrased in this way, it's clear that the calculation of DNA loop formation kinetics is very complicated. However, such *ab initio* calculations are not our goal.

Our goal is to develop a simple characterization for the looping behavior seen in TPM experiments, in a way that is also relevant for looping behavior *in vivo*, and that also lets us observe differences in behavior as we change system parameters. We imagine that, for a free DNA chain, polymer dynamics repeatedly brings binding sites into juxtaposition at some rate, with a certain probability that any such encounter leads to formation of a bound state. The product of the attempt rate and the probability per encounter is an average loop formation rate. If we suppose that each encounter's probability for binding is small, then it is reasonable to expect that overall loop formation (and breakdown) processes should be monomolecular reactions describable with simple exponential kinetics.

Turning from free DNA to the case of DNA tethering a large reporter bead, we first note that the presence of the bead does not by itself alter the thermal force fluctuations on the looping part of the DNA; these equilibrium fluctuations are determined by equipartition applied to the entropic elasticity of the semiflexible polymer chain. It is true that bead-surface repulsion can tend to stretch the DNA, altering the equilibrium constant for loop formation [17]. However, this entropic force falls off rapidly when the distance from the polymer's endpoint to the surface exceeds the bead diameter; it can be minimized by choosing small enough beads (or by replacing the bead by a functionalized colloidal gold particle [40]).

Notwithstanding the above remarks about equilibrium, we should expect that the large, sluggish bead will significantly alter looping *kinetics*. But precisely because the bead is slowly diffusing, whenever it is close to the attachment point of the DNA to the wall, it is likely to stay close for many milliseconds. During this time, the bead offers no significant obstacle to the same thermally-driven chain rearrangements that bring the operator sites of free DNA into juxtaposition. Hence, we may expect that when the bead is close to the wall attachment point, DNA loop formation will proceed as if the DNA were free in solution (or part of a larger bacterial genome)‡. In the contrary case (the bead is far from the attachment point), loop formation is geometrically forbidden. A similar argument suggests that loop breakdown kinetics should not be altered by the presence of the reporter bead.

‡ More precisely, we are assuming that the time required for the polymer tether to explore its available conformations is much less than the time for the bead to diffuse a significant fraction of its total range of motion.

Equation (6) embodies the above ideas, together with the idea that the bead wanders in and out of the range allowed for looping, subject to distributions $T_{\mathrm{loop}}$ and $T_{\mathrm{unl}}$ that themselves can be found from the observed behavior of permanently looped or unlooped tethers. Thus we expect that the parameters $\tau_{\mathrm{LF}}$ and $\tau_{\mathrm{LB}}$ appearing in (6) will, when the model is fit to TPM data, also give a good guide to looping rates for DNA free in solution. In contrast, simply applying the windowing/threshold method to data does not correct for the expected slowing-down of loop formation due to the presence of the bead. Indeed, applying that method to our data leads to a inferred $\tau_{\mathrm{LF}}$ values significantly slower than the one we will obtain in section 5 below.

Although the preceding paragraphs have argued that our approach is reasonable, certainly it is somewhat crude. For example, once the bead center is observed to be at a certain distance from the attachment point, this distance amounts to a stretching of the DNA chain. The expected rate for loop formation will be some decreasing function of this stretching, but not of course a step function, as we assumed in (6).

Another simplification we have made is to ignore the unobserved height variable $z$, in effect treating the bead motion as diffusion in two dimensions. Although in *free* Brownian motion all three coordinates perform independent random walks, in our problem the presence of the wall and tether couple $x, y$, and $z$. To some extent, our technique of obtaining $T_{\mathrm{unl}}$ and $T_{\mathrm{loop}}$ in (6) directly from control data will correct for this effect, but the criterion for loop formation to be possible really depends on the full 3D separation $\sqrt{x^2 + y^2 + z^2}$, not on $\rho = \sqrt{x^2 + y^2}$ as assumed in (6). A better analysis might treat $z$ as another hidden (unobserved) variable.

We also ignore rotatory Brownian motion of the bead. The ability of the tether to form loops actually depends on the distance between the two DNA end points. One endpoint, where the DNA attaches to the microscope slide, is fixed. We are using the bead center location as a proxy for the DNA-bead attachment point, but really the latter depends on both the former and also on the angular orientation of the bead. Again, a better analysis might treat this orientation as another hidden variable.

Our justifications for all three of the above simplifications are simply that *(a)* although the pdfs for projected bead position in the looped and unlooped states overlap partially, they are nevertheless fairly distinct, allowing reliable state identification even when we simplify some of the details in the model; *(b)* the projected-step distribution functions $T_{\mathrm{unl}}$ and $T_{\mathrm{loop}}$ that we extract from control-experiment data do have the qualitative form that we would expect for two-dimensional diffusion in an effective spring trap [39] (see section 4.3); and *(c)* changing our choice of the cutoff $\rho_{\mathrm{max}}$ in the analysis does not significantly change our inferred values of the rate constants (data not shown). Despite these encouraging observations, however, other experimental tests of the method would

certainly be desirable, for example, analyzing the kinetics of loop formation in identical tethers attached to different-sized beads, to check that similar values of $\tau_{\mathrm{LF}}$ and $\tau_{\mathrm{LB}}$ emerge. It may also be desirable to refine our mathematical model to account for some of the above issues; we leave these refinements to future work.

### 4.3. Implementation

*4.3.1. Obtaining step distribution functions*  Even granting the simplifications proposed in the preceding subsections, it would be a daunting task to determine the appropriate step distribution functions $T_{\mathrm{unl}}$ and $T_{\mathrm{loop}}$ appearing in (6) *a priori* (directly from theory). For example, bead–wall hydrodynamic interactions depend on the bead's height, which is not observed; the tether couples the unobserved bead orientational fluctuations to the observed position fluctuations; and so on. We circumvent these difficulties by empirically determining the $T_{\mathrm{unl}}$ and $T_{\mathrm{loop}}$ from experimental control data for the two states. After these functions are obtained, we confirm that the simple model of tethered-particle dynamics we construct from them indeed reproduces some nontrivial features of the real control data (figures 4 and 5). Finally we examine dynamic-looping data, and adjust the two remaining free parameters $\tau_{\mathrm{LB}}$ and $\tau_{\mathrm{LF}}$ in (6) until the log-likelihood, $\ln[P_{\mathrm{tot}}(\mathcal{O})]$, is a maximum.

In order to obtain $T_{\mathrm{unl}}(\mathbf{r}|\mathbf{r}')$ from the unlooped control data, we first note that this function must be symmetric under rotations of both $\mathbf{r}$ and $\mathbf{r}'$ about the attachment point by a common angle. Thus we only need to find this function for $\mathbf{r}'$ on the $\hat{\mathbf{x}}$-axis, at some radial distance $\rho'$. Starting from a time series for unlooped DNA (no repressor protein present), we thus select all of the points in the time series for which the bead center's distance from the anchor point, $\rho'$, lies in a particular range. Next, we find the rotation in the plane that brings $\mathbf{r}'$ to the $\hat{\mathbf{x}}$-axis, and apply that rotation also to $\Delta \mathbf{r} = \mathbf{r} - \mathbf{r}'$, the bead's vector displacement to its position on the following video frame. Finally, we build up a 2-dimensional histogram of the observed displacements $\Delta \mathbf{r}$, and normalize to obtain $T_{\mathrm{unl}}(\mathbf{r}|\mathbf{r}')$. We then repeat the process, producing histograms for all observed initial distances $\rho'$. Using data obtained from about 30 minutes of bead observation, we found that we could divide the observed range of $\rho'$ into 30 intervals and still have reasonable statistics in the histograms; figures 7–8 shows typical examples for two values of $\rho'$. We applied a similar procedure to the permanently-looped control data to obtain $T_{\mathrm{loop}}(\mathbf{r}|\mathbf{r}')$.

The step-probability distributions (figure 8) obtained in this way show that at small $\rho'$ there is no preferred direction for the next time step; for larger $\rho'$, the tether is stretched and exerts a restoring force on the bead, so the step distribution shows a bias to diffuse toward the attachment point (the $-\hat{\mathbf{x}}$ direction). We now seek a convenient analytical representation of these distributions, both for computing the likelihood function $P_{\mathrm{tot}}(\mathcal{O})$,
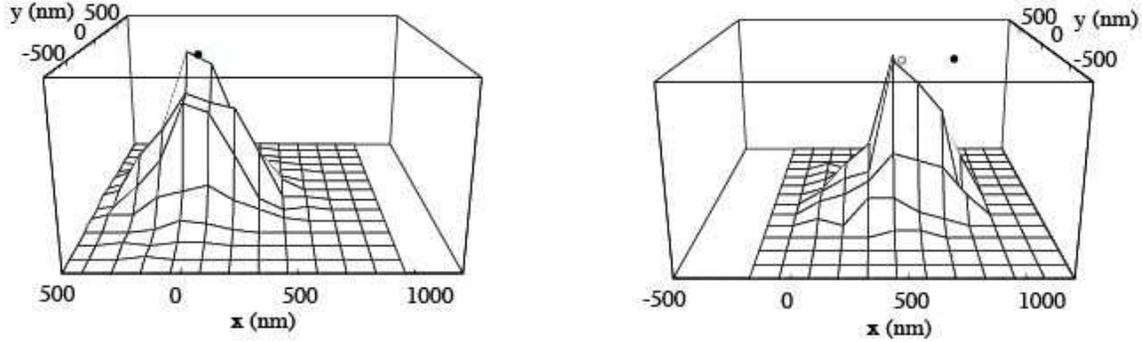
**Figure 7.** 2D histograms of unlooped control data at two values of $\rho'$ near (*left*) and far (*right*) from the anchor point (0,0) after rotation onto the $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ axes (see text). The dots on the upper x-y plane indicate the initial position $\rho'$ (*full circle*) and the mean midpoint of the final position $\mu(\rho')$ (*open circle*). For $\rho'$ near the anchor point the two dots coincide.
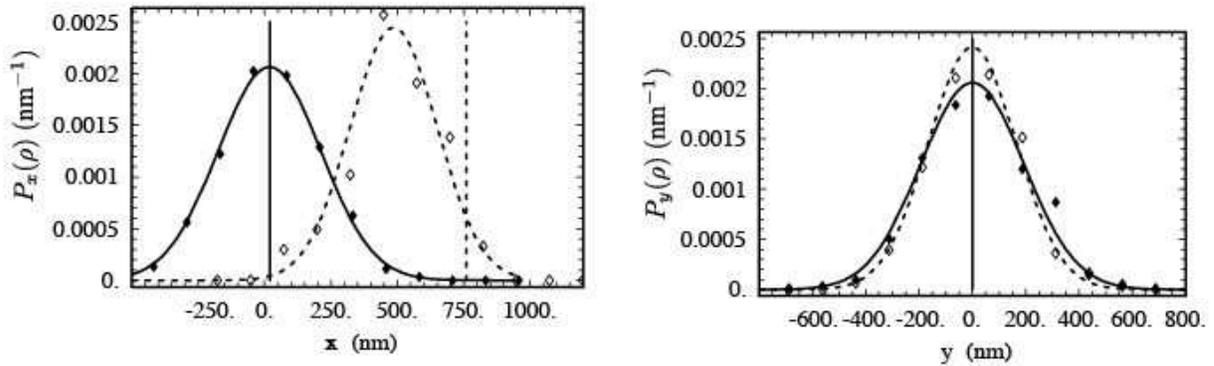


**Figure 8.** Projections of the two distributions in figure 7 onto the $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ axes, together with Gaussian distributions chosen to idealize them. *Left:* the *vertical lines* represent two choices for the initial bead position; *dots* represent the corresponding distributions of bead positions on the next video frame. Note the shift in the mean $\hat{\mathbf{x}}$ at larger $\rho'$ (*open*) compared to shorter $\rho'$ (*full*). *Right:* No such shift is observed in the $\hat{\mathbf{y}}$ direction.

and also for simulation purposes.

Each distribution is seen to be roughly a 2D Gaussian, with one principal axis along the radial direction to the attachment point. After rotating $\mathbf{r}'$ to lie along the $\hat{\mathbf{x}}$ axis as described above, the principal axes of the distribution are the $\hat{\mathbf{x}}$- and $\hat{\mathbf{y}}$-axes. The center point also lies on the $\hat{\mathbf{x}}$-axis, and is increasingly shifted from $\rho'\hat{\mathbf{x}}$ toward the anchor point $(0,0)$ as $\rho'$ increases. Accordingly, we can characterize $T_{\text{unl}}(\mathbf{r}|\mathbf{r}')$ for each fixed $\mathbf{r}' = (\rho',0)$ by finding its mean $\langle x \rangle$ and the variances in the $x$ and $y$ directions. The mean $\langle y \rangle$ equals zero (see for example figure 8), as it must by rotational invariance. Thus we
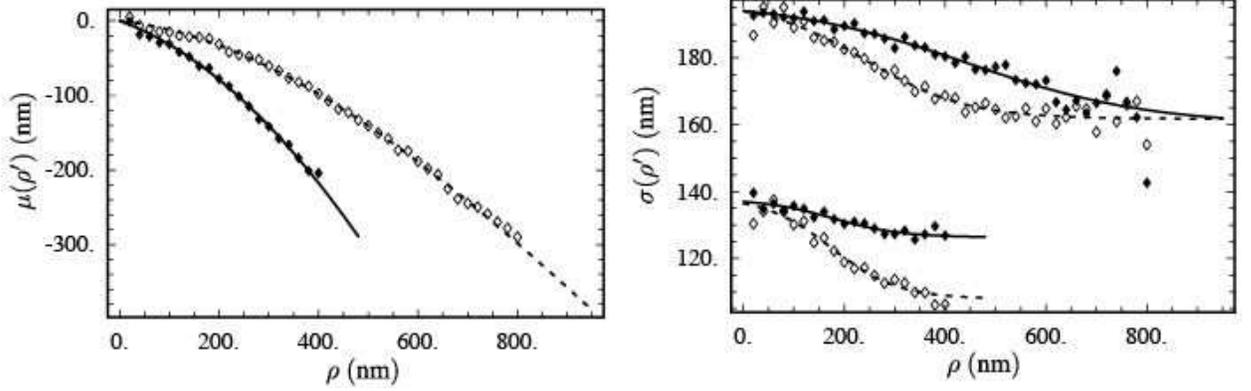
**Figure 9.** Empirical fit functions for the mean (*left*) and variance (*right*) of the 2D histograms (see e.g. figure 7) for experimental control data corresponding to unlooped (*solid* symbols) and looped (*open* symbols) tether states. Corresponding fit parameters for the functions of (7) are located in Appendix B.

seek three functions of $\rho'$ to characterize the histograms: the mean $\mu(\rho') \equiv \langle x \rangle_{\rho'}$, and the variances $\sigma_x^2(\rho')$ and $\sigma_y^2(\rho')$. Examples of these functions for illustrative values of $\rho'$ appear in figure 9.

Again, the mean shift function $\mu$ reflects the drift of the bead toward the central point under the influence of a restoring force from the tether's entropic elasticity. As expected, it vanishes at $\rho' = 0$ and becomes more negative with increasing $\rho'$. (The variances have weaker dependences on $\rho'$.) We constructed three interpolating functions to represent our results, which we took to be a 3rd-order polynomial for $\mu$ and sigmoids for $\sigma_{x,y}$ (see figure 9):

$$
\begin{aligned}
\mu_x(\rho) &= a_0 + a_1 \rho + a_2 \rho^2 + a_3 \rho^3 \\
\sigma_x^2(\rho) &= b_0/(1 + e^{(\rho - b_1)/b_2}) + b_3 \\
\sigma_y^2(\rho) &= c_0/(1 + e^{(\rho - c_1)/c_2}) + c_3
\end{aligned} \tag{7}
$$

Using these fit functions, we can represent the observed step probabilities starting from the point $\mathbf{r}' = (\rho', 0)$ as the product of 1D Gaussian distributions in $x$ and $y$:

$$
T_{\mathrm{unl}}(\mathbf{r}|\mathbf{r}') = G_x(x|\rho') \cdot G_y(y|\rho') \quad \text{where} \tag{8}
$$

$$
G_x(x|\rho') = (2\pi\sigma_x^2(\rho'))^{-1/2} \exp\left( \frac{-(x - \mu(\rho'))^2}{2\sigma_x^2(\rho')} \right) \tag{9}
$$

$$
G_y(y|\rho') = (2\pi\sigma_y^2(\rho'))^{-1/2} \exp\left( \frac{-y^2}{2\sigma_y^2(\rho')} \right) \tag{10}
$$

For arbitrary $\mathbf{r}'$ (not necessarily on the $\hat{\mathbf{x}}$ axis), we evaluate the probability by rotating $\mathbf{r}'$ to the $\hat{\mathbf{x}}$-axis, rotating $\mathbf{r}$ by the same amount, and evaluating (8) on the components of the

rotated $\mathbf{r}$.

The procedure summarized in (8–10) is conceptually simple. However, we chose to improve the accuracy of our calculations by using a small elaboration. We noted that, like any Gaussian, the distribution defined above is nonzero for any $x$ and $y$. In reality, however, the DNA tether sets an absolute limit on $\rho$ beyond which the probability must be exactly zero. Not surprisingly, we found that following the procedure outlined above yielded simulated time series that occasionally violated this limit. Although the effect of this error may be minor for the unlooped step distribution, we reasoned that for the looped distribution it could interfere with looping state identification.

Accordingly, we modified our formula for $T_{\text{unl}}(\mathbf{r}|\mathbf{r}')$ to account for the limit in an approximate (and computationally inexpensive) way: We replaced (9) by a truncated Gaussian function. That is, we took $G_x(x|\rho')$ to be zero for $x > \rho_{\text{max}}$, and a Gaussian with modified parameters for $x < \rho_{\text{max}}$. The modified parameters were chosen in such a way that the truncated Gaussian would again have the mean $\mu(\rho')$ and variance $\sigma_x^2(\rho')$ shown in figure 9. That is, for each value of $\rho'$, the $\mu$ and $\sigma_x^2$ determined empirically from the data were not used directly; instead we found a new Gaussian, with modified parameters $\tilde{\mu}$ and $\tilde{\sigma}_x^2$, which has mean $\mu$ and variance $\sigma_x^2$ when the probability of values greater than $\rho_{\text{max}}$ is set to zero. Details of this transformation appear in Appendix B.

*4.3.2. Optimization*   Initially, the optimum lifetimes were found by evaluating $P_{\text{tot}}(\mathcal{O})$ on an evenly-spaced logarithmic grid of values for $\tau_{\text{LF}}$ and $\tau_{\text{LB}}$, and the maximum $P_{\text{tot}}^*(\mathcal{O})$ was found. The resulting likelihood surface is smooth (figure 10), so the peak likelihood can be determined more precisely by fitting a 2D quadratic in the neighborhood of the optimum lifetimes. The uncertainty of the optimum lifetimes corresponding to $\ln[P_{\text{tot}}^*(\mathcal{O})] - 2$, i.e., enclosing 97% of the probability, were estimated along the principal axes of this 2D quadratic to account for any correlation between the estimated lifetimes. In order to facilitate this iterative process, an automated simplex solver routine was implemented to find the maximum [41].

*4.3.3. Simulation strategy* Simulations of bead motion, with and without dynamic looping, were performed in *Mathematica* to test the DHMM model. Each step of the simulation (*a*) first determined whether or not to remain in the current looped state, and then (*b*) the next spatial position was determined appropriate for the particular loop state.

In more detail, (*a*) If the initial state was looped, a pseudorandom number was used to determine whether to transition to the unlooped state with probability $\frac{\Delta t}{\tau_{\text{LB}}}$. If the initial state was unlooped, and if $\rho < \rho_{\text{max}}$, then a transition to the looped state was allowed with probability $\frac{\Delta t}{\tau_{\text{LF}}}$. Next (*b*) a $(\Delta x, \Delta y)$ pair was drawn from the appropriate distribution
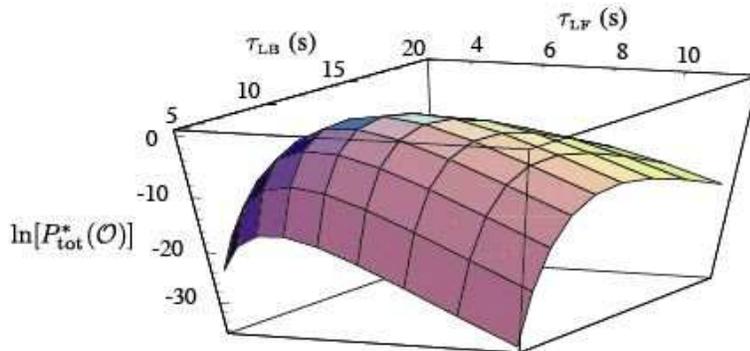
**Figure 10.** Evaluation of $\ln[P^*_{\mathrm{tot}}(\mathcal{O})]$ on a logarithmically-spaced grid of $\tau_{\mathrm{LF}}, \tau_{\mathrm{LB}}$ lifetimes corresponding to data from figure 2.

obtained in section 4.3.1. That is, $\Delta y$ was Gaussian distributed, and similarly for $\Delta x$ except that steps resulting in $\rho > \rho_{\mathrm{max}}$ were discarded and the step repeated in order to achieve a truncated Gaussian as discussed earlier.

## 5. Results and discussion

Section 4.3 outlined how we determined $T_{\mathrm{unl}}(\mathbf{r}|\mathbf{r}')$ and $T_{\mathrm{loop}}(\mathbf{r}|\mathbf{r}')$ from the control data, then used these functions to simulate tethered particle motion. In this section we compare one-state simulations with control data and two-state simulations with dynamic looping data. We also use DHMM to locate a change in looping dynamics, possibly indicating a change in protein occupancy of one or both of the *lambda* operators.

### 5.1. One-state modeling

To validate our approach to equilibrium, tethered Brownian motion, we wrote a simple simulation using the step-distribution functions obtained from adjacent video frames (section 4.3) and compared the resulting trajectories with the experimental control data. The resulting simulated time series $\{\rho_t\}$ are difficult to distinguish from actual data by inspection (not shown), so we compared the equilibrium properties of the motion using the radial probability distributions (figure 4) and the dynamic properties using the autocorrelation function (figure 5). Both of these non-trivial checks agree fairly well with the actual data, although for unknown reasons the equilibrium distribution of the longer tether is captured better by the model than the shorter tether.

We also applied our two-state DHMM approach to permanently-unlooped experimental control data, to see if the algorithm would incorrectly report any looping

transitions. Instead, the algorithm correctly reported loop-formation times that were proportional to the length of the data set (that is, consistent with infinity), and loop-breakdown times approaching the sampling interval (that is, consistent with zero). Simulations of unlooped motion behaved similarly. As expected, DHMM applied to permanently-looped experimental control data and simulations did not detect any false loop breakdown events with the trends for loop breakdown and formation lifetimes exchanged from the unlooped case. Another consistency check was to verify that the step distribution functions originally calculated from the experimental control data section 4.3.1 were the same for the simulations.

Finally, we found the maximum likelihood transition sequence corresponding to the optimal lifetime values [25]. No false looping transitions were reported for either of the control data sets, indicating that our rate for false positives for loop formation and breakdown is low.

## 5.2. Dynamic looping

We applied our two-state DHMM algorithm to the part of the time series in figure 2 that displays dynamic looping (the region between 650 and 1750 seconds). The algorithm reported $\tau_{\mathrm{LF}} = 5.8\,\mathrm{s} \pm 25\%$ and $\tau_{\mathrm{LB}} = 9.9\,\mathrm{s} \pm 25\%$; it also determined the most-likely sequence of state transitions (see Fig. 2 of [1]). To see whether our model really detects dynamic transitions between the two tether lengths in this parameter regime, we simulated a 20-minute dynamic looping data set with these same lifetime values, applied DHMM to the simulated data, and compared the reported lifetimes to those we had input to the model. We also know each of the transitions in the simulated data, so we could check the ability of our algorithm to reconstruct these. Indeed, applying DHMM to simulations yielded reported lifetimes that agree with these values within uncertainty: The agreement was within 20% of the known answer (see table 1). The maximum-likelihood state transition sequence corresponding to the optimum lifetimes successfully detects $\sim 85$–$90\%$ of the known transitions, with missing events usually less than $\sim 1\,\mathrm{s}$ in duration.§

We also simulated looping with lifetimes different from those seen in the experimental data, see Appendix C. 10 additional simulations, similar to those reported in table 1, indicate that scatter in the lifetimes determined by DHMM is no worse than

§ Note however that, like any Hidden Markov method, DHMM is not specifically sensitive to missing events. Indeed in DHMM, event identification is performed *after* the lifetimes are determined, and thus does not affect the determination of the lifetime. Instead of binning identified events and fitting the resulting histogram, DHMM directly maximizes the likelihood that a function describes the entire data set. If the time scale of events is too short compared to other time scales in the problem, DHMM reports that fact via its estimates of errors in the fit parameters.

**Table 1.** Comparison of results for three simulations, their average, and data (from figure 2). The inferred lifetimes for the data were used as input for the simulations. All trials had a total number of points $M = 29\,424$. The second number in the event column is the (known) number of events. Lifetimes are in seconds and other terms are nondimensional.

|  | # Events | $\tau_{\mathrm{LF}}$ | $\tau_{\mathrm{LB}}$ | $\ln[P_{\mathrm{tot}}(\mathcal{O})]$ | $\frac{1}{M}\ln[P_{\mathrm{tot}}(\mathcal{O})]$ |
|---|---|---|---|---|---|
| Data | 50 | 5.8 | 9.9 | $-374\,089$ | $-12.714$ |
| Sim. #1 | 68/75 | 4.9 | 9.2 | $-376\,951$ | $-12.811$ |
| Sim. #2 | 57/64 | 6.4 | 9.3 | $-378\,173$ | $-12.853$ |
| Sim. #3 | 62/77 | 5.6 | 8.3 | $-378\,228$ | $-12.855$ |
| Average Sim. | 62.3/72 | 5.6 | 9.0 | $-377\,784$ | $-12.840$ |

the minimum expected from variability due to the finite sample size (data not shown). Simulations with longer lifetimes $\tau_{\mathrm{LF}} = \tau_{\mathrm{LB}} = 20\,\mathrm{s}$ and a total time of $\approx 60\,\mathrm{minutes}$ to give more transitions were similarly successful (agreement of lifetimes within 10% and 92% success in detecting events, data not shown).

As discussed in section 3, in both the threshold method and traditional HMM the time resolution depends on how the data are analyzed (i.e., window size and degree of thinning respectively). To demonstrate the robustness of the DHMM method, we subdivided our data set (taken at a frame rate of $(20\,\mathrm{ms})^{-1}$) into two subsets with $\Delta t = 40\,\mathrm{ms}$, and also into four subsets with $\Delta t = 80\,\mathrm{ms}$ and computed lifetimes for all of these subsets. All lifetimes agree within the uncertainty of the method (data not shown; note that (7–10) must be re-calculated for different $\Delta t$).

## 5.3. Detection of very long-lived state transitions

So far, we have used the peak of the likelihood function $P_{\mathrm{tot}}(\mathcal{O})$ and its vicinity to determine the optimum looping lifetimes and their uncertainty. We can take further advantage of the likelihood function to assess the uniformity of the dynamics. Our motivation for undertaking this study is that in a real DNA-looping system there are certainly more than two discrete states: E.g., individual repressor proteins can bind and unbind to their operator sites [14]. Indeed, the data studied in this paper (from [1]) came from a system with two sets of three operators, leading to a large set of potential occupancy patterns. Presumably the obvious change in behavior in figure 2 at $t = 700\,\mathrm{s}$ reflects the arrival of another repressor at an operator site, enabling loop formation. But there seems also to be a less obvious transition in the data around $t = 1250\,\mathrm{s}$, from one looping regime to another one with different kinetics. Can DHMM help us to locate such changes objectively?
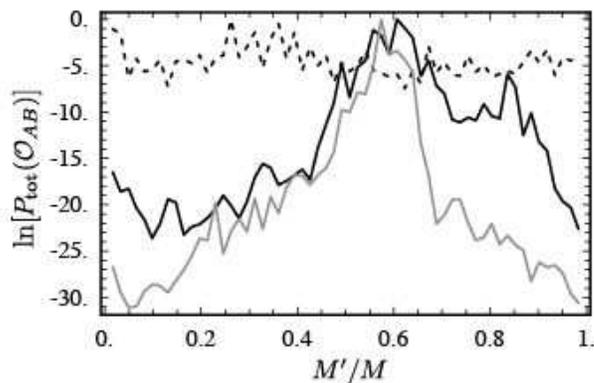
**Figure 11.** See text. The peak in $\ln P_{\text{tot}}(\mathcal{O}_{AB})$ (*solid line*) suggests that the data from figure 2 is composed of two regions at $M' \simeq 61\%M$ with different kinetics. The *gray/dotted* lines are the results from simulated data with/without a transition at $M' \simeq 61\%M$, respectively. For comparison purposes, each likelihood has had its peak vertically shifted to 0 by subtracting a constant.

We reasoned that if the data had uniform, two-state looping behavior for the entire recording, then the log-likelihood of the whole would be equal to the sum of the log-likelihoods of its parts, and also the best-fit lifetimes should come out roughly equal for each part separately as for the whole. To test this, we adapted a procedure used by Ref. [42]: We again restricted the data in figure 2 to the region after $650\,\text{s}$, but this time we also divided the data into two regions, $A$ from 1 to $M'$ and $B$ from $M'$ to $M$, and then separately applied DHMM on the two regions to determine $\ln P_{\text{tot}}(\mathcal{O}_{AB}) \equiv \ln P_{\text{tot}}(\mathcal{O}_A) + \ln P_{\text{tot}}(\mathcal{O}_B)$. The resulting $\ln P_{\text{tot}}(\mathcal{O}_{AB})$ has a single, very sharp peak at a particular value of $M'$ (figure 11). In agreement with visual inspection, the optimal value of $M'$ is roughly two-thirds of the way through the retained data, that is, around $t = 1300\,\text{s}$ in figure 2. Moreover the reported optimal lifetime values in the two subsets are quite different: $(\tau_{\text{LF}}, \tau_{\text{LB}})_A \simeq (10\,\text{s}, 20\,\text{s})$ and $(\tau_{\text{LF}}, \tau_{\text{LB}})_B \simeq (3\,\text{s}, 4\,\text{s})$. When we repeated the procedure with simulated data, with a transition at $M' = 0.6M$ from looping with these $(\tau_{\text{LF}}, \tau_{\text{LB}})_A$ and $(\tau_{\text{LF}}, \tau_{\text{LB}})_B$ values, we found a peak in $\ln P_{\text{tot}}(\mathcal{O}_{AB})$ that was very similar to the one found with the experimental data (figure 11). In contrast, when we applied the procedure to simulated constant, two-state dynamics, we found as expected that $\ln P_{\text{tot}}(\mathcal{O}_{AB})$ was insensitive to $M'$.

## 6. Conclusion and outlook

Our goal was to enhance the ability of tethered particle experiments to study DNA looping kinetics *in vitro*. For illustration, we used a simplified 2-state model that includes the

one looped and one unlooped tether state, i.e. we did not account for the various possible occupancies of the repressor protein on the DNA binding sites, but our method can readily be extended to include such details. Other DNA looping systems, e.g., *lac*, could also be studied with DHMM, including multiple-state models to study the multiple length looped states recently observed [14].

Previously, threshold methods were applied to the data to quantify the kinetic rates between various looped and unlooped states. This method involves filtering the data to extract the transitions from the noisy diffusive motion of the bead, then fitting a (single or double) exponential to the tail of a histogram of the dwell times. Both filtering and histogramming discard potentially useful information; moreover, at least in the system we studied, the choice of filter window can influence the reported results. Our DHMM method avoids any such steps.

We reasoned that hidden Markov modeling would be a good way to learn about the hidden conformation of the DNA tether from the observed motion of the bead, because the observed motion can be statistically quantified and simple models used to describe the unobserved state of the tether. We represent the uninteresting diffusive motion empirically from control experiments, and then model the looping assuming exponential kinetics with unknown parameters corresponding to the loop formation and breakdown lifetimes. Then we maximize the likelihood that the experimental data (or a simulation) came from our proposed model. Our method is implemented in a computationally efficient code, available in the online supplement to [1]. Since there is no filtering and no binning of the data in DHMM, the kinetic parameters can be determined unambiguously. If desired, the most-likely transition state sequence can also be determined [1].

It is easy to obtain unlooped control data to train the DHMM algorithm by omitting repressor protein during the experiment; however looped control data can be more challenging. For the *lambda* system considered here, infrequent yet long-lived looped states made this a relatively simple task. In other systems alternatives exist, including separate experiments on constructs with shorter length tethers corresponding to the expected looped length or, if available, mutant repressor proteins with stronger affinity for DNA that result in permanently looped tethers. In our system, we verified the robustness of DHMM to the model of the looped state by varying the fit parameters in (7) by $\pm 10\%$ and noting that $\tau_{\mathrm{LB}}$ and $\tau_{\mathrm{LF}}$ remained unchanged (data not shown).

Another advantage of DHMM is its ability to, at least partially, compensate an experimental bias inherent in the tethered particle method: Loops cannot form between successive measurements if the DNA is in an extended conformation due to the time for the bead to diffuse to a location closer to the anchor point. This effect inflates the observed loop formation times relative to the case of interest (free DNA in solution); indeed, in our

simulations, where the lifetimes are known *a priori*, we found that this effect inflates $\tau_{\mathrm{LF}}$ by $\sim 30\%$. Our DHMM model allowed us to compensate for it, by allowing loops to form only when $\rho < \rho_{\mathrm{max}}$.

The recent incorporation of single–particle tracking into the TPM method was essential for us, because it allows rapid and precise measurements of bead position that are required for DHMM. In particular, the ability to simultaneously track *multiple* tethered beads is helpful for removing instrumental drift [16]. Future experiments could in principle remove drift from the data entirely by simultaneously tracking a fixed, fiducial marker object.

We applied DHMM to one illustrative experimental dataset of *lambda* DNA with 200 nM cI and determined loop formation and breakdown lifetimes of ~6 and 10 s respectively. One surprisingly biologically relevant application of this work is that, at physiological [cI] corresponding to the lysogenic state, the loop is not permanently closed. We also found it interesting that these lifetimes are neither representative of all the beads that were observed in [1] (data not shown), nor even for the entire observation time of any single bead (see figure 2). Prior to adding cI, most beads have nearly identical tethered diffusive motion; however, after addition of cI, the kinetics of looping varied widely. Some beads were mostly unlooped with occasional looping events, some beads were the inverse of this, others showed long periods of dynamic looping, and some like figure 2 seemed to show very sharply-defined changes between long-lived (often $> 10$ min) regimes of homogeneous behavior. One hypothesis to explain these long-time looping trends is that the occupancy of cI protein among the 6 *lambda* binding sites changes, resulting in periods with more or less stable loops. For example, when all 6 sites are occupied a very stable loop might form, whereas 4-sites occupancy could result in a less stable, but still detectable, loop. Future experiments with fewer operators, or perhaps fluorescent cI protein, could be used to test this hypothesis directly.

One technique for considering such complex kinetic scenarios is to use a more elaborate state diagram; however, we explored a different approach that might be appropriate if cI proteins are binding and unbinding on a time scale much slower than loop formation, as we indeed observed. We reasoned that in this case, DNA looping data could be adequately represented by a concatenation of 2-state models, each with different kinetics, rather than by a far more elaborate model with many states. There appear to be three such regions in the data shown in figure 2. We removed the first, mostly unlooped region, to focus on the faster dynamics of the later region, which we split in all possible ways into two subregions. We analyzed each subregion separately using DHMM, and found the partition that resulted in the highest total likelihood, which was also much higher than if the two regions were assumed to be one homogeneous kinetic regime. The sharp

peak in figure 11 indicates that DHMM is a sensitive tool to localize such subtle transitions in time.

## Acknowledgments

## 7. Glossary and notation

See table 2 for mathematical symbols.

Tethered Particle Motion (TPM): Class of biophysical experiments that infer the properties of a polymer tether, such as DNA, by tracking the position of a $\mu$m-size bead that is attached to one end of the tether.

DNA Looping: protein-stabilized loops of DNA that form and break when protein binds to specific sites along the DNA in vivo and in vitro.

Diffusive Hidden Markov Model (DHMM): a modification of traditional hidden Markov methods that accounts for the diffusive properties of the TPM particle to determine kinetic rates for DNA loop formation and breakdown in single molecule experiments of DNA looping.

Threshold Method: an analysis method for determining kinetic rate constants from a signal by identifying transitions between states, and then fitting the dwell times between these state transitions to a model (such as an exponential).

Window: The width in time of the filter applied to a noisy time trace in order to elucidate the states for identification in the Threshold method.

| | |
|---|---|
| $\mathbf{r} = (x, y)$ | drift-subtracted, projected bead center position, relative to anchoring point |
| $\rho, \rho_{\max}$ | length of $\mathbf{r}$, and its maximum observed value |
| $z$ | height of tethered particle |
| $t, \Delta t$ | time and time difference, measured in units of $40\,\mathrm{ms}$ |
| $M;\ M' < M$ | total number of data points in a time series; and in a selected subset |
| $\mathcal{O} = \{\mathbf{r}_t\}$ | time series of observed $\mathbf{r}$ |
| $q, \mathcal{Q} = \{q_t\}$ | discrete variable describing DNA looping state, and corresponding time series |
| $P(\mathbf{r}_t, q_t)$ | joint probability of bead position and looping state at time $t$ |
| $\pi(\mathbf{r}, q)$ | same as $P(\mathbf{r}_1, q_1)$ |
| $p_{\mathrm{bead}}(\mathbf{r}|q)$ | pdf of bead position given looping state, neglecting diffusive correlations in $\mathbf{r}$ |
| $T(q|q')$ | looping state transition matrix, neglecting this distribution's dependence on observed position |
| $\tau_{\mathrm{LF}}, \tau_{\mathrm{LB}}$ | time constants for loop formation and breakdown (defined in (4,6)) |
| $\tau_{\mathrm{diff}}$ | diffusion time for bead to traverse its range of motion |
| $W$ | filter time constant used in windowing/threshold method (section 3.1) |
| $P_{\mathrm{tot}}(\mathcal{O}), P_{\mathrm{tot}}^*(\mathcal{O})$ | predicted likelihood for entire time series in a given model (5), and its optimum value |
| $T_{\mathrm{unl}}(\mathbf{r}|\mathbf{r}'), T_{\mathrm{loop}}(\mathbf{r}|\mathbf{r}')$ | step probability distributions for unlooped and permanently looped tethers respectively |
| $\mathsf{T}_{\mathrm{HMM}}(\rho),\ \mathsf{T}_{\mathrm{DHMM}}(\mathbf{r}|\mathbf{r}')$ | $2 \times 2$ matrices defined in (4, 6) |
| $\mu_x, \sigma_x^2, \sigma_y^2$ | functions of $\rho$ characterizing $T(\mathbf{r}|\mathbf{r}')$ (see (7)) |
| $G_x(x|\rho),\ G_y(y|\rho)$ | gaussian distributions chosen to represent $T(\mathbf{r}|\mathbf{r}')$ (see (8)) |
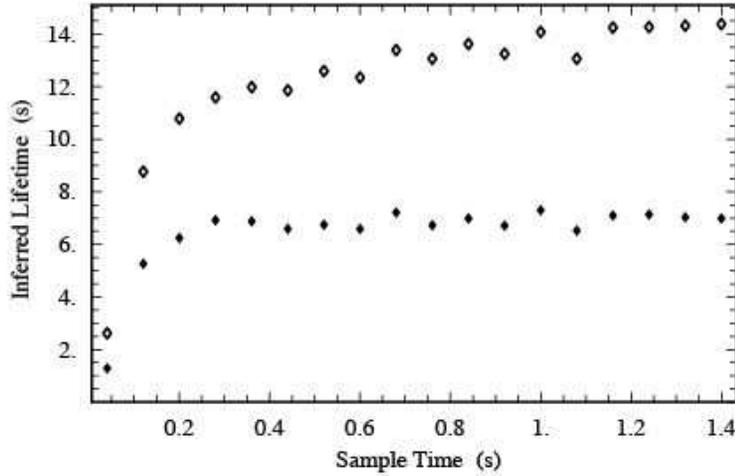
**Figure A1.** Hidden Markov analysis of the data from figure 2 illustrate how the diffusive time scale of the bead and tether dominate the inferred lifetimes when the measurement sampling interval is faster than the time scale for bead/tether diffusion. If the sampling time is artificially slowed down by thinning the data then the inferred lifetimes plateau approximately to the values obtained by DHMM (*solid/open* symbols refer to the inferred lifetimes for loop formation/breakdwon). For a thinning of '2' all of the odd points are separated from the even and then these two data sets are concatenated to reconstitute the original length trace. An analogous process was applied for higher order thinning. We verified (data not shown) that HMM analysis of the concatenated recording gave results approximately equal to the average of the results if each thinned trace was considered individually.

## Appendix A. Bead diffusion in standard HMM

As discussed in section 3.2, applying the standard hidden Markov analysis (2) to tethered particle experiments resulted in unrealistically short lifetimes for loop formation and breakdown that are on the same order as $\tau_{\text{diff}}$. The most probable looping sequence for $\mathcal{Q}$ corresponding to these fast lifetimes can be calculated (by a method described in [25]), and is consistent with the bead moving diffusively between regions of large and small $\rho$, effectively masquerading as very fast looping events. That is, the spurious reported transitions reflect the fact that successive video frames are not really independent measurements of the underlying tether state.

One can reduce this problem by thinning the data, thereby decreasing the influence of diffusion and making successive points more independent of each other. Repeating the HMM analysis, we find that the reported lifetimes increase. If this thinning process is repeated, the calculated lifetimes eventually plateau to a value roughly consistent with the

time scale of the looping events as identified by eye (figure A1). The looping sequence corresponding to these longer lifetimes also agrees with the looping events observed by eye in the raw data.

## Appendix B. Implementation details

Section 4.3.1 describes how we summarized empirical data, like those in figure 7, with a step-distribution function $G(x, y|\rho')$. Ideally $G$ should be chosen to be a function that vanishes when $x^2 + y^2$ exceeds $(\rho_{\max})^2$, and falls smoothly to zero as that boundary is approached. To make our calculations tractable, we instead simplified by taking $G$ to be the product of a cutoff, shifted, 1D-Gaussian in $x$ times an ordinary Gaussian in $y$. Examination of many graphs like figure 7 convinced us that this simplification adequately represented our empirical histograms. Moreover, we reasoned that, since we rotated axes to make the initial position lie along the $x$-axis, an extra excursion along $x$ was more likely to violate the tether condition than one along $y$. Perhaps most important, we checked that small changes to our choice of the empirical function $G$ have little effect (see section 6).

To implement efficient calculation of the truncated Gaussian $G_x(x|\rho')$, we evaluate a look–up table for the Gaussian with mean, variance, and normalization ($\tilde{\mu}(\rho'), \tilde{\sigma}^2(\rho')$ and $\tilde{N}$) such that when $\rho > \rho_{\max}$ this Gaussian is zero and satisfies the mean and variance ($\mu(\rho')$ and $\sigma_x^2(\rho')$) determined empirically from data in [1]:

$$\int_{-\infty}^{\rho_{\max}} dx \frac{1}{\tilde{N}} e^{-(x-\tilde{\mu})^2/(2\tilde{\sigma}^2)} = 1, \quad \int_{-\infty}^{\rho_{\max}} dx \frac{x}{\tilde{N}} e^{-(x-\tilde{\mu})^2/(2\tilde{\sigma}^2)} = \mu(\rho') \quad \text{(B.1)}$$

$$\int_{-\infty}^{\rho_{\max}} dx \frac{x^2}{\tilde{N}} e^{-(x-\tilde{\mu})^2/(2\tilde{\sigma}^2)} = \sigma_x^2(\rho') \quad \text{(B.2)}$$

Such look-up tables were evaluated at 100 values of $\rho'$ for both (unlooped, looped) tether states using (7) and parameters: a0=(0,0), a1= (-0.068, -0.238), a2= (-5.0e-4, -7.9e-4), a3 = (1.52e-7, 6.30e-8), b0= (35.1,30.6), b1=(161.75,107.95), b2=(242.3,173.8), b3=(100.16,66.42), c0=(37.11, 11.43), c1=(159.8,126.3), c2=(444.88, 177.06), c3=(180.72,67.86), where $\rho$ is measured in nm.

## Appendix C. DHMM simulations

We set up a $5 \times 5$ log-spaced grid of different combinations of lifetimes centered on the optimum lifetimes we obtained from the experimental data. Then we generated simulated data sets using those assumed lifetimes, and applied DHMM, to see how well it could extract them (figure C1). The agreement over the range ($1/4\times$ to $4\times$) was good to within the error bars estimated by the curvature of the log-likelihood function (figure 10). These
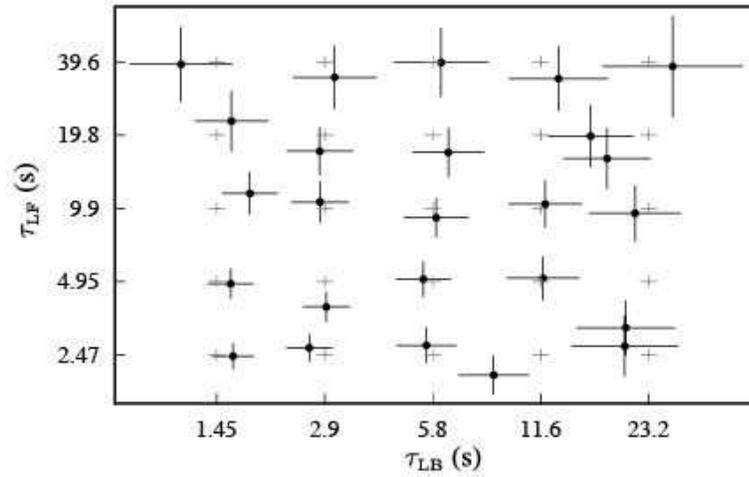
**Figure C1.** 25 simulations were performed with input lifetimes $\tau_{\mathrm{LF}}$ and $\tau_{\mathrm{LB}}$ distributed on a log-spaced grid *(gray crosses)* centered at the optimum lifetimes that we obtained in the text from experimental data. Each simulated data set had length $M$ equal to that of the experimental data. Next, DHMM was performed for each simulated data set, to determine the most-likely lifetimes *(black symbols)*.

error bars in part reflect the finite sample size $M$, which is often limited by experimental considerations. To check that last assertion, we generated 10 simulated data sets all with the same "true" lifetimes and of the same length $M$ as the experimental data discussed in the paper. Then we found the variation in the lifetimes deduced knowing the transition times, which represents the minimal variation, due to finite sample size. Finally we calculated the best-fit lifetimes from the simulated data using DHMM, and found the variation was similar to this minimal value (data not shown).

## References

[1] J. F. Beausang, C. Zurla, C. Manzo, D. Dunlap, L. Finzi, and P. C. Nelson. DNA looping kinetics analyzed using diffusive hidden Markov model. *Biophysical Journal*, 92:L64–L66, 2007. Explicit *Mathematica* code implementing our method can be found in the online supplement `http://www.biophysj.org/cgi/content/full/biophysj.107.104828/DC1`.

[2] D. A. Schafer, J. Gelles, M. P. Sheetz, and R. Landick. Transcription by single molecules of RNA–polymerase observed by light-microscopy. *Nature*, 352(6334):444–448, 1991.

[3] H. Yin, R. Landick, and J. Gelles. Tethered particle motion method for studying transcript elongation by a single RNA–polymerase molecule. *Biophysical Journal*, 67(6):2468–2478, 1994.

[4] L. Finzi and J. Gelles. Measurement of lactose repressor-mediated loop formation and breakdown in single DNA-molecules. *Science*, 267(5196):378–380, 1995.

[5] J. Gelles, H. Yin, L. Finzi, O. K. Wong, and R. Landick. Single-molecule kinetic-studies on DNA-transcription and transcriptional regulation. *Biophysical Journal*, 68(4):S73–S73, 1995.

[6] G. Zocchi. Force measurements on single molecular contacts through evanescent wave microscopy. *Biophys. J.*, 81:2946, 2001.

[7] K. M. Dohoney and J. Gelles. $\chi$-sequence recognition and DNA translocation by single RecBCD helicase/nuclease molecules. *Nature*, 409(6818):370–374, 2001.

[8] F. Vanzi, S. Vladimirov, C. R. Knudsen, Y. E. Goldman, and B. S. Cooperman. Protein synthesis by single ribosomes. *RNA*, 9:1174–1179, 2003.

[9] M. Singh-Zocchi, S. Dixit, V. Ivanov, and G. Zocchi. Single-molecule detection of DNA hybridization. *Proc. Natl. Acad. Sci. USA*, 100:7605, 2003.

[10] N. Pouget, C. Dennis, C. Turlan, M. Grigoriev, M. Chandler, and L. Salomé. Single-particle tracking for DNA tether length monitoring. *Nucl. Acids Res.*, 32:e73, 2004.

[11] S. F. Tolic-Norrelykke, A. M. Engh, R. Landick, and J. Gelles. Diversity in the rates of transcript elongation by single RNA polymerase molecules. *Journal of Biological Chemistry*, 279(5):3292–3299, 2004.

[12] S. Dixit, M. Singh-Zocchi, J. Hanne, and G. Zocchi. Mechanics of binding of a single integration-host-factor protein to DNA. *Physical Review Letters*, 94(11):118101–(1–4), 2005.

[13] S. Blumberg, A. Gajraj, M. W. Pennington, and J. C. Meiners. Three-dimensional characterization of tethered microspheres by total internal reflection fluorescence microscopy. *Biophysical Journal*, 89(2):1272–1281, 2005.

[14] O. K. Wong, M. Guthold, D. A. Erie, and J. Gelles. Interconvertible lactose repressor–DNA looped complexes revealed by single-molecule experiments. Unpublished.

[15] C. Zurla, A. Franzini, G. Galli, D. D. Dunlap, D. E. A. Lewis, S. Adhya, and L. Finzi. Novel tethered particle motion analysis of cI protein-mediated DNA looping in the regulation of bacteriophage lambda. *Journal of Physics-Condensed Matter*, 18(14):S225–S234, 2006.

[16] P. C. Nelson, C. Zurla, D. Brogioli, J. F. Beausang, L. Finzi, and D. Dunlap. Tethered particle motion as a diagnostic of DNA tether length. *Journal of Physical Chemistry B*, 110(34):17260–17267, 2006.

[17] D. E. Segall, P. C. Nelson, and R. Phillips. Volume-exclusion effects in tethered-particle experiments: Bead size matters. *Physical Review Letters*, 96(8):088306–(1–4), 2006.

[18] H. Qian and E. L. Elson. Quantitative study of polymer conformation and dynamics by single-particle tracking. *Biophysical Journal*, 76:1598–1605, 1999.

[19] B. van den Broek, F. Vanzi, D. Normanno, F. S. Pavone, and G. J. L. Wuite. Real-time observation of DNA looping dynamics of type IIE restriction enzymes NaeI and NarI. *Nucleic Acids Research*, 34(1):167–174, 2006.

[20] S. A. McKinney, C. Joo, and T. Ha. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophysical Journal*, 91(5):1941–1951, 2006.

[21] B. Revet, B. von Wilcken-Bergmann, H. Bessert, A. Barker, and B. Muller-Hill. Four dimers of lambda repressor bound to two suitably spaced pairs of lambda operators form octamers and DNA loops over large distances. *Current Biology*, 9(3):151–154, 1999.

[22] I. B. Dodd, A. J. Perkins, D. Tsemitsidis, and J. B. Egan. Octamerization of lambda cI repressor is needed for effective repression of $P_{RM}$ and efficient switching from lysogeny. *Genes and Development*, 15(22):3013–3022, 2001.

[23] A. Hochschild. The $\lambda$ switch: cI closes the gap in autoregulation. *Current Biology*, 12:R87–R89, 2002.

[24] F. Vanzi, C. Broggio, L. Sacconi, and F. S. Pavone. Lac repressor hinge flexibility and DNA looping: single molecule kinetics by tethered particle motion. *Nucleic Acids Research*, 34(12):3409–3420, 2006.

[25] L. R. Rabiner. A tutorial on hidden Markov-models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[26] F. Qin, A. Auerbach, and F. Sachs. A direct optimization approach to hidden Markov modeling for single channel kinetics. *Biophysical Journal*, 79(4):1915–1927, 2000.

[27] F. Qin, A. Auerbach, and F. Sachs. Hidden Markov modeling for single channel kinetics with filtering and correlated noise. *Biophysical Journal*, 79(4):1928–1944, 2000.

[28] D.A. Smith, W. Steffen, R.M. Simmons, and J. Sleep. Hidden-Markov methods for the analysis of single-molecule actomyosin displacement: The variance-hidden-Markov method. *Biophys. J.*, 81:2795–2816, 2001.

[29] L. S. Milescu, A. Yildiz, P. R. Selvin, and F. Sachs. Maximum likelihood estimation of molecular motor kinetics from staircase dwell-time sequences. *Biophysical Journal*, 91(4):1156–1168, 2006.

[30] H. Yang and X. S. Xie. Statistical approaches for probing single-molecule dynamics photon-by-photon. *Chemical Physics*, 284(1-2):423–437, 2002.

[31] H. Yang, G. B. Luo, P. Karnchanaphanurach, T. M. Louie, I. Rech, S. Cova, L. Y. Xun, and X. S. Xie. Protein conformational dynamics probed by single-molecule electron transfer. *Science*, 302(5643):262–266, 2003.

[32] M. Andrec, R. M. Levy, and D. S. Talaga. Direct determination of kinetic rates from single-molecule photon arrival trajectories using hidden Markov models. *Journal of Physical Chemistry A*, 107(38):7454–7464, 2003.

[33] G. F. Schroder and H. Grubmuller. Maximum likelihood trajectories from single molecule fluorescence resonance energy transfer experiments. *Journal of Chemical Physics*, 119(18):9920–9924, 2003.

[34] L. S. Milescu, A. Yildiz, P. R. Selvin, and F. Sachs. Extracting dwell time sequences from processive molecular motor data. *Biophysical Journal*, 91(9):3135–3150, 2006.

[35] J. P. Hoogenboom, W. K. den Otter, and H. L. Offerhaus. Accurate and unbiased estimation of power-law exponents from single-emitter blinking data. *J. Chem. Phys.*, 125:204713–(1–12), 2006.

[36] D. Colquhoun and F. J. Sigworth. Fitting and statistical analysis of single-channel records. In B. Sakmann and E. Neher, editors, *Single-Channel Recording*, pages 191–263. Plenum, New York, 1983.

[37] P. R. Bevington and D. K. Robinson. *Data reduction and error analysis for the physical sciences*. McGraw–Hill, Boston, 3rd edition, 2003.

[38] D.R. Cox and D.V. Hinkley. *Theoretical Statistics*. Chapman and Hall, New York, 1974.

[39] J. F. Beausang, C. Zurla, L. Finzi, L. Sullivan, and P.C. Nelson. Elementary simulation of tethered Brownian motion. *Am. J. Phys.*, 75:520–523, 2007.

[40] A. R. Dunn and J. A. Spudich. Dynamics of the unbound head during myosin V processive translocation. *Nature Structural and Molecular Biology*, 14(3):246–249, 2007.

[41] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, Cambridge, 2nd edition, 1992.

[42] L. P. Watkins and H. Yang. Detection of intensity change points in time-resolved single-molecule measurements. *Journal of Physical Chemistry B*, 109(1):617–628, 2005.