Principal-components analysis of shape fluctuations of single DNA molecules

Adam E. Cohen* and W. E. Moerner

Department of Chemistry, Stanford University, Stanford, CA 94305

Edited by Robert J. Silbey, Massachusetts Institute of Technology, Cambridge, MA, and approved March 15, 2007 (received for review October 25, 2006)

Thermal fluctuations agitate molecules in solution over a broad range of times and distances. By passively watching the shape fluctuations of a thermally driven biomolecule, one can infer properties of the underlying interactions that determine the motion. We applied this concept to single molecules of fluorescently labeled λ-DNA, a key model system for polymer physics. In contrast to most other single-molecule DNA experiments, we examined the unstretched, equilibrium state of DNA by using an anti-Brownian electrokinetic trap to confine the center of mass of the DNA without perturbing its internal dynamics. We analyze the long-wavelength conformational normal modes, calculate their spring constants, and measure linear and nonlinear couplings between modes. The modes show strong signs of nonlinear hydrodynamics, a feature of the underlying equations of polymer dynamics that has not previously been reported and is neglected in the widely used Rouse and Zimm approximations.

DNA dynamics | single molecule | hydrodynamic interactions | polymer physics

The simplest model of a linear polymer is a chain of beads joined by springs. In the Rouse model (1), each bead is a Brownian diffuser with the same drag and diffusion coefficients it would have in the absence of other beads. This model neglects the fact that each bead is subject to the time-varying flow fields produced by the other diffusing beads. This hydrodynamic interaction (HI) renders the underlying dynamics nonlinear. The Zimm model (2) includes hydrodynamics but restores linearity through a mean-field approximation: each bead is made to interact with the average conformation of its neighbors. Subsequent work has applied sophisticated mathematical techniques to calculate corrections resulting from fluctuating hydrodynamics (3–5), but the overall significance of internal HIs to polymer dynamics remains unresolved (6).

The HI is the dominant long-range force for biomolecules in aqueous buffers: it couples motion of one part of a molecule to motion of possibly remote parts of the same molecule. Thus, understanding HIs is crucial to understanding the rates of molecular events that involve large-scale conformational change such as folding of proteins and RNA, packaging of DNA, motion of molecular motors, and motion of DNA-binding proteins.

Long before polymers were studied at the single-molecule level, many clever experiments applied light scattering and neutron scattering as indirect probes of polymer dynamics (7, 8). However, these experiments were (i) limited to probing only the lowest one or two internal relaxations and (ii) only yielded second-order ensemble-averaged correlation functions without measuring the entire distribution of underlying states. This second property of scattering techniques makes them insensitive to deviations from the linearized Zimm theory.

Higgins and Benoit (9) and Quake et al. (10) used laser tweezers to study the dynamics of partially extended DNA in solution yet failed to find deviations from the Zimm theory. This negative result is not surprising, because extending a molecule weakens the internal HIs, rendering the nonlinearity harder to detect. Subsequent experiments have probed DNA under a wide range of twists and extensions (for reviews see refs. 11 and 12). In contrast, the equilibrium (i.e., unstretched) dynamics of single DNA molecules have received relatively little attention (13–15) because of the challenge of following a molecule as it diffuses away from the field of view.

Experimental Method

Our experimental method consists of (i) acquiring a large number of images showing the free-solution shape fluctuations of single fluorescently labeled molecules of DNA [see supporting information (SI) Movie 1]; (ii) identifying the conformational normal modes (analogous to the wavefunctions of an electron or the normal modes of a drum); and (iii) determining whether the dynamics in these modes can be fit to a linear model, as required by the Zimm theory. The observed deviations from Zimm theory verify that HI effects are present in DNA and that HI may play an important role in the dynamics of more complex biomolecular systems.

To study a single molecule in equilibrium, one would like to eliminate the motion of the center of mass without affecting internal motions. Active feedback provides a means to accomplish this elimination. One tracks the Brownian motion of a molecule and then imposes a body force that counteracts this motion. A variety of schemes have been proposed (16–18) and implemented (19–22), differing in the method of tracking and the source of the restoring force. Here we apply our anti-Brownian electrokinetic (ABEL) trap that uses video tracking and electrokinetic feedback and is capable of trapping objects as small as individual proteins in solution (23). The hardware and software have been described in detail (19, 20, 24, 25). In brief, the molecule to be trapped is confined to a thin fluid layer (∼1 μm thick, slightly larger than the radius of gyration $R_g$ for λ-DNA (∼700 nm)) in a glass microfluidic cell. An automated video tracking system follows the Brownian displacements (via fluorescence microscopy) and, for every frame, applies feedback voltages to the cell to induce an electrokinetic drift that approximately cancels the Brownian motion. Motions at frequencies higher than the update rate are not affected. Molecules of double-stranded λ-DNA fluorescently labeled with YOYO-1 (Molecular Probes) were held in the ABEL trap, and a two-dimensional projection of their conformational motions was recorded with video microscopy at a time resolution of 4.5 ms per frame. Twenty-one separate molecules were trapped, each for between 9 and 18 s (yielding between 2,000 and 4,000 images per molecule; total data set: 58,421 frames). Several frames showing the shape fluctuations that are the focus of this article are shown in Fig. iii.
with the full video given in SI Movie 1 along with details of data analysis and calculations. The molecules were labeled with a uniform density of fluorophores, so the fluorescence intensity at each point in an image was proportional to the density of DNA averaged over the point-spread function of the microscope at a corresponding point in the sample.

An important question is whether the trapping feedback fields affect the conformation or dynamics of the DNA. To test for such interactions experimentally, we calculated the correlation between the applied voltage and the measured shape of the molecule as follows. Let $I(r, t)$ be the intensity distribution (after removing center-of-mass motion). Let $\langle I(r, t) \rangle$ be the time-average distribution and $\delta I(r, t) = I(r, t) - \langle I(r, t) \rangle$ be the instantaneous deviation from this average. The feedback voltages are $V_x$ and $V_y$. We found that $\langle V_x(t_2)\delta I(r, t_1) \rangle \approx 0$ and $\langle V_y(t_2)\delta I(r, t_1) \rangle \approx 0$ to within the experimental uncertainty for all $t_2, t_1$, and $r$. Thus, coupling between the feedback voltage and the conformation is small enough to be neglected here. This finding is consistent with the theoretical argument that electrokinetic force and drag act uniformly along the DNA backbone (26) and the experimental observation that the free-solution mobility of DNA is independent of contour length or conformation (27). We cannot rule out the possibility that higher-order effects modify the conformation to a small extent, but our current measurements are not sensitive to such perturbations.

Finite element simulations show that the electric field is homogeneous to within $0.1\%$ over the size of the DNA molecule (25).

Another possible source of bias in the data is the finite z thickness of the trapping region, which is only slightly larger than the radius of gyration of $\lambda$-DNA (1 $\mu$m vs. 700 nm). This confinement is expected to have little effect on the distribution of states; excluded volume interactions are very weak in $\lambda$-DNA, so a modest confinement of the random walk in one dimension does not significantly affect the random walks in the perpendicular dimensions. This assumption is confirmed by the analysis below, which shows...
that the DNA shape is well described by a pure random walk. Of
greater concern is the impact of the walls of the trap on the
dynamics. The “no-slip” condition on the walls leads to a screening
of HI on a length scale of \( \approx 1 \) \( \text{\textmu} \text{m} \). Thus, the HIs described below
will be slightly stronger in a truly three-dimensional polymer. The
diffusion coefficient of the center of mass of \( \lambda \)-DNA in the ABEL
trap is 24% lower than in bulk (28). Internal HIs propagate over
shorter distances than does the disturbance generated by diffusion
of the center of mass. Thus, the perturbation to internal HI effects
is expected to be <24%.

**Principal-Components Analysis**

We now analyze the shapes of \( \lambda \)-DNA, first considering only
the distribution of shapes and then the dynamics (i.e., how one shape
becomes another). How can one describe the variability among
the observed shapes without selecting an arbitrary descriptive
statistic as the quantity of interest? Principal-components analysis
(PCA) provides a systematic expansion procedure for
course-graining over atomic degrees of freedom while preserving
the large-distance dynamics that are relevant for many
functions. Furthermore, PCA is unbiased in the sense that it uses
the data to determine the characteristic motions without requiring
the experimenter to specify a model of the underlying
process.

The process of PCA consists of (i) obtaining the covariance
matrix for some randomly fluctuating \( N \)-dimensional quantity and
(ii) calculating the eigenvectors of this covariance matrix. The
eigenvectors with the largest eigenvalues are called the principal
elements of the fluid are identical), so there is no need to
approach takes advantage of a symmetry of the fluid (i.e., that
is generally preferred because of its simplicity. The Eulerian
calculates fluid properties (e.g., pressure, velocity, shear) as a
trajectory of distinct fluid elements (imagine tracer parti-
would a projection onto any other basis with \( m \) elements.

There are two approaches to PCA on a material system, which,
in analogy to fluid dynamics, we call the Lagrangian and Eulerian approaches. In Lagrangian fluid dynamics, one follows the trajectories of distinct fluid elements (imagine tracer particles) as a function of time. In Eulerian fluid dynamics, one calculates fluid properties (e.g., pressure, velocity, shear) as a function of position in some fixed reference frame. The Eulerian approach leads to the well known Navier–Stokes equations and is generally preferred because of its simplicity. The Eulerian approach takes advantage of a symmetry of the fluid (i.e., that all elements of the fluid are identical), so there is no need to follow a particular fluid element (29).

A similar situation prevails when one wishes to perform PCA on the fluctuations of a material body. There are two ways one can construct the covariance matrix. One can follow the trajectory \( x(t) \) of mass element \( i \) and calculate covariance matrices such as \( C(i, j, \tau) = \langle x(t)x(\tau \pm \tau) \rangle \). This approach is necessary when considering a multicomponent system such as a protein, in which two elements \( i \) and \( j \) may not be interchangeable (e.g., they have different chemical properties). We call this the Lagrangian approach. For a homogeneous system, however, one can adopt a reduced description, considering only the density \( p(x) \). Then, one can calculate covariance matrices such as \( C(x_1, x_2, \tau) = \langle p(x_1,t)p(x_2, t + \tau) \rangle \). We call this the Eulerian approach.

Starting with the pioneering work of Karplus and Koshlick (30),
PCA has been widely applied to analyze molecular dynamics
trajectories, but always in the Lagrangian perspective. Such
an approach is productive, because the secondary structure of proteins
is fairly rigid, so many motions involve entire domains. For a review
of PCA of proteins, see refs. 31 and 32. The standard Rouse and
Zimm models of polymer dynamics are also developed in the
Lagrangian language (following every mass element). Indeed,
performing PCA on a random walk in the Lagrangian perspective
yields precisely the Rouse polymer modes.

For determining many dynamical properties of polymers, only
the density distribution is needed. The present experiments are not
sensitive to the underlying sequence of base pairs, so we approxi-
mate the DNA as a homogeneous polymer. Furthermore, we do not
know which piece of the polymer contributes to each piece of the
image. Thus, it is most appropriate to perform PCA in the Eulerian
perspective.

We have performed PCA (33) on 58,421 video images of DNA to
determine spatially separated parts of the DNA that fluctuate in
synchrony (see SI Appendix). Fig. 1b shows the first 16 PCs for
\( \lambda \)-DNA. Starting at the upper left, the dominant PC is a radial
breathing mode, followed by a pair of degenerate modes in which the
molecule stretches along one axis and contracts along an
orthogonal axis, followed by more complex deformations. Each PC
is indexed by \((n, l)\), where \( n \) is the number of radial nodes, and \( l \) is
the number of azimuthal nodes. PCA implies a picture of the
molecule as a gel-like solid with a spectrum of long-wavelength
collective motions.

Each PC has associated with it an eigenvalue, \( \lambda_p \), that is equal to
the fraction of the variance of the entire data set that falls along the
PC. The equipartition theorem implies that \( \frac{1}{2}k_B\lambda_p = \frac{1}{2}k_B T \),
where \( k_B \) is the Boltzmann constant, and \( T \) is the temperature, from which
one can extract the stiffness, \( k_p \), of mode \( p \). The persistence length
is much less than our optical resolution, so the stiffness is entirely
caused by the entropic cost of deforming the molecule. The
stiffnesses of the first 45 PCs (shown in Fig. 1d) follow an unex-
plained semiregular pattern. The PCs and the spectrum of eigen-
values appear qualitatively similar to atomic wavefunctions and
energy levels familiar from quantum mechanics, but the details of
the shape and the underlying equations are completely different.

Use of the PCs achieves a large reduction in the amount of data
required to describe the conformational fluctuations compared with
the raw images. Although each image contains 1,024 pixels,
90% of the variance in the data set is contained in the first 34 PCs
(Fig. 1e). The remaining variance is mostly caused by measurement
noise. By working in the PC basis, we suggest that numerical
simulations of polymer dynamics could be rendered more efficient.
Rather than simulating the trajectory of each mass element, one
could simulate the dynamics of the PCs. This approach would only
be beneficial for studying near-equilibrium fluctuations; otherwise
the PCs cease to be an efficient basis.

**PCs of a Random Walk**

Here we develop a semianalytical description of the observed
PCs. As a minimal model we assume that each image recorded by
the camera shows a pure two-dimensional random walk. In
this section we derive the PCs of a \( d \)-dimensional random walk
with a fixed center of mass. Although these results are presented
in the context of polymer physics, a coarse-grained description
of the shapes of random walks may prove useful in other
disciplines as well.

Consider a one-dimensional random walk of \( N \) Gaussian steps
joining \( n + 1 \) mass elements, each step of variance \( a^2 \). We will work
in the limit of large \( N \) and small \( a \), keeping \( Na^2 = 1 \), and keeping
the center of mass of the walk fixed at the origin. We think of the
polymer as a density distribution \( \rho(x) \). In the Eulerian perspective,
the covariance matrix depends on fluctuations about the mean
density distribution, \( \langle \rho(x) \rangle \), so we start by discussing this distribution.

The mean distribution of density about the center of mass of
a random walk is not Gaussian (34). Each segment obeys a
Gaussian density distribution about the center of mass, but the
width of this distribution varies along the chain: the ends wander
further than the middle. The total density distribution is the sum
of many Gaussian distributions of distinct widths and, thus, is not
a Gaussian distribution. Yamakawa (34) showed that the total
density distribution is
\[
\langle \rho(x) \rangle = \sum_{x=0}^{N} \frac{3}{2F(x)} \exp \left[ -\frac{3x^2}{2F(x)} \right] \tag{1}
\]

with
\[
F(x) = 3 \frac{x^2}{N^2} - 3 \frac{x}{N} + 1. \tag{2}
\]

This function looks qualitatively like a Gaussian distribution near \( x = 0 \) but has fat tails relative to a Gaussian distribution. For a multidimensional Gaussian random walk, the displacements along orthogonal axes are statistically independent, so the total probability density is the product of the one-dimensional probability densities along each of the axes. We previously showed that Eq. 1 agrees well with the observed density distribution of λ-DNA, whereas a Gaussian distribution does not (28).

Now we calculate the density–density covariance of an ensemble of one-dimensional random walks in the Eulerian perspective. The covariance matrix is
\[
C(x_1, x_2) = \langle \rho(x_1) \rho(x_2) \rangle - \langle \rho(x_1) \rangle \langle \rho(x_2) \rangle \tag{3}
\]

The second term on the right-hand side of Eq. 3 is obtained from Eq. 1. The challenging task is to calculate the first term, \( \langle \rho(x_1) \rho(x_2) \rangle \).

If we assume that each piece of the random walk contributes a point-like density, the density at position \( x_1 \) is
\[
\rho(x_1) = \sum_{\alpha=0}^{N} \delta(x_1 - x_\alpha),
\]

where \( \alpha \) is an index of the mass elements. The product of the densities at two positions is
\[
\rho(x_1) \rho(x_2) = \sum_{\alpha, \beta=0}^{N} \delta(x_1 - x_\alpha) \delta(x_2 - x_\beta).
\]

We take the Fourier transforms of the above equation with respect to \( x_1 \) and \( x_2 \) to obtain
\[
\rho(k_1) \rho(k_2) = \sum_{\alpha, \beta=0}^{N} \exp[i(k_1 x_\alpha + k_2 x_\beta)].
\]

The position of mass element \( \alpha \) is then expanded in Rouse modes as
\[
x_\alpha = \sum_{\nu=1}^{\infty} c_\nu \cos \left( \frac{\nu \pi x}{N} \right)
\]

and similarly for \( x_\beta \). Leaving out the \( \nu = 0 \) term guarantees that the center of mass remains fixed at \( x = 0 \) (i.e., \( \sum x_\alpha = 0 \) for all \( \{c_\nu\} \)). Thus, we have
\[
\rho(k_1) \rho(k_2) = \sum_{\alpha, \beta=0}^{N} \exp \left[ i \sum_{\nu=1}^{\infty} c_\nu \left( k_1 \cos \left( \frac{\nu \pi x}{N} \right) + k_2 \cos \left( \frac{\nu \pi x}{N} \right) \right) \right].
\]

The coefficients \( c_\nu \) are Gaussian distributed and statistically independent, with \( \langle c_\nu c_\mu \rangle = 0 \) and \( \langle c_\nu \rangle = 0 \). After taking the average over all conformations (i.e., all \( \{c_\nu\} \)), we obtain
\[
\langle \rho(k_1) \rho(k_2) \rangle = \sum_{\alpha, \beta=0}^{N} \exp \left[ -\sum_{\nu=1}^{\infty} \frac{1}{(\nu \pi)^2} \left( k_1 \cos \left( \frac{\nu \pi x}{N} \right) + k_2 \cos \left( \frac{\nu \pi x}{N} \right) \right) \right]. \tag{4}
\]

The next challenge is to evaluate the sum
\[
S = \sum_{\nu=1}^{\infty} \frac{1}{(\nu \pi)^2} \left( k_1 \cos \left( \frac{\nu \pi x}{N} \right) + k_2 \cos \left( \frac{\nu \pi x}{N} \right) \right)^2. \tag{5}
\]

A similar sum appears in the theory of scattering from polymer solutions (35), but in that case the sum depends on only one \( k \) variable. Eq. 4 becomes the formula for the static structure factor when \( k_2 = -k_1 \). This difference arises because scattering experiments always probe spatially averaged quantities. Direct imaging experiments, on the other hand, allow us to compare fluctuations at pairs of distinct points, \( x_1 \) and \( x_2 \), or alternatively at distinct \( k \) vectors, \( k_1 \) and \( k_2 \).

When \( k_1 = -k_2 \), the sum in Eq. 5 evaluates to \( S = k_1^2 |\alpha - \beta|/2 \) (35). In the more general case where \( k_1 \) and \( k_2 \) vary independently, the sum evaluates to
\[
S = \frac{1}{12} \left[ 2k_2^2 F(\alpha) + 2k_2^2 F(\beta) + k_1 k_2 G(\alpha, \beta) \right], \tag{6}
\]

where \( F(\cdot) \) is as defined in Eq. 2, and
\[
G(\alpha, \beta) = 3 \left( \frac{\alpha}{N} + \frac{\beta}{N} - 1 \right)^2 + 3 \left| \frac{\alpha}{N} - \frac{\beta}{N} \right|^2 - 2. \tag{7}
\]

Substituting Eq. 6 into Eq. 4 and taking the inverse Fourier transform yields
\[
\langle \rho(x_1) \rho(x_2) \rangle = \sum_{\alpha, \beta=0}^{N} \frac{6}{\pi} \frac{16 F(\alpha) F(\beta) - G(\alpha, \beta)^2}{16 F(\alpha) F(\beta) - G(\alpha, \beta)^2} \exp \left[ -\frac{24 F(\beta) x_1^2 - 24 F(\alpha) x_2^2 + 12 G(\alpha, \beta) x_1 x_2}{16 F(\alpha) F(\beta) - G(\alpha, \beta)^2} \right]. \tag{7}
\]
in the DNA (such as excluded volume) that has been left out of the model.

**Dynamics of the PCs**

The PCA is only sensitive to shape deformations that occur within the same video frame, so it provides no information on the dynamics (i.e., how a fluctuation at one time impacts the fluctuations at a later time). If the video images of the DNA were randomly reordered, the results of PCA would not change. To obtain a more detailed picture of the dynamics, we decomposed each image into the basis of PCs and examined the time dependence of the mode amplitudes, $a_p(t)$.

The amplitude in each PC ebbs and flows stochastically as thermal fluctuations and viscous damping add and remove energy. The modes will decay with characteristic relaxation times that are related to the eigenvalues of the covariance matrix $\rho_{pq} = \langle a_p(t) a_q(t') \rangle$, where the indices $p$ and $q$ may be truncated at a small value (e.g., 15 in Fig. 2a). If the PCs were also the eigenstates of time evolution, then $\rho_{pq}$ would be diagonal for all $t$. However, Fig. 2a shows that off-diagonal terms arise. The only significant off-diagonal elements connect mode $p = (n \pm 1, l)$ to mode $q = (n, l)$, suggesting that conservation of azimuthal node number is a selection rule for DNA conformational transitions.

The autocorrelations of the PC amplitudes [i.e., the diagonal elements $\tilde{\rho}_{pp}(\tau)$] show nonexponential decay in time, a phenomenon due partially to the mode-mixing shown in Fig. 2a and partially to the nonlinear interaction discussed below. Nonetheless, a characteristic relaxation time can be associated with each PC by examining its slope near zero time lag by using the formula

$$\frac{1}{\tau_p} = \frac{1}{\delta \tau} \frac{\tilde{\rho}_{pp}(1) - \tilde{\rho}_{pp}(2)}{\tilde{\rho}_{pp}(1)}.$$  

The values of $\tilde{\rho}_{pp}$ at lags 1 and 2 were used here (rather than at lags 0 and 1) to avoid contaminations from shot noise and other $\delta$-correlated noise sources. These relaxation times are shown in Fig. 2b and are well fit by a power-law $\tau_p \sim p^\alpha$, with $\alpha = -0.55 \pm 0.05$ (95% confidence interval).

Why do the relaxation times of the PC modes scale like $\tau_p \sim p^{-1/2}$ while the relaxation times of the Zimm modes scale like $p^{-3/2}$? In the Zimm model, the modes are sinusoids, like the vibrations of a violin string, with a characteristic wavenumber $k_p \sim p$. Zimm showed that a fluctuation with a wavenumber of $k$ has a relaxation time $\tau \sim k^{-3/2}$. In the PC model the fundamental motions are those of a three-dimensional spherically symmetric elastic continuum. In $k$ space, the number of modes with a wavenumber less than $k$ in magnitude scales as $p \sim k^3$, and so the $p$th mode will have a wavenumber $k_p \sim p^{1/3}$. The relation $\tau \sim k^{-3/2}$ still applies, whence $\tau_p \sim p^{-1/2}$. Thus, the essential difference between the Zimm and PC models is that the Zimm model deals with the vibrations of a one-dimensional string, whereas the PC model deals with the vibrations of a three-dimensional elastic continuum.

Now we quantify nonlinear interactions between the modes. The approach is to fit a linear model to the dynamics and then to look for higher-order correlations among the residuals. If the time evolution of the system were linear and Markovian (as required by both the Rouse and Zimm models), then the vector of amplitudes in each eigenstate would evolve according to a Langevin equation,

$$a(k + 1) = Ma(k) + \xi(k),$$

where $M$ is a transition matrix, $\xi$ is a vector of Gaussian white noise describing the effect of thermal fluctuations, and $k$ is the frame index ($t = k \delta t$) (see SI Appendix). We considered only the first 15 eigenvectors and estimated the elements of $M$ by a least-squares fit to the data. The time series of the residuals $\xi(k)$ was examined for signs of nonlinear dynamics. For example, a conformation-dependent internal friction would lead to a nonzero value of the third-order correlation function:

$$\tilde{\rho}_{pq}(\tau) = \frac{\langle \xi^*_p(t + \tau) a_q(t) \rangle}{\text{var}(\xi_p)\text{var}(a_q)^{1/2}}.$$  

This correlation function characterizes the effect of amplitude in mode $q$ affecting the friction (and hence the thermal fluctuations $\xi$) in mode $p$. Fig. 2c shows that for several $p$ and $q$, $\tilde{\rho}_{pq}(\tau) \neq 0$, indicating that the ansatz of Eq. 8 is only approximately true. Excitation in the $(1, 0)$ mode leads to the largest nonlinear effects, strongly affecting the dynamics in the mode $(1, 0)$ itself as well as the two $(2, 1)$ modes and weakly affecting the dynamics in the two $(1, 2)$ modes. Details of the calculation are given in SI Appendix. These nonzero correlations show that the intrinsic...
nonlinear couplings in the DNA-shape dynamics are directly observable in our measurements. These effects are not predicted by the linearized Zimm model and cannot be detected by traditional scattering techniques.

We have shown that active trapping of single molecules of DNA in the ABEL trap provides rich dynamical information about the shapes and fluctuations of the molecules. The results presented here serve as a benchmark against which to test analytical theories and numerical simulations of polymer dynamics. The selection rules for conformational transitions (Fig. 2a) and the pattern of nonlinear coupling between modes (Fig. 2c) both await theoretical explanation.

In our analysis we found strong signs of dynamic internal interactions (forces depending on the relative motion of chain elements) but not of static internal interactions (forces depending on the relative position of chain elements). We justified neglect of static interactions on the basis of the good agreement between the PCA eigenstates of the data and those of a pure random walk. The appearance of features that resemble condensed globules, or pearls, in the movies is, we believe, an artifact of the finite resolution of the imaging; all features smaller than a diffraction-limited cutoff take on a globular appearance. The possibility remains that PCA is particularly sensitive to the effects of static interactions and that a different statistical description would be preferable. Such static interactions could arise from solvent-induced attraction, electrostatic repulsion, or excluded volume. However, the present experiment was performed under “good-solvent” conditions, for which the DNA is expected to adopt a random coil conformation.

It will be interesting to see how the descriptive statistics presented here change when a trapped molecule of DNA is subjected to physical and chemical perturbations, such as changes in pH, temperature, ionic strength, or the addition of proteins that interact with DNA. In particular, it would be interesting to add condensing agents to study the dynamics of the coil-to-globule transition (36), which is a topic of much interest for polyelectrolytes (37). One may also use more complex labeling schemes, such as labeling the ends of the polymer with a different color from the center. We hope that our experiments on DNA will spur efforts toward a better understanding of the role of HIs in biomolecular processes.

Materials and Methods

DNA Preparation. Double-stranded λ phage DNA (Molecular Probes) was dissolved in a buffer of 10 mM Tris-HCl, 10 mM NaCl, and 1 mM EDTA (pH 8.0). The fluorescent dye YOYO-1 (Molecular Probes) was added at a concentration of 1 dye/10 bp of DNA, and the mixture was incubated at room temperature in the dark for 30 min. An oxygen-scavenger mixture was added because of the presence of statistical noise in the images. Eigenvectors 11–14 are nearly degenerate and arrived from each frame. In a small fraction of the frames (<5%), a second DNA molecule was seen floating through the field of view. In these frames, the pixels affected by the second molecule were manually set to the background level. Images were shifted to remove residual center-of-mass fluctuations uncompensated by the ABEL trap by using a bicubic interpolation to localize the center of mass to less than the pixel size. The total intensity of each frame was normalized to account for the slow rate of photobleaching of the YOYO-1 during the trapping period. For the present analysis, the data from all 21 molecules was aggregated except for the third-order correlation functions in Fig. 2c.

The 1,024 × 1,024 equal-time covariance matrix, $C(x_i, x_j)$, was numerically diagonalized in Matlab (Mathworks, Natick, MA). The eigenvectors with the largest eigenvalues were converted to 32 × 32 images. Eigenvectors 11–14 are nearly degenerate and arrived mixed together because of the presence of statistical noise in the covariance matrix. The space of four-dimensional rotations among these eigenvectors was manually searched to find linear combinations that had manifest symmetry. These linear combinations were taken to be the “true” eigenvectors.

We thank John Brauman, Joel E. Cohen, and Kit Werley for a careful reading of this manuscript and Willy Wyat of Applied Biosystems for samples of an antiadsortion polymer. This work was supported in part by the National Science Foundation Grant CHE-0554681 and National Institutes of Health Grant 1R21-RR023149. A.E.C. acknowledges the support of a Hertz Fellowship.

37. Dobrynin AV, Rubinstein M, Eigenvectors 11–14 are nearly degenerate and arrived mixed together because of the presence of statistical noise in the covariance matrix. The space of four-dimensional rotations among these eigenvectors was manually searched to find linear combinations that had manifest symmetry. These linear combinations were taken to be the “true” eigenvectors.