Modeling the relaxation of internal DNA segments during genome mapping in nanochannels

Aashish Jain,¹ Julian Sheats,¹ Jeffrey G. Reifengerger,² Han Cao,² and Kevin D. Dorfman^{1,a)}

¹*Department of Chemical Engineering and Materials Science, University of Minnesota-Twin Cities, 421 Washington Ave. SE, Minneapolis, Minnesota 55455, USA*

²*BioNano Genomics, 9640 Towne Centre Drive, Suite 100, San Diego, California 92121, USA*

(Received 23 July 2016; accepted 4 October 2016; published online 13 October 2016)

We have developed a multi-scale model describing the dynamics of internal segments of DNA in nanochannels used for genome mapping. In addition to the channel geometry, the model takes as its inputs the DNA properties in free solution (persistence length, effective width, molecular weight, and segmental hydrodynamic radius) and buffer properties (temperature and viscosity). Using pruned-enriched Rosenbluth simulations of a discrete wormlike chain model with circa 10 base pair resolution and a numerical solution for the hydrodynamic interactions in confinement, we convert these experimentally available inputs into the necessary parameters for a one-dimensional, Rouse-like model of the confined chain. The resulting coarse-grained model resolves the DNA at a length scale of approximately 6 kilobase pairs in the absence of any global hairpin folds, and is readily studied using a normal-mode analysis or Brownian dynamics simulations. The Rouse-like model successfully reproduces both the trends and order of magnitude of the relaxation time of the distance between labeled segments of DNA obtained in experiments. The model also provides insights that are not readily accessible from experiments, such as the role of the molecular weight of the DNA and location of the labeled segments that impact the statistical models used to construct genome maps from data acquired in nanochannels. The multi-scale approach used here, while focused towards a technologically relevant scenario, is readily adapted to other channel sizes and polymers. *Published by AIP Publishing.* [<http://dx.doi.org/10.1063/1.4964927>]

I. INTRODUCTION

The advent of genome mapping in nanochannels^{1–6} has spurred interest in understanding the properties of semiflexible polymers confined in channels smaller than their persistence length. In this technology, DNA molecules hundreds of kilobase pairs (kbp) long, containing sequence specific fluorescence probes, are electrokinetically injected into an array of tens of thousands of nanochannels (Fig. 1(a)). The width of these square channels is around 45 nm,² which is roughly the same as the persistence length of DNA in a high ionic strength buffer.⁷ Under such strong confinement, the labels on the DNA backbone order linearly and their locations can be “read” using optical microscopy (Fig. 1(b)). Importantly, the genomic data arising in these experiments do not have any global hairpins in the DNA, which would scramble the ordering of the labels.⁸ The combination of extensive work in theory,^{9–14} simulation,^{12–34} and experiments^{8,34–41} have produced a relatively complete picture of the thermodynamics of DNA confined below its persistence length. However, considerably less is known about the dynamics of DNA in small nanochannels,^{26,29,36,42,43} in particular, the internal segmental dynamics. In the

^{a)}Electronic mail: dorfman@umn.edu

FIG. 1. Schematic of the mapping of a strongly confined DNA molecule in a channel onto a one-dimensional bead-spring chain. (a) Typical experimental image obtained during genome mapping in nanochannels.^{2,40} The DNA are stretched in an array of circa $45 \text{ nm} \times 45 \text{ nm}$ nanochannels. This image shows a small fraction of the array. Two images are taken in quick succession; the blue color indicates the DNA backbone (labeled with YOYO-1) and the cyan color indicates the sequence-specific fluorescent probes. (b) Zoom image of one of the DNA molecules. (c) Kymograph of the position of the labels in panel (b) as a function of time.⁴⁰ (d) Fine-grained simulation approach. The DNA is modeled as a discrete wormlike chain of touching beads,¹⁷ parameterized by the persistence length l_p , effective width w of the chain, and hydrodynamic radius a ,⁴⁴ confined in a square channel of width H . Segment-segment and segment-wall interactions are incorporated through a confined Green's function method.⁴⁵ The cyan beads are intended to represent the labels in panel (a). (e) The fine-scale model is simulated using a pruned-enriched Rosenbluth method^{27,46,47} for a long contour length L to determine the mean fractional extension, the variance about the mean fractional extension, and the friction coefficient per unit length. (f) The outputs of the fine-scale model are used to parameterize a coarse-grained, Rouse-like model⁴⁸ by a spring constant k_s , equilibrium spring length X_s , and bead friction coefficient ζ , all obtained from the PERM simulation. A normal-mode analysis of the Rouse model provides the autocorrelation decay time τ for the distance between beads, which can be compared to the decay time for distances in kymographs similar to panel (c).

present contribution, we develop a multi-scale model that allows us to study efficiently the internal dynamics of DNA when confined in a nanochannel in the absence of global hairpins.

We are particularly interested in the internal dynamics of DNA due to its relevance to the practice of genome mapping. During these experiments,^{1,2} a two-color image similar to Fig. 1(a) is acquired using a fixed exposure time for each color, typically 150 ms per color with a brief delay in switching the laser line and optical filters. The locations of the fluorescent labels are then measured with respect to a reference point, for example, the edge of the entire DNA molecule or one of the other labels, to remove effects of DNA diffusion or drift. There are several sources of error in the measurement of separation distance in these experiments including ones due to static effects such as topological events,⁴¹ mutations, channel size variations, and dynamic thermal fluctuations. The latter are the main focus of the present work.

As seen in the kymograph in Fig. 1(c), obtained using a movie of the label positions rather than a single snapshot,⁴⁰ the positions of the labels in genome mapping fluctuate under thermal energy, and these fluctuations can impact the accuracy of the resulting DNA map since a single snapshot of 150 ms would average over 4–5 frames of this kymograph. If the acquisition time for a snapshot image is long compared to the relaxation time (decorrelation time) for the

distance between a given label and the reference point, then the measured label position will be close to its thermodynamic average value. This is the optimal result for genome mapping, since the resulting conversion from physical distance (nm) to genomic distance (base pairs, bp) only requires knowledge of the average extension of the chain, which can be obtained by theory,^{10,14} simulation,^{31,32} or experimental calibration with a known molecular weight standard. Unfortunately such long exposure times are prohibitive due to photobleaching, which reduces the fluorescence signal of the labels, and photocleavage of the molecule,⁴⁹ which reduces the integrity of the DNA molecule and effectively terminates the experiment. Moreover, the accuracy inherent in increasing the exposure time must be balanced against the concomitant lower throughput. Since the autocorrelation function (ACF) is presumably an exponentially decaying quantity, there are diminishing returns to increasing the exposure time, even if photobleaching and photocleavage are minimized. Thus, it behooves us to understand how the relaxation time for label-label distances is affected by the genomic distance between them, their location within the DNA chain, and the overall molecular weight of the DNA molecule being imaged.

DNA, as a model polymer, affords the unparalleled ability to study the properties of polymers in confinement in a very direct manner by imaging single molecules. However, there is a paucity of work in the literature focusing on the dynamics of DNA in nanochannels, in particular, channels below the persistence length. When available, such studies tend to focus on the dynamics of the entire molecule. For example, the seminal work by Reisner *et al.*³⁶ provided the first data on the relaxation time of λ -DNA over a wide range of channel sizes, results that were later captured by simulations and models.^{26,42} Recent work on the internal dynamics has focused on either the equilibrium^{50,51} or non-equilibrium^{52,53} behavior in channels much larger than the persistence length, and thus describe physics in confinement regimes that do not apply to genome mapping. There are a few studies involving the fluctuations of internal segments of barcoded DNA in channel sizes relevant for genomic mapping.^{8,18,40} However, in each case, only the static thermodynamic properties were measured.

Developing a model for the dynamics of internal fluctuations of DNA molecules during genome mapping poses substantial challenges due to the wide separation of length scales and the computational costs for modeling hydrodynamics in confinement. Explicitly, molecules used for genome mapping are typically greater than 150 kbp in length (and up to 1 megabase pair) to provide a sufficient number of fluorescent labels to align the molecule to the consensus genome map. However, in order to accurately model the configuration of such a chain in a 45 nm channel using a discrete wormlike chain model, we require the sub-persistence length resolution of the chain configuration. The persistence length of DNA corresponds to approximately 150 bp; using 15 beads per persistence length implies that modeling even the smallest DNA molecules used for genome mapping requires 15 000 beads. Simulating the dynamics of such a large chain in the absence of confinement, where there is a closed-form result for the hydrodynamic interactions, is already a daunting challenge and normally handled by resorting to some pre-averaging approximation.^{54,55} The situation is even more difficult in a channel due to the hydrodynamic interactions. While these interactions are screened over a length scale close to the channel size,^{44,56} there is no closed-form solution for the Green's function describing the hydrodynamic interactions within the screening volume. Although the methods for including hydrodynamic interactions in confinement have advanced substantially,^{45,57–59} simulating the dynamics of a confined chain with hydrodynamic interactions is considerably slower than its unconfined counterpart.

Fortunately, the internal dynamics do not need to be resolved at such a fine length scale to produce meaningful information for the practice of genome mapping in nanochannels. Barcoded genomic DNA in the genome mapping system used here contain segments defined by barcode labels separated by distances that range from 0.5 to 80 μm on molecules that are at least 17 μm in length and often much longer. We developed the multi-scale approach outlined in Fig. 1 to produce a model that provides dynamic information resolved within this range for DNA that do not have global hairpin folds, which are the ones of interest for genome mapping. Importantly, our model only uses input parameters that can be obtained from bulk experiments or theories for unconfined polyelectrolytes.

Explicitly, our model takes as input the properties of the DNA (molecular weight, persistence length, effective width, and segmental hydrodynamic radius), all of which are readily available from bulk experimental data or theories that describe their dependence as a function of ionic strength.^{60–63} We also require the system temperature and buffer viscosity, again readily available from bulk experimental data. As illustrated in Fig. 1(d), these parameters serve as the inputs to a discrete wormlike chain model, which we have used extensively^{17,19,25,27,28,30,32,39,44,64} to study the properties of confined polymers in channels near the persistence length. This model resolves DNA at the length scale of the model's bond length, which is approximately 3 nm, and easily reaches $O(10^4)$ beads when simulated using the pruned-enriched Rosenbluth method (PERM).^{46,47} Keep in mind that these are thermodynamic simulations, which provide, *inter alia*, expectations for the average chain extension and the variance about the average, but no information about the polymer dynamics. We also combine the PERM simulations with a confined hydrodynamics calculation to compute the friction coefficient of the confined chain with sub-persistence length spatial resolution.⁴⁴ This fine-scale, confined DNA model has been tested using experimental data for the chain extension^{17,39,65} and its longest relaxation time⁴² and yields results that agree well with experimental data.

The results of the PERM simulations are then used as inputs to the one-dimensional, Rouse-like model of coupled, overdamped harmonic oscillators seen in Fig. 1(f), analogous to previous work by Riehn and coworkers for DNA under weak confinement.^{50,51} Importantly, our diffusion data for the fine-scale model allows us to pick a coarse-graining length scale where all of the hydrodynamic interactions occur within the Rouse beads, with the walls screening the interaction between Rouse beads. The advantage of this approach to simulating confined DNA,^{50,51} even if small errors are introduced by the coarse-graining, is that the coarse-grained model is substantially simpler than the original fine-scale model, reducing the number of beads by a factor of $O(10^3)$ and reducing the dimensionality from 3D to 1D. Equally important, our coarse-grained model captures both the segment-segment and segment-wall hydrodynamic interactions at the local scale but, due to, the screening from the walls, does not require any hydrodynamic interactions between Rouse beads for the modest degrees of coarse-graining that we use here. The model is thus simple to simulate using Brownian dynamics, and many of the interesting properties of the model are readily obtained using a standard normal-mode analysis.⁴⁸

With this overall approach in mind, we begin with a detailed description for constructing and parameterizing the Rouse-like model. The conclusion of this analysis is a closed-form expression for the relaxation time between two segments on the chain, where all of the parameters in the expression emerge from simulating the fine-scale model. The overall approach we use here is general and readily applied to other channel sizes or shapes, as needed, and we focus on the nanochannels used in recent experiments on time-series data during genome mapping.⁴⁰ We reanalyze these data⁴⁰ to test the ability of the model to predict the trends in the relaxation time as a function of the distance between labels on the DNA and the distance of the labels from the ends of the molecule. Our results indicate that our model, while quite simple, reproduces both the trends in the experimental data, as well as the order-of-magnitude for the relaxation time. We then use the model to investigate in more detail the effects of DNA molecular weight, label separation, and label location, and we connect these results to practical implications for genome mapping. Our discourse concludes with perspectives on the further potential uses of both our approach and the particular model developed here.

II. MODELING APPROACH

A. One-dimensional Rouse-like model

Following previous work,^{50,51} the coarse-grained model in Fig. 1 represents the highly confined linear chain by a one-dimensional overdamped harmonic oscillator chain model. This model has an analogy to the Rouse model for polymer chains in free solution,⁴⁸ in which a polymer molecule is represented by beads connected through harmonic springs with zero equilibrium spring length, and where each bead experiences random thermal force and a drag force.

One of the differences between the Rouse model and what we call, for notational clarity, a *Rouse-like model* for a nanochannel-confined chain is that the equilibrium length of each spring in the nanochannel confined chain is nonzero.

The model consists of N_b beads connected by $N_s = N_b - 1$ springs. As such, the equation of motion governing the motion of Rouse bead $i = 1, 2, \dots, N_b$ is

$$\zeta \frac{dR_i}{dt} = k_s [(R_{i+1} - R_i - X_s) - (R_i - R_{i-1} - X_s)] + f_i(t), \quad (1)$$

with hypothetical end beads $R_0 = R_1 - X_s$ and $R_{N_b+1} = R_{N_b} + X_s$, i.e., with a free-end boundary condition at both ends. Here, ζ is the friction coefficient of a bead, R_i is the axial position of bead i in the one-dimensional model, k_s is the spring constant, and X_s is the equilibrium rest length of the spring. Although the factor X_s cancels out in Eq. (1), we retain it to emphasize that the springs have non-zero rest lengths. The random force $f(i)$ acting on bead i has the properties

$$\langle f_i(t) \rangle = 0, \quad (2)$$

and

$$\langle f_i(t_1) f_k(t_2) \rangle = 2\zeta k_B T \delta(t_1 - t_2) \delta_{ik}, \quad (3)$$

where k_B is Boltzmann's constant, T is the absolute temperature, $\delta(t_1 - t_2)$ is a Dirac delta function, and δ_{ik} is the Kronecker delta function. It is clear from the form of Eq. (1) that the Rouse model ignores excluded volume and hydrodynamic interactions between beads, assumptions that will be justified shortly.

B. Parameterization via a fine-scale model

The Rouse-like model involves four parameters: the number of beads N_b , the bead friction ζ , the spring constant k_s , and the equilibrium spring length X_s . The aim of our multi-scale approach in Fig. 1 is to relate these quantities to known properties of the DNA molecule (contour length L , persistence length l_p , effective width w , and segmental hydrodynamic radius a), the buffer (viscosity η and temperature T), and the channel width H . The number of beads is obtained from L and the corresponding mean extension X of the chain

$$N_b = \frac{X}{X_s} + 1, \quad (4)$$

which can be readily solved for N_b once X_s is known. The remaining quantities need to be obtained from a fine-scale model.

To this extent, we model the DNA as a discrete wormlike chain of N_b^P touching beads,¹⁷ with a bond length equal to the hydrodynamic diameter $2a$,⁴⁴ and simulate this model using PERM.^{27,46,47} There is a bending penalty between trios of continuous beads of the form

$$U_{\text{bend}} = -k_B T \kappa (1 - \cos \theta), \quad (5)$$

where θ is the angle between the beads. The bending penalty κ is related to the persistence length via⁴⁴

$$\frac{l_p}{w} = \frac{\kappa}{\kappa - \kappa \coth(\kappa) + 1}. \quad (6)$$

Hard-core repulsion is enforced between all beads and between the beads and the walls over a distance w . This wall interaction implies an effective channel size $H_{\text{eff}} = H - w$ available to

the beads. Each bead also has a hydrodynamic radius a that is generally not equal to the effective width and, for DNA, conveniently satisfies the inequality $a < w$.⁶⁶

The standard PERM simulations for a confined chain²⁷ furnish the average extension of the chain in the axial direction, X , and the variance about the mean, δX^2 , as a function of L . We also use PERM simulations to compute a Kirkwood diffusivity D_K in the axial direction,⁴⁴ using a grid-based method to obtain the hydrodynamic interactions between segments and between the DNA and the walls.⁴⁵ We have recently shown that the Kirkwood approximation improves as confinement increases,⁶⁷ and we obtained a good agreement between experiments and simulations for the relaxation time of the entire DNA chain by invoking the Kirkwood approximation.⁴² For our purposes here, it proves convenient to recast the Kirkwood diffusivity as an axial friction coefficient ζ^P , where the superscript denotes a quantity obtained from PERM simulations. PERM is a chain growth algorithm,⁴⁶ and natively outputs thermodynamic averages as a function of contour length. As a result, the output(s) of our PERM simulations are the functions $X(N_b^P)$, $\delta X^2(N_b^P)$, and $\zeta^P(N_b^P)$, where N_b^P is the number of PERM beads.

To construct the map between the discrete wormlike chain model and the Rouse-like model, we need to determine an appropriate number of PERM beads N_b^{P*} to coarse-grain into a single Rouse bead (or a Rouse spring). There are three important criteria that we consider to choose N_b^{P*} . First, there should be no hydrodynamic interactions beyond the coarse-graining length scale because the Rouse model does not include any bead-bead hydrodynamic interactions. In other words, the value of N_b^{P*} should be sufficiently large so that all hydrodynamic interactions occur, on average, within a length scale represented by a Rouse bead representing the friction of N_b^{P*} PERM beads. This criterion will be used in setting the lower bound on the choice of N_b^{P*} . Second, the number of Rouse beads should be large enough so that end effects emerging from using a fixed value of ζ for each Rouse bead in Eq. (1) are minimized. This criterion will set the upper bound on the choice of N_b^{P*} . This upper bound depends on the chain length; for example, it can be very small for small chains. This introduces a limitation in our multi-scale model that only the chains of lengths greater than a particular length can be studied (so that the upper bound is always higher than the lower bound). Finally, the third criterion pertains to the static phenomena: There should be enough PERM beads to represent a Rouse spring such that the distribution for the extension of N_b^{P*} PERM beads is approximately Gaussian, as required by the harmonic nature of a Rouse spring. This criterion sets a second lower bound on the choice of N_b^{P*} ; naturally, the lower of these two lower bounds must be satisfied. The details of the procedure to choose N_b^{P*} , by satisfying all three criteria, are discussed in Sec. III A.

Once we have selected a value of N_b^{P*} , the mapping between the fine-scale model and the coarse-grained model is straightforward. The friction of a Rouse bead is

$$\zeta = \zeta^P(N_b^{P*}), \quad (7)$$

which is obtained directly from the PERM data. As is customary in the Rouse model, we assign the same friction ζ to each bead, including the end ones to facilitate the normal mode analysis and ultimately arrive at a closed-form result. The mean extension of the spring between two beads is also obtained directly from the PERM data as

$$X_s = X(N_b^{P*}). \quad (8)$$

To obtain the spring constant, we treat each spring as a dumbbell. From the equipartition theorem, the spring constant is⁴²

$$k_s = \frac{k_B T}{\delta X^2(N_b^{P*})}. \quad (9)$$

While we have focused so far on criteria for selecting N_b^{P*} , it is worthwhile to mention two more important limitations to our model. First, the choice of the coarse-grained length scale naturally sets a lower bound on the time resolution of the model. In order for the Rouse-like

model to be valid, the degrees of freedom modeled by a bead-spring pair need to be able to relax such that their thermodynamic averages, used to compute the Rouse parameters via Eqs. (7)–(9), are representative values. To a good approximation, this time scale is set by the relaxation time for a bead-spring dumbbell, $t_r \approx \zeta/(2k_s)$. In dynamic simulations, Eq. (1) typically needs to be integrated using a time-step smaller than this relaxation time for numerical stability. This does not pose any conceptual problems, provided that one only considers dynamics that take place at a time exceeding t_r . A similar caveat applies for a normal-mode analysis of the Rouse-like model. Second, the Rouse-like model assumes no backfolding of the chain at the length scale of the coarse-grained model. Experiments and simulations suggest that this is a reasonable assumption. While there does exist a small population of folds and knots in experimental data, they can be easily removed from the ensemble by scanning the backbone signal and looking for anomalous bright regions.⁴¹ Moreover, folded chains are short-lived, and tend to be removed simply by waiting several minutes after loading the DNA and acquiring data. It is also possible that there are small folds within the DNA that are not detected in experiments due to the low signal-to-noise ratio for the DNA backbone. Fortunately, simulations of confined DNA using realistic parameters for nanochannel mapping²³ suggest that these folds are small compared to our coarse-grained length scale. As a result, small S-loop shapes that appear due to thermal fluctuations are already incorporated into the coarse-grained model, in an average sense, by the equilibrium spring length and the spring constant. The extension data produced these from the aforementioned simulations,²³ which used a Metropolis Monte Carlo method, were recently confirmed by PERM simulations.³¹ Thus, it seems reasonable to assume that the local configurational data (which are not easily obtained in a PERM simulation) are also reproduced.

C. Extension autocorrelation function

For our analysis, we focus on the extension autocorrelation function (ACF) of the distance between segments in the Rouse-like model. This quantity is related to the relaxation time of this segment of the chain. We provide a brief outline of the derivation here;⁶⁸ the complete derivation is provided in the [supplementary material](#). The analysis is similar to that of the standard Rouse model, with the exception that the conversion to normal modes requires shifting the positions of bead R_i to $R_i - iX_s$ to map the finite-extension chain onto a chain with zero average extension.⁵¹ If the extension of a segment between beads a and b is denoted as $x_{ab} = R_b - R_a$, where R_b and R_a are the positions of beads b and a , respectively, then the normalized ACF for the mean extension, $Z_{ab}(t)$, of that segment of the chain is

$$Z_{ab}(t) = \frac{\langle (x_{ab}(t) - \langle x_{ab} \rangle) (x_{ab}(0) - \langle x_{ab} \rangle) \rangle}{\langle (x_{ab}(0) - \langle x_{ab} \rangle) (x_{ab}(0) - \langle x_{ab} \rangle) \rangle}, \quad (10)$$

where $\langle \dots \rangle$ is the mean value. Equation (10) satisfies the properties of a normalized ACF, i.e., $Z_{ab}(0) = 1$ and $Z_{ab}(\infty) = 0$. The former result is obvious and the latter result is due to the relaxation at infinite time, i.e., $\langle x_{ab}(\infty) x_{ab}(0) \rangle = \langle x_{ab}(\infty) \rangle \langle x_{ab}(0) \rangle$. If we define $\tilde{x}_{ab} = x_{ab} - \langle x_{ab} \rangle$, then Eq. (10) can be rewritten as

$$Z_{ab}(t) = \frac{\langle \tilde{x}_{ab}(t) \tilde{x}_{ab}(0) \rangle}{\langle \tilde{x}_{ab}(0) \tilde{x}_{ab}(0) \rangle}. \quad (11)$$

The quantity $\langle \tilde{x}_{ab}(t) \tilde{x}_{ab}(0) \rangle$, computed in the [supplementary material](#), is⁶⁸

$$\langle \tilde{x}_{ab}(t) \tilde{x}_{ab}(0) \rangle = \frac{4k_B T}{N_b k_s} \sum_{j=1}^{N_b-1} p_j q_j \exp \left[-\frac{t}{\tau_j} \right], \quad (12)$$

where

$$p_j = \left(1 - \cos \left[\frac{\pi j}{N_b} \right] \right)^{-1}, \quad (13)$$

$$q_j = \sin^2 \left[\frac{\pi(a+b-1)j}{2N_b} \right] \sin^2 \left[\frac{\pi(b-a)j}{2N_b} \right], \quad (14)$$

and

$$\tau_j = \frac{\zeta}{2k_s} p_j, \quad (15)$$

is the relaxation time of the j th Rouse-like mode. Clearly, many single exponential decays contribute to the decay of the mean extension normalized ACF. The relaxation time is given by the integral of the normalized ACF from time $t=0$ to a final time $t=T_e$

$$\tau = \int_{t=0}^{T_e} Z_{ab}(t) dt, \quad (16)$$

which leads to

$$\tau = \frac{\sum_{j=1}^{N_b-1} p_j q_j \tau_j (1 - \exp[-T_e/\tau_j])}{\sum_{j=1}^{N_b-1} p_j q_j}. \quad (17)$$

In theory, the upper limit on the integral is $T_e = \infty$.⁶⁸ However, when we compare this model to experimental data, we will treat T_e as the total experimental acquisition time, i.e., the number of frames multiplied by the frame exposure time t_e . Note that each independent Rouse-like mode j represents the local motion of the chain, which includes N_b/j beads with relaxation time approximated by τ_j . However, the actual relaxation time τ of segment (a, b) depends on all modes, i.e., all τ_j .

III. RESULTS AND DISCUSSION

A. Model parameterization

We focus here on the experimental system used in recent experiments on the dynamics of DNA during genome mapping.⁴⁰ The channel size is $H=40$ nm. The viscosity of the buffer was measured using an AR-G2 rheometer (TA instrument) with the cone-plate geometry⁶⁹ to be approximately 0.0022 Pa s at room temperature. The ionic strength of the buffer is approximately 100 mM. From Dobrynin's theory,^{70,71} the corresponding persistence length is $l_p = 52.2$ nm and, from Stigter's theory,⁶⁰ the effective width is $w = 5.6$ nm. The resulting effective channel size available to the chain is $H_{\text{eff}} = 34.4$ nm. For the hydrodynamic radius of a PERM bead, we use the value $a = 1.47$ nm.^{44,66} This value of bead radius was determined previously from the diffusivity of unconfined, short DNA and Yamakawa and Fujii's theory for the diffusivity of an ideal wormlike chain.⁷² Subsequent simulations of the unconfined Kirkwood diffusivity using this value of the hydrodynamic radius and our Kirkwood diffusion methodology led to excellent agreement with experimental data.⁶⁶ The free-solution value of the hydrodynamic radius applies in confinement as well because the wall effects are incorporated through far-field interactions.^{44,45} Using these input parameters, we conducted PERM simulations using 100 000 tours with a total length of $N_b^P = 15\,000$ beads per tour. For wall hydrodynamic interactions between segments and wall, we use the same grid-based method and the same grid parameters as used in the past.⁴⁴

Overall, the idea behind the parameterization is to find a reasonable coarse-graining length scale, N_b^{P*} , such that the criteria of having (i) no hydrodynamic interactions between Rouse

beads, (ii) minimal errors due to extra friction on end Rouse beads, and (iii) harmonic potential for the Rouse spring in the Rouse-like model are satisfied, as outlined in Sec. II B. We first start with finding a lower bound on N_b^{P*} based on criterion (i). Figure 2 shows the results for the friction coefficient per bead, ζ^P/N_b^P , as a function of N_b^P obtained from PERM simulations. The friction coefficient converges very quickly, consistent with arguments that hydrodynamic interactions are screened outside the length scale of the channel.⁵⁶ From these data, we select the lower bound on N_b^{P*} as 500 PERM beads because, beyond this value, the friction coefficient becomes approximately extensive in N_b^P .

To find the upper bound set by criterion (ii), we calculate in Fig. 3 the relaxation time of a whole chain of fixed length X , using Eq. (17), for varying values of the coarse-graining parameter N_b^{P*} . Equivalently, the number of Rouse beads is

$$N_b = \frac{N_b^P}{N_b^{P*}} + 1, \quad (18)$$

which is similar to Eq. (4). Values of other Rouse parameters k_s and ζ are easily mapped from the PERM results (shown in Figs. 2 and 4) for a given value of X . For example, $N_b^P = 7000$ PERM beads corresponds to $X = 18.18 \mu\text{m}$, and resulting $k_s^{\text{wc}} = 221.6 \text{ nN/m}$ and $\zeta^{\text{wc}} = 153.8 \text{ nN s/m}$ for the whole chain; the superscript “wc” stands for whole chain. Note that the smallest DNA considered in the genomic experiment in this work is $17 \mu\text{m}$, and hence we used $X = 18.18 \mu\text{m}$ (which is close to the minimum length used in experiments) to show the end effects. The spring constant is $k_s = (N_b - 1) k_s^{\text{wc}}$ and the friction coefficient is $\zeta = \zeta^{\text{wc}}/N_b$.

Results for the relaxation time τ for $X = 18.18 \mu\text{m}$ are shown in the left panel of Fig. 3. There is a strong end effect when $N_b \lesssim 10$ for all chain lengths, indicated by dashed green horizontal line in Fig. 3. Using $N_b = 10$ as the upper bound implies that the upper bound is $N_b^{P*} \approx 750$, which is indicated by the blue dashed vertical line. In the middle and right panel of Fig. 3, we also show the end effects for two other larger chains to verify that the minimum

FIG. 2. PERM results for the friction coefficient per bead as function of N_b^P . The black dashed vertical line indicates the cutoff $N_b^P = 500$, beyond which the friction coefficient becomes almost extensive in N_b^P , confirming the screening of hydrodynamic interactions.

FIG. 3. Relaxation time of three different chains, corresponding to $N_b^P = 7000, 10\,000, 15\,000$, for varying coarse-graining parameter N_b^{P*} . The left y-axis is for the relaxation time, indicated by blue circles, and the right y-axis is for the number of Rouse beads, indicated by the green curve corresponding to a given coarse-graining parameter N_b^{P*} . In all three plots, the blue dashed vertical lines indicate a reasonable choice of N_b^{P*} , and the green dashed horizontal lines indicate $N_b = 10$, which is a good choice for the minimum number of beads that must be present in a Rouse chain. The red circles in all three plots show the acceptable values of relaxation times.

value of $N_b = 10$ is a good choice, bearing in mind that the smallest chain in our experiments provides the most stringent upper bound.

Finally, to test criterion (iii), the top panel of Fig. 4 shows examples of the probability density functions for the chain extension, $\psi(X)$, for $N_b^P = 2, 20, 200, 2000$. Because the chain has a finite extension, $\psi(X)$ in all cases deviates from the Gaussian behavior, as evidenced by the normal probability plots shown in the bottom panel of Fig. 4 for respective chain lengths. This non-Gaussian behavior has also been seen in experimental studies in the past.^{8,40}

For our purposes, we want to determine the length scale for which the distribution of extension is as close as possible to the Gaussian distribution. In order to see the closeness between the empirical cumulative distribution function (CDF) of our extension data and the standard normal CDF, we calculate the root mean square of error (RMSE) between the two for a range of N_b^P in Fig. 5. In general, there is a steady decrease in RMSE all the way up to $N_b^P = 15\,000$. However, since the reduction in RMSE is quite small (0.09–0.04) over four decades of N_b^P , and the minimum value of RMSE occurs well above the upper bound we have already set, criterion (i) alone provides a lower bound on N_b^{P*} , namely, $N_b^{P*} = 500$ PERM beads.

There is a very narrow window between the lower bound (500) and the upper bound (750) on the choice of N_b^{P*} . A choice of 750 for N_b^{P*} would result in a more accurate value of the friction coefficient (see Fig. 2), but a less accurate value for τ due to end effects (see Fig. 3); the opposite would be true if N_b^{P*} was 500. Throughout this paper, we use $N_b^{P*} = 600$ as the

FIG. 4. Top panel: PERM results for the probability density function, $\psi(X)$, for four different chain lengths. Bottom panel: cumulative distribution function (CDF) for respective chain lengths. The blue solid line and the red dashed line indicate the empirical CDF and standard normal CDF, respectively. $\langle X \rangle$ and δ_X are the mean and standard deviation of X , respectively.

FIG. 5. The root mean square error (RMSE) between the empirical CDF and standard normal CDF for the extension data, as a function of N_b^P .

coarse-grained length scale to represent one spring in the Rouse-like model. Similarly, the friction coefficient of 600 PERM beads represents the friction coefficient of a single Rouse bead. The values of Rouse-like model parameters at this coarse-graining length scale are listed in Table I. Unlike the three parameters listed in Table I, the fourth parameter in the model, N_b , is set by the length of the chain. For clarity, Table II summarizes the conversion from PERM beads to Rouse springs. By way of example, the particular chain in Fig. 1(b) has an extension of $28.62 \mu\text{m}$ that corresponds to roughly 17 springs or, equivalently, $N_b = 18$ beads.

B. Comparison with the experimental data

To test the model, we compared the predictions of Eq. (17) using the parameters in Table I to experimental data on genomic DNA in nanochannels. We recently obtained a large videomicroscopy data set of labeled *E. coli* genomic DNA molecules that were analyzed primarily for static properties.⁴⁰ Here, we return to the same data set⁴⁰ and now extract the dynamic data. In these experiments, the DNA molecules were driven electrophoretically into square nanochannels of size $H = 40 \text{ nm}$, using the version 1 chip from BioNano Genomics, in a buffer of ionic strength of approximately $I = 0.1 \text{ M}$. A typical image of the DNA in the nanochannels appears in Fig. 1(a). The extension of the DNA molecules were in the range $17 - 124 \mu\text{m}$. Twenty-seven videos were acquired, each consisting of 330 frames of the label channel with a frame exposure time of 30 ms. An image of the YOYO-stained DNA backbone was taken immediately after the final image of the movie. The custom image processing program DM-static, which is available from BioNano Genomics, was used to analyze the raw images. The resulting time series were filtered for blinking and non-resolved labels using a set of filters developed

TABLE I. Mapped parameters for the Rouse-like model for coarse-graining parameter $N_b^{P*} = 600$.

Parameter	Value	Units
X_s	1.64	μm
k_s	2878	nN/m
ζ	13.27	nN s/m

TABLE II. Length scales in the multi-scale model. For the PERM model, the bond length is twice the hydrodynamic radius, $2a = 2.94$ nm, and the fractional extension is $X/L = 0.8421$. The conversion from nanometers to basepairs is 0.3 nm per base pair, neglecting any additional extension due to intercalation.

Entity	Contour length [bp]	Extension [μm]	PERM beads	Rouse springs
PERM bead	9.8	...	1	...
Rouse spring	6492	1.64	600	1
Example (Fig. 1(b))	113 288	28.62	11 560	$\cong 17$

separately from the DM-static algorithm.⁴⁰ The final data set includes a total of 1718 molecules and 18 589 internal segments.

Each molecule produces a kymograph, similar to the example in Fig. 1(c), that provides the position in the channel for each label as a function of time. The normalized ACF for the experimental data is computed as

$$Z_{ab}(t) = \frac{\sum_{j=1}^{T_f-t_f} (x_{ab}(j+t_f) - \langle x_{ab} \rangle) (x_{ab}(j) - \langle x_{ab} \rangle)}{\sum_{j=1}^{T_f} (x_{ab}(j) - \langle x_{ab} \rangle)^2}, \quad (19)$$

where x_{ab} is the separation distance between two labels a and b , T_f is the total number of frames, and j and t_f denote the lag/frame index. This definition of the normalized ACF is simply a discretized version of Eq. (10). In Eq. (19), the time series of the separation of label positions are known from experimental data, whereas in the Rouse-like model, normal modes are used to define label or bead positions as described in the [supplementary material](#). Although the buffer used in these experiments contains multiple additives to enhance photostability,⁴⁰ some of the labels will blink on and off. In the event that a label blinks off, we set the corresponding terms in the sums in Eq. (19) to zero when they involve that frame. We then fit the normalized ACF to a multiple exponential function to obtain the relaxation time τ . We found that some of the fits were clearly not exponential. Most likely, these fits were due to transient large compression or extension. However, we cannot make a definitive analysis because we only have a single image of the DNA backbone at the end of the movie. These outliers were removed from the data set; more details about the criteria to remove these outliers are provided in the [supplementary material](#).

We chose to work with genomic DNA extracted from *E. coli*,⁴⁰ rather than a model system (e.g., barcoded λ -phage DNA) for two reasons. First and foremost, genomic mapping in nano-channels is intended to acquire information from large genomic DNA molecules, so these samples are most relevant to the intended application. Second, it is relatively easy to acquire a very large data set over a wide range of molecular weights and label separation distances with a sample extracted from an organism. The downside of our approach is that essentially every molecule that we image is a unique sample, with a different molecular weight and different label pattern. As a result, we are forced to bin these data in order to obtain sufficient statistics. In what follows, we chose to bin the molecules first as a function of their extension X . Then, within a particular X bin, we sub-bin the data in terms of the label separation distance x_{ab} . We only considered bins which contained at least 20 data points. All of the values involved in the experimental data analysis are listed in Table III. Additional details about the experimental protocol are provided in the [supplementary material](#). While our model suggests that there is a further dependence on the location of the labels with respect to the end of the molecule, we do not have enough data to support a third level of binning. We will consider the effect of label location from the model in Sec. III C.

TABLE III. Parameters in experimental data analysis.

Experimental parameter	Value
X bin size	$9.27 \mu\text{m}$ (37 kbp)
x_{ab} bin size	$6.18 \mu\text{m}$ (25 kbp)
Number of segments before (after) filtering	18 589 (12 173)
Number of molecules before (after) filtering	1718 (480)
Final range of x_{ab}	$[0.5, 46.5] \mu\text{m}$
Final range of X	$[17, 62] \mu\text{m}$

By way of example, Fig. 6 displays the normalized ACF, Z_{ab} , for small distances between labels on small molecules (Fig. 6(a)) and large distances between labels on large molecules (Fig. 6(b)). Note that the mean for a given time lag is obtained by averaging the value of Z_{ab} of all segments for a particular combination of X and x_{ab} bins. We chose to represent the uncertainty in the data by the standard deviation over the values of Z_{ab} within a given bin. There are certainly errors in the measurement of x_{ab} for any particular label pair due to the imaging optics and the method for extracting the label positions. In addition to propagation of this error to Z_{ab} , the autocorrelation calculation also has errors arising from the limited sample size and blinking labels. Nevertheless, the dominant source of variance within a given bin likely arises from the distribution of molecular weights and label distances within that bin.

Overall, Fig. 6 makes it clear that longer segments relax more slowly than smaller segments, as would be expected. However, in both cases there are substantial variances in the correlation, especially at long lags. We attribute this phenomenon to two likely factors: (i) the finite duration of the videos, which limits the number of periods that can be realized for long Rouse modes, and (ii) transient compression or extension.

Our theoretical analysis produced a result in Eq. (17) for the time constant characterizing the decay in the autocorrelation. To compare this formula with the experimental data, we made a multiple exponential fit

$$Z_{ab}(t) = \frac{\sum_{k=1}^{N_e} A_k e^{-t/\theta_k}}{\sum_{k=1}^{N_e} A_k}, \quad (20)$$

FIG. 6. Examples of normalized extension autocorrelation functions, Z_{ab} [Eq. (19)], as a function of time for two bins. (a) DNA molecules with extensions $X = [18, 28] \mu\text{m}$ and label separation distances $x_{ab} = [0.5, 6] \mu\text{m}$. (b) DNA molecules with extensions $X = [37, 46] \mu\text{m}$ and label separation distances $x_{ab} = [31, 37] \mu\text{m}$. The solid lines indicate the average value of Z_{ab} in the bin; the shading indicates the standard deviation of Z_{ab} within the bin.

to the first 100 data points in plots similar to Fig. 6 for each label pair combination. In Eq. (20), N_e is the number of exponentials, which we choose to be 10 in this work, and A_k and θ_k are the coefficients and time constants for the k th exponential, respectively. After fitting the ACF using multiple exponential function, we use Eq. (16) with $T_e = 10$ s to calculate τ . Figure 7 shows the relaxation time data for every molecule and every pair of labels with segment lengths that lie within the ranges listed in Table III.

There are only a few data points for $x_{ab} > 37.1 \mu\text{m}$, hence they are not shown in the figure. The spread in the data is quite large, consistent with the shading in the typical binned ACFs plotted in Fig. 6. Also, there are substantially more data points for smaller X and x_{ab} than for larger ones. The solid lines in all panels in Fig. 7(a) show a linear model regression for τ as a function of X . For all six segment bins, we observe a positive correlation between τ and X as quantified by the corresponding p -values. The p -values less than 0.05 indicate the rejection of the null hypothesis that τ does not depend on X , at the $\alpha = 0.05$ confidence level. We also perform a statistical comparison of the slopes of all segment size bins using an Analysis of Covariance at the $\alpha = 0.05$ level, followed by multiple individual pair comparisons using Tukey's honest significant difference criterion. These comparisons are shown in Table IV. All but three of the slope pairs are significantly different.

Although Fig. 7 shows almost all of the data we obtained, it is difficult to (i) draw clear inferences about the dependencies of τ on X and x_{ab} from all of the data points, and (ii) compare the experimental data with Rouse-like model results. Therefore, we also bin our data in X . For each bin, we then compute the median, and the first and the third quantiles, which are depicted in Fig. 8(a). Poorer statistics are seen for τ in Fig. 8(b) for larger values of X and x_{ab} because, as evident from Fig. 7, there are relatively fewer data points for larger values of X and x_{ab} .

Figure 8 indicates that (i) for a given molecule extension X , the relaxation time of a segment within that molecule increases with the length of the segment, and (ii) for a given

FIG. 7. Experimental data for the relaxation time as a function of X for six different segment size bins indicated on each panel. Each data point corresponds to a single pair of labels obtained in the experiment. The value of X is the extension of the DNA molecule containing that label pair. Black solid lines indicate a linear regression through the data points and p -values indicate that the slopes in each case are statistically significantly different from zero.

TABLE IV. Analysis of Covariance with multiple individual pair comparisons for the values of the slope of τ vs. X for different x_{ab} ; the values of x_{ab} shown are the center values of corresponding bins. p -values are shown for each pair, where bold font indicates statistically significant difference, and normal font represents no difference. Confidence is at the $\alpha = 0.05$ level and Tukey's honest significant difference criterion was used for comparisons.

x_{ab} [μm]	9.3	15.5	21.7	27.9	34.1
3.1	0.037	<0.001	<0.001	<0.001	<0.001
9.3		0.179	<0.001	<0.001	0.004
15.5			0.002	0.016	0.023
21.7				0.982	0.406
27.9					0.698

segment length x_{ab} , the relaxation time also increases with increasing size of the entire molecule. Both the trends in the data are captured by the Rouse-like model in Fig. 8(b), where τ is calculated using Eq. (17) with the parameters in Table I. For comparison with the experiments, we use $T_e = 10$ s in Eq. (17), which is the experimental video duration. Figure 8 shows a qualitative agreement between the experimental results and the Rouse-like model predictions. A similar level of experimental agreement was achieved by Carpenter *et al.*⁵⁰ and Karpusenko *et al.*,⁵¹ who used a simple one-dimensional overdamped oscillator chain model to predict and compare the dynamics of density fluctuations in DNA in larger channels. Therefore, our work corroborates the use of simplified models to describe these experiments.

Quantitatively, our Rouse-like model results are roughly within a factor of 2 compared to our experimental data. This quantitative discrepancy can be partially attributed to the excess friction coefficient assigned to end beads. This excess friction coefficient leads to overestimation of the relaxation time. Adjusting ζ for the end beads would result in a lower value of τ , albeit with a substantial increase in the complexity of the normal mode analysis. Moreover, on the experimental side, the limited video duration results in only a lower bound on the τ for larger segments and/or molecules. Thus, we expect larger τ for those segments and molecules in the experimental data. Additionally, for short segments with very short relaxation times, in many cases our calculated value will be merely an upper bound due to the duration of a frame, here optimized at 0.033 ms.

It is important to understand the effect of the overall length of a molecule on the relaxation dynamics of a given internal segment. Figure 8 shows that a segment of size x_{ab} relaxes much faster if it is a part of smaller DNA molecule, as opposed to being part of the larger DNA molecules. In general, this trend is captured by the model because the number of Rouse-like modes, and thus the relaxation time, increases with molecule length. Moreover, the molecule length

FIG. 8. Relaxation time as a function of the extension X of the DNA molecules for different label separation distances x_{ab} in (a) experiments and (b) the Rouse-like model in Eq. (17). For (a), the data points are the median, and the error bars are constructed using the first and the third quantiles of the τ distribution within a given bin.

dependence of the relaxation time increases with segment length, x_{ab} , as evidenced by the rising slopes in Fig. 8.

C. Effects of molecular weight and label locations

The relatively good agreement between the model and experiments in Fig. 8 allows us to use the model to investigate, in more detail, effects of molecular weight, label separation, and label location. These are challenging parameters to examine by experiment. In principle, one could use modern molecular biology methods to design molecules with precise molecular weights and label locations. Indeed, such a strategy proved useful in the development of genome mapping in nano-channels.¹ However, constructing a set of molecules sufficient to study the physics over a wide region of the phase space would be prohibitively expensive, let alone the additional costs required for obtaining and analyzing the autocorrelation data. Thus, the model developed here presents an ideal opportunity to explore the phase space, keeping in mind that the accuracy of any conclusions obtained is likely to have an accuracy similar to that seen in Fig. 8.

Using the Rouse-like model, we can study the effect of segment position on τ in detail by choosing various values of a and b in Eq. (17). In this analysis we set $T_e = \infty$ in order to capture the much longer relaxation time of very long molecules in the theoretical limit. Figure 9(a) plots τ as a function of the normalized relative position on the chain from the end

$$d_{\text{rel}} = \left(\frac{a+b}{2} - 1 \right) \frac{X_s}{X}, \quad (21)$$

for four different molecule lengths and for $x_{ab} = 3.29 \mu\text{m}$, which is the mean extension of a segment consisting of two Rouse springs. The four different molecules in Fig. 9(a) correspond to Rouse chains of $N_b = 10, 20, 40, 80$. Since there is no directionality in our model at equilibrium, these results are symmetric about the center of the chain, and we only plot the left side. Similarly, Fig. 9(b) displays the dependence of τ on the position of the segment for various segment lengths x_{ab} (corresponding to $b - a = 2, 4, 6, 8, 16, 32, 64$) in a fixed chain length of $X = 129.8 \mu\text{m}$ (corresponding to $N_b = 80$). The results indicate that, for a given segment length, locations on the edge of the chain relax much faster than locations in the center of the chain. This dynamic inhomogeneity has also been observed before in the context of polymer melts of polystyrene and polyisoprene.^{73,74} The latter studies used atomistic simulations to study the influence of position of certain segments of polystyrene and polyisoprene on the dynamics of segments. This end effect becomes stronger as the molecule length increases. In order to see the effect of molecule length on end effects, we plot the ratio of the relaxation time of center segments to that of the outermost segments, τ_{ratio} , as a function of X in Fig. 9(c). For smaller DNA, this ratio is of the order one and the end effects are not strong. This suggests that the

FIG. 9. (a) Effect of the segment position on the relaxation dynamics for segments of size $x_{ab} = 3.29 \mu\text{m}$ ($b - a = 2$) for four different molecule lengths. (b) Effect of segment position on the relaxation dynamics for segments of seven different lengths in a molecule of length $X = 129.8 \mu\text{m}$. (c) The ratio of the relaxation times of center segments to that of the outermost segments, τ_{ratio} , as a function of X for $x_{ab} = 1.64 \mu\text{m}$ ($b - a = 1$), which is the shortest segment. For large X , there is an apparent power law with an exponent close to unity.

averaging of τ for all positions in our experimental data analysis is valid only for the shorter molecules. This effect likely contributed to the increased variance in the relaxation time for the larger segments in Fig. 8(a), because these segments are more likely to be present on larger molecules.

D. Impact on genome mapping

The consensus genome map for a biological sample is constructed by obtaining images similar to Fig. 1(b) for a large number of molecules such that the total number of base pairs imaged exceeds at least $30\times$ the size of the genome. For a human genome, this corresponds to approximately 10^{11} base pairs or, approximately, 30 m of DNA. This measurement is accomplished in a single day's experiment using the version 2 chip available from BioNano Genomics. The fluorescent labels on each of the fragments within this football-field worth of DNA are then assembled by finding all of their overlapping regions. The assembly is done in a bioinformatic pipeline, where a set of statistical weights is applied to each individual molecule.⁷⁵ As stated previously, there are number of sources of error, such as topological events, mutations, channel size variations, and dynamic fluctuations, that together constitute an overall error in label distance measurements. This overall error is then used to determine the statistical weight to assign to a particular measurement when constructing the genomic map. The information gained in this work studying the relaxation time as a function of segment length, molecule length, and segment position isolates the effects of dynamic fluctuations, which can improve the determination of the statistical weights. For example, the exposure time of 150 ms used for standard genome mapping roughly bisects the data in Fig. 8(a). This means that a snapshot fails to cover the fluctuation of a single relaxation for most of the larger segments. Whereas for the smaller segments, their extension distribution will be fairly completely sampled within one exposure time. As such, a larger statistical weight can be applied to shorter segments. We give a rough example calculation in the [supplementary material](#) that further amplifies this application of our model.

Considering the effect of position, the outermost segments will relax faster than the center segments, as can be seen from the general trends obtained solely from the Rouse-like model in Fig. 9. As a result, the distribution of the barcode distances is narrower for outermost segments. Thus, more statistical weight can be given to the outermost segments and less can be given to the center segments.

Figure 9(a) also shows that the relaxation time for outermost segments displays a weak dependence on X for larger X . For clarity, this trend is displayed in Fig. 10(a) across a wide range of molecular weights for three different segment lengths. A similar conclusion about outermost segments was made by Stephanou *et al.*⁷⁶ but in the context of entangled polymer melts

FIG. 10. Effect of extension X on the relaxation times of internal segments of three different lengths ($b - a = 3, 7, 15$) located at: (a) outermost, and (b) center positions.

using primitive-path analysis on the trajectories obtained from molecular dynamics atomistic simulations. These results suggest that the accuracy in the measurements of barcode distances near the outermost segments do not decrease by increasing X . In other words, even for very large DNA chains, the accuracy level for outermost segments is the same as it would be in a small DNA chain. However, as shown in Fig. 10(b), the relaxation time strongly depends on X for center segments, suggesting a decrease in accuracy with an increase in X . This suggests that a larger statistical weight can be applied to outermost segments as compared to center segments.

IV. CONCLUDING REMARKS

In this work, we have developed a multi-scale approach to model the dynamics of internal segments of nanoconfined DNA that ultimately produces a one-dimensional, Rouse-like model using input parameters that are readily available from the properties of unconfined bulk DNA and its buffer. While the limitations of the experimental data preclude a thorough model validation, this simple one-dimensional model captures the general trends in experimental data for DNA during genome mapping in nanochannels. This parameterized model provides further insights into the effects of molecular weight and label location that would be both expensive and difficult to obtain via experiments. Our results also impact the practice of genome mapping in nanochannels by identifying label pairs where the measurement of the label locations is likely to be impacted by a relaxation process that is slow compared to the imaging time. Indeed, as genome mapping moves towards ever longer molecules, the effects identified here will become increasingly important in optimizing the system for throughput and accuracy.

The modeling approach we used here is generic for channel-confined, semiflexible chains and thus easily extended to study other regimes of confinement or other polymers. In weaker confinement, however, it may prove convenient to move to a less resolved fine-scale model for computational efficiency.⁷⁷ Likewise, while we found that normal-mode analysis was ideal for the problem we considered here, it is quite easy to simulate the Rouse-like model by Brownian dynamics. One should use caution, however, when using the model developed here to study strongly non-equilibrium processes such as the nanodozer assay.⁵² In order for the model to be valid, the statistics of the fine-scale model (which are obtained from equilibrium simulations) need to remain valid during simulation of the coarse-grained model. While this limitation does not impose any constraints on studying the equilibrium or near-equilibrium properties of channel-confined DNA, it constrains the model's use when non-equilibrium effects are important. Finally, the model developed here opens up the opportunity for *in silico* engineering of genome mapping technologies. Simulations of the Rouse-like model, with appropriate parameterization, furnish data on the dynamics of the label positions. Given the speed of these simulations, they offer an enticing avenue for optimizing operating parameters (exposure time, DNA molecular weight, label pattern) and device design for maximum throughput and accuracy.

SUPPLEMENTARY MATERIAL

See [supplementary material](#) for a derivation of the autocorrelation time for the Rouse-like model, additional information on the fitting of the experimental data, and the effect of exposure time.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (Grant No. R01-HG006851). We thank Seunghwan Shin for the measurement of the buffer viscosity. Part of this work was carried out in the College of Science and Engineering Polymer Characterization Facility, University of Minnesota, which has received capital equipment funding from the NSF through the UMN MRSEC program under Award No. DMR-1420013. Computational resources were provided in part by the Minnesota Supercomputing Institute at the University of Minnesota. Jeffrey G. Reifenger and Han Cao are employees of BioNano Genomics, which is commercializing nanochannel genome mapping.

- ¹S. K. Das, M. D. Austin, M. C. Akana, P. Deshpande, H. Cao, and M. Xiao, *Nucleic Acids Res.* **38**, e177 (2010).
- ²E. T. Lam, A. Hastie, C. Lin, D. Ehrlich, S. K. Das, M. D. Austin, P. Deshpande, H. Cao, N. Nagarajan, M. Xiao, and P.-Y. Kwok, *Nat. Biotechnol.* **30**, 771 (2012).
- ³A. R. Hastie, L. Dong, A. Smith, J. Finklestein, E. T. Lam, N. Huo, H. Cao, P.-Y. Kwok, K. R. Deal, J. Dvorak, M.-C. Luo, Y. Gu, and M. Xiao, *PLoS One* **8**, e55864 (2013).
- ⁴M. Levy-Sakin and Y. Eberstein, *Curr. Opin. Biotechnol.* **24**, 690 (2013).
- ⁵H. Cao, A. R. Hastie, D. Coa, E. T. Lam, Y. Sun, H. Huang, X. Liu, L. Lin, W. Andrews, S. Chan, S. Huang, X. Tong, M. Requa, T. Anantharaman, A. Krogh, H. Yang, H. Cao, and X. Xu, *GigaScience* **3**, 34 (2014).
- ⁶M. Pendleton, R. Sebra, A. W. C. Pang, A. Ummat, O. Franzen, T. Rausch, A. M. Stutz, W. Stedman, T. Anantharaman, A. Hastie, H. Dai, M. H. Y. Fritz, H. Cao, A. Cohain, G. Deikus, R. E. Durrett, S. C. Blanchard, R. Altman, C.-S. Chin, Y. Guo, E. E. Paxinos, J. O. Korbel, R. B. Darnell, W. R. McCombie, P.-Y. Kowk, C. E. Mason, E. E. Schadt, and A. Bashir, *Nat. Methods* **12**, 780 (2015).
- ⁷C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith, *Science* **265**, 1599 (1994).
- ⁸W. F. Reinhart, J. G. Reifengerger, D. Gupta, A. Muralidhar, J. Sheats, H. Cao, and K. D. Dorfman, *J. Chem. Phys.* **142**, 064902 (2015).
- ⁹T. Odijk, *Macromolecules* **16**, 1340 (1983).
- ¹⁰T. Odijk, *Phys. Rev. E* **77**, 060901 (2008).
- ¹¹J. Z. Y. Chen, *Macromolecules* **46**, 9837 (2013).
- ¹²Y. Yang, T. W. Burkhardt, and G. Gompper, *Phys. Rev. E* **76**, 011804 (2007).
- ¹³F. Wagner, G. Lattanzi, and E. Frey, *Phys. Rev. E* **75**, 050902(R) (2007).
- ¹⁴T. W. Burkhardt, Y. Yang, and G. Gompper, *Phys. Rev. E* **82**, 041801 (2010).
- ¹⁵P. Cifra, *J. Chem. Phys.* **131**, 224903 (2009).
- ¹⁶P. Cifra, Z. Benkova, and T. Bleha, *J. Phys. Chem. B* **113**, 1843 (2009).
- ¹⁷Y. Wang, D. R. Tree, and K. D. Dorfman, *Macromolecules* **44**, 6594 (2011).
- ¹⁸T. Su, S. K. Das, M. Xiao, and P. K. Purohit, *PLoS One* **6**, e16890 (2011).
- ¹⁹Y. Wang, W. F. Reinhart, D. R. Tree, and K. D. Dorfman, *Biomicrofluidics* **6**, 014101 (2012).
- ²⁰P. Cifra, *J. Chem. Phys.* **136**, 024902 (2012).
- ²¹P. Cifra and T. Bleha, *Soft Matter* **8**, 9022 (2012).
- ²²Z. Benkova and P. Cifra, *Macromolecules* **45**, 2597 (2012).
- ²³L. Dai, S. Y. Ng, P. S. Doyle, and J. R. C. van der Maarel, *ACS Macro Lett.* **1**, 1046 (2012).
- ²⁴R. Chang and K. Jo, *J. Chem. Phys.* **136**, 095101 (2012).
- ²⁵W. F. Reinhart, D. R. Tree, and K. D. Dorfman, *Biomicrofluidics* **7**, 024102 (2013).
- ²⁶Y.-L. Chen, *Biomicrofluidics* **7**, 054119 (2013).
- ²⁷D. R. Tree, Y. Wang, and K. D. Dorfman, *Phys. Rev. Lett.* **110**, 208103 (2013).
- ²⁸A. Muralidhar, D. R. Tree, and K. D. Dorfman, *Macromolecules* **47**, 8446 (2014).
- ²⁹Y.-L. Chen, Y.-H. Lin, J.-F. Chang, and P.-K. Lin, *Macromolecules* **47**, 1199 (2014).
- ³⁰A. Muralidhar, D. R. Tree, Y. Wang, and K. D. Dorfman, *J. Chem. Phys.* **140**, 084905 (2014).
- ³¹A. Muralidhar and K. D. Dorfman, *Macromolecules* **49**, 1120 (2016).
- ³²A. Muralidhar, M. Quevillon, and K. D. Dorfman, *Polymers* **8**, 79 (2016).
- ³³C. Manneschi, P. Fanzio, T. Ala-Nissila, E. Angeli, L. Repetto, G. Firpo, and U. Valbusa, *Biomicrofluidics* **8**, 064121 (2014).
- ³⁴C. Manneschi, E. Angeli, T. Ala-Nissila, L. Repetto, G. Firpo, and U. Valbusa, *Macromolecules* **46**, 4198 (2013).
- ³⁵J. O. Tegenfeldt, C. Prinz, H. Cao, S. Chou, W. W. Reisner, R. Riehn, Y. M. Wang, E. C. Cox, J. C. Sturm, P. Silberzan, and R. H. Austin, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 10979 (2004).
- ³⁶W. Reisner, K. J. Morton, R. Riehn, Y. M. Wang, Z. Yu, M. Rosen, J. C. Sturm, S. Y. Chou, E. Frey, and R. H. Austin, *Phys. Rev. Lett.* **94**, 196101 (2005).
- ³⁷K. Jo, D. M. Dhirga, T. Odijk, J. J. de Pablo, M. D. Graham, R. Runnheim, D. Forrest, and D. C. Schwartz, *Proc. Natl. Acad. Sci. U. S. A.* **104**, 2673 (2007).
- ³⁸Y. Kim, K. S. Kim, K. L. Kounovsky, R. Chang, G. Y. Jung, J. J. de Pablo, K. Jo, and D. C. Schwartz, *Lab Chip* **11**, 1721 (2011).
- ³⁹D. Gupta, J. Sheats, A. Muralidhar, J. J. Miller, D. E. Huang, S. Mahshid, K. D. Dorfman, and W. Reisner, *J. Chem. Phys.* **140**, 214901 (2014).
- ⁴⁰J. Sheats, J. G. Reifengerger, H. Cao, and K. D. Dorfman, *Biomicrofluidics* **9**, 064119 (2015).
- ⁴¹J. G. Reifengerger, K. D. Dorfman, and H. Cao, *Analyst* **140**, 4887 (2015).
- ⁴²D. R. Tree, Y. Wang, and K. D. Dorfman, *Biomicrofluidics* **7**, 054118 (2013).
- ⁴³T. Cui, J. Ding, and J. Z. Y. Chen, *Phys. Rev. E* **78**, 061802 (2008).
- ⁴⁴A. Muralidhar and K. D. Dorfman, *Macromolecules* **48**, 2829 (2015).
- ⁴⁵R. M. Jendrejack, J. J. de Pablo, and M. D. Graham, *J. Chem. Phys.* **116**, 7752 (2002).
- ⁴⁶P. Grassberger, *Phys. Rev. E* **56**, 3682 (1997).
- ⁴⁷T. Prellberg and J. Krawczyk, *Phys. Rev. Lett.* **92**, 120602 (2004).
- ⁴⁸P. E. Rouse, Jr., *J. Chem. Phys.* **21**, 1272 (1953).
- ⁴⁹B. Akerman and E. Tuite, *Nucleic Acids Res.* **24**, 1080 (1996).
- ⁵⁰J. H. Carpenter, A. Karpusenko, J. Pan, S. F. Lim, and R. Riehn, *Appl. Phys. Lett.* **98**, 253704 (2011).
- ⁵¹A. Karpusenko, J. H. Carpenter, C. Zhou, S. F. Lim, J. Pan, and R. Riehn, *J. Appl. Phys.* **111**, 024701 (2012).
- ⁵²A. Khorsid, P. Zimny, D. Tetreault-La Roche, G. Massarelli, T. Sakaue, and W. Reisner, *Phys. Rev. Lett.* **113**, 268104 (2014).
- ⁵³A. Khorsid, S. Amin, Z. Zhang, T. Sakaue, and W. Reisner, *Macromolecules* **49**, 1933 (2016).
- ⁵⁴J. G. Kirkwood and J. Riseman, *J. Chem. Phys.* **16**, 565 (1948).
- ⁵⁵B. H. Zimm, *J. Chem. Phys.* **24**, 269 (1956).
- ⁵⁶F. Brochard and P. G. de Gennes, *J. Chem. Phys.* **67**, 52 (1977).
- ⁵⁷J. P. Hernández-Ortiz, J. J. de Pablo, and M. D. Graham, *J. Chem. Phys.* **125**, 164906 (2006).

- ⁵⁸J. P. Hernández-Ortiz, J. J. de Pablo, and M. D. Graham, *Phys. Rev. Lett.* **98**, 140602 (2007).
- ⁵⁹M. D. Graham, *Annu. Rev. Fluid Mech.* **43**, 273 (2011).
- ⁶⁰D. Stigter, *Biopolymers* **16**, 1435 (1977).
- ⁶¹J. Skolnick and M. Fixman, *Macromolecules* **10**, 944 (1977).
- ⁶²T. Odijk, *J. Polym. Sci. Pol. Phys.* **15**, 477 (1977).
- ⁶³A. Dobrynin, *Macromolecules* **38**, 9304 (2005).
- ⁶⁴D. R. Tree, Y. Wang, and K. D. Dorfman, *Phys. Rev. Lett.* **108**, 228105 (2012).
- ⁶⁵D. Gupta, J. J. Miller, A. Muralidhar, S. Mahshid, W. Reisner, and K. D. Dorfman, *ACS Macro Lett.* **4**, 759 (2015).
- ⁶⁶D. R. Tree, A. Muralidhar, P. S. Doyle, and K. D. Dorfman, *Macromolecules* **46**, 8369 (2013).
- ⁶⁷A. Jain and K. D. Dorfman, *Biomicrofluidics* **9**, 024112 (2015).
- ⁶⁸S. Yang, J. B. Witkoskie, and J. Cao, *J. Chem. Phys.* **117**, 11010 (2002).
- ⁶⁹L. Bai, J. W. Fruehwirth, X. Cheng, and C. W. Macosko, *Soft Matter* **11**, 5282 (2015).
- ⁷⁰A. V. Dobrynin, *Macromolecules* **39**, 9519 (2006).
- ⁷¹C.-C. Hsieh, A. G. Balducci, and P. S. Doyle, *Nano Lett.* **8**, 1683 (2008).
- ⁷²H. Yamakawa and M. Fujii, *Macromolecules* **6**, 407 (1973).
- ⁷³R. Faller, *Macromolecules* **37**, 1095 (2004).
- ⁷⁴Q. Sun and R. Faller, *Macromolecules* **39**, 812 (2006).
- ⁷⁵A. Valouev, "Shotgun optical mapping: A comprehensive statistical and computational analysis," Ph.D. thesis (University of Southern California, 2006).
- ⁷⁶P. S. Stephanou, C. Baig, G. Tsolou, V. G. Mavrantzas, and M. Kröger, *J. Chem. Phys.* **132**, 124904 (2010).
- ⁷⁷L. Dai, J. R. C. van der Maarel, and P. S. Doyle, *Macromolecules* **47**, 2445 (2014).