Distribution of distances between DNA barcode labels in nanochannels close to the persistence length

Wesley F. Reinhart,^{1,a)} Jeff G. Reifenberger,² Damini Gupta,¹ Abhiram Muralidhar,¹ Julian Sheats,¹ Han Cao,² and Kevin D. Dorfman^{1,b)}

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

²BioNano Genomics, 9640 Towne Centre Dr., Ste. 100, San Diego, California 92121, USA

(Received 16 November 2014; accepted 25 January 2015; published online 10 February 2015)

We obtained experimental extension data for barcoded *E. coli* genomic DNA molecules confined in nanochannels from 40 nm to 51 nm in width. The resulting data set consists of 1 627 779 measurements of the distance between fluorescent probes on 25 407 individual molecules. The probability density for the extension between labels is negatively skewed, and the magnitude of the skewness is relatively insensitive to the distance between labels. The two Odijk theories for DNA confinement bracket the mean extension and its variance, consistent with the scaling arguments underlying the theories. We also find that a harmonic approximation to the free energy, obtained directly from the probability density for the distance between barcode labels, leads to substantial quantitative error in the variance of the extension data. These results suggest that a theory for DNA confinement in such channels must account for the anharmonic nature of the free energy as a function of chain extension. © 2015 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4907552>]

I. INTRODUCTION

One of the outstanding questions in polymer confinement concerns the statistics of a wormlike chain, such as DNA, confined in a nanochannel whose size D is commensurate with the persistence length, l_p , of the polymer. The importance of this problem extends beyond chemical physics, as these channel sizes form the basis for an emerging method of genomic mapping.^{1–3} Here, the DNA is stretched by confinement in a nanochannel^{4,5} and the genomic information is then read by fluorescence microscopy, either using sequence-specific probes,^{6–11} the absence of fluorescence due to localized melting of AT-rich regions,^{12,13} or competitive binding.¹⁴ Genome mapping in nanochannels uses intact genomic DNA molecules that are hundreds of kilobases (and even megabases) in contour length. These massive stretches of labeled DNA greatly facilitate the assembly of genomic maps for anchoring sequencing data,² and they are especially important for the analysis of structural variations.^{3,15}

Understanding the probability distribution governing the polymer extension between labeled points along nanochannel-confined DNA is critical to accurately map the physical distance X (in nm) between probes to their genomic distance, L (in base pairs (bp)). However, the theory surrounding DNA confinement for channel sizes $D \approx l_p$ is contentious.^{16–22} It would be greatly beneficial to have a fundamental understanding of the thermodynamics of DNA confinement in circa 50 nm channel sizes to guide the development of the emerging genome barcoding technology.

In the present contribution, we use a genome mapping system^{1,2,23} to make massively parallel measurements of DNA extension at the single-molecule level on barcoded DNA molecules. The approach begins with sheared, labeled genomic DNA from *E. coli*. These DNAs are of random lengths, which we do not know *a priori*, but the set of labels on the DNA (their barcode) corresponds to a specific sequence motif. These DNAs are stretched in a nanochannel, leading to images similar to Fig. 1. By aligning the barcodes on these molecules to the reference genome, we are able to obtain the contour length (genomic distance) between the probes. The subsequent analysis involves thousands of molecules per channel size, with the labels on these molecules containing hundreds of thousands of measurements of the distance between barcode label pairs. This high data density on single molecules allows us to construct estimates for the probability distribution $P(X; L)$ for the extension X for a given contour length L between the labels. Provided that the distance between labels is sufficiently long, this information is equivalent to the extension of a chain of length L .²⁴ We thus not only address the moments of the distribution for X , as is normally done in theoretical modeling, but also the underlying theoretical approach, which makes certain assumptions about the nature of this probability distribution.

II. METHODS

A. Genome mapping in nanochannels

Genomic DNA was extracted from the MG1655 strain of *E. coli* and according to commercial protocols using the IrysPrep Reagent Kit (BioNano Genomics, Inc.) as reported elsewhere in the context of human cells.¹⁵ The labeled DNA has fluorescent dUTP inserted at the Nt.BspQI nicking site

^{a)}Present address: Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544, USA.

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

FIG. 1. Optical images obtained during genome mapping in nanochannels. (a) Composite, false color image of the DNA backbone (green) and the sequence-specific probes (red). (b) DNA backbone. (c) Sequence-specific probes.

and YOYO-1 (Invitrogen) in the backbone at a ratio of 40 bp per dye molecule. Image data were collected on a custom-built epifluorescence microscope consisting of an Olympus IX-71 microscope body with an Olympus 60x air NA 0.9 objective. The system uses two coherent lasers: an OBIS 488 nm, 150 mW laser to image the YOYO-stained DNA molecule and a Sapphire 532 nm, 200 mW laser for the fluorescent probes at the nick sites. Fluorescence is collected through a high speed filter wheel (Finger Lakes Instrumentation) with a 525/50 or a 579/34 filter (Semrock), depending on the laser color. Molecules and labels were imaged with a Zyla 5.5 sCMOS camera (Andor). Autofocus is maintained during imaging with a CRISP system and a z -axis motor (Applied Scientific Instrumentation). The xy -stage is also from Applied Scientific Instrumentation. Finally, the entire system is controlled with Micromanager.²⁵

A version 1 chip available from BioNano Genomics was used with the *E. coli* sample. The DNA was loaded into the nanochannels with a custom electrophoresis script available from BioNano Genomics as part of the Irys system. Once the DNA was loaded in the array of approximately square nanochannels, the DNA was imaged once at each position with the 532 nm and 488 nm lasers for 150 ms for each color.

Individual reads were mapped to the MG1655 *E. coli* reference using a custom algorithm from BioNano Genomics that follows published procedures.²⁶ A prealignment filter step was applied requiring molecules to be greater than 50 kilobases. The alignments of molecules whose p -value was less than 10^{-4} were kept for further analysis. Note that the alignment process is flexible enough to handle incomplete labeling of the chain; the alignment process provides the location of the dots visible in images such as Fig. 1, even if some of the expected dots are not present.

In order to avoid possible artifacts that might occur due to incomplete labeling or actual genomic variations (such as an insertion or a deletion), we scanned through the barcode patterns for all of the molecules that met the alignment and total molecular weight cutoffs and extracted each section of the molecule that contains at least 5 contiguous perfect matches. In some cases, the entire molecule aligns to the genome. This

is the best case scenario, where we obtain measurements from barcode labels that are displaced by hundreds of thousands of base pairs. In other cases, there are portions of the molecule aligned perfectly to the genome, separated by regions whose barcode patterns indicate insertions or deletions (or, possibly, missing labels that lead to the same genomic conclusion). In those cases, we extracted the aligned regions and treated them as individual “sub-molecules,” thereby excising the region containing uncertain genomic information.

B. Uniform extension filter

The alignment to the reference genome is a requisite for our analysis, since it allows us to determine the contour length of the portion of the barcoded DNA molecule that is aligned. The alignment algorithm is statistical in nature, and thus returns a confidence (p -value) that the molecule was correctly aligned. The approach described above is a conservative threshold to avoid unnecessary data attrition—we expect that the probability distributions for the extension between labels will be somewhat broad, so we do not want to remove these thermal fluctuations from our ensemble through an overly restrictive threshold for alignment to the genome. The tradeoff for this approach is that we may be including molecules that are not aligned to the proper location in the genome. To help remove these molecules, we implemented a second test that identifies molecules where the barcode does not indicate uniform extension of the chain.

To construct this “uniform extension filter,” we note that the physical positions X_i of the barcode labels $i = 1, 2, \dots, k$ on a linearized molecule should be proportional to the genomic distance L_i between them. It is thus convenient to represent these data for a given molecule as $k \times 1$ vectors $\mathbf{X}$ and $\mathbf{L}$. As both of these quantities are known after alignment to the reference genome, we used linear regression to obtain the best fit, $X_{\text{fit}}(L) = aL + b$, for each molecule that aligns to the genome. We then computed the 2-norm of the residual vector

$$\Delta\mathbf{X} = \mathbf{X} - \mathbf{X}_{\text{fit}}\{\mathbf{L}_i\} \quad (1)$$

for each molecule, where $\mathbf{X}_{\text{fit}}\{\mathbf{L}_i\}$ is the vector of points obtained by evaluating X_{fit} at each point L_i in $\mathbf{L}$. Naturally, thermal fluctuations will give rise to small deviations from linearity, so we would expect that any given pair of labels along a “linear” molecule will exhibit the small (but random) deviations about the averaged fractional extension of the chain as a whole seen in Fig. 2(a). However, we also identified some molecules, such as the one shown in Fig. 2(b), that are typically linear but have occasional dips. These dips are not consistent with the formation of a knot or bunching of the chain; if that were the case, we would expect to have two lines in the plot with the same slope but a negative offset in X at the position L corresponding to the location of the bunch. We propose that molecules such as the one in Fig. 2(b) are incorrect alignments to the reference genome that were not identified by our initial filters.

To systematically identify molecules such as the one in Fig. 2(b) from our large data set of single molecules, we use the 2-norm of the residual vector, $\|\Delta\mathbf{X}\|_2$. Figure 3 plots the distribution of $\log_{10}\|\Delta\mathbf{X}\|_2$, overlaid with a normal

FIG. 2. Examples of (a) a linearized molecule and (b) a molecule with some barcode labels that are not uniformly stretched. Case (b) typically occurs in around 0.1% of the measurements.

distribution with an average value of 2.5 and standard deviation of 0.4 for comparison. Along with visual inspection, several statistical tests for normality at the 5% significance level (Anderson-Darling, $p = 0.78$, one-sample Kolmogorov-Smirnov, $p = 0.92$, and Pearson χ -squared, $p = 0.60$) indicate that the distribution of $\|\Delta\mathbf{X}\|_2$ is highly log-normal. We removed all molecules with z -scores higher than 3 from subsequent analysis. Note that we are only concerned about positive z -scores; negative z -scores indicate very small deviations from the linear fit. As we can see from the location of the dashed line for $z = 3$ in Fig. 3, this filter typically removed a few molecules out of several thousands for a given channel size. These molecules produced obvious outliers in the probability distributions.

C. Constructing probability distributions

Our experimental approach provides thousands of molecules per experiment, with each molecule containing numerous labels. Table I summarizes the measurement statistics for the five different nanochannel sizes used in our experiments. To construct probability densities for $P(X; L)$, we certainly need to bin in terms of the physical distance between barcode labels, X , since this is a continuous variable. Each nick site often contains several fluorescently labeled dUTP incorporation events, which emit roughly 1000-3000 photons during the 150 ms imaging time. This results in a localization of the point-spread function from the emit-

FIG. 3. Histogram of 2-norm of residuals from linear regression of X_i versus L_i on each molecule (open black bars) with standard normal probability distribution overlaid (red line). The dashed vertical line indicates the cutoff for $z = 3$.

TABLE I. Summary of the single molecule experimental statistics. The columns correspond to (i) the number of molecules that aligned to the genome and passed the uniform extension filter; (ii) the total number of labels used to construct the probability densities; (iii) the number of barcode label pairs used to construct the probability densities; (iv) the number of distinct genomic distances corresponding to these label pairs; and (v) the number of 250 bp bins containing at least 500 measurements per bin. Note that the number of non-zero bins is much larger than those with 500 measurements; for instance, the 51 nm channel has 1279 non-zero bins. The 500 measurement cutoff is used for fitting the distributions to avoid propagating sampling errors into the fitting for low-occupancy bins.

Width (nm)	Molecules	Labels	Label pairs	Genomic distances	Bins
40	6304	94 267	326 276	7 194	393
42	3109	51 773	154 067	5 211	206
43	4557	70 305	253 081	6 758	292
49	1839	29 371	87 010	4 623	141
51	9598	195 404	807 345	11 222	313

ting fluorophores of roughly 4 nm–9 nm depending on the brightness of the individual sites.^{27–29} While the imaging time is similar to the decorrelation time for λ -DNA in 100 nm channels,³⁰ the successful application of localization methods for measuring positions with sub-diffraction limit spatial resolution^{27–29} indicates that motion blur is not a substantial problem. Nevertheless, we expect that there should be some error in the localization due to motion of the fluorophores during imaging. We thus chose a bin size of 25 nm, which is a conservative estimate for the uncertainty in the resolution of the label location by microscopy. The genomic distance between barcode labels L is a discrete variable, corresponding to the number of base pairs between a given pair of barcode labels. While it is possible, in principle, to construct probability distributions without binning in L , we found that such binning was necessary to provide enough data points per bin. We thus used a bin size of 250 bp, which is roughly the size of a persistence length.

We chose not to analyze any pairs of labels on a barcode separated by genomic distances shorter than 2500 bp, since the rate of false negatives arising from proximity to the resolution limit increases substantially below this point. Indeed, due to pixelation of the image and the diffraction limit of visible light, the resolution limit of two spots is around 2000 to 3000 bp. In later tasks, where we fit the distribution with a normal-inverse Gaussian distribution, we only use those bins containing at least 500 measurements. Table I reports the number of such bins for each channel size.

III. RESULTS

We will focus in detail on the data obtained in the 43 nm channel, as the data obtained in other channel sizes are qualitatively similar and provided as supplementary material.³¹ The key quantity of interest is the probability density $P(X, L)$ that a pair of barcode labels separated by a genomic distance L (using 250 bp bin sizes) will be separated by a distance X (using 25 nm bin sizes) down the channel. To make the data easier to visualize, we chose to subtract

FIG. 4. One-dimensional probability density functions for each bin of 250 bp in genomic distance for the 43 nm channel. Color scale is logarithmic. White pixels correspond to bins with probability density less than 10^{-5} . Similar plots for the other channel sizes are included as supplementary material.³¹

the average distance, $\langle X \rangle$, for plotting purposes. As a result, Fig. 4 corresponds to $P(X - \langle X \rangle, L)$; the corresponding plot of $P(X, L)$ is linearly skewed since $\langle X \rangle \sim L$.³² A horizontal slice of Fig. 4, $P(X - \langle X \rangle; L)$, is the probability density function for the physical distance between labels using all molecules containing a pair of labels separated by a genomic distance L falling into that bin, normalized to unity using the number of observations in that bin. Note that the shorter distances occur with a much higher frequency than their larger counterparts, since a given molecule contains many labels (see Fig. 1). As such, our measurements of the probability densities at shorter genomic lengths are more reliable than those for longer genomic lengths.

The probability densities in Fig. 4 are not Gaussian distributed. Rather, it is clear from inspection that the probability densities are skew-left. While it is not readily apparent from Fig. 4, the tails of the probability densities also tend to be heavier than those predicted by a Gaussian distribution. For example, Fig. 5 shows three slices from Fig. 4, all of which

FIG. 5. Normal-inverse Gaussian probability density functions [Eq. (2)] fitted to binned experimental data at three genomic lengths for the 43 nm channel. The noise in the experimental data increases with genomic distance due to reduced sample size and broadening of the probability density.

exhibit both skew-left behavior and semi-heavy tails. Since the probability distribution is skewed, the mean value of X is less than the mode of X (i.e., the peak in the distribution is shifted to the right of the origin of the abscissa).

Probability distributions that have some qualities of a Gaussian distribution but also exhibit skew (and excess kurtosis) may be modeled using a normal-inverse Gaussian distribution. For some random variable u , the normal-inverse Gaussian distribution has the form^{33,34}

$$P(u; \alpha, \beta, \mu, \delta) = \frac{\alpha \delta K_1(\alpha \epsilon)}{\pi \epsilon} e^{\delta \gamma + \beta(u - \mu)}, \quad (2)$$

where $\gamma^2 = \alpha^2 - \beta^2$, $\epsilon^2 = \delta^2 + (u - \mu)^2$, and $K_1(z)$ is the modified Bessel function of the third order and index 1. This distribution is described by four parameters: (i) the location parameter, μ ; (ii) the tail heaviness parameter, α ; (iii) the asymmetry parameter, β ; and (iv) the scale parameter, δ . In what follows, it is important to keep in mind that the parameters $(\alpha, \beta, \delta, \mu)$ of Eq. (2) are themselves functions of the genomic distance and the channel size. The normal-inverse Gaussian's moment generating function³⁴

$$M(u; \alpha, \beta, \mu, \delta) = \exp \left[\delta \left\{ \sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + u)^2} \right\} + \mu u \right] \quad (3)$$

makes calculation of the expectations $E[u^n]$ straightforward by repeated differentiation of Eq. (3),

$$E[u^n] = \left. \frac{d^n M}{du^n} \right|_{u=0}. \quad (4)$$

There are a number of possible ways to fit Eq. (2) to the experimental data. We found that the most robust approach was to work with the centered probability distribution for each bin in L , $P(X - \langle X \rangle; L)$. This corresponds to $u = X - \langle X \rangle$, where $\langle X \rangle$ was computed for a particular value of L directly from the distribution data. Since by definition, the expectation $E[X - \langle X \rangle] = 0$, this information constrains μ from the mean of the normal-inverse Gaussian distribution for a given genomic distance,

$$E[X - \langle X \rangle] = \mu + \frac{\delta \beta}{\gamma} = 0. \quad (5)$$

A multivariable least-squares fitting routine then produced the remaining parameters (α, β, δ) for a given genomic separation L . In the fitting procedure, we only work with those bins that contain at least 500 measurements. Three representative probability density functions and their best-fit probability densities for different genomic lengths are shown in Figure 5. Note that $\beta < 0$ since the distribution is skew-left, while the remaining parameters are positive.

In order for the normal-inverse Gaussian to be a realistic model for the separation between barcode labels, it must produce the correct molecular weight scaling for the variance. We thus require that the variance from the normal-inverse Gaussian distribution satisfies

$$E[(X - \langle X \rangle)^2] \equiv \delta^2 X = \frac{\delta \alpha^2}{\gamma^3} \sim L. \quad (6)$$

We have obtained the values of α , β , and δ (and thus γ) for each value of L from a separate multivariate fit, without

FIG. 6. Test of the molecular weight scalings in Eq. (6) for the 43 nm channel. The red line corresponds to linear scaling with a prefactor of 1. Similar plots for the other channel sizes are included as supplementary material.³¹

enforcing these relationships *a priori*. As a result, satisfying Eq. (6) serves as a *a posteriori* verification that the fitting procedure produces physically reasonable results as a function of molecular weight.

Figure 6 tests the molecular weight scaling from the fits for the 43 nm channel. The variance exhibits linear behavior over the central portion of the plot, consistent with the polymer models for confinement.^{16,32,35} There is some deviation at the smaller molecular weights, which we can attribute to the difficulty in accurately measuring small distances between labels and the systematic error introduced by 250 bp bins for small molecular weights, and scatter at higher molecular weights, which arises from the sampling error (even after binning the data). It is possible that the deviation from linearity at smaller distances is physically relevant, since simulations clearly indicate that the asymptotic scaling for the variance in extension is not achieved until the chain is long enough, of the order of 185 kbp, and that the variance increases sub-linearly with molecular weight for small molecular weights.²² This is under the assumption that the variance of extension of segments of a long chain behaves similarly to the variance observed in single molecules of the same molecular weight. However, the uncertainty in the experimental data does not allow us to draw any conclusions about subtle deviations from linearity in Fig. 6. We obtained results similar to Fig. 6 for the other channel sizes, which are included as supplementary material.³¹

In our analysis, we are using the normal-inverse Gaussian distribution as a way to reduce the enormous amount of extension data produced by our experiments to a minimal number of parameters that still captures the essential physics. In doing so, we do not propose that there is necessarily a fundamental reason for the chain extension to have a normal-inverse Gaussian distribution. Rather, the normal-inverse Gaussian distribution is a natural progression from the simple Gaussian model to one with more degrees of freedom. Indeed, the normal-inverse Gaussian reduces to a Gaussian distribution in the case $\beta = 0$, $\alpha \rightarrow \infty$ and $\delta/\alpha = \sigma^2$, where σ is the standard deviation of the Gaussian distribution.³⁴

Our analysis produced values of $(\alpha, \beta, \delta, \mu)$ for all of the bins listed in Table I, leading to 1345 sets of parameters for given combinations of L and D . The values of these parameters for all channels are included as supplementary material,³¹ and hopefully provide a compact, useful resource for comparison between theory and experiment.

While the normal-inverse Gaussian is a convenient description of the data (and superior to the other distributions that we tested for this purpose), one might question the basic premise of using the normal-inverse Gaussian distribution in the first place; it seems logical that the distances between labels should be described by a probability density with compact support,³⁶ since their distances are bounded from below by zero extension and above by the contour length of the DNA between labels. In response, we note that although the normal-inverse Gaussian distribution has semi-heavy tails, the decay of the probability density is sufficiently fast to make it a physically realistic model in practice. The key is the skew of the probability density, which puts much of the weight below the most probable value of X and thus minimizes the probability of overextension. Indeed, the significantly skew-left distributions that we obtained implied a probability of chain over-extension $\Pr(X > L) \approx 10^{-12}$, for long chains, and under-extension $\Pr(X < 0) \approx 10^{-8}$, for short chains. These probabilities are sufficiently low to be considered negligible for any practical purposes. The only issues with the fitting protocol arise for very short distances between labels, where the bin size becomes a reasonable fraction of the molecular weight. In this case, the molecules with the largest value of L in the bin may appear overextended from the fitting protocol, which averages over all molecules within a 250 bp range, even though inspection of the data indicates that the longest chains inside these bins for short barcode label distances are indeed extended below their contour length.

IV. DISCUSSION

A. Physical basis for a skewed probability density

The key result of our analysis is that the distance between barcode labels on DNA confined inside a nanochannel is not Gaussian distributed. In particular, we found that it is skewed towards extensions that lie below the most probable extension. The physical basis for the skew in the probability density is explained by Fig. 7. For a chain in the Odijk regime,³² corresponding to $D \ll l_p$, the most probable configuration is similar to Fig. 7(a); the DNA is highly extended but has occasional fluctuations due to thermal energy. In such an extended state, thermal fluctuations to a higher extension must be small since the chain cannot extend beyond its contour length. If the chain is confined to a channel where $D \approx l_p$, the extension can adopt a value substantially smaller than $\langle X \rangle$ either by fluctuations in the deflection segment length because D is not small relative to l_p or the backfolding mechanism illustrated in Figs. 7(b) and 7(c). In the most dramatic case, shown in Fig. 7(b), the chain forms a large-scale global hairpin fold consisting of tens of thousands of base pairs. Such an event would lead to a very strong skew in the distribution. However, such molecules cannot be the source of the skew in our

FIG. 7. Physical mechanisms for introducing a skew-left fluctuation in the distance between barcode labels relative to (a) a chain with a typical configuration corresponding to extensions close to the average extension, $\langle X \rangle$. (b) Global hairpin fold. (c) Bunching of the chain through nearby local folds (S-loops) or knotting.

distribution—they are likely excluded from our measurement because our approach requires that the barcodes align to the reference genome so that we can identify the DNA molecule. As we see in Fig. 7(b), a global hairpin tends to disorder the location of the barcode labels in the axial direction, and thus leads to poor alignment with the reference genome. This does not preclude the existence of large global hairpins; rather, they probably are not present in our ensemble because, if they exist, they cannot be aligned to the genome. At smaller scales, it is possible to reduce the extension by bunching (Fig. 7(c)) interspersed with regions of ordered deflection segments. These bunches could be analogous to the S-loop model for DNA in nanochannels near the persistence length,²⁰ but they could be knots as well.^{37,38} We do not have the optical resolution to examine the conformation of the chain at the scale required to demonstrate the presence of an S-loop or a knot, so we will use the more generic term “bunching” to refer to the stored length. This is a nucleation phenomenon, with a certain probability of formation based on the energetic cost to form and then confine the bunched region of the chain.

A natural question to ask is whether the data support formation of bunches via a nucleation phenomenon. If this were the case, then we would expect the skewness of the distribution,

$$\text{skewness} = E \left[\left(\frac{X - \langle X \rangle}{\sqrt{\delta^2 X}} \right)^3 \right], \quad (7)$$

to differ sharply depending on the distance between labels. For short distances between labels, there would be no bunches because the nucleation probability is very low and we have a finite amount of measurements. As the distance between labels increases, these bunches would eventually nucleate for sufficiently long chains. Thus, we would expect no skew for the shorter distances and a sensible skew for the longer distances. Figure 8 presents the skewness of the probability distribution computed from the normal-inverse Gaussian fits,

$$\text{skewness} = \frac{3\beta}{\alpha\sqrt{\delta\gamma}}, \quad (8)$$

FIG. 8. Skew of the distribution as a function of the distance between barcode labels for the 43 nm channel. Similar plots for the other channel sizes are included as supplementary material.³¹

as a function of the distance between barcode labels for the 43 nm channel. The trend is similar for the other channels and included as supplementary material.³¹ We posit that the skewness is essentially independent of barcode label distance because (i) the distances have a much larger range than the skew and (ii) the uncertainty in the measurement of the third centered-moment is rather large, so the statistical significance of any trend in the data is very low.

It appears that the mechanism in Fig. 7 is biasing the distribution towards shorter extensions is not a rare event. This suggests that the skew in the overall extension arises primarily from the underlying distribution for the size of a deflection segment for $D \approx l_p$, rather than from the nucleation of a large concentration fluctuation due to an S-loop or a knot. In the Odijk theory³² for very small channels $D \ll l_p$, the deflection segments are strongly stretched pieces of DNA containing a contour length $\lambda = D^{2/3}l_p^{1/3}$ aligned at an angle δ with respect to the channel axis. As the confinement weakens towards $D \approx l_p$, bending within a deflection segment should affect fluctuations in the mean span of the molecule; the energetic penalty for bending has now decreased to the extent that small deviations along the chain backbone are quite likely even if global hairpins do not form frequently. The distribution for the size of a deflection segment would be skew-left, since (i) the maximum fluctuation in length is bounded on the right by the maximum contour length of the DNA comprising the deflection segment and (ii) there is more entropy gained by a fluctuation to a shorter mean span than a longer mean span. In principle, we should be able to identify the large fluctuations in the DNA concentration characteristic of an S-loop or a knot from the small fluctuations in DNA concentration characterizing fluctuations in the extension of a deflection segment using the intensity distribution of the YOYO-stained backbone. This is the subject of ongoing work.

This qualitative model for a skew in the fluctuations towards smaller extensions by relatively frequent, small bending events is supported by Odijk’s theory for confinement in channels smaller than the persistence length.¹⁶ As noted in the previous paragraph, in the classic Odijk regime,³² the

DNA is organized into deflection segments of typical length $\lambda = D^{2/3}l_p^{1/3}$ that are linearly ordered throughout the channel. The probability of backfolding of the deflection segments is governed by the z parameter for a hairpin fold, which Odijk defined as¹⁶

$$\xi = \frac{gw}{D_{\text{eff}}^{5/3}l_p^{1/3}}, \quad (9)$$

where g is the global persistence length of the chain,³⁹ w is the effective width of the backbone (including electrostatic interactions), and $D_{\text{eff}} = D - w$ is the cross-sectional width available to the chain. Excluded volume suppresses backfolding when $\xi > 1$. From numerical simulations, large scale backfolding of the type illustrated in Fig. 7(b) tends to be substantial only in the case $\xi < 0.2$,²² consistent with the approximate nature of the scaling arguments leading to Eq. (9).

To estimate ξ for our experiments, we use $D_{\text{eff}} = 40$ nm as a representative channel size. The persistence length $l_p = 58$ nm and effective width $w = 10$ nm were estimated from Dobrynin's theory⁴⁰ and Stigter's theory,⁴¹ respectively, in TBE 0.5 $\times$ buffer ($I \approx 2.4 \times 10^{-2}$ M). For the global persistence length, we take advantage of a recent calculation²² indicating that $g = 190$ nm for this value of D_{eff}/l_p . As a result, the experiments described here correspond to $\xi = 47.7$. Repeating the calculation leads to $\xi = 47.7, 30.6, 24.9, 8.7$, and 6.5 for the five different channels (in increasing order). We thus expect that the formation of a global hairpin of the type in Fig. 7(b) to be an extremely rare event. Indeed, this is the reason why genome mapping in nanochannels works, since global hairpins destroy the linear ordering of the genomic information in the nanochannel. Nevertheless, the absence of global hairpins does not preclude the possibility of small scale bunching of the chain configuration beyond what would be expected from the conformation of a deflection segment.³² Moreover, while excluded volume does not impact the formation of hairpins for very stiff chains where $\xi \ll 1$,^{16,22,39} excluded volume does play a role for $\xi \approx 1$.²² At the marginal value of ξ in our experiments, such bunching is possible.

While the bunching mechanism or bending within a deflection segment seems to be plausible explanations for the skew in the probability distribution, they do not explain the occasional micron-sized deviations from the mean extension in Fig. 4. These large deviations are indeed very rare events; the intensity scale bar in Fig. 4 is logarithmic and most of these outliers correspond to a single observation. Moreover, although there appear to be a number of outliers at different genomic distances, these typically correspond to many pairs of labels on the same molecule. Most likely, the outliers arise from some combination of incomplete barcode labeling and substantial deformation of the chain that nonetheless still leads to reasonable alignment to the reference genome. From a genomics standpoint, these molecules are irrelevant since their contribution to the final genome map is rendered negligible by obtaining sufficient coverage of the genome, which is easily accomplished in practice.^{1,2} From a physics standpoint, we feel that our experimental approach is a reasonable compromise between including the full range of possible physical rearrangements (knots, folds, loops) while ensuring that we can identify the sequence of the DNA. We certainly

could have removed these outliers by setting a higher threshold for genome alignment or linearity of $\langle X \rangle$ versus L . However, if we set these thresholds too high, then we will begin to exclude small bending that is certainly physically reasonable. The need to balance these two objectives is an intrinsic shortcoming of our genomic approach. In our opinion, this shortcoming is more than offset by the tremendous throughput of the methodology.

Our discussion has focused exclusively on fluctuations in the chain configuration, either due to formation of S-loops/knots or fluctuations in the lengths of a deflection segment. Such configurational fluctuations arise solely due to the impenetrability of the channel wall, which we treat as a infinite repulsive potential energy barrier. We should bear in mind that there may be softer (and possibly attractive potentials) that could alter the DNA configuration as well. For example, incomplete ligation of the DNA following the insertion of the fluorescent dUTP labels could lead to hydrophobic interactions with the channel walls. It is also possible that electrostatic interactions in the double layer surrounding the channel walls could affect the configuration of the chain.

B. Comparison with Odijk's theories

Inasmuch as the ξ value for our channels lies at the boundary between the two Odijk regimes, one consisting of linearly ordered deflection segments³² (classic Odijk regime) and the other where these deflection segments can backfold¹⁶ (backfolded Odijk regime), we would expect the average extension and variance to lie between these two limiting cases. To make this comparison, we used the values of $l_p = 58$ nm and $w = 10$ nm cited above. For the contour length, we used a rise of 0.34 nm/bp, which does not account for any extension due to intercalation. The physical channel width, D , was obtained from SEM measurements of two nanochannels of thousands in the array. The lack of statistical power in such a measurement, along with inherent variation from one channel to the next in the fabrication process, means that the channel sizes in Table I and Fig. 9 should be treated as approximations. Thus, we do not attribute any significance to the trends with respect to channel size over such a narrow and uncertain range in D , and focus only on the magnitude of the predictions from the theories.

In the classic Odijk regime, the extension is given by³²

$$\langle X/L \rangle = 1 - 0.18274(D_{\text{eff}}/l_p)^{2/3}, \quad (10)$$

with the prefactor corresponding to a square channel.³⁵ Figure 9(a) shows that, at first glance, Eq. (10) appears to be a good fit to the experimental data. However, corrections to L for intercalation,^{30,42} even at rather low dye:bp ratios, would tend to reduce the fractional extension by increasing L , so we would expect that the classic Odijk regime is in fact an overestimate of the experimental data. This result is consistent with expectations at $\xi \approx 1$, since the classic Odijk regime is only valid for $\xi \gg 1$. It is also consistent with recent experimental data for DNA confined by convex lens-induced nanoscale templating.⁴³ The predicted extension

FIG. 9. Comparison of experimental data (blue squares), the harmonic approximation (green circles, see Sec. IV C), and Odijk's theories. (a) Fractional extension, where the solid line is Eq. (10). (b) Variance, where the solid line is Eq. (12) and the dashed line is Eq. (13). The data points for the experimental data and the harmonic approximation are the average value obtained over all bins; the standard error is smaller than the data points.

in the backfolded Odijk regime is^{16,22}

$$\langle X/L \rangle = 0.92\xi^{1/3}, \quad (11)$$

where the prefactor was obtained by numerical simulations in square channels.²² This prediction gives unphysical extension greater than unity. This result is also expected, since Eq. (11) is only valid for $\xi < 0.2$.²²

For the variance, Fig. 9(b) shows that the predictions for the classic Odijk regime,³⁵

$$\delta^2 X = 0.00956 \frac{D_{\text{eff}}^2}{l_p} L, \quad (12)$$

and the backfolded Odijk regime,^{16,22}

$$\delta^2 X = 0.25 L g, \quad (13)$$

bracket the experimental data. Again, the prefactors in these equations were obtained by simulations,^{22,35} and the global persistence length³⁹ was computed by a formula that accounts for thermal fluctuations.²² The prediction in Eq. (12) is valid for chains consisting of linearly ordered deflection segments that exhibit thermal fluctuations about the average length of a deflection segment. In contrast, the prediction in Eq. (13) is valid for chains where the deflection segments can backfold in the channel many times and form global hairpins. In as much as the experimental data for the skewed probability distributions suggest that the chain exhibits small bunches nucleated along the chain or bending within deflection segments, it is reasonable that the variance of these distributions would lie somewhere between the two limiting cases.

A previous experimental study¹⁹ of barcode label distances in somewhat larger channels also indicated that

the variance about the mean extension is larger than the prediction for the Odijk regime in Eq. (12). However, the data in those experiments further suggested that the variance increases very weakly with molecular weight, albeit over a much narrow range of molecular weights. This previous result is at odds with our results, where two pieces of evidence indicate a linear dependence of $\delta^2 X$ on L : (i) the negligible standard error in Fig. 9 and (ii) the data in Fig. 6 and similar data in the supplementary material.³¹ The most likely discrepancy between our measurements and prior work¹⁹ is the advancement in the barcoding technology; we were able to confirm the molecular weight scaling for the variance only by studying a very wide range of molecular weights with the large data set afforded by our high throughput method.

C. Evaluation of a harmonic approximation

While the probability distribution for the extension between barcode labels is clearly skewed, it is worth assessing how the skew impacts the accuracy of a harmonic approximation for the extension of the confined chain. If we represent the confined chain by a bead-spring dumbbell model,^{44,45} the probability density for the end-to-end distance is governed by a spring fluctuating at equilibrium under thermal energy,

$$p(X) = Z^{-1} \exp \left[-\frac{F(X)}{k_B T} \right], \quad (14)$$

where $F(X)$ is the free energy for an extension X and Z is a normalization constant. For small fluctuations about the equilibrium extension of the chain, it is common to model the confinement free energy by a harmonic potential

$$F(X) = \frac{1}{2} k_s (X - \langle X \rangle)^2. \quad (15)$$

In the latter, k_s is the spring stiffness, reflecting changes in the configurational entropy due to extension/compression of the spring and $\langle X \rangle$ is the equilibrium extension of the spring. There are two important properties of the harmonic potential to keep in mind. First, the average extension corresponds to the minimum of the free energy, $\langle X \rangle = X_{\text{min}}$, and fluctuations about that average extension are governed by

$$\delta^2 X = \frac{k_B T}{k_s}. \quad (16)$$

Second, if the spring potential is harmonic, then the probability distribution for the extension is Gaussian. Note that even if the function $F(X)$ is not harmonic, these assumptions about the mean extension and its variance are often invoked, for example, in Flory theory.

To test the accuracy of a harmonic spring model to predict the first two moments of the probability distribution governing the extension, we inverted Eq. (14) to obtain the free energy corresponding to the probability densities from our experiments in the 43 nm channel. Following the harmonic approximation, we assumed that X_{min} is the mode of the distribution (and thus the extension corresponding to the minimum value of the free energy, F_{min}) and then computed a

FIG. 10. Comparison of (a) the average extension and (b) the variance about the average extension for a harmonic approximation about the minimum free energy (black squares) and the experimental data (blue circles). These data correspond to the 43 nm channel. Note that (b) is a log-log plot, so the deviation between the model and experimental data is larger than in the linear plot of (a). Similar plots for the other channel sizes are included as supplementary material.³¹

harmonic approximation to this free energy about $X_{\min}$,

$$F(X) \approx F_{\min} + \frac{1}{2} k_s (X - X_{\min})^2 + \dots, \quad (17)$$

using the normal-inverse Gaussian fits to smooth the experimental data. As we can see in Fig. 10(a), the harmonic spring model is a reasonable approximation for the average extension, with an error of around 0.5%. Even though the probability densities are skew-left, and thus $X_{\min} \neq \langle X \rangle$, these values are relatively close. As we move to the second moment, we see in Fig. 10(b) that the deviation between the harmonic model and the experimental data increases substantially to 64%. This highlights the importance of the skewness of the distribution; making a harmonic approximation for the spring force in a dumbbell model overestimates the stiffness of the spring, leading to a decrease in the variance. Obviously, the harmonic approximation, which is symmetric, cannot predict a skewed distribution. This is a potential problem for Flory-theory approaches to describing confinement.

We repeated the harmonic approximation analysis for all of the channel sizes and the results are included as supplementary material.³¹ Figure 9 summarizes the key conclusion of our analysis of the mean extension and its variance. The conclusions we drew from Fig. 10 for the 43 nm channel are robust to channel size. The fractional extension produced by the harmonic model is slightly different than the experimental data, but the deviation between the model and experiments increases when we consider the variance. The difference between the model and experiments is smallest for the longest molecular weights, but these barcode label distances correspond to the range where we have the least data.

V. CONCLUSIONS

In the present contribution, we have used a genome mapping system to measure the probability density $P(X; L)$ for the distance X between two barcode labels on a DNA molecule separated by some distance L in channels ranging from 40 nm to 51 nm in width. Our analysis relied on a genomic method, where we aligned the barcoded molecules to the reference genome to obtain their molecular weights and the locations of the barcode labels along the chain.

The benefit of this approach is obvious from the volume of data obtained in these experiments. Typical single molecule experiments, for example, recent measurements in tapered nanochannels³⁰ or confinement by convex lens-induced nanoscale templating,⁴³ obtain statistics from tens of molecules. In contrast, we were able to obtain data from tens of thousands of molecules. Indeed, it is likely that the single molecule data presented here exceeds all of the single molecule data presented in all previous experimental studies of DNA confinement in nanochannels. There is simply no way to construct probability densities without obtaining a very large quantity of data. However, our method for reaching a high data density does come with a potential source of systematic error. Explicitly, the genome alignment is a statistical algorithm²⁶ and thus provides only a confidence that a given molecule was correctly aligned. In our analysis, we set a relatively lenient threshold for accepting the alignment to the genome to ensure that we did not inadvertently exclude molecules that have allowed thermal fluctuations in their extension. We then implemented a linearity filter that identified molecules with dips in their values of X vs. L , which likely correspond to misalignments to the reference genome. The latter step also requires a threshold, since the residuals from the linear regression in Fig. 3 are log-normal distributed. Clearly, changing the threshold for this step will change the quantitative results. However, we expect the qualitative results to be unchanged; decreasing the threshold would increase the number of outliers in the distribution, similar to the ones we saw in Fig. 4, and thus further increase the skew. We should also bear in mind that the DNA comes from a biological sample, so there is always the possibility that the genomes of the bacteria are not identical to the reference. In this case, even a perfect alignment algorithm would still lead to some error in the measurements of L . We are presently working on a method of measuring the distribution for X without requiring alignment to the genome, thereby alleviating the key shortcoming of the method used here.

Even with these potential sources of systematic error, we posit that the skew-left distribution for $P(X; L)$ is a robust result, as it is consistent with phenomenological models for channel confinement near the persistence length²⁰ as well as the physics underlying Odijk's theory.^{16,22} However, this skew-left behavior poses challenges for any theory attempting to describe DNA confinement when backfolding is permitted but large scale folding is a rare event (i.e., $\xi \approx 1$). In particular, it seems that a theory based on a harmonic approximation to the free energy about its minimum will incur substantial error for higher-order moments such as the variance about the mean extension. We anticipate that future modeling efforts

will benefit greatly from considering the full probability distribution for the extension, rather than simply focusing on the two lowest moments of the distribution. We anticipate that refinements in the experimental approach used here, when combined with large-scale simulations, will point the way forward to a predictive theory of DNA confinement near the persistence length.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (CBET-1262286) and the National Institutes of Health (R01-HG006851). This work was carried out in part using computing resources at the University of Minnesota Supercomputing Institute. J.G.R. and H.C. are employees of BioNano Genomics, which is commercializing nanochannel genome mapping.

- ¹E. T. Lam, A. Hastie, C. Lin, D. Erlich, S. K. Das, M. D. Austin, P. Deshpande, H. Cao, N. Nagarajan, M. Xiao, and P.-Y. Kwok, *Nat. Biotechnol.* **30**, 771 (2012).
- ²A. 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).
- ³M. Levy-Sakin and Y. Eberstein, *Curr. Opin. Biotechnol.* **24**, 690 (2013).
- ⁴H. Cao, J. O. Tegenfeldt, R. H. Austin, and S. Y. Chou, *Appl. Phys. Lett.* **81**, 3058 (2002).
- ⁵J. O. Tegenfeldt, C. Prinz, H. Cao, S. Chou, 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).
- ⁶K. Jo, D. M. D'ingra, 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).
- ⁷S. K. Das, M. D. Austin, M. C. Akana, P. Deshpande, H. Cao, and M. Xiao, *Nucleic Acids Res.* **38**, e177 (2010).
- ⁸Y. Eberstein, N. Gassman, S. Kim, J. Antelman, Y. Kim, S. Ho, R. Samuel, X. Michalet, and S. Weiss, *Nano Lett.* **9**, 1598 (2009).
- ⁹M. Xiao, A. Phong, C. Ha, T.-F. Chan, D. Cai, L. Leung, E. Wan, A. L. Kistler, J. L. DeRisi, P. R. Selvin, and P.-Y. Kwok, *Nucleic Acids Res.* **35**, e16 (2007).
- ¹⁰R. K. Neely, P. Dedeker, J. Hotta, G. Urbanaviciute, S. Klimasauskas, and J. Hofkens, *Chem. Sci.* **1**, 453 (2010).
- ¹¹M. Baday, A. Cravens, A. Hastie, H. Kim, D. E. Kudeki, P.-Y. Kwok, M. Xiao, and P. R. Selvin, *Nano Lett.* **12**, 3861 (2012).
- ¹²W. Reisner, N. B. Larsen, A. Silahtaroglu, A. Krisentsen, N. Tommerup, J. O. Tegenfeldt, and H. Flyvbjerg, *Proc. Natl. Acad. Sci. U. S. A.* **107**, 13294 (2010).
- ¹³R. L. Welch, R. Sladek, K. Dewar, and W. Reisner, *Lab Chip* **12**, 3314 (2012).
- ¹⁴A. N. Nilsson, G. Emilsson, L. K. Nyberg, C. Noble, L. Svensson Stadler, J. Fritzsche, E. R. B. Moore, J. O. Tegenfeldt, T. Ambjornsson, and F. Westerlund, *Nucleic Acids Res.* **42**, e118 (2014).
- ¹⁵H. Cao, A. R. Hastie, D. Cao, E. T. Lam, Y. Sun, H. Huang, X. Liu, L. Lin, W. Andrews, S. Chan, S. Huang, X. Tong, M. Requa, T. Anatharaman, A. Krogh, H. Yang, H. Cao, and X. Xu, *GigaScience* **3**, 34 (2014), available online at <http://www.gigasciencejournal.com/content/3/1/34>.
- ¹⁶T. Odijk, *Phys. Rev. E* **77**, 060901 (2008).
- ¹⁷P. Cifra, Z. Benkova, and T. Bleha, *J. Phys. Chem. B* **113**, 1843 (2009).
- ¹⁸Y. Wang, D. R. Tree, and K. D. Dorfman, *Macromolecules* **44**, 6594 (2011).
- ¹⁹T. Su, S. K. Das, M. Xiao, and P. K. Purohit, *PLoS ONE* **6**, e16890 (2011).
- ²⁰L. Dai, S. Y. Ng, P. S. Doyle, and J. van der Maarel, *ACS Macro Lett.* **1**, 1046 (2012).
- ²¹D. R. Tree, Y. Wang, and K. D. Dorfman, *Phys. Rev. Lett.* **110**, 208103 (2013).
- ²²A. Muralidhar, D. R. Tree, and K. D. Dorfman, *Macromolecules* **47**, 8446 (2014).
- ²³M. O'Brien, V. B. Searles, C. M. Dickens, D. Astling, D. Albracht, A. C. Y. Mak, Y. Y. Y. Lai, C. Lin, C. Chu, T. Graves, P.-Y. Kwok, R. K. Wilson, and J. M. Sikel, *BMC Genomics* **15**, 387 (2014).
- ²⁴A. Muralidhar, D. R. Tree, Y. Wang, and K. D. Dorfman, *J. Chem. Phys.* **140**, 084905 (2014).
- ²⁵A. Edelstein, N. Amodaj, K. Hoover, R. Vale, and N. Stuurman, *Curr. Protocols Mol. Biol.* **92**, 14.21.1 (2010).
- ²⁶A. Valouev, "Shotgun Optical Mapping: A Comprehensive Statistical and Computational Analysis," Ph.D. thesis, University of Southern California, 2006.
- ²⁷N. Bobroff, *Rev. Sci. Instrum.* **57**, 1152 (1986).
- ²⁸R. E. Thompson, D. R. Larson, and W. W. Webb, *Biophys. J.* **82**, 2775 (2002).
- ²⁹A. Yildiz, J. N. Forkey, S. A. McKinney, T. Ha, Y. E. Goldman, and P. R. Selvin, *Science* **300**, 2061 (2003).
- ³⁰D. 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).
- ³¹See supplementary material at <http://dx.doi.org/10.1063/1.4907552> for (i) probability densities, similar to Fig. 4, for all channel sizes; (ii) variance data, similar to Fig. 6, for all channel sizes; (iii) skewness plots, similar to Fig. 8, for all channel sizes; (iv) result from the harmonic approximation, similar to Fig. 10, for all channel sizes; and (v) tabulated normal-inverse Gaussian distribution fitting parameters for all channel sizes and all binned genomic lengths.
- ³²T. Odijk, *Macromolecules* **16**, 1340 (1983).
- ³³O. Barndorff-Nielsen, *Proc. R. Soc. London, Ser. A* **353**, 401 (1977).
- ³⁴O. Barndorff-Nielsen, *Scand. J. Stat.* **24**, 1 (1997), available online at <http://www.jstor.org/stable/4616433>.
- ³⁵T. W. Burkhardt, Y. Yang, and G. Gompper, *Phys. Rev. E* **82**, 041801 (2010).
- ³⁶Y. Wang, W. F. Reinhart, D. R. Tree, and K. D. Dorfman, *Biomicrofluidics* **6**, 014101 (2012).
- ³⁷C. Micheletti and E. Orlandini, *Soft Matter* **8**, 10959 (2012).
- ³⁸C. Micheletti and E. Orlandini, *ACS Macro Lett.* **3**, 876 (2014).
- ³⁹T. Odijk, *J. Chem. Phys.* **125**, 204904 (2006).
- ⁴⁰A. V. Dobrynin, *Macromolecules* **39**, 9519 (2006).
- ⁴¹D. Stigter, *Biopolymers* **16**, 1435 (1977).
- ⁴²L. K. Nyberg, F. Persson, B. Akerman, and F. Westerlund, *Nucleic Acids Res.* **41**, e184 (2013).
- ⁴³D. J. Berard, F. Michaud, S. Mahshid, M. Jalal Ahamed, C. M. J. McFaul, J. S. Leith, P. Berube, R. Sladek, W. Reisner, and S. R. Leslie, *Proc. Natl. Acad. Sci. U. S. A.* **111**, 13295 (2014).
- ⁴⁴R. B. Bird, C. F. Curtiss, R. C. Armstrong, and O. Hassager, *Dynamics of Polymeric Liquids, Volume 2, Kinetic Theory* (John Wiley & Sons, Inc., 1986), Vol. 2.
- ⁴⁵M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Oxford University Press, 1986).