Non-Markovian Transport of DNA in Microfluidic Post Arrays

Nicolas Minc, Jean-Louis Viovy, and Kevin D. Dorfman*

Laboratoire Physicochimie-Curie, CNRS/UMR 168, Institut Curie, 26 Rue d'Ulm, F-75248 Paris Cedex 5, France

(Received 29 October 2004; published 18 May 2005)

We present an analytically solvable model for the transport of long DNA through microfluidic arrays of posts. The motion is a repetitive three-part cycle: (i) collision with the post and extension of the arms; (ii) rope-over-pulley post disengagement; and (iii) a random period of uniform translation before the next collision. This cycle, inspired by geometration, is a nonseparable (Scher-Lax) continuous-time random walk on a lattice defined by the posts. Upon adopting a simple model for the transition probability density on the lattice, we analytically compute the mean DNA velocity and dispersivity in the long-time limit without any adjustable parameters. The results compare favorably with the limited amount of experimental data on separations in self-assembled arrays of magnetic beads. The Scher-Lax formalism provides a template for incorporating more sophisticated microscale models.

DOI: 10.1103/PhysRevLett.94.198105 PACS numbers: 87.15.Tt, 05.40.Fb

Microfluidic post arrays are a powerful alternative to gel electrophoresis for separating long DNA. For example, arrays of self-assembled magnetic bead columns [1,2] (the Ephesia system) or microfabricated pillars [3] can separate DNA tens of kilobase pairs (kbp) long in minutes [1,2] or even seconds [3]. At the same time, the uniformity of these arrays and the availability of fluorescence microscopy techniques makes these systems ideal for fundamental studies of DNA dynamics in obstacle courses [4]. It is no surprise, then, that many researchers have turned their attention towards post collisions of long DNA. For single posts, there exists a growing body of experimental work [5–7], complimented by theoretical analyses and numerical simulations [8–10]. There are also analogous experimental [4] and numerical [11] studies on DNA dynamics in arrays. While these studies have produced a wealth of information, it is not always clear how to utilize this information to guide device design.

To address this need, we previously proposed [12] an analytically solvable Markovian model for averaging single-post data. When the DNA moves through a row in the post array, we assumed that it collides with the post with some probability Π_c and, if it collides, it is retained for an average time τ. Otherwise, the DNA moves between post rows with its unhindered velocity $U = \mu_0 E$, with μ_0 the free-solution mobility and E the electric field strength. If Π_c is independent of the field and $\tau \sim E^{-1}$, then this model predicts [12] that the band broadening scales like E^{-1} and the resolution is independent of the field, in agreement with experiments [2]. Nevertheless, the model cannot make quantitative predictions without functional forms for τ and Π_c.

In order to determine τ and Π_c, and simultaneously test the validity of the Markovian model, Minc *et al.* undertook a single-molecule study [4] of T4 DNA (168.9 kbp) in self-assembled magnetic arrays of various post densities. They confirmed the aforementioned dependencies of τ and Π_c on E, but found that the Markovian model is invalid for high density arrays used in separations [2,3]. This implies that an analytical model for guiding real device design must account for the close proximity of the posts and the resultant memory between collisions.

In this Letter, we develop such a non-Markovian model for DNA motion in post arrays, inspired by geometration theory [13,14]. Although we focus our comparison with experimental results on the Ephesia system, the model is valid for *any* of the post arrays used in practice [1–3,5,6]. We treat the DNA dynamics as a repetitive cycle consisting of the three steps shown in Fig. 1: (i) collision with the post

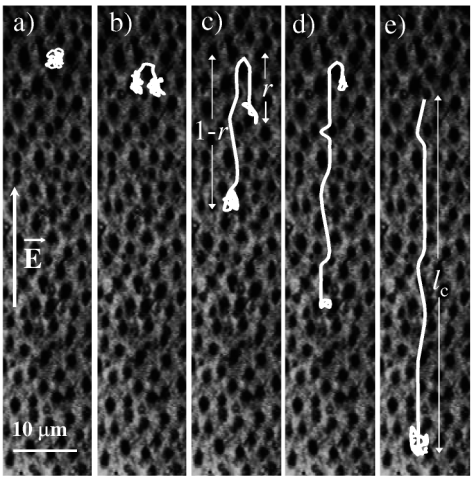

FIG. 1. Schematic of a post collision superimposed on a self-assembled array of magnetic beads [2]: (a) The DNA collides with the post and begins step (i). (b) The coil unravels and extends into two arms. (c) When the two arms are extended, the DNA begins step (ii). r is the fraction of monomers in one arm. (d) The DNA disengages by a rope-over-pulley mechanism. (e) Step (iii) begins when the DNA disengages from the post, extended close to its full contour length l_c, and ends with the next collision. The DNA cannot collide with any of the posts between its leading and trailing ends. The schematic highlights the incomplete extension observed in experiments [4,6].

and extension into two arms; (ii) a rope-over-pulley disengagement; and then (iii) a random period of translation before colliding with the next post. The model differs from geometration [14] in the third step. Geometration assumes the next collision always occurs a distance l_c downstream, where l_c is the contour length of the DNA. This is reasonable in gels, where the ''posts'' are extremely dense, but not in microfluidic arrays [2]. Randomizing the third step is nontrivial, since it couples the random duration of a cycle to the random distance traversed, making the continuous-time random walk (CTRW) nonseparable [15].

To analyze this walk, we need to derive the transition probability density, $\psi(n, t)$, for moving through n post rows in time t during a cycle. For steps (i) and (ii), we adopt the preaveraged model of Popelka *et al.* [14], which permits a closed form for $\psi(n, t)$. The preaveraging refers to step (i), where the DNA collides with the post and uncoils into two arms in the time $t_1 = l_c/U$. In the second step, the difference in the electrical force on each arm drives the disengagement from the post, requiring the time

$$t_2 = \frac{l_c}{2U} \ln\left(\frac{1}{|2r - 1|}\right), \tag{1}$$

where $r \in [0, 1]$ is the initial fraction of DNA in one of the arms. We model r as a uniformly distributed random variable [14,16], noting that the logarithmic divergence of t_2 at $r = 1/2$ is integrable [8]. Once the DNA disengages from the post, we assume it travels n rows in the time $t_3 = na/U$ before it collides again. The total time for a cycle of distance na is then $t_1 + t_2 + t_3$.

We also need to specify the probability that the DNA collides in the nth row. The simplest model [2] assumes that the probability of colliding in a given row is equal to the density of posts, $\rho = d/a$, where d is the post diameter and a is the center-to-center post spacing. As we can see in Fig. 1, we also need to account for the fact that the fully extended molecule cannot collide with any of the $n^* = l_c/a$ post rows between its leading and trailing ends. Thus, the probability of colliding in row $n < n^*$ is zero, and the probability of colliding in row $n \geq n^*$ is $\rho(1 - \rho)^{n-n^*}$. Geometration [14] corresponds to $\rho = 1$.

Although the key parameters in the model (ρ, a, n^*, and U) appear unambiguous thus far, in practice there is some uncertainty in their definitions. The post density and spacing are well-defined for microfabricated arrays [3,5,6], but not for quasiregular arrays like Fig. 1. For the latter, we use the average post diameter and center-to-center spacing, which does not account for array disorder.

Disorder also lends uncertainty to n^*, as the exact number of ''excluded posts'' varies from collision to collision. Relaxation of the DNA introduces further uncertainty into n^*. As indicated in Fig. 1(e), the leading end of the DNA tends to relax before the molecule disengages from the post [6], so the true value of n^* may be less than l_c/a. Indeed, even the precise value of l_c for stained DNA is debatable [17]. However, the DNA must relax partially before the next collision, so these errors may cancel out.

Perhaps the greatest source of ambiguity is universally using $U = \mu_0 E$ as the velocity of each step. During the collision (t_1 and t_2), the polymer has a limited number of conformations, so it is unlikely that the arms move at the same velocity as a random coil. Simulations [18] show that the friction coefficient not only depends on the DNA conformation, but that the deviation from the random coil friction coefficient is length dependent. In translation (t_3), dense arrays force the DNA to move in a zigzag manner, reducing the value of U relative to $\mu_0 E$. Moreover, there may be a small electro-osmotic flow, which both reduces U and varies from experiment to experiment. On the other hand, the strand tends to be oriented in the direction of the field shortly after leaving the posts, which leads to a transient where the effective velocity is greater than $\mu_0 E$. To avoid introducing any adjustable parameters, we proceed here with the prescribed definitions of ρ, n^*, and U, keeping in mind their limitations.

The model predicts an average hooking time $\tau = (3/2)l_c/U$. Using some representative values for T4 DNA [19], we predict the slope of τ versus E^{-1} to be 42 cm s/V, independent of ρ. The experimental value is also independent of ρ, but its numerical value (14 cm s/V) is lower [4]. The agreement is reasonable given the simplicity of the model, and we believe that uncertainty in U is the main source of error. The model also predicts that the average number of rows traversed between collisions, $\langle n \rangle$, is $n^* + (1 - \rho)/\rho$. As indicated in Table I, the agreement for low densities is remarkable. At the highest density, we suspect that the uncertainty in n^* between collisions is a major source of error. A more complex collision model, which accounts for relaxation between collisions, lateral diffusion in the matrix, the effect of DNA elasticity [7], and array disorder, might lead to better agreement with the experimental data over the full range of parameters.

From this microscale model, the transition probability density, $\psi(n, t)$, that a given cycle results in the DNA molecule traveling through n rows in the time t is

$$\psi(n, t) = \frac{2\rho U}{l_c}(1 - \rho)^{n-n^*} \exp\left[-\frac{2U}{l_c}\left(t - \frac{l_c + na}{U}\right)\right], \tag{2}$$

TABLE I. Comparison of the model and experimental results [4] on single T4 DNA molecules for the distance between collisions. Experimental values are presented in brackets next to the theoretical prediction.

Data set	Post diameter d (μm)	Post spacing a (μm)	Rows to collision $\langle n \rangle$
1	1.7	13.0	11.9 (12.6 ± 2.1)
2	1.3	6.3	14.9 (14.1 ± 4.2)
3	1.0	3.4	23.4 (14.0 ± 1.8)

valid for $n \geq n^*$ and $t \geq (l_c + na)/U$. Otherwise, $\psi(n, t) = 0$.

The cyclic motion is a Scher-Lax [20] CTRW, which can be solved exactly in Fourier-Laplace space. With s the Laplace variable and k the Fourier variable, the walk has the solution [21]

$$p(k, s) = \frac{1 - \Lambda(k, 0)}{s[1 - \Lambda(k, s)]}, \tag{3}$$

where $\Lambda(k, s)$ is the Fourier-Laplace transform of $\psi(n, t)$,

$$\Lambda(k, s) = \sum_{n \geq n^*} e^{-ikna} \int_{(l_c+na)/U}^{\infty} dt e^{-st} \psi(n, t). \tag{4}$$

The mean velocity, $\bar{U}^*$, and dispersivity, $\bar{D}^*$, are then computed by the following [21]: (i) differentiating Eq. (3) an appropriate number of times with respect to k; (ii) setting $k = 0$ in the derivatives; (iii) expanding the result for small s and inverting to the time domain; and finally (iv) using the standard relationships between the Fourier transform of p, the moments of p, and the parameters $\bar{U}^*$ and $\bar{D}^*$. With some algebraic manipulations, we ultimately arrive at the mean velocity

$$\frac{\bar{U}^*}{U} = \frac{2(\rho n^* + 1 - \rho)}{5\rho n^* + 2(1 - \rho)} = \left(1 + \frac{1}{\langle n \rangle} \frac{\tau U}{a}\right)^{-1}. \tag{5}$$

As in the Markovian case [12], the retardation from the posts is the product of the average collision rate, $1/\langle n \rangle$, and the ratio of trapping to convection, $\tau U/a$. The dispersivity,

$$\frac{\bar{D}^*}{Ua} = \frac{\rho(n^*)^2[10 + \rho(2n^* - 11) + (1 - n^*)^2\rho^2]}{[5\rho n^* + 2(1 - \rho)]^3}, \tag{6}$$

is markedly different from the Markovian model [12].

Before we compare Eqs. (5) and (6) to the experimental data, it is useful to consider some limiting cases. When $\rho = 1$, we recover the geometration results [14] $\bar{U}^* = 2U/5$ and $\bar{D}^* = Ul_c/125$. When the posts vanish, $d \to 0$, we recover the free-solution velocity with no post-induced dispersion. In the single-post limit ($a \to \infty$), $\bar{U}^* = U$ and $\bar{D}^* = Ul_c/5$, clearly illustrating the Nixon-Slater [8] ''catastrophic'' dispersion. When the posts are very dilute, the number of collisions is insufficient to alter the average velocity, but the scarcity of collisions results in a large dispersion. As a consequence, separations in very dilute post arrays would yield broad, overlapping bands.

With our microscale model, $\bar{U}^*$ and $\bar{D}^*$ scale linearly with U. Consequently, the model captures the experimental scaling laws discussed at the outset; the band broadening scales like E^{-1} and the resolution is independent of the field. Moreover, we capture the increase in band broadening with molecular weight [2].

We now compare our theoretical predictions to single-molecule experiments with T4 DNA. The experimental values are the results of a CTRW calculation based upon the measured distribution of passage times through a 110 μm section of the array [4]. As Table II demonstrates, the velocity estimates are good, whereas we overestimate the dispersivity. A sensitivity analysis shows that our values for $\bar{U}^*$ and $\bar{D}^*$ are most sensitive to a and increase in sensitivity with decreasing l_c. While some of the disagreement between theory and experiment may be attributed to uncertainty in the parameter values, we suspect that the simplicity of our model plays a more important role. For $\bar{U}^*$, the experimental value for the densest array is below the minimum value (0.4) predicted by the theory. This is likely due to the zigzag motion in dense arrays. Indeed, untrapped DNA in this array appeared to move between collisions with a mobility closer to $\mu_0/2$ [4]. Our predicted dispersivity is too high because the microscale model predicts much longer collisions than were observed in experiments. A more sophisticated trapping model should increase the agreement between the theory and experiment.

TABLE II. Comparison of the model and experimental results [4] on single T4 DNA molecules for $\bar{U}^*$ and $\bar{D}^*$. Experimental values are presented in the brackets next to the theoretical prediction.

Data set	Mean velocity $\bar{U}^*/U$	Dispersivity $\bar{D}^*/Ua$
1	0.59 (0.53 ± 0.08)	0.26 (0.027 ± 0.005)
2	0.47 (0.58 ± 0.11)	0.19 (0.154 ± 0.054)
3	0.43 (0.34 ± 0.06)	0.21 (0.054 ± 0.011)

While single-molecule experiments are useful in assessing the validity of the microscale model, in practice we are most interested in capturing the separation resolution. For a separation length L, the resolution is defined as [22]

$$R_s = \frac{|\bar{U}_1^* - \bar{U}_2^*|}{\bar{U}_1^* + \bar{U}_2^*} \sqrt{\frac{L}{16}\left(\frac{\bar{U}_1^*}{\bar{D}_1^*} + \frac{\bar{U}_2^*}{\bar{D}_2^*}\right)}. \tag{7}$$

In Fig. 2, we plot the predicted resolution per square root of the number of post rows for a hypothetical separation between λ, 2λ, and T4 DNA. The limited experimental data are not in disagreement with the model. The sensitivity of our predictions to the parameters increases as the difference in l_c decreases, which impacts our ability to make accurate predictions in sequencing applications. We also gain insight into the optimization of the arrays, since there is a plateau as the posts are moved farther apart. More widely spaced posts imply a larger device, so our result indicates an optimal post spacing on the order of 8 μm.

We have shown here how the Scher-Lax CTRW can be used to analyze the non-Markovian motion of DNA in post arrays. Using a simple model for ψ without any adjustable parameters, this model captures the macroscale dynamics, most importantly the separation resolution. We envision that more sophisticated functions for ψ will appear in the near future as more data in the spirit of Randall and Doyle [7] become available. In particular, a revised microscale model could incorporate different types of collisions [7,10]

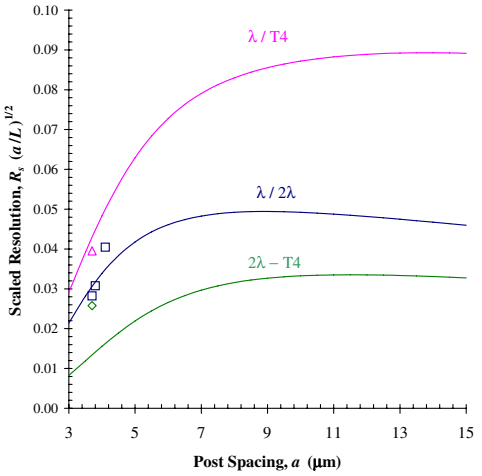

FIG. 2 (color online). Separation resolution per post row as a function of the post spacing for $d = 1.2$ μm. The error in the prediction from this choice of d is approximately 15%–30%. The curves correspond to theoretical predictions for the resolution between λ (48.5 kbp), 2λ (97 kbp), and T4 DNA (168.9 kbp). The experimental points are resolutions obtained for d between 1.0 and 1.4 μm. △, λ − T4; □, $\lambda - 2\lambda$; and ◇, 2λ − T4.

and DNA acceleration after collisions [4], as well as more precisely quantify the dynamics between post collisions (i.e., relaxation, precise quantification of n^*, etc.). Moreover, the Scher-Lax CTRW can easily be modified to include a different ψ for the first step [23], allowing us to rigorously incorporate the effect of the DNA injection into the matrix. We envision the model presented here will serve as a template for future studies of this type.

We are grateful to Martin Z. Bazant of MIT for a useful correspondence on CTRWs. K. D. D. acknowledges support from the International Human Frontier Science Program Organization. This work was supported by IFCPAR Project No. 2705-1 and the CNRS Microfluidics Network.

*Electronic address: Kevin.Dorfman@curie.fr

[1] P. S. Doyle *et al.*, Science **295**, 2237 (2002).
[2] N. Minc *et al.*, Anal. Chem. **76**, 3770 (2004).
[3] N. Kaji *et al.*, Anal. Chem. **76**, 15 (2004).
[4] N. Minc *et al.*, Electrophoresis **26**, 362 (2005).
[5] W. D. Volkmuth and R. H. Austin, Nature (London) **358**, 600 (1992).
[6] W. D. Volkmuth *et al.*, Phys. Rev. Lett. **72**, 2117 (1994).
[7] G. C. Randall and P. S. Doyle, Phys. Rev. Lett. **93**, 058102 (2004).
[8] G. I. Nixon and G. W. Slater, Phys. Rev. E **50**, 5033 (1994).
[9] E. M. Sevick and D. R. M. Williams, Phys. Rev. E **50**, R3357 (1994); E. M. Sevick and D. R. M. Williams, Phys. Rev. Lett. **76**, 2595 (1996); P. André, D. Long, and A. Ajdari, Eur. Phys. J. B **4**, 307 (1998).
[10] P. M. Saville and E. M. Sevick, Macromolecules **32**, 892 (1999).
[11] P. D. Patel and E. S. G. Shaqfeh, J. Chem. Phys. **118**, 2941 (2003).
[12] K. D. Dorfman and J.-L. Viovy, Phys. Rev. E **69**, 011901 (2004).
[13] J. M. Deutsch, J. Chem. Phys. **90**, 7436 (1989).
[14] S. Popelka *et al.*, J. Chromatogr., A **838**, 45 (1999).
[15] Krawczyk *et al.*, [24] attempted a similar extension of geometration by introducing a phenomenological coefficient for the mean distance between collisions. However, they assumed that the dispersion is the sum of that from the random cycle time and a Poisson-distributed number of collisions, rather than rigorously accounting for the nonseparability of the walk.
[16] Different distributions could be chosen, such as a Gaussian one [8], albeit at the expense of complicating the ensuing analysis.
[17] S. B. Smith, L. Finzi, and C. Bustamante, Science **258**, 1122 (1992); T. T. Perkins *et al.*, Science **268**, 83 (1995); O. B. Bakajin *et al.*, Phys. Rev. Lett. **80**, 2737 (1998).
[18] M. Kenward and G. W. Slater, Eur. Phys. J. E **14**, 55 (2004).
[19] For comparison with Ephesia data, we use $\mu_0 = 2.5 \times 10^{-4}$ cm^2/V s [2]. The l_c are estimated using the λ-DNA l_c of 16.3 μm as a baseline, adjusting for the number of base pairs, and then increasing l_c by 25% to account for elongation by YOYO-1.
[20] H. Scher and M. Lax, Phys. Rev. B **7**, 4491 (1973).
[21] B. H. Hughes, *Random Walks in Random Environments* (Oxford University Press, New York, 1995).
[22] J. C. Giddings *Unified Separation Science* (Wiley, New York, 1991).
[23] G. H. Weiss and R. J. Rubin, Adv. Chem. Phys. **52**, 363 (1983).
[24] M. J. Krawczyk, J. Dulak, and K. Kułakowski, Electrophoresis **23**, 182 (2002).