Preserving the Sequence of a Biopolymer's Monomers as They Enter an Electrospray Mass Spectrometer

William Maulbetsch, Benjamin Wiener, William Poole, Joseph Bush, and Derek Stein

Physics Department, Brown University, Providence, Rhode Island 02912, USA

(Received 1 September 2015; revised manuscript received 19 September 2016; published 17 November 2016)

This paper investigates how faithfully an electrospray mass spectrometer reports the order of monomers of a single biopolymer in the context of two sequencing strategies. We develop a simplified one-dimensional theoretical model of the dynamics of Brownian particles in the Taylor cone of an electropray source, where free monomers drift towards the apex in an elongational force gradient. The likelihood that neighboring particles will invert their order decreases near the apex because the strength of the force gradient increases. Neighboring monomers on a stretched biopolymer should be cleaved by photofragmentation within about 3 nm of the apex if they are to enter the mass spectrometer in sequence with 95% probability under typical experimental conditions. Alternatively, if the monomers are cleaved processively at milliseconds-long intervals by an enzyme, their sequence will be faithfully reported with 95% confidence if the enzyme is within about 117 nm of the apex.

DOI: 10.1103/PhysRevApplied.6.054006

I. INTRODUCTION

If small particles in solution are arranged in a row and then released, Brownian motion [1,2] scrambles their order over time. In situations where the initial sequence is important, one can attempt to preserve it by applying an elongational force gradient that pulls neighboring particles apart along the row. Here, we theoretically evaluate the effectiveness of this approach. We analytically model the dynamics of a pair of Brownian particles in a one-dimensional linear force gradient and calculate the probability that their order will invert as a function of time, the initial separation distance, and the strength of the force gradient.

Our study is motivated by a desire to obtain information from the temporal sequence of ions registered by a mass spectrometer, in particular, the monomers cleaved from the end of a single biopolymer. The electric field and fluid flow gradients that pull ions into a mass spectrometer generate a force gradient that can help to preserve the native sequence [3]. We use a simplified theoretical model to predict the conditions under which a mass spectrometer will faithfully report the sequence of DNA nucleotides. Although we originally conceived this application of mass spectrometry as a way to sequence DNA, the same principle applies to ribonucleic acid (RNA) and protein sequencing. As we discuss, the sequencing strategy seems best suited for proteins.

The sequence preservation model we develop can be applied in different contexts where information is encoded in the ordering of micrometer- and nanometer-scale objects in solution. For example, biomolecules, viruses, and cells all carry biological information, while nanoparticles, microscopic beads, and fluid droplets can transport cargo. There have been significant efforts to develop microfluidic devices to sort these objects in recent years [4,5], and it is straightforward to generate elongational force gradients in such devices with fluid flows [6], electrophoretic forces [7], or magnetic forces [8]. Thus, in cases where Brownian motion presents a fundamental problem, applying a force gradient may offer a simple solution to preserving a sequence of interest.

II. THEORETICAL MODEL

We consider the dynamics of two distinguishable, noninteracting Brownian particles in one dimension. The particles move in a concave potential which generates a linear force gradient. A particle at position x feels a force kx, where k is the strength of the force gradient. The system is illustrated in Fig. 1.

In addition to the gradient force, each particle experiences a fluctuating thermal force $f(t)$, which is the cause of Brownian motion [1]. The exact course of $f(t)$ in time t cannot be known, so $f(t)$ is treated as a stochastic variable. We assume, as is conventional, that the distribution of $f(t)$ is Gaussian, its average is $\langle f(t)\rangle = 0$, and its autocorrelation function is $\langle f(t)f(t')\rangle = 2\xi k_B T\delta(t-t')$, where ξ is the drag coefficient of a particle, $k_B T$ is the thermal energy, and $\delta(t)$ is Dirac's δ function.

Finally, a moving particle experiences a viscous drag force $-\xi[dx(t)/dt]$. We assume the system to be in the overdamped low-Reynolds-number regime; therefore, inertia can be ignored. The balance of the applied, thermal, and viscous drag forces gives the equation of motion for a single particle, $\xi[dx(t)/dt] = kx(t) + f(t)$.

We are interested in the ordering of the particles; therefore, we consider the separation between them $X \equiv x_2 - x_1$, where x_1 and x_2 are the positions of particles 1 and 2, respectively. Note that when X changes sign, the original order of the particles reverses.

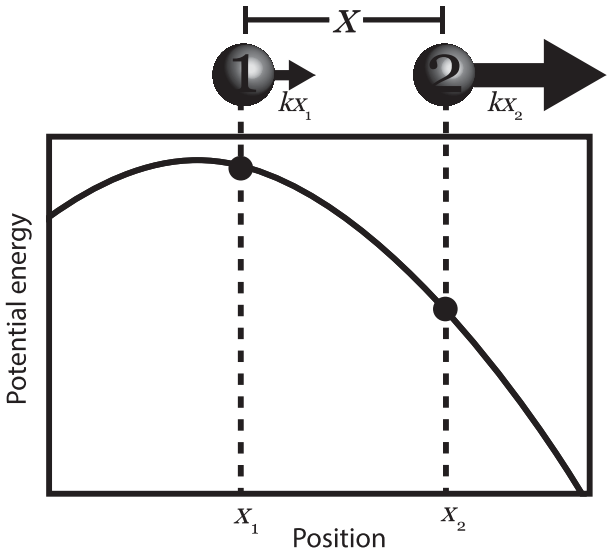

FIG. 1. Basic model. Two Brownian particles are at positions x_1 and x_2 in a repulsive force gradient which results from a concave potential energy landscape. The net force driving the two particles apart is proportional to $|X|$, their separation distance.

If ξ is the same for both particles, we can subtract the equation of motion for particle 1 from that for particle 2 to obtain the equation describing the dynamics of the separation,

$$\xi\frac{dX(t)}{dt} = kX(t) + F(t), \tag{1}$$

where $F(t) \equiv f_2(t) - f_1(t)$ is a new stochastic force corresponding to the difference between the thermal forces acting on each particle. $f_1(t)$ and $f_2(t)$ are uncorrelated, and it can be shown that the average and the autocorrelation function of $F(t)$ are, respectively,

$$\langle F(t)\rangle = 0, \tag{2}$$

$$\langle F(t)F(t')\rangle = 4\xi k_B T\delta(t-t'). \tag{3}$$

According to Eq. (1), the force gradient drives particles apart at a rate proportional to X. The force gradient is "elongational" because the farther apart the particles are, the faster they are driven apart. This statement is true even when both particles move in the same direction; the force on the leading particle always exceeds that on the trailing one by an amount proportional to X.

This situation is similar to an Ornstein-Uhlenbeck process or to a small bead held in an optical trap [9]. The crucial difference is the sign of the applied force; whereas a particle undergoing an Ornstein-Uhlenbeck process or held in an optical trap feels a restoring force proportional to its displacement, particles in an elongational force gradient are pulled apart by a force proportional to their displacement. Despite this important difference, we can solve Eq. (1) as we would the standard Ornstein-Uhlenbeck problem [1]. Our approach follows a familiar one for the motion of a Brownian particle in a harmonic trap [10].

The dynamics of the interparticle separation distance are driven in part by the unknowable random thermal force $F(t)$, so we must treat the problem probabilistically. Thus, we solve for the Green function $G(X, X_0; t)$, which gives the likelihood of finding two particles, initially a distance X_0 apart, separated by a distance between X and $X + dX$ after a time t. The Green function for this system is [1,10]

$$G(X, X_0; t) = \frac{1}{\sqrt{\frac{4\pi k_B T}{k}\left(e^{[(2kt)/\xi]} - 1\right)}} \times \exp\left[-\frac{(X - X_0 e^{(k/\xi)t})^2}{\frac{4k_B T}{k}\left(e^{[(2kt)/\xi]} - 1\right)}\right]. \tag{4}$$

Equation (4) differs from the standard Ornstein-Uhlenbeck solution in that the sign of the force gradient is flipped, and factors of 2 appear because Eq. (4) describes the separation between two diffusing particles.

A. The probability of particles crossing

When the two particles cross, the sign of X flips. The probability that particles initially separated by X_0 will have crossed after a time t, $P(X_0; t)$, is obtained from Eq. (4) by integrating $G(X, X_0; t)$ over all negative values of X,

$$P(X_0; t) = \int_{-\infty}^{0} G(X, X_0; t)dX. \tag{5}$$

Equation (5) can be rewritten in terms of the error function $\mathrm{erf}[q] \equiv \int_0^q e^{-t^2}dt$ as

$$P(X_0; t) = \frac{1}{2}\left(1 - \mathrm{erf}\left[\frac{1}{\sqrt{1 - e^{-[(2kt)/\xi]}}}\sqrt{\frac{kX_0^2}{4k_B T}}\right]\right). \tag{6}$$

We can gain physical insight and simplify Eq. (6) by introducing two dimensionless variables. The first $\lambda \equiv \sqrt{(kX_0^2/4k_B T)}$ parametrizes the strength of the force gradient driving the particles apart. λ is the work required to bring the particles from their initial positions to their midpoint against the force gradient, divided by $k_B T$. The second variable $\tau \equiv [(4k_B Tt)/\xi X_0^2]$ is the time nondimensionalized by the mean time a particle takes to diffuse across half the initial separation distance in the absence of a force gradient. The probability $P(X_0; t)$ is expressed in terms of these natural variables as

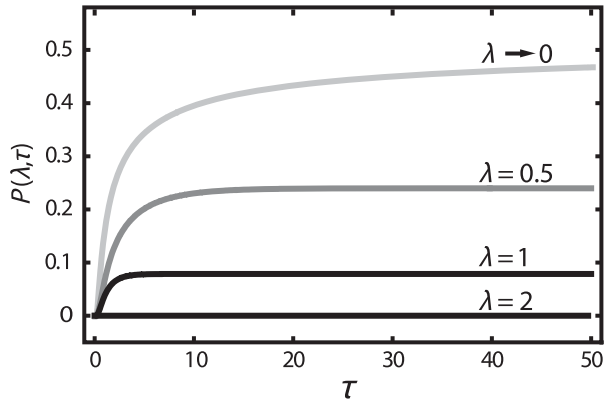

FIG. 2. Dependence of the crossing probability on τ for various fixed values of λ ranging from vanishingly small to 2.

$$P(\lambda;\tau) = \frac{1}{2}\left(1 - \mathrm{erf}\left[\frac{\lambda}{\sqrt{1 - e^{-2\lambda^2\tau}}}\right]\right). \tag{7}$$

Figure 2 shows the dependence of $P(\lambda,\tau)$ on τ for various fixed values of λ. Note that in the absence of a force gradient ($\lambda \to 0$), $P(\lambda,\tau)$ eventually reaches a plateau at 0.5. This means that freely diffusing particles have an equal chance of crossing one another or not, corresponding to the complete randomization of order. Conversely, for a sufficiently strong force gradient ($\lambda = 2$), the particles rarely reverse their order, even after an arbitrarily long time. The rapid rise in monomer crossing probability in Fig. 2 results from particles needing sufficient time to diffuse together from their initial positions. The subsequent probability saturation occurs because as particles move apart, the growing repulsive force between them makes a future crossing ever more unlikely. In Fig. 2, the crossing probability approaches a plateau beyond $\tau = 1$ because after a long time, the particles have either crossed already or they never will. The strength of the force gradient is what determines the likelihood of having crossed in the limit of long times. These observations show that the ordering of particles can be preserved to any desired level of confidence by applying a sufficiently strong force gradient.

III. PRESERVING THE ORDER OF DNA MONONUCLEOTIDES ENTERING A MASS SPECTROMETER

In this section, we apply our model to the motion of the ordered monomers of a biopolymer, such as a DNA, RNA, or protein, near the tip of an electrospray ion source. We envision sequencing a single molecule by cleaving its monomers from one end, delivering them one by one and in order into a mass spectrometer, and identifying them by their unique charge-to-mass ratios. For this strategy to succeed, the order of the monomers must be preserved to a high degree of confidence until they enter the mass spectrometer for identification. The calculations that follow focus on DNA nucleotides, which are the monomers of DNA. They also suppose that the required confidence level is 95%, which corresponds to Q13 bases in the standard measure of quality factor [11]. It is easy to repeat these calculations for any desired confidence level or for a different type of biopolymer. The implications of our results for biopolymer sequencing are discussed later.

We consider two distinct methods for cleaving nucleotides from a DNA strand. The first is to use laser light to photofragment a stretched DNA strand. The second is to have an exonuclease cleave the DNA enzymatically and processively, as was first proposed by Keller and co-workers [12]. We show that in both cases the nucleotides must be cleaved within less than a micrometer of the electrospray tip in order for their order to be preserved; in the case of photofragmentation, the critical distance is only a few nanometers. The precise distance depends significantly on the initial separation between liberated monomers, which depends, in turn, on the mechanism by which the polymer is cleaved.

A. Elongational force gradients in an electrospray ion source

The electrospray ionization technique transfers ions into a mass spectrometer from the liquid inside a capillary tube that tapers to a needlelike tip. A voltage is applied between the liquid and an electrode located a short distance in front of the tip. When the voltage is large enough, the electric fields it generates deform the liquid meniscus into a pointed shape called a Taylor cone [13]. Ions escape the Taylor cone from the apex in a charged fluid jet or by the mechanism of ion evaporation [3]. Figure 3 sketches an electrospray ion source with two nucleotides in it.

Inside the Taylor cone, a nucleotide experiences a combination of electric and viscous forces. The electric force increases in a straightforward manner as the nucleotide approaches the apex. The *average* viscous force, which is linked to the flow rate in the jet, also grows in a straightforward manner as a nucleotide approaches the apex. But the action of the electric fields on the induced charge at the Taylor cone surface also generates circulating flow patterns that complicate the situation substantially [14]. We first estimate the conditions needed for preserving the order of nucleotides by making simplifying assumptions about the electrospray and by considering only the average electrical and viscous forces. We later discuss the effect of the circulating flow component.

To find the strength of the electric force, we take the Taylor cone to be an ohmic liquid whose resistivity ρ is constant and uniform. The ions emitted from the apex carry a total current I that must be supplied by a current density J inside the Taylor cone. Far away from the jet, the spherically symmetric current density is $J(r) = I/\{2\pi[1 - \cos(\alpha)]r^2\}$ [15], with r the distance from the apex of the Taylor cone along its axis and $\alpha = 49.3°$ the

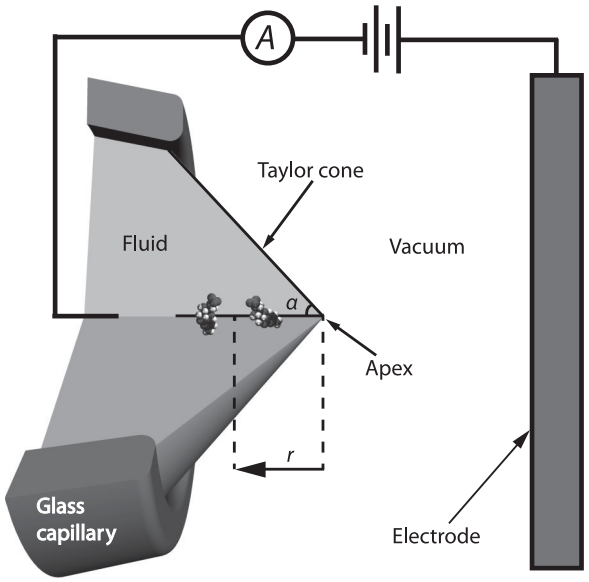

FIG. 3. Schematic of an electrospray ion source. A Taylor cone forms at the tip of a capillary needle when a high voltage is applied between the conducting fluid inside and a nearby electrode held in vacuum. A pair of DNA mononucleotides approach the apex. The sketch shows r and α.

characteristic half angle of a Taylor cone [13]. The electric field $E(r)$ is related to $J(r)$ and ρ by Ohm's law, $E(r) = \rho J(r)$. Each nucleotide in solution carries the charge of an electron $-e$, so it experiences an electrical force

$$f_e = -eE(r) = -\frac{e\rho I}{2\pi[1-\cos(\alpha)]r^2}. \tag{8}$$

Fluid flows also arise in electrosprays due to the flow of ions in charged interfacial layers. A narrow jet typically shoots from the apex of the Taylor cone. The volume flow rate Q must be supplied by a sink flow $U(r) = Q/\{2\pi[1-\cos(\alpha)]r^2\}$, where $U(r)$ is the flow velocity averaged over the section of the Taylor cone at radius r; note that it has the same form as the electric current density. The average viscous force on a nucleotide is $f_v = \xi U(r)$, where ξ is the nucleotide's viscous drag coefficient. De La Mora and Locertales [15] have shown theoretically and experimentally that under generic electrospray conditions, I and Q are related by

$$|I| = f(\epsilon)\sqrt{\gamma Q/\rho\epsilon}, \tag{9}$$

where γ is the surface tension of the liquid interface, ϵ is the dielectric constant, and $f(\epsilon) \approx 18$ for high-ϵ fluids like water ($\epsilon = 80$) and formamide ($\epsilon = 111$). Using the above relations and Eq. (8), we find an expression for the total force

$$f = -\frac{e\rho I}{2\pi[1-\cos(\alpha)]r^2}\left(1+\frac{I}{I_0}\right). \tag{10}$$

The first term in Eq. (10) corresponds to the electric force, which grows stronger in proportion with I. The second term corresponds to the viscous force, which grows in proportion with I^2; $I_0 = f(\epsilon)^2 e\gamma/\xi\epsilon$ is the characteristic current at which the viscous force reaches the same strength as the electric force and is typically on the order of a nanoampere.

As a nucleotide approaches the apex, the force on it increases as the inverse square of r, rather than depending linearly on r, as our theoretical model assumes. To apply the model, therefore, we Taylor expand (different Taylor) f and find the strength of the linearized elongational force gradient at r,

$$k \approx (\nabla \cdot f)|_r = \frac{2e\rho I}{2\pi[1-\cos(\alpha)]r^3}\left(1+\frac{I}{I_0}\right). \tag{11}$$

The force gradient diverges at the apex of the Taylor cone according to Eq. (11), so the order of two particles can be preserved to an arbitrary confidence level by releasing them close enough to the tip. Here we estimate R_{95}, the maximum distance from the apex at which two ordered nucleotides can be released and expected to enter the mass spectrometer in their original order with 95% confidence. Recall that the likelihood two particles will invert their order is greatest after long times. Thus, if we take the limit $t \to \infty$ of Eq. (7), we obtain an expression for the upper bound on the expected disorder, $P(\lambda) = \frac{1}{2}(1-\mathrm{erf}[\lambda])$. Setting $P = 0.05$ and numerically solving for λ leads to

$$k_{95} \approx 5.4 \times \frac{k_B T}{X_0^2}. \tag{12}$$

We combine Eqs. (11) and (12) to obtain the value of R_{95} from estimates of X_0. The value of X_0 depends on the particular technique used to cleave monomers from a DNA strand, as we show. Also note that the linearization in Eq. (11) systematically underestimates the difference in forces acting on a pair of particles. Consequently, the value of R_{95} we obtain is a conservative (short) estimate of the actual safe distance. The ratio of the leading nonlinear correction to the linearized gradient force is $3X_0/2R_{95}$.

B. Cleaving the monomers of a stretched polymer by photofragmentation

It may be possible to cleave a DNA polymer into monomer-sized pieces by irradiating it with ultraviolet light as it approaches the apex of the Taylor cone. DNA tends to break into fragments when a molecule absorbs more than a few electron volts of energy; mass spectrometric studies on nucleic acids in vacuum have found that DNA fragments primarily at the glycosidic bond, which holds a base to the sugar-phosphate backbone,

and secondarily along the backbone [16]. Furthermore, DNA has strong optical absorption bands near 200 and 280 nm in wavelength.

We suppose that a single strand of DNA approaches the apex of the Taylor cone and becomes stretched out by the force gradient before incident light causes the molecule to fragment, releasing neighboring bases essentially simultaneously. The initial separation distance of the bases is $X_0 = 5.9$ Å, the linear distance between them along the backbone of a single stranded DNA molecule [17]. Plugging this initial separation distance into Eq. (12), we find the critical force gradient to be $k_{95} = 0.068\ \mathrm{N\,m^{-1}}$.

Let us estimate a typical R_{95} from Eq. (11), given $k_{95} = 0.068\ \mathrm{N\,m^{-1}}$. In our laboratory, we routinely generate electrosprays from 2M solutions of sodium iodide (NaI) in formamide, similar to what Gamero-Castano and De La Mora reported [18]. The solution resistivity is $\rho = 0.45\ \Omega\,\mathrm{m}$, and the surface tension is $58\ \mathrm{mN\,m^{-1}}$. The viscous drag coefficient for the DNA monomer adenosine monophosphate has been measured experimentally in water and found to be $\xi \approx 9.1 \times 10^{-12}\ \mathrm{N\,s\,m^{-1}}$ [19]. Since ξ should be proportional to the viscosity of the fluid, we expect ξ to be 3.3 times greater in formamide than in water, giving $\xi \approx 3.0 \times 10^{-11}\ \mathrm{N\,s\,m^{-1}}$. With these parameters, Eq. (10) predicts $I_0 = 0.91$ nA for the electrospray current at which the electric and viscous contributions to the total force on a monomer are equal. Electrospray currents are typically higher than I_0, so the viscous force from the sink flow is the larger effect. Here we consider a typical current of $I = 5$ nA ($I/I_0 \approx 5.5$). The critical force gradient is achieved at $R_{95} \approx 3.2$ nm, the greatest distance from the apex at which photofragmentation can occur while preserving the monomer order with 95% confidence. We also note that the nonlinearity in the force gradient is significant at that location—the leading correction to the force difference between neighboring monomers is 0.28 times the linearized term.

It is, in principle, possible to control the critical distance by changing the force gradient through I or ρ; in practice, however, these approaches can have at best a modest impact on R_{95}. As the electrospray current is increased from $I = 1$ nA to $I = 10$ nA, for example, R_{95} rises only from 1.3 to 4.9 nm. If the resistivity is additionally increased tenfold to $\rho = 4.5\ \Omega\,\mathrm{m}$, R_{95} rises to 2.7 nm for $I = 1$ nA and to 10.5 nm for $I = 10$ nA. Clearly, neighboring monomers must be cleaved within a few nanometers of the electrospray apex if their order is to be preserved.

We note that at such short distances, an important assumption underlying our model of the Taylor cone no longer holds. Equation (8) is a good description of the electric fields inside the Taylor cone far from the apex, where charge relaxation by conduction is fast compared with convective transport [3,14]. However, convective charge transport becomes comparable with conduction close to the apex, on length scales comparable to or smaller than [3,15]

$$r^* \approx (\rho Q \epsilon \epsilon_0)^{1/3}. \tag{13}$$

Using Eqs. (9) and (13), we find $r^* \approx 3.1$ nm. The fact that $R_{95} \approx r^*$ suggests that the corrections to Eq. (11) are nonnegligible, and, therefore, the values of R_{95} we find are only approximate.

C. Cleaving DNA with an exonuclease

An alternative to photofragmentation is to allow an enzyme to processively cleave the nucleotides of a single-stranded DNA molecule. This idea comes from Keller and co-workers [12,20–22], who first proposed a single-molecule DNA sequencing technique in 1989. In their scheme, exonuclease I would be held in a fluid flow so that the released monomers would drift downstream through an optical focal volume where each one would be identified by fluorescence spectroscopy. The main problem this approach faced was the low signal-to-noise ratio obtained using optical spectroscopy. That problem might be solved by instead using mass spectrometry, which can easily detect and analyze a single ion.

Following this sequencing strategy, an exonuclease I molecule is immobilized a distance R_{exo} from the apex of the Taylor cone and made to processively cleave the nucleotides from a single DNA strand [12]. The kinetics are stochastic, giving rise to a distribution of intervals τ between subsequent nucleotide cleavages. Werner *et al.* [23] measured the distribution of cleaving rates and fit the data to a truncated Gaussian with a mean duration of $\langle\tau\rangle = 6.7$ ms and cutoffs for cleaving rates slower than 10 nucleotides/ sec. Through a simple change of variables, we convert the distribution of cleavage rates to a distribution of cleavage times

$$P_{\mathrm{exo}}(\tau) = \frac{1}{\sqrt{2\pi}\sigma_k \tau^2} \exp\left[-\frac{(\frac{1}{\tau} - k_0)^2}{2\sigma_k^2}\right], \tag{14}$$

where $\sigma_k = 63$ nucleotides/ sec is the measured rate variance, and $k_0 = 97$ nucleotides/ sec is the mean rate.

It might be tempting here to apply our model for order preservation in an elongational force gradient, but Eq. (7) does not apply well to this situation. To understand why, consider the average distance a monomer diffuses in a time $\langle\tau\rangle$, $\sqrt{2D\langle\tau\rangle} \approx 1.4\ \mu$m [10]. Compare this with the distance X_{drift} that a monomer drifts in the electric fields inside the Taylor cone in $\langle\tau\rangle$ when released 1.4 μm away from the apex. We find X_{drift} by equating Eq. (10) to the viscous drag force $-\xi[dX(t)/dt]$ and integrating,

$$X_{\text{drift}} = \sqrt{2D\langle\tau\rangle} - \left((\sqrt{2D\langle\tau\rangle})^3 - \frac{3e\rho I\langle\tau\rangle(1+I/I_0)}{2\pi\xi[1-\cos(\alpha)]} \right)^{\frac{1}{3}} \approx 136 \text{ nm}. \quad (15)$$

Thus, diffusive motion typically overwhelms drift in the time between cleavages. Drift becomes important only very close to the apex of the Taylor cone. The location where drift and diffusion are comparable effects depends on I; that location is only about 24 nm from the apex when $I = 1$ nA but grows to about 1.36 μm when $I = 10$ nA.

The arrival of monomers at the apex is better thought of as a first passage problem where only diffusion matters. When a free monomer diffuses to the apex of the Taylor cone, it is ejected into the mass spectrometer, never to return. If that happens before the next monomer is cleaved, then their relative order is preserved. Alternatively, if the first monomer does not reach the apex before the next monomer is cleaved, the former can diffuse upstream and flip the ordering. We compute the distance from the apex where a cleaved monomer stands a 95% chance of reaching the apex within τ. This approach establishes a lower bound on R_{95}. It is only a lower bound because we neglect the fluid flow and the electrophoretic drift which both help to bring the monomers towards the apex and preserve their order and because some fraction of monomers that do not reach the apex within τ will nevertheless reach it before the next monomer does.

The probability that a Brownian particle will not have diffused past some point x_c after time τ is called the survival probability $P_s(x_c, \tau)$ and is given by [24]

$$P_s(x_c, \tau) = \text{erf}\left(\frac{x_c}{2\sqrt{D\tau}}\right). \quad (16)$$

In our sequencing strategy, $x_c = R_{\text{exo}}$ is the distance from the exonuclease to the Taylor cone apex. We find the total survival probability $P_s(R_{\text{exo}})$ and also account for the stochastic exonuclease cleavage kinetics by multiplying the survival probability in Eq. (16) with the distribution of τ in Eq. (14) and then integrating over all possible τ,

$$P_s(R_{\text{exo}}) = \int_0^\infty P_s(R_{\text{exo}}, \tau) P_{\text{exo}}(\tau) d\tau. \quad (17)$$

We evaluate Eq. (17) numerically and obtain the value of R_{exo} for which $P_s(R_{\text{exo}}) = 0.05$, in accordance with our 95% confidence criterion. We find that the monomer order is preserved if the exonuclease I is located within $R_{95} = 117$ nm of the apex.

D. The influence of circulating flows in the Taylor cone

Barrero *et al.* [14] investigated circulating fluid flows inside the Taylor cone and showed that these flows are a fundamental feature of electrosprays at low and high Reynolds number. The electrospray voltage draws ions to the liquid interface, which obtains a net charge within a thin interfacial layer. The tangential component of the electric field exerts an electric stress on the fluid at the surface of the Taylor cone, pulling it toward the apex. The fluid returns up the axis of the Taylor cone, creating a meridional circulation. At low Reynolds number, the characteristic velocity of the circulating flow is

$$U_c(r) \sim \left(\frac{\gamma\epsilon_0\rho^2 I^2}{r^3\eta^2}\right)^{\frac{1}{2}}, \quad (18)$$

where γ is the surface tension, ϵ_0 is the permittivity of vacuum, and η is the fluid viscosity. Equation (18) is obtained by balancing the electric stress with the fluid shear across the Taylor cone at r. U_c can be large compared with other components of a monomer's velocity, and circulating flows have the potential to shuffle the order of nearby monomers. Let us consider the two sequencing strategies separately.

In the first case, we find that monomers need to be cleaved by photofragmentation within 3.2 nm of the apex for $I = 5$ nA. At such short distances from the apex, the velocity of a monomer from the combined effects of electrophoresis and the sink flow exceeds the characteristic velocity of the circulating flow; the net motion toward the apex, driven by the total force in Eq. (10), exceeds U_c at distances shorter than $r = 5.4$ nm. Furthermore, as we have already discussed, the assumption that charge relaxation by conduction is fast compared with convective transport does not hold at distances comparable to $r^* \approx 3.1$ nm; the development of the circulating flow rests on that assumption [3,14]. Given these considerations, we speculate that sink flow and electrophoretic motion are the dominant effects at the short distances required to preserve the order of photofragmented monomers. Numerical studies are likely required to understand the dynamics in detail, but such studies are beyond the scope of this paper.

In the case of exonuclease sequencing, the circulating flow at R_{95} is fully developed and fast compared with the other velocity components; however, its influence is somewhat counterintuitive. The circulating flow tends to *enhance* order preservation by increasing the effective diffusivity of monomers and thereby decreasing their first passage time at the apex. The underlying mechanism is similar to Taylor-Aris diffusion: When a free monomer can diffuse laterally, it will randomly sample the different streamlines of a circulating flow. The randomness in its axial motion becomes magnified as it rides the various streamlines. The effective diffusion coefficient D_{eff} is related to the thermal diffusion coefficient $D_0 = k_b T/\xi$ and the Péclet number $Pe \sim rU_c/D_0$ as [25]

$$D_{\text{eff}} = D_0(1 + \beta Pe^2), \quad (19)$$

where β is a factor that depends on the geometry and the flow ($\beta = \frac{1}{48}$ for Poiseuille flow in a cylinder). Under typical experimental conditions, $Pe \approx 50$ at R_{95}, so the enhancement in the effective diffusion coefficient should be substantial. We, therefore, conclude that our earlier finding is likely an underestimate of R_{95}. We do not attempt to make a more accurate estimate here.

Note that Eq. (19) is only valid in the limit where the monomer has enough time to fully explore the flow in the transverse direction. Since a monomer will diffuse about 2 μm perpendicular to the Taylor cone axis in between subsequent cleavages, it is reasonable to expect it to fully sample the circulating flow from a starting distance of 117 nm from the apex.

E. The implications for sequencing single biopolymers

Our analysis sheds light on the feasibility of sequencing single biopolymers using electrospray ionization mass spectrometry. To implement Keller's idea of sequencing DNA with the help of exonuclease I, the enzyme needs to be placed within about 100 nm of the apex of the ion source's Taylor cone, otherwise, the sequence of bases becomes significantly scrambled by Brownian motion. This sequencing seems technically feasible. It is straightforward to make electrospray sources with tip diameters on the 100-nm scale by pulling glass capillaries [26], and one can imagine fixing exonuclease I enzymes to the tip's surface. Those enzymes are within about a tip diameter of the apex of the Taylor cone, which is close enough to ensure that the sequence of bases is reliably preserved. A similar approach to single-molecule protein sequencing is possible by replacing the exonuclease I with an AAA + family protease that processively degrades proteins and releases their constituent amino acids one by one [27]. Since the rate at which such proteases typically operate is slower than that of exonuclease I, we expect better than 95% sequence preservation when operating 100 nm from the apex.

Using photofragmentation instead of an enzyme to cleave the monomers of a biopolymer presents different challenges. Most significantly, the photofragmentation must be carried out within a few nanometers of the apex of the Taylor cone. An electrospray source with a diameter that small has yet to be demonstrated. We further note that our model may not accurately describe such a small electrospray source, whose structure is expected to deviate from a perfect Taylor cone at such short distances from the apex; that is the scale at which the cone typically transitions into a thin jet of fluid [15,28], which ultimately breaks up into charged droplets.

New fabrication techniques offer a possible solution to the challenge of preserving the monomer order. One can create an electrospray source featuring a nanotube made of carbon or boron nitride at the tip, whose aperture has a diameter in the single-nanometer range. It has been shown that nanotubes with diameters less than 10 nm can be incorporated into a chip-based nanofluidic device [29]; the same technique has been used to insert nanotubes into pulled glass capillary tips [30]. Such a nanotube electrospray source can additionally prevent bases from swapping their order by allowing too little room inside the tube for monomers to diffuse past one another.

Finally, we remark that other technical challenges must be overcome in order for the biopolymer sequencing strategy to be viable. Immobilizing a single enzyme within tens of nanometers of a desired location is not trivial, though solutions to similar problems have been developed in the context of zero-mode-waveguide DNA sequencing [31]. Another challenge for DNA sequencing with an exonuclease is that the technique needs to be massively parallelized in order to process the vast number of nucleotides in a genome in a reasonable amount of time. This challenge is somewhat daunting because while there has been progress toward miniaturizing mass spectrometry [32,33], there remains considerable distance to go before it can be parallelized.

We believe that the biopolymer sequencing strategy considered here is best suited for proteins. Proteins, which are typically a few hundred to a few thousand amino acids long, are 6–7 orders of magnitude shorter than DNA genomes; therefore, parallelization is not required for high throughput. Furthermore, *de novo* sequencing of proteins, which is the direct determination of a protein's sequence in the absence of a reference sequence, remains difficult even by state-of-the-art methods. At present, two different methods are commonly employed. The first is Edman degradation [34], which provides the most reliable protein sequences but requires short protein fragments and slow, costly chemical cycles. The typical sequence read length is 10 to 20 amino acids, and the process can take around 20 min per amino acid and cost around $70 per amino acid. The second is based on mass spectrometry and algorithmic reconstructions of multiply fragmented protein sections [35]. This method can sequence polypeptides that are tens to thousands of amino acids long. Analyses are still slow and expensive, however, taking weeks to complete at a cost of around $10 per amino acid. Overall, protein sequencing is still orders of magnitude slower and more costly than DNA sequencing, which is presently nearing the $1000 per genome milestone. This is why *de novo* protein sequencing is less common an objective than protein identification, whereby parts of a molecule's sequence are compared against a database.

The challenge of protein sequencing stems partly from the fact that there are 20 amino acids to discriminate, as opposed to only four DNA bases. This places a heavier burden on the resolution of the sensor. Mass spectrometry arguably offers the best hope for a single-molecule protein sequencing technology, as it is uniquely capable of

identifying all 20 amino acids. The strategy considered here, therefore, has the potential to greatly improve the state of the art. We envision an entire denatured protein passing through a nanoscale aperture leading into a mass spectrometer, which sidesteps the need to prepare purified small protein fragments or to run slow and expensive chemical cycles. Long read lengths will further reduce the computational postprocessing needed to stitch together small fragments. Finally, the speed of sequencing by mass spectrometry with photofragmentation is, in principle, limited by the count rate of the single-ion detector, which can exceed 5×10^8 Hz [36].

IV. CONCLUSION

We analyze the use of elongational force gradients for preserving the linear order of particles against the randomizing effects of Brownian motion. The analytic expression we derive for the probability that two particles will invert their order after being released from a known initial separation depends on two rescaled parameters, one related to the strength of the force gradient and the other to the time interval following the release of the particles. The model can be easily applied to a variety of micro- and nanoscale situations where Brownian motion competes with a spatially varying force. We apply our model to the case of DNA nucleotides being released within the Taylor cone of an electrospray ion source in order to evaluate the feasibility of a sequencing strategy for single biopolymers. The model makes simplifying assumptions about the dynamics inside the Taylor cone.

The force gradient inside the electrospray ion source can preserve the ordering of DNA monomers cleaved from a parent strand of DNA. If the exonuclease I enzyme sequentially cleaves the monomers of DNA from a parent strand within about 100 nm of the tip of the electrospray source, the sequence of monomers entering the mass spectrometer is 95% preserved. If the monomers are cleaved from a stretched parent DNA strand by photofragmentation, the corresponding distance from the tip is between 1 and 10 nm, depending on the experimental conditions. The sequencing strategy is best suited for proteins because they are composed of 20 different monomers and relatively short, which takes advantage of the resolution of mass spectrometry while minimizing the importance of parallelization. Our theoretical model can be improved in the future by accounting for finite-size objects, interactions between them, and the detailed motion of the fluid and particles in three dimensions.

ACKNOWLEDGMENTS

This work is supported by National Institutes of Health Grant No. NHGRI R21HG005100-01 and by Oxford Nanopore Technologies, Ltd. D. S. declares a financial relationship with Oxford Nanopore Technologies, Ltd.

[1] George E. Uhlenbeck and Leonard S. Ornstein, On the theory of the Brownian motion, Phys. Rev. **36**, 823 (1930).
[2] Ming Chen Wang and George Eugene Uhlenbeck, On the theory of the Brownian motion II, Rev. Mod. Phys. **17**, 323 (1945).
[3] Juan Fernández de La Mora, The fluid dynamics of Taylor cones, Annu. Rev. Fluid Mech. **39**, 217 (2007).
[4] Nicolas G. Green and Hywel Morgan, Dielectrophoretic separation of nano-particles, J. Phys. D **30**, L41 (1997).
[5] L. Jullien, A. Lemarchand, and H. Lemarchand, Diffusion of reactive species tuned by modulated external fields: Application to high performance chromatography, J. Chem. Phys. **112**, 8293 (2000).
[6] Thomas T. Perkins, Douglas E. Smith, and Steven Chu, Single polymer dynamics in an elongational flow, Science **276**, 2016 (1997).
[7] Chia-Fu Chou, Jonas O. Tegenfeldt, Olgica Bakajin, Shirley S. Chan, Edward C. Cox, Nicholas Darnton, Thomas Duke, and Robert H. Austin, Electrodeless dielectrophoresis of single-and double-stranded DNA, Biophys. J. **83**, 2170 (2002).
[8] Alexander van Reenen, Arthur M. de Jong, Jaap M. J. den Toonder, and Menno W. J. Prins, Integrated lab-on-chip biosensing systems based on magnetic particle actuation—A comprehensive review, Lab Chip **14**, 1966 (2014).
[9] Karel Svoboda and Steven M. Block, Biological applications of optical forces, Annu. Rev. Biophys. Biomol. Struct. **23**, 247 (1994).
[10] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Oxford University Press, New York, 1986).
[11] Brent Ewing and Phil Green, Base-calling of automated sequencer traces using Phred. II. Error probabilities, Genome Res. **8**, 186 (1998).
[12] James H. Jett, Richard A. Keller, John C. Martin, Babetta L. Marrone, Robert K. Moyzis, Robert L. Ratliff, Newton K. Seitzinger, E. Brooks Shera, and Carleton C. Stewart, High-speed DNA sequencing: An approach based upon fluorescence detection of single molecules, J. Biomol. Struct. Dyn. **7**, 301 (1989).
[13] Geoffrey Taylor, Disintegration of water drops in an electric field, Proc. R. Soc. A **280**, 383 (1964).
[14] A. Barrero, A. M. Ganan-Calvo, J. Davila, A. Palacio, and E. Gomez-Gonzalez, Low and high Reynolds number flows inside Taylor cones, Phys. Rev. E **58**, 7309 (1998).
[15] J. Fernández De La Mora and I. G. Loscertales, The current emitted by highly conducting Taylor cones, J. Fluid Mech. **260**, 155 (1994).
[16] Jin Wu and Scott A. McLuckey, Gas-phase fragmentation of oligonucleotide ions, Int. J. Mass Spectrom. **237**, 197 (2004).
[17] Steven B. Smith, Yujia Cui, and Carlos Bustamante, Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded dna molecules, Science **271**, 795 (1996).
[18] M. Gamero-Castano and J. Fernández De La Mora, Direct measurement of ion evaporation kinetics from electrified liquid surfaces, J. Chem. Phys. **113**, 815 (2000).
[19] Martin Dworkin and Kenneth H. Keller, Solubility and diffusion coefficient of adenosine 3': 5'-monophosphate, J. Biol. Chem. **252**, 864 (1977).

[20] W. Patrick Ambrose, Peter M. Goodwin, James H. Jett, Mitchell E. Johnson, John C. Martin, Babetta L. Marrone, Jay A. Schecker, Charles W. Wilkerson, Richard A. Keller, Alberto Haces *et al.*, Application of single molecule detection to DNA sequencing and sizing, Ber. Bunsen-Ges. Phys. Chem. **97**, 1535 (1993).

[21] Lloyd M. Davis, Frederic R. Fairfield, Carol A. Harger, James H. Jett, Richard A. Keller, Jong Hoon Hahn, Letitia A. Krakowski, Babetta L. Marrone, John C. Martin, Harvey L. Nutter *et al.*, Rapid DNA sequencing based upon single molecule detection, Genet. Anal. Biomol. Eng. **8**, 1 (1991).

[22] James H. Werner, Hong Cai, James H. Jett, Linda Reha-Krantz, Richard A. Keller, and Peter M. Goodwin, Progress towards single-molecule DNA sequencing: A one color demonstration, J. Biotechnol. **102**, 1 (2003).

[23] James H. Werner, Hong Cai, Richard A. Keller, and Peter M. Goodwin, Exonuclease I hydrolyzes DNA with a distribution of rates, Biophys. J. **88**, 1403 (2005).

[24] Sidney Redner, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, England, 2001).

[25] Brian J. Kirby, *Micro- and Nanoscale Fluid Mechanics: Transport in Microfluidic Devices* (Cambridge University Press, Cambridge, England, 2010).

[26] Elizabeth M. Yuill, Niya Sa, Steven J. Ray, Gary M. Hieftje, and Lane A. Baker, Electrospray ionization from nanopipette emitters with tip diameters of less than 100 nm, Anal. Chem. **85**, 8498 (2013).

[27] Robert T. Sauer and Tania A. Baker, AAA + proteases: ATP-fueled machines of protein destruction, Annu. Rev. Biochem. **80**, 587 (2011).

[28] Alfonso M. Gañán-Calvo, On the general scaling theory for electrospraying, J. Fluid Mech. **507**, 203 (2004).

[29] Alessandro Siria, Philippe Poncharal, Anne-Laure Biance, Rémy Fulcrand, Xavier Blase, Stephen T. Purcell, and Lydéric Bocquet, Giant osmotic energy conversion measured in a single transmembrane boron nitride nanotube, Nature (London) **494**, 455 (2013).

[30] Eleonora Secchi, Sophie Marbach, Antoine Niguès, Derek Stein, Alessandro Siria, and Lydéric Bocquet, Massive radius-dependent flow slippage in carbon nanotubes, Nature (London) **537**, 210 (2016).

[31] Michael J. Levene, Jonas Korlach, Stephen W. Turner, Mathieu Foquet, Harold G. Craighead, and Watt W. Webb, Zero-mode waveguides for single-molecule analysis at high concentrations, Science **299**, 682 (2003).

[32] Steve Arscott and David Troadec, A nanofluidic emitter tip obtained by focused ion beam nanofabrication, Nanotechnology **16**, 2295 (2005).

[33] Eric Wapelhorst, Jan-Peter Hauschild, and Jörg Müller, Complex MEMS: A fully integrated TOF micro mass spectrometer, Sens. Actuators A Phys. **138**, 22 (2007).

[34] Victoria Pham, Jake Tropea, Suzy Wong, James Quach, and William J. Henzel, High-throughput protein sequencing, Anal. Chem. **75**, 875 (2003).

[35] Hanno Steen and Matthias Mann, The abc's (and xyz's) of peptide sequencing, Nat. Rev. Mol. Cell Biol. **5**, 699 (2004).

[36] A. S. Tremsin, J. V. Vallerga, J. B. McPhate, and O. H. W. Siegmund, Optimization of high count rate event counting detector with microchannel plates and quad timepix readout, Nucl. Instrum. Methods Phys. Res., Sect. A **787**, 20 (2015).