# Symmetry Breaking in Polynucleotide Translocation Through Biological Nanopores Utilizing a Brownian Ratchet and Direct Single-Molecule Manipulations.

Thomas Anthony Pologruto

Department of Biophysics, Harvard University, Cambridge, MA 02138.

## Abstract

Rapid nanopore discrimination of polynucleotides has wide applicability to high-throughput DNA sorting and low copy-number DNA sequencing. Meller *et al.* [[1]] have shown that polymers of the same length but with different base composition are discernible using single-channel recordings of the translocation duration (lifetime) and the residual ionic current. The event diagram for polymer translocation indicate that some polymers traverse the pore more slowly than other identical polynucleotides, as indicated by additional peaks in the pore lifetime histogram. Kasianowicz *et al.* [[2]] first described this curious result and they speculated that the three lifetimes they observed corresponded to distinct physical situations. The two longer lifetimes reflected complete polymer translocation in two different orientations (namely 5’ ® 3’ vs. 3’ ® 5’), while the shortest lifetime peak was attributed to polymers that did not completely traverse the pore. Lubensky and Nelson [[3]] then provided a theory for polynucleotide traversal through a pore with which it has strong interactions. Their model accounted for the presence of the two longer lifetime peaks by noticing that an asymmetrical potential for the pore-polymer interaction could discriminate between the two orientations by producing a different force, and hence a distinct traversal time, for each orientation of the polymer. The alternate explanation of the multiple peaks is that each peak corresponds to a different secondary structure of the polymer. The latter hypothesis is supported by recent data [1] accumulated from monitoring polynucleotides traversing an a-hemolysin pore at different temperatures.

In order to resolve the debate, we are proposing the introduction of a ratcheting asymmetry that will perturb the lifetime histogram in a specific way depending on which hypothesis is correct. The natural 5’/3’ asymmetry of the polymer is hidden in these translocation experiments. By exploiting this intrinsic asymmetry, it is straightforward to understand if local (*orientational*) or global (*conformational*) effects are responsible for the multiple peaks in the translocation lifetime histogram. The exploitation of this asymmetry will also allow the prediction of which lifetime corresponds to which orientation, something paramount when this technology is used for sequencing. We will further predict the magnitude of the perturbation caused by the ratcheting using the theory of Lubensky and Nelson. Finally, we will show how to construct suitable substrates for force measurements using an optical trap. These direct measurements will determine the stall force for a given polymer translocating through the pore in each possible orientation. The stall force reflects the pore potential and dictates the traversal time for a given polynucleotide and orientation. Direct determination of this quantity for different orientations of the same polymer will compliment the ratcheting data, since if the stall force is the same for each direction that implies the potential is also the same, supporting the *orientational* hypothesis. By repeating these force measurements with different polymers in the same orientation, it is possible to determine the effect of secondary structure on the stall force. When the ratcheting mechanism is introduced, the stall force of the Brownian ratchet will be determined in a similar fashion. Comparison of the orientational and ratcheting data will yield a definitive answer as to the factors responsible for the multiple peaks in the lifetime histogram. This system is also a very good model for protein translocation through the endoplasmic reticulum, and is one of the first models that allow the force of a Brownian Ratchet, a mechanism implicated in numerous cellular processes, to be measured directly and precisely.

## Introduction

The ability to sort and sequence different polynucleotides rapidly is paramount to nearly all fields of biology. From the quick identification of mutations in the genetic code to elucidating the mechanism of an important transcription factor, DNA sequencing has aided in nearly every major biological breakthrough in the last twenty years. Modern sequencing requires macroscopic samples of double stranded DNA (dsDNA), associated priming sites, and the use of fluorescent dideoxy nucleotides. The fluorescent probes have helped in the automation of the sequencing process, making high-throughput sequencing feasible and efficient.

Whole genome (shotgun) sequencing has become commonplace using parallel bulk sequencers, and a genome sequence is assembled after six to ten times the total number of bases has been sequenced, using a complex mathematical assembly process. While the shotgun sequencing approach to DNA sequencing is effective, it is essentially a brute-force technique that has associated statistical, systematic, and repetitive sequence errors. In addition, genome assembly is limited by islands of contiguous sequences overlapping, with a resolution governed by the automated sequencers (currently approximately a 1,000 base maximum contiguous sequence is possible).

## Experimental Background: Initial Observations

The use of single-channel recordings to sequence single stranded DNA (ssDNA) or RNA has recently become feasible. This alternate sequencing approach uses single molecules of ssDNA (RNA) that pass through a narrow biological pore, and the base sequence of the ssDNA (RNA) is translated into electronic signals directly and quickly. This method can be made high-throughput via nanofabrication of channels and associated shuttling devices. In addition, it is possible to sequence longer sequences in a continuous fashion than with the conventional sequencers, minimizing some of the errors previously described, which are associated with shotgun sequencing [[4]]. The small sample size utilized by single-channel recordings (single molecule resolution possible) and its use for rapid discrimination between different polynucleotides makes these techniques very appealing to a wide range of scientists.

The first experiments that showed the feasibility of such single-channel polynucleotide discrimination were performed by Kasianowicz *et al.* [2] and a synopsis of their work is given here. The experiments involved: 1) imbedding a single membrane channel in a lipid bilayer, 2) passing a steady current through the channel, 3) adding a sample of polynucleotides to the anode (*cis*) side of the membrane, and 4) monitoring the current blockage durations as a function of the number of translocation events as the ssDNA passes from the *cis* into the *trans* (cathode) chamber. Each of these steps is elucidated in more detail below.

The use of a membrane channel that can remain open for several seconds was a prerequisite to driving ssDNA through a membrane. The mitochondrial voltage dependent ion channel and the heptameric *Staphylococcus aureus* a-hemolysin both can remain open for extended periods and allow the passage of current across an otherwise insulating lipid bilayer. A recording from a single channel is a paradigm of neurobiology, and many of these techniques are borrowed from that field. A diphytanoyl-PC bilayer was formed across a .1mm-diameter orifice in a 25-mm thick Teflon partition separating the two buffer filled compartments [[5]]. The a-hemolysin monomers spontaneously form heptameric pores when imbedded in a membrane. a-Hemolysin pores assemble in the bilayer partition in less than 5-minutes and assurance of single channel insertion was realized by extensive chamber flushing after incorporation [[6]]. The structure of the *Staphylococcus aureus* a-hemolysin was solved by X-ray diffraction and reveals a mushroom shaped complex with a 10-nm long channel and a 1.5-nm diameter pore at its narrowest constriction [[7]]. This constriction is roughly the same size as the diameter of a ssDNA strand. Two views of the a-hemolysin pore are given in **Figure 1**.

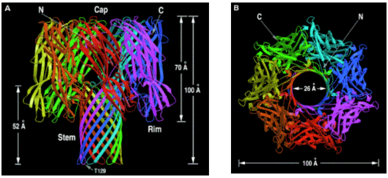


**Figure 1** Structure of the a-hemolysin pore from Song *et al.* (1996). [7]

After the pore is inserted, a transmembrane potential of 110-140 mV is applied across the pore. The anode (negative) is designated as the *cis* chamber, while the cathode (positive) is designated as the *trans* chamber. The experimental set-up is depicted graphically in **Figure 2.** When the channel is open, a current will flow through the pore in proportion to the applied voltage. Since the channel is so narrow compared to the large buffer chambers on each side of the membrane, it has a much larger electrical resistance, and most of the applied potential will fall across the pore region. In a typical experiment, a –125 mV potential was applied and a –125 pA current flowed through the pore, implying an Ohm's resistance of $10^9$ W. A typical recording of the voltage and current is given in **Figure 3**.

Once a steady current had been established, polynucleotides of mean length 210 bases were placed into the *cis* compartment up to a final concentration of 500 nM. These were initially poly[U] RNA molecules, but later a variety of different polymers was used [1]. These polymers caused blockages in the current that lasted for discrete amounts of time, while in the absence of polymers, virtually no blockages were recorded [2]. The blockage of current was attributable to polymers occupying the pore and thus preventing ions from passing through. In these blocking events, the current through the pore dropped dramatically, from 120 pA to only 10-15 pA. A diagram of these blockage events is depicted in **Figure 4.** Corroborating biochemical evidence showed the amount of ssDNA or RNA that accumulated in the *trans* chamber depended on the number of channels in the membrane and was proportional to the total number of events under peaks 2 and 3 in the lifetime histogram (**Figure 5**). Thus, the current blockages must represent ssDNA/RNA traversing the membrane through the a-hemolysin pore(s) [2].

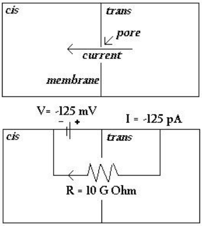


**Figure 2** Schematic of the buffer chamber and the direction of the applied current and voltage. The bottom picture is the circuit equivalent of the buffer chamber in the absence of any polynucleotides, but after a potential is applied.

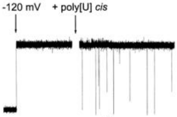


**Figure 3** Voltage and current recorded as a function of time from a single pore in the absence of any polynucleotides. Figure and legend are taken from [2, Figure 1].

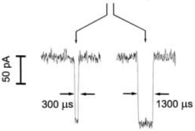


**Figure 4** After the addition of polynucleotides to the *cis* compartment, discrete events where the current dropped dramatically could be recorded, as seen in (**4a**). The expanded view (**4b**) reveals the individual blockage durations, which were typically a few hundred msec. Figure and legend are taken from [2, Figure 1].

The duration of the pore current blockages was characterized and binned, producing a histogram where the number of events lasting a given duration is plotted versus the pore lifetimes. Some blockages lasted much longer than a second and were attributable to polymers "stuck" in the pore, apparently due to secondary structure outside of the pore. Reducing or inverting the applied voltage could clear these events. These longest lifetime blockages were seen with poly [U], poly[A], poly[C], poly[dT], poly[dC], as well as with a synthetic ssDNA composed of poly[dA,dT,dC] [2]. In the analysis, these peaks were not input into the histogram.

The pore lifetime histogram recorded for the poly[U] (210 base) RNA is given in **Figure 5.** This histogram reveals some surprising results. The mixture of polynucleotides is homogenous and monodisperse, and yet three characteristic lifetimes appear for polymer translocation. For extended polymers, the blockade lifetimes should be proportional to the length of the polymer and inversely proportional to the voltage applied. Analysis of the peaks in **Figure 5** as a function of length and inverse voltage shows that while the polymers in peaks 2 and 3 display the expected dependence on these parameter, peak 1 events were invariant to them (**Figure 6**). Blunt-ended, dsDNA also showed very short pore lifetimes analogous to peak 1, since these molecules could not fit through the pore because of their size. These observations led to the conclusion that peak 1 corresponds to polymers that are not fully translocated through the pore, but rather enter and exit the pore on the *cis* side.

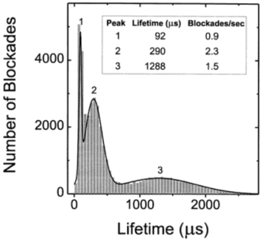


**Figure 5** Histogram of the blockage events as a function of their lifetime (msec). Three peaks are observed. The shortest lifetime peak (peak 1: mean lifetime of 92 msec) is not sensitive to polymer size or applied voltage, and represents polymers that are not completely translocated through the pore. The two longer lifetime peaks (peaks 2 & 3: mean lifetimes of 290 and 1288 msec, respectively) are dependent on polymer length and inversely proportional to applied voltage. They represent full polymer translocation through the pore. The smooth envelope and the mean lifetimes were determined by fitting the data to a sum of three Gaussian curves. Figure and legend are taken from [2, Figure 2].

While partial translocation events are expected, the presence of two longer lifetime peaks displaying characteristic voltage and length dependence is puzzling. Why should a mono-disperse ensemble of equal-length polymers traverse the same pore at such different rates? Peaks 2 & 3 contain almost the same number of events, and are distinct from each other. The connected curve in **Figure 5** is determined by the sum of three Gaussian curves, indicating each peak corresponds to a physically different situation, with a different associated mean and variance.

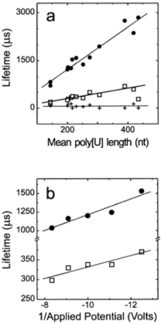


**Figure 6** Poly[U]-induced channel blockages were directly proportional to the length of the polymer and inversely proportional to the applied voltage, as seen in (**6a**) and (**6b**), respectively. The plots (**6a**) show lifetimes for peaks 1 (+), 2([]), and 3(·) in an experiment with V = -120 mV with 13 size selected poly[U]s and (**6b**) for peaks 2([]) and 3(·) with poly[U] of mean length 215 bases at a given inverse voltage. Figure and legend are taken from [2, Figure 3].

The authors suggest that the natural asymmetry of the polynucleotide is responsible for peaks 2 & 3 in the lifetime histogram. Namely, one of the peaks corresponds to traversal in the 5’ ® 3’ direction while the other peak corresponds to translocation in the 3’ ® 5’ direction. This concept, which accounts for the presence of multiple lifetimes because of the asymmetry of the polynucleotide, will henceforth be referred to as the *orientational* hypothesis.

# Theoretical Background: Overview

The physical basis for the asymmetry of the pore-polymer potential motivated Lubensky and Nelson [3] (hereafter referred to as LN) to formulate a theoretical description of these experiments; namely, to study driven polynucleotide translocation through a narrow pore that has strong interactions with the polymer. In order to better understand the experiments to be proposed later in this paper, we shall go through the basic arguments and derive the key parameters from LN’s theory. This is essentially the theory behind the *orientational* hypothesis.

Many polymers are successfully modeled as flexible “beads-on-a-string” or “links-in-a-chain”, where the conformational entropy of the backbone dominates all the relative dynamics of the system, including its approach to equilibrium as well as its final equilibrium state [[8]]. This type of analysis breaks down when the detailed interactions of the “links” or “beads” are considered, or when the monomer/polymer interacts with an external potential. In the experiments described above, the external electric field and the a-hemolysin pore provide the external potential with which the polynucleotides interact. These potentials arise because the polymers are charged and because the diameter of the pore is commensurate with the chemical repeat unit of the polynucleotide (both are on order of » .16 nm). Therefore, the potential that arises due to the detailed chemical interactions of the polymer with the pore cannot be ignored, and in fact will dominate the dynamics of the system (as we shall show later).

LN examined the motion of a homopolymer threaded through a narrow pore with which it has strong interactions. The reasons LN give for assuming the pore interactions are dominant in these translocation experiments are threefold. The lifetime histogram data for the initial experiments described above are easiest to interpret if the polymer's speed is determined by the interactions of the pore with the polymer, since this is where molecular scale asymmetries could be important. Second, recent experiments with different polynucleotides show the variability in pore lifetimes that is strongly dependent on base composition, and thus it is not unreasonable to conclude that the specific chemical interactions of the pore with the polymer give rise to such variance. Lastly, the absolute value of the mean translocation lifetime shows that even the fastest polynulceotide moves far more slowly than simple estimates of hydrodynamic drag would suggest [3]. The drag is due to the viscosity of the fluid, h, and is proportional to the velocity of the polymer , n. If you model the polymer as having radius r and the pore as having radius R, the drag force per unit length on the part of the polymer in the pore is 2phrn/(R – r), and the drag from the parts of the polymer outside the pore is 2 x 6phbn, where b is the Kuhn length of the molecule [[9]]. Equating this drag force to the driving force *F* (this will defined in more detail later), the polymer would be expected to move at $10^8$ nt/sec, 100 times faster than observed (**Figure 6**). Thus, LN postulated that the potential of the pore and its interactions with the polymer should account for these observations and discrepancies.

LN took the pore to be sufficiently small that only a completely unfolded polymer could pass through the pore, an assumption realized in the experiments described. They only consider the dynamics once the polymer has inserted in the pore, and ignore how the polymer inserted there. The essential reduction to make the problem tractable is pushing it from three-dimensions (3D) to one-dimension (1D). LN accomplish this by showing that in the presence of a force driving the polymer through the pore, there should be a regime where the polymeric degrees of freedom outside of the pore can be neglected, and the system then becomes one-dimensional. A two-tiered picture is presented: a coarse-grained macroscopic description and a detailed microscopic model.

The macroscopic model is very general and resembles the more familiar diffusion with drift and conservation equations of fluid mechanics. This macroscopic picture has all of the relevant details of the actual potential (either electric or pore) engulfed in a few parameters, and it treats the translocation process as essentially a one-dimensional diffusion process. The microscopic model predicts the value and form of the macroscopic parameters. LN modeled the potential of the pore-polymer interaction as a tilted washboard potential, whose periodicity mimics that of the polynucleotide's sugar-phosphate backbone. This combination of macroscopic one-dimensional diffusion with drift coupled with a microscopic picture of the pore potential demonstrates how a simple physical mechanism can account for several striking features of the data from **Figure 5**.

# Theoretical Background: Macroscopic Coarse-Grained Description

The macroscopic model given by LN for the one-dimensional diffusion of the polymer is a very general result that requires few assumptions, aside from those usually given in fluid mechanics, and does not rely directly on the microscopic dynamics of the system. The macroscopic model is of considerable more interest and is much more relevant to our discussions here because it will be used to calculate the distribution of passage times for a polymer being driven through the pore. We shall concentrate on the macroscopic picture for most of this paper. The details of the microscopic picture are not necessary for understanding the arguments presented here, but are nevertheless fascinating and elucidating, and they will be gone over briefly.

The most important parameter introduced by LN is that of the force *F* driving the polymer through the pore. *F* is defined as the mean force required to immobilize a given monomer in the pore, where the average is taken over time and over all the monomers in a given polymer. This definition of *F* is reminiscent of what is called the stall force $F_S$ in experiments done using optical trapping with RNA polymerase [16] or kinesin [17]. In these trapping experiments, $F_S$ is defined as the minimum force required to halt the translocation of the protein motor. The similarity between *F* and $F_S$ and the latter's appealing name justify calling *F* the stall force of the polynucleotide, where the subscript is dropped for convenience. *F* is how much force we would need to assert on the polynucleotide to stop it from moving once it has entered and has begun to translocate through the pore. Note that the definition of *F* precludes it from including neither forces that average to zero when taken over the ensemble nor any forces due to hydrodynamic drag. In a equilibrium view of the process, we can take exp($Fa/k_BT$) as the ratio of the probability that the polymer will move forward one base to the probability that it will move backward one base, where a is the chemical repeat distance of the polynucleotide backbone (*a* » .6 nm) and $k_BT$ is the unit of thermal energy.

We can make an estimate of *F* if the only force driving the polymer through the pore is the potential across the membrane. We have already seen that the narrow pore dimensions make its resistance very large, and therefore the monomers in the pore should feel nearly the entire applied potential. Since each monomer has unit charge e, the amount of energy gained by moving a monomer through the pore from the *cis* to *trans* chambers is eV, where V is the applied potential. Since the length of the monomer is *a*, the force this monomer feels should be on the order of eV/*a*. If we apply a 125 mV potential, we have the following as a rough estimate of *F:*

$$F \approx \frac{eV}{a} \approx 5\frac{k_B T}{a} \qquad (1)$$

This value for *F* is an upper bound, but shows that the force can be considerable when expressed in appropriate units (~ 33 pN at room temperature).

The main purpose of the coarse-grained description of polymer diffusion is the calculation of the distribution of blockage times (like those seen in **Figure 5**). The main quantity of interest will be the probability P(x,t) that a given contour length x of the polymer's backbone has passed through the pore at time t. The value of x ranges from 0 (when the polymer has just enters the pore on the *cis* side) to L (when the entire polymer length L has exited the pore on the *trans* side). L can be written equivalently as L = N*a*, where *a* is the length of a monomer (.6 nm) and N is the number of bases in the polynucleotide (N is 210 for the poly[U] RNA in **Figure 5**). In order for this model to make sense, the length L must be greater than *a*, and the dynamics of the counterions must be fast compare dot the polymer dynamics. Both of these conditions are satisfied in these experiments [1,2].

The conservation of the probability density P(x,t) will govern the macroscopic dynamics of polymer diffusion. LN found that the probability current j(x,t) (which is the probability density per unit time) and the probability density P(x,t) must satisfy the familiar hydrodynamic conservation equation:

$$\frac{\partial P(x,t)}{\partial t} + \frac{\partial j(x,t)}{\partial x} = 0 \qquad (2)$$

which implies that the probability current should be linear in P(x,t). The simplest way to express this is by only allowing the lowest order terms in the expansion of j(x,t) to be considered. Thus, keeping the zero and first order derivatives of P(x,t), we can write that:

$$j(x,t) = vP(x,t) - D\frac{\partial P(x,t)}{\partial x} \qquad (3)$$

where we have allowed a linear term because of the driving. This resembles Fick's first law for mass transport in 1D. If we conserve the probability current (by substituting equation 3 into equation 2), we derive the familiar diffusion equation with drift for a 1D particle:

$$\frac{\partial P(x,t)}{\partial t} = D\frac{\partial^2 P(x,t)}{\partial x^2} - \nu\frac{\partial P(x,t)}{\partial x} \qquad (4)$$

The constants D & n that appear in (4) are the effective diffusion constant and the average drift velocity of the polymer, respectively. The actual values of these constants are derived from the microscopic physics of the pore potential and the electric field bias, and both can vary in a complicated way with respect to the applied field(s) [3].

The microscopic model that would predict D & n involves the pore potential, and the calculations and analysis are greatly simplified if the potential is invariant under a translation by an integer multiple of monomers (length *a*) in either direction. This symmetry in the potential becomes transparent if we are able to neglect the degrees of polymer freedom outside the pore, which in turn should be feasible if the pore polymer interactions are sufficiently strong, an assumption LN have already justified. It is a well-known paradigm in condensed matter physics that the symmetry and the conservation laws for the system will govern the long-length scale dynamics [3]. Here, we see that the coarse-grained hydrodynamic description implies a conservation of probability density, while the microscopic pore potential is symmetric under translations. By looking at each of these, then, we should be able to get a good description of the polymer dynamics on the time scales relevant to the experiments in question [1,2].

Using dimensional analysis, we can observe in equation (4) that there exists a characteristic length scale for the diffusion of the polymer. This becomes easier to visualize if we re-write (4) with the following coordinate transforms: t ® nt = t', x ® x':

$$\frac{\partial P(x',t')}{\partial t'} = \frac{D}{\nu}\frac{\partial^2 P(x',t')}{\partial x'^2} - \frac{\partial P(x',t')}{\partial x'} \qquad (5a)$$

$$\frac{\partial P(x',t')}{\partial t'} = l_d\frac{\partial^2 P(x',t')}{\partial x'^2} - \frac{\partial P(x',t')}{\partial x'} \qquad (5b)$$

We define $l_d$ = D/n as the diffusive length, and all of the macroscopic parameters are related to this length scale. Physically, if you examined the polymer on length scales smaller than $l_d$ you would not notice the driving by the electric field (which is governed by n) but only see the diffusion of the chain, whereas on length scales greater than $l_{d,}$ the driving would dominate the motion. If we did indeed have a 1D particle diffusing in a field under the influence of a uniform force *f*, we would see that an Einstein relation would hold, and that n = D*f*/($k_B$T) and $l_d$ = $k_B$T/*f*. This implies a revised definition of $l_d$ as the length over which the driving force *f* does one $k_B$T of work. We shall use *f* as the generic driving force for the rest of our discussions.

In order to calculate the distribution of passage times, LN assumed that the length of the polymer L was much larger than $l_d$, which is justified in these experiments [1,2]. LN defines Y(t) as the probability that a polymer takes time t to move completely through the pore. A rough idea of the first cumulants of this distribution can be reached via simple calculations. If a polymer of length L moves with average speed n, the expected time for passage should be <t> » L/n and the variance in the distance traveled in a time <t> is $(Dx)^2$ = 2D<t> (from the definition of a random walk in 1D), which implies the variance in the lifetimes would go as $(Dt)^2$ º <(t - <t>)$^2$> » $(Dx)^2/n^2$ which is equivalent to $(Dt)^2$ = 2DL/$n^3$. In the limit of an infinitely long polymer, these crude approximations are actually quite good. Conversely, in the limit where L » $l_d$, the driving should negligible since the polymer only "feels" the driving on scales longer than $l_d$, and the mean traversal time should approach the n = 0 limit of $<t>_{n=0}$ = $L^2$/6D.

The explicit calculation of Y(t) was formulated as a first-passage problem [3,[10]]. The solving of equation (4) on the interval [0,L] is required for the calculation of Y(t), with the absorbing boundary condition that P(0) = P(L) = 0 as well as an initial starting point, $x_0$. In these calculations, the interesting limit will be $x_0$ ® 0, implying that the polymer starts completely on the *cis* side of the pore. Specifically, the probability current at L, j(L), will yield the probability per unit time that the polymer will leave the pore from the *trans* side, while -j(0) is the probability per unit time that the polymer will exit from the *cis* side which it entered.

LN went through the calculation and derived Y(t), <t>, $(Dt)^2$ explicitly from equation (5b) and the given boundary conditions [3, Appendix A]. However, considering Y(t) itself is much more illuminating, LN wrote this as an infinite series expansion, which is the only form possible for arbitrary L/$l_d$. If you neglect the terms in this expansion that become exponentially small as L/($l_d$nt) ® ¥, a simple expression for Y(t) can be obtained:

$$\Psi(t) \cong \frac{\nu}{2}\sqrt{\frac{l_d}{\pi}}\left(\frac{L^2}{l_d(\nu t)^{5/2}} - \frac{2}{(\nu t)^{3/2}}\right)\exp\left(-\frac{(\nu t - L)^2}{4\nu t l_d}\right) \qquad (6)$$

This expression in equation (6) is not valid for large values of t where a negative probability would be attained. However, for t near the maximum in the distribution Y(t) where nt/L ~ O(1), it is accurate to within a percent for L/$l_d$ as small as 4. For larger values L/$l_d$, the qualitative aspects of Y(t) are still valid. The graph of equation (6) is a skewed Gaussian with a mean and maximum that are well separated. This is depicted in **Figure 7**.

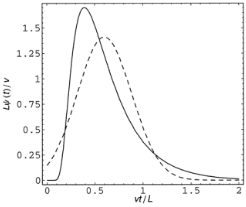


**Figure 7** The solid curve is the normalized distribution of passage times LY(t)/ n plotted versus nt/L, where each parameter is dimensionless. The dashed curve is a Gaussian with the same mean and variance as Y(t). Figure and legend are taken from [3, Figure 2].

Notice how the curve has a single peak, and that the maximum and shape of the curve are dictated by the specific values of D and n. Thus, each peak in **Figure 5** must correspond to a different physical situation with correspondingly different values for D and n. This corroborates the idea that the different lifetimes exhibited in **Figure 5** represent the different physical orientations of the polymer translocating through the pore [1,2]. The exact nature of the interaction that could cause such a difference in translocation times is explored in the next section (**Theoretical Background: Microscopic Model of the Pore**).

The interesting parameters of Y(t), as **Figure 7** illustrates, are the maximum of the distribution $t_{max}$ and the width of the peak dt, the latter of which is defined as dt º $(t_R – t_L)/2$, and $t_R$ and $t_L$ satisfy $Y(t_R,t_L) = e^{-1/2}\, Y(t_{max})$. A very useful expression for $t_{max}$ can now be obtained:

$$t_{max} = \frac{L}{v}\left(1 - 5\frac{l_d}{L} + \frac{17}{2}\frac{l_d^2}{L^2} + 32\frac{l_d^3}{L^3} + \ldots\right) \tag{7}$$

which can be seen to converge on <t> = L/n in the limit as L ® ¥. However, for finite L, $t_{max}$ falls away rapidly from this asymptotic value. Interestingly, the value of $dt/t_{max}$ depends only on $L/l_d$ and not on the macroscopic parameters D and n, so a measurement of $dt/t_{max}$ can yield an expression for $l_d$, assuming L is known. From the data in [2], LN proposed that $l_d$ was on the order of 40 nucleotides.

The macroscopic picture just presented describes the parameters that are measured in the lab, namely the distribution of passage times Y(t) and its peak $t_{max}$. These two parameters are perturbed by experimental details such as base composition, applied voltage, the length of the polymer, as well as by the addition of any other driving force(s). When we consider the ratcheting perturbation in the **Specific Aim I** section of this paper, we will use the macroscopic model of LN to predict the form and degree of this disturbance. The main idea from LN's macroscopic model is that each peak in the distribution of passage times corresponds to a different physical situation manifest in different values of the parameters D and n.

The physics of calculating the diffusion coefficient and drift-velocity, (which are the macroscopic parameters D and n, respectively) have been left unanswered in the above analysis, and indeed are not accessible by such a coarse-grained model. Rather, a detailed microscopic calculation of the pore potential is necessary.

# Theoretical Background: Microscopic Model of the Pore

LN postulate that there is a driving force *F* and a potential U(x) that govern the free energy function of the pore F(x), where x is the polymer translocation parameter and is the only dynamical variable considered. F(x) is defined as U(x) – Fx, where LN have assumed that the potential is not dependent on the driving force, *F*, an assumption that does not alter the conclusions drawn in a significant way. U(x) is the pore-polymer potential and it is only dependent on the x, the fraction of the polymer translocated. For a homopolymer (like poly[U]) the potential should have a periodicity that is commensurate with the chemical repeat distance *a* of the polymer. In the case of polynucleotides, *a* = .6 nm = 1 nucleotide (nt). The problem is now formally that of a point particle diffusing in a periodic potential U(x) driven be a constant force *F.* The probability P(x) of finding such a particle at a point x is given by a Smoluchowski equation that has the form:

$$\frac{\partial P(x)}{\partial t} = D_0 \frac{\partial}{\partial x}\left(\frac{\partial P(x)}{\partial x} + \frac{U'(x) - F}{k_B T} P\right) \tag{8}$$

where $D_o$ is the bare diffusion constant, and is different from D (the parameter in the macroscopic model) which defines the polymer's motion on length scales smaller than $l_d$. D includes the effects of U(x) and could vary in a complicated way with the applied field. Here, we assume that $D_o$ is unaffected by the counterionic flow. Equation (8) can be solved and complicated expressions for D and n can be obtained [3, Appendix B]. These calculations are not relevant here, so they will be omitted.

On the contrary, the predicted form of the potential U(x) is very interesting and illuminating. A very easy way to understand a periodic potential is via a sawtooth potential function, which is defined by the two parameters $U_o/ k_BT$ and the asymmetry parameter a. $U_o$ is the peak height of the potential defined as the difference between the value of U(x) at its maximum and minimum. The asymmetry parameter perturbs the sawtooth potential in such a way that if you moved your finger over it in one direction versus the other, the slope of the teeth would be different in each direction.

The asymmetry parameter a is defined such that when a = ½, the potential is perfectly symmetrical, and when a = 0 or 1, the potential has maximum asymmetry. Alternatively, we can say that if the distance between peaks (or troughs) is *a*, then the distance from one minimum to the next maximum is a*a*. **Figure 8** shows the skewed sawtooth potential and the relevant parameters. While no *a priori* information is available about the form or value for a, the value of $U_o$ should be of the order of a few $k_BT$, with $Fa/ k_BT$ » 5 in the experiments described above [1,2]. The full potential F(x) is thus defined in terms of $D_o$, *a*, a, the pore potential height $U_o/ k_BT$, and the driving term $F/ k_BT$.

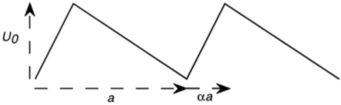


**Figure 8** Sketch of the sawtooth potential and relevant asymmetry and periodic parameters. Figure and legend are taken from [3, Figure 5].

The key to LN's microscopic model, which accounts for the large differences in traversal lifetimes, is that the probability for a particle to diffuse over a barrier is proportional to the *exponential* of the potential height. Thus, small differences in the potential height $U_o$ can lead to large deviations in the macroscopic lifetimes. If we refer to **Figure 5**, we can see that if the 5'/3' orientation is the cause for the multiple peaks in the lifetime histogram, it is necessary that an atomic scale asymmetry (namely the 2 distinct polymer backbone interactions with the narrow pore) be responsible for the form of the histogram. The mechanism by which the seemingly small 5'/3' asymmetry causes a macroscopic perturbation necessitates that the interaction potential be sensitive to such molecular detail, as well as be magnified greatly by small differences. The proposed unsymmetrical sawtooth potential and the exponential dependence of the crossing probability on the energy satisfy both of these requirements, making the *orientational* hypothesis reasonable and well grounded theoretically. **Figure 9** shows how by applying a reflection and a tilting, the asymmetric potential can lead to different barrier heights for each orientation, while keeping the average gradient fixed.

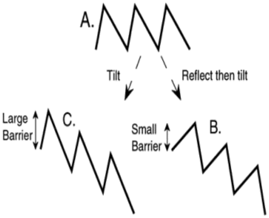


**Figure 9** Sketch of the reflection and tilting of the asymmetric sawtooth potential. The tilting reveals the polymer interacting with the pore (in either orientation), whereas the reflection determines any *orientational* effects leading to different traversal times. While the peak in the lifetime histogram with the longer lifetime corresponds to the direction with the larger barrier (**9c**), the prediction of which direction of translocation corresponds to which of the peaks (and hence potentials) is left unresolved in LN's model. Figure and legend are taken from [3, Figure 6].

.

To get an idea of how this sawtooth potential can reflect molecular asymmetries, LN showed that if U(x) is unsymmetrical, the forces $F$ and $-F$ will lead to different barrier heights, and thus to different mean drift speeds for the diffusing polymer. This change of sign of $F$ does not exactly correspond to changing the polymer orientation, because the orientation of the pore, the field, and the polymer must be considered together. **Figure 10** shows all the possible configurations of pore and polymer with a fixed electric field applied. The changing of the potential from U(x) ® U(-x) in LN's model corresponds to changing the direction of the pore (between A & B or C & D).

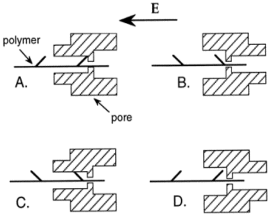


**Figure 10** The four possible orientations of polymer and pore in the case of a fixed electric field. The transformation from U(x) ® U(-x) in LN's model corresponds with the pairs A & B and C & D, whereas the changing of polymer direction for fixed field corresponds to pairs A & C and B & D as in the experiments previously described [1,2]. Pairs A & B and C & D should have the same response to an applied voltage, whereas other pairs (like the observed A & C, B & D) correspond to different potentials. Figure and legend are taken from [3, Figure 7].

## Theoretical Background: Conclusions

The microscopic model provides a good framework in which to understand the interactions of the pore and the polymer, and shows how very detailed molecular interactions and asymmetries can generate experimentally observable results. The macroscopic model uses the effective diffusion constant D and the drift velocity n, both of which need to be determined from the microscopic potential, to predict the form of the observed data. In particular, LN showed that each peak in the lifetime histogram (**Figure 5**) corresponds to a different physical situation, with distinct values for D and n. They supported this claim by formulating their microscopic potential to be asymmetric, and thus showed how the orientation of the polymer (and only the part of the polymer in the pore) can dictate the traversal lifetime, in support of the *orientational* hypothesis put forth previously [2].

In the model, LN demonstrated the importance of the quantity $F$, which governs the lifetimes. They suggest that since a large value of $Fa/k_BT$ is important for the potential sequencing of polynucleotides by pore techniques, it is very important to determine this value experimentally. They go on to say that if one could exert a non-electrical force on the polymer strong enough that its mean velocity through the pore fell to 0, this would yield a direct measurement of $F$. This is precisely the purpose of this proposal (see **Specific Aim II**).

# Experimental Background: Further Observations

While the data from [2] and the theory of LN [3] are supportive of the *orientational* hypothesis, recent results suggest that the orientation of the polymer in the pore may be irrelevant to the macroscopic dynamics of polymer [1]. In these studies, different ssDNA polymers are driven through an a-hemolysin pore by a field, similar to the experiments described previously [2]. By using a variety of different sequences, Meller *et al.* [1] have showed that rapid discrimination of ssDNA polynucleotides is feasible. Different polymers can be discerned by analysis of their "event diagrams," which are plots of their translocation duration versus blockade current for an ensemble of events [1]. The translocation duration is the same as the lifetimes previously discussed (**Figure 5**), while the blockade current reflects the residual current that passes through the pore when the polymer is present.

There are three statistical parameters that define a given polymer in these experiments: the most probable translocation current $I_P$, the most probable translocation duration $t_P$, and the characteristic dispersion of values for an individual translocation duration, $t_T$ [1]. $I_P$ can be defined in terms of $I_B$ by examining a typical blockage event as depicted in **Figure 11**. $I_B$ is defined as the normalized blockade current level and is equal to $<I_{event}>/<I_{open}>$, where $I_{event}$ is the average current during a blockage, and $I_{open}$ is the average current from 150 msec before to 150 msec after the blockage. These parameters can be visualized in **Figure 11**. For a given ensemble of polymers, $I_P$ is simply the most probable current derived from a histogram of all the $I_B$'s. The blockage duration is also defined in **Figure 11**, analogous to its previous definition.

The first experiments done were with poly[dA]$_{100}$ or poly[dC]$_{100}$ and each of the polynucleotides was characterized by the duration of the blockage $t_D$ they produced as well as the residual ionic current $I_B$. Events correspond to the troughs in **Figure 11**, and can be scatter-plotted with the $t_D$ for a given translocation event plotted versus the $I_B$ for that event, as depicted in **Figure 12a**, producing the "event diagram" previously mentioned. Notice in these scatter-plots that there are two distinct groups of events, corresponding to each of the polymers, and there is less than 1% overlap between the two clusters, indicating that good discrimination between polymers is readily achievable.

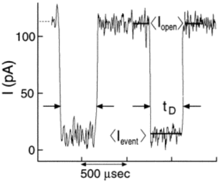


**Figure 11** A typical current trace showing two DNA translocation events (within the pairs of facing arrows) when a 120-mV gradient is applied across a membrane with a single a-hemolysin pore inserted. This is essentially the same data as from **Figure 4**. Figure and legend are taken from [1, Figure 1].

By taking vertical and horizontal slices through the event diagram and counting the number of events that lie in these slices, two histograms can be constructed. The horizontal slices will produce a lifetime histogram analogous to **Figure 5**. The vertical slices produce a residual current histogram. The residual current and lifetime histograms are given in **Figure 12b** & **12c**, respectively. The events for each of the polynucleotides separate into two groups when plotted on the histograms, as were previously observed [2]. The group with the lower residual ionic current is defined as "group 1," while the remaining events are defined as "group 2." The majority of events are included in "group 1" for any given polymer.

The current histograms are fit by the sum of 2 Gaussian curves, with the poly[dA] group 1 (80% of the total number of events) having a peak at $I_{P1}$ = .115 and the poly[dA] group 2 (20% of the total number of events) having a peak at $I_{P2}$ = .152. For the poly[dC] ssDNA, we have group 1 (70% of the events ) with a peak current of $I_{P1}$ = .125 and group 2 (30% of the events) with a $I_{P2}$ = .157. If we take only the group 1 events for the polymers, we can compute the lifetime histogram as mentioned previously. This histogram shows a clear peak, which is defined as $t_{P1}$ (**Figure 12c**) and the difference between peaks for the 2 different polymers is large. The ratio of $t_{P1}$ (dA)/ $t_{P1}$ (dC) = 2.8, where $t_{P1}$ (dA) = 330 msec (translocation rate of 3.3 msec/base) and $t_{P1}$ (dA) = 120 msec (translocation rate of 1.23 msec/base). The distribution of lifetimes ($t_D$) is Gaussian for $t_D < t_{P1}$, but for $t_D > t_{P1}$ the distribution is best fit by an exponential with time constant $t_T$. This time constant is much longer for poly[dA]$_{100}$ events then for poly[dC]$_{100}$ events, with $t_{T1}/t_{P1}$ = .6 for the former and $t_{T1}/t_{P1}$ = .2 for the latter, indicating a larger temporal dispersion for the poly[dA]$_{100}$ polymers (i.e. a higher $t_{T1}/t_{P1}$ value).

The discrimination between different polynucleotides does extend beyond homopolymers. In experiments with poly[dAdC]$_{50}$ and poly[dA]$_{50}$[dC]$_{50}$ , which have identical base composition but different sequences, the event diagrams observed are extremely overlapped (**Figure 13a)**, and the corresponding distribution of blockade currents $I_B$ and blockade lifetimes $t_{D1}$ (group 1 lifetimes) were also convoluted (**Figure 13b & 13c**). These distributions did show the characteristic double peak in the current histogram (i.e. division into groups 1 and 2 was again possible) and each of the events in these groups displayed different mean lifetimes $t_{P1}$ & $t_{P2}$.

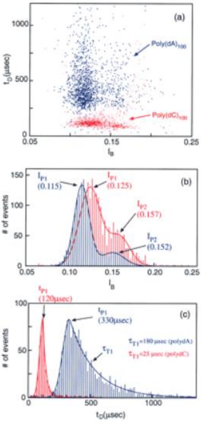


**Figure 12** **(a)** Event diagram showing translocation duration versus blockade level for poly[dA]$_{100}$ (blue) and poly[dC]$_{100}$ (red) at 20.0°C. **(b)** The current histograms derived from vertical slicing of the event diagram. **(c)** Lifetime histogram derived from the horizontal slicing of only those events in the event diagram that fall into group 1. In both **(b & c)** the solid lines are fits by Gaussian curves, or by exponential curves for the points in the lifetime histogram that have lifetimes greater than the peak value ($t_D > t_{P1}$) in **(c)**. Figure and legend are taken from [1, Figure 2].

Despite these similarities, the group 2 distribution of blockade lifetimes for each of the polymers was quite different, with the group 2 exponential time constants for the poly[dA]$_{50}$[dC]$_{50}$ and poly[dAdC]$_{50}$ polymers being $t_{T2}$ = 260 msec and 65 msec, respectively (**Figure 13d**). Therefore, since the events did show different $t_{T2}$'s and $t_{P2}$'s, they were discernible owing to their unique statistical properties. When a similar set of experiments was conducted with poly[dT]$_{50}$[dC]$_{50}$ and poly[dTdC]$_{50}$, unique event diagrams were again realized, with different values for $I_P$, $t_P$, and $t_T$. A summary of the statistical parameters for the various polymers is given **Table 1**.

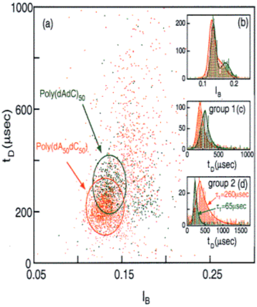


Figure 13 (a) Event diagram at 20°C for poly[dA]$_{50}$[dC]$_{50}$ (green) and poly[dAdC]$_{50}$ (orange). The ovals represent > 95% of the events are located within their boundaries. These events overlap much more so than for the homopolymers depicted in Figure 12. (b) The blockade current histogram shows the overlap between the polymers, as well as the characteristic peak doubling. (c) The blockage lifetime histogram for group 1 events, with heavily overlapped peaks. (d) The blockage lifetime histogram for group 2 events, showing distinct peaks as well as vastly different exponential lifetimes. Here, the group 2 exponential time constants for the poly[dA]$_{50}$[dC]$_{50}$ and poly[dAdC]$_{50}$ polymers are $t_{T2}$ = 260 msec and 65 msec, respectively, a four-fold difference. This is what distinguishes the two polymers most convincingly. This figure and legend were taken from [1, Figure 3].

Meller *et al.* [1] went on to show how a mixture of poly[dA] and poly[dC] molecules can be separated based on their lifetimes ($t_D$'s). **Figure 14** shows the data they recorded, and how their computer program could assign 98% of the polymers as either poly[dA] or poly[dC], evidencing that this technology is useful for bulk sorting of polynucleotides. Moreover, they showed that the polymers behave as 2 independent populations, with the statistical properties of one not being affected by the presence of the other. They also concluded that incorporation of more statistical parameters from **Table 1** would lead to discrimination of a wider variety of polymers.

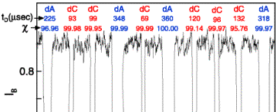
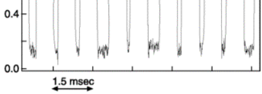


Figure 14 Representative current trace showing 10 events recorded from a mixture of equal molar concentrations of poly[dA] and poly[dC] molecules where assignment of an event to a polymer is done based on the statistics already recorded for the individual polynucleotides. This figure and legend were taken from [1, Figure 5].



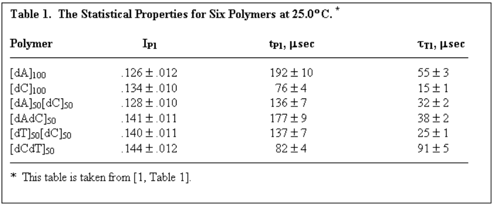
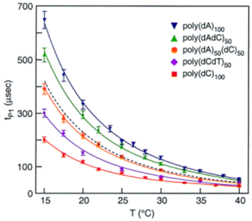


**Table 1. The Statistical Properties for Six Polymers at 25.0°C.** *

| Polymer | $I_{P1}$ | $t_{P1}$, µsec | $\tau_{T1}$, µsec |
|---|---|---|---|
| $[dA]_{100}$ | .126 ± .012 | 192 ± 10 | 55 ± 3 |
| $[dC]_{100}$ | .134 ± .010 | 76 ± 4 | 15 ± 1 |
| $[dA]_{50}[dC]_{50}$ | .128 ± .010 | 136 ± 7 | 32 ± 2 |
| $[dAdC]_{50}$ | .141 ± .011 | 177 ± 9 | 38 ± 2 |
| $[dT]_{50}[dC]_{50}$ | .140 ± .011 | 137 ± 7 | 25 ± 1 |
| $[dCdT]_{50}$ | .144 ± .012 | 82 ± 4 | 91 ± 5 |

* This table is taken from [1, Table 1].

Since the translocation process is closely related to the thermal energy of the reservoir (as LN [3] have shown), and is hence modeled as a diffusion process, the effect that temperature has on the statistical parameters is revealing. The poly[dA] and poly[dC] molecules were again used to see what effects temperature have on the event diagrams and on the statistical parameters $I_B$, $t_D$, and $t_T$. For both polymers, $t_{P1}$ and $t_{P2}$ decrease strongly with temperature, in the range from 15°C to 40°C. Other properties were specific to a given polymer.



While the increasing temperature causes the poly[dC] group 1 and 2 events to merge above 25°C, the poly[dA] molecules remain in two groups throughout the temperature range sampled. The relative number of events in group 2 decreases from 50% at 15°C to about 20-25% at 40°C, and these group 2 events become more dispersed at low temperature. Analysis with other polymers reveals a very beautiful scaling law, where the peak lifetime $t_{P1}$ (for group 1 events) scales like ~ $a/(T^2 + b)$, where T is in °C, a is a constant that is different for each polymer, and b is an additive constant. Thus, as can be seen in **Figure 15**, discrimination of polymers diminishes at high temperatures, and low temperatures are better suited to identification from an ensemble of ssDNA or RNA.

Figure 15 Dependence of the group 1 peak lifetimes on temperature. All the measurements were performed at 120 mV. This figure and legend were taken from [1, Figure 7].

## Experimental Background: Conclusions

The data derived from the temperature studies (Figure 15) is remarkable because it suggests an alternate hypothesis for the presence of the multiple peaks in the lifetime histogram (Figure 5). We have already gone through the theory of LN [3] which described in painstaking detail how to account for the data observed by Kasianowicz *et al.* [2] in terms of the pore potential and the 2 distinct 5'/3' orientations of the polymer. This *orientational* hypothesis for the presence of two peaks in the lifetime histogram was supported by the available data, and LN neglected the degrees of freedom of the polymer outside the pore when formulating their model, because of the strong pore potential.

If the degrees of freedom of the polymer outside the pore matter when considering how the lifetimes will be distributed, the theory of LN will break down to some extent, particularly the microscopic model. The macroscopic model, however, will still be valid, as it makes no assumptions on the details of the potential. The data from [1] suggest that the multiple lifetimes arise because of differing conformations of the polymer, rather than different orientations of the polymer. Specifically, since each polymer folds into a unique three-dimensional structure, and because the unwinding of this secondary structure is necessary before translocation occurs, it is possible that it is precisely the conformation of the given polynucleotide that determines its dynamics. There are four lines of evidence to support this *conformational* hypothesis.

First, all polymers that contain long purine stretches ($poly[dA]_{100}$) have $t_{P2}$ and $t_{T2}$ values that show a very strong temperature dependence. Unlike the uniform $T^{-2}$ scaling of $t_{P1}$ vs. T, the $t_{P2}$ vs. T plots show irregular features, and for lower T the value of $t_{P2}$ diverges to longer times. This is not observed in the $poly[dAdC]_{50}$ polymers, which cannot form purine-purine base stacking structures. If the secondary structure is what is limiting the traversal times, then as the temperature is lowered, the secondary structure would stabilize, and hence the unwinding of that structure would be more difficult. This leads directly to a divergence of the $t_{P2}$ lifetimes to larger values, as is observed here for $poly[A]_{100}$, but not observed for $poly[dAdC]_{50}$, because they are unable to form the pronounced base stacking of the poly[dA] polymers, and therefore as the temperature is lowered, no stabilization of secondary structure occurs.

Second, if we assume secondary structure denaturing is what is limiting translocation of the polymers, the dispersion of group 2 events, characterized by $t_{T2}$, is a direct reflection of that unfolding. For poly[dA] molecules, which have a large amount of base stacking, as the temperature is decreased, the dispersion also increases, indicating that it is indeed secondary structure unfolding that is crucial for translocation to occur.

Third, the number of events associated into groups 1 and 2 should depend on temperature, as is observed. If base stacking limits diffusion of the polymer, then the number of events in peak 2 (longer lifetime) should increase as the temperature decrease, since more folded polymers would traverse the pore slowly. This is observed for poly[dA]$_{100}$, which has 20% of events in group1 at 25°C and 45% in group 2 at 15°C.

Lastly, in polymers where only weak pyrimidine interactions are possible, like poly[dCdT]$_{50}$ and poly[dC]$_{50}$[dT]$_{50}$, only one group is observed, even at 15°C.

Despite the evidence for the *conformational* hypothesis, some data is left unaccounted in this model. First, the longest lifetime peaks are not input into the histogram or event diagrams, since these last for several seconds and are attributable to blockages by the secondary structure of the polymer (i.e. the polymer gets “stuck” in the pore) [2]. To clear these blockages, it is necessary to reduce or invert the current. This is seen with many different polymers, and must be accounted for by the secondary structure of the polymer interacting with the pore. Thus, if it is secondary structure causing these longest lifetime blockages, then the peaks shifts must reflect another phenomena, perhaps the pore-polymer interactions.

Second, polymers with weak secondary structure, like poly[U]$_{100}$ or poly[dC]$_{100}$, do display 2 peaks in the event diagrams for both the blockage current and the pore lifetimes [1,2], which is hard to explain if the *conformational* hypothesis is invoked, since the last thing that supported this theory is the converse of what this data suggests.

Third, if the lifetimes depend on the secondary structure, then there should not be a difference in the $I_B$’s for the different groups. The current that trickles through the pore when a polymer is inserted is related to the exact chemistry and interactions in the a-hemolysin pore and how the bases fit through this narrow .16 nm constriction. The part of the polymer in the pore (~10 nucleotides) dominates the electrical response of the system, regardless of the secondary structure outside the pore. Referring to Table 1, we see that all of the $I_{P1}$’s for the different polymers are the same (within experimental error). This supports the idea of the polymer backbone (not individual bases) interacting strongly with the pore, since these $I_{P1}$ values are the same regardless of base composition of the polymer, but do differ for the two groups (i.e. $I_{P1}$ is different from $I_{P2}$) for many polymers [1]. The difference between $I_{P1}$ and $I_{P2}$ for a single polymer is larger than the difference between the $I_{P1}$’s for different polymers. This result is easier to explain if you believe that each orientation of the polymer interacts in a distinct was with the pore, allowing a different amount of current to pass through the pore, rather than the structure of the polymer outside the pore playing a dominant role.

Lastly, the entire translocation procedure is dependent on temperature, and while the scaling of the various statistical parameters is unique, the actual fact that at higher temperatures the polymers move faster (with smaller $t_P$’s) is expected. LN had shown that if the parts of the polymer outside the chamber are neglected, the problem is one-dimensional, and the diffusion rate is dependent on the temperature in a very complicated way. Since the details are very complex in LN’s model, the $T^{-2}$ scaling may arise if a different pore potential is used, perhaps one that more accurately describes molecular scale interactions. The “black box” of the pore potential allows flexibility in adjusting it to incorporate new data, like that observed here [1].

In conclusion, the reason why the distribution of passage times has multiple peaks is left unresolved. Both the *orientational* and *conformational* hypotheses have data to support them, but direct determination of the polymer orientation as it moves through the pore has not been established, making the discrimination between these theories difficult. Indeed, the 5’/3’ asymmetry of the polynucleotides is “hidden” in these experiments, since the event diagrams cannot discern one orientation from the other. This obstacle needs to be overcome in order for a resolution of this debate to be achieved. Moreover, if the *orientational* hypothesis proves to be correct, there needs to be a way to distinguish which peak corresponds to which orientation. This is particularly important if this technology is to be used for sequencing ssDNA molecules. Lastly, a direct method for determining the stall force of the translocating polymer is valuable for support of both theory and experiment, and it can be a powerful tool to deconvolute what is happening as these polymers move from the *cis* to the *trans* chamber.

Specific Aim I: *Establishing the mechanism by which a polynucleotide traverses an a-hemolysin pore when driven by an electric field, utilizing a Brownian ratchet mechanism to uncover the hidden 5’/3’ polymer asymmetry, and thus resolving which translocation model is correct.*

Specific Aim I: Exploitation of the 5’/3’ Asymmetry

In order to determine the mechanism of translocation, we are proposing the exploitation of the base pairing potential of the ssDNA being driven through the pore in order to uncover the direction that the polymer is moving through the membrane. This can be accomplished in a direct way because the pore is narrow and prevents dsDNA from moving through [2]. By using this fact and the geometry of the chamber, it is possible to ratchet the ssDNA through the pore in an orientation dependent way, such that only the 5’ ® 3’ translocation direction is ratcheted, while the 3’ ® 5’ traversals are left unaffected.

A polymer with the sequence 5’-poly[dC]$_{10}$ [dA]$_{90}$-3’, which has been shown to display a characteristic event diagram via previous experiments, is used [1]. An ensemble of these polymers is placed into the *cis* chamber and a current is applied, forcing them through the a-hemolysin pore, where they then end up in the *trans* chamber. This is the standard experiment we have been describing. The blockage current histogram and event diagram can be computed from analysis of the current data [1,2].

The problem here is that the force generated by the electric potential is only proportional to the charge on the polymer, not to its orientation. Thus, when a single translocation event is observed, there is no way to tell in which of the 2 possible orientations that traversal occurred. This is why the asymmetry of the polymer is said to be "hidden" in these experiments, because the force we use to drive the polymer is not dependent on the orientation of the polymer. The pore potential is assumed to be asymmetric, as argued by LN [3], but because the details of this potential are not known, the prediction of the orientation based on the pore potential alone is vexing.

However, by exploiting the sequence asymmetry of the polymer (i.e. the fact that the sequence reads differently in each direction), as opposed to the backbone 5'/3' asymmetry just mentioned, it is possible to ratchet the polymer through the pore in an unsymmetrical way. This is accomplished by first placing a given amount of ssDNA complimentary to only the 5' end of the polymer being translocated on the *trans* side of the buffer chamber. As the polymer begins to move through the pore with the 5' end first (for the polymer here, with the poly[dC]$_{10}$ portion first), it will be base-pairing with the complimentary DNA placed in the *trans* compartment, which would be poly[dG]$_{10}$.

By using poly[dG]$_{10}$ molecules, we can bias the random walk in an orientation specific way. Namely, as the first 10 bases of the 5'-poly[dC]$_{10}$ [dA]$_{90}$-3' move through the pore, the *trans* chamber will see either a poly[dA] or a poly[dC] sequence. However, the poly[dG]$_{10}$ molecules will only base-pair to the poly[dC]$_{10}$ portion of the translocating polymer, and not to the poly[dA]$_{90}$ portion. If the polymer does go through with the 5' end (poly[dC]) first, this base-pairing will prevent the polymer from moving back through the pore into the *cis* chamber, because the pore is too small to accommodate dsDNA, hence driving translocation. If the poly[dA] portion moves through first, no base-pairing can occur, and the polymer will move through the pore unperturbed by the ratcheting mechanism.

Of course, the poly[dC]$_{10}$ sequence will still move through the pore eventually, even if the 3'-poly[dA] portion goes first. However, since the poly[dC]$_{10}$ part of the polynucleotide will only bind to the poly[dG]$_{10}$ molecules once the entire polymer has traversed the pore, no ratcheting (i.e. prevention of backward motion) can occur. Hence, this mechanism will only perturb the speed of the 5'® 3' orientation moving through the pore, leaving the 3'® 5' direction unaffected. This process is depicted in Figure 16.

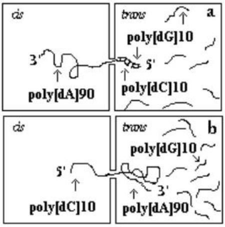


Figure 16 The ratchet setup for polymer translocation in each orientation. (a) Translocation in the 5'® 3' direction where the poly[dC]$_{10}$ portion moves through the pore first. The poly[dG]$_{10}$ molecules bind to the polymer's poly[dC] region, preventing backwards traversal, owing to the narrowness of the a-hemolysin channel. (b) Translocation in the 3' ® 5' direction where the poly[dA]$_{90}$ portion moves through first and no base pairing occurs. The ratcheting applies an effective force to the polymer, altering its average velocity and hence its translocation speed in only the 5' ® 3' traversal direction.

The mechanism of ratcheting depicted in Figure 16 can be extended to include multiple binding sites or larger binding sites, so experimental variations to maximize the perturbation are easily accomplished. All of the polymers may be synthesized in large quantities from a solid-state support, and the variations in sequence and size used are virtually unlimited [1,2]. The question then becomes what kind of data should we expect when we introduce such a ratchet? In order to answer this question, it is useful to review the mechanism of a Brownian ratchet and how it generates a force on the polymer being pulled through the pore.

## Specific Aim I: The Brownian Ratchet Model

The rectification of random thermal motion in order to do useful mechanical work is an important concept in biophysics. This idea arose because some biological processes seemed to generate forces, but do not involve molecular motors that hydrolyze nucleotides in a mechano-chemical cycle [[11]]. Such processes include filopodia extension and protein translocation through the endoplasmic reticulum, as well as actin-myosin motors. The ratcheting of polynucleotides through an a-hemolysin pore by the method described above is also an example of a Brownian ratchet, since nucleotide hydrolysis is absent from the process. Rather, the chemical energy gained from the binding of the poly[dG]$_{10}$ molecules on the *trans* side of the chamber is harnessed to bias the polymer walk in the *cis* to *trans* direction, generating a real force on the polynucleotide.

The Brownian ratchet model of Peskin *et al.* [11] in one-dimension follows a similar formulation to LN's macroscopic theory [3]. If a particle diffuses in one-dimension with some diffusion coefficient D, the root mean square displacement from the origin (where x = 0) is given by the random walk law: $<x^2>= 2Dt$, where t is the average time it takes for the particle to move from the origin to x. Thus, the mean time T it takes for the particle to diffuse from x = 0 to x = d is given by $T = d^2/2D$.

Now, if we subdivide the interval from x = 0 to x = L into N pieces of length d, we will have N = L/d. We can define the N boundaries by their coordinates: x = n x d, n = 1,2,3… N. Now, if each boundary is a "ratchet", that implies that particles can pass freely from the left to the right through the boundary, but once they cross the boundary, they cannot move back across it. Formally, the boundary is said to be absorbing from the left, but reflecting to the right. The physical mechanism for this boundary condition in our model (Figure 16) is the binding of the *trans* poly[dG]$_{10}$ molecules to the polymer's 5' poly[dC]$_{10}$ end. Notice that this definition of the ratchet precludes the 3'® 5' orientation from being ratcheted even though it has the poly[dC]$_{10}$ on its distal end, since once that last poly[dC]$_{10}$ passes through the pore, the polymer is fully translocated, and the mechanism for ratcheting ceases to exist (i.e. reflection is impossible).

Since the time to diffuse a length d is given by $T_d = d^2/2D$, the total time to diffuse a distance $L = N \times d$ is simply $T = N \times T_d = N \times d^2/2D = L\,(d/2D)$. If we re-arrange the last equality and solve for the velocity of the particle (L/T) being ratcheted at intervals d long, we see that:

$$v' = \frac{2D}{\delta} \qquad (9)$$

where ½' is the velocity of a perfect Brownian ratchet, and is thus an upper bound for the velocity of a particle driven in this manner. Notice that as the ratcheting interval d decreases, the velocity increases, because the frequency of smaller random thermal steps grows more rapidly then the step size shrinks. Factors such as the non-zero probability for the polymer to move back across the boundary and the presence of a force opposing the ratchet need to be considered. In our model, the pore potential or the secondary structure of the polymer is opposing the ratcheting and the electric field. Also, we have that d = 10*a* = 6 nm for our model.

The model of the Brownian ratchet for polymer translocation has been examined, and the main results are discussed here [11]. Brownian motion will cause the polymer to move through the pore and back again, unless biased in some way. In our model, the electric force and the ratchet pull the polymer through the pore, whereas the pore potential opposes this motion. The potential of the pore is necessary for the ratchet to produce any work, since it presents a load on the polymer.

The macroscopic model of ratcheting involves another diffusion equation. We consider a rod diffusing in one dimension along the x-axis with diffusion coefficient D. There is a force –*f* imparted on the end of the rod producing a drift velocity –(*f*/z) = -(D*f*/k$_B$T), where z is the frictional drag coefficient. The rod carries sites for ratcheting of length d that are equally spaced. It is assumed that the ratchet site boundaries are free to cross the origin (*trans* side of the pore) from left to right. In our poly[dC]$_{10}$[dA]$_{90}$ polymer, there is only one boundary point, the 3' most dC monomer. This boundary only exists when the molecule traverses the pore with the poly[C] portion (5' end) first, since if the poly[dA] portion (3' end) goes first, the ratchet boundary falls on the end of the polymer, which is equivalent to it falling on the very front of the polymer, where no poly[dG] can bind. Thus, no ratcheting force is produced for the 3' ® 5' direction, only in the 5' ® 3' direction.

For an imperfect ratchet, there is some probability p that the polymer will reflect from the ratchet boundary, and some probability (1 – p) for it to be absorbed back through the boundary. In our model, this probability is governed by the strength of the binding of the poly[dG]$_{10}$ to the poly[dC]$_{10}$ molecule, as will be discussed shortly. The Bernoulli variable X(t) indicates the position of the first boundary to the right of the origin. For the poly[dC]$_{10}$[dA]$_{90}$ molecule, this variable indicates the position of the 3' most (dC) monomer, supporting the idea that only one direction will be ratcheted. X(t) is always in the open interval (0, d). By this definition of X(t), we have a continuous random walk on a circular domain with a partially reflecting condition at the origin. This problem is formulated much like in the model of LN, and the details are omitted here [11, Appendix B].

The fact that the poly[dC]$_{10}$ portion binds to the poly[dG]$_{10}$ molecule with a given probability, governed by the T$_M$ for the complex, suggests we define p in terms of the binding constant for these oligonucleotides. If we assume that S$_0$ is the free state (ssDNA) and S$_1$ the bound state (dsDNA), then we have an equilibrium established between the two states:

$$S_0\,(ssDNA) \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} S_1\,(dsDNA) \qquad (10)$$

We assume only those molecules in the state S$_1$ are ratcheted (since they are the only double stranded DNA molecules). Therefore, sites in the state S$_0$ are free to pass through the origin in both directions, but sites in the state S$_1$ are reflected. If p is the probability of finding a ratchet in state S$_1$ (which is equivalent to the probability of the DNA being double as opposed to single stranded) is given as:

$$p = \frac{k_{on}}{(k_{on} + k_{off})} \qquad (11)$$

where k$_{on}$ and k$_{off}$ are the rate constants for the formation and denaturation of dsDNA between poly[dC]$_{10}$ and poly[dG]$_{10}$, respectively. The ratio of (1 – p)/p = k$_{off}$ / k$_{on}$ = K, which is called the dissociation constant of the complex.

With these assumptions, one can derive the velocity n' of a ratchet as well as the (stall) force *f'* it exerts on the polymer by solving the associated diffusion equations [11, Appendix B]. The results are given below.

$$v' = \frac{2D}{\delta}\left[\frac{\frac{1}{2}\omega^2}{\left(\frac{(e^{\omega}-1)}{1-K(e^{\omega}-1)} - \omega\right)}\right] \qquad (12a)$$

$$f' = \frac{k_B T}{\delta}\ln\left(1 + \frac{1}{K}\right) \qquad (12b)$$

Here w is the normalized work done by the ratchet against the load opposing translocation, w = *f*d/ k$_B$T. If we assume that the force (of the pore potential) is on the order of *f* » k$_B$T/*a* that makes w » 5, again an overestimate for our model [3]. The limitations of equation (12a) are that both k$_{on}$ and k$_{off}$ must be very fast and the ratchet must be inelastic [11].

The determination of the stall force and velocity for the ratcheted polymers can be facilitated by calculating the melting temperature T$_M$ of the dsDNA formed (namely, the complex between poly[dG]$_{10}$ and poly[dC]$_{10}$). It has been established that the T$_M$ can be calculated for PCR primers according to the following formula:

$$T_M = 81.5 + 16.6(\log_{10}[K^+]) + .41(\%G{+}C) – 675/\text{length} \qquad (13)$$

which implies that for our dsDNA in 1M [KCl], with 100% G+C content, and of length 10 bases, that T$_M$ = 55°C [[12]]. The T$_M$ reflects the temperature at which 50% of the DNA is single stranded. This duplex is very stable, owing to its large T$_M$.

We can also calculate the thermodynamic parameters for dsDNA formation from sequence consideration [12]. At 25°C in 1M NaCl at pH = 7.0, the DG = -13.9 kcal/mole, which implies that K = $10^{-16.81}$ and pK = 16.81 (where pK = -log$_{10}$K). We will use this value to compute the velocity and stall force for the Brownian ratchet, using equations (12a & 12b). We will assume that our diffusion coefficient does not deviate far from the value for small particles in water, $10^{-8}$ cm$^2$ sec$^{-1}$ and that w = 5, d = 6 nm, and K = $10^{-16.81}$. Substituting these values into equations (12a & 12b), we get that n' » .003 cm/s and *f'* » 25.5 pN. Both of these are expected to be larger than the ones actually observed, but show the magnitude of the velocity and the strength of the ratcheting force are comparable to the corresponding field parameters (n ~ .006 cm/s and *f* ~ 33 pN, respectively) and that this perturbation will be experimentally observable.

### Specific Aim I: Determining the Mechanism of Polymer Translocation

Now that we have shown that the ratchet can be used to move one orientation of the polymer through the pore quicker than the other, the data collected with the free polymer versus the ratcheted polymer is very illuminating. The experimental procedure is as follows.

First, the poly$[dC]_{10}$ $[dA]_{90}$ molecule is used in a series of experiments as outlined previously [1]. The event diagram, lifetime histogram, and current histogram can be computed for this polymer. The lifetime histogram would have the generic form given in Figure 17a. Here, we see the characteristic double peak and partitioning of events into groups 1 and 2. Now, we add poly$[dG]_{10}$ molecules to the *trans* chamber up to final concentration of 10 ¼M (20-fold excess when compared to the poly$[dC]_{10}$ $[dA]_{90}$ polynucleotides in the *cis* chamber). We then observed the event diagram and histograms again.

Because the ratcheting causes only the 5’ end to be “pulled’ through, the lifetime histogram will be perturbed in a very different way depending on which translocation hypothesis is correct. To see this, we should examine what Figure 17a (or Figure 5) tells us about the polymers moving through the pore.

Since there are 2 peaks present in the lifetime histogram, we can assume that the events under those peaks correspond to different physical situations, as LN have made clear [3]. Each hypothesis offers a different view of exactly what these different physical situations refer. In the *orientational* hypothesis, group 1 events would represent all the polymers in the ensemble that passed through the pore in the 5’ ’ 3’ direction (as we shall assume for the argument presented here), while the group 2 events would represent the polymers translocating in the 3’’ 5’ direction (of course this designation is arbitrary, since the asymmetry of the polymer is hidden in these experiments, but it will facilitate further discussions). Alternatively, if the *conformational* hypothesis were correct, the events in group 1 would represent polymers in the ensemble with less secondary structure, and hence quicker lifetimes. Group 2 would then include those polymers with more secondary structure. The distinction between these theories presents itself concisely once we introduce the ratcheting to uncover the hidden 5’/3’ asymmetry of the polymer.

If the *orientational* hypothesis is correct, then once the poly$[dG]_{10}$ molecules are added to the *trans* chamber, we should observe our group 1 events to shift to smaller average lifetimes. The reason for this is that the group 1 events correspond to 5’-first translocation of the polymer, and hence the ratcheting will aid in the traversal of this ensemble of polymers, as is depicted by the shift of $t_{P1}$ to $t_{p1'}$ in Figure 17b. The ratchet will not affect the events in group 2, and hence their peak $t_{P2}$ remains the same. By observing which of the events (either group 1 or 2) shift their lifetimes to lower values and knowing which orientation of the polymer gets ratcheted, it is easy to see which peak corresponds to which translocation direction. In our model, we would see the group 1 events shift because these events correspond to the 5’-first translocation, which is the direction that is aided by the ratcheting. The group 2 events represent 3’ first traversal, with no ratcheting and no mean lifetime shift. In a real experiment, we would obviously have to observe the peak shift and then imply which group corresponds to that orientation.

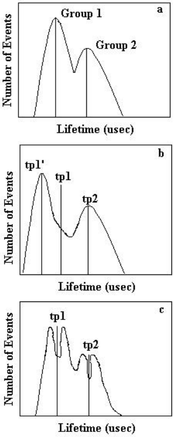


**Figure 17 (a) The predicted shape of the lifetime histogram for the poly$[dC]_{10}$ $[dA]_{90}$ molecules in the absence of a ratchet. Group 1 & 2 events can either represent the orientation of the polymer or the polymer’s secondary structure. For example, group 1 events could represent those polymers in the ensemble that move through the pore in the 5’ ’ 3’ direction (*orientational* hypothesis) or the events in group 1 could represent a weaker secondary structure of the polymer (*conformational* hypothesis). Group 2 events would thus represent the 3’ ’ 5’ orientation or a more pronounced secondary structure, depending on which theory is correct. (b) If the *orientational* hypothesis is correct, when the ratchet is applied, the 5’ ’ 3’ orientation would be ratcheted, and hence the lifetime peak $t_{P1}$ for this orientation would shift to a new value, $t_{P1'}$, as seen here. Notice that $t_{P2}$ was not disturbed. This data would indicate that the *orientational* hypothesis was correct, and also tell which group of events corresponded to which orientation. (c) If secondary structure were paramount, then each of the group 1 and 2 events would be composed of polymers that moved through the pore in either direction, with the different lifetimes reflecting varying amounts of secondary structure. Here, we could assume that the events in groups 1 & 2 would have an equal number of traversals in each orientation, hence the peak splitting.**

If we assume the *conformational* hypothesis is true, then we would observe our original lifetime histogram to be “split.” This is reminiscent of peak splitting in NMR spectroscopy. Essentially, the reason for this splitting is that each of the group 1 and 2 events would be composed of 50% 5’first and 50% 3’first traversals, and the ratchet only makes the 5’-first translocations faster. We can see this data represented in Figure 17c. We have the peaks splitting into 2 peaks, with one peak remaining at the same mean lifetime while the other shifts to a faster lifetime because of the ratcheting.

As we have seen, the event diagram of the poly$[dC]_{10}$ $[dA]_{90}$ polynucleotide will change under the ratcheting in a unique way depending on which hypothesis is correct. The observation of peak shifting would imply the *orientational* hypothesis is correct, while the observation of peak splitting would imply the *conformational* hypothesis is correct. In addition, if the orientation is what causes the multiple lifetime peaks, this method will allow the determination of which peak corresponds to which orientation, something very important if this technology is to be used for sequencing.

We can use the theory of LN to estimate the magnitude of the peak shifting or splitting we observe, since we have already determined the drift velocity of the polymer due to ratcheting (equation (12a)). Equation 7 tells us the value of the most probably traversal time, $t_{max}$:

$$t_{P1} \approx t_{\max} = \frac{L}{v}\left(1-5\frac{l_d}{L}+\frac{17}{2}\frac{l_d^2}{L^2}+32\frac{l_d^3}{L^3}+\ldots\right) \qquad (7)$$

in terms of the three parameter L, $l_d$, and ½, and $t_{P1}$ is the experimental value derived from the lifetime histogram (Figure 17a). L equals 100 for these polymers and $l_d$ is estimated to be 40 [3], so we have that the ratio $l_d/L = .4$, leaving ½ as the only free parameter. We have seen that the computation of ½ is difficult, since it is related in a non-trivial way to the applied field and pore potential. However, we can see that if we assume ½ is constant for the polymer without the ratchet, we can make the transform of ½ ’ ½ + ½’, where ½’ is the ratchet velocity that only affects the 5’’ 3’ traversal direction. Then we have that $t'_{max}$, the peak lifetime of the ratcheted polymers, is given by:

$$t_{p1'} \approx t'_{\max} = \frac{L}{v+v'}\left(1-5\frac{l_d}{L}+\frac{17}{2}\frac{l_d^2}{L^2}+32\frac{l_d^3}{L^3}+\ldots\right) \qquad (7a')$$

We can factor out a 1/½ term from equation (7a’) to get the ratio of $t'_{max}$ to $t_{max}$ in terms of ½ and ½’:

$$t_{p1'} \approx t'_{\max} = \frac{1}{1+\frac{v'}{v}}\frac{L}{v}\left(1-5\frac{l_d}{L}+\frac{17}{2}\frac{l_d^2}{L^2}+32\frac{l_d^3}{L^3}+\ldots\right) = \frac{1}{1+\frac{v}{v'}}t_{\max} \qquad (7b')$$

$$\frac{t'_{\max}}{t_{\max}} = \frac{1}{1+\frac{v'}{v}} = \left(1-\frac{v'}{v}+\frac{v'^2}{v^2}-\frac{v'^3}{v^3}+\ldots\right) \qquad (7c')$$

This expression (7c’) is only valid for ½’/ ½ < 1, which is satisfied here. We can calculate this ratio explicitly since we can sum the series (7c’) with ½’/ ½ = .003/.006 = ½ yielding that $t'_{max}/t_{max}$ = T.

Therefore, for our data set in the *orientational* hypothesis, we would see the value of $t_{P1}$ shift to 66% of its un-ratcheted value, while the $t_{P2}$ peak would not shift at all. For the *conformational* theory, we would see the split peak move to 66% of its initial value for both peaks. Notice that equation (7c’) gives a qualitatively correct answer, since we would expect the ratcheted peak to move to lower lifetimes, implying that the ratio $t'_{max}/t_{max} < 1$, as is observed here.

This model is ideal, and the results will likely be more convolved than those described. However, by being able to distinguish the 2 orientations of the polymer in this way will yield the necessary data to determine how the polymer moves through the pore, and what factors are most significant to that traversal, both of which are important when this technology is used for sorting and sequencing of polynucleotides.

# Specific Aim II: *To use direct application of tension via optical trapping to determine the stall force of the polymer moving through the a-hemolysin pore under the direction of both a field and a Brownian ratchet.*

### Specific Aim II: Optical Tweezers Substrate Construction

As has been pointed out by LN, the actual value of the force driving the polymer through the pore is very important, since it governs the velocity of the polymer, the regime of validity of the one-dimensional model, and the strength of the ratchet and pore potential [3]. It is thus advantageous to determine this force directly.

Here we propose the construction of a set of substrates to be used in single molecule experiments utilizing optical trapping. Trapping is used because it can determine the force exerted on or by a very small molecule with great precision by the application of radiation pressure. The basics of trapping are simple: the trapping force is governed by a Hooke’s Law relationship, where the distance from the center of the trap is proportional to the force on the particle being trapped, with the “spring’ constant or trap stiffness reflecting the light intensity gradient or laser power [[13]]. By measuring this displacement, one can compute the force on the bead at any given time. Alternatively, increasing the trap power at a fixed gradient (stiffness) also increases the force [20]. In these experiments, a single beam trap is used, and substrates suitable for trapping are constructed as follows.

It is advantageous to measure the force the pore and the field exert on the polymer in each orientation for the same polymer, as well as for the same orientation for different polymers. To accomplish this, we are proposing the construction of 2 polymers that have the ssDNA poly[dC]$_{10}$ [dA]$_{90}$ sequences protruding from a dsDNA tether, which is attached to a 1-¼m diameter polystyrene bead via a biotin-streptavidin linkage. The dsDNA tether would be a restriction endonuclease opened plasmid with a length of 5 kb (~ 1.7 ¼m) with a biotinated end for attachment to the bead. The actual substrate can be constructed utilizing the sticky-end of the plasmid for annealing to a double-stranded DNA molecule possessing the poly[dC]$_{10}$ [dA]$_{90}$ sequence. If the sticky-end (distal to the bead) of the plasmid has the sequence:

5’ATGGCA 3’
3’TA---- 5’

we would construct the following dsDNA oligonucleotide that would be complimentary to this sticky end, in analogy with how molecular cloning is done:

5’ ----GCGCGCGCGGC 3’
3’ CCGTCGCGCGCGCCG(dA)$_{90}$(dC)$_{10}$ 5’

where we see the above dsDNA molecule has three parts: a complimentary overlap to the plasmid stick-end, a 10-bp GC rich core that holds the strands together and provides a substrate for the ligase, and the ssDNA poly[dC]$_{10}$ [dA]$_{90}$ molecule that will traverse the pore. We can anneal and ligate this molecule to the plasmid by standard methods [[14]], producing our desired substrate, which is depicted in **Figure 18a**. Attachment to beads can be performed after ligation has occurred and has been shown to be correct by gel electrophoresis. A substrate similar to this one was recently used by *Wuite et al.* [[15]] to study T7 DNA polymerase activity via optical trapping tensiometry.

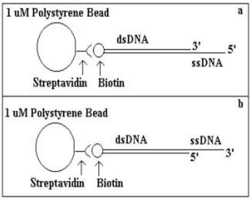


**Figure 18 (a)** The fully assembled molecule that can only traverse the pore in the 5’ ’ 3’ direction, which is with the poly[dC]$_{10}$ portion first. The bead is attached to the opposite end of the polymer via a biotin-streptavidin linkage, as shown. The bead is locked in an optical trap tensiometer, and the distance the bead moves from the trap center is used as a measure of the force the polymer is generating. The force required to halt the polymer is the desired stall force, *f*. **(b)** Same set-up as in **(a)** but with the 3’ end moving through first.

Notice that this molecule has a free ssDNA 5’end. When we use these molecules in our pore traversal experiments, we have again uncovered the hidden 5’/3’ asymmetry of the polymers, since these molecules can only pass through the pore in the 5’ ’ 3’ direction. We can also construct a molecule that can only pass through the pore in the 3’ ’ 5’ direction (**Figure 18b**), simply by changing our synthesized oligonucleotide to the following:

5’ ----GCGCGCGCGGC(dC)$_{10}$(dA)$_{90}$ 3’
3’ CCGTCGCGCGCGCCG 5’

and ligating as previously described. We can adjust the lengths of the base-pairing GC region to accommodate longer sequences that make ligation more efficient, and the plasmid tether used can also be varied depending on the experimental conditions. This latter variation is easily accomplished because most cloning vectors possess multiple cloning regions with the same restriction sites.

The method given above is very general, and using different ssDNA molecules is also easily accomplished. By placing different ssDNA portions on the templates for trapping, it is also possible to look at secondary structural effects on the stall force, as will be seen shortly.

## Specific Aim II: Stall Force Determination

In order to determine the stall force, an electric field is applied and a few substrates attached to beads are added to the *cis* chamber. In order to do trapping, a specially adapted buffer chamber needs to be constructed under the microscope, such that a current can be passed through the pore while it is under the scope. Once a current has been established, we can again record the current as a function of time, and observe the pore blockages. However, once a pore blockage has begun to occur, it will last a very long time, because the dsDNA portion of the substrate will not be able to pass through the pore. Since it is possible to see the bead, we will know a polymer has inserted in the pore when the bead appears tethered to the membrane. Once the molecule is “stuck” in the pore, we can trap the bead with infrared laser tweezers.

There are two ways to calculate the stall force of the polymer. In the first method, the strategy is to first trap the bead at very low power and to move the bead away from the pore, attempting to dislodge the polymer. If the trap falls off the bead, then the laser power is too low to overcome the force of the electric field and the polymer-pore interactions. Application of gradually stronger laser power and repetition will finally yield a lowest power, and hence force, that the trap must be set at in order for the bead (polymer) to be moved an infinite distance (i.e. a distance greater than the length of the substrate) away from the membrane. This is indeed the desired stall force of the polymer being driven. The use of an optical trap in this way to compute the stall force has been used previously to study the molecular motors kinesin and RNA polymerase [[16],[17],[18]], as well as for stretching DNA [[19]].

The problem with this methodology is that the dsDNA tether has some tertiary structure reflecting the bends of the DNA duplex. We would thus expect to observe a “first” pseudo-stall force, which is relatively weak since a 10 pN force is sufficient to make dsDNA fully extended, until the bead is moved out to a distance equal to the contour length of the dsDNA tether [15]. This force simply reflects the stretching out of the dsDNA tether, and does not reflect the pore potential. Above this pseudo-stall force, we would expect another minimum force that will pull the bead completely away from the membrane. This is precisely the stall force of the polymer with respect to the field and the pore potential.

The second method involves trapping the bead at high laser power and moving the bead away from the membrane to straighten out the dsDNA structure manually. As mentioned above, the tertiary bending structure of dsDNA mostly disappears when a tension greater than 10pN is applied [15]. This is easily accomplished with a stiff trap and a way to control the position of the laser (usually via an AOD) with great precision [13], leaving the ssDNA embedded the pore. The exact geometry of the system is known, and the distance between the bead and the membrane can be computed with great accuracy using video microscopy. Once the bead center is the contour length of the DNA away from the membrane, most of the bending of the dsDNA should be removed, and the ssDNA in the pore should then feel the force exerted by the trap, which is conveyed through the dsDNA. The dsDNA does not over-stretch until a force of over 60 pN is applied, making this method valid in the force window from 10-60 pN [15].

Utilizing the latter trapping arrangement, it is possible to see if the ssDNA is in the pore by moving the bead away from the pore at 10 nm intervals, pulling the ssDNA from the *trans* side to the *cis* side of the chamber. Once a single step is taken, the trap power is reduced to about 10 pN, and the bead should move out of the trap center. This movement of the bead would reflect the ssDNA being pushed through the pore by the field. Repetition of the above would suggest that the polymer is indeed in the pore, and not interacting with the membrane or pore in a non-specific way.

Once it has been established that the ssDNA is in the pore, we can again pull the ssDNA out of the pore 10 nm, and then maintain that trap power and bead position. We can then lower the trap power until the bead moves out of the center of the trap. By monitoring the bead position as a function of trap stiffness, it is possible to compute the force the trap needs to apply to keep the polymer from moving a certain distance through the pore. This again is precisely a measure of the stall force *f* of the polymer being driven through the pore, as LN have defined it [3]. A diagram for each orientation being pulled by the tweezers is given in **Figure 19**.

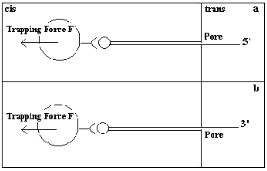


**Figure 19** The trap is designed to apply a force on the bead (F), which is transmitted through the dsDNA tether to the ssDNA in the pore. Notice that these polymers are stuck in the pore, owing to the dsDNA being too large to traverse the pore. The various directions are depicted in **(a) (**5' first orientation) and **(b) (**3'first orientation).

## Specific Aim II: Using the Stall Force to Determine the Method of Translocation

Now that we have seen how the substrates are constructed and how to measure the stall force for any of them, we can make some conclusions based on the data we expect to acquire. The same polymer (such as our ssDNA poly$[dC]_{10}$ $[dA]_{90}$) is made to go through the pore in only one orientation, and a stall force is computed for each of the orientations. If the stall force for each orientation were the same, that implies both "see" the same potential regardless of their translocation direction, which would refute the *conformational* hypothesis. Conversely, if the stall force for each orientation of the polymer were unique, that would support the idea that the pore interactions with the asymmetric backbone are dominant, supporting the theory of LN [3] as well as the *conformational* hypothesis. The orientation with the larger stall force would then correspond to the group 1 events in **Figure 17a**, and a direct comparison to the ratcheting data will confirm this conclusion. The absolute value of the stall force is very important for LN's theory, so that can be checked as well once *f* is determined experimentally.

By using a ssDNA poly$[dA]_{100}$ sequence versus a ssDNA poly$[dCdT]_{50}$ sequence, and determining their respective stall forces, it is again possible to see if the *orientational* hypothesis is correct. These polymers have very different secondary structure, and also display extremely different event diagrams [1]. Since the stall force reflects the polymer-pore interactions, these polymers would, if the *conformational* hypothesis were correct, have identical stall forces, and all the differences in their event diagrams would be derived from their secondary structure. By comparing the stall forces of these polymers in each orientation, it should then be possible to distinguish directly the validity of each of the traversal theories.

In addition to changing the sequence of the ssDNA going through the pore and its orientation, one can also change the ionic strength of the buffer, the applied voltage, and the concentration of the ssDNA. For each condition, a stall force can be computed, and the relationship between the pore potential and any of these parameters can be determined, yielding a wealth of information as to what contributes to the pore potential most significantly. More specifically, an analytical potential surface can be computed with these parameters as coordinates.

## Specific Aim II: The Stall Force of a Brownian Ratchet

We can combine our trapping experiments with the ratcheting experiments in a simple way, by adding anti-sense molecules to the *trans* chamber. For any polymer, since the pore traversal orientation is pre-determined by the geometry of the substrate, anti-sense oligonucleotides can be added to the *trans* chamber that will ratchet that polymer through the pore. The stall force can be determined with and with out the ratcheting perturbation, and the difference between the 2 values would approximate the stall force of the ratchet. Varying the ratcheting length d can be done by changing the length of the oligonucleotides on the *trans* side, providing a means to verify the equations given for the stall force (equation 12b). Also, by changing the base composition of the ratcheting units, and hence changing the K for duplex formation, the stall force can be computed as a function of K, again verifying equation (12b).

We have already seen that the average velocity of the ratchet depends in a very complicated way on the dimensionless load and dissociation constant (equation (12a)) and that it is possible to derive the velocity of the ratchet from translocation experiments alone. Here, we know the average velocity of the polymer without the ratchet, and we can recalculate this velocity after ratcheting takes place. The difference in velocities of the polymer, which would be directly derived from the lifetime histogram and its perturbation (Figure 17b), is the effective velocity due to the Brownian ratchet. Again, verification of equation (12a) by varying K and by actually determining *f* directly (as described above) is easily accomplished in this system. Therefore, nanopore polymer translocation, aside from being a possible way to sort and sequence DNA, is well suited to the study of Brownian ratchets, a ubiquitous mechanism in biology that lacks much direct experimental support.

## Conclusion

This proposal has shown the experimental data and theoretical models accumulated about driven polynucleotide traversal through narrow pores, and it has grappled with the conflicting theories that govern the process. Understanding the mechanism of traversal is a prerequisite for utilization of nanopore techniques for sorting and sequencing polynucleotides, since the data would lack any meaningful interpretation in its absence. The data and theory that support the competing *orientational* and *conformational* hypotheses have been explored in detail. We have shown how the 5'/3' asymmetry of the polymer is hidden in these translocation experiments, since the driving force is only dependent on the charge of the polymer, not on its orientation. To resolve the issue, we propose to resolve this hidden asymmetry by the introduction of a Brownian ratchet. We then demonstrated how the observed data would distinguish between the different theories of translocation. We have also proposed a series of direct optical trapping experiments to compliment the ratcheting data. Implications for the general mechanism of ratcheting have also been explored by these latter single-molecule techniques.

## References

---

[1] Meller, A., Nivon, L., Brandin, E., Golovchenko, J., & Branton, D. (2000) Rapid nanopore discrimination between single polynucleotide molecules. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1079-1084.

[2] Kasianowicz, J.J., Brandin, E., Branton, D., & Deamer, D.W. (1996) Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13770-13773.

[3] Lubensky, D.K & Nelson, D.R. (1999) Driven polymer translocation through a narrow pore. *Biophysical Journal* **77**, 1824-1838.

[4] Branton, D. & Golovchenko, J. (2000) http://mcb.harvard.edu/branton/nanopore.html

[5] Kasianowicz, J.J. & Bezrukov, S.M. (1995) Protonation dynamics of the alpha-toxin ion channel from spectral analysis of pH-dependent current fluctuations. *Biophysical Journal* **69**, 94-105.

[6] Montal, M. & Mueller, P. (1972) Formation of bimolecular membranes from lipid monolayers and a study of their electrical properties. *Proc. Natl. Acad. Sci. U.S.A.* **69**, 3561-3566.

[7] Song, L., Hobaugh, M.R., Shustak, C., Cheley, S., Bayley, H., & Gouaux, J.E. (1996) Structure of the *Staphylococcus aureus* a-hemolysin, a heptameric transmembrane pore. *Science* **274**, 1859-1866.

[8] de Gennes, P.G. (1979) Scaling Concepts in Polymer Physics. Cornell University Press, Ithaca, NY.

[9] Lee, N., and Obukhov, S. (1996) Diffusion of a polymer chain through a thin membrane. *J. Phys. II (France)* **6**, 195-204.

[10] Risken, H. (1984) The Fokker-Planck Equation. Springer-Verlag, Berlin.

[11] Peskin, C.S., Odell, G.M., and Oster, G.F. (1993) Cellular motions and thermal fluctuations: the brownian ratchet. *Biophysical Journal* **65**, 319-324.

[12] Breslauer, K.J., Frank, R., Blocker, H., & Marky, L.A. (1986) Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8746-8750. & http://www.biotechlab.nwu.edu/OligoCalc.html.

[13] Block, S.M. (1992) Making light work with optical tweezers. *Nature* **360**, 493-495.

[14] Sambrook, J., Fritsch, E.F., & Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, *Second Edition*. Cold Spring Harbor Press, NY.

[15] Wuite, G.J.L., Smith, S.B., Young, M., Keller, D., and Bustamante, C. (2000) Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. *Nature* **404**, 103-106.

[16] Wang, M.D., Schnitzer, M.J., Yin, H., Landick, R., Gelles, J., Block, S.M. (1998) Force and velocity measured for single molecules of RNA polymerase. *Science* **282**, 902-907**.**

[17] Svoboda, K., Schmidt, C.F., Schnapp, B.J., & Block, S.M. (1993) Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **365**, 721-727.

[18] Kuo, S.C. & Sheetz, M.P. (1993) Force of single kinesin molecules measured with optical tweezers. *Science* **260**, 32-34.

[19] Bustamante, C., Marko, J.F., Siggia, E.D., Smith, S.B. (1994) Entropic elasticity of lambda-phage DNA. *Science* **265**, 1599-1600.

[20] Ashkin, A., Schutze, K., Dviedic, J.M., Euteneuer, U., & Schliwa, M. (1990) Force generation in organelle transport measured *in vivo* by an infrared laser trap. *Nature* **348**, 346-348.