

Determination of Free Energy Profiles for the Translocation of Polynucleotides through α -Hemolysin Nanopores using Non-Equilibrium Molecular Dynamics Simulations

Hugh S. C. Martin,[†] Shantenu Jha,^{*,‡} Stefan Howorka,[†] and Peter V. Coveney^{*,†}

Centre for Computational Science and Department of Chemistry, University College London, 20 Gordon Street, London, United Kingdom, and Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana 70803

Received February 21, 2009

Abstract: The translocation of polynucleotides through transmembrane protein pores is a fundamental biological process with important technological and medical relevance. The translocation process is complex, and it is influenced by a range of factors including the diameter and inner surface of the pore, the secondary structure of the polymer, and the interactions between the polymer and protein. In this paper, we perform nonequilibrium constant velocity-steered molecular dynamics simulations of nucleic acid molecule translocation through the protein nanopore α -hemolysin and use Jarzynski's identity to determine the associated free energy profiles. With this approach we are able to explain the observed differences in experimental translocation time through the nanopore between polyadenosine and polydeoxycytidine. The translocation of polynucleotides and single nucleotides through α -hemolysin is investigated. These simulations are computationally intensive as they employ models with atomistic level resolution; in addition to their size, these systems are challenging to study due to the time scales of translocation of large asymmetric molecules. Our simulations provide insight into the role of the interactions between the nucleic acid molecules and the protein pore. Mutated protein pores provide confirmation of residue-specific interactions between nucleotides and the protein pore. By harnessing such molecular dynamics simulations, we gain new physicochemical insight into the translocation process.

1. Introduction

The translocation of nucleic acid strands through confined protein pores has biological relevance in instances such as transfer of antibiotic resistance genes between bacteria,^{1–3} phage infection,⁴ and the uptake of oligonucleotides into kidney tissue.⁵ Moreover, the passage of nucleic acids through pores is also of biotechnological and diagnostic relevance. For these applications, a single nanopore is inserted into a lipid bilayer, and individual negatively charged

nucleic acids are electrophoretically driven through the pore. The passages of strands lead to detectable fluctuations in the ionic pore current. Data from these single channel current recording (SCCR) experiments provide information on polymer length, orientation, and composition for polymers such as single stranded DNA and RNA.^{6–11} The capability of SCCR to reveal information on translocating DNA strands may provide an avenue for faster and cheaper genetic sequencing. Progress toward single base detection has been realized; for example, a modified nanopore has been constructed to yield a base-specific current response from isolated ribonucleoside and deoxyribonucleoside monophosphates,¹² while modified SCCR techniques provide measurable current responses to single mutations in a static nucleic

* Corresponding authors. E-mail: sjha@cct.lsu.edu (S.J.) and p.v.coveney@ucl.ac.uk (P.V.C.).

[†] University College London.

[‡] Louisiana State University.

acid polymer.^{13,14} Recently, oligonucleotides have been covalently modified with chemical tags as a step towards base-specific sensing in nanopore current recordings.¹⁵

Understanding the microscopic processes of nucleic acid translocation through nanopores is crucial in improving SCCR techniques and apparatus for sequencing DNA. Using molecular dynamics simulations of the translocation process, it is in general possible to retrieve kinetic and structural information that cannot be obtained solely through experiments. Experiments investigating the translocation of nucleic acid under the influence of a transmembrane potential suggest that the process typically takes hundreds of microseconds to tens of milliseconds.⁷ But accurately simulating biological processes and systems with atomistic resolution remains a challenge for many reasons, not least of which are the very substantial computational resources required. Even with state-of-the-art high-end computers, performing simulations with atomistic resolution for such large systems over the required time scales remains currently infeasible. Applying an artificially high potential difference in the simulations to induce faster translocation disrupts the lipid membrane; applying a high uniform electrostatic field to only the translocating atoms fails to translocate nucleic acid polymers through the protein nanopore.¹⁶ Thus, if these events are to be effectively investigated, novel approaches and better algorithms are required in order to bridge the gap between time scales over which the translocation events occur and those which are accessible using simple equilibrium simulations.

Nonequilibrium simulations offer one possible approach to increasing the time scales over which physical phenomena may be investigated with atomistic resolution. Various types of nonequilibrium simulations can be designed to effect translocation at higher speeds, allowing the full translocation process to be studied. Steered molecular dynamics¹⁷ is a type of nonequilibrium simulation that applies either a constant force to an atom or group of atoms or connects an atom or center of mass of a group of atoms via a harmonic spring (governed by a force or spring constant, k) to a constraint position, which is moved at a constant velocity. Constant velocity-steered molecular dynamics (cv-SMD) has the advantage of a well-defined wall clock and simulated time frame for a given translocation distance; cv-SMD is used in this paper to induce high-speed translocation.

This paper exploits nonequilibrium molecular dynamics simulations to gain microscopic insight into nucleic acid translocation through protein pores from the calculation of free energy profiles. It is desirable to obtain the free energy profile of the translocation process due to its central role in determining the thermodynamic properties of a system; it can also be used to ascertain kinetic properties. Free energy is an equilibrium property; in order to compute the free energy from a nonequilibrium process such as cv-SMD, we use Jarzynski's equality (JE).¹⁸ By obtaining and analyzing the free energy profiles of the translocation process, information is revealed about the energy barriers that a translocating nucleotide molecule experiences, and physical insight into the nature and origins of these barriers is gained.

In this paper, the protein nanopore we use is α -hemolysin (α HL), depicted in Figure 1. α HL is a heptameric protein

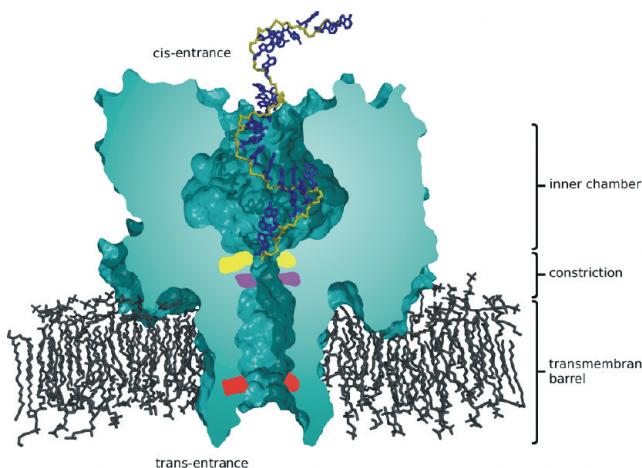


Figure 1. Figure representing the starting configuration of a 3' led A_{25} translocation simulation. The heptameric protein pore α HL (green) is inserted into a lipid bilayer (black). Features of the translocating molecule include the backbone of A_{25} (dark yellow) and the nucleic acid bases (blue). The *trans*-entrance is at the bottom of the pore; taking the *trans*-entrance of α HL as a reference point at 0 Å, other notable features include protein residue Leu-135 at 13 Å (red), Met-113 at 43 Å (pink), Lys-147 at 45 Å (light-yellow), and the *cis*-entrance at the top of the protein at 95 Å. The *cis*-entrance is 28 Å in diameter. The wide section of the pore running from the *cis*-entrance to residue Lys-147 is termed the inner chamber and is up to 46 Å wide. The constriction marked by residues Lys-147 and Met-113 is 14 Å wide, while the transmembrane barrel runs from the constriction to the *trans*-entrance and is around 20 Å wide. The *trans*-entrance is 24 Å wide. The C3' carbon atom of the 3' end nucleotide of A_{25} is aligned with the center of mass of the C_α atoms of protein residue 111, which lies at the mouth of the constriction, just above residue Lys-147. For the sake of clarity, water molecules and sodium and chloride ions are not displayed (they are found along the entire length of the pore).

pore that has been extensively studied in experiments and computer simulations.^{10–16,19–21} Of particular note is the lumen and the two entrances of the protein pore. Going from the top to the bottom of the pore Figure 1, the lumen at the *cis*-entrance is 28 Å in diameter, widens at the inner chamber, and then narrows to 13 Å at the inner constriction. The diameter of the transmembrane barrel increased to 20 Å at the *trans*-entrance Figure 1. The narrow inner constriction and also the transmembrane barrel is the bottleneck for the transport of matter through the pore. We explore the protein pore translocation of the nucleic acid strands polyadenosine, poly(A), and polydeoxycytidine, poly(dC), which are single strands of RNA and DNA, respectively. Poly(A) and poly(dC) molecules of 100–200 bases in length exhibit a 20-fold difference in translocation time through α HL in SCCR experiments.⁷ In order to gain insight into the experimentally observed difference, we simulate translocation of these two polynucleotides using cv-SMD. We also translocate single nucleotides A_1 and dC_1 to discern the relative contributions to the free energy profiles. We then investigate aspects of the translocation process using site-specific mutations within α HL.

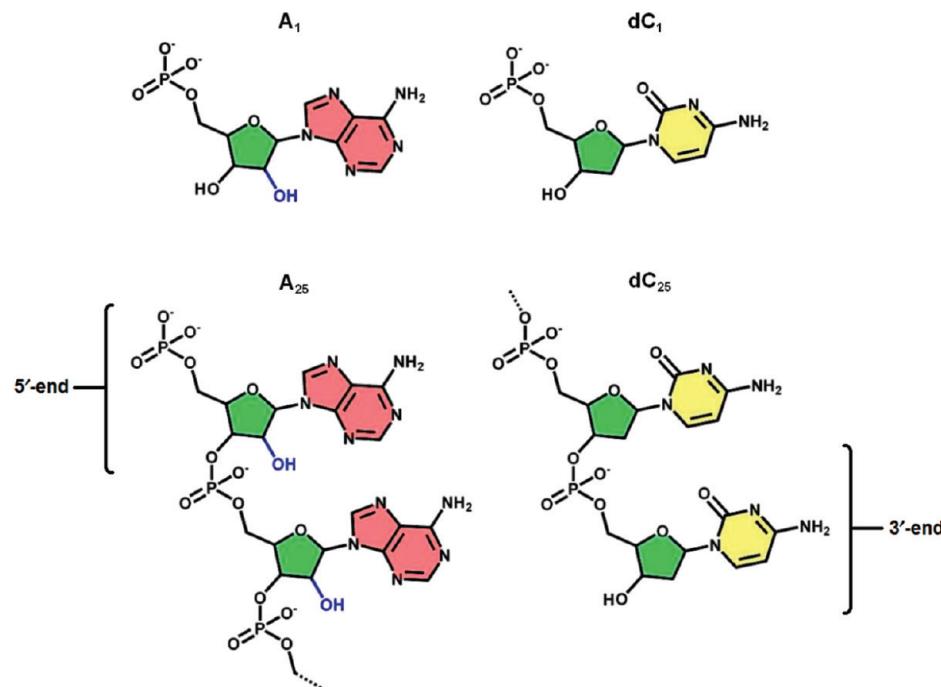


Figure 2. Molecular structures of A₂₅, A₁, dC₂₅, and dC₁. Structural features include the extra hydroxyl group associated with RNA (blue), the pentose base (green), the adenine base (light red) and the cytosine base (yellow). The 3' and 5' ends of the polynucleotides are also labeled for reference.

In the Methods section, we provide details of the model and techniques used to perform our simulations as well as some of the underlying theoretical formalism. In the Results and Discussion section, we present a detailed analysis of the results of our simulations; in the process, we discuss the optimization of various key parameters, visually inspect molecular configurations at equilibrium, and present the free energy profiles from polynucleotide and single nucleotide translocation through α HL in its wild and mutated forms. In the Conclusion section, we present our conclusions.

2. Methods

This section describes the computational construction of the transmembrane nanopore model and its associated components. We discuss Jarzynski's equality and explain our method for extracting free energy profiles from the simulations. We also describe the handling of errors for the simulations.

2.1. Model Construction. The α HL crystallographic structure coordinates were taken from Protein Data Bank (PDB) entry 7AHL. Atoms missing from the PDB file (residues dLys-30, gLys-30, aLys-75, dLys-240, fLys-283, and aArg-66) were reconstructed using psfgen, which is part of the NAMD2²² structure building module. α HL residues were mutated when required using the VMD²³ plug-in mutate. The protein was inserted into a patch of 150 Å × 150 Å pre-equilibrated and solvated 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer using the VMD plug-in membrane, aligned to the xy-plane plane. The center of mass of the hydrophobic belt of α HL (residues 118–126 and 132–142) was aligned with the center of mass of the lipid bilayer. Lipid and water molecules overlapping the protein were removed. The system was solvated in a

water box of pre-equilibrated TIP3P²⁴ water molecules using the VMD plug-in solvate with a 2.4 Å buffer distance between water molecules applied. The aqueous solution was set at 1 M NaCl using the VMD plug-in ionize. The dimensions of the lipid membrane and water box were chosen to minimize interactions between periodic images of the molecules due to the use of three-dimensional periodic boundary conditions. Figure 1 shows α HL inserted in a lipid membrane as in our models.

The protonation states chosen were consistent with SCCR recording pH range, typically between 7.5 to 8.5.^{6,11} The protonation states of the amino acid and phosphate groups were selected as follows: the amine groups of lysine and arginine residues were protonated, carrying a positive charge; interchain phosphate groups were unprotonated, carrying a single negative charge; and finally, terminal phosphate groups on the single nucleotide molecules were unprotonated, carrying a double negative charge.

The poly(A) and poly(dC) molecules (molecular structures depicted in Figure 2) were constructed using the AMBER module nucgen.²⁵ Nucleic acid polymers with 25 bases were constructed; the length was selected in order to maximize the influence of the nucleic acid polymer helix throughout the simulation, allowing comparisons to experimentally used lengths of 100–200 bases, while keeping the number of atoms in the model reasonably low. Double stranded conformations of the nucleic acid polymers were first built, the coordinates of which were derived from fiber diffraction studies.²⁶ A selected strand of the double helix was then removed from the PDB file, while the remaining single stranded nucleic acid polymer was charge neutralized with sodium counterions using the AMBER module xleap.²⁵ The nucleic acid polymer was then orientated as necessary using

the VMD module orient,²³ and the C3' carbon atom of the leading nucleotide was aligned with the center of the α carbon atoms (C_α) of protein residue 111. Water molecules that were overlapping the inserted polynucleotide were removed. After pulling the SMD atom from the constriction to the *trans*-entrance, the 25-base polynucleotide will have more than 10 bases remaining in the wide inner chamber. Our simulations with a solvated polynucleotide show that 10 bases are sufficient for it to form a helical structure in poly(A). An example of the starting position of the polynucleotides is shown in Figure 1.

Single nucleotide PDB files of adenosine monophosphate (A_1) and deoxycytidine monophosphate (dC_1) were also obtained (PDB identifiers AMP and DCM, respectively), and topology files were modified accordingly to produce accompanying PSF files. The molecular structures of these molecules are depicted in Figure 2.

The final models consisted of 328 000 and 262 000 atoms for the 25-base polynucleotide and single nucleotide models, respectively. The dimensions of the models were $145 \times 145 \times 154 \text{ \AA}$ for the 25-base nucleic acid polymer system and $145 \times 145 \times 121 \text{ \AA}$ for the single nucleotide system. The protein and nucleic acid molecules were at a minimum of 12 \AA from the edge of the simulation cell, ensuring a minimum distance of 24 \AA from the periodic images was maintained.

Translocation was chosen to be led by the 3' end (see Figure 2) of the polynucleotide for all the simulations in this paper. As shown by several groups, a difference is observed in the translocation times of nucleic acid polymers depending on whether they are threaded by the 3' or 5' end of the nucleotide chain.^{11,27,28} Nucleic acid bases projecting out along the phosphate backbone must flatten toward the backbone in order to pass through a tight constriction such as found in α HL. Nucleic acid bases tend to point toward the 5' end; therefore, they must bend through a larger angle than that of the 3' threaded polymers in order for the molecule to translocate through the pore. This results in an energetically less favorable process for 5' translocation than for 3' threaded instances.¹¹ Thus, 3' led translocation was selected to compare poly(A) and poly(dC). Given the nature of cv-SMD pulling, the stressing of bond angles, such as that noted for the deformation in the helical shape of the nucleic acid polymers, is likely to occur in the angle of the bases along the backbone. This influences the 5' led polymers more than the 3' ones as the bases could be made to invert orientation toward the 3' end by the force of translocation.

Translocation simulations in this paper were focused on SMD atom movement through the constriction and transmembrane barrel of α HL (see Figure 1). The dimensions of this part of the pore present the greatest resistance to a translocating molecule.¹⁹ The starting position of the 25-base polymer is shown in Figure 1; the SMD atom of the leading nucleotide is aligned with the center of the C_α atoms of the glutamic acid residue 111 at the top of the constriction. The dimensions of the inner chamber of α HL allow for the presence of a polynucleotide helix, whereas the smaller diameters of the constriction and transmembrane barrel do not allow it. Hence, at these constricting sections of the pore,

Table 1. The Translocation Molecules and Pore Types Simulated in This Work^a

system name	α HL type	nucleotide base	nucleotide units
A_1 -WT	Wild type	Adenine	1
A_{25} -WT	Wild type	Adenine	25
dC_1 -WT	Wild type	Deoxycytosine	1
dC_{25} -WT	Wild type	Deoxycytosine	25
dC_1 -Mut	Mutant	Deoxycytosine	1
dC_{25} -Mut	Mutant	Deoxycytosine	25

^a Wild type indicates α HL with no mutated residues, and mutant indicates α HL mutant L147M.

the helix of a translocating polynucleotide is required to unwind in order to pass through.

Constraints placed on the movement of protein amino acid residues are necessary in order to prevent the entire pore from being dragged downwards due to the cv-SMD applied force. To ensure that specific interactions between amino acids and nucleotides could take place, the side chains on the protein amino acids were left unconstrained to allow reorientation. Constraints of 0.5 N/m were, therefore, limited to the C_α atoms of the protein amino acid residues.

2.2. Simulation Details. Simulations were performed using the molecular dynamics simulation package NAMD version 2.6.²² NAMD is designed for scalable high-performance simulations on parallel machines. The CHARMM²⁹ force field was applied using all-hydrogen parameter files for CHARMM22 proteins and CHARMM27 lipids and nucleic acids. Periodic boundary conditions were applied, and the particle mesh Ewald method³⁰ was used to calculate electrostatic forces with a dielectric constant $\epsilon_0 = 1$ computed over $128 \times 128 \times 128 \text{ \AA}$ grids. Van der Waals energies were calculated with a smooth $10\text{--}12 \text{ \AA}$ cutoff. Constant temperature was maintained at 295 K using Langevin forces.³¹ Constant pressure was maintained by Nosé–Hoover Langevin piston pressure control³² at 1.01325 bar . A 2 fs time step was used with SHAKE turned on. So as to eliminate high-energy atom clashes, models were energy minimized for 2 ps until the gradient tolerance was below 20. The temperature was raised by 5 K every 50 fs for 3 ps to a final temperature of 295 K . Prior to insertion of the nucleic acid, the α HL model was equilibrated in the NpT ensemble for 1.3 ns with the backbone of the protein constrained and for a further 4 ns without constraints. After insertion of the nucleic acid molecules, the model was equilibrated in the NpT ensemble for 4 ns with the alpha carbons of the protein constrained and the ribose unit's C3' atom of the leading nucleotide of the nucleic acid molecule fixed. Translocation was limited to 1 ns per simulation due to resource constraints and to control binning errors. An overlap of 0.2 ns between sequential simulations was performed to enable removal of startup artifacts.

In order to sample the ensemble, multiple simulations of the nucleic acid molecule translocating past the same section of the pore were required. The initial configurations used to perform these translocation samples were obtained by capturing snapshots of the atomic positions and velocities, separated by 0.2 ns at equilibrium, with the SMD atom position fixed and the C_α protein atoms restrained.

Table 2. Parallel Scaling Performance and Efficiency Data^a

processors	queen bee		HPCx	
	wall clock/ time step (s)	efficiency	wall clock/ time step (s)	efficiency
32	0.29	1.00	0.47	1.00
64	0.16	0.91	0.25	0.95
128	0.11	0.67	0.13	0.90
256	0.06	0.65	0.07	0.83
512	0.04	0.48	0.04	0.74

^a The performance is reported in terms of the wall clock time taken in seconds per timestep and the efficiency relative to 32 processor simulations for a 328 000 atom model. Efficiency, E , at i number of processors is calculated as $E(N_i; N_{32}) = (t_{32}N_{32})/(t_iN_i)$, where N_i is the number of processors, N_{32} is the reference number of processors, t_{32} is the wall clock/timestep at 32 processors, and t_i is the wall clock/timestep using i number of processors. Data are presented for both the LONI/TeraGrid queen bee and HPCx machines.

Table 1 lists the key system configurations simulated. The molecular structures of the translocating molecules are depicted in Figure 2.

The simulations performed to generate the results we report here required more than 800 000 CPU hours. It is important to note that these were required on high-end tightly coupled machines, including those on HPCx (<http://www.hpcx.ac.uk/>), NGS (<http://www.grid-support.ac.uk/>), TeraGrid (<http://www.teragrid.org/>), and LONI (<http://www.loni.org/>). For the simulations reported here, we used between 256 and 512 processors per simulation. We list the parallel scaling performance and efficiency data for an atomistically detailed model containing 328 000 atoms in Table 2. See Supporting Information for an in-depth discussion of the computational resources used.

2.3. Jarzynski's Equality. In this work we have used Jarzynski's equality (JE)¹⁸ to extract free energies from the nonequilibrium simulations. JE equates the change in equilibrium free energy, ΔF , to the average ensemble work, W , by the following equality:

$$e^{-\beta\Delta F} = \langle e^{-\beta W} \rangle \quad (1)$$

Here, the angular brackets, $\langle \rangle$, indicate an ensemble average, and β is the inverse temperature, $1/k_B T$, where k_B is Boltzmann's constant. Thus, JE provides a means of equating the equilibrium free energy to the average work done in a nonequilibrium process.¹⁸

To derive Jarzynski's equality, consider a process changing a parameter, λ , of a system from equilibrium point 0 at a time zero to Λ at time τ . As a consequence of the second law of thermodynamics, the average work done, W , on the system cannot be smaller than the Helmholtz free energy (F) difference between the initial and final states of λ :

$$\Delta F = F(\lambda_t) - F(\lambda_0) \leq \langle W \rangle \quad (2)$$

Jarzynski showed that the free energy difference between the two states can be related to the work done to transition between them, W , by the equality shown in Equation 1. This is known as the Jarzynski's equality (JE).¹⁸ A rearrangement

of Equation 1 can be used to obtain the Helmholtz free energy difference in the form:

$$F_{\lambda(\tau)} - F_{\lambda(0)} = -\frac{1}{\beta} \log \langle \exp[-\beta W(\tau)] \rangle \quad (3)$$

The ensemble average of the exponential work ($\exp[-\beta W(\tau)]$) is taken over the ensemble of trajectories of the process. A problem in using Jarzynski's equality is that, due to the exponential, the free energy is dominated by small values of the work. In the absence of very extensive sampling, these small amounts of work are not included in the ensemble average of the exponential work. Therefore, instead of estimating the exponential of work directly, a cumulant expansion is often employed. The general form for the cumulant expansion of a logarithm of an ensemble average of an exponential (as in Equation 3) is

$$\log \langle e^x \rangle = \langle x \rangle + \frac{1}{2!} (\langle x^2 \rangle - \langle x \rangle^2) + \frac{1}{3!} (\langle x^3 \rangle - 3\langle x^2 \rangle \langle x \rangle + 2\langle x \rangle^3) + \dots \quad (4)$$

Inserting Equation 4 into Equation 3 gives

$$F_{\lambda(\tau)} - F_{\lambda(0)} = \langle W(\tau) \rangle + \frac{\beta}{2!} (\langle W(\tau)^2 \rangle - \langle W(\tau) \rangle^2) + \frac{\beta}{3!} (\langle W(\tau)^3 \rangle - 3\langle W(\tau)^2 \rangle \langle W(\tau) \rangle + 2\langle W(\tau) \rangle^3) + \dots \quad (5)$$

This series can be terminated at a chosen order to give an approximate formula for the Helmholtz free energy difference of the process. Using the cumulant expansion reduces the sampling error of estimating the ensemble average of the exponential directly. However, a truncation error is introduced by ignoring higher order terms. For this work, we initially established that a first-order cumulant expansion was valid (see Supporting Information for analysis comparing the first- and the second-order expansions).

The free energy as a function of a reaction coordinate is also known as the potential of mean force (PMF). For cv-SMD simulations, in order for the free energy profile to be expressed as a function of a well-defined reaction coordinate, a stiff spring is required to ensure that the SMD atom closely follows the constraint position.

2.4. Free Energy from cv-SMD Simulations. In this section we explain how the data output from cv-SMD simulations performed using NAMD are converted into free energy profiles using the JE. NAMD outputs the force (f) experienced by the harmonic spring connecting the SMD atom(s) and the constraint. The work can be calculated from the force according to the following force versus extension integral:

$$W = \sum_{i=1}^{N_S} f_i \Delta x_i \quad (6)$$

Here, Δx_i is the distance pulled at force f_i , and N_S is the number of intervals (bins) in the sum. In this paper, the work is calculated along the trajectory using a typical bin size $\Delta x = 0.25 \text{ \AA}$, enabling the resolution of well-defined peaks; as a noise reduction technique, we also use $\Delta x = 0.75 \text{ \AA}$.

The work can be plotted as an integral of the whole trajectory, which we refer to as a global plot, or each bin can be treated independently, representing the work done in that bin alone, which we refer to as a local plot. A global plot is, therefore, essentially a cumulative local plot.

Free energy profiles calculated from the average ensemble work are derived in principle by averaging over an infinite number of samples. In practice, they must be determined by a computationally feasible number of samples, while still effectively sampling the trajectory. The number of samples used in each free energy profile calculation is listed in the figures. As we show in sections 3.3, 3.4 and 3.5, the local representation of work and free energy shows peaks while the global representation gives a clearer comparison between the cumulative energy barriers to the translocation of two different systems, such as poly(A) and poly(dC). See Supporting Information for representations of the data obtained during the calculation process.

2.5. Error Handling. Just as in any experiment, multiple sources of error arise when using simulations to calculate free energies. In addition to the usual errors arising from numerical imprecision and approximations, there are errors due to the finite sampling used to compute ensemble averages and the choice of simulation parameters. Additional sources of error arise due to the finite approximations that are made in the expression used for computing the free energy as well as from the binning and smoothening of data. Errors can never be completely eliminated; for a fixed computational effort, at best, the errors can be controlled and their influence minimized. That is why, in this work, we have taken considerable care to discern the effect of the different types of errors, and we have planned our simulations accordingly so as to minimize their influence.

As described in the section 2.3, by increasing the force constant of the harmonic spring connecting the SMD atom and constraint position, the systematic errors due to deviations of the SMD atom from the reaction coordinate are reduced. However, a high spring constant also introduces statistical noise. An optimal balance between systematic and statistical error was found and is reported in section 3.1. Also, as discussed in section 2.3, the statistical error from applying JE, in its direct exponential form, is reduced by the use of the cumulant expansion formula, which introduces a systematic error due to truncation of the higher order terms. The systematic error can be controlled through this method to be smaller than the statistical error from exponential averaging, thus, providing a better overall estimate.

In determining the values of the work and energy averaged over multiple samples, there are two types of errors that must be accounted for. The first arises from the use of a coarse grained value for the work. The mean value is referred to as the bin value, and the error that arises from replacing the individual data points with the average taken over the bin, i.e., binning, is referred to as the bin error. The second error arises from taking the average of the bin value over multiple samples (sample-to-sample error); this error is statistical in nature and can be reduced with a greater number of samples. A challenge arises when the number of samples required to lower the statistical error is impractically large. Where both

error types arise, they are combined according the propagation of errors.

Additionally, as we explore in section 3.1, a high pulling speed introduces undesirable conformational changes. Thus, a balance must be found between performing a significant degree of translocation on a given resource and obtaining correct and accurate data.

3. Results and Discussion

In this section, we present the details of the optimization of parameters, free energy profiles, and other data from our MD simulations and the physical insight gained from them. In section 3.1 we describe the optimization of the cv-SMD spring constant and the chosen pulling speed. Section 3.2 presents the differences between the conformations of A_{25} and dC_{25} , relating the variations to experimentally and computationally observed trends. Section 3.3 compares the free energy profiles for A_{25} and dC_{25} polynucleotide translocation. In section 3.4, we present the free energy profiles for A_1 and dC_1 , comparing them to the polynucleotide free energy profiles. In section 3.5, we perform translocation through a mutated α HL pore, comparing the free energy profiles of the mutant and wild type pores.

3.1. Parameter Optimization. As described in section 2.1, there are many parameters that need to be carefully selected for effective cv-SMD simulations. For example, the choice of spring constant and pulling speed is especially sensitive to the size and conformational behavior of the pulled molecule, requiring explicit attention for the translocation of polymers. As demonstrated by Park and Khalili-Araghi,³³ a stiff spring connecting the SMD atom and the constraint position is required in order to use JE to extract the free energy correctly. The positions of the SMD atom and the constraint position may be significantly different with a low spring constant, causing greater deviation from the reaction coordinate. A stiff spring is essential; however, the statistical noise in the force output is higher for a larger force constant.

Figure 3a provides data which we use to find the optimized value of the spring constant for the translocation of dC_{25} in wild type α HL. At a pulling speed of 0.04 Å/ps, spring constants 50, 100, and 200 kcal/mol are shown to approximate a stiff spring as the SMD atom closely follows the constraint position. A value of 100 kcal/mol was chosen to ensure the spring was stiff without introducing excessive statistical noise. Figure 3b shows that a 100 kcal/mol spring constant exhibits similar behavior at pulling speeds of 0.04 and 0.004 Å/ps. The sampling rate for the SMD force was every 40 fs at 0.04 Å/ps and every 400 fs at 0.004 Å/ps.

The statistical and systematic errors tend to increase as the SMD atom moves further from its starting position. Lower spring constants allow the SMD atom to lag behind the constraint position as the translocation length increases. As shown in Figure 3a, the position of the SMD atom for 1.4, 7, and 50 kcal/mol spring constants shows an increase in lag between the SMD atom and the constraint position at the end of 4 Å of translocation. Translocations with 100 and 200 kcal/mol spring constants show a negligible increase in lag at the end of 4 Å. The lag controlled by segmenting the

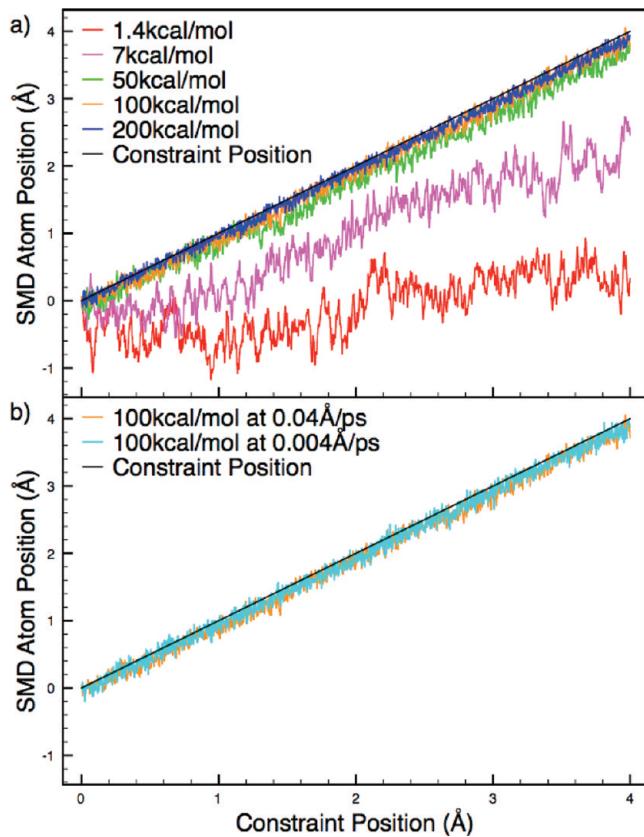


Figure 3. Analysis of the effect of spring constant values on the lag between the SMD atom and the constraint position. (a) The position of the SMD atom and the constraint position (\AA) plotted against the constraint position (\AA) for simulations with differing spring constants. The simulations were run at 0.04 \AA/ps using spring constants of 1.4, 7, 50, 100, and 200 kcal/mol. The harmonic spring must be approximately stiff in order for the free energy to be accurately calculated along a well-defined reaction coordinate. This plot shows that the spring could be safely considered stiff at 100 kcal/mol, which was used for all subsequent simulations. (b) The SMD atom and constraint positions using a 100 kcal/mol spring constant pulled at 0.04 or 0.004 \AA/ps show that the spring is approximately stiff at both speeds, with only a marginal increase in stiffness at the slower speed.

trajectory into 1 ns (4 \AA) sections and piecing them together to form the full translocation length. Therefore, with a spring constant at 100 kcal/mol, the errors are controlled by resetting the constraint position to the SMD atom every 4 \AA .

As shown by Wells et al.,¹⁶ if the system is too far from equilibrium, the translocation of a large deformable molecule in a cv-SMD simulation can result in undesirable conformational changes. This is due to translocation forces being applied to only a small part of a molecule. If the molecule moves too fast, then there is insufficient time for relaxation forces to return the structure to its equilibrium state. To reduce this effect, the speed of cv-SMD simulations for large molecules should be kept to a minimum. The challenge is to balance the desire for higher speeds with the need for correctness and accuracy. Figure 4 shows the change in separation between the first and second and the first and third nucleotides as the translocation of A_{25} proceeds at various speeds. At slower speeds, the backbone has more time to

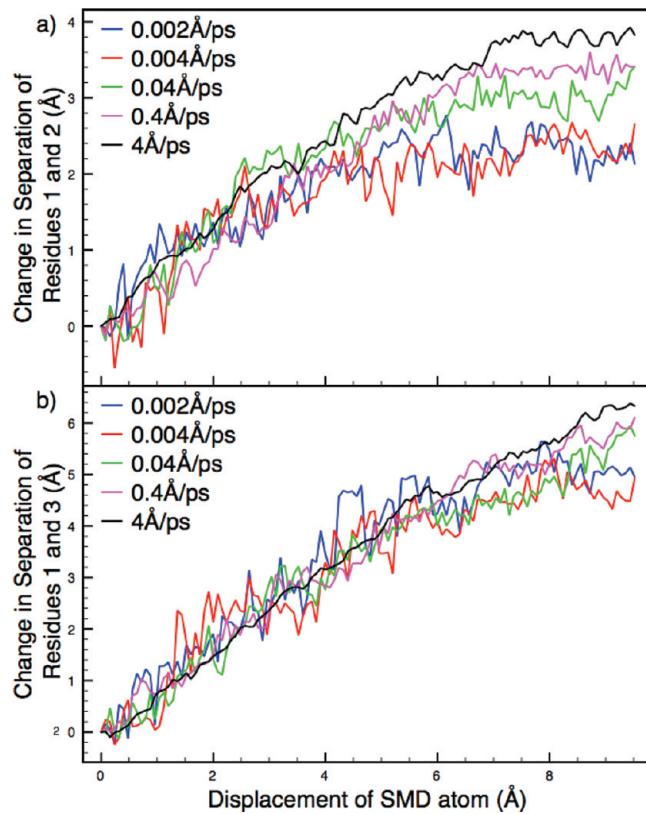


Figure 4. The effect of pulling speed on the relaxation of the conformation of translocating nucleic acid polymers. The plots show the change in z-axis separation between the pulled atom ($\text{C3}'$ carbon atom of the front-most nucleotide) and the $\text{C3}'$ atom of the (a) second and (b) third nucleotides during an A_{25} pulling simulation at different speeds. At pulling speeds where the system is close to equilibrium (not shown), the separation between the first and other nucleotides in the polymer will not tend to increase. Both plots show that, at slower speeds, nucleotides further along the polymer from the pulling atom have time to relax conformation and to move closer to the pulled atom, reducing the stretching of the polymer. At higher pulling speeds, the polymer stretches more, making it desirable to keep the speed to a minimum. At a pulling speed of 0.002 \AA/ps , the computational expense for our chosen translocation distances is too high; instead, 0.004 \AA/s was chosen for simulations in this paper.

relax from deviations from its normal conformations as a result of the pulling, thereby keeping nucleotides closer to each other during translocation. The increased separation is more pronounced as the nucleotide under consideration is further along the chain from the pulled atom, highlighting the need to keep the pulling speed to a minimum. Taking these factors into account, a speed of 0.004 \AA/ps was selected for the simulations reported in this paper, which is roughly a 1 000 times faster than translocation speeds found in experiment. While Akeson et al.⁷ hypothesized that the unwinding of a poly(A) helix gives rise to a long translocation time, it should be noted that at a translocation speed of 0.004 \AA/ps , it is difficult to preserve the helical structure of the polynucleotide in the nonconstricting inner chamber.

3.2. Base Stacking and Secondary Structure. In this section, we examine the conformations of A_{25} and dC_{25} at equilibrium. Our simulations are consistent with the predic-

tion that poly(A) exhibits a greater degree of base stacking than poly(dC) molecules,^{34–37} which is thought to contribute to the experimentally observed 20-fold higher translocation time of poly(A) compared to poly(dC).⁷ Poly(A) possesses a helical structure with a diameter that is greater than the constriction and transmembrane barrel of α HL. According to Akeson et al.,⁷ the helix is believed to unwind during translocation, posing an energy barrier. It is known that overlap of the aromatic π -orbitals in the nucleic acid bases causes them to stack, giving rise to the helical conformation of the backbone. While the CHARMM27 forcefield does not explicitly describe π -orbital overlap, it accounts for the base stacking that results from the interactions between aromatic groups. “[Base-stacking occurs due to the overlap of the aromatic π -orbitals, therefore the two aromatic rings on purine bases (such as adenine) give rise to a greater degree of stacking compared to that of the one aromatic ring found in pyrimidine bases (such as cytosine).] While the CHARMM27 forcefield does not explicitly describe π -orbital overlap, it accounts for the base stacking that results from the interactions between aromatic groups. Polynucleotides can adopt a helical shape as a result of base-stacking, thus poly(A) is expected to be helical in solution, whereas poly(dC) is expected to take a more random conformation. The predicted poly(A) helical structure has a diameter that is greater than the constriction and transmembrane barrel of α HL. According to Akeson et al.,⁷ the helix is believed to unwind during translocation, posing an energy barrier, while poly(dC) experiences a lower activation barrier to unwinding and translocates through constricting pores faster.”

At equilibrium, our systems containing poly(A) or poly(dC) exhibit a difference in mean conformation. As can be seen from Figure 5, which shows the average structure during a 2.6 ns simulation at equilibrium, A_{25} displays a helical shape and the bases align in a well stacked formation. On the other hand, dC_{25} displays a less defined helical backbone with reduced stacking of its bases.

As reported by Wells et al.,¹⁶ previous MD simulations indicate that base stacking can influence the translocation time of nucleic acid polymers. These authors found that a break in the base stacking passing through the constriction caused a temporary slowing down in translocation due to steric obstruction. Our own explorative simulations of stretched and unstretched polynucleotides indicate that base stacking can occur even in polynucleotides with a linear conformation. Therefore, energy barriers associated with the rupture of base stacking do not necessarily imply that helical unwinding has taken place. Base stacking and helical unwinding may, thus, represent two energetic barriers to translocation.

While the average structures show a difference in base stacking and conformations of A_{25} and dC_{25} , we will find in sections 3.3 and 3.4 that peaks in the local work and free energy profiles do not appear to correlate to stacked or unstacked sections along A_{25} or dC_{25} .

3.3. Simulations of Polynucleotide Translocation. We present the free energy profiles from the translocation of 25 base poly(A) and poly(dC) nucleotides (systems A_{25} –WT

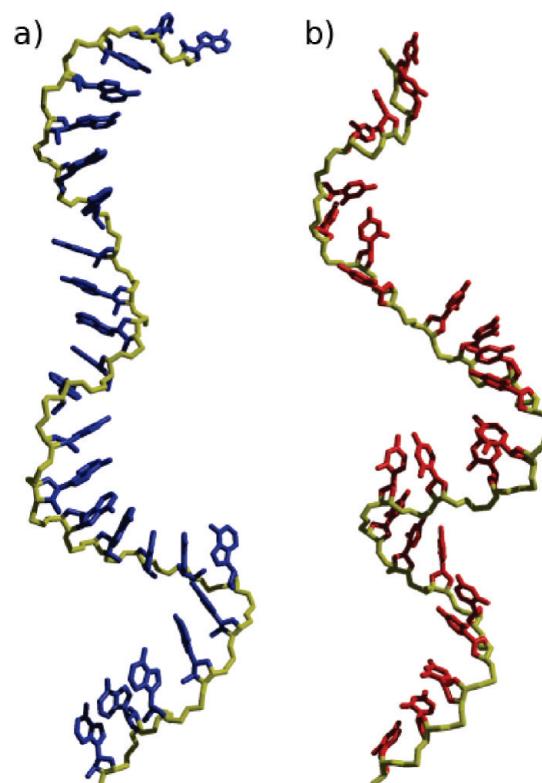


Figure 5. The average structures of (a) A_{25} and (b) dC_{25} over a 2.6 ns period at equilibrium in systems A_{25} –WT and dC_{25} –WT, respectively (see Table 1). The C3' atom of the front-most nucleotide is fixed in the center of the C_α atoms of the protein residue 111 at the top of the α HL constriction. The polynucleotides lie in both the inner chamber and outside the *cis*-entrance of α HL as shown in Figure 1. The A_{25} bases stack to a high degree, resulting in the more pronounced helical shape of its backbone than in dC_{25} . The dC_{25} structure does exhibit base stacking, though this is significantly less than that of A_{25} and is reflected in the more random configuration of the backbone. The increased stacking in A_{25} is due to the greater number of aromatic π -orbitals in adenine compared to cytosine, resulting in a more favorable π -orbital overlap.

and dC_{25} –WT (see Table 1)). The two polymers were shown, in the previous subsection 3.2, to exhibit different conformational behavior at equilibrium. We investigate the use of cv-SMD to produce free energy profiles from the translocation of these polymers in order to determine whether the profiles reflect experimentally determined longer translocation time for poly(A) compared to poly(dC).⁶ In addition, the simulations were conducted to gain physical insight into the differences.

The free energy profiles for a translocation length of 48 Å are shown in Figure 6 for A_{25} and dC_{25} with two samples performed for each trajectory system. Here, the SMD atom at the 3' end of the nucleic acid polymer was pulled from the top of the constriction to the bottom of the transmembrane barrel. The pore dimensions, as listed in Figure 1, indicate that the steric barriers to translocation occur mainly within this region. See Supporting Information for snapshots of the polynucleotide conformations during such a cv-SMD simulation.

Figure 6a shows the local free energy profiles of A_{25} and dC_{25} . Taking error bars into consideration, the two profiles

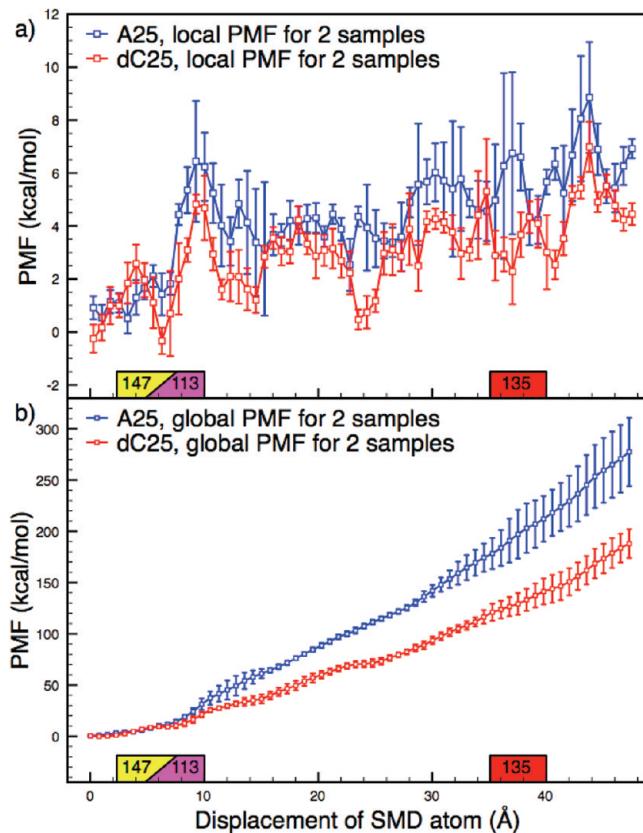


Figure 6. (a) Local free energy profiles of A_{25} and dC_{25} translocation from the top of the constriction to the bottom of the *trans*-entrance; each profile was derived from two samples. Labeled along the *x*-axis are protein residues Met-147, Lys-113, and Leu-135. The residue labels span 5 Å from the pulled atom to when first phosphate atom passes the labeled residue. The plot shows poor separation of A_{25} and dC_{25} when considering the error bars, though a general trend, where A_{25} has a higher free energy profile than dC_{25} is observed. An overall increase in the free energy profile is observed from left to right, which is expected as more nucleotides enter the confining dimensions of the constriction and transmembrane barrel. (b) Global free energy profile of A_{25} and dC_{25} translocation from the top of the constriction to the bottom of the *trans*-entrance; each profile was derived from two samples. The plot shows good discrimination of A_{25} and dC_{25} beyond the error bars.

are poorly separated when plotted in the local form, although A_{25} tends to exhibit a higher free energy than dC_{25} . Large peaks found in the local work profiles of individual samples are not as pronounced in the local free energy profiles consisting of multiple samples. The size of these peaks varies considerably from sample-to-sample and, other than the peak at 9 Å, their positions also vary.

The global free energy plots for the 48 Å translocation of A_{25} and dC_{25} are shown in Figure 6b. A_{25} displays a higher free energy profile than dC_{25} , with the discrimination well beyond the error bars. The separation between the means of the profiles continues to grow throughout the translocation process. The error bars are larger for A_{25} than for dC_{25} suggesting a higher sample-to-sample variation in the former case as the binning errors represent a small portion of the total errors. The higher free energy profile for A_{25} compared

to dC_{25} is in qualitative agreement with the longer experimental translocation times for A_{25} .⁷

Jensen et al. estimated the mean first passage time of glycerol translocation through the protein pore aquaglyceroporin GlpF from the reconstructed free energy difference of the translocation event.³⁸ This method relates the translocation time to the integral of the exponential free energy difference along the reaction coordinate, with constants representing the friction coefficient and inverse temperature. Using this method, we compute the relative translocation times of A_{25} and dC_{25} from the data in Figure 6b over the complete range. By assuming that the friction coefficients of A_{25} and dC_{25} translocation through the pore are similar, we find that the estimated translocation times have an approximate ratio of 25:1. Due to the exponential dependence of the translocation time on the free energy difference, reliable estimates of the translocation time without precise parameters such as friction coefficients are difficult. However, it is worth noting that our analysis is consistent with the 20:1 translocation time ratio of poly(A) and poly(dC) found in experiments.⁷

Figure 7 shows the free energy profiles of smaller segments of the trajectory, constructed from more samples to represent the ensemble average. Two parts of the pore are examined in this way, chosen to exhibit high energy barriers in the 48 Å translocation: the constriction, which is around residues Met-113 and Lys-147 (Figure 7a and b), and the transmembrane barrel, specifically around residue Leu-135 (Figure 7c and d). The diameter of the constriction is smallest at residues Lys-147 and Met-113, while the transmembrane barrel is smallest at leucine residue Leu-135; these residues are marked in Figure 1.

The free energy profiles in Figure 7 tend to show an increased energetic barrier to translocation of A_{25} relative to dC_{25} . Due to high variation in the single sample work profiles, the local free energy plots exhibit large overlapping error bars between A_{25} and dC_{25} (Figure 7a and c). The high variation in the single samples also leads to less pronounced peak shapes in the local free energy profiles. Exploration of peaks and local barriers to translocation is, therefore, best performed through examination of the local work profiles from single samples (see further below). In contrast to the local profiles, the global profiles show good separation between A_{25} and dC_{25} around protein residue Leu-135 (Figure 7d) and poor separation at the constriction (Figure 7b), although they point toward A_{25} exhibiting a higher free energy profile by the end of the trajectory. The error bars in the global free energy profiles again show greater sample-to-sample variation for A_{25} than for dC_{25} (Figure 7c and d).

We explored the molecular reason for the high peaks in the profiles of single samples. We visually inspected a molecular model of A_{25} at the inner pore constriction and observed that the phosphate groups of the polynucleotide strand interacts electrostatically with the positively charged amine group of residue Lys-147 (Figure 8). See Supporting Information for further investigation of the phosphate–lysine interaction.

The local single sample work profiles (not shown) contain higher peaks correlating to electrostatic phosphate–lysine

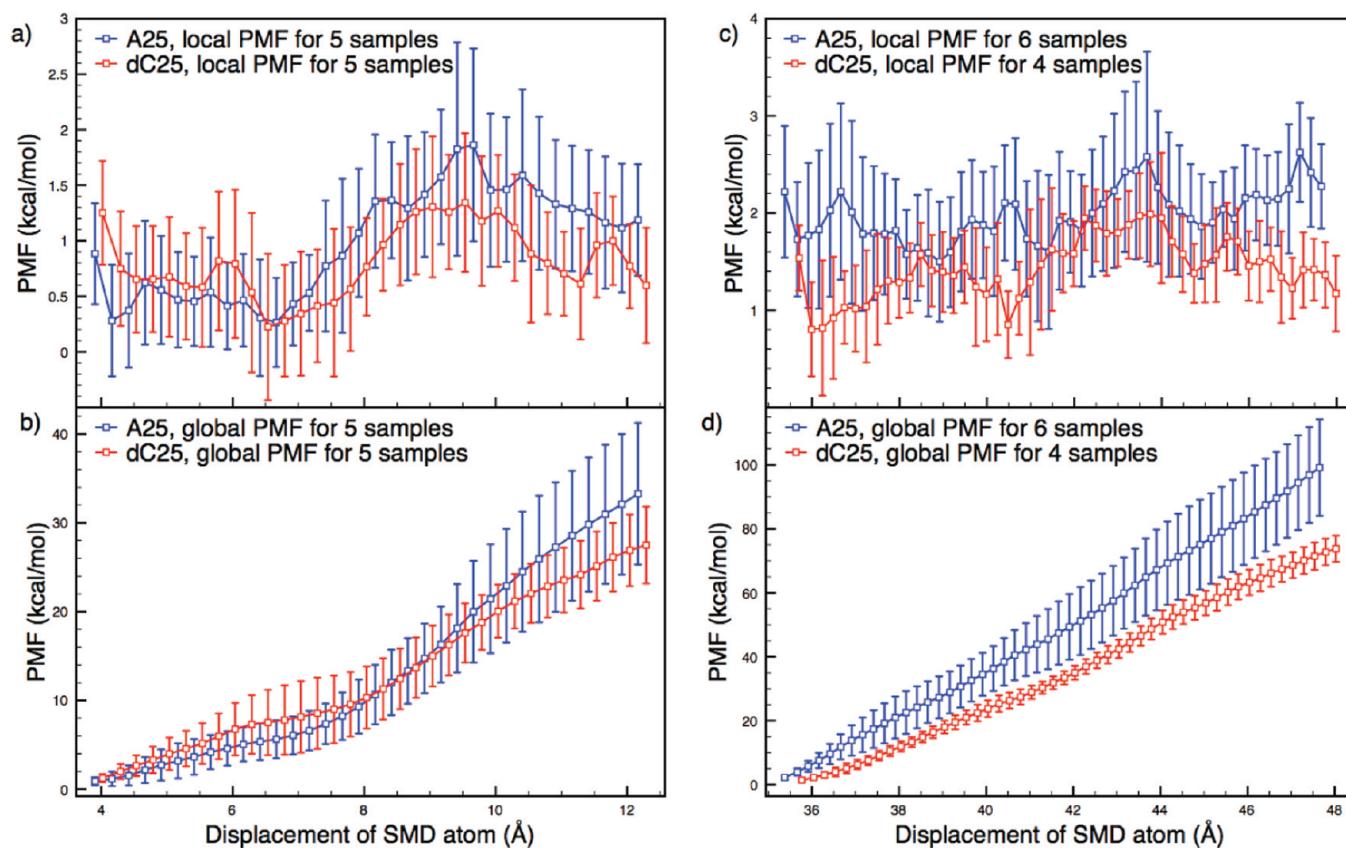


Figure 7. Local and global free energy profiles at the constriction and in the transmembrane barrel with increased sampling. (a) Local free energy profiles of A₂₅ and dC₂₅ translocating at the constriction. The statistics are based on 5 samples for each A₂₅ and dC₂₅. The separation is not beyond the error bars though both show a broad peak between 7 and 12 Å. (b) Global free energy profiles of A₂₅ and dC₂₅ translocating at the constriction. These show a tendency toward discrimination between A₂₅ and dC₂₅, albeit not beyond the error bars. (c) Local free energy profiles of A₂₅ and dC₂₅ translocating in the transmembrane barrel; there are 4 samples for dC₂₅ and 6 samples for A₂₅. The free energy of A₂₅ and dC₂₅ are the same to within errors. (d) Global free energy profiles of A₂₅ and dC₂₅ translocating in the transmembrane barrel. These show separation of A₂₅ and dC₂₅ beyond the error bars. The error bars in both global plots indicate the greater variation between samples of A₂₅ compared to those of dC₂₅, indicating a higher degree of sample-to-sample variation.

interactions for A₂₅ than for dC₂₅. This suggests a more accessible and/or a more electrostatically charged phosphate group on A₂₅. Base stacking or base size may play a role in making the A₂₅ more accessible by pushing the phosphate group closer to the sides of the pore. Figure 2 shows the molecular structures of the translocating molecules, 2' hydroxyl group shown on the pentose sugar of A₂₅ may contribute to the electrostatics and sterics of the phosphate group. This hydroxyl group is inherent to RNA polymers such as A₂₅, and exhibits an electrostatic/steric repulsion between the hydroxyl group and the 3' attached phosphate group. See Supporting Information for further investigation of the effect of the RNA–hydroxyl group on the polynucleotide phosphates.

3.4. Simulations of Single Nucleotide Translocation. In this section we present the free energy profiles from single nucleotide translocation in systems A₁–WT and dC₁–WT (see Table 1). The free energy profiles derived from polynucleotide translocation presented in the previous subsection represent combined energy barriers from the interactions of multiple nucleotides with the protein pore. To gain a clearer picture of the contributions from specific interactions, we performed translocation simulations on single nucleotides, and we present the resulting free energy profiles

here. By obtaining both the polymer and single nucleotide profiles, physical insight can be inferred with greater confidence.

As shown in Figure 9 for both A₁ and dC₁, the majority of activity occurs in the constriction region between 2 and 12 Å. For A₁, the increase in the free energy profile is broad, whereas for dC₁, there is a large, sharp increase between 8 and 12 Å. Inspection of the molecular models around the sharp peak in dC₁ shows that the increase clearly correlates with a strong electrostatic interaction occurring between the phosphate group of dC₁ and the amine group of Lys-147 (see Supporting Information for an animation of this interaction).

Comparing the molecular models of A₁ and dC₁ during equilibration and translocation, we find a clear difference in the local sodium ion positions. During equilibration of A₁ and dC₁ (not shown), sodium ions can be seen to approach both molecules. A sodium ion approaches the A₁ molecule, associates itself between the phosphate and the purine base, and remains there for the rest of the equilibration as well as for the start of the translocation simulations. For dC₁, sodium ions approach the molecule but do not appear to associate with the pyrimidine base, moving away before the end of the simulation. During translocation (not shown), the A₁

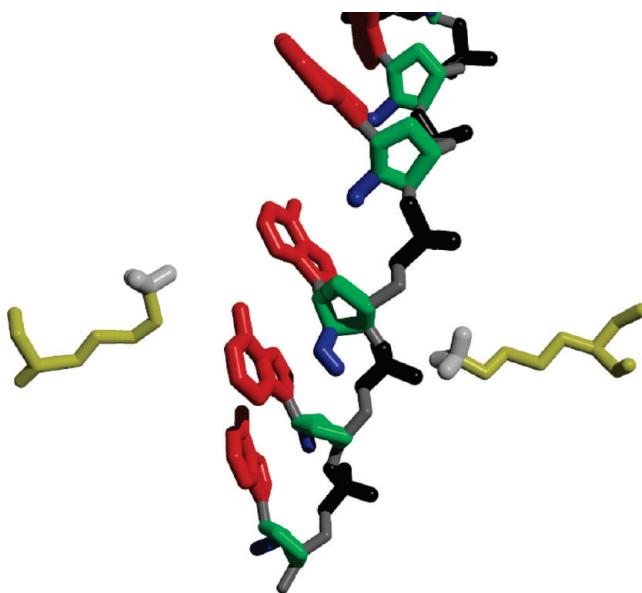


Figure 8. Molecular representation of the interaction between the nucleic acid phosphate groups of A_{25} and the amine group of residue Lys-147. The polynucleotide is shown in the middle with a lysine amino acid shown from either side of the constriction. The hydrogen atoms, other than those on the amine of lysine and the 2' hydroxyl groups, have been omitted for clarity. Notable features include the polynucleotide's adenine group (red), pentose base (green), and the hydroxyl group (blue), which are colored to match Figure 2. Also shown is the lysine residue (yellow), its amine group (white), and the polynucleotide's backbone phosphate groups (black). The lysine group nearest the phosphates of the polynucleotide shows a reorientation of its side chain toward the phosphate, while the lysine on the other side remains in a conformation that is observed during equilibration.

associated sodium ion stays close to the phosphate group until the phosphate group moves past residue Lys-147; at this point the sodium ion remains above the constriction, while A_1 translocates down the pore. By contrast, dC_1 does not possess an associated sodium ion at any point during the translocation. This sequence of events occurred in all instances of A_1 and dC_1 translocation that was performed, though further sampling of these models would be required in order to draw strong conclusions about the influence of the aromatic bases on the associated ions. Aromatic groups have a known tendency to form strong interactions with cations,³⁹ just as we observe for A_1 . Intuitively, one would expect the interacting sodium ion of A_1 to diminish the effective charge on the phosphate group and, thereby, reduce the interaction with the positively charged amine group of lysine. By contrast, it is anticipated that the uncoordinated phosphate of dC_1 will interact strongly with the lysine side chain, as is apparent in the sharp and repeatable peak in the energy profile in Figure 9a. See Supporting Information for an animation of the A_1 associated ion during translocation.

Figure 10 represents the global free energy profiles for the translocation of A_1 and dC_1 . The figure shows that the sample-to-sample variance is much greater for A_1 translocation than for dC_1 . The error bars in Figure 10 are roughly 4–5 times larger for A_1 .

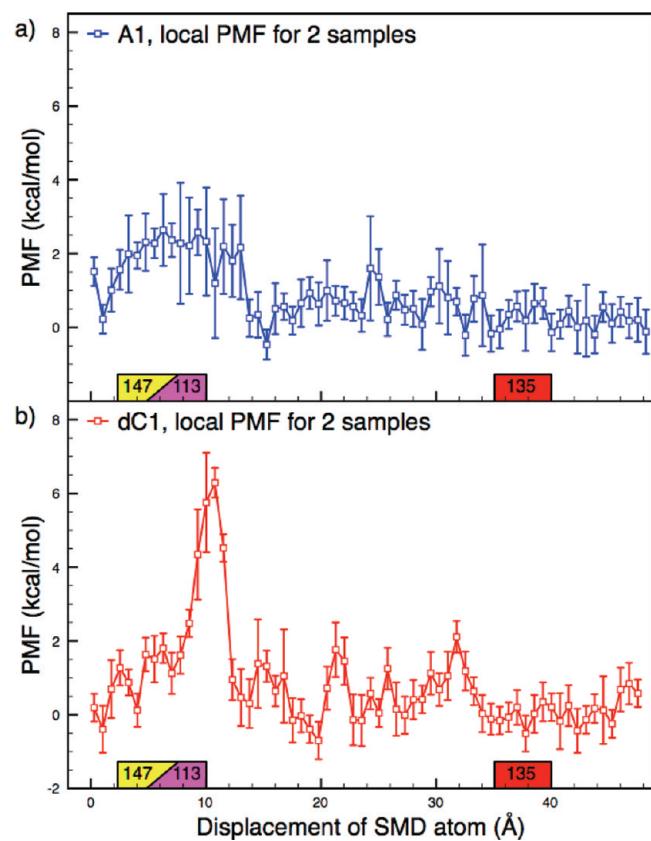


Figure 9. Local free energy profiles for the A_1 and dC_1 translocation, from the top of the constriction to the bottom of the transmembrane barrel, are derived from two samples. (a) Local free energy profile for A_1 . A broad peak is shown between 2 and 12 Å which corresponds to the constriction of α HL. The error bars in this region are large, indicating a high degree of sample-to-sample variation. There is little notable activity exhibited in the transmembrane barrel. (b) Local free energy profile for dC_1 . A sharp peak is observed at 10 Å, and an examination of the model at this point reveals a strong interaction occurring between the phosphate group of dC_1 and the amine group of residue Lys-147 at the top of the constriction. The error bars in this region are small, indicating a low degree of sample-to-sample variation. Examination of this region in A_1 shows some degree of interaction between the phosphate group of A_1 and the amine group of residue Lys-147, though it is considerably reduced compared to dC_1 .

By comparing the gradients of the global energy profiles from the single nucleotide and nucleic acid polymer simulations (Figures 10 and 6b), contributions to the free energy profile from the constriction and transmembrane barrel can be deduced. Of note is the high gradient beyond 30 Å in the polynucleotide global plots compared to considerably lower gradients at this stage in the single nucleotide simulations. The low profile gradients for single nucleotide translocation in the transmembrane barrel give an indication of the energy barriers experienced by the leading nucleic acid nucleotide for polynucleotide translocation in this region. Therefore, the high gradients for transmembrane polynucleotide translocation indicate that increased resistance to translocation is occurring further along the polymer as other nucleotides are passing through the constriction.

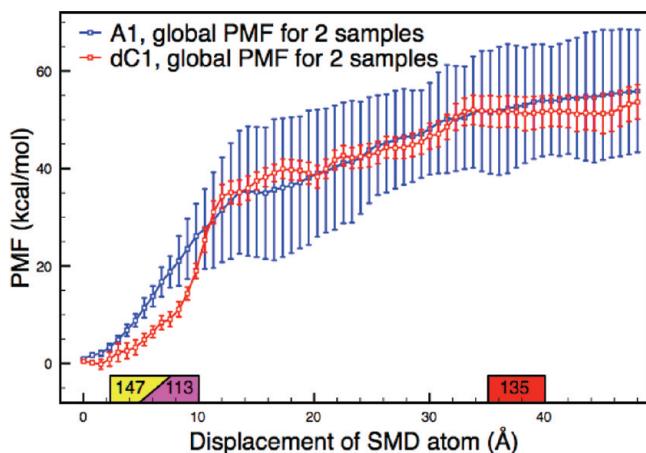


Figure 10. Global free energy profiles of the A_1 and dC_1 translocation, from the top of the constriction to the bottom of the transmembrane barrel, are derived from two samples. The gradients of the plots show that the majority of activity for single nucleotides occurs at the constriction. The plots are similar throughout the translocation process, showing separation outside of the error bars only at the start of the constriction. For the majority of the trajectory, the error bars are larger for A_1 than for dC_{25} by 4- to 5-fold, showing a larger sample-to-sample variance for A_1 , which is in line with findings for the polynucleotide simulations.

3.5. Effect of α -HL Mutation. In this section, we present the free energy profiles from polymer and single nucleotide translocation in the dC_1 –Mut and dC_{25} –Mut systems (see Table 1). From the single nucleotide and polynucleotide simulations in the unmutated α HL pore (see Table 1, wild type systems), the phosphate–lysine interaction appears to play a significant role. To confirm that this interaction gives rise to the observed peaks in the free energy profiles, we perform nucleotide translocation through an α HL pore where the residue giving rise to the interaction is mutated. Figure 11 compares the free energy profiles from translocation through the mutant to that of the wild type pore.

Figure 11 shows the local and global free energy profiles from dC_1 and dC_{25} translocation through wild type α HL and α HL where residue Lys-147 is mutated into methionine. Methionine is similar in size to lysine. Therefore, the mutation is not expected to alter the energetic barriers of the constriction beyond removal of the charged amine group on lysine. The free energy profiles show a profound difference between the wild type and mutant for dC_1 translocation in both the local and global forms of the profile (Figure 11a and b), confirming that the phosphate–lysine interaction, indeed, gives rise to the peaks in the wild type profiles. dC_{25} shows a similar global profile for the wild type and mutant (Figure 11c and d); while a major difference is not observed in the local profile, a change in shape comparable to that seen in dC_1 is observed. As shown in previous sections, dC_{25} exhibits the weakest phosphate–lysine interaction of the four nucleotides investigated in this paper, resulting in smaller peak sizes in the free energy profiles. This accounts for the reduced differentiation between the global plots of dC_{25} translocation in the mutant and wild-type pores. Initial attempts at A_1 and A_{25} translocation through α HL mutant L147 M did not show a significant

difference in the free energy profiles compared to that of the translocation through wild type α HL.

The degree to which the phosphate–lysine interaction plays a role in single channel current recording experiments is unknown. Factors indicating that the interaction could be significant include the observation that there are approximately 12 positively charged protein residues accessible to a molecule translocating in α HL and that current-driven translocation is about 1000 times slower than in the simulations reported here. Conversely, under conditions pertaining in experimental translocation experiments there exists a rapid ionic current flow, not present in our simulations, which may reduce the influence of the electrostatic interaction. In SCCR experiments, it has been found that increasing the number of positive charges inside the α HL pore increases the capturing frequency of polydeoxyribonucleotide translocations,²⁰ indicating an electrostatic relationship between positively charged residues and negatively charged translocating polynucleotides. Our simulations show a high sensitivity to local ion positions, giving rise to a greater sample-to-sample variance in the free energy profiles. Therefore, in order to draw comparisons between noncurrent driven translocation simulations and SCCR experiments, careful consideration of the role played by ions is required.

4. Conclusions

Jarzynski's equality (JE) in conjunction with constant velocity-steered molecular dynamics (cv-SMD) has been used to compute the free energy profiles from atomistic molecular dynamics simulations of nucleic acid translocation through the protein nanopore α -hemolysin. Obtaining the free energy profiles for such large atomistically detailed systems, including long deformable translocating molecules, is unprecedented.

We present the free energy profiles from the translocation of adenine and deoxycytosine based polynucleotides. The results are in qualitative agreement with experimental data, showing poly(A) experiences a greater resistance to translocation through α HL than poly(dC). By obtaining the translocation free energy profiles for the translocation of single nucleotides, we show that the majority of peaks representing significant barriers to translocation occur as a result of interactions with the inner constriction. The single nucleotide profiles exhibit stronger electrostatic interactions for deoxycytosine than for adenine, contrary to the polymer translocation profiles. Through inspection of the molecular positions and conformations and by plotting the radial distances of phosphate and lysine groups, we are able to show that a phosphate–lysine interaction between the protein and nucleic acid makes a dominant contribution to the translocation energy barrier. By mutating the interacting lysine group into a methionine (L147M), we are able to confirm that the phosphate–lysine interaction is, indeed, responsible for sharp peaks in the free energy profiles.

By examination of the molecular structures and investigation of the contributions of local solvated cations we can account for many of the patterns observed in the free energy profiles. We attribute the stronger phosphate–lysine interactions in systems containing poly(A) to a more electrostatically charged and sterically accessible phosphate group on the polyribonucleotide. This occurs due to an

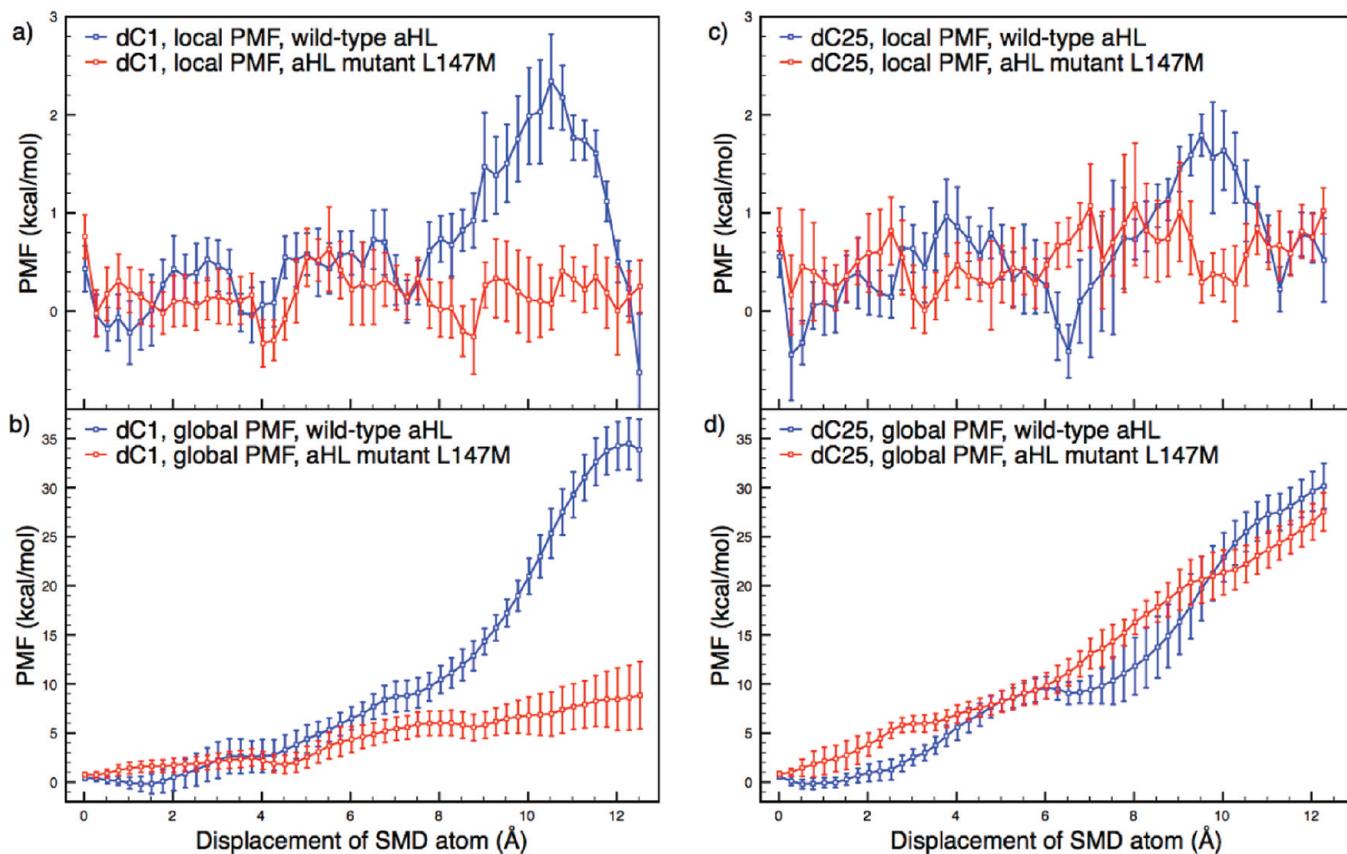


Figure 11. Free energy profiles for dC₁ and dC₂₅ translocation, through wild type α HL and mutant L147M, are derived from two samples. Residue Lys-147 was mutated due to an observed interaction with dC₁ that appeared to cause a strong peak in the local free energy profile. (a) Local free energy profiles for dC₁ translocation through the constriction of wild type α HL and mutant L147M. The profiles show that, with the lysine residue mutated, the interaction does not occur, and the peak is not present in the free energy profile. (b) Global free energy profiles for dC₁ translocation through the constriction of wild type α HL and mutant L147M; the difference between the two models is represented well in this global form. (c) Local free energy profiles for dC₂₅ translocation through the constriction of wild type α HL and mutant L147M. The local profiles do not manifest a good separation between the wild type and mutated pore but do show different shapes. (d) Global free energy profiles for dC₂₅ translocation through the constriction of wild type α HL and mutant L147M. Here the global profiles do not show a discernible difference between the two pores. The peak present between 8 and 12 Å in for dC₁-WT is also seen for dC₂₅-WT at the same position, though somewhat less pronounced. Simulations of the translocation of A₁ and A₂₅ through the α HL mutant L147 M did not show a significant difference in the free energy profiles compared to translocation through the wild type pore.

energetically unfavorable interaction between the phosphate group at the 3' carbon and the neighboring 2' hydroxyl group. The hydroxyl group is not present in polydeoxyribonucleotides such as poly(dC). The higher sample-to-sample variation we observe in the free energy profiles of poly(A) translocation is due to variation in the local cation position combined with the enhanced sensitivity of poly(A) relative to poly(dC) due its greater electrostatic charge.

This relationship between the phosphate and hydroxyl group does not occur in the nucleotide monomers due to the much larger distance between the phosphate at carbon 5' and the hydroxyl group at carbon 2'. For the translocation of adenine and deoxycytosine monophosphates, we find that the combination of the negatively charged phosphate group and the more aromatic adenine base causes a strong association with sodium ions, leading to a reduced phosphate–lysine interaction in adenine monophosphate. For deoxycytosine monophosphate translocation, the less aromatic cytosine base does not associate with local sodium ions, leading to an unbound phosphate group

and a stronger phosphate–lysine interaction. By providing evidence of the importance of phosphate–lysine interactions, our simulations set the stage for experimental confirmation.

In the future, it would be interesting to take advantage of the advent of even more substantial – petascale – computational resources⁴⁰ to examine slower translocation speeds and higher sampling for the free energy profiles. A comparison of poly(A) to poly(dA) would help to clarify the effect of the RNA–hydroxyl group on the phosphate–lysine interaction. Such data would illuminate the contribution of steric friction on translocation free energy profiles under these conditions. Single channel current recording data on an L147 M α HL mutant would indicate the influence of the phosphate–lysine interaction in experiments.

Overall, our study has yielded free energy profiles for translocation using constant velocity-steered molecular dynamics at speeds considerably higher than experiment, but nonetheless provides detailed explanations for the differences observed between selected nucleotide molecules. The findings furnish novel explanations for experimentally observed

translocation time differences. While our simulations are not performed under ideal conditions for close comparison with experimental results, they represent the state-of-the-art achievable with existing high-end computing resources and point the way to future enhancements that will become possible at the petascale and beyond.

Acknowledgment. We thank Daniel Keedy for his support with force field parametrization of the single nucleotide models. This work was supported by the U.K. Engineering and Physical Science Research Council (EPSRC) through the RealityGrid project (GR/R67699, EP/C536452/1), which also provided access to HPCx (<http://www.hpcx.ac.uk/>). TeraGrid (<http://www.teragrid.org/>) access was made possible through the National Science Foundation MRAC grants TG-DMR070014N and TG-DMR070013N and DAC grant TG-ASC070019T; significant computational resources on LONI were also made available via a grant to S.J. We are grateful to EPSRC and UCL Department of Chemistry for funding H.S.C.M.'s Ph.D. studentship.

Supporting Information Available: Discussions of computational resources utilized, Jarzynski's equality, snapshots of polynucleotide conformations during translocation, further investigation of the phosphate–lysine interaction, investigation of Crooks fluctuation theorem alongside JE, animations related to single nucleotide translocation, and a listing of significant residues in the α HL pore. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) Mullen, C.; Kilstup, M.; Blaese, R. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 33–37.
- (2) Davies, J. *Science* **1994**, *264*, 375–382.
- (3) Koonin, E.; Makarova, K.; Aravind, L. *Annu. Rev. Microbiol.* **2001**, *55*, 709–742.
- (4) Dreiseikelmann, B. *Microbiol. Mol. Biol. Rev.* **1994**, *58*, 293–316.
- (5) Hanss, B.; Leal-Pinto, E.; Bruggeman, L.; Copeland, T.; Klotman, P. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1921–1926.
- (6) Kasianowicz, J.; Brandin, E.; Branton, D.; Deamer, D. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13770–13773.
- (7) Akeson, M.; Branton, D.; Kasianowicz, J.; Brandin, E.; Deamer, D. *Biophys. J.* **1999**, *77*, 3227–3233.
- (8) Meller, A.; Nivon, L.; Brandin, E.; Golovchenko, J.; Branton, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1079–1084.
- (9) Meller, A.; Nivon, L.; Branton, D. *Phys. Rev. Lett.* **2001**, *86*, 3435–3438.
- (10) Butler, T.; Gundlach, J.; Troll, M. *Biophys. J.* **2006**, *90*, 190–199.
- (11) Mathe, J.; Aksimentiev, A.; Nelson, D.; Schulten, K.; Meller, A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 12377–12382.
- (12) Clarke, J.; Wu, H. C.; Jayasinghe, L.; Patel, A.; Reid, S.; Bayley, H. *Nature Nanotechnol.* **2009**, *4*, 265–270.
- (13) Howorka, S.; Cheley, S.; Bayley, H. *Nat. Biotechnol.* **2001**, *19*, 636–639.
- (14) Ashkenasy, N.; Sánchez-Quesada, J.; Ghadiri, M.; Bayley, H. *Angew. Chem., Int. Ed.* **2005**, *44*, 1401–1404.
- (15) Mitchell, N.; Howorka, S. *Angew. Chem., Int. Ed.* **2008**, *47*, 5565–5568.
- (16) Wells, D.; Abramkina, A. *J. Chem. Phys.* **2007**, *127*, 125101–1–125101–10, A, A .
- (17) Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (18) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.
- (19) Aksimentiev, A.; Schulten, K. *Biophys. J.* **2005**, *88*, 3745–3761.
- (20) Maglia, G.; Restrepo, M.; Mikhailova, E.; Bayley, H. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 19720–19725.
- (21) Hornblower, B.; Coombs, A.; Whitaker, R.; Kolomeisky, A.; Picone, S.; Meller, A.; Akeson, M. *Nat. Methods* **2007**, *315*–317.
- (22) Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283–312.
- (23) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (24) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926.
- (25) Pearlman, D.; Case, D.; Caldwell, J.; Ross, W.; Cheatham III, T.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1–41.
- (26) Arnott, S.; Smith, P.; Chandrasekharan, R. *CRC Handbook of Biochemistry and Molecular Biology: Nucleic Acids*; CRC Press: Cleveland, OH, 1972.
- (27) Wanunu, M.; Chakrabarti, B.; Mathé, J.; Nelson, D.; Meller, A. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2008**, *77*, 31904.
- (28) Wiggin, M.; Tropini, C.; Tabard-Cossa, V.; Jetha, N.; Marziali, A. *Biophys. J.* **2008**, *95*, 5317.
- (29) Feller, S.; MacKerell, A. *J. Phys. Chem. B* **2000**, *104*, 7510–7515.
- (30) Batcho, P. *J. Chem. Phys.* **2001**, *115*, 4003.
- (31) Brünger, A. *X-Plor Version 3.1: A System for X-Ray Crystallography and NMR*; Yale University Press: New Haven, CT, 1993.
- (32) Martyna, G.; Tobias, D.; Klein, M. *J. Chem. Phys.* **1994**, *101*, 4177.
- (33) Park, S.; Khalili-Araghi, F. *J. Chem. Phys.* **2003**, *119*, 3559.
- (34) Adler, A.; Grossman, L.; Fasman, G. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *57*, 423–430.
- (35) Holcomb, D.; Tinoco Jr, I. *Pept. Sci.* **1965**, *3*, 121–133.
- (36) Brahms, J.; Michelson, A.; Van Holde, K. *J. Mol. Biol.* **1966**, *15*, 467–88.
- (37) Leng, M.; Felsenfeld, G. *J. Mol. Biol.* **1966**, *15*, 455–66.
- (38) Jensen, M.; Park, S.; Tajkhorshid, E.; Schulten, K. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6731.
- (39) Ma, J.; Dougherty, D. *Chem. Rev.* **1997**, *97*, 1303–1324.
- (40) Saksena, R.; Boghosian, B.; Fazendeiro, L.; Kenway, O.; Manos, S.; Mazzeo, M.; Sadiq, S.; Suter, J.; Wright, D.; Coveney, P. *Philos. Trans. R. Soc. London, Ser. A* **2009**, *367*, 2557–2571.