

Free-Energy Calculations Reveal the Subtle Differences in the Interactions of DNA Bases with α -Hemolysin

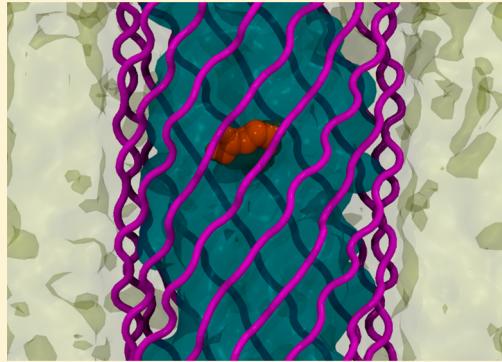
Richard M. A. Manara,[†] Andrew T. Guy,[†] E. Jayne Wallace,[‡] and Syma Khalid*,[†]

[†]Chemistry, Faculty of Natural and Environmental Sciences, University of Southampton, Southampton SO17 1BJ, United Kingdom

[‡]Oxford Nanopore Technologies Ltd., Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford OX4 4GA, United Kingdom

Supporting Information

ABSTRACT: Next generation DNA sequencing methods that utilize protein nanopores have the potential to revolutionize this area of biotechnology. While the technique is underpinned by simple physics, the wild-type protein pores do not have all of the desired properties for efficient and accurate DNA sequencing. Much of the research efforts have focused on protein nanopores, such as α -hemolysin from *Staphylococcus aureus*. However, the speed of DNA translocation has historically been an issue, hampered in part by incomplete knowledge of the energetics of translocation. Here we have utilized atomistic molecular dynamics simulations of nucleotide fragments in order to calculate the potential of mean force (PMF) through α -hemolysin. Our results reveal specific regions within the pore that play a key role in the interaction with DNA. In particular, charged residues such as D127 and K131 provide stabilizing interactions with the anionic DNA and therefore are likely to reduce the speed of translocation. These regions provide rational targets for pore optimization. Furthermore, we show that the energetic contributions to the protein–DNA interactions are a complex combination of electrostatics and short-range interactions, often mediated by water molecules.



■ INTRODUCTION

In recent years, DNA sequencing using nanopores has received much scientific scrutiny due to its promise in next-generation sequencing.¹ The early promise has been realized by Oxford Nanopore Technologies in 2011–2012, and initial devices continue to be upgraded and improved to push the boundaries of what this technology can achieve. Nanopore based sequencing is based upon the same principles as stochastic biosensors:² an engineered protein with a nanoscale aperture is placed in a membrane that separates two chambers containing ionic solution. A voltage is then applied across the chambers, inducing the movement of ions through the pore. This ionic movement is detected as an electrical current. Any charged analyte molecules in the solution will also move through the pore, according to the electric field. In doing so, they cause partial blockage of the current. The extent of the current blockage can be used to identify the molecules, as the level of blocking is specific to a given analyte. Furthermore, it is also possible to calculate the concentration of the analyte by the frequency of blocking events and also the strength of binding from the duration of the event.³ This technique provides information at the single-molecule level and is even sensitive enough to discriminate between enantiomers of the same molecule.⁴ Due to their polyanionic nature, DNA strands can also be driven electrophoretically through a nanopore by applying a potential difference. Each of the four DNA bases blocks the current by a different amount, allowing the sequence

of the strand to be determined.^{1c} While this is the basic principle, for strand sequencing additional complexity arises due to the base currents being context dependent.^{1g}

Both proteinaceous and solid-state nanopores are currently being developed for sequencing DNA.^{1a,e,f,g} To date, much of the research on protein-based nanopores has focused on α -hemolysin, (aHL), a pore forming toxin from *Staphylococcus aureus*. aHL is a homoheptameric protein composed of two main domains: a 14-stranded transmembrane β -barrel and a large extramembranous cap with an internal vestibule.⁶

While aHL is a robust protein that is resistant to changes in pH and temperature over practical ranges, issues related to the high voltage thresholds required for DNA translocation render wild-type aHL unsuitable for incorporation into a sequencing device.⁷ To optimize aHL for DNA sequencing it is essential to understand on a molecular level how DNA behaves within the pore. Previously we have shown that DNA within both wild-type and mutant aHL pores can adopt a range of conformations. These conformations may sometimes lead to nonlinear DNA translocation⁸ i.e. the DNA bases do not necessarily exit the pore in the same order in which they entered. This phenomenon was primarily observed when charged residues were introduced into the barrel lumen. One way to overcome this potential problem is to ratchet the DNA

Received: December 1, 2014

Published: January 13, 2015



into the pore with an enzyme, so it is always fully threaded through the pore.^{1a} In other studies, it has been demonstrated that mutation of the lumen-lining residues of the aHL β -barrel can have a marked effect on the rate of DNA translocation.^{1d}

In this investigation we seek to identify the main protein–DNA interaction sites in the transmembrane region of wild-type aHL and to calculate the energetic difference between the interactions of the protein with the four DNA bases. We use molecular dynamics (MD) simulations, since they enable us to explore the energetic relationships between the structure and dynamics of molecular systems at atomistic resolution. For example, MD has allowed investigation into protein and membrane dynamics and ligands binding to receptors.⁹ A major strength of MD is that it enables the direct study of energetics through techniques such as BAR¹⁰ and the Weighted Histogram Analysis Method (WHAM).¹¹ Free energy techniques have previously been used by Coveney and co-workers to investigate polynucleotide translocation through aHL, using alternative techniques to those used here.¹² The authors focused upon the free energy change involved in translocation, as opposed to characterizing all of the interaction sites. They did however predict that the ring of K147 residues at the constriction site of aHL forms the dominant binding site.

In order to calculate the energetic difference between the interactions of the protein with the DNA bases we perform free energy calculations of nucleic acid fragments through aHL, using umbrella sampling and WHAM.¹¹

METHODS

Simulations were performed in the GROMACS¹³ package version 4.5.5,¹⁴ using the GROMOS 53a6 force field¹⁵ and the SPC water model.¹⁶

We simulate aHL using the pore model described previously.¹⁷ We truncate the protein to only include the transmembrane β -barrel, place the pore in a methane slab, and then solvate in neutralizing 1 M NaCl solution. Importantly the constriction site identified both by Coveney and Bayley is retained in the truncated barrel.^{12,18} Our model pore reduces the total atom count (including water and ions) from hundreds of thousands of atoms, to approximately 20000 atoms. Hence, this reduction in the system size leads to a decrease in the required simulation time, rendering our extensive free energy calculations feasible. As the system is a conformationally stable β -barrel, we do not expect the use of restraints on the protein to influence the PMF. The model system has previously been validated.¹⁷ The system was equilibrated in the NPT ensemble for 100 ns, using the Nosé–Hoover thermostat¹⁹ and Parrinello–Rahman barostat²⁰ to 310 K and 1 bar prior to the umbrella sampling frame setup. Once the substrates were inserted, the systems were energetically minimized using the steepest descent algorithm. The simulations were run between 150 to 250 ns with the first nanosecond discarded for pressure and temperature equilibration. Constraints were used for bond lengths using the LINCS algorithm.²¹ Nonbonded interactions were treated with a cutoff of 1.2 nm, and the Particle Mesh Ewald method was used for long-range electrostatics.²²

We model the DNA molecule by fragmenting the mononucleotide into a phosphate and the bases, shown in Figure 1. Fragmenting the mononucleotides in this way enables us to overcome the issue of long autocorrelation times expected for molecules as complex as mononucleotides, enabling well-resolved and converged PMFs to be obtained.

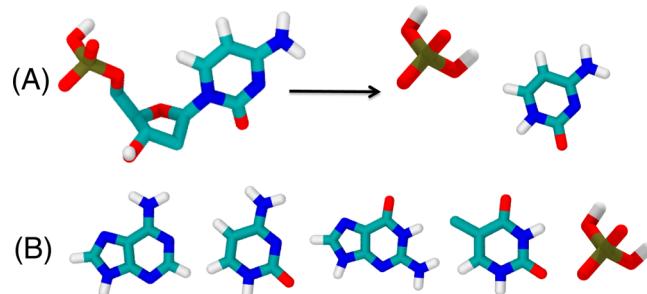


Figure 1. A. Fragmentation of a mononucleotide into a phosphate and base. B. The nucleotide fragments studied in our simulations. From left to right: adenine, cytosine, guanine, thymine, and phosphate fragments. The molecules are colored as follows, oxygen in red, carbon in cyan, nitrogen in blue, hydrogen in white, and phosphate in brown.

The phosphate group was modeled as the single deprotonated form of phosphoric acid, given that 70% of the molecule is in this form at neutral pH. Furthermore, this form carries a net charge of -1 , similar to the nucleic acid backbone. The structures of the nucleotide fragments were constructed using the nucleotide parameters from GROMOS 53a6 as a reference, and the additional hydrogen charge was adjusted so that the total charge was zero for the DNA bases and -1 for the phosphate.

In house scripts were used to set up approximately 160 umbrella sampling windows separated by 5 Å along the principal axis, corresponding to the center of the aHL barrel, as shown in Figure 2. Each window was simulated up to a maximum of 250 ns in duration, giving a maximum simulation time of 40 μ s for each nucleotide fragment. The fragments were restrained along the principle axis of the protein, using a force of 1000 kJ mol $^{-1}$ ·nm $^{-1}$ remaining free to move laterally within the pore. The bias of the restraining force was removed using the weighted histogram analysis method (WHAM) via the

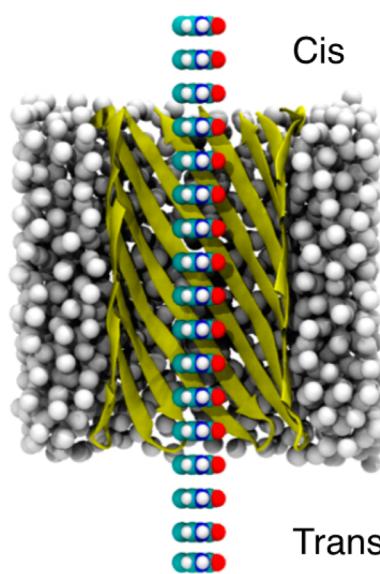


Figure 2. Thymine umbrella sampling windows showing how the position of the base is varied along the principal (z dimension) axis of the protein barrel. Only a selection of windows is shown for clarity, and water and ions are omitted. The methane slab is shown as white spheres, and yellow ribbons represent the protein backbone.

GROMACS program g_WHAM.²³ WHAM was used to construct the cyclic potential of mean force (PMF) curves. For WHAM,¹¹ the number of bins was equal to the number of simulation windows. Autocorrelation times were also calculated using g_WHAM. Bootstrap analysis was performed using the Bayesian bootstrap method with an average of 500 bootstraps. The protein pore dimensions were analyzed using HOLE.²⁴ Visualization was performed with the VMD software package.

RESULTS AND DISCUSSION

Before discussing the energetic profiles of the various DNA fragments, it is useful to clarify the locations of the residues with side chains pointing into the lumen of the β -barrel; these are shown in Figure 3. In subsequent PMF profiles, the Gibbs

$z=2.4 \text{ nm}$	E111	K147
	Y	L
	M113	T145
	S	H
	T115	G143
	L	I
	T117	S141
	Y	V
$z=0.0 \text{ nm}$	G119	N139
	F	A
	N121	G137
	G	I
	N123	L135
	V	G
	T125	G133
	G	I
	D127	K131
	D128	G130
$z=-2.6 \text{ nm}$		T129

Figure 3. Residues within the aHL model pore, shown for one monomer in the β -barrel. Numbered residues are solvated, and their side chains point into the interior of the barrel. The approximate z coordinate of the residues is shown.

free energy is plotted against the z coordinate. The cis and trans entrance to the β -barrel correspond to $z = 2.4 \text{ nm}$ and -2.6 nm , respectively. The center of the pore ($z = 0.0 \text{ nm}$) corresponds to the location of the C- α atoms of G119.

GUANINE PERMEATION ENERGETICS AND PORE DIMENSIONS

The pore radius profile provides an intuitive starting point when considering barriers to translocation. Here we will compare the pore radius against the PMF profile for guanine (Figure 4).

We know that the narrowest constriction within the pore comprises the residues K147 and E111. This constriction has been identified as the major base recognition point, although other recognition points have also been identified.^{1g}

The energetic peaks and troughs, in general, correspond to regions in which the pore is wider and narrower, respectively. Specifically, the cis mouth of the pore provides an energetically favorable region for the bases; this is the narrowest region of the pore. The widest region of the pore is at $z = 0.6$ to 0.9 nm , where T117 residues are flanked predominantly by glycine residues. This region is energetically less favorable presumably

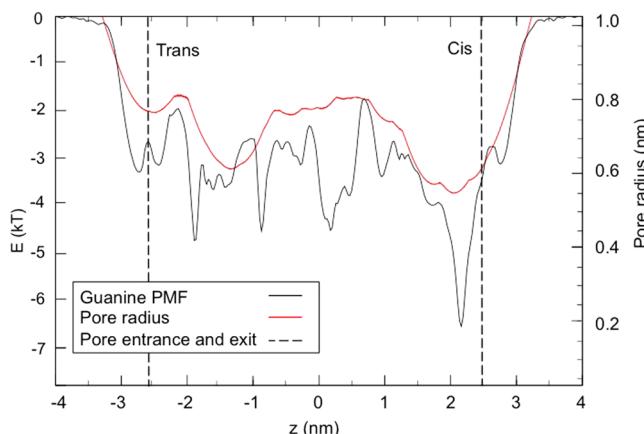


Figure 4. aHL pore radius profile overlaid on the guanine PMF profile.

due in part to the lack of dispersion interactions between the base and the pore-lining residues. While the pore shape does not account for all of the features in the PMF profile it does show the importance of the contribution of the van der Waals interactions and solvation to the energetic profiles.

GUANINE INTERACTION WITH AHL

Closer inspection of the features observed in the PMF profile provides additional details of their molecular origins. First, we discuss the guanine PMF in detail, before performing a comparison with the other three bases of DNA. The PMF profile reveals no significant barrier to permeation along the entire β -barrel (Figure 4). Instead, there are a number of points that correspond to regions of particularly favorable energetics. We shall discuss each of these regions in turn.

$z = -2.0 \text{ nm}$ corresponds to the location of residues D127 and D128. We find that there are two distinct modes of interaction available for D127, either D127 forms a salt bridge with residue K131, located approximately 0.6 nm from the trans exit of the pore, or deprotonated D127 can form hydrogen bonds to guanine (Figure 5). The base-protein interaction (where an interaction is defined as base-protein distance $\leq 0.3 \text{ nm}$) between guanine and either D127 or D128 is present in 75% of the 250 ns simulation. The lifetime of each hydrogen-bonding interaction has duration of less than 15 ns, with 1–4 hydrogen bonds existing simultaneously at any given time (Figure 5). In this region there are on average ~ 20 water molecules within 0.4 nm of the base, these water molecules that freely interchange between hydrogen bonding with the base, the protein, and other water molecules. Thus, guanine is stabilized in this region through hydrogen bonding with D127, D128, and water molecules within the pore.

At $z = -0.9 \text{ nm}$, the side chains of residues N123, N121, and L135 form a pocket in which the five-membered ring of guanine can fit (Figure 6). Importantly, hydrogen bonds are not formed with either the protein or water while the base is within this pocket. Thus, the energetic stabilization of the base in this region is a result of dispersion interactions due to the shape complementarity of the guanine and the three-residue pocket.

At $z = 0.0 \text{ nm}$, guanine is able to form hydrogen bonds with residues N121 and N139. These residues are located on adjacent strands. The lateral distribution of guanine in the umbrella sampling calculation shows a distinct heptameric distribution, indicating the base is not stuck between any two specific β -strands of the protein but samples all seven β -strands

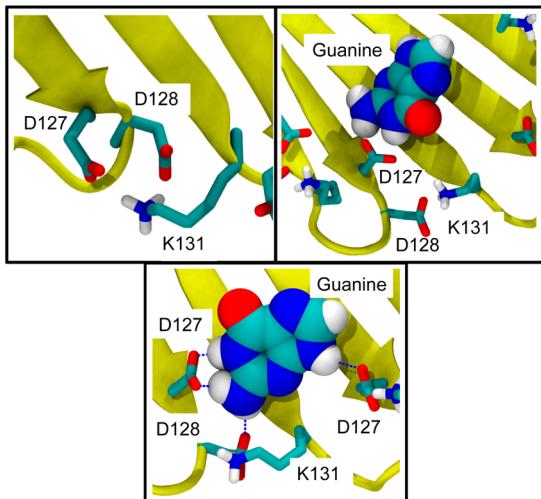


Figure 5. Alternative binding modes available to D127. Both aspartic acid residues D127 and D128 forming salt bridges with K131 (top left) and D127 hydrogen bonding with guanine, while D128 is still engaged in a salt bridge with K131 (top right). Guanine is able to form 4 hydrogen bonds simultaneously with two flanking D127 residues and a lower D128 (bottom). The yellow ribbons represent the protein backbone, the guanine is shown as VDW spheres, and the protein side chains are shown in the stick representation.

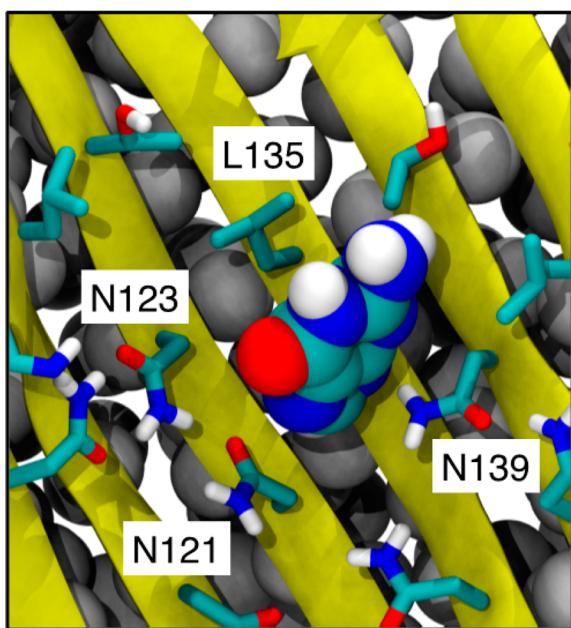


Figure 6. Shape complementarity between the guanine and protein side chains provides favorable energetics in the region $z = -0.9$ nm.

(Figure 7). The base shows a preference for being located between the side chains of N139, orientated such that the long axis of the molecule is parallel to the N139 side chains and the oxygen is pointing into the lumen of the pore. Guanine is stabilized in this region by forming hydrogen bonds with N139. Occasionally guanine forms hydrogen bonds with two N139 residues from different monomers simultaneously, see Figure 7. In this arrangement it can also form water mediated hydrogen bonds with side chains of N121. Alternatively it can form direct hydrogen bonds to N121; the latter are not however generally observed simultaneously with hydrogen bonds to N139. In the

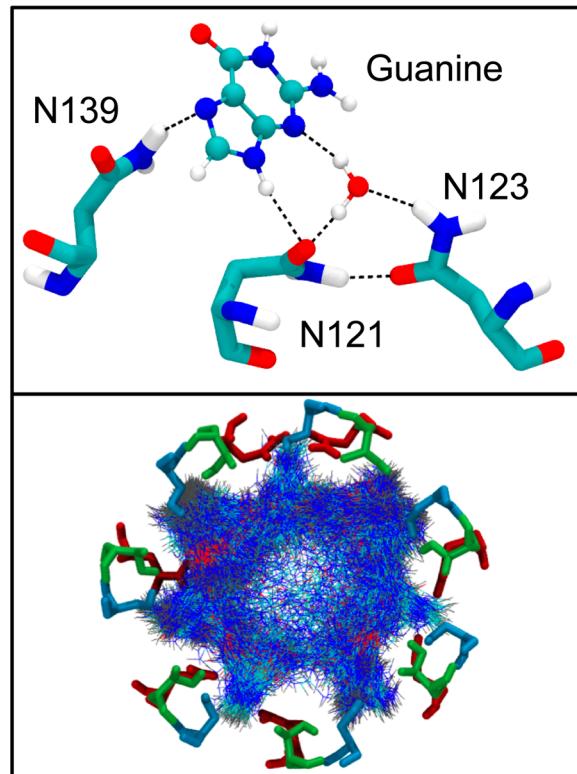


Figure 7. Top: Alternative hydrogen-bonding modes available to guanine at $z = 0.0$ nm. Direct hydrogen bonds formed simultaneously to N139 and N121 residues are shown as well as a water-mediated hydrogen bonding interaction to N123. Bottom: guanine locations across the trajectory superimposed on each other (viewed down the protein principle axis from the cis entrance) shows good sampling of the heptamer, the sticks are asparagine residues, red for N139, green for N121, and blue for N123.

region $z = 0.0\text{--}0.6$ nm guanine forms hydrogen bonds with residues T117 and S141 which are located on adjacent strands.

At $z = 2.4$ nm, K147 and E111 form the narrowest constriction in the pore. We have previously shown that this region provides a barrier for translocation of DNA single strands owing to the hydrogen-bonding and permanent electrostatic interactions of the DNA strands and the side chains of the protein residues.⁸ Here our calculations show that the adjacent position of the acidic and basic side chains can lead to a mimicking of the Watson–Crick hydrogen bonding arrangement with guanine, see Figure 8. It is useful to point out here that for the guanine nucleotide, K147 may preferentially form a salt bridge with the phosphate group rather than the oxygen of guanine, but due to the homoheptameric architecture of aHL, both interactions may be possible simultaneously, see the Supporting Information for evidence of sampling the heptamer. Alternative binding modes are also observed, with the amine group of guanine hydrogen bonding to E111. We note that Stoddart et al. showed that mutation of residues E111 and K147 to asparagine leads to weakened recognition of bases in this region^{1g} and also demonstrated the importance of mutation of M113.²⁶ They surmised that this is likely due to reduced interaction between the pore and the DNA when the charged residues are mutated. Our analysis of the energetics of DNA-pore interactions in this region support their view. Maglia et al. have shown the impact on ssDNA translocation by charged residues within the aHL barrel; our results are in

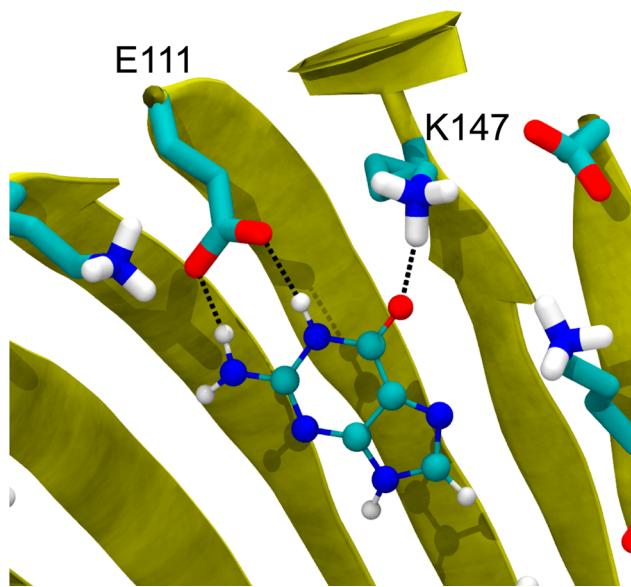


Figure 8. Hydrogen bonding between guanine and K147 and E111, at the narrowest point of the aHL pore.

qualitative agreement with their observations as charged residues lead to features in the PMF curves.⁷

■ COMPARATIVE FREE ENERGY PROFILES OF THE DNA BASES

In order to understand the molecular origins of the ability of aHL to discriminate between the bases of DNA, it is important to compare their individual energetic profiles. In addition to guanine, we have also constructed the PMF profiles of adenine, cytosine, thymine, and the phosphate moiety, see Figure 9. In

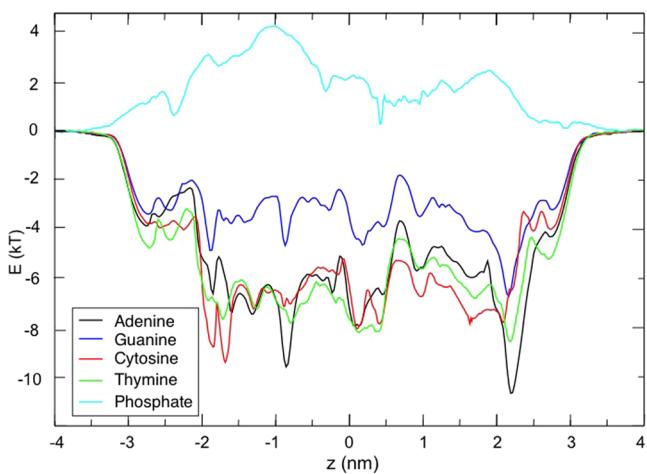


Figure 9. PMF profiles for all four DNA bases and the phosphate group.

the Supporting Information we also provide a short discussion of the full cytosine mononucleotide and show that the profiles are not altered in most regions, by addition of the sugar moiety.

The PMF profiles of all four bases share a similar shape profile; the most notable difference is that the guanine PMF is at higher energy than the other bases in all regions within the barrel. The other 3 bases are generally within thermal error (1 kT is approximately 2.5 kJ mol^{-1} at 310 K). Given their size and chemical similarities it is reasonable to expect the energetic

profiles of the purines to be comparable to each other and likewise for the pyrimidine molecules, which is generally the case. The shapes of the free energy curves for adenine and guanine are similar to each other, and likewise those of thymine and cytosine are similar to each other.

The most striking difference between the purine and pyrimidine profiles is in the region $z = -0.9 \text{ nm}$. As previously explained when discussing the guanine profile, the purine profiles feature a trough in this region that arises from the shape complementarity of the five-membered ring and the pocket formed from the side chains of N123, L135 and N139. This energetic well is not observed in the PMF profiles of the pyrimidines, presumably due to the lack of a five-membered ring in cytosine and thymine.

It is interesting to note that at the narrowest constriction within the protein pore, near E111/K147 (at $z = 2.0\text{--}2.1 \text{ nm}$), the energetics for the adenine-protein interaction are more favorable than for the other bases. We hypothesize that this may be the molecular origin of the slower translocation speed of polyA DNA strands compared to polyC strands, reported from previous experimental and computational studies.²⁷

■ POTENTIAL OF MEAN FORCE PROFILE OF THE PHOSPHATE GROUP

For DNA sequencing applications, either DNA strands or individual nucleotides will be detected. Hence, it is useful to consider the energetics of the charge-carrying phosphate group, H_2PO_4^- (Figure 9). As expected, the energetic profile for phosphate generally follows the opposite trend to the hydrophobic bases. The phosphate experiences an energetic barrier throughout the length of the barrel.

The peak in the PMF profile occurs at $z = -1.0 \text{ nm}$. This corresponds to a leucine (L135) and an asparagine (N123) residue and is at a narrow region in the barrel. Given the hydrophobic nature of leucine, there are fewer protein-phosphate hydrogen bonds and no electrostatic interactions possible here.

Interestingly at $z = 2.4 \text{ nm}$ where the pore narrows at the cis mouth, there is no appreciable well or barrier. This is most likely due to a cancellation of electrostatic attraction/repulsion with the charged residues K147 and E111, respectively.

Our data shows very different energetics experienced by phosphate in aHL compared to the phosphate transporter OprP, which is lined with arginine residues.²⁸ The phosphate group alone is generally at unfavorable energies throughout the hemolysin barrel, compared with favorable energetics within OprP.

■ CONVERGENCE AND ERROR ANALYSES

It is important to evaluate the convergence of the simulations and to estimate errors associated with the PMF curves in order to be confident about their validity. Figure 10 shows bootstrap analysis of the adenine PMF curve. Encouragingly the errors are negligible in the bulk water regions and have a maximum of $\sim \pm 1 \text{ kT}$ in the protein pore. These values compare favorably with similar studies of free energies reported in the literature.²⁹ The histograms, see the Supporting Information (SI), show overlap between all of the umbrella sampling windows, indicating that none of the regions along the principal axis of the pore are unsampled. Autocorrelation times of the energies (SI) are small (below 5 ns on average) compared to simulation length (150–250 ns) and have therefore been covered

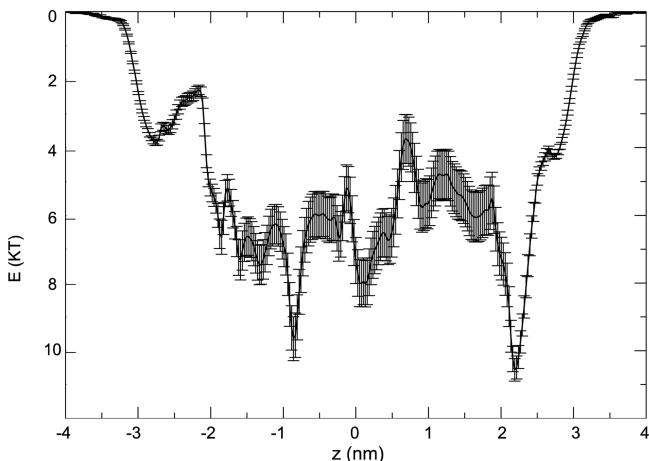


Figure 10. Bootstrap analysis of the adenine fragment PMF. The profile is an average of 500 bootstraps and shown with an associated error estimate.

extensively. The noncyclized PMFs, i.e. the raw data, are presented in the SI.

It is important to place the results into context by considering the limitations of the present study. These arise primarily due to the use of a model pore system, in which only the β -barrel region of the protein has been sampled. However, given that base discrimination occurs within the barrel, for the purposes of this study neglecting the vestibule is reasonable. For a consideration of the energetics of base entry into the pore, it would be essential to incorporate a representation of the pore vestibule into the model.

CONCLUSIONS

In conclusion, we have shown that all the DNA bases experience favorable energetics throughout the β -barrel. The trough in the PMF profiles occurs at the cis entrance, the narrowest region of the barrel. At the cis entrance, there is a ring of glutamate and lysine residues that provide not only a steric constriction but also the possibility for various hydrogen-bonding and electrostatic interactions. In addition, there is potential for discrimination of purines by exploiting their size and shape. For example, our simulations have revealed a region in which the purines are stabilized by virtue of the shape complementarity of the pocket, which is formed by the side chains pointing into the lumen of the barrel and the five-membered ring of the bases. Interestingly, the PMF profiles reveal that the permeation of guanine throughout the pore is less energetically favorable than the other bases. Unfortunately, the energetics of guanine rich regions are not straightforward to study experimentally, due to the formation of the G_quadruplex.^{23,24} Future work will include a consideration of the guanine monophosphate nucleotide to further explore the origins of the free-energy differences exhibited by this base compared to the others.

We expect the energetics of a strand of ssDNA to be similar to the predicted PMFs presented here. However, due to the increase in flexibility compared to the fragments we predict phase space would not be as thoroughly explored within the same simulation time scales.

Overall, the free energy landscape for DNA base permeation through the transmembrane region of aHL is rather more complex than, for example, a phosphate ion within the phosphate specific channel OprP.²⁸ In general, the dimensions

of the aHL barrel are such that the narrower regions provide greater stability for the DNA within the pore. Although there are no energetic barriers to base permeation, there are a number of regions that are more favorable than others. We predict that DNA translocation will be slightly slower in these regions, due to stabilizing protein–DNA interactions. Importantly, optimization of the aHL barrel for improved base discrimination is most likely to be achieved by mutations in these regions to improve the differential “binding” of the four bases. Our simulations have shown that the combinations of protein–base interactions present in this system are difficult to predict *a priori* as they are a complex mix of steric and electrostatic interactions, often also involving contribution from local water molecules. Also, given the inherent flexibility of ssDNA and the multiple binding modes observed in this study, it is conceivable for strand sequencing that multiple binding modes may exist simultaneously; this is particularly likely for DNA that is ratcheted in with an enzyme and fully threaded across the pore. Consequently free energy characterization via molecular simulation provides an invaluable tool for facilitating modification and optimization of nanopores for sequencing applications.

ASSOCIATED CONTENT

S Supporting Information

The autocorrelation times; bootstrap analysis; histograms; and noncyclized profiles. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: S.Khalid@soton.ac.uk.

Author Contributions

All authors have given approval to the final version of the manuscript.

Funding

Richard Manara and Andrew Guy are funded by Oxford Nanopore Technologies.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge the use of the University of Southampton high performance computing resources, Iridis3 and Iridis4. We would also like to thank J. Parkin and H. Bayley for helpful discussions.

ABBREVIATIONS

aHL, α -hemolysin; MD, molecular dynamics; PMF, potential of mean force; WHAM, weighted histogram analysis method

REFERENCES

- (a) Schneider, G. F.; Dekker, C. DNA sequencing with nanopores. *Nat. Biotechnol.* **2012**, *30* (4), 326–328. (b) Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93* (24), 13770–13773. (c) Clarke, J.; Wu, H. C.; Jayasinghe, L.; Patel, A.; Reid, S.; Bayley, H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **2009**, *4* (4), 265–270. (d) Rincon-Restrepo, M.; Milthallova, E.; Bayley, H.; Maglia, G. Controlled Translocation of Individual DNA Molecules through Protein Nanopores with Engineered Molecular Brakes. *Nano Lett.* **2011**, *11* (2), 746–750.

- (e) Venta, K.; Shemer, G.; Puster, M.; Rodriguez-Manzo, J. A.; Balan, A.; Rosenstein, J. K.; Shepard, K.; Drndic, M. Differentiation of Short, Single-Stranded DNA Homopolymers in Solid-State Nanopores. *ACS Nano* **2013**, *7* (5), 4629–4636. (f) Wallace, E. V. B.; Stoddart, D.; Heron, A. J.; Mikhailova, E.; Maglia, G.; Donohoe, T. J.; Bayley, H. Identification of epigenetic DNA modifications with a protein nanopore. *Chem. Commun.* **2010**, *46* (43), 8195–8197. (g) Stoddart, D.; Maglia, G.; Mikhailova, E.; Heron, A. J.; Bayley, H. Multiple Base-Recognition Sites in a Biological Nanopore: Two Heads are Better than One. *Angew. Chem., Int. Ed.* **2010**, *49* (3), 556–559.
- (2) Bayley, H.; Cremer, P. S. Stochastic sensors inspired by biology. *Nature* **2001**, *413* (6852), 226–230.
- (3) Dekker, C. Solid-state nanopores. *Nat. Nanotechnol.* **2007**, *2* (4), 209–215.
- (4) (a) Boersma, A. J.; Bayley, H. Continuous Stochastic Detection of Amino Acid Enantiomers with a Protein Nanopore. *Angew. Chem., Int. Ed.* **2012**, *51* (38), 9606–9609. (b) Kang, X. F.; Cheley, S.; Guan, X. Y.; Bayley, H. Stochastic detection of enantiomers. *J. Am. Chem. Soc.* **2006**, *128* (33), 10684–10685.
- (5) Aksimentiev, A.; Heng, J. B.; Timp, G.; Schulten, K. Microscopic kinetics of DNA translocation through synthetic nanopores. *Biophys. J.* **2004**, *87* (3), 2086–2097.
- (6) Song, L. Z.; Hobaugh, M. R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. E. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **1996**, *274* (5294), 1859–1866.
- (7) Maglia, G.; Restrepo, M. R.; Mikhailova, E.; Bayley, H. Enhanced translocation of single DNA molecules through alpha-hemolysin nanopores by manipulation of internal charge. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (50), 19720–19725.
- (8) Guy, A. T.; Piggot, T. J.; Khalid, S. Single-Stranded DNA within Nanopores: Conformational Dynamics and Implications for Sequencing; a Molecular Dynamics Simulation Study. *Biophys. J.* **2012**, *103* (5), 1028–1036.
- (9) Durrant, J. D.; McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9*, 71.
- (10) Bennett, C. H. Efficient Estimation of Free-Energy Differences from Monte-Carlo Data. *J. Comput. Phys.* **1976**, *22* (2), 245–268.
- (11) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. 1. The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021.
- (12) Martin, H. S. C.; Jha, S.; Howorka, S.; Coveney, P. V. Determination of Free Energy Profiles for the Translocation of Polynucleotides through alpha-Hemolysin Nanopores using Non-Equilibrium Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2009**, *5* (8), 2135–2148.
- (13) Bekker, H.; Berendsen, H. J. C.; Dijkstra, E. J.; Achterop, S.; Vondrumen, R.; Vanderspoel, D.; Sijbers, A.; Keegstra, H.; Reitsma, B.; Renardus, M. K. R. Gromacs - a Parallel Computer for Molecular-Dynamics Simulations. *Physics Computing '92* **1993**, 252–256.
- (14) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29* (7), 845–54.
- (15) Oostenbrink, C.; Soares, T. A.; van der Vegt, N. F. A.; van Gunsteren, W. F. Validation of the 53A6 GROMOS force field. *Eur. Biophys. J. Biophys.* **2005**, *34* (4), 273–284.
- (16) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. SPC. *Intermolecular Forces*; 1981; p 331.
- (17) Bond, P. J.; Guy, A. T.; Heron, A. J.; Bayley, H.; Khalid, S. Molecular Dynamics Simulations of DNA within a Nanopore: Arginine-Phosphate Tethering and a Binding/Sliding Mechanism for Translocation. *Biochemistry* **2011**, *50* (18), 3777–3783.
- (18) Stoddart, D.; Heron, A. J.; Mikhailova, E.; Maglia, G.; Bayley, H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (19), 7702–7707.
- (19) Cheng, A. L.; Merz, K. M. Application of the Nose-Hoover chain algorithm to the study of protein dynamics. *J. Phys. Chem.* **1996**, *100* (5), 1927–1937.
- (20) Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50* (5), 1055–1076.
- (21) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472.
- (22) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.
- (23) Hub, J. S.; de Groot, B. L.; van der Spoel, D. g_wham-A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, *6* (12), 3713–3720.
- (24) Smart, O. S.; Neduvilil, J. G.; Wang, X.; Wallace, B. A.; Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graphics* **1996**, *14* (6), 354–60, 376.
- (25) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33–8, 27–8.
- (26) Stoddart, D.; Heron, A. J.; Klingelhoefer, J.; Mikhailova, E.; Maglia, G.; Bayley, H. Nucleobase Recognition in ssDNA at the Central Constriction of the alpha-Hemolysin Pore. *Nano Lett.* **2010**, *10* (9), 3633–3637.
- (27) (a) Meller, A.; Nivon, L.; Brandin, E.; Golovchenko, J.; Branton, D. Rapid nanopore discrimination between single polynucleotide molecules. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (3), 1079–1084. (b) Wells, D. B.; Abramkina, V.; Aksimentiev, A. Exploring transmembrane transport through alpha-hemolysin with grid-steered molecular dynamics. *J. Chem. Phys.* **2007**, *127* (12) 125101.
- (28) Pongprayoon, P.; Beckstein, O.; Wee, C. L.; Sansom, M. S. P. Simulations of anion transport through OprP reveal the molecular basis for high affinity and selectivity for phosphate. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (51), 21614–21618.
- (29) Yang, C.; Kim, E.; Pak, Y. Potential of Mean Force Simulation by Pulling a DNA Aptamer in Complex with Thrombin. *Bull. Korean Chem. Soc.* **2012**, *33* (11), 3597–3600.