


Cite this: *Nanoscale*, 2021, **13**, 1673

# Translocation of flexible and tensioned ssDNA through *in silico* designed hydrophobic nanopores with two constrictions†

Punam Rattu,<sup>ID</sup> Bastien Belzunces, Taylor Haynes, Chris-Kriton Skylaris<sup>ID</sup> and Syma Khalid<sup>ID</sup> \*

Protein-inspired nanopores with hydrophobic constriction regions have previously been shown to offer some promise for DNA sequencing. Here we explore a series of pores with two hydrophobic constrictions. The impact of nanopore radius, the nature of residues that define the constriction region and the flexibility of the ssDNA is explored. Our results show that aromatic residues slow down DNA translocation, and in the case of short DNA strands, they cause deviations from a linear DNA conformation. When DNA is under tension, translocation is once again slower when aromatic residues are present in the constriction. However, the lack of flexibility in the DNA backbone provides a narrower window of opportunity for the DNA bases to be retained inside the pore *via* interaction with the aromatic residues, compared to more flexible strands. Consequently, there is more variability in translocation rates for strands under tension. DNA entry into the pores is correlated to pore width, but no such correlation between width and translocation rate is observed.

Received 29th June 2020,  
Accepted 20th December 2020

DOI: 10.1039/d0nr04890a

rsc.li/nanoscale

## Introduction

DNA sequencing using nanopores is now a well-established ‘next generation sequencing’ method.<sup>1</sup> Over the years there has been much attention devoted to optimizing the apertures within the membrane, with the aim of controlling conformation and translocation rate of DNA to improve the resolution of the devices.<sup>2–4</sup> The apertures are encased within nanoscale pores which are based on proteins or synthetic materials. Protein engineering of the *S. aureus* toxin  $\alpha$ -hemolysin (a heptamer that forms a 14-stranded  $\beta$ -barrel in the membrane) followed by electrophysiology measurements revealed that basic residues introduced into the lumen of the pore of  $\alpha$ -hemolysin slowed down the translocation of ssDNA.<sup>2</sup> Subsequent simulation studies of simplified models of the  $\alpha$ -hemolysin pore region showed that the strong electrostatic interactions between the basic residues and the acidic phosphate groups of the DNA led to transient tethering of the DNA to the pore, which for short DNA strands resulted in undesirable DNA conformations that may impact on the accuracy of the read.<sup>5</sup> The approach of using model pores was then employed to study hydrophobic hourglass-shaped pores. These pores were based

on a 14-stranded  $\beta$ -barrel architecture and contained a central constriction formed by amino acid residues GAVLVAG.<sup>6</sup> It was shown that ssDNA was slow to enter these pores, however once inside it remained in a largely linear conformation. However, in the absence of the constriction region, the DNA was observed to coil up somewhat. Ideal behaviour for DNA sequencing is that the DNA translocation is slow, and the DNA conformation remains largely linear so the bases pass through the constriction region in the correct order. Inspired by experimental work on engineered  $\alpha$ -hemolysin which showed some advantages in having two constriction regions within a nanopore,<sup>4</sup> here we have designed a set of four nanopores, each with two constriction regions. The protein pores used for DNA sequencing usually have large vestibule regions at the end through which DNA enters. However, solid-state nanopores such as those made from graphene do not have such regions.<sup>7</sup> The outer surface of our pores is cylindrical; thus they are shaped like graphene pores, but are proteinaceous. We explore the impact of pore geometry and chemical nature of the constriction on DNA translocation rate and conformation, with the earlier stated goals of achieving slow translocation of DNA in a linear conformation. ssDNA translocation through these pores, under an applied electric field, is investigated for two scenarios: short (12 bases) DNA strands, and long DNA strands under tension to retain a linear conformation.

Our results show that tryptophan and phenylalanine residues inside the pore interact with the DNA to slow down DNA

School of Chemistry, University of Southampton, Highfield Campus, Southampton, SO17 1BJ, UK. E-mail: S.Khalid@soton.ac.uk

†Electronic supplementary information (ESI) available. See DOI: 10.1039/d0nr04890a

translocation under an applied electric field. This effect is observed both with short strands of DNA that coil/bend due to the interaction, and also longer DNA strands under tension that cannot alter their conformation to any significant extent.

## Computational methods

The hydrophobic pore models used in this work were those constructed using Modeller<sup>8</sup> and PyMOL<sup>9</sup> and were based on the pores first reported by Sansom and co-workers.<sup>10</sup> The ssDNA model used was a 12-nucleotide polyA strand, and was generated using the 3DNA package.<sup>11</sup> The C $\alpha$  atoms of the pores were restrained using a harmonic potential with force constant 1000 kJ mol<sup>-1</sup> nm<sup>2</sup>.

Simulations were performed using GROMACS version 2018.3 and the GROMOS 53a6 forcefield with Berger lipid definitions.<sup>12–14</sup> Pores were inserted into bilayers of 505 DPPC lipids, solvated using the SPC water model and ions added to a concentration of 1.0 M.<sup>15</sup> Additional ions were added to neutralize systems prior to simulation. Positional restraints with force constant 1000 kJ mol<sup>-1</sup> nm<sup>2</sup> were applied to the phosphate moieties of DPPC lipids to prevent movement in the *z* dimension. Long-range electrostatics were treated using the Particle Mesh Ewald method with a short-range cutoff of 1.4 nm.<sup>16</sup> The van der Waals interactions were truncated at 1.4 nm with long-range dispersion corrections applied to the energy and pressure. The temperature of the systems was maintained at 310 K using the *v*-rescale thermostat and a coupling constant of 0.1 ps for both short and continuous DNA systems. The continuous DNA was simulated in the NVT ensemble, whereas the short DNA was simulated in the NPT ensemble. Pressure was maintained at 1 bar semi-isotropically using the Parrinello-Rahman barostat and a time constant of 1 ps.<sup>17</sup> All bonds were constrained using the LINCS algorithm enabling a timestep of 2 fs.<sup>18</sup> A constant electric field was applied across the whole simulation cell, where reverse field simply means that the charged particles are subjected to a force in the opposite direction to the original field. Periodic boundary conditions were applied in three dimensions as in our own previous work on pores based on  $\alpha$ -hemolysin and the work of other groups studying similar phenomena.<sup>5,19–21</sup> Repeat simulations were performed by extracting system coordinates from random timepoints from the last 100 ps of the equilibration run to initiate the simulations. Therefore, the starting coordinates as well as the initial velocities are different for the repeat runs. Analysis was conducted using GROMACS utilities and locally written code. Pore radius profiles for the pore models were obtained using HOLE.<sup>22</sup> Molecular graphic images were produced using the Visual Molecular Dynamics, VMD package.<sup>23</sup> Interaction energies were calculated by post-processing the trajectories within GROMACS. The interaction energies of the 'gate' formed in some of the tensioned DNA simulations were computed using density functional theory (DFT) implemented within the

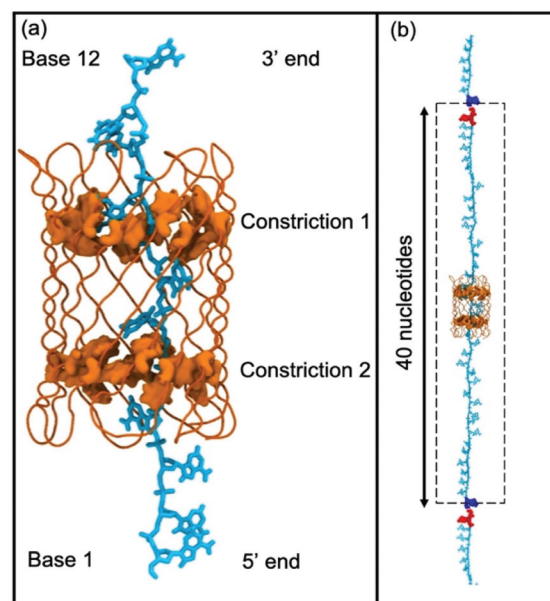
ONETEP code.<sup>24,28</sup> Full details of the DFT methods are provided in the ESI.†

## Results and discussion

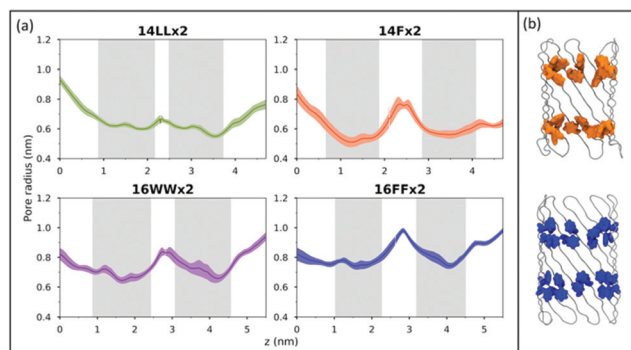
### Model nanopores in an applied electric field

The 4 nanopores studied here mimic  $\beta$ -barrel proteins. Each barrel is composed of either 14 or 16 antiparallel strands and contains 2 hydrophobic regions. The details of the pores are as follows: 14LLx2: 14-stranded barrel with 2 constrictions, each composed of 2 rings of LEU residues; 14Fx2: 14-stranded barrel with 2 constrictions, each composed of 1 ring of PHE residues; 16FFx2: 16-stranded barrel with 2 constrictions, each composed of 2 rings of PHE residues; and 16WWx2: 16-stranded barrel with 2 constrictions, each composed of 2 rings of TRP residues (Fig. S1†). For all pores the constriction nearest the entrance to the pore is termed constriction 1, while the constriction closest to the pore exit is termed constriction 2. Two lengths of DNA were used in this study: a 12-nucleotide strand of finite length and a 40-nucleotide strand bonded to itself through periodic boundaries, both of which have a polyA sequence (Fig. 1).

We explored the behavior of the pores under an applied electric field of 0.15 V nm<sup>-1</sup>. The dimensions of the pores fluctuate only by a small amount under these conditions, due to side chain flexibility (Fig. 2).



**Fig. 1** Panel (a) shows the naming conventions adopted throughout this study, and the initial position of the 12-nucleotide DNA (cyan) in the 14Fx2 pore (orange) for the translocation studies (the constriction residues are shown in surface representation). Panel (b) shows the continuous strand of DNA under tension. The terminal bases are colored blue and red to provide an illustration of the bonding across periodic boundaries (depicted by the dashed lines) to give a continuous strand of DNA.



**Fig. 2** Panel (a) shows the average pore radius profiles for the pores under  $0.15 \text{ V nm}^{-1}$  electric field, (no DNA present). Standard deviations are also shown. The areas shaded in grey correspond to the constriction regions. Panel (b) shows 'side on' views of the 14Fx2 (orange) and 16FFx2 (blue) pores respectively.

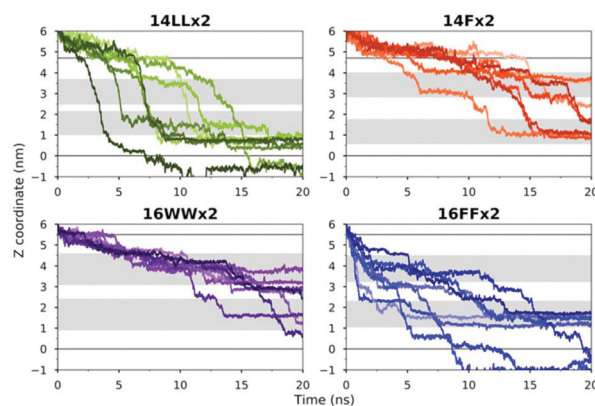
The radii of the constriction regions of the 14-stranded pores are  $\sim 0.47 \text{ nm}$  and  $0.58 \text{ nm}$  in 14Fx2 and 14LLx2 respectively, with the 14Fx2 pore having a larger separation between the constrictions (Fig. 2). The radii of the constriction regions in the 16-stranded pores are  $\sim 0.70 \text{ nm}$  and  $0.62 \text{ nm}$  for 16FFx2 and 16WWx2 respectively. The water flux through the open pores (no DNA present) is provided in Table S1.† Interestingly we observe no water flux in the absence of DNA through the 14LLx2 pore.

### Entry of short DNA strands into hydrophobic pores under an applied electric field

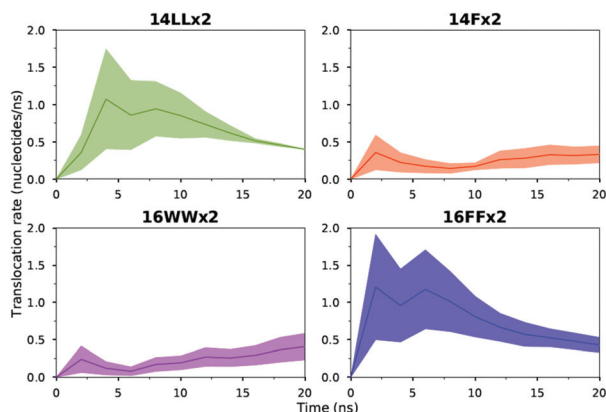
The translocation of ssDNA through these pores was investigated. Two different initial distances from the pore were investigated: (1) the DNA terminal base (5' end) is initially located at the entrance of the pore (2) the DNA is pre-threaded through the pore, such that bases 1–3 (where 1 is the terminal base at the 5' end) are already outside the pore and only base 12 has not yet entered the pore. We have previously shown that the DNA will not enter 14-stranded hydrophobic pores (containing 1 constriction region), without an electric field, and rarely even with an applied electric field.<sup>6</sup> Here we observed that in all 8 simulations under an applied electric field of  $0.15 \text{ V nm}^{-1}$  in which the DNA is initially located at the mouth of the 14-stranded pores ( $4 \times 14\text{Fx2}$  and  $4 \times 14\text{LLx2}$ ), DNA did not enter the pore at all. In contrast to this, in all 8 simulations of the 16-stranded pores ( $4 \times 16\text{FFx2}$  and  $4 \times 16\text{WWx2}$ ), DNA only failed to translocate past constriction 1 once (movement of DNA as a function time for these simulations is shown in Fig. S2†). The simulations show that there is a clear barrier to DNA entry into the 14-stranded pores, but not the 16-stranded pores. This is likely to be related to the entropic penalty incurred upon entering a more confined geometry. Given that DNA entry into the 14-stranded  $\alpha$ -hemolysin is clearly known to occur, our results here suggest that a vestibule region would be advantageous to facilitate entry into our 14-stranded pores but is not needed for the 16-stranded pores.

### Translocation of short DNA strands under an applied electric field

We next explored the simulations in which DNA had been pre-threaded through the pore. Eight independent simulations were performed for each pore. The position of the center of mass of the 3' terminal end base as a function of time was calculated to characterize the DNA translocation rate (Fig. 3). The center of mass data is quoted rather than DNA translocation rate as sometimes the DNA is unable to move through the pore because one end is physically occluding the exit or due to interactions with TRP residues at the entrance of the pore. This is nothing to do with the nature of the constriction regions but is rather a function of the short DNA strands used here. In all eight simulations of the 14LLx2 pore, we found the DNA to take between 4–17 ns to exit constriction 2, and in two of the simulations the DNA exited the pore completely (by 8 ns and 15 ns). Where it was unable to exit the pores, this was a consequence of the DNA occluding the pore at the exit due to interactions with TRP residues. Translocation was slower for the 14Fx2 pore, with the DNA remaining threaded through either both constrictions (in two simulations) or constriction 2 only (in six simulations). For the 16-stranded pores, DNA did not exit constriction 2 in eight and five of the 16WWx2 and 16FFx2 simulations respectively (Fig. 3). For the former this was the case even when the simulations were extended to 40 ns (Fig. S3†). The behaviour for 16FFx2 was more variable, with translocation within 10 and 15 seconds also seen. Thus in 24 independent simulations of pores containing aromatic residues and DNA moving in the 3' to 5' direction, DNA was not retained in the constriction region only twice (both during simulations of 16FFx2). The translocation rates are shown in Fig. 4, here the slower rates through 14Fx2 and 16WWx2 are clear. There is more variation in the 14LLx2 and 16FFx2 rates, but both show faster translocation compared to 14Fx2 and 16WWx2. The cumulative number of nucleotides



**Fig. 3** DNA translocation time in the short strand DNA simulations, measured as the Z coordinate of the center of mass of base 12 (3'-terminal end) as a function of time for all 8 simulations of each pore in which DNA is pre-threaded into the pore. The pore constriction regions are marked by dark grey areas, the solid lines represent the mouths of the pores.

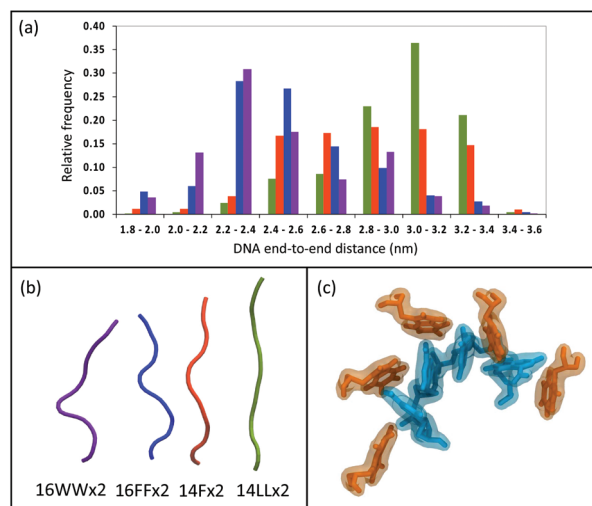


**Fig. 4** DNA translocation rate in short strand DNA simulations, shown as the average rate at which nucleotides exited constriction 2 of model pores as a function of time. Standard deviations are shown.

that exit the constriction region as a function of time is provided in Fig. S4,<sup>†</sup> which again shows the slower movement of nucleotides through 14Fx2 and 16WWx2. Interestingly when the field was reversed so that DNA was moving in the 5' to 3' direction (8 simulations of each system), translocation was slowest through the 16WWx2 pore (Fig. S5 and S6<sup>†</sup>). DNA was unable to fully exit even constriction 1 in any of the simulations (Fig. S7<sup>†</sup>). Fastest translocation was observed through the 14LLx2 pore.

We next sought to establish the origins of these differences in translocation rate by considering the DNA-pore interactions and the conformational behaviour of the DNA strand. We measured the end-to-end distance of the DNA for a 6-nucleotide DNA segment from portions of the trajectories for each pore during which all bases of the segment are located inside the pore. The segments outside the pore are not included in the analysis as interactions with the mouth of the pore or the membrane may mask the effect of the interior of the pore; the latter is our focus here. The relative frequency of DNA end-to-end distances and representative snapshots of the most frequently occurring conformations in each pore are shown in Fig. 5.

We consider the 16-stranded barrels first. The overall end-to-end distance of DNA conformations inside these pores varies between 1.8–3.6 nm. The aromatic residues within the constrictions in the 16-stranded barrels show definite interactions with the DNA. These interactions have two consequences; they slow down the translocation of DNA and also cause it to deviate from a linear conformation. In general, these interactions arise as the aromatic sidechains form pockets within which the DNA bases can fit (Fig. 5). In both pores, the interaction with the residues in the constriction regions causes the DNA to coil such that up to 4 bases are often interacting simultaneously with residues of a particular constriction. The end-to-end distance of the most frequently observed DNA conformations in the 16FFx2 pore varies from 2.2–2.6 nm, while this is 2.2–2.4 nm for 16WWx2. We note



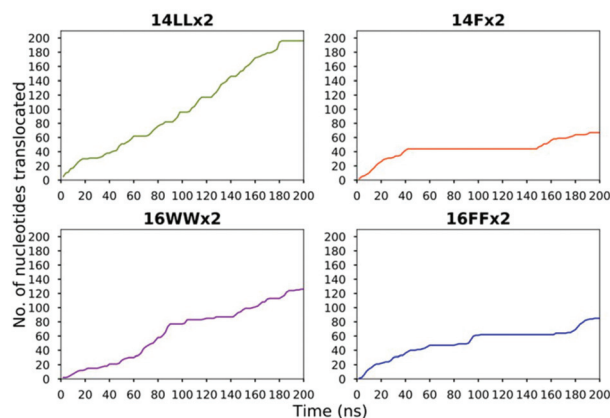
**Fig. 5** DNA conformation in the short strand DNA simulations. Panel (a) shows the distribution of end-to-end (phosphate to phosphate) distances of DNA bases within each pore. Representative backbone conformations are provided in panel (b). Panel (c) shows 4 × DNA bases (cyan) interacting with TRP residues (orange) of the 16WWx2 pore. The DNA bases are seen slotting into the gaps or pockets between the TRP residues.

here that the DNA end-to-end distance is 3.4 nm at the beginning of the simulations. In the 16WWx2 pore, in some cases when the bases at the 3' terminus break free from constriction 1 they rapidly move towards constriction 2, as they are unhindered outside of the constrictions. However, due to the bases already within constriction 2 not being able to escape, there are occasionally up to 6 bases within this constriction leading to retention of the DNA inside the pore, and alterations in its conformation. Examples of the stepwise movement of the DNA between the two constriction regions is seen in Fig. 3. This effect is both less pronounced and less consistent for the 16FFx2 pore, leading to faster translocation times compared to 16WWx2 in general, although there is some variation. Moving onto the 14-stranded pores; within 14LLx2 DNA is in a largely linear conformation throughout the time it spends inside the pore (Fig. 5). The major end-to-end distance of DNA conformations range from 2.4 to 3.4 nm (with ~58% of DNA from all 8 simulations corresponding to conformations with end-to-end distance of 3.0–3.4 nm). There is little interaction with the pore as the DNA moves through, despite the pore being narrow thus, there is no tethering effect. This results in relatively fast translocation rates while also maintaining the DNA in a largely linear conformation, such that there are only 1–2 bases inside the constriction region at any point. DNA within the 14Fx2 has clear interactions with the PHE residues that comprise the pore constrictions. The major DNA conformations of the DNA inside the 14Fx2 pore have end-to-end distances within the range 2.3–3.4 nm (the same range as 14LLx2). However, while the distribution is skewed towards the more extended conformation for 14LLx2, for 14Fx2 the distribution is more even, showing a greater propensity for more compact structures (but

less so than the 16-stranded pores). This is commensurate with the DNA translocation rate observed for this pore, which is slower than 14LLx2. The results here clearly indicate that the interaction with aromatic residues leads to less linear DNA structures and slower translocation rates. In order to quantify the DNA-pore interactions, we postprocessed our trajectories to calculate the interaction between a DNA nucleotide and two side chains of the constriction region. This corresponds to portions of two trajectories for each pore in which the DNA is interacting with the residues and is retained within the pore for the TRP and PHE pores, and simply two representative trajectories for 14LLx2 given we do not observe any DNA retention due to interaction with the LEU residues. The interaction energies are given in Table 1 and representative snapshots are provided in Fig. S8.† The values show greater interaction for the aromatic residues compared to LEU as expected given the DNA translocation behaviour.

### Translocation of long, tensioned DNA strands under an applied electric field

It has been previously shown for solid state nanopores that upon entering a narrow pore from a less confined geometry, the DNA is under tension and in an extended conformation.<sup>25</sup> Proteins that have previously been used for DNA sequencing often have large vestibule regions from which the DNA enters the narrower pore where the sequence is read.<sup>2,26,27</sup> Thus, in the interests of exploring how DNA would behave in our model pores if it was entering from a vestibule, we performed a set of simulations in which the DNA is under tension and therefore the conformation is forced to be more extended. A longer strand of polyA DNA composed of 40 nucleotides was bonded across periodic boundaries in the *z* dimension to give a continuous polymer threaded through the pores. The DNA was simulated in the NVT ensemble and, as the box size is kept constant, the DNA is essentially under tension and cannot coil/kink. The systems were subject to an applied field of 0.09 V nm<sup>-1</sup>. The continuous nature of the DNA enables us to perform one long simulation (200 ns) for each pore and observe multiple DNA translocation events. The cumulative number of nucleotides exiting constriction 2 as a function of time (Fig. 6) show DNA translocation is uniformly rapid through 14LLx2 in these simulations. Plots of the movement of the terminal base show that DNA translocation takes ~40 ns for translocation of 40 nucleotides, but that the DNA is halted

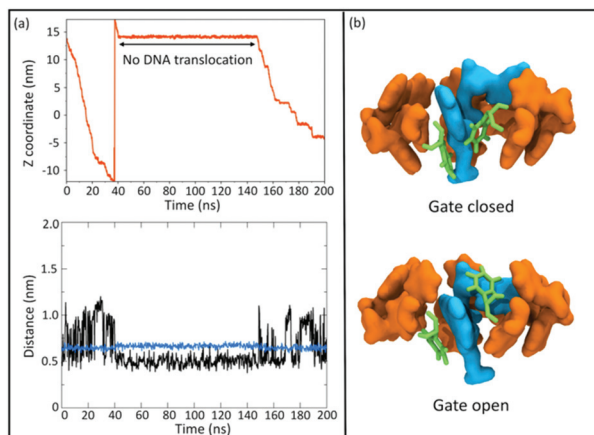


**Fig. 6** Translocation of continuous ssDNA shown as the cumulative number of nucleotides exiting constriction 2 of model pores as a function of time.

from ~180–200 ns (Fig. S9†). Inspection of this trajectory revealed that this halted motion is a consequence of interactions with a TRP residue at the mouth of the pore not the LEU residues of the constriction (Fig. S10†). Extension of the simulation showed that by 208 ns the DNA broke free of this interaction and DNA movement was restored (Fig. S10†). Thus, once again the LEU residues are unable to retain the DNA in the pore. The rate of DNA movement is less uniform through the pores with aromatic constrictions (Fig. 6), and in general slower compared to 14LLx2. For example, the first full translocation of 40 nucleotides occurred within 40 ns through 14Fx2 (same rate as through 14LLx2), but the second translocation event did not fully complete within the 40–200 ns timeframe of the remainder of the simulation (Fig. S9†). In particular, during the time interval ~42–145 ns, there is no appreciable translocation of ssDNA through this pore. Inspection of the trajectory revealed two neighbouring PHE residues in constriction 1 that ‘trap’ one of the DNA bases through non-specific steric interactions, forming a ‘gate’ (Fig. 7). When these residues rearranged such that the separation between their aromatic rings was >~0.5 nm, the base was observed to move out of the constriction under the influence of the electric field. The correlation in the distance between the PHE residues forming the ‘gate’ and the DNA translocation rate is clear (Fig. 7). We computed the interaction energy between the two gating PHE residues and the captured nucleotide using DFT; when the gate is closed (*i.e.* DNA is stationary) the interaction energy is –11.4 kcal mol<sup>-1</sup>, and this is reduced to –7.5 kcal mol<sup>-1</sup> when the gate opens and is further reduced to –2.2 kcal mol<sup>-1</sup> as the nucleotide moves away (Fig. S11†). Interestingly, energy decomposition analysis reveals that the smallest contribution to the PHE-nucleotide stabilisation is from charge-transfer. The individual contributions to the interaction energy are provided in the ESI (Table S4†). In order to further explore the dynamics of the gate, we initiated a further set of six simulations from the point at which the gate is closed. We performed two simulations each with applied electric fields of

**Table 1** DNA-pore interactions energies. In each case two constriction region sidechains are interacting with 1 DNA nucleotide. Standard deviations are given in parentheses

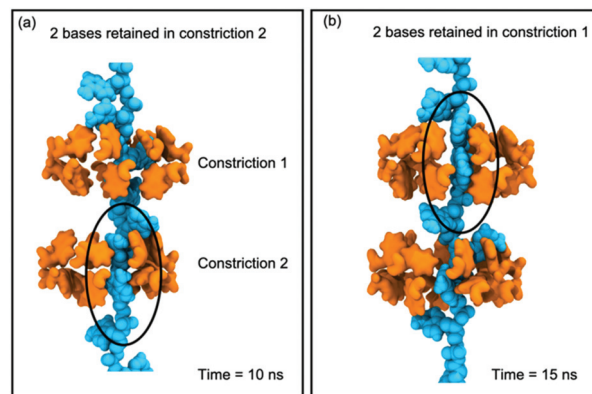
Model pore	Interaction energy (kcal mol <sup>-1</sup> ) (SD)	
	Simulation 1	Simulation 2
14LLx2	–3.6 (0.3)	–4.6 (0.1)
14Fx2	–8.3 (0.3)	–9.1 (0.2)
16FFx2	–7.7 (0.2)	–8.7 (0.1)
16WWx2	–9.2 (0.3)	–15.6 (0.5)



**Fig. 7** Panel (a) shows the center of mass movement of base 40 (3' end) of the continuous DNA model as a function of time through the 14Fx2 pore (top). The bottom of panel (a) shows the distance between the aromatic rings of two neighboring PHE residues (black). The distance between the backbone C $\alpha$  atoms of these residues (blue) is shown for comparison. Panel (b) provides a molecular view of the same PHE residues (green) in the context of other residues in constriction 1 (orange) and their interaction with two DNA nucleotides that occupy the constriction from time = 42–145 ns (cyan).

0.08, 0.09, and 0.10 V nm<sup>-1</sup> (Fig. S12–S14<sup>†</sup>). In each case, the gate opened in 1 simulation but not the other. In 1 simulation with a field of 0.10 V nm<sup>-1</sup>, a third PHE residue was observed to also participate in the gate, and the DNA was held for 100 ns (Fig. S15<sup>†</sup>). This shows the stochastic nature of the DNA–protein interactions when DNA is under tension. Similarly, for the 16-stranded pores, DNA retention times can be correlated with interactions with aromatic residues in the constriction. The aromatic residues are observed to form tight pockets around the DNA bases (Fig. S16<sup>†</sup>) through non-specific interactions. Steps in the translocation correspond to bases moving out of a pocket, and others being retained either within the same constriction or the other one (Fig. 8).

The faster translocation events observed for these pores (full translocation in ~40 ns), correlate with the timescales for full translocation through the 14LLx2 pore. Visual inspection of the trajectories reveals that in these cases of rapid translocation, the DNA is unable to move into a tight-fitting pocket. There are two factors that contribute to this; (a) the DNA is moving under the field and (b) it has very little conformational freedom in the backbone. Consequently, there is a chance that the DNA will not be retained within a pocket formed by neighbouring aromatic residues due to the initial interaction not being optimal, and given it cannot alter its backbone conformation to optimize/improve the interaction, it is able to translocate rapidly past the constriction. Interestingly, the pores containing PHE residues in their constrictions retained tensioned ssDNA for longer periods than 16WWx2. We reason that this likely due to the greater flexibility of the PHE side chains compared to the bulkier TRP side chains (root mean



**Fig. 8** Panel (a) shows two bases caught in a pocket formed by TRP residues in constriction 2, and panel (b) shows the same system 5 ns later, now two bases are caught within a pocket in constriction 1. While there are none in the pockets of constriction 2 (1 base is present inside constriction 2, but this is a loose association, rather than a tight fit into a pocket). This leads to stepwise translocation of the ssDNA.

square fluctuation data is provided in Fig. S17<sup>†</sup>), which enable the former to rapidly form a pocket around DNA bases before they can move through the pore.

## Conclusions

In conclusion, we find that DNA enters the 16-stranded pores readily, but not the 14-stranded pores under an applied electric field, indicating that DNA entry into nanopores is strongly influenced by the pore width. We note here that proteins used for DNA sequencing usually have vestibule regions from which the DNA enters the pore, whereas in our simulations entry (or not) was from bulk solution and thus for pores without vestibules we predict that the cut-off size for easy DNA entry from bulk solutions is ~16 strands. Short strands of DNA move through 14LLx2 unhindered and in a linear/extended conformation. The presence of aromatic residues in the constriction slows down the DNA translocation rate and also leads to DNA deviating from a linear conformation. The slowest translocation rates are observed for 16WWx2. This pore is intermediate in width between the other two pores with aromatic constriction regions. Thus, the nature of the residues at the constriction region has a strong influence over DNA translocation rates for short strands, but there is no correlation between pore width and translocation rate for the pores studied here. For DNA strands under tension, once again the fastest translocation rates are for 14LLx2. Pores containing aromatic residues, in particular PHE, can retain the DNA *via* interactions with the bases for extended periods thereby slowing down translocation. However, if at least one base is not retained within a pocket formed of aromatic residues, the DNA translocation rate is the same as within 14LLx2. In terms of design principles for nanopores used within DNA sequencing devices, our results suggest pores engineered to include TRP for flexible DNA strands and PHE for DNA under tension can slow down

translocation rates. Overall, our 14 and 16-stranded pores with PHE constrictions are the most promising as they slow down DNA translocation (in the case of 16FFx2 mostly when the DNA is under tension), while also not causing too many conformational changes. The 16-stranded pores have the advantage of DNA entering readily, whereas the 14-stranded pores retain the DNA in a more linear conformation and slow down the DNA whether under tension or not. The addition of a vestibule to the 14Fx2 pore would likely facilitate DNA entry, given we know that DNA readily enters the 14-stranded barrel of  $\alpha$ -hemolysin. This will be explored in future designs.

## Funding

PR and TH are funded by Oxford Nanopore Technologies. SK is funded by EPSRC. BB is Funded by the Leverhulme Trust.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

The authors thank Mark Sansom and Shanlin Rao for helpful discussions. BB is supported by the Leverhulme Trust and PR is supported by Oxford Nanopore Technologies Ltd. The authors acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton and the use of the UK national supercomputer, ARCHER granted *via* the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim (<http://hecbiosim.ac.uk>), supported by EPSRC (grant no. EP/R029407/1), in the completion of this work. The Michael supercomputer was used for the DFT calculations for which the authors thank the MMM Hub.

## References

- 1 M. Ayub and H. Bayley, *Curr. Opin. Chem. Biol.*, 2016, **34**, 117–126.
- 2 G. Maglia, M. R. Restrepo, E. Mikhailova and H. Bayley, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 19720–19725.
- 3 M. Rincon-Restrepo, E. Mikhailova, H. Bayley and G. Maglia, *Nano Lett.*, 2011, **11**, 746–750.
- 4 D. Stoddart, G. Maglia, E. Mikhailova, A. J. Heron and H. Bayley, *Angew. Chem., Int. Ed.*, 2010, **49**, 556–559.
- 5 P. J. Bond, A. T. Guy, A. J. Heron, H. Bayley and S. Khalid, *Biochemistry*, 2011, **50**, 3777–3783.
- 6 T. Haynes, I. P. S. Smith, E. J. Wallace, J. L. Trick, M. S. P. Sansom and S. Khalid, *ACS Nano*, 2018, **12**, 8208–8213.
- 7 S. J. Heerema and C. Dekker, *Nat. Nanotechnol.*, 2016, **11**, 127–136.
- 8 R. Sanchez and A. Sali, *Methods Mol. Biol.*, 2000, **143**, 97–129.
- 9 The PyMOL Molecular Graphics System, *Version 2.0 Schrödinger*, LLC.
- 10 J. L. Trick, E. J. Wallace, H. Bayley and M. S. Sansom, *ACS Nano*, 2014, **8**, 11268–11279.
- 11 X. J. Lu and W. K. Olson, *Nucleic Acids Res.*, 2003, **31**, 5108–5121.
- 12 S. Pall, M. J. Abraham, C. Kutzner, B. Hess and E. Lindahl, *Lect. Notes Comput. Sci. Eng.*, 2015, **8759**, 3–27.
- 13 C. Oostenbrink, T. A. Soares, N. F. van der Vegt and W. F. van Gunsteren, *Eur. Biophys. J.*, 2005, **34**, 273–284.
- 14 O. Berger, O. Edholm and F. Jahnig, *Biophys. J.*, 1997, **72**, 2002–2013.
- 15 C. D. Berweger, W. F. van Gunsteren and F. Mullerplathe, *Chem. Phys. Lett.*, 1995, **232**, 429–436.
- 16 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
- 17 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
- 18 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.
- 19 A. T. Guy, T. J. Piggot and S. Khalid, *Biophys. J.*, 2012, **103**, 1028–1036.
- 20 R. M. A. Manara, S. Tomasio and S. Khalid, *Nanomaterials*, 2015, **5**, 144–153.
- 21 S. Bhattacharya, I. M. Derrington, M. Pavlenok, M. Niederweis, J. H. Gundlach and A. Aksimentiev, *ACS Nano*, 2012, **6**, 6960–6968.
- 22 O. S. Smart, J. G. Neduvilil, X. Wang, B. A. Wallace and M. S. P. Sansom, *J. Mol. Graphics*, 1996, **14**, 354–360.
- 23 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.
- 24 C. K. Skylaris, P. D. Haynes, A. A. Mostofi and M. C. Payne, *J. Chem. Phys.*, 2005, **122**, 84119.
- 25 N. A. W. Bell, K. Chen, S. Ghosal, M. Ricci and U. F. Keyser, *Nat. Commun.*, 2017, **8**, 380.
- 26 I. M. Derrington, T. Z. Butler, M. D. Collins, E. Manrao, M. Pavlenok, M. Niederweis and J. H. Gundlach, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 16060–16065.
- 27 R. M. Manara, E. J. Wallace and S. Khalid, *Sci. Rep.*, 2015, **5**, 12783.
- 28 J. C. A. Prentice, J. Aarons, J. C. Womack, A. E. A. Allen, L. Andrinopoulos, L. Anton, R. A. Bell, A. Bhandari, G. A. Bramley, R. J. Charlton, R. J. Clements, D. J. Cole, G. Constantinescu, F. Corsetti, S. M.-M. Dubois, K. K. B. Duff, J. M. Escartin, A. Greco, Q. Hill, L. P. Lee, E. Linscott, D. D. O'Regan, M. J. S. Phipps, L. E. Ratcliff, Á. R. Serrano, E. W. Tait, G. Teobaldi, V. Vitale, N. Yeung, T. J. Zuehlsdorff, J. Dziedzic, P. D. Haynes, N. D. M. Hine, A. A. Mostofi, M. C. Payne and C.-K. Skylaris, *J. Chem. Phys.*, 2020, **152**, 174111.