

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

Punam Rattu,^a Bastien Belzunces,^a Taylor Haynes,^a Chris-Kriton Skylaris,^a and Syma Khalid^{*a}

^aSchool of Chemistry, University of Southampton, Highfield Campus, Southampton, SO17 1BJ

Supplementary information.

Methods for Density Functional Theory Calculations

Three system snapshots were extracted from the MD simulations in which (a) a nucleotide is captured within a gate formed by two PHE residues (b) the configuration of the PHE residues has changed so the gate is open and (c) when the nucleotide has started to move away from the gate. The coordinates of the PHE side chains and the nucleotide were retained and the geometry of each frame was optimised using linear-scaling Density Functional Theory (DFT) in ONETEP^{1,2}, using the Perdew-Burke-Ernzerhof exchange-correlation functional³ with the D2 Grimme dispersion correction (PBE-D2)⁴. Open boundary conditions via the multigrid Poisson-Boltzmann solver were used in a (125 nm)³ cubic simulation cell. Norm-conserving pseudopotentials were used. The psinc basis set (which is equivalent to a plane wave basis set) was employed with a kinetic energy cut-off of 800 eV. No cut-off was applied to the density kernel and 8.0 Bohr radii were used for the nonorthogonal generalized Wannier functions (NGWFs)⁵.

Prior to calculation the carbonyl and amine groups were replaced by hydrogen atoms (H in place of C and N). All the missing hydrogen atoms from the united-atom MD simulations were added using the PDB Reader tool of the CHARMM-GUI webserver^{6,7}. All the systems bore an overall charge of -1 *e* due to the phosphate moiety of the nucleotide.

Interaction energies (E_{int}) were calculated as the difference between the total DFT energy of the complex formed from the nucleotide (35 atoms) and the two PHE residues (36 atoms) ($E_{(adenine-residue)}$) and the energies from the two fragments constituting that complex, according to equation 1:

$$E_{int} = E_{(adenine-residue)} - E_{(adenine)} - E_{(residue)}$$

Equation 1.

The Energy Decomposition Analysis (EDA) approach enables separation of interaction energies into chemically relevant components for greater understanding of the origins of stabilisation effects. The ONETEP version of EDA^{8,9}, based on the Localised Molecular Orbital (LMO)¹⁰ and Absolutely LMO (ALMO)¹¹ approaches, allows to decompose the interaction energy into three main contributions: frozen ($E_{(FRZ)}$, sum of electrostatics, exchange, repulsion, and correlation energies), polarisation ($E_{(pol)}$), and charge transfer ($E_{(CT)}$), according to equation 2:

$$E_{int} = E_{(FRZ)} + E_{(pol)} + E_{(CT)}$$

Equation 2.

Results

Table S1. Average mean water and ion flux in model pores from simulations under an electric field of 0.15 Vnm^{-1} , in the absence of DNA. Standard deviations are given in parentheses.

Model pore	Average mean flux (ns^{-1}) (SD)		
	Water	Na^+	Cl^-
14LLx2	0	0	0
14Fx2	24 (3.0)	1.2 (0.4)	1.9 (0.2)
16FFx2	55 (0.8)	3.0 (0.8)	5.2 (0.4)
16WWx2	45 (4.7)	1.3 (0.5)	4.6 (0.1)

Table S2. A summary of all short strand simulations in which ssDNA was initially placed at the entrance to the pore, under an electric field of 0.15 Vnm^{-1} .

Model pore	Final DNA location (number of simulations)		
	DNA above or in constriction 1	DNA above or in constriction 2	DNA at pore exit
14LLx2	4	0	0
14Fx2	4	0	0
16FFx2	0	0	4
16WWx2	0	1	3

Table S3. A summary of all short strand simulations in which ssDNA was pre-threaded into the pore, under an electric field of 0.15 Vnm^{-1} .

Model pore	Final DNA location (number of simulations)		
	DNA above or in constriction 1	DNA above or in constriction 2	DNA at pore exit
14LLx2	0	0	8
14Fx2	2	6	0
16FFx2	0	5	3
16WWx2	2	6	0

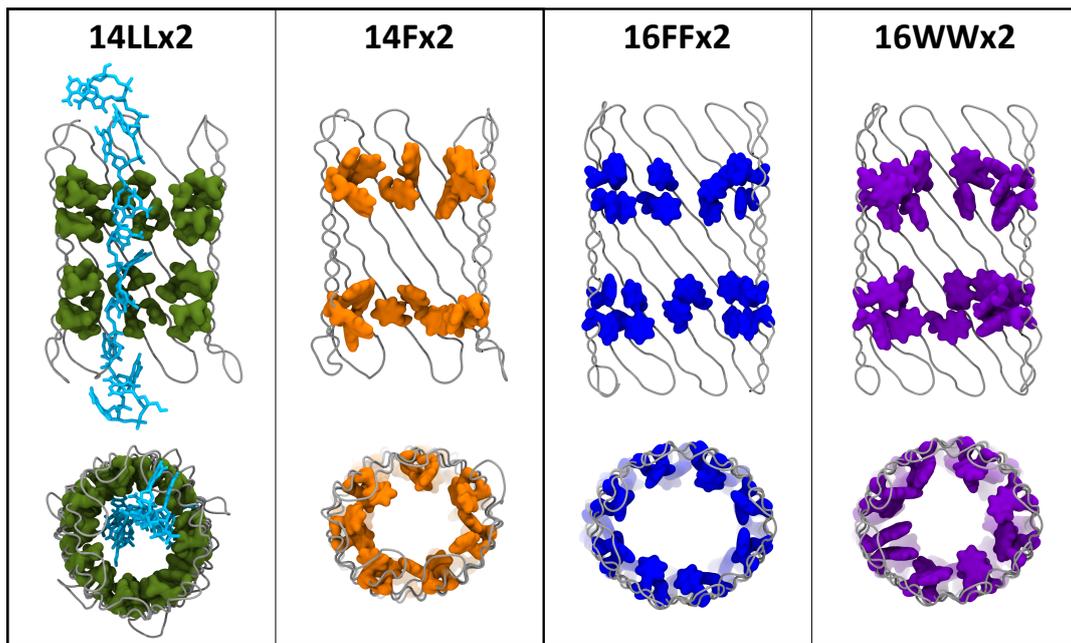


Fig. S1. Cross-sectional and birds-eye views of the model pores used in this study. The 14LLx2 pore is shown with pre-threaded DNA strand (cyan).

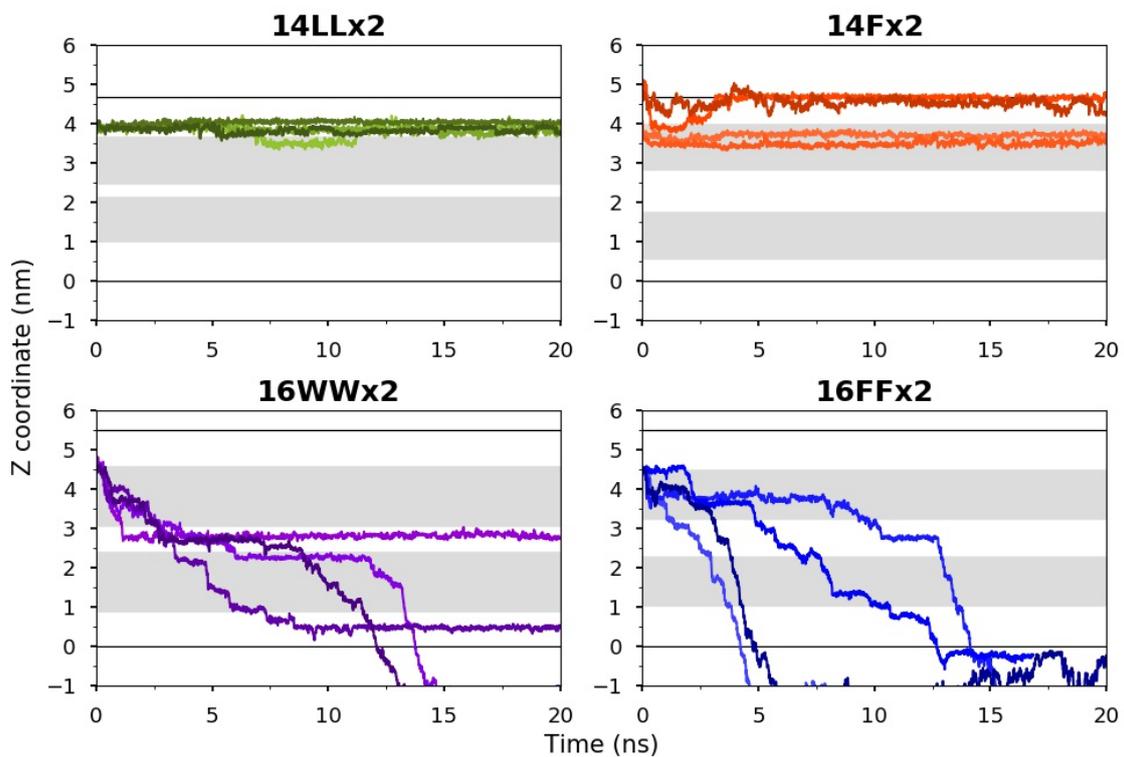


Fig. S2. DNA translocation time in the short strand DNA simulations, measured as the Z coordinate of the center of mass of base 1 (5'-terminal end) as a function of time for all 4 simulations of each pore in which DNA is initially located at the entrance of the pore. The pore constriction regions are marked by dark grey bands, the solid lines represent the mouths of the pores.

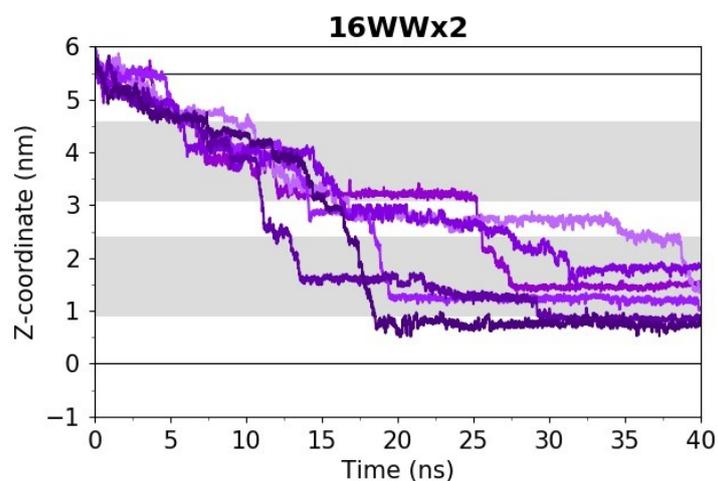


Fig. S3. DNA translocation time in the short strand DNA simulations, measured as the Z coordinate of the center of mass of base 12 (3'-terminal end) as a function of time for 6 simulations of 16WWx2 pore extended to 40 ns. The pore constriction regions are marked by dark grey bands, the solid lines represent the mouths of the pores.

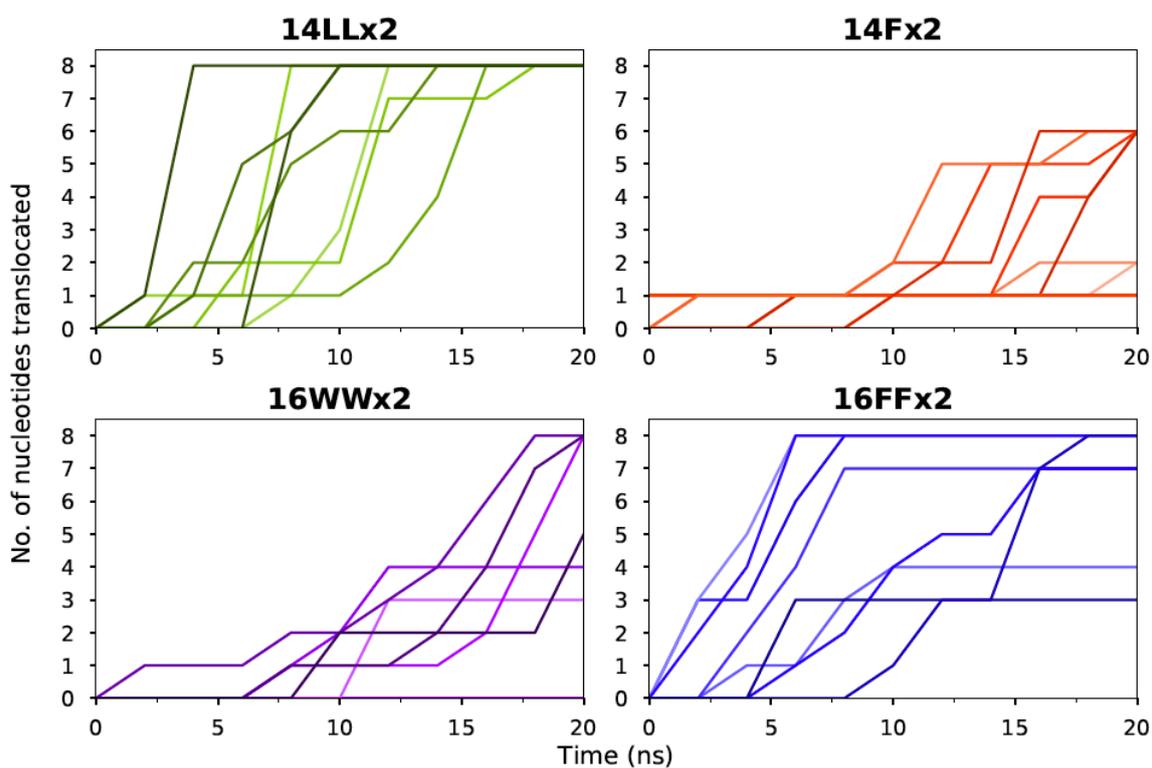


Fig. S4. Translocation of the short DNA strands shown as the cumulative number of nucleotides translocated through constriction 2 of the model pores as a function of time.

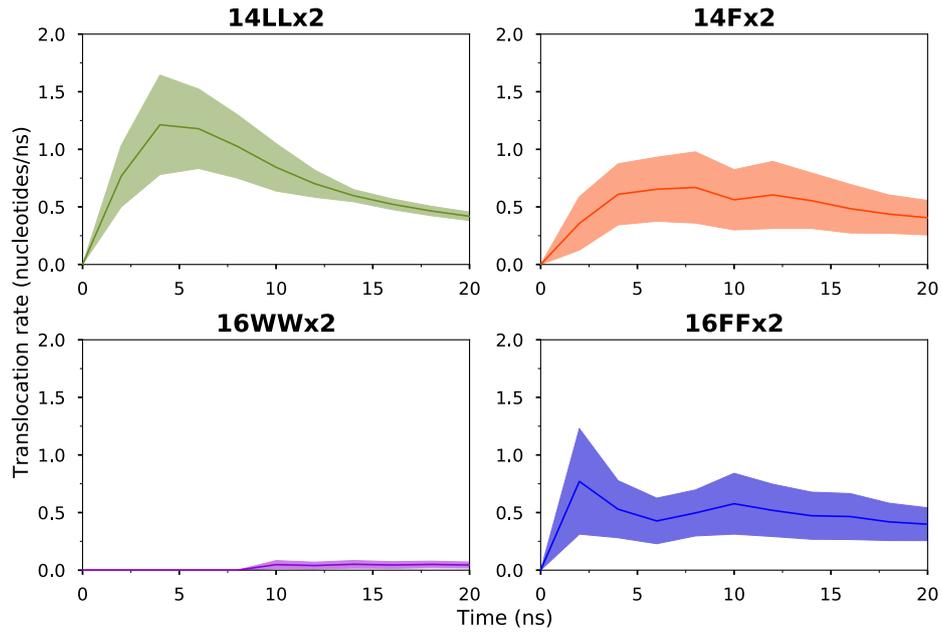


Fig. S5. DNA translocation rate in short strand DNA simulations, with DNA moving in the 5'-to-3' direction, shown as the average rate at which nucleotides exited constriction 2 of model pores as a function of time. Standard deviations are shown.

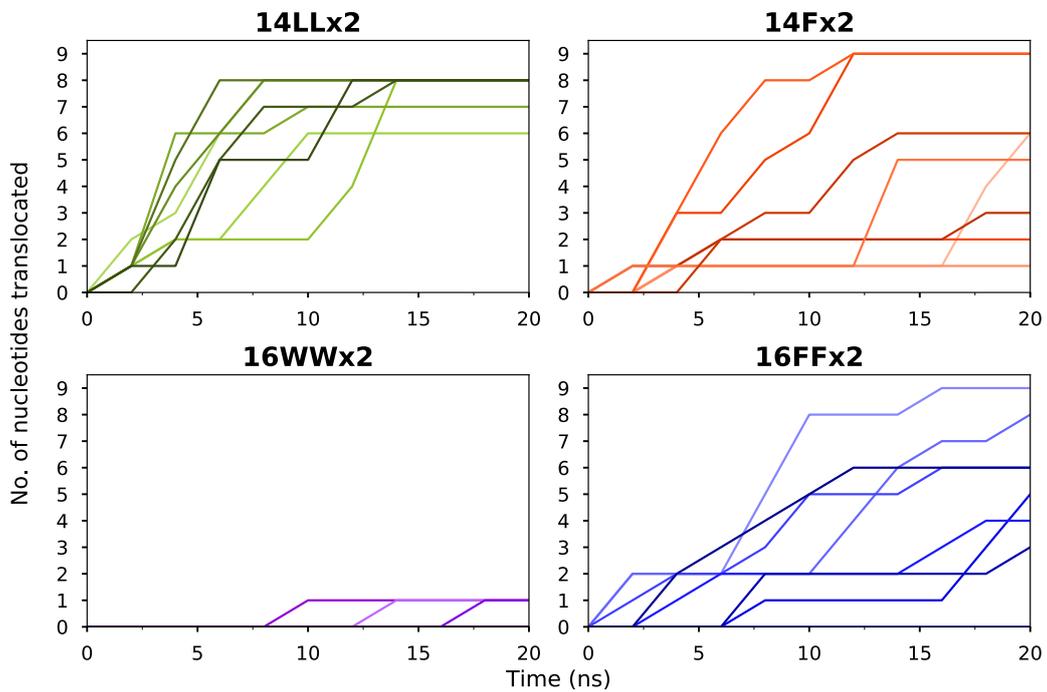


Fig. S6. Translocation of the short DNA strands in the 5'-to-3' direction, shown as the cumulative number of nucleotides translocated through constriction 2 of the model pores as a function of time.

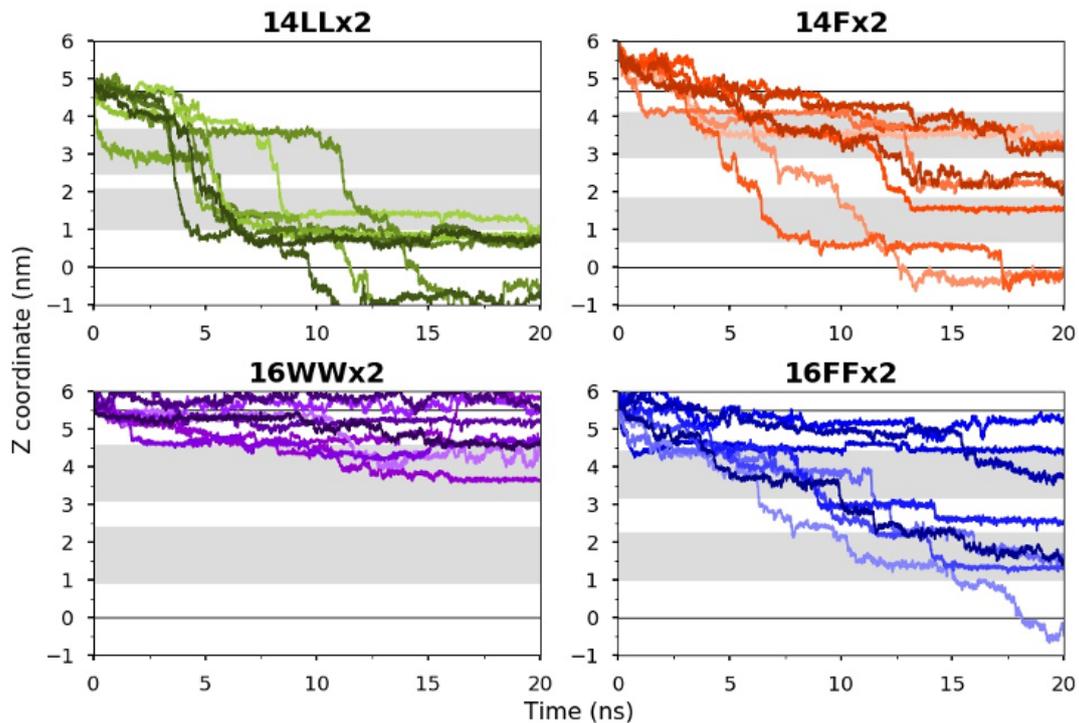


Fig. S7. DNA translocation time in the short strand DNA simulations, with DNA moving in the 5'-to-3' direction, measured as the Z coordinate of the center of mass of base 1 (5'-terminal end) as a function of time for all 8 simulations of each pore in which DNA is pre-threaded into the pore. The pore constriction regions are marked by dark grey bands, the solid lines represent the mouths of the pores.

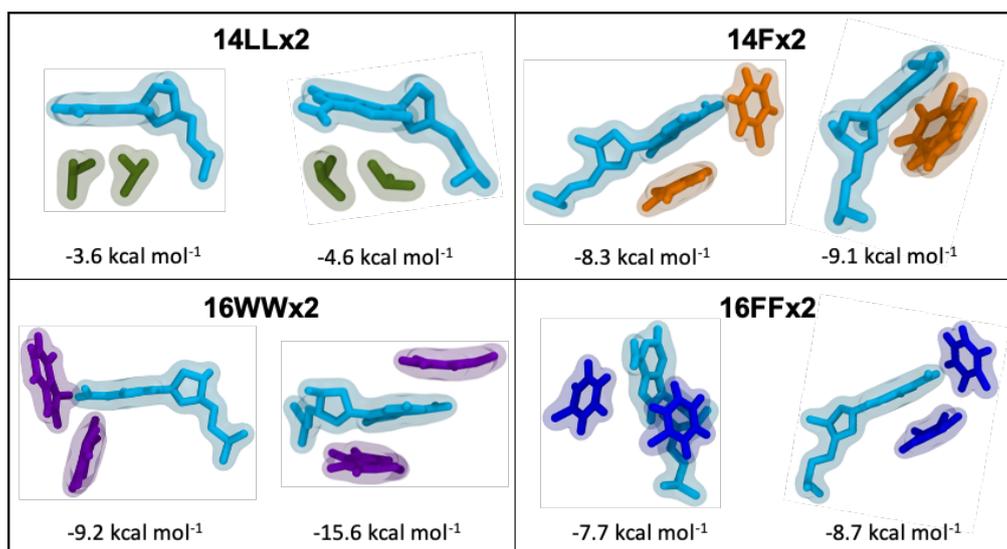


Fig. S8. Interaction energies between a DNA nucleotide (cyan) and side chains of two residues of the constriction regions of the model pores, from two trajectories of each pore.

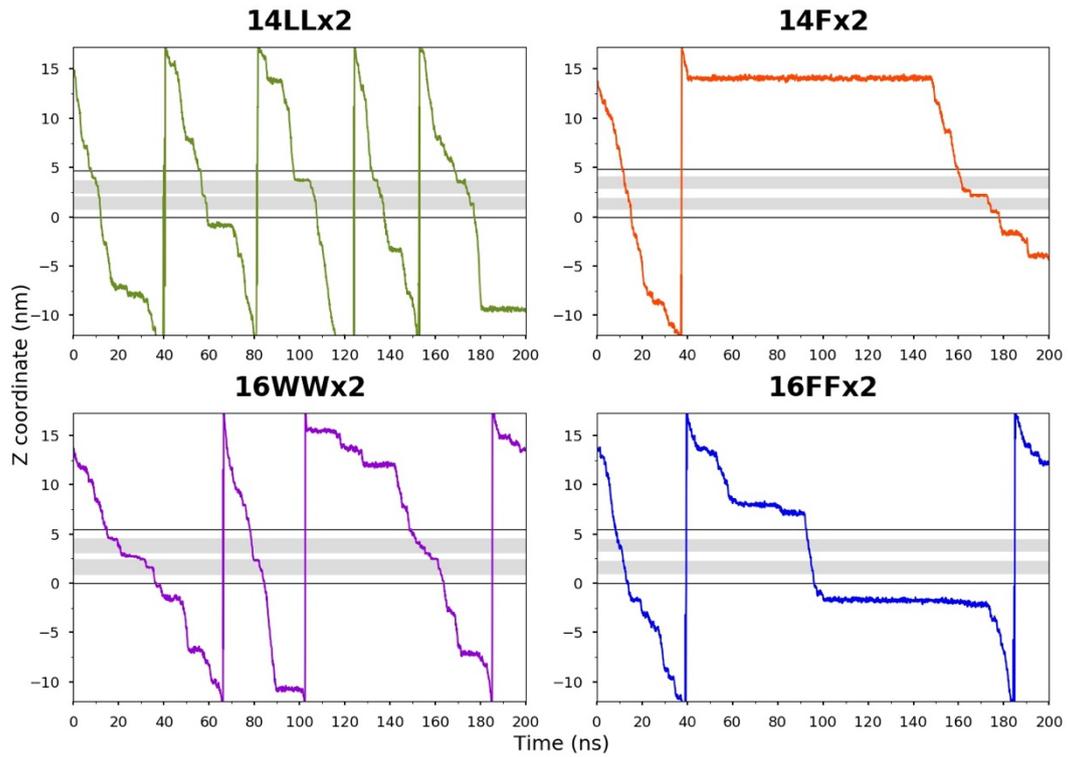


Fig. S9. Translocation of continuous ssDNA under tension through all four pores. The center of mass of the base furthest from constriction 1 at the start of the simulations (3'-terminal end) plotted against time. The shaded regions represent the pore constrictions and the solid black lines represent the mouths of the pores.

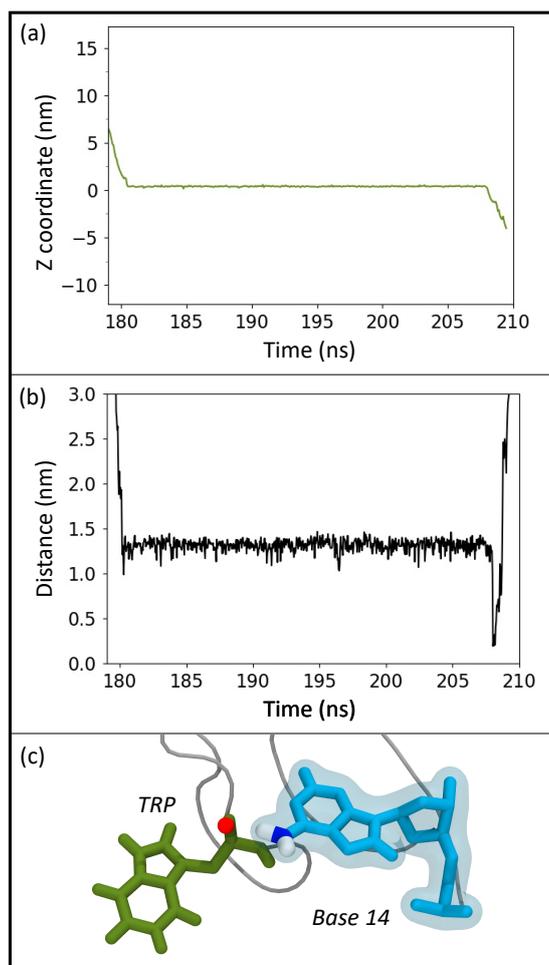


Fig. S10. DNA translocation is halted in 14LLx2 pore due to a base interacting with a TRP residue in the pore mouth. Panel (a) shows the center of mass movement of base 14 of the continuous DNA during 180-210 ns. The distance between base 14 and a TRP residue is shown in panel (b), and the molecular view of their interaction is shown in panel (c).

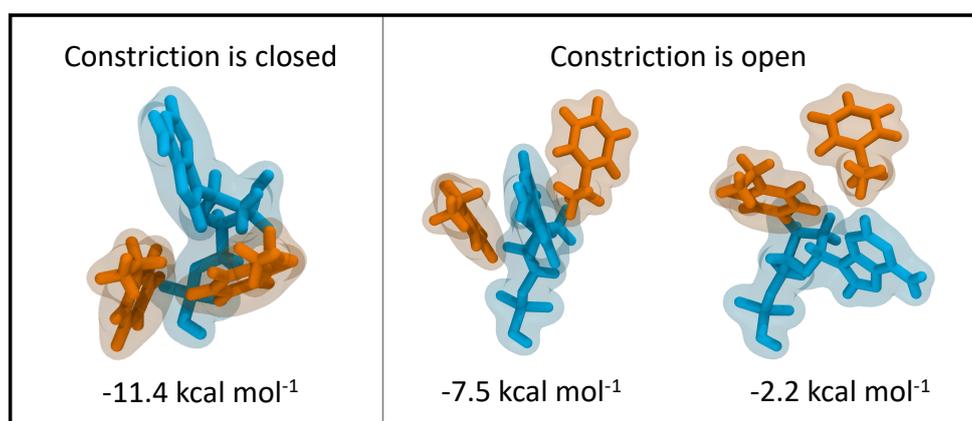


Fig. S11. The interaction energies between the gating PHE residues and captured nucleotide during translocation of continuous DNA through 14Fx2 pore is shown, the values are calculated for closed and open gate conformations using DFT.

Table S4. Energy Decomposition Analysis (at the PBE level) of the interaction between the gating PHE residues and captured nucleotide, for closed and open gate conformations of the PHE gate.

	kcal mol ⁻¹		
	Constriction is closed	Constriction is open	
$E_{(int)}$ (PBE-D2)	-11.4	-7.5	-2.2
$E_{(int)}$ (PBE)	-4.6	2.9	-0.2
$E_{(D2)}$	-6.8	-10.4	-2.0
$E_{(FRZ)}$	-0.8	4.1	0.2
$E_{(pol)}$	-3.9	-1.6	-0.7
$E_{(CT)}$	0.2	0.4	0.2

$E_{(int)}$: Interaction energy with (PBE-D2) and without (PBE) dispersion correction.

$E_{(D2)}$: Grimme dispersion correction contribution such as: $E_{D2} = E_{int}(\text{PBE-D2}) - E_{int}(\text{PBE})$

$E_{(FRZ)}$: Frozen interaction energy.

$E_{(rep)}$: Repulsion interaction energy.

$E_{(pol)}$: Polarisation.

$E_{(CT)}$: Charge transfer.

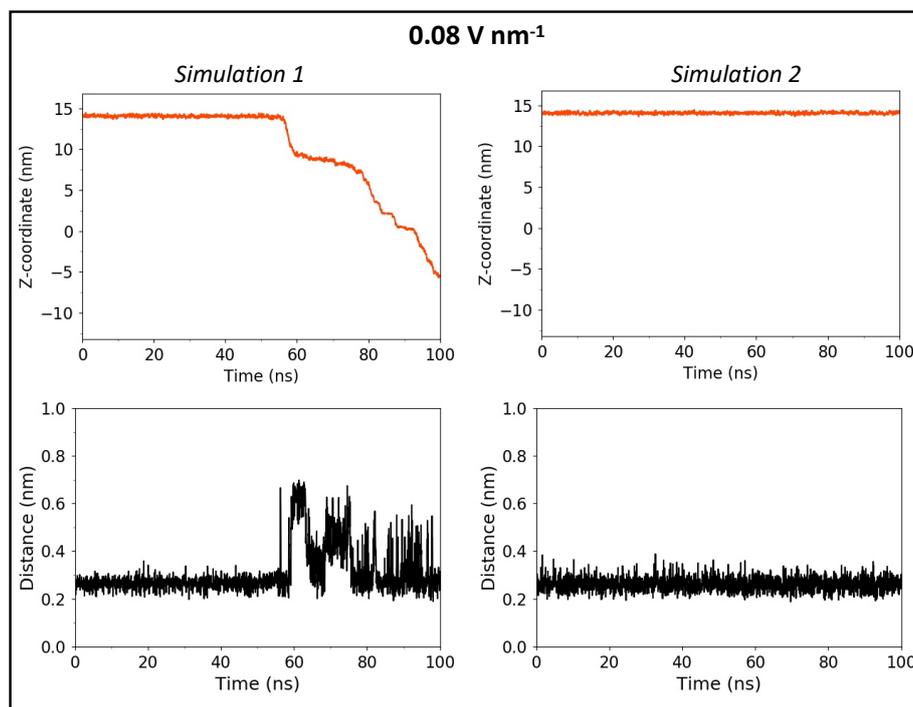


Fig. S12. The center of mass movement of base 40 (3'-terminal end) of the continuous DNA through 14Fnx2 pore in two simulations in 0.08 V nm⁻¹ electric field is shown (orange), starting with a nucleotide trapped in gate formed by two PHE residues. The distance between PHE sidechains forming the gate is shown in black.

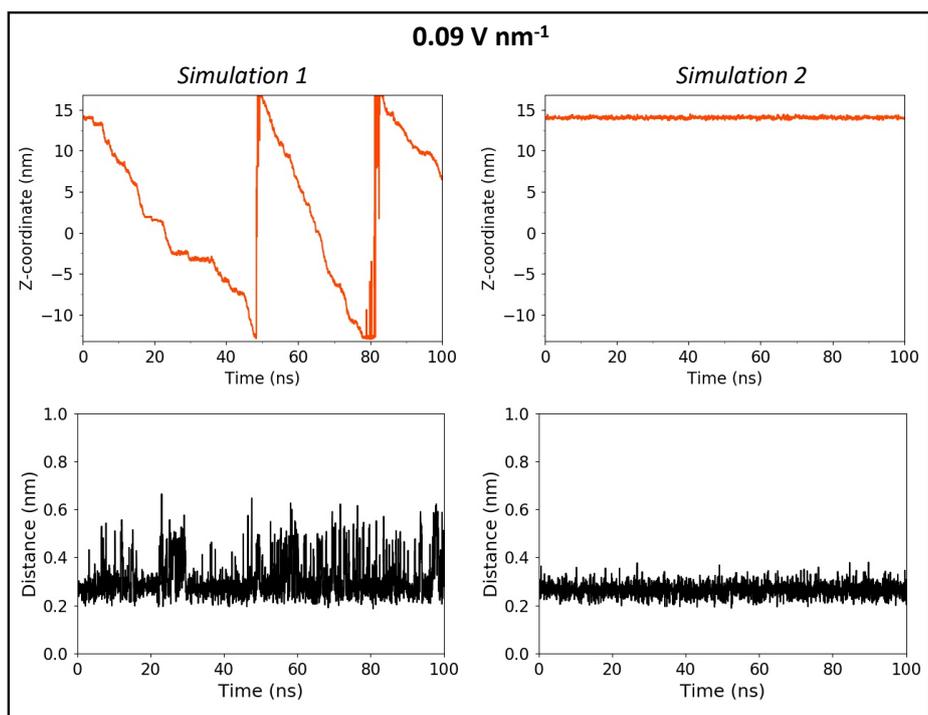


Fig. S13. The center of mass movement of base 40 (3'-terminal end) of the continuous DNA through 14Fx2 pore in two simulations in 0.09 V nm^{-1} electric field is shown (orange), starting with a nucleotide trapped in gate formed by two PHE residues. The distance between PHE sidechains forming the gate is shown in black.

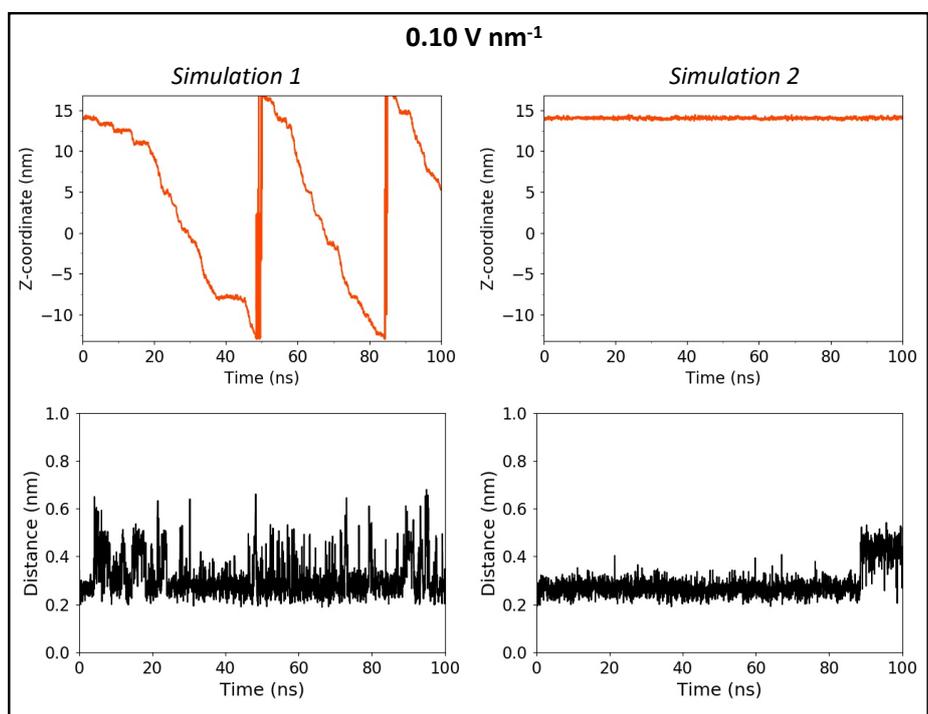


Fig. S14. The center of mass movement of base 40 (3'-terminal end) of the continuous DNA through 14Fx2 pore in two simulations in 0.10 V nm^{-1} electric field is shown (orange), starting with a nucleotide trapped in gate formed by two PHE residues. The distance between PHE sidechains forming the gate is shown in black.

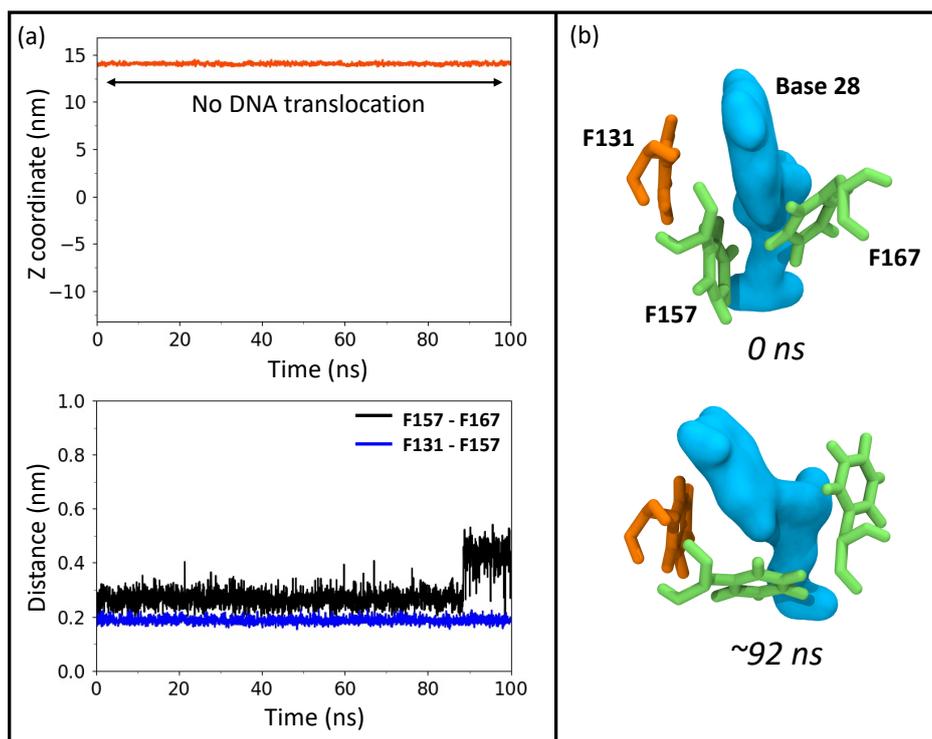


Fig. S15. Panel (a) shows the center of mass movement of base 40 (3'-terminal end) of the continuous DNA model as a function of time through the 14Fx2 pore (top), in 0.10 V nm^{-1} applied electric field. The bottom of panel (a) shows the distance between the aromatic rings of two pairs of neighbouring PHE residues. Panel (b) provides a molecular view of the same PHE residues and their interaction with a DNA nucleotide that occupies the constriction for 100 ns.

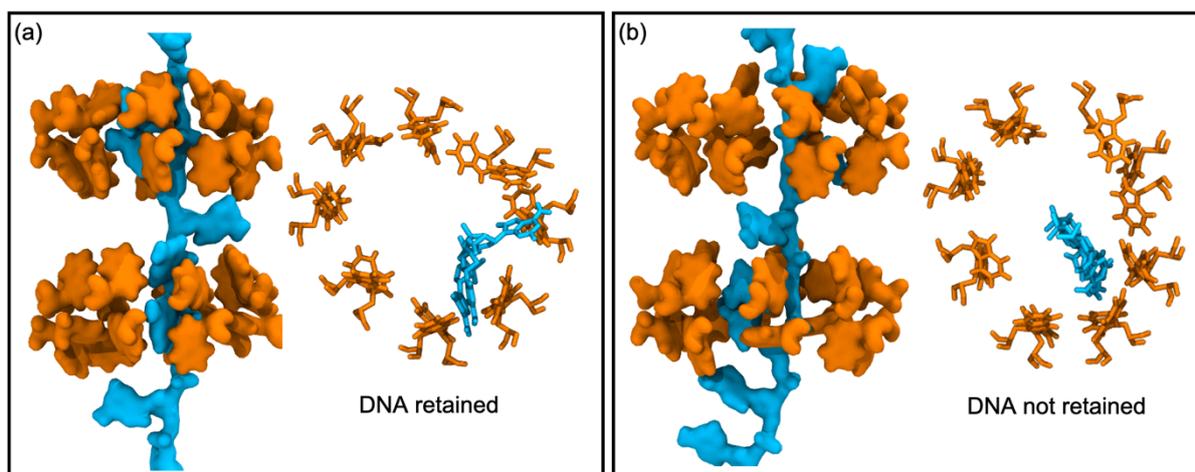


Fig. S16. Two types of ssDNA translocation are observed through pores containing aromatic residues, rapid ($\sim 40 \text{ ns}$) and much longer. Panel (a) shows DNA bases (blue) retained within a pocket formed by TRP residues (orange) in constriction 2, top and side views this leads to tethering of the DNA to this region and thus slower translocation. Panel (b) shows a scenario in which the bases are unable to move into such a pocket leading to faster translocation.

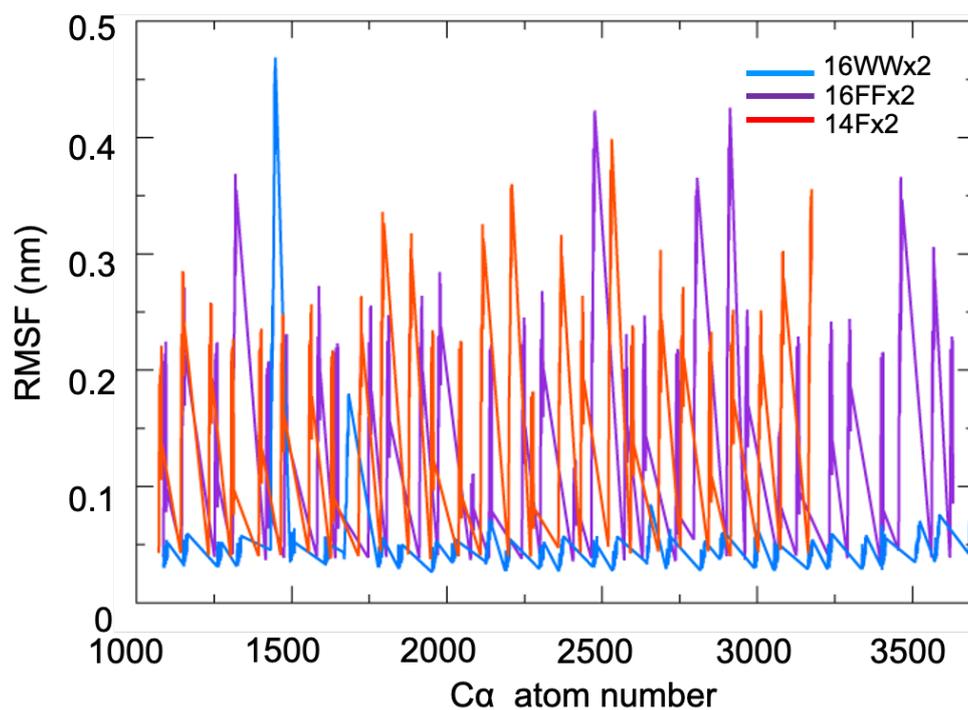


Fig. S17. Root mean square fluctuations of the side chains of aromatic residues within the constriction regions of 16WWx2, 16FFx2 and 14Fx2. There is marked lower flexibility in the TRP side chains compared to the PHE sidechains.

1. C. Skylaris, P. D. Haynes, A. A. Mostofi, M. C. Payne. *J. Chem. Phys.*, 2005, **122**, 084119.
2. J.C.A. Prentice, J. Aarons, J. C. Womack, et al. *J. Chem. Phys.*, 2020, **152**, 174111.
3. J. P. Perdew, K. Burke, and M. Ernzerhof. *Phys. Rev. Lett.*, 1996, **77**, 3865-3868.
4. S. Grimme. *J. Comput. Chem.*, 2006, **27**, 1787-1799.
5. C. Skylaris, A. A. Mostofi, P. D. Haynes, O. Dieguez, M. C. Payne. *Phys. Rev. B*, 2002, **66**, 035119.
6. S. Jo, T. Kim, V.G. Iyer, and W. Im. *J. Comput. Chem.*, 2008, **29**, 1859-1865.
7. S. Jo, X. Cheng, S.M. Islam, et al. *Adv. Protein Chem. Struct. Biol.*, 2014, **96**, 235-265.
8. M.J.S. Phipps, T. Fox, C.S. Tautermann, and C. Skylaris. *Chem. Soc. Rev.*, 2015, **44**, 3177-3211.
9. M.J.S. Phipps, T. Fox, C.S. Tautermann, and C. Skylaris. *J. Chem.*, 2016, **12**, 3135-3148.
10. P. Su and H. Li. *J. Chem. Phys.*, 2009, **131**, 014102.
11. R.Z. Khaliullin, E.A. Cobar, R.C. Lochan, A.T. Bell, and M. Head-Gordon. *J. Phys. Chem. A*, 2007, **111**, 8753-8765.