Supplemental Information

Nanopore sequencing with proteins: synchronization and dischronization of molecular dynamics simulations with laboratory and industrial developments

Zongan Wang[1,3,6,7,10,11] (王宗安), Yuqian Liu[3,4,7] (刘宇倩), Tao Zeng[1,2] (曾涛), Denghui Li[3] (李登辉), Xiao Shi[1,2] (师虓), Jiawen Zhang[1,2] (章佳文), Hao Wang[1,2] (王灏), Liuxin Shi[1,2] (史刘歆), Jiayuan Zhang[2,5] (张嘉远), Yinqi Bai[3,6] (白寅琪), Yuning Zhang[1,2,10] (张宇宁), Yuliang Dong[1,2,7,9,10] (董宇亮)

[1]State Key Laboratory of Genome and Multi-omics Technologies, BGI Research, Shenzhen 518083, China
[2]BGI Hangzhou CycloneSEQ Technology Co., Ltd., Hangzhou 310030, China
[3]BGI Research, Sanya 572025, China
[4]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[5]State Key Laboratory of Genome and Multi-omics Technologies, BGI Research, Hangzhou 310030, China
[6]Hainan Technology Innovation Center for Marine Biological Resources Utilization (Preparatory Period), BGI Research, Sanya 572025, China
[7]Shenzhen Key Laboratory of marine biology genomics, BGI Research, Shenzhen 518083, China
[8]Shenzhen Key Laboratory of Bioenergy, BGI Research, Shenzhen 518083, China
[9]Shenzhen Engineering Laboratory for Molecular Enzymology, BGI Research, Shenzhen 518083, China
[10]Correspondence: wangzongan@genomics.cn (Z. W.), dongyuliang@genomics.cn (Y. D.), zhangyuning@genomics.cn (Y. Z.).
[11]Lead contact.

Supplemental table 1. Detailed List of research articles of studying αHL related to DNA sequencing with simulations.

Ref.	Nanopore	Year	Solute	Simulated system: P + M + W + N + salt + E/F	Simulation type: Length of production run	Force field
[1]	αHL	2004	/	P (rigid) + implicit solvent + KCl + E (-150 to 150 mV)	GCMC/BD + PNP: 13.875 µs in total	CHARMM22
[2]	αHL	2005	/	Truncated P (rigid) + W + KCl + E (125 mV)	MD + PNP: 1 ns	Cornell 95[3]
[4]	αHL	2005	/	P + M (DPPC) + W + KCl + E	AAMD: ~50ns	CHARMM27
[5]	αHL	2005	poly(dA)$_{58}$	P + M (DPPC) + W + N + KCl + E	AAMD: 50ns * 2 systems	CHARMM27
[6]	αHL, nanotube	2006	ss-poly(dC)$_{45}$, (dA)$_{8}$, (dC)$_{8}$, hetero-8mers	P (CG, rigid) + implicit solvent + N (CG) + E (120 mV)	BD + PNP: 100s µs	Customized
[7]	αHL (WT & mutant)	2007	/	P + implicit solvent + KCl + E (±120 mV)	MD + PNP: NA	Cornell 95
[8]	αHL	2007	ss-poly(dA)$_{58}$, ss-poly(dA)$_{58}$, ss-poly(dAdC)$_{29}$	P + M (DPPC) + W + N + KCl + E (1.2 - 13.2 V only applied to ssDNA)	AA G-SMD: Tens of ns	CHARMM27
[9]	αHL (WT & mutant)	2009	ss-poly(A)$_{25}$, ss-poly(dC)$_{25}$, A1, dC1	P + M (POPC) + W + N + NaCl + F	cvSMD+JE: Tens of ns	CHARMM27
[10]	αHL (WT & 8 mutants)	2011	ss-poly(G)$_{6}$	Truncated P + M (methane slab) + W + N + NaCl + E (~600 mV)	AAMD: >20 ns per traj.	GROMOS96
[11]	αHL, a cylindrical pore	2012	ss-poly(dA)$_{n}$, ss-poly(dC)$_{n}$	GCMC/BD: P (rigid) + implicit solvent + N (CG) + KCl + E (50, 120, 200, 300 mV) AAMD: P + M (POPC) + W + N (CG) + KCl + E	GCMC/BD: 1 - 5 µs AAMD: 100 ns	Customized
[12]	αHL (WT & 3 mutants)	2012	ssDNA (Dickerson-Drew 12mer)	Truncated P + M (methane slab) + W + N (CG) + NaCl + E (300 mV)	MD: 100 ns * 24 trajs.	AMBER99, AMBER99 + ParmBSC0, CHARMM27, GROMOS96 53A6
[13]	αHL	2014	ss-poly(A)$_{25}$, ss-poly(dC)$_{25}$, A1, dC1	P + M (POPC) + W + N + NaCl + F	cvSMD+JE: Several 1-ns sims. ABF: 0.1 – 1 M integration steps	CHARMM22/27
[14]	αHL (WT & mutant), Si_3N_4	2014	ss-poly(dA)$_{40}$, ss-poly(dC)$_{40}$, ss-poly(dT)$_{40}$ (2 orient.)	P (rigid) + implicit solvent + N (CG) + KCl + E (250 mV)	GCMC/BD: 1 µs * 25 sims.	Customized
[15]	αHL	2015	DNA bases & phosphate group	Truncated P + M (methane slab) + W + N + NaCl + F	US: 40 µs per fragment	GROMOS 53a6
[16]	αHL	2015	CMP	P + M (DMPC) + W + N + NaCl + E (350 mV)	AAMD: 5 ns * 3600 sims. (9 dist. * 2 orient. * 20 per)	GROMOS 53a6

(Continued from previous page)

[17]	αHL (WT & 7 mutants)	2016	AAMD: ss-poly(dA)$_{20}$, ss-poly(dC)$_{20}$ GCMC/BD: ss-poly(dA)$_{40}$, ss-poly(dC)$_{40}$	AAMD: Truncated P + M (DPPC) + W + N + KCl + E (600 mV) GCMC/BD: P + implicit solvent + N (CG) + KCl + E (250 mV)	AAMD: 50 ns per sim. (10 conf. per pore, 2 orient. per ssDNA) GCMC/BD: several ms	CHARMM27, Customized
[18]	αHL, ClyA	2017	/	P + M (DMPC) + W (CG) + NaCl	MARTINI: 20 / 22 µs for ClyA / αHL AAMD: 100 ns	MARTINI v2.2
[19]	αHL	2018	/	P + M (DPhPC / POPC) + W + KCl + E (600 mV)	AAMD: 250 ns * 11 trajs.	CHARMM36
[20]	αHL	2024	ss-poly(dA)$_{10\text{-}19}$	P (CG) + M (DPPC) + W (CG) + N (CG) + NaCl + F	MARTINI: 24 µs in total	MARTINI 2.2p
[21]	αHL, AeL, lysenin	2025	/	Truncated P (CG) + M (PBD$_{11}$PEO$_{8}$ + DPhPC at 1:1 molar ratio, CG)	MARTINI: 18 µs per pore (3 traj. per. pore, 6 µs per traj.)	Martini 3
[22]	αHL	2025	ss-poly(dA)$_{10,\ 16}$, ss-poly(dC)$_{10,\ 13,\ 16}$	P (CG) + M (DPPC, CG) + W + N (CG) + NaCl + F	MARTINI: 36 µs in total	MARTINI 2.2p

Acronyms are in accordance with Figure 4: P, pore; M, membrane; W, water molecule; N, nucleic acid or nucleotide; E, electric field; F, mechanical force. Abbreviations: GCMC, Grand-canonical Monte-Carlo; BD, Brownian dynamics; PNP, Poisson-Nernst-Plank electrodiffusion algorithm; AA, all-atom; MD, molecular dynamics; CG, coarse-grained; G-SMD, grid steered molecular dynamics; WT, wide-type; cvSMD, constant velocity steered molecular dynamics; JE, Jarzynski's equality; ABF, adaptive biasing force; US, umbrella sampling; NA, not available; DPPC, 1,2-dipalmitoyl-sn-glycero-3-phosphocholine; POPC, 1-palmitoyl-2-oleoylphosphatidylcholine; DMPC, 1,2-dimyristoyl-sn-glycero-3-phosphocholine; DPhPC, 1,2-diphytanoyl-sn-glycero-3-phosphocholine (CHARMM-GUI[23] code: PHPC); AeL, aerolysin; DOPC, 1,2-Dioleoyl-sn-glycero-3-phosphocholine; PBD-PEO, poly(1,2-butadiene)-b-poly(ethylene oxide).

Supplemental table 2. Detailed list of research articles of studying MspA and CsgG related to DNA sequencing with simulations.

Ref.	Nanopore	Year	Solute	Simulated system: P + M + W + N + salt + E/F	Simulation type: Length of production run	Force field
[17]	M1-MspA & 3 mutants	2012	poly(dC), poly(dT), ssDNA (random 58mer)	P (full & truncated) + M (POPC) + W + N + KCl + E (1.2 V, 180 mV)	AAMD: ~100 µs in total	CHARMM27
[18]	M1-MspA & mutant	2015	ss-poly(dA)$_{40}$, ss-poly(dC)$_{40}$, ss-poly(dT)$_{40}$	P + implicit solvent + N (CG) + KCl + E (-120 - 180 mV)	GCMC/BD: NA	Customized
[19]	A96R-MspA, M2-MspA	2015	AMP, GMP, CMP, TMP	Truncated P + M (methane slab) + N + W + NaCl + F	US: 250 ns per window	GROMOS 53a6
[20]	M1-MspA	2016	Truncated: 11mer 5’-trans / 12mer 3’-trans for poly(dT), poly(dA), poly(dC) and mixed Full: 5’-trans ss-poly(dA)$_{40}$	Truncated P + M (POPC) + W + N + KCl + E (180 mV) Full P + phi29 DNAP + M (POPC) + W + N + dsDNA (12bp) + KCl + E (180 mV)	AAMD Truncated: >100 µs Full: 130 ns	CHARMM36
[21]	mutant of M1-MspA	2021	ss-poly(dT)$_{20}$ with one dA/dC/dG replaced at different locations	P + M (DOPC) + W + N + KCl + E (0.8 V)	AAMD: 24 ns per replica * 3 replicas * 10 systems	CHARMM36
[22]	αHL, M1-MspA, CsgG, AeL	2020	ss-poly(dA)$_{15}$	P + M (POPC) + W + N + KCl + E (300 mV)	AAMD: >100ns per system	CHARMM36
[23]	CsgG, CsgG + CsgF	2021	ss-poly(dA)$_{12}$, ss-poly(dT)$_{12}$	AAMD: P + M (POPC) + W + KCl + E (~0.9 V, ~1.6 V) SMD: P + M (POPC) + W + N + KCl + F	AAMD: 4450 ns in total SMD: 1160 ns in total	CHARMM36m

Acronyms of (P, M, W, N, E/F) are in accordance with Figure 4. Abbreviations: A/T/C/GMP, adenosine / thymidine / cytidine / guanosine monophosphate; The rest are the same as in Supplemental table 1.

References

1. Noskov SY, Im W and Roux B. Ion permeation through the alpha-hemolysin channel: theoretical studies based on Brownian dynamics and Poisson-Nernst-Plank electrodiffusion theory. *Biophys J* 2004; 87: 2299-2309. DOI: 10.1529/biophysj.104.044008.

2. Cozmuta I, O'Keeffe JT, Bose D, et al. Hybrid MD-Nernst Planck model of α-hemolysin conductance properties. *Molecular Simulation* 2005; 31: 79-93. DOI: 10.1080/08927020412331308467.

3. Cornell WD, Cieplak P, Bayly CI, et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* 1995; 117: 5179-5197. DOI: 10.1021/ja00124a002.

4. Aksimentiev A and Schulten K. Imaging alpha-hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophys J* 2005; 88: 3745-3761. DOI: 10.1529/biophysj.104.058727.

5. Mathe J, Aksimentiev A, Nelson DR, et al. Orientation discrimination of single-stranded DNA inside the alpha-hemolysin membrane channel. *Proc Natl Acad Sci U S A* 2005; 102: 12377-12382. DOI: 10.1073/pnas.0502947102.

6. Muthukumar M and Kong CY. Simulation of polymer translocation through protein channels. *Proc Natl Acad Sci U S A* 2006; 103: 5273-5278. DOI: 10.1073/pnas.0510725103.

7. James O'Keeffe IC, Deepak Bose, Viktor Stolc. A predictive MD-Nernst–Planck model for transport in alpha-hemolysin: Modeling anisotropic ion currents. *Chemical Physics* 2007; 342: 25-32. DOI: 10.1016/J.CHEMPHYS.2007.09.013.

8. Wells DB, Abramkina V and Aksimentiev A. Exploring transmembrane transport through alpha-hemolysin with grid-steered molecular dynamics. *J Chem Phys* 2007; 127: 125101. DOI: 10.1063/1.2770738.

9. Martin HS, Jha S, Howorka S, et al. 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: 2135-2148. DOI: 10.1021/ct9000894.

10. Bond PJ, Guy AT, Heron AJ, et al. Molecular dynamics simulations of DNA within a nanopore: arginine-phosphate tethering and a binding/sliding mechanism for translocation. *Biochemistry* 2011; 50: 3777-3783. DOI: 10.1021/bi101404n.

11. De Biase PM, Solano CJ, Markosyan S, et al. BROMOC-D: Brownian Dynamics/Monte-Carlo Program Suite to Study Ion and DNA Permeation in Nanopores. *J Chem Theory Comput* 2012; 8: 2540-2551. DOI: 10.1021/ct3004244.

12. Guy AT, Piggot TJ and Khalid S. Single-stranded DNA within nanopores: conformational dynamics and implications for sequencing; a molecular dynamics simulation study. *Biophys J* 2012; 103: 1028-1036. DOI: 10.1016/j.bpj.2012.08.012.

13. Martin HS, Jha S and Coveney PV. Comparative analysis of nucleotide translocation through protein nanopores using steered molecular dynamics and an adaptive biasing force. *J Comput Chem* 2014; 35: 692-702. DOI: 10.1002/jcc.23525.

14. De Biase PM, Markosyan S and Noskov S. Microsecond simulations of DNA and ion transport in nanopores with novel ion-ion and ion-nucleotides effective potentials. *J Comput Chem* 2014; 35: 711-721. DOI: 10.1002/jcc.23544.

15. Manara RM, Guy AT, Wallace EJ, et al. Free-energy calculations reveal the subtle differences in the interactions of DNA bases with alpha-hemolysin. *J Chem Theory Comput* 2015; 11: 810-816. DOI: 10.1021/ct501081h.

16. Manara RM, Tomasio S and Khalid S. The Nucleotide Capture Region of Alpha Hemolysin: Insights into Nanopore Design for DNA Sequencing from Molecular Dynamics Simulations. *Nanomaterials (Basel)* 2015; 5: 144-153. DOI: 10.3390/nano5010144.

17. Bhattacharya S, Derrington IM, Pavlenok M, et al. Molecular dynamics study of MspA arginine mutants predicts slow DNA translocations and ion current blockades indicative of DNA sequence. *ACS Nano* 2012; 6: 6960-6968. DOI: 10.1021/nn3019943.

18. De Biase PM, Markosyan S and Noskov S. BROMOC suite: Monte Carlo/Brownian dynamics suite for studies of ion permeation and DNA transport in biological and artificial pores with effective potentials. *J Comput Chem* 2015; 36: 264-271. DOI: 10.1002/jcc.23799.

19. Manara RM, Wallace EJ and Khalid S. DNA sequencing with MspA: Molecular Dynamics simulations reveal free-energy differences between sequencing and non-sequencing mutants. *Sci Rep* 2015; 5: 12783. DOI: 10.1038/srep12783.

20. Bhattacharya S, Yoo J and Aksimentiev A. Water Mediates Recognition of DNA Sequence via Ionic Current Blockade in a Biological Nanopore. *ACS Nano* 2016; 10: 4644-4651. DOI: 10.1021/acsnano.6b00940.

21. Yu M, Si W, Zeng T, et al. Unveiling the Microscopic Mechanism of Current Variation in the Sensing Region of the MspA Nanopore for DNA Sequencing. *J Phys Chem Lett* 2021; 12: 9132-9141. DOI: 10.1021/acs.jpclett.1c02414.

22. Zhou W, Qiu H, Guo Y, et al. Molecular Insights into Distinct Detection Properties of alpha-Hemolysin, MspA, CsgG, and Aerolysin Nanopore Sensors. *J Phys Chem B* 2020; 124: 1611-1618. DOI: 10.1021/acs.jpcb.9b10702.

23. Rattu P, Glencross F, Mader SL, et al. Atomistic level characterisation of ssDNA translocation through the E. coli proteins CsgG and CsgF for nanopore sequencing. *Comput Struct Biotechnol J* 2021; 19: 6417-6430. DOI: 10.1016/j.csbj.2021.11.014.