

Cite this: *Chem. Sci.*, 2021, 12, 15750

All publication charges for this article have been paid for by the Royal Society of Chemistry

Controlled movement of ssDNA conjugated peptide through *Mycobacterium smegmatis* porin A (MspA) nanopore by a helicase motor for peptide sequencing application†

Zhijie Chen,[‡] Zhenqin Wang,[‡] Yang Xu, Xiaochun Zhang, Boxue Tian^{ID} and Jingwei Bai^{ID}*

The lack of an efficient, low-cost sequencing method has long been a significant bottleneck in protein research and applications. In recent years, the nanopore platform has emerged as a fast and inexpensive method for single-molecule nucleic acid sequencing, but attempts to apply it to protein/peptide sequencing have resulted in limited success. Here we report a strategy to control peptide translocation through the MspA nanopore, which could serve as the first step toward strand peptide sequencing. By conjugating the target peptide to a helicase-regulated handle-ssDNA, we achieved a read length of up to 17 amino acids (aa) and demonstrated the feasibility of distinguishing between amino acid residues of different charges or between different phosphorylation sites. Further improvement of resolution may require engineering MspA-M2 to reduce its constriction zone's size and stretch the target peptide inside the nanopore to minimize random thermal motion. We believe that our method in this study can significantly accelerate the development and commercialization of nanopore-based peptide sequencing technologies.

Received 7th August 2021
Accepted 2nd November 2021

DOI: 10.1039/d1sc04342k

rsc.li/chemical-science

Introduction

Protein sequencing is an indispensable tool for discovering new proteins and characterizing the post-translational modifications to elucidate their functional dynamics.^{1,2} Unfortunately, current methods, such as Edman degradation and mass spectrometry, are inefficient, complicated to execute, and expensive due to instrument costs.^{3–6} Protein fluorosequencing has the advantage of reading millions of oligo-peptides simultaneously, but its detection is limited to those amino acid residues amenable to fluorophore labeling.^{7,8} Recently, the nanopore technology has emerged as a highly sensitive and inexpensive DNA sequencing method,⁹ which prompted a flurry of research aimed to apply it to protein identification.^{9–11} In fact, several studies have already demonstrated the feasibility of protein unfolding and translocation through either aerolysin or solid-state nanopores.^{12–15} Nivala *et al.* utilized the ClpX unfoldase as a motor protein to regulate the speed of polypeptide translocation through the α -hemolysin pore.^{16,17} Despite the improvement, the strategy was unable to accurately identify

individual amino acid residues due to the low spatial sensitivity of the α -hemolysin pore, as well as the overly fast and uneven translocation of the peptide strand.

Compared to well-studied nucleic acid motors, such as polymerases, helicases, translocases, motor proteins that bind to and transport polypeptides have not been extensively characterized.^{18,19} To address the poor availability of reliable peptide motor proteins, we hypothesized that nanopore-based peptide transport could be controlled by conjugating the peptide to a single-stranded DNA (ssDNA) handle under the regulation of a DNA helicase. In fact, similar strategies have previously been employed in the investigation of enzyme dynamics,²⁰ direct microRNA sequencing, and epigenetic analysis.^{21,22} We selected MspA-M2 (D90N/D91N/D93N/D118R/E139K/D134R) as the model nanopore because its constriction is located at the very bottom of the channel, which could theoretically lead to better read length.^{23,24} We demonstrated that peptides up to 17 amino acid residues could be threaded through the constriction of MspA-M2. Based on the temporal pattern of the generated ionic current, we were able to differentiate point mutations based on their charges and locations in the backbone, as well as phosphorylation sites. Further improvement could be achieved by shrinking the constriction zone and stretching the peptide chain with an enzyme or applying osmotic flow.^{25,26} Combined, these strategies may provide a feasible pathway to efficient, low-cost peptide sequencing.

School of Pharmaceutical Sciences, Tsinghua University, 100084, Beijing, China.
E-mail: jingwbai@tsinghua.edu.cn

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d1sc04342k

‡ These authors contributed equally to this work.

motor and thus resulted in continuous, rather than intermittent, strand movement. Additionally, although the length of the constriction of MspA-M2 is only 0.6 nm, which is on par with the average length of the interval between two consecutive nucleotides in a stretched DNA sequence, molecular dynamic studies have previously indicated that the neighbouring nucleotides could also occlude the nanopore passage due to thermal motion, resulting a *k*-mer length of at least 3 nucleic acid bases.^{29–31} Because the average peptide strand has a shorter unit distance and a more flexible backbone than ssDNA, a greater number of neighbouring amino acid residues can contribute to the blocking current simultaneously in nanopore protein sequencing compared to nucleotides in DNA sequencing. As a result, the increase in the *k*-mer length led to signal degeneration by smoothing what would have been step-like signatures of consecutive amino acids into a continuous slope.

Determination of sequencing read length

Since the MTA-h can only move along ssDNA and stop at the peptide conjugate site, this limits the sequencing read length, defined as the number of amino acid residues that can be pulled back through the constriction. Therefore, the read length is governed by the distance between the helicase and the constriction, which is limited to about 70 angstroms for MspA-M2, roughly equivalent to the combined length of 13 nucleotides or 17 amino acids (with the 1-nm long linker).³² To verify the read length of peptide sequencing, we synthesized a series of polypeptides in different lengths (Table S3†), and determined whether each full-length peptide could be completely led back through the constriction by monitoring the polyT induced flat current signature after the peptide signal (Fig. 2, S5†). To compare the signal profiles of different peptides, we measured the scaled ratio of the translocation-induced ionic current to the open-pore current ($I_{\text{res}}/I_{\text{open}}$, ESI†), as shown in Fig. 2. Indeed, conjugate strands in which the peptide segment was shorter than 17 amino acids produced a polyT signal at about 0.19 I_{open} in 70% to 90% of the sequencing events (Fig. 2a, c). For the 17-aa peptide, the probability of observing the polyT signal in the blocking current profile dropped to 50% (Fig S5e, f†), and the translocation of polyT was rarely observed for peptides with 18 amino acids or more (Fig. 2b, c). These results indicated that our method could detect peptides with lengths up to 17 amino acids, consistent with the structural analysis of MspA-M2.³³

To further improve the read length, we envisaged a tandem arrangement of two MTA-h units to achieve coordinated ssDNA transport over a longer distance (Fig. 3a). Because the length of the ssDNA exceeds that of a single MspA-M2 channel, it will thread through the bottom MTA-h and enter the top helicase unit before the peptide conjugate site reaches the bottom MTA-h and causes it to cease the strand transport. At this point, the top MTA-h will continue to pull the sequence upward until it stops at the ssDNA-peptide junction, ultimately leading to increased sequencing length. Unfortunately, although our initial experiments showed that the dual-MTA-h construct had a greater binding affinity for ssDNA, suggesting that both helicase units could bind to the nucleic acid sequence (Fig. S2†), we

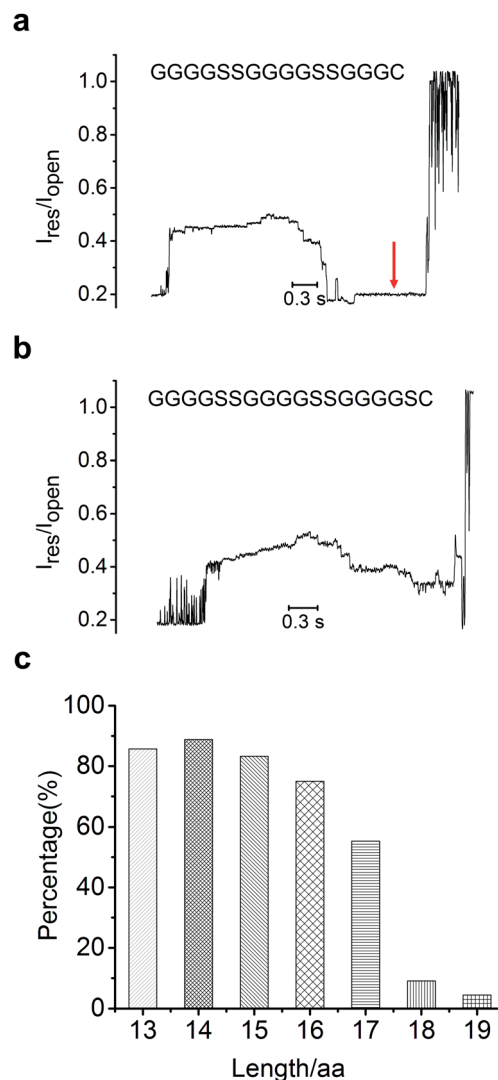


Fig. 2 Ionic current signals of peptide length of 16 amino acids (a) and 18 amino acids (b); the red arrow pointed to the signal from the polyT sequence at the 5' end of the sandwiched 3'-ssDNA-5'-N-peptide-C-3'-polyT-5' conjugates. (c) Percentage of translocation events in which the polyT signal was detected for peptides of each length. The open-pore current is typically around 150 pA under the experimental conditions employed.

did not observe an increase in read length. Interestingly instead, the blocking current profile showed two or more peptide translocation signals in a row, providing evidence that the same peptide was repeatedly sequenced (Fig. 3b). These observations combined prompted us to theorize that the strand dissociation from the bottom MTA-h at the ssDNA-peptide junction resulted in unregulated nanopore translocation; however, as the strand backtracked away from the MTA-h units, the dual-helicase system had a better chance to recapture the ssDNA and thus re-initiate the pulling process, leading to repeated sequencing of the same peptide (Fig. 3a). Of note, a similar strategy has been incorporated in the Single Molecule, Real-Time (SMRT) Sequencing platform to improve the accuracy of single-molecule DNA sequencing from about 90% to over



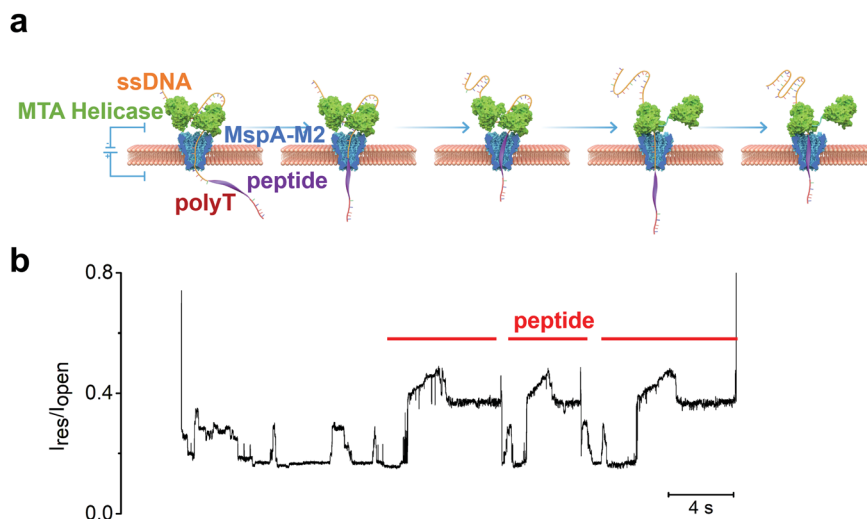


Fig. 3 (a) Illustration of the mechanism for repeated sequencing with a dual-helicase construct. (b) The experimental result showing at least three repeated peptide translocation events; the sequencing signal of ssDNA (6–89T from Table S2†) can also be observed between two high current signatures induced by peptide sequence (GGGGFGGGGSSGGGSSGGSGGC).

99%.³⁴ Therefore, the dual-MTA-h system can potentially improve the accuracy of peptide sequencing by repeatedly reading the same strand and analysing the signal consensus.³⁵

Differentiation of amino acid residue

Based on above experimental setup, we then investigated whether we could distinguish between different types of amino acids. We first mutated the serine at the fifth position (designated as S_5) from the N-terminus of the 23-aa polypeptide (GGGGSGGGSSGGGSSGGSGGC) to an array of neutral/hydrophobic amino acid residues and characterized the nanopore translocation of the resultant conjugate strands (Fig. 4, S6, S7, S8, S9†). The current profiles of the ssDNA-peptide conjugates showed a similar pattern as the control described earlier, composed sequentially of an initial steep increase to ~ 0.4 – 0.5 I_{open} depending on the specific mutations, a moderate further increase to ~ 0.5 – 0.6 I_{open} , and a subsequent drop to ~ 0.4 – 0.5 I_{open} . As a result, we could not determine which neutral/hydrophobic amino acid replaced S_5 simply by the profile of the peptide-induced current signal. This result might have stemmed from the fact that the constriction of MspA-M2 is not sufficiently narrow to allow accurate discrimination of different amino acids solely based on their molecular contours. We also found MspA-M2 to be able to accommodate the simultaneous passage of a ssDNA and a peptide strand (Fig. S12†).

On the other hand, more pronounced changes were observed when S_5 was mutated to a negatively charged amino acid residues such as an aspartate or glutamate. The current profiles of both mutants showed a "small hump" when the mutant amino acid traveled through the constriction, after which the signal slightly attenuated before reverting back to the original current profile (Fig. 4c, d). We then replaced S_5 with double aspartates or glutamates (designated as S_5DD and S_5EE) to ascertain whether this would lead to further signal augmentation. As seen in Fig. 4e and f, the current profiles of both S_5DD and S_5EE

exhibited an even more pronounced hump that now exceeded the aforementioned peak amplitude levels that we observed in S_5D and S_5E . In addition, the hump in the current of mutant G_9E , in which G_9 was replaced with glutamate, occurred later than its counterpart in S_5E or S_5D (Fig. 4h). These data suggested that the sequential way that the amino acid residues passed through the nanopore constriction played a key role in shaping the signal hump associated with the negatively charged aspartate or glutamate.¹³

Previous studies have revealed, rather counter-intuitively, that nucleotides C and T, although smaller in size than A and G, were more effective in blocking the MspA-M2 nanopore.^{28–31} For example, despite being smaller than poly-A, poly-T generated a lower ionic current when translocating through MspA-M2. Molecular dynamics simulations suggested that thymine had a greater tendency to stack inside the nanopore channel than adenine due to its smaller aromatic rings, which restricted the traverse of ions. Based on the same principle, we hypothesized that the blocking current of a peptide might also be derived mainly from the curving and folding of its backbone chain. Therefore, the observation that negatively charged amino acid residues generated greater ionic currents could be because they were driven downward in the electrical field, which stretched the peptide chain and reduced its thermal motion.³⁶ This, in turn, de-blocked the constriction and allowed the passage of more ions.

Because a phosphorylation site carries two negative charges under physiological conditions, we reasoned that its translocation through MspA-M2 should produce a similar change in ionic current as S_5DD and S_5EE . Indeed, the current of the polypeptide (GGGGSGGGSSGGGSSGGSGGC) phosphorylated at S_5 began with a steep rising, followed by an extended, high-rising hump, whose amplitude level exceeded those of S_5E and S_5D , and was on par with those of S_5DD and S_5EE (Fig. 4l). Meanwhile, for the mutant with a phosphorylated $G_9(pS)$, the



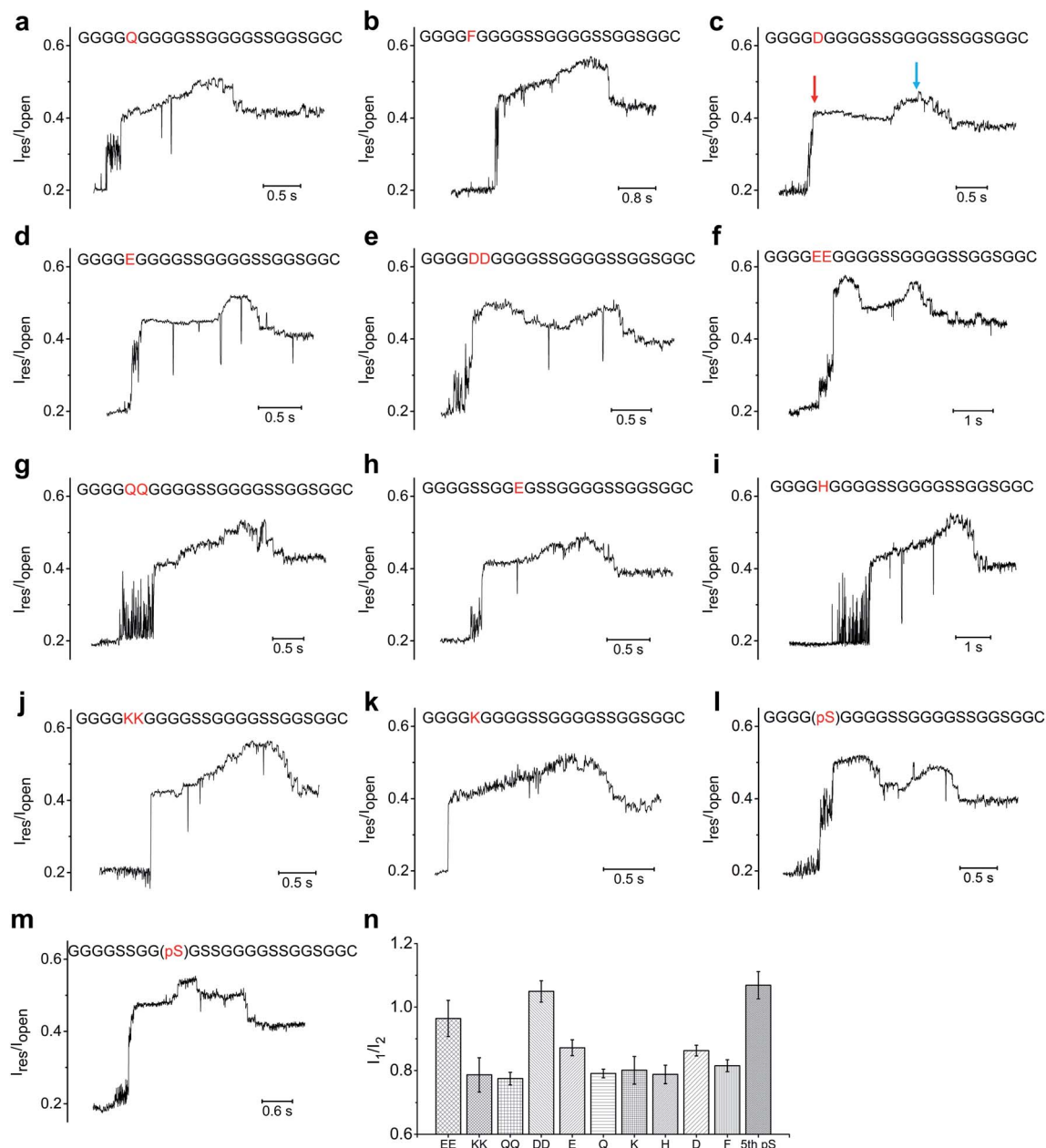


Fig. 4 The ionic current profile of different mutants on polypeptide (GGGGSGGGSSGGSSGGSGGC): (a) S₅Q; (b) S₅F; (c) S₅D; (d) S₅E; (e) S₅DD; (f) S₅EE; (g) S₅QQ; (h) S₅SS, G₉E; (i) S₅H; (j) S₅KK; (k) S₅K; (l) phosphorylation of S₅; (m) S₅SS, phosphorylation of G₉S. (n) Comparison between I_1 , which is measured based on the amplitude level of the initial steep current level rising, and I_2 , which is the amplitude level of the main peak. The blue arrows point to the main peak; the red arrows point to the type of initial current rising (hump) induced by the negatively charged mutations or phosphorylation.

signal hump moved to the middle of the peptide-induced signal region (Fig. 4m). Taken together, the data lent credence to the utility of our method in distinguishing between different phosphorylation sites, which could potentially extend to other post-translational modifications.

A careful examination of the blocking current at the ssDNA-peptide transition zone revealed a dense array of upward spikes for all the peptide mutants we tested, except for S₅R and S₅K (Fig. 5a–e). The average dwell time of each spike was shorter than 1 ms, far below the typical duration of MTA-h-induced step

motion (Fig. 5f). We reasoned that these spikes were generated by the mechanical oscillation of the ssDNA-peptide strand in resemblance to the movement of a strained spring upon release. The electrical field distribution can be roughly estimated by $E = j/\sigma$, where E is the electric field, j is the ionic current density, and σ is the conductivity. Based on the equation, the electrical field is the strongest at the constriction region, where the ionic current density is the highest.³¹ During the translocation of the conjugate strands, the ssDNA entered the pore first and behaved largely as a stretched spring due to the collective



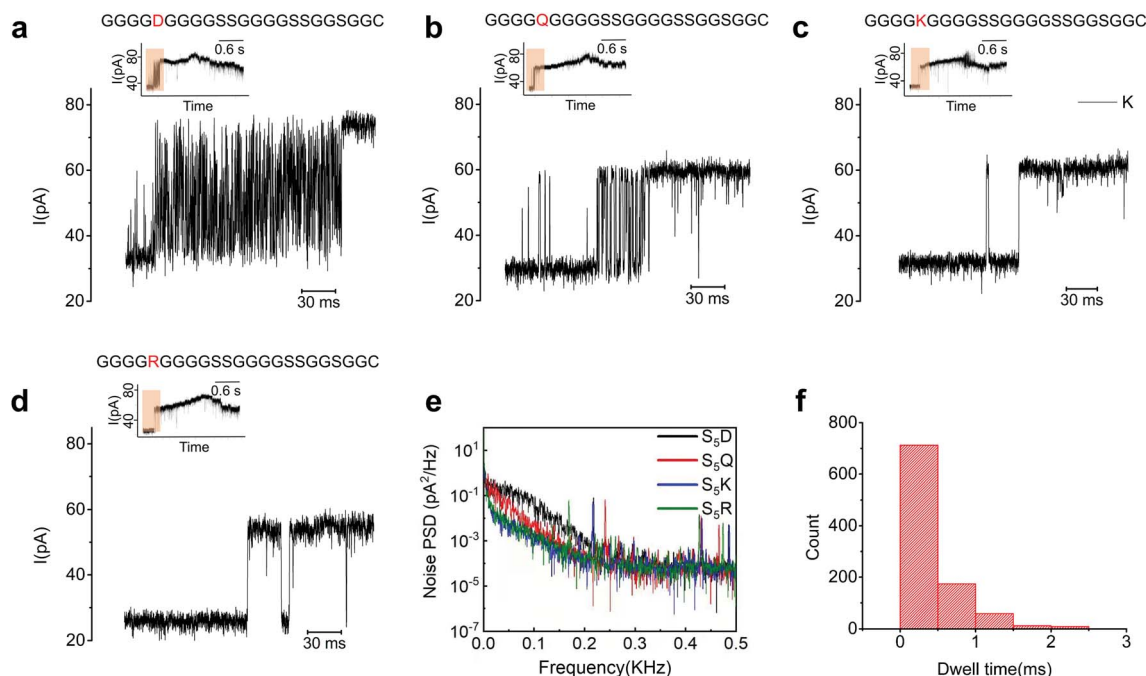


Fig. 5 The ionic current profile at the transition region. (a) S₅D and (b) S₅Q showing intense upward spikes, compared to S₅K (c) and S₅R (d) with few upward spikes. The signal transition region is marked by brown in the inserted figures. (e) Characterization of the noise for S₅D (black), S₅Q (red), S₅K (blue), and S₅R (green). Below 300 Hz, the noise power spectral density of S₅D and S₅Q is higher than S₅K and S₅R, contributed mainly by the upward spikes with frequency characteristics ranging from tens to a few hundred hertz. (f) Statistics showing that the dwell time of the upward spikes in (a) is in the range of a few milliseconds, which is much shorter than that of the helicase-induced step movement.

modulation of MTA-h and the electrical field. As the peptide segment, which carried much fewer net charges, approached the constriction, the electrostatic force attenuated, which, similar to the release of a strained spring, caused the ssDNA to oscillate. The oscillation dragged part of the peptide in and out of the nanopore in rapid succession, manifested in upward current spikes. In the case of S₅R and S₅K, the electrical field exerted an upward force on the positively charged amino acid residues, which could accelerate the peptide translocation and thus mitigate the oscillation. The absence of upward spikes was also observed when G₉ was replaced by R, but not when a position too far away from the peptide N-terminus was mutated (R₁₉), possibly due to the weakening of the electrical field (Fig. S13[†]). These findings suggested that the ionic current generated by peptide translocation is governed by a complex set of mechanisms. Further studies, including more detailed molecular dynamics simulations, are needed to elucidate the transition states of various amino acid residues inside the MspA-M2 nanopore.

Conclusion

Here, we have developed a strategy to regulate peptide translocation through the MspA-M2 nanopore and demonstrated its utility in detecting single amino acid residues. The fact that the constriction of MspA-M2 is located at the very bottom of its pore channel allows a read length of up to 17 amino acid residues. The nanopore system that we established could distinguish

between amino acid residues with different charge states, and enabled the analysis of post-translational modifications such as serine phosphorylation. We also demonstrated that by using a dual-helicase system, a single peptide could be repeatedly sequenced, potentially leading to significant improvement in sequencing accuracy. Unfortunately, unlike nanopore-based DNA sequencing, peptide translocation did not generate a clear and reproducible step-like ionic current in this study, which precluded sequential discrimination of individual amino acid residues. This observation might be due to the smoothing effect of thermal motion on the blocking current generated by different amino acid residues in an unstrained peptide. Notably, there have recently been studies that achieved step-like blocking current by applying a similar strategy to peptides with a high percentage of negatively charged amino acid residues, confirming the importance of stretching the peptide inside the nanopore.^{37,38} Potential solutions include engineering MspA-M2 to improve the spatial discrimination of its constriction zone, and stretching the peptide chain by applying electro-osmotic flow or *via* an enzymatic approach. These improvements could pave the way for our strategy to eventually achieve high-throughput, low-cost single molecule peptide sequencing, with enormous potential in scientific discovery and clinical applications.

Data availability

Further data are accessible upon request.



Author contribution

J. B. was responsible for the overall research design. Z. C. and Z. W. conducted all the experiments for the study, with input from Y. X., Z. C., Z. W., J. B. analysed experimental data, with input from X. Z and B. T., Z. C. and Z. W. generated all figures used in this study. J. B., Z. C. and Z. W. prepared the manuscript. All authors discussed the results and contributed to the final manuscript.

Conflicts of interest

A patent application has been filed by Tsinghua University, which contains the technologies described in this study, with J. B. and Z. C. listed as the inventors.

Acknowledgements

This research is financially supported by the National Natural Science Foundation of China (Grant No. 61871250), the Tsinghua-Peking Center for Life Sciences and the Ministry of Science and Technology of China (2019YFA0707000). We also thank Qitan Technology for providing the gene of MTA helicase and for advising on experiment design and execution.

References

- 1 K. A. Walsh, L. H. Ericsson, D. C. Parmelee and K. Titani, *Annu. Rev. Biochem.*, 1981, **50**, 261–284.
- 2 L. Restrepo-Pérez, C. Joo and C. Dekker, *Nat. Nanotechnol.*, 2018, **13**, 786–796.
- 3 C. V. Bradley, D. H. Williams and M. R. Hanley, *Biochem. Biophys. Res. Commun.*, 1982, **104**, 1223–1230.
- 4 H. Zhong, Y. Zhang, Z. Wen and L. Li, *Nat. Biotechnol.*, 2004, **22**, 1291–1296.
- 5 M. Miyashita, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 4403–4408.
- 6 J. R. Yates, *Nat. Methods*, 2011, **8**, 633–637.
- 7 Y. Yao, M. Docter, J. Ginkel, D. Ridder and C. Joo, *Phys. Biol.*, 2015, **12**, 05003–05010.
- 8 J. Swaminathan, A. A. Boulgakov, E. T. Hernandez, A. M. Bardo, J. L. Bachman, J. Marotta, A. M. Johnson, E. V. Anslyn and E. M. Marcotte, *Nat. Biotechnol.*, 2018, **36**, 1076–1082.
- 9 B. M. Venkatesan and R. Bashir, *Nat. Nanotechnol.*, 2011, **6**, 615–624.
- 10 P. Waduge, *ACS Nano*, 2017, **11**, 5706–5716.
- 11 M. Pastoriza-Gallego, *J. Am. Chem. Soc.*, 2011, **133**, 2923–2931.
- 12 F. Piguet, H. Ouldali, M. Pastoriza-Gallego, P. Manivet, J. Pelta and A. Oukhaled, *Nat. Commun.*, 2018, **9**, 966–978.
- 13 H. Ouldali, K. Sarthak, T. Ensslen, F. Piguet, P. Manivet, J. Pelta, J. C. Behrends, A. Aksimentiev and A. Oukhaled, *Nat. Biotechnol.*, 2020, **38**, 176–181.
- 14 D. S. Talaga and J. Li, *J. Am. Chem. Soc.*, 2009, **131**, 9287–9297.
- 15 L. Restrepo-Pérez, S. John, A. Aksimentiev, C. Joo and C. Dekker, *Nanoscale*, 2017, **9**, 11685–11693.
- 16 J. Nivala, D. B. Marks and M. Akeson, *Nat. Biotechnol.*, 2013, **31**, 247–250.
- 17 J. Nivala, L. Mulroney, G. Li, J. Schreiber and M. Akeson, *ACS Nano*, 2014, **8**, 12365–12375.
- 18 C. J. LaBreck, S. May, M. G. Viola, J. Conti and J. L. Camberg, *Front. Mol. Biosci.*, 2017, **4**, 26–38.
- 19 J. C. Whitman, B. H. Paw and J. Chung, *Hematol. Transfus. Cell Ther.*, 2018, **40**, 182–188.
- 20 I. M. Derrington, J. M. Craig, E. Stava, A. H. Laszlo, B. C. Ross, H. Brinkerhoff, I. C. Nova, K. Doering, B. I. Tickman, M. Ronaghi, J. G. Mandell, K. L. Gunderson and J. H. Gundlach, *Nat. Biotechnol.*, 2015, **33**, 1073–1075.
- 21 S. Yan, X. Li, P. Zhang, Y. Wang, H. Y. Chen, S. Huang and H. Yu, *Chem. Sci.*, 2019, **10**, 3110–3117.
- 22 J. Zhang, S. Yan, L. Chang, W. Guo, Y. Wang, Y. Wang, P. Zhang, H. Y. Chen and S. Huang, *iScience*, 2020, **23**, 100916–100948.
- 23 J. M. Carter and S. Hussain, *Wellcome Open Res.*, 2017, **2**, 23–33.
- 24 W. Zhou, H. Qiu, Y. Guo and W. Guo, *J. Phys. Chem. B*, 2020, **124**, 1611–1618.
- 25 E. L. Bonome, F. Cecconi and M. Chinappi, *Microfluid. Nanofluid.*, 2017, **21**, 96–104.
- 26 M. Boukhet, F. Piguet, H. Ouldali, M. Pastoriza-Gallego, J. Pelta and A. Oukhaled, *Nanoscale*, 2016, **8**, 18352–18359.
- 27 E. A. Manrao, *Nat. Biotechnol.*, 2012, **30**, 349–353.
- 28 A. H. Laszlo, I. M. Derrington, B. C. Ross, H. Brinkerhoff, A. Adey, I. C. Nova, J. M. Craig, K. W. Langford, J. M. Samson, R. Daza, K. Doering, J. Shendure and J. H. Gundlach, *Nat. Biotechnol.*, 2014, **32**, 829–833.
- 29 A. H. Laszlo, I. M. Derrington, B. C. Ross, H. Brinkerhoff, A. Adey, I. C. Nova, J. M. Craig, K. W. Langford, J. M. Samson, R. Daza, K. Doering, J. Shendure and J. H. Gundlach, *Nat. Biotechnol.*, 2014, **32**, 829–833.
- 30 S. Bhattacharya, I. M. Derrington, M. Pavlenok, M. Niederweis, J. H. Gundlach and A. Aksimentiev, *ACS Nano*, 2012, **6**, 6960–6968.
- 31 S. Bhattacharya, J. Yoo and A. Aksimentiev, *ACS Nano*, 2016, **10**, 4644–4651.
- 32 N. V. Bhagavan and C.-E. Ha, in *Essentials of Medical Biochemistry*, ed. N. V. Bhagavan and C.-E. Ha, Academic Press, San Diego, 2015, 2nd edn, pp. 31–51, doi: DOI: 10.1016/B978-0-12-416687-5.00004-X.
- 33 M. Faller, M. Niederweis and G. E. Schulz, *Science*, 2004, **303**, 1189–1192.
- 34 K. J. Travers, C. S. Chin, D. R. Rank, J. S. Eid and S. W. Turner, *Nucleic Acids Res.*, 2010, **38**, e159.
- 35 A. Rhoads and K. F. Au, *Genomics, Proteomics Bioinf.*, 2015, **13**, 278–289.
- 36 G. Di Muccio, A. E. Rossini, D. Di Marino, G. Zollo and M. Chinappi, *Sci. Rep.*, 2019, **9**, 6440–6447.
- 37 S. Yan, J. Zhang, Y. Wang, W. Guo, S. Zhang, Y. Liu, J. Cao, Y. Wang, L. Wang, F. Ma, P. Zhang, H.-Y. Chen and S. Huang, *Nano Lett.*, 2021, **21**, 6703–6710.
- 38 H. Brinkerhoff, A. S. W. Kang, J. Liu, A. Aksimentiev and C. Dekker, *Science*, 2021, DOI: 10.1126/science.abl4381.

