Article

Sequential reading of a stepwise-shortened peptide immobilized on nanopore

https://doi.org/10.1038/s41586-026-10881-1

Received: 9 July 2025

Accepted: 2 July 2026

Published online: 29 July 2026

Check for updates

Jialu Chen (陈佳璐)[1,2,3], Hanhan Zhang (张含含)[1,2,3], Kefan Wang (王可凡)[1,2,3], Tian Li (厉田)[1,2], Wen Sun (孙雯)[1,2], Zixuan Wang (王子璇)[1,2], Xinmeng Gao (高心萌)[1,2], Zhenyuan Cao (曹祯媛)[1,2], Yusheng Ouyang (欧阳羽升)[1,2], Lang Yao (姚琅)[1,2], Yifan Wang (王逸凡)[1,2], Yunqi Xiao (肖云麒)[1,2], Ruidong Li (李瑞冬)[1,2], Xingwang An (安兴旺)[1,2], Xinyi Dai (代馨怡)[1,2], Lulu Zhao (赵露露)[1,2], Lu Qian (钱璐)[1,2], Panke Zhang (张盼科)[1] & Shuo Huang (黄硕)[1,2]✉

Accurate decoding of peptide sequences is crucial in proteomics. However, achieving this goal is a technical challenge, owing to the compositional and structural complexity of peptides[1]. Studies inspired by the success of nanopore nucleic acid sequencing have shown that nanopore-based techniques can also be applied to peptide sequencing[2,3]. The key is to generate narrowly distributed, consistent and sequence-dependent events during sequential nanopore readout. Here we introduce a nanopore-based strategy termed transient pore analyte looping (tPAL). We develop an engineered *Mycobacterium smegmatis* porin A (MspA) nanopore that is dual modified with a nickel-ion-bound nitrilotriacetic acid (NTA-Ni) adapter and the target peptide. This distinctive sensing configuration enables precise recognition of the N terminus of the immobilized peptide by multiple re-readings. With the aid of cholesterolized aminopeptidase, the immobilized peptide can be shortened sequentially in single-amino-acid increments, yielding sequence-dependent, stepwise and narrowly distributed signal alterations that provide clues to allow peptide sequence decoding. Our strategy achieves single-amino-acid resolution and effectively identifies single-amino-acid mutations, post-translational modifications and unnatural-amino-acid insertions—indicative of its versatility in nanopore proteomics and chiral peptide analyses.

Proteins are the primary functional molecules in cells, executing a wide range of biological functions[4]. The amino acid sequence of a protein is the fundamental determinant of its structure and function. Decoding protein sequences is therefore essential for understanding biological mechanisms and pathological processes[5]. However, conventional protein-sequencing techniques, such as Edman degradation[6] and mass spectrometry[7], struggle to detect low-abundance proteins and post-translational modifications (PTMs), owing to limited detection sensitivity and dynamic range[8]. Developing single-molecule protein sequencing, which enables direct, high-resolution and PTM-sensitive analysis of individual proteins, is therefore essential[9].

Several promising technologies for single-molecule protein sequencing have emerged, including tunnelling currents[10,11], single-molecule force spectroscopy[12], reverse translation[13], fluorescence[14–17] and nanopore[18–20]. Although these methods have made considerable progresses, they still face technical challenges. Strategies based on tunnelling currents have successfully detected individual amino acids and peptides[10,11], and single-molecule force spectroscopy has achieved peptide fingerprinting[12]. But a practical peptide-sequencing method based on these principles has yet to be demonstrated. The reverse-translation strategy allows peptide sequencing by translating peptides into DNA barcodes[13], but can recognize only four amino acids because of the limited availability of specific antibodies. Fluorescence-based methods, such as those that use consecutive Edman degradation[14] or aminopeptidase digestion[15], also result in incomplete sequence coverage. For Edman-degradation-based fluorosequencing, only a few fluorescently labelled amino acids can be recognized, and errors from incomplete labelling, photobleaching or failed Edman reactions are inevitable[14]. Quantum-Si's method identifies the N terminus of a peptide using dye-labelled N-terminal amino acid recognizers[15], but the scarcity of efficient recognizers limits identification to 15 amino acids[21].

Nanopore, a powerful tool for single-molecule analysis, has achieved substantial breakthroughs in nucleic acid sequencing[22–25]. The success of nanopore-based nucleic acid sequencing has accelerated the development of nanopore-based protein sequencing. Strategies analogous to nucleic acid strand sequencing have been adapted for peptides and proteins, giving rise to approaches such as nanopore-induced phase-shift sequencing (NIPSS)[18,19,26–28], unfoldase-mediated protein sequencing[20,29–31] and electroosmotic flow-retarded protein translocation[32–34]. However, accurate decoding of peptide sequences from

[1]State Key Laboratory of Analytical Chemistry for Life Sciences, School of Chemistry, Department of Laboratory Medicine, Nanjing Drum Tower Hospital, Nanjing University, Nanjing, China. [2]Chemistry and Biomedicine Innovation Center (ChemBIC), Nanjing University, Nanjing, China. [3]These authors contributed equally: Jialu Chen, Hanhan Zhang, Kefan Wang. ✉e-mail: shuo.huang@nju.edu.cn

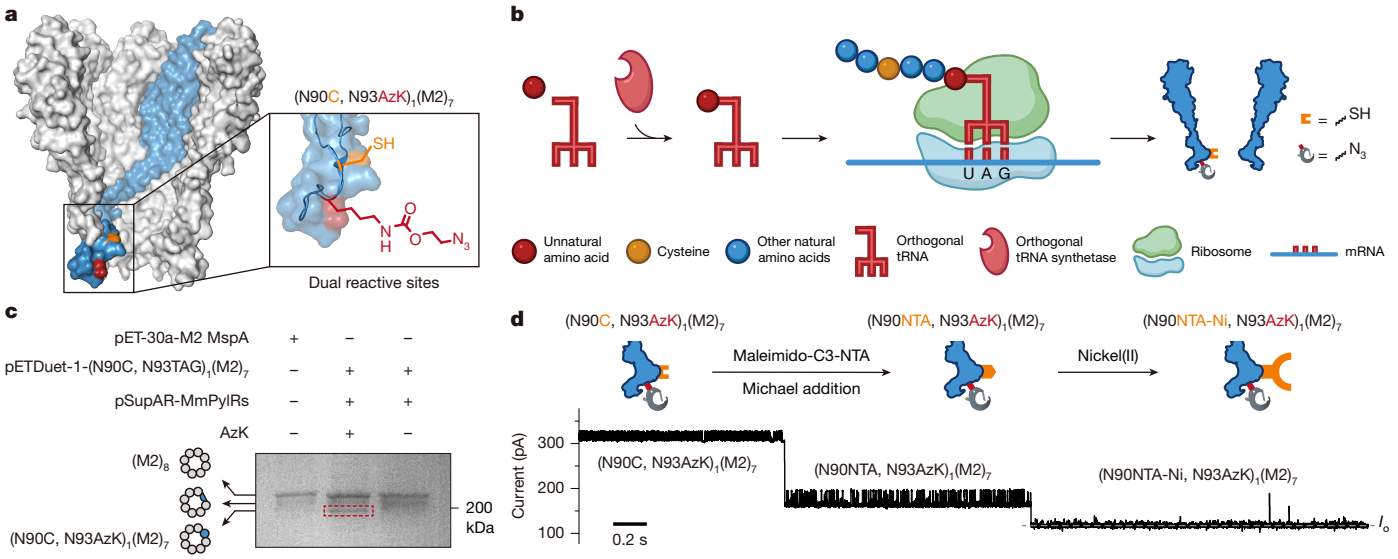


Fig. 1 | Construction of a hetero-octameric MspA containing dual reactive sites. a, Structure of $(N90C, N93AzK)_1(M2)_7$. $(N90C, N93AzK)_1(M2)_7$ is a hetero-octameric MspA composed of seven units of M2 MspA-D16H6 (grey) and one unit of N90C, N93AzK MspA-H6 (blue). The N90C, N93AzK MspA-H6 monomer contains a sole thiol group at site 90 and a sole azido group at site 93. **b**, Schematic of N^6-((2-azidoethoxy)carbonyl)-L-lysine (AzK) incorporation into a hetero-octameric MspA. In brief, an unnatural amino acid is specifically recognized by an orthogonal aminoacyl-tRNA synthetase and is attached to the orthogonal tRNA. The aminoacyl-tRNA participates in translation at the ribosome in response to the amber codon (UAG), enabling site-specific incorporation of AzK into the MspA monomer. This monomer, together with seven M2 MspA-D16H6 monomers, forms $(N90C, N93AzK)_1(M2)_7$. **c**, Gel electrophoresis. Left lane, homo-octameric M2 MspA-D16H6. Middle lane, the product when M2 MspA-D16H6 and N90C, N93AzK MspA-H6 were co-expressed and assembled. Right lane, the co-expression product without the addition of AzK. The cartoons on the left show possible hetero-octameric MspA assemblies. Grey circles represent M2 MspA-D16H6 monomers; blue circle represents a N90C, N93AzK MspA-H6 monomer; blue semicircle represents a truncated N90C, N93AzK MspA-H6 monomer. $(N90C, N93AzK)_1(M2)_7$, the desired MspA hetero-octamer, appears as the third band in the middle lane (marked with a dashed red box). For gel source data, see Supplementary Fig. 55. Three independent experiments were performed with similar results. **d**, Single-molecule observation of NTA-Ni modification. With a $(N90C, N93AzK)_1(M2)_7$ inserted, maleimido-C3-NTA was added to the *cis* chamber with a final concentration of 200 μM and nickel sulfate was added to the *trans* chamber with a final concentration of 50 μM. Two steps of current drop were sequentially observed, indicating the success of NTA modification and nickel chelation at site 90. Part of the schematic in **b** was created in BioRender; Chen, J. https://BioRender.com/vh22pfk (2026).

these nanopore signals remains challenging, owing mainly to the limited spatial resolution of nanopore sensors and the wide dispersion of corresponding nanopore events as a result of unconstrained peptide motion within the nanopore lumen.

Mycobacterium smegmatis porin A (MspA), an octameric conical nanopore renowned for its high resolution[35,36], is used extensively in single-molecule sequencing[18,19,23,37] and sensing[38–40]. Notably, a hetero-octameric MspA modified with nickel-ion-bound nitrilotriacetic acid (NTA-Ni)—termed MspA-NTA-Ni—has been shown to distinguish 20 amino acids and their post-translational modifications[41]. This finding also suggests that the N terminus of a peptide can interact with the NTA-Ni adapter, thereby enabling accurate discrimination of various peptides through stochastic sensing. Building on this insight, here we develop a strategy in which a peptide is irreversibly attached to an MspA nanopore equipped with an NTA-Ni adapter, which provides structural constraint and allows multiple re-readings of the same peptide. With the assistance of aminopeptidase, the immobilized peptide is shortened in a stepwise manner, with each step of cleavage corresponding to changes in the nanopore events—thus generating sequence-dependent peptide events.

A hetero-MspA with dual reactive sites

To achieve dual modifications within a single nanopore, a hetero-octameric MspA bearing two bioorthogonal reactive sites was first synthesized using genetic codon expansion (GCE)[42] (Methods). Experimentally, two genes encoding, respectively, M2 MspA-D16H6 and N90C, N93TAG MspA-H6 were inserted into a pETDuet-1 co-expression vector (Supplementary Table 1 and Supplementary Fig. 1). After expression with N^6-(2-azidoethoxy)carbonyl-L-lysine (AzK), octameric MspA assemblies with varying ratios of two types of monomer were formed. The desired MspA assembly $(N90C, N93AzK)_1(M2)_7$, comprising seven M2 MspA-D16H6 monomers and one N90C, N93AzK MspA-H6 monomer (Fig. 1a), was purified from the octamer mixture by gel separation (Fig. 1c and Supplementary Fig. 2a,b).

The $(N90C, N93AzK)_1(M2)_7$ nanopore has two bioorthogonal reactive sites—a thiol group at position 90 and an azido group introduced by GCE in response to the amber codon at position 93 (Fig. 1b)—allowing dual modifications within the same pore. The thiol group at the pore's constriction can be used for modifying a reactive adapter via a Michael addition reaction[39,41]. Concurrently, the azido group situated near the pore constriction acts as an anchoring site for the target analyte through a click reaction.

$(N90C, N93AzK)_1(M2)_7$ was characterized by single-channel recording. The results of spontaneous pore insertions (Supplementary Fig. 2c) and current–voltage (I–V) curves (Supplementary Fig. 3 and Supplementary Table 2) indicate that the introduced AzK of site 93, which is located outside the pore constriction, does not notably affect the pore's insertion ability and the open pore current. An NTA-Ni adapter was first introduced to position 90 of the nanopore by mixing $(N90C, N93AzK)_1(M2)_7$ with maleimido-C3-NTA and nickel sulfate (Methods). The resulting chelate is termed $(N90NTA\text{-}Ni, N93AzK)_1(M2)_7$. The stepwise modification was also monitored in real time by single-channel recording (Fig. 1d). The success of $(N90NTA\text{-}Ni, N93AzK)_1(M2)_7$ construction could be confirmed by the characteristics of its open pore current (Supplementary Fig. 4).

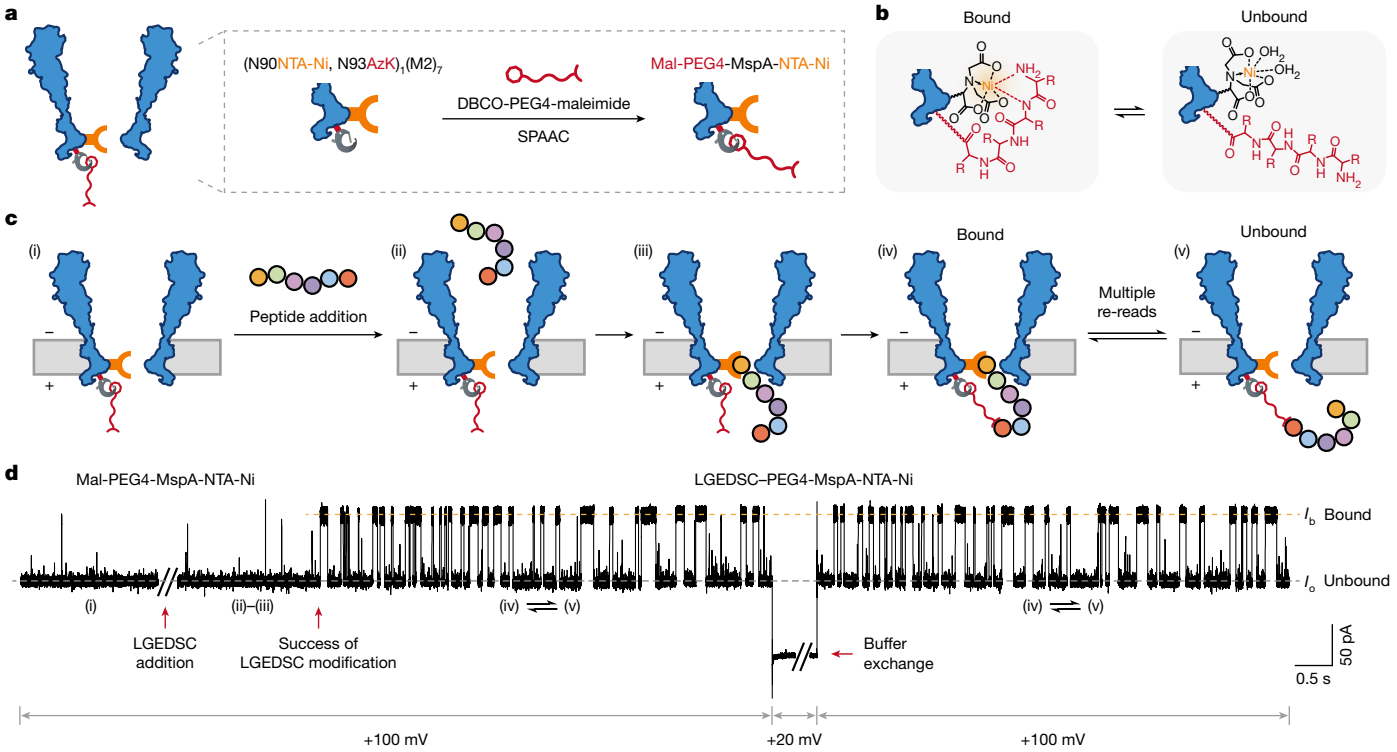


Fig. 2 | Multiple re-reads of a single peptide by tPAL sensing. a, Construction of Mal-PEG4-MspA-NTA-Ni. The azido group of (N90NTA-Ni, N93AzK)$_1$(M2)$_7$ reacts with a DBCO-PEG4-maleimide through a strain-promoted alkyne-azide cycloaddition (SPAAC) reaction to generate Mal-PEG4-MspA-NTA-Ni. **b**, tPAL sensing mechanism. When a peptide is immobilized to the nanopore, its N terminus can reversibly bind to the NTA-Ni adapter to repetitively report characteristic binding events. **c**, Construction of peptide–PEG4-MspA-NTA-Ni. With a Mal-PEG4-MspA-NTA-Ni inserted (i), peptides containing a C-terminal cysteine are added to the *cis* chamber (ii). When the N terminus of the peptide is transiently grabbed by the NTA-Ni adapter (iii), the C-terminal cysteine can react with the maleimide group at site 93 (iv), forming a peptide-conjugated MspA nanopore. The N terminus of the immobilized peptide can still reversibly interact with the NTA-Ni adapter (iv and v), allowing multiple re-reads of the same peptide. **d**, Real-time conjugation of LGEDSC peptide and Mal-PEG4-MspA-NTA-Ni. The measurements were performed as described in the Methods. LGEDSC was added to the *cis* chamber with a final concentration of 20 μM. The success of LGEDSC immobilization reports reversible current-switching events (iv and v). After that, the buffer in *cis* was exchanged with fresh buffer to remove all remaining peptides. The switching did not disappear after buffer exchange, confirming the success of LGEDSC immobilization to the pore. I_o stands for the current level of the unbound state of LGEDSC–PEG4-MspA-NTA-Ni, which is consistent with the open pore current of Mal-PEG4-MspA-NTA-Ni. I_b stands for the current level of the bound state of LGEDSC–PEG4-MspA-NTA-Ni. The average time required for the conjugation of LGEDSC at a final concentration of 20 μM is 25 ± 8 s (n = 5), and this can be further accelerated by increasing the peptide concentration.

Multiple re-reads of a single peptide

To facilitate peptide immobilization, (N90NTA-Ni, N93AzK)$_1$(M2)$_7$ was subsequently treated with dibenzocyclooctyne (DBCO)-PEG4-maleimide (Fig. 2a and Methods). The resulting complex, termed Mal-PEG4-MspA-NTA-Ni, bears a maleimide group at position 93, allowing the attachment of a cysteine-containing peptide via a Michael addition reaction. Because the modification site is positioned outside the pore constriction, the incorporation of DBCO-PEG4-maleimide does not induce a notable change in pore conductance (Supplementary Table 3 and Supplementary Fig. 4).

The conjugation of the peptide was observed in real-time using single-channel recording. With a single Mal-PEG4-MspA-NTA-Ni inserted, an open pore current, I_o, was first recorded (state i; Fig. 2c,d). After the peptide was added to the *cis* chamber, the N terminus of a single peptide was first captured by the NTA-Ni adapter, bringing the peptide's thiol group into proximity with the maleimide group on the nanopore (states ii and iii; Fig. 2c,d). This assists the conjugation of the peptide to the pore, forming peptide–PEG4-MspA-NTA-Ni. For clarity, the detailed nomenclature of this configuration is also illustrated in Supplementary Fig. 5.

Here, an LGEDSC peptide was used as the model. The successful modification of LGEDSC resulted in frequent transitions between two current states, corresponding to the bound and unbound states of a single LGEDSC with the NTA-Ni adapter (states iv and v; Fig. 2c,d). Peptide events remained stable after buffer exchange in the *cis* chamber, confirming that the peptide was irreversibly immobilized on the nanopore (Supplementary Fig. 6). However, peptide immobilization was not achieved in the absence of DBCO-PEG4-maleimide modification (Supplementary Fig. 7).

In the case of a peptide–PEG4-MspA-NTA-Ni, the terminal amino-N and the amide-N of the immobilized peptide can form a five-membered chelate with the NTA-Ni adapter, creating a transient pore analyte loop (Fig. 2b). We term this sensing strategy 'transient pore analyte looping' (tPAL). The repetitive binding and dissociation of the immobilized peptide with the NTA-Ni adapter during tPAL sensing facilitate multiple re-readings of the same peptide, enabling precise recognition of its N terminus.

To quantitatively characterize peptide events, key parameters such as the open pore current (I_o), event current (I_b), blockage amplitude ($\Delta I = I_b - I_o$), event noise standard deviation (SD), event dwell time (t_{off}) and inter-event duration (t_{on}) are defined in Supplementary Fig. 8. A significantly higher event appearance rate was observed with tPAL than with stochastic sensing, owing to the proximity of the immobilized LGEDSC to the NTA-Ni adapter (Supplementary Fig. 9). Even at an extremely low applied voltage of +5 mV, tPAL events were still observable (Supplementary Fig. 10), indicating that their generation is governed by spontaneous chemical reactions. Electrophoretic forces and electroosmotic flow, however, might also act to regulate the event appearance frequency.

Peptide–PEG4-MspA-NTA-Ni can also be batch-prepared by mixing (N90NTA-Ni, N93AzK)$_1$(M2)$_7$ with DBCO-functionalized peptides. A variety of pre-mixed peptide–PEG4-MspA-NTA-Ni can also spontaneously insert into the membrane (Supplementary Figs. 11 and 12). However, the insertion efficiency is significantly reduced. Thus, all peptide–PEG4-MspA-NTA-Ni used in subsequent experiments was assembled in real time, unless specified otherwise.

To assess the versatility of tPAL sensing, two additional peptides differing at the N terminus were examined. Similar to the observations with LGEDSC, high-frequency tPAL events were clearly observed for RGEDSC and DGEDSC (Supplementary Fig. 13a–c). The events of these peptides were clearly discernible (Supplementary Fig. 13d,e) and showed excellent reproducibility (Supplementary Fig. 14 and Supplementary Table 5). A comparison of tPAL events and those observed with stochastic sensing (Supplementary Fig. 15) showed that they had highly similar characteristics.

Furthermore, various linkers and conjugation sites were investigated (Supplementary Figs. 16 and 17). The results indicate that the peptide–PEG4-MspA-NTA-Ni configuration, offering a higher resolution and event appearance rate, is optimum for tPAL sensing. Therefore, all subsequent experiments were done using this configuration.

Single-amino-acid resolution

After establishing the tPAL methodology, its sensitivity towards single-amino-acid mutations was investigated further. Twenty XTRSC peptide variants, with mutations introduced at their N terminus, were tested. Representative events of these peptides obtained at +100 mV are shown in Extended Data Fig. 1b. All events exhibit clear characteristics, most of which can be directly distinguished by visual observation. Notably, CTRSC, HTRSC and PTRSC each generated two types of sensing events, distinguishing them from other peptides. To achieve unambiguous discrimination of all 20 XTRSC peptides, a multi-voltage sweeping was performed (Extended Data Fig. 1a). It uses the anchoring capability of tPAL to enable repeated measurements of the same peptide under different voltage conditions. Even if some peptides exhibit similar signals at a certain voltage, differences might emerge in other voltages, thereby enabling their unambiguous discrimination. Experimentally, each peptide was subjected to programmed cycling voltage recordings at +60 mV, +100 mV and +140 mV. Representative events of these peptides at different voltages are presented in Supplementary Figs. 18–23. The core event parameters at different voltages are summarized in Supplementary Table 6.

A corresponding machine-learning analysis was performed with combined multi-voltage data. Among all classifiers, the quadratic support vector machine (SVM) model achieved the highest validation accuracy of 98.0% (Supplementary Table 7). By contrast, when using a dataset from a single voltage (+60 mV, +100 mV or +140 mV), the validation accuracy decreased to 92.9%, 90.1% and 86.1%, respectively. The true positive rate for most peptides improved significantly when multi-voltage data were included (Supplementary Fig. 24). Feature importance analysis further showed that all three voltages contributed significantly to discrimination (Supplementary Table 8). In addition, the multi-voltage sensing data of all 20 XTRSC peptides were processed using the *t*-distributed stochastic neighbour embedding (t-SNE) algorithm, reducing the dimensionality from 27 to 2 for visualization (Extended Data Fig. 1d). The results clearly demonstrate the high resolution of tPAL technology and the effectiveness of multi-voltage sweeping. However, for the sake of clarity, all subsequent experiments present only data acquired at +100 mV, unless otherwise stated.

Stepwise-shortened peptides

Inspired by the tPAL sensing mechanism, in which the event feature is typically dominated by the N-terminal region, we hypothesized that sequence-dependent events could be generated through the sequential cleavage and recognition of the peptide's N terminus. To evaluate the feasibility of this approach, we tested a series of synthetic peptides mimicking progressively shortened sequences, ranging from LRGESC to C (Extended Data Fig. 2a–g). Six distinct event types were effectively differentiated (Extended Data Fig. 2h,i and Supplementary Figs. 25 and 26), with a machine-learning accuracy of 99.5% (Supplementary Fig. 27 and Supplementary Table 9). Notably, cysteine had the longest event dwell time and the shortest inter-event duration, compared with the other peptides (Extended Data Fig. 2j and Supplementary Table 5). Furthermore, the difference between tPAL-sensing and stochastic-sensing events became increasingly pronounced as the peptide length decreased (Supplementary Fig. 28). This is likely to result from the increased contribution of the C-terminal PEG linker to the signal as the peptide shortens.

To determine the dynamic range of peptide length by tPAL, additional peptides, including Pep12K (LKLRGDDGSESC), Pep14 (LMLMLRGDDGSESC) and Pep22 (LMLMLKRDDGSDASRGDKGESC), were tested in the same conditions (Supplementary Fig. 29). Event frequency generally decreases with increasing peptide length (Supplementary Fig. 30), whereas other intrinsic characteristics, such as sequence composition, charge and conformation, can also influence peptide binding kinetics. These findings imply that the tPAL method has the potential to detect longer peptides, provided that they have an accessible N terminus. Moreover, peptides can be attached to the nanopore not only through a C-terminal cysteine residue but also through a cysteine residue in the middle region, as shown with Pep16 (LMLMLRGDDGSESCGG) (Supplementary Fig. 31). For long peptides, downward spikes resulting from transient translocation of the tethered peptide were observed in addition to characteristic tPAL events, but these did not interfere with measurements (Supplementary Figs. 32 and 33).

Sequencing by enzymatic cleavage

Streptomyces griseus aminopeptidase (sgAP)[43] was used to catalyse the sequential cleavage of the immobilized peptide in real time. It is anticipated to specifically cleave the N-terminal amino acid when the immobilized peptide is released from the NTA-Ni adapter, shortening the peptide by one amino acid. The newly exposed N terminus is then recognized by the NTA-Ni adapter, yielding distinct binding-event features to report successful cleavage. Consequently, sequence-dependent events resulting from the stepwise enzymatic digestion were observed (Fig. 3b). However, the mobile aminopeptidase in the *trans* chamber has limited access to the immobilized peptide, leading to low cleavage efficiency. To mitigate this issue, sgAP was modified with cholesterol to enhance its enrichment on the lipid membrane (Methods, Supplementary Table 10 and Extended Data Fig. 3). This complex, referred to as sgAP-dA15-chol (Fig. 3a), does not affect the intrinsic activity of the aminopeptidase. However, it increases the local enzyme concentration, thereby enhancing the cleavage efficiency. In addition, sgAP-dA15-chol does not interact with the NTA-Ni adapter, thus avoiding interference in event acquisition (Supplementary Fig. 34).

To experimentally validate this strategy, a heptapeptide (LEFTRSC) was subjected to analysis. In the presence of sgAP-dA15-chol, the heptapeptide exhibited a series of sequence-dependent events during stepwise cleavage, including seven distinct event stages labelled as aa1–aa7 (Fig. 3c,f). By contrast, using mobile sgAP, only the aa1 and aa2 stages were observed, and no further enzymatic cleavage was detected during extended monitoring (Extended Data Fig. 4 and Supplementary Table 11), emphasizing the improved efficacy of the sgAP-dA15-chol. Notably, the binding events at the aa7 stage, characterized by extended dwell times and short inter-event intervals, were identical to those previously characterized by cysteine (Extended Data Fig. 2g), indicating that the peptide digestion has been completed and the observed

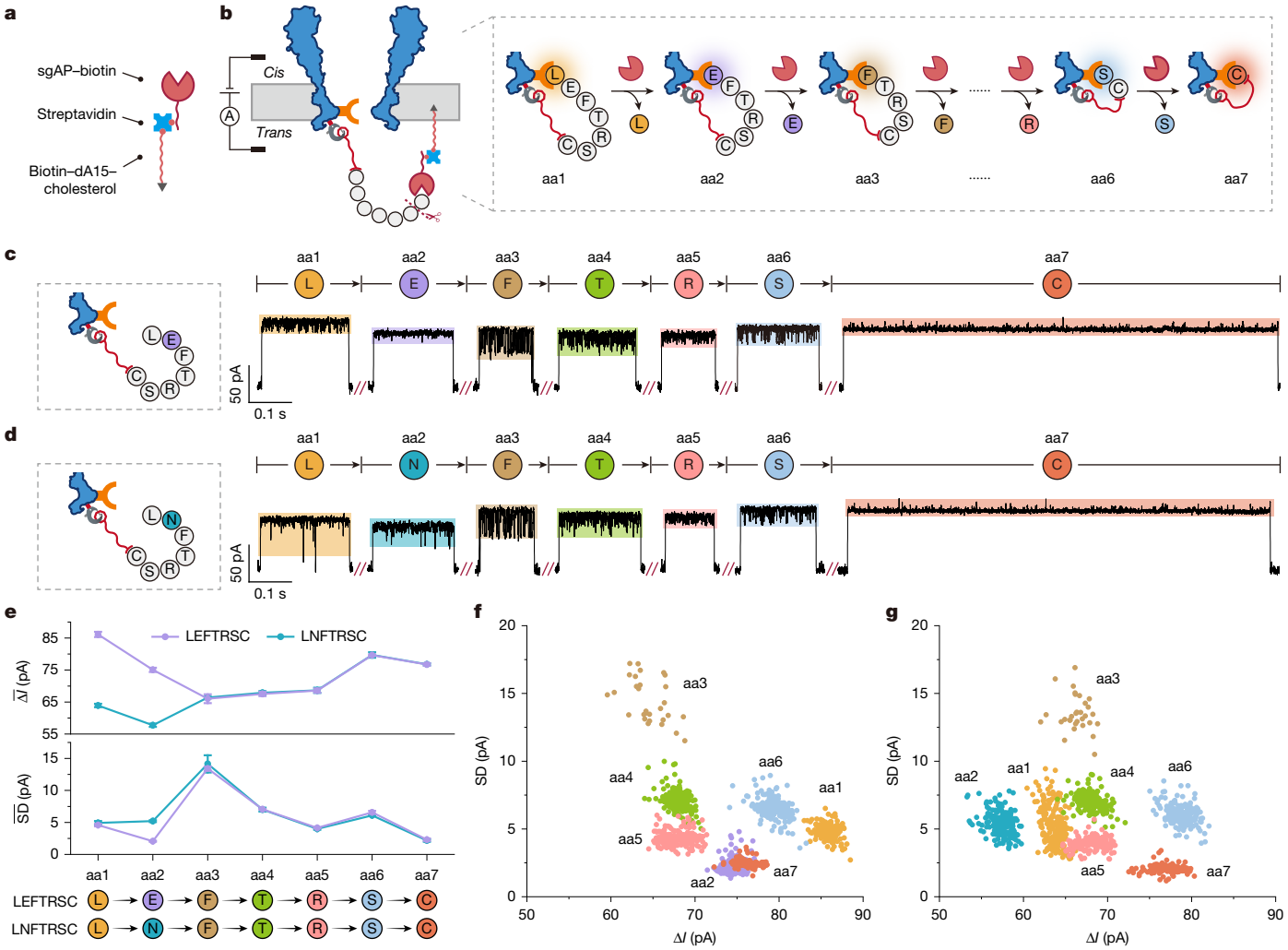


Fig. 3 | Production of sequence-dependent events by real-time enzymatic digestion. a, Composition of sgAP-dA15-chol. sgAP-dA15-chol is a ternary complex comprising sgAP–biotin, streptavidin and biotin–dA15–cholesterol. sgAP–biotin is a biotinylated aminopeptidase generated by BirA (Methods). Biotin–dA15–cholesterol is an oligonucleotide containing 5′ biotin and 3′ cholesterol modifications. **b**, Schematic depicting the sequential cleavage of the conjugated peptide. The sgAP-dA15-chol added to the *trans* chamber can spontaneously insert into the membrane, allowing the enrichment of aminopeptidases near the peptide-conjugated nanopore. Aminopeptidase can specifically cleave the N-terminal amino acid of the peptide, shortening the peptide by one amino acid at a time. Successive cleavages of the same peptide result in sequential exposure of amino acids for recognition with the NTA-Ni adapter, producing characteristic signal changes that are sequence dependent. **c**,**d**, Representative events obtained by sequential cleavage of LEFTRSC (**c**) and LNFTRSC (**d**). The measurements were performed as described in the Methods. For each peptide, six steps of enzymatic cleavages were observed. Consequently, seven stages, referred to as aa1–aa7, were observed. Events produced at each stage are marked with corresponding colour bands. The corresponding N-terminal amino acid of each stage is labelled above each event. **e**, Plots of $\overline{\Delta I}$ (top) and $\overline{SD}$ (bottom) at different stages. Data are mean ± s.d. from three independent measurements. The cartoons below the plot illustrate the amino acid sequence of each peptide. The plots converged once the mutated sites of both peptides were cleaved. **f**,**g**, Scatter plots of Δ*I* versus SD of results acquired with LEFTRSC (**f**) and LNFTRSC (**g**). The event clusters of aa1–aa7 are labelled with corresponding colours.

count of distinct stages is consistent with the length of the peptide. The result of TRSC sequential cleavage was identical to the aa4–aa7 stages acquired from LEFTRSC (Supplementary Fig. 35), further confirming that the observed changes in events are indeed due to enzymatic cleavages of the bound peptide.

At each stage, the N terminus of the peptide can be read repeatedly, efficiently accumulating a large number of events for downstream data analysis. In contrast with other nanopore-based peptide resampling strategies[19,20], the tPAL strategy is strictly unidirectional, and peptide backstepping is impossible. Unlike resistive pulse sensing, a tPAL event is generated only when the peptide N terminus is chemically constrained, resulting in more consistent event features that are highly desirable for downstream sequence decoding.

To further ascertain the resolution and generalization of the tPAL strategy, an additional peptide, LNFTRSC, was subjected to analysis. Seven distinct stages were observed throughout the sequential cleavage process, culminating in the recognition of cysteine (Fig. 3d,g). Three separate trials acquired with LEFTRSC or LNFTRSC are shown in Supplementary Figs. 36 and 37, affirming the reliability and consistency of the sequential cleavage observed by tPAL. Both heptapeptides are identical in their amino acid sequence except for the second position from the N terminus. Therefore, after the cleavage of the second residue, all remaining sequences of the heptapeptides align, leading to identical event patterns in all subsequent stages (Fig. 3e and Supplementary Table 13).

Peptides that contain histidine or proline are unique in that they generate two distinct types of event. After the cleavage of histidine or proline, these unique event features disappear immediately, providing an indicator for determining that histidine or proline was present in the previous stage of observation (Supplementary Figs. 38–40).

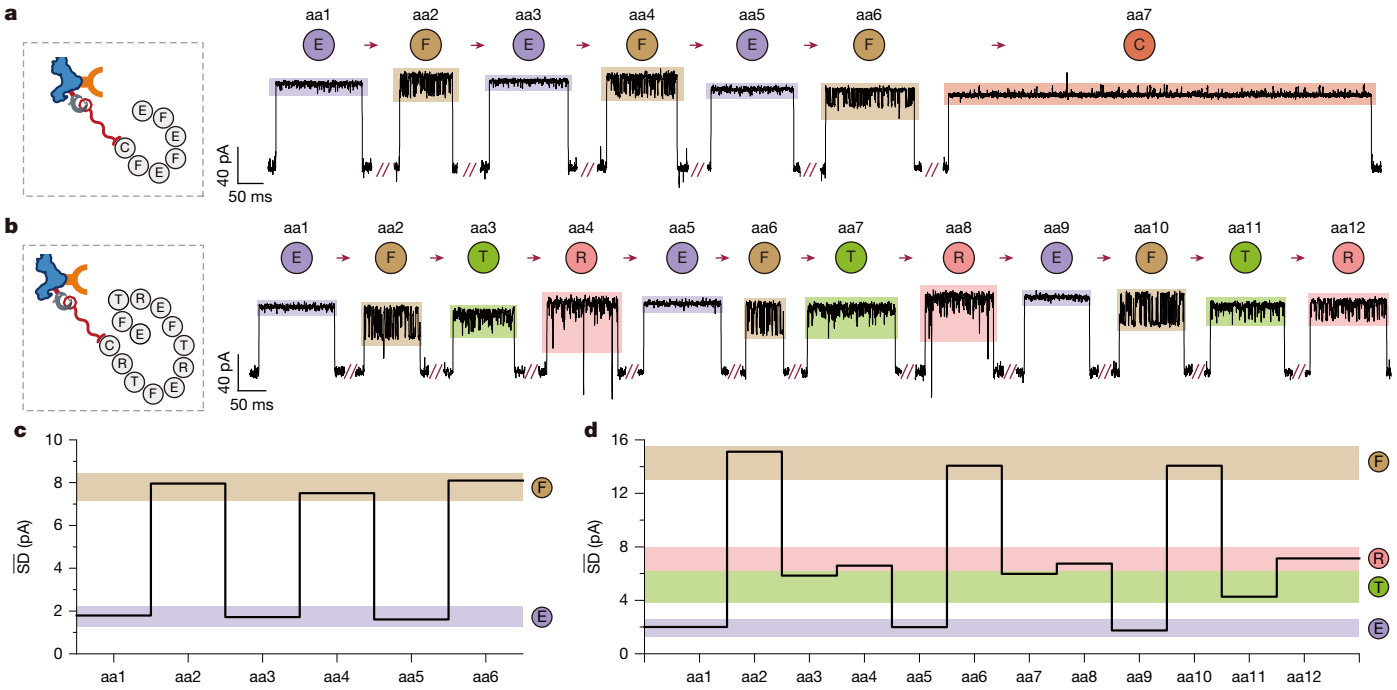


Fig. 4 | Sequential cleavage of peptides containing tandem repeats. **a**,**b**, Representative events obtained by sequential cleavage of EFEFEFC (**a**) and EFTREFTREFTRC (**b**). All measurements were performed as described in the Methods. The corresponding N-terminal amino acid of each stage is marked above each trace. **c**, Plot of $\overline{SD}$ at the aa1–aa6 stages, derived from sequential cleavage of EFEFEFC. **d**, Plot of $\overline{SD}$ at the aa1–aa12 stages, derived from sequential cleavage of EFTREFTREFTRC.

sgAP demonstrates broad specificity; it is capable of cleaving not only neutral amino acids but also acidic and basic ones, including proline at the N terminus. However, it exhibits low enzymatic activity towards specific motifs, such as the penultimate proline and glycine–glutamic acid (Supplementary Figs. 40–42). Owing to its enzymatic selectivity, the digestion time for each amino acid varies (Supplementary Table 14). These hydrolysis kinetics can also serve as a characteristic feature for amino acid recognition.

Analysis of peptides with PTMs and UAAs

To demonstrate the identification of PTMs using tPAL, we tested Pep12Mo (L(Mo)LRGDDGSESC), Pep12Kac (L(Kac)LRGDDGSESC) and Pep12Kme3 (L(Kme3)LRGDDGSESC), which contain oxidation, acetylation and trimethylation modifications, respectively (Extended Data Fig. 5 and Supplementary Fig. 43). All peptides exhibited three cleavages, resulting in four distinct stages. However, owing to the limited efficiency of sgAP on subsequent sequences, no further enzymatic cleavage on RGDDGSESC was detected. Notably, lysine trimethylation and acetylation, which are a pair of PTMs with nearly identical mass, can be clearly differentiated by tPAL (Supplementary Fig. 43). Compared with their unmodified counterparts Pep12M (LMLRGDDGSESC) and Pep12K (LKLRGDDGSESC), significant differences in binding events were observed at aa1 and aa2, whereas the event features of aa3 or aa4 were indistinguishable between the modified and the unmodified peptides (Extended Data Fig. 5c,d,g,h and Supplementary Table 15), consistent with PTM localization at the second position. The consistency of these results was confirmed by carrying out three independent trials for each peptide (Supplementary Figs. 44–48).

A phosphorylated heptapeptide, L(pY)FTRSC, and its reference peptide, LYFTRSC, were also subjected to testing. However, owing to the resistance of the phosphorylated residue to sgAP, complete enzymatic cleavage of L(pY)FTRSC was not achieved. Despite this, when phosphorylated tyrosine was exposed at the N terminus, the resulting sensing events exhibited characteristic low noise and high current amplitude, enabling accurate identification of the phosphorylation site. An enzyme cocktail strategy was applied for phosphopeptide measurement (Supplementary Fig. 49). When the phosphorylation site reached the N terminus, lambda phosphatase was introduced to specifically hydrolyse the phosphate group at the N-terminal residue, thereby enabling sgAP to continue cleaving the peptide to achieve full-length tPAL measurement. Although not investigated in detail in this study, other PTMs such as sulfation, hydroxylation, and glycosylation can also be detected using this method.

To assess the applicability of the tPAL strategy to analyses of peptides containing unnatural amino acids (UAAs), two model peptides, F(NA)TRSC and F(AzK)TRSC, were investigated. F(NA)TRSC is a hexapeptide with a naphthylalanine residue (Extended Data Fig. 6a), whereas F(AzK)TRSC contains an azido-lysine at the equivalent position (Extended Data Fig. 6c). Using sgAP-dA15-chol, both peptides were read sequentially until cysteine was reached (Extended Data Fig. 6b,d). This underscores the potential of the tPAL strategy for identifying UAAs.

The tPAL strategy is, in principle, also suitable for D-peptides. To validate this, a series of synthetic D-peptides, including D-FYAESC, D-YAESC and D-AESC, which represent progressively shortened peptides starting from D-FYAESC, was analysed using tPAL. Notably, all of these peptides exhibited fully distinguishable events (Extended Data Fig. 7). These results suggest that sequential cleavage and recognition of D-peptides could be achieved through the development of D-aminopeptidases in the future. However, because of the inaccessibility of corresponding D-peptide exopeptidases, sequential enzymatic cleavage of a D-peptide is not demonstrated here.

Peptides of repeating sequence

To show the strong correlation between tPAL events and peptide sequence, a series of peptides composed of tandem repeats was subjected to tPAL measurements. For instance, the peptide EFEFEFC exhibited an alternating pattern of low- and high-noise events during

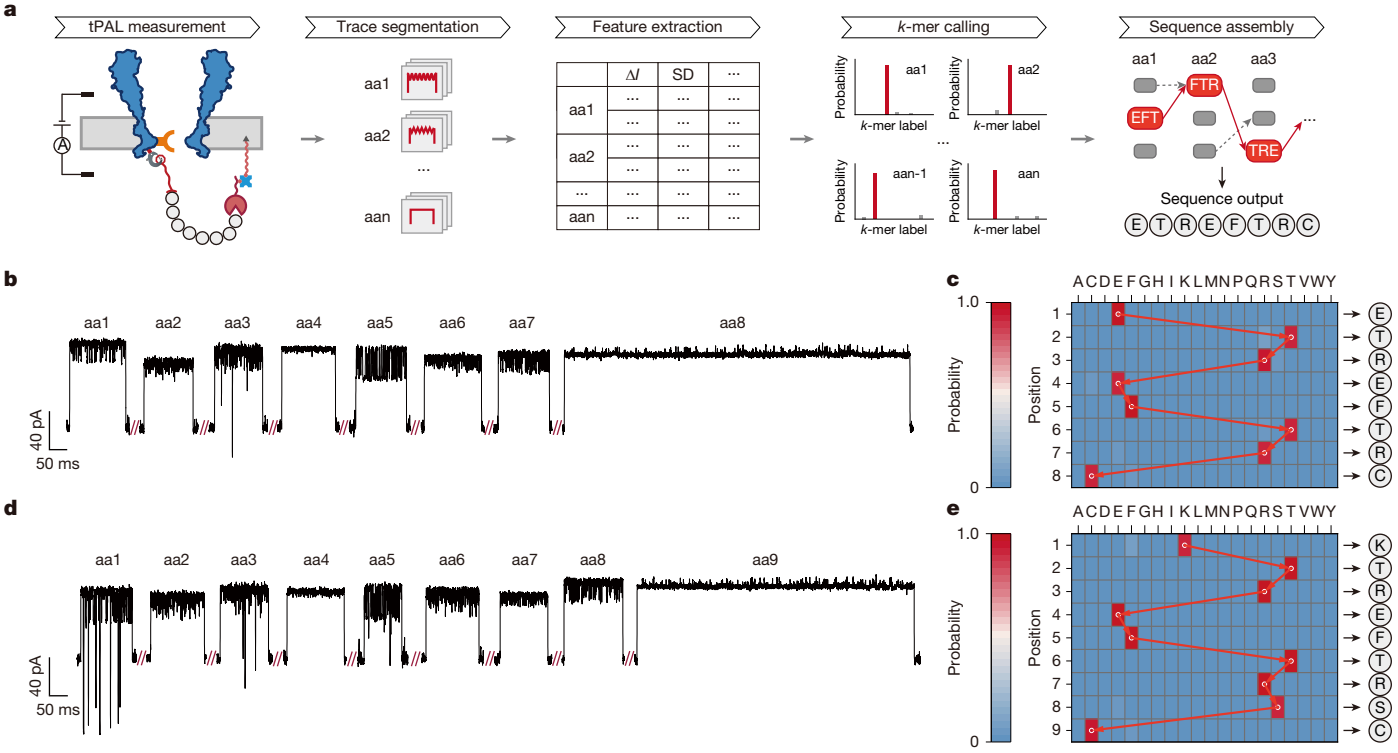


Fig. 5 | Decoding peptide sequence from tPAL events. a, Workflow for machine-learning-assisted sequence analysis. In brief, the raw data obtained from peptide sequential cleavage is first divided into different stages. Feature extraction is then performed for sensing events within each stage to generate the feature matrix. Subsequently, a machine-learning model pre-trained with a *k*-mer dataset is used for event prediction, yielding *k*-mer labels and probability distributions for each stage. Finally, sequence assembly is done on the basis of the sequence overlap between *k*-mers from adjacent stages, outputting the complete peptide sequence. **b**,**d**, Representative events acquired with ETREFTRC (**b**) and KTREFTRSC (**d**) during sequential cleavage. The measurements were performed as described in the Methods. **c**,**e**, Amino acid probability heat maps derived from ETREFTRC (**c**) and KTREFTRSC (**e**) measurements. The probabilities of 20 amino acids at each position were calculated by normalizing the frequencies of all *k*-mers covering that position in prediction results. The optimal path (red polyline) connects the amino acid with the highest probability at each position, forming the predicted peptide sequence on the right side of each heat map.

sequential cleavage, corresponding to the exposure of either glutamic acid or phenylalanine at the N terminus (Fig. 4a,c). Similarly, FEFEFEC produced a signal pattern analogous to that of EFEFEFC, but in reverse order (Extended Data Fig. 8). For EFTREFTREFTRC—a peptide composed of repeating EFTR motifs—a distinct periodic signal pattern was detected (Fig. 4b,d). Within each repeating unit, when the same N-terminal residue was exposed, the resulting events had similar noise characteristics, with only minor differences in current amplitude, despite variations in their sequence length. The only exception was the aa12 stage, in which the sensing events differed significantly from those at the aa4 and aa8 stages, owing to differences in the downstream amino acid composition and the linkage to the pore. Furthermore, given the high resolution of the pore, this method could also resolve short homopolymeric sequences, as demonstrated with the peptide AAAAC (Extended Data Fig. 9). However, it might be challenging to sequence peptides containing long homopolymeric segments with tPAL.

Direct decoding of peptide sequence

To further evaluate the spatial resolution of tPAL, three peptides that differed only at the fifth residue (LEFTNSC, LEFTLSC and LEFTFSC) were tested (Supplementary Fig. 50). A comparison of the events generated during enzymatic cleavage revealed distinct features among the three peptides at the aa3–aa5 stages, which could be directly attributed to the mutation at the fifth position. However, the signals at the aa1 and aa2 stages were highly similar across all three peptides, indicating that the events at these stages are not affected by the mutation. Additional cases with identical N-terminal amino acid compositions for the first three residues also showed high signal similarity (Supplementary Fig. 51), suggesting that the first three N-terminal amino acids have a dominant role in determining event features, with only minor contributions from subsequent residues. Moreover, the influence of a given residue decreases progressively with increasing distance from the N terminus. This suggests that peptide sequences can be decoded from tPAL results directly, through a 3-mer-based strategy (Fig. 5a). Such an approach is conceptually similar to the base-calling methodology used in nanopore nucleic acid sequencing, in which *k* values of 6–9 are commonly used to reliably identify the nucleic acid sequence[44,45], corresponding to a reference database covering $4^6 = 4{,}096$ to $4^9 = 262{,}144$ combinations. By contrast, the high spatial resolution achieved by tPAL enables peptide sequence decoding at 3-mer level. The corresponding database for all possible 3-mer combinations contains $20^3 = 8{,}000$ entries, readily achievable using modern bioinformatics pipelines through efficient pattern recognition algorithms and machine-learning frameworks. This major advantage of tPAL is achieved through enzymatic cleavage and repeated N-terminus readout of cleaved peptides, which simultaneously simplifies the decoding architecture and mitigates stochastic errors in *k*-mer calling.

As a proof of concept, a miniaturized *k*-mer database was first constructed using a set of pre-labelled 3-mer, 2-mer and 1-mer events collected from results acquired during the sequential cleavage of EFEFEFC, EFTREFTREFTRC, ETRSC, KTRSC and FTRSC (Supplementary Fig. 52). The wide neural network model, which achieved the highest validation accuracy of 95.9%, was selected for subsequent predictions. The data acquired during sequential cleavage of ETREFTRC were analysed (Fig. 5b), and the trained model was used to predict the *k*-mer

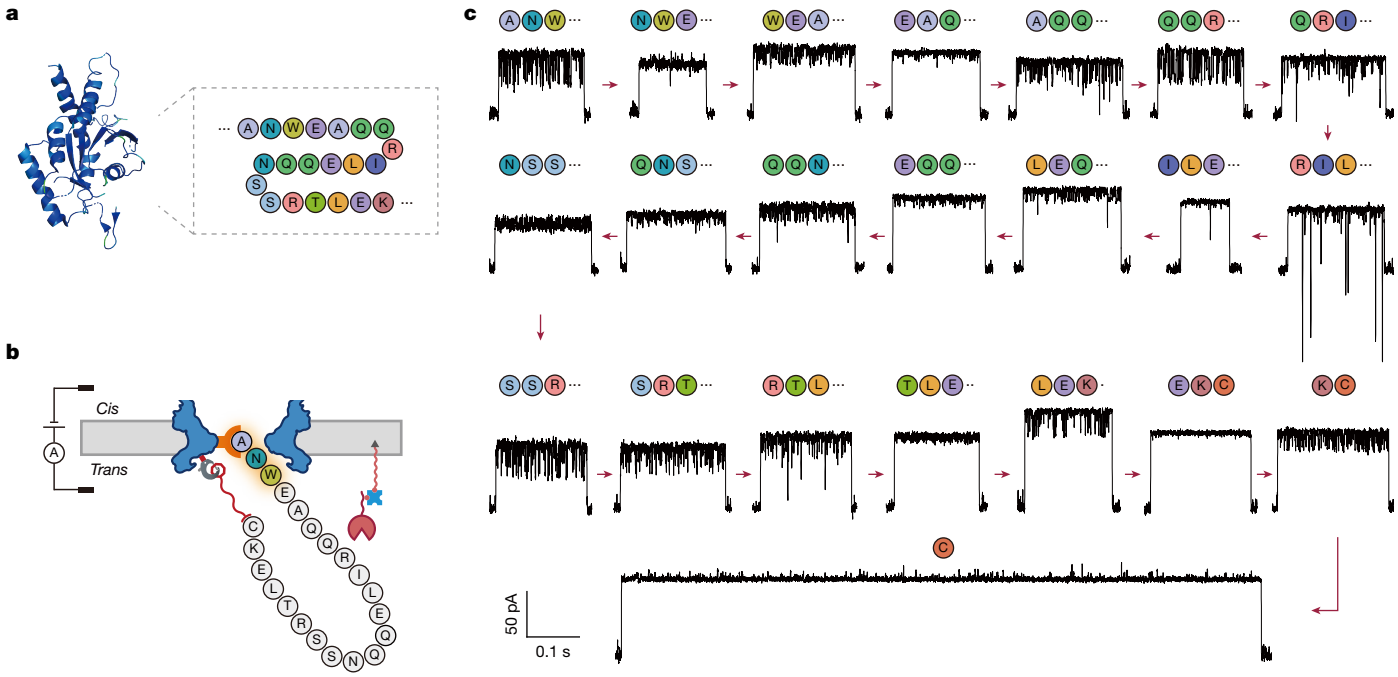


Fig. 6 | Sequential cleavage of a long peptide. a, A peptide derived from the human septin 7 sequence (residues 409–429; sequence ANWEAQQRILEQQNSSRTLEKC) was selected for tPAL analysis. **b**, Schematic depicting sequential cleavage of ANWEAQQRILEQQNSSRTLEKC using sgAP-dA15-chol. The tPAL events generated at each stage mainly encode information on the first three amino acids at the N terminus. **c**, Representative events acquired with ANWEAQQRILEQQNSSRTLEKC during sequential cleavage. Stepwise cleavage of the peptide from the N terminus to the C-terminal cysteine residue was fully observed, resulting in 22 distinct stages. The corresponding 3-mer read at each stage is labelled above each event for clarity. In the last two stages, only two or one amino acids remain uncleaved. Notably, subtle differences in event features can be observed for the same terminal amino acid (such as QQR and QRI), owing to the influence of its sequence context, a property that proves advantageous for resolving short homopolymers (Extended Data Fig. 9).

identity of each stage. The amino acid prediction probabilities at each position were calculated from the predicted *k*-mers, and an amino acid probability heat map was generated accordingly (Fig. 5c and Methods).

Notably, the tPAL events in each stage encode information not only about the N-terminal amino acid of that stage but also about a few downstream residues. Therefore, even if a small portion of events are misclassified, the errors can be corrected through the contextual information provided by neighbouring *k*-mers. Specifically, an approach analogous to the de Bruijn graph[46] was developed for sequence reconstruction (Methods). As illustrated in Extended Data Fig. 10, although a few *k*-mers were misidentified, they rarely formed complete connections with upstream and downstream *k*-mers. By identifying the optimal path that spans from the beginning to the end, the correct sequence can be reconstructed.

To validate the generalizability of this decoding method, the tPAL data from KTREFTRSC (Supplementary Video 1) and FTREFTRC were also analysed. In both cases, the peptide sequences were reconstructed successfully (Fig. 5d,e and Supplementary Figs. 53 and 54). Although demonstrated here using a miniaturized *k*-mer database constructed from events collected with existing peptides, further expansion of the *k*-mer database and optimization of machine-learning algorithms holds considerable promise for practical sequencing applications.

The capacity of tPAL to resolve long peptides of random sequence was validated with ANWEAQQRILEQQNSSRTLEKC. This peptide sequence is derived from an enzymatically cleaved fragment of human septin 7 with an appended C-terminal cysteine, mimicking the tPAL analysis of native proteins after proteolytic hydrolysis. As shown in Fig. 6, 22 distinct event stages, each encoding the sequence information of multiple corresponding N-terminal residues rather than a single N-terminal amino acid, were clearly identified during sequential enzymatic cleavage. This provides evidence of the successful tracking of prolonged consecutive cleavage steps, and confirms the method's potential for protein sequencing.

Conclusion

We developed a nanopore-based peptide analysis tool termed tPAL (Supplementary Video 2). With tPAL, an MspA nanopore is concurrently modified with an NTA-Ni adapter and a peptide. The N terminus of the immobilized peptide can be read repeatedly through reversible coordination interactions with the NTA-Ni adapter. Stepwise shortening of the N-terminal amino acid by aminopeptidase during tPAL measurement enables the generation of peptide-sequence-dependent events. Improved cleavage efficiency is achieved through a cholesterol-modified aminopeptidase, owing to its enrichment near the membrane. Using this principle, tPAL enables the sequential cleavage of peptides containing single-amino-acid mutations, PTMs and UAAs. Corresponding sequence-dependent event features were observed, aligning with the sequential alterations. A read length of up to 22 amino acids was achieved, with longer reads also deemed feasible but not experimentally examined herein. A specialized data-processing pipeline was developed and used to demonstrate the feasibility of sequence decoding.

However, sgAP exhibits a sequence-specific substrate preference, leading to inefficient hydrolysis in the presence of certain amino acids, peptide sequences or PTMs. Fortunately, the tPAL set-up allows for the simultaneous adaptation of multiple aminopeptidases or other enzymes in the same assay, which should address this issue. For highly structured peptides with an inaccessible N terminus, denaturant-mediated structural unfolding[33] can be used to enable detection. Although demonstrated here with synthetic peptides, the tPAL method also holds great promise for the analysis of natural

proteins. Proteins can be first hydrolysed into peptide fragments using endopeptidases, and then immobilized to a nanopore through photocatalytic C-terminal-specific labelling methods[47] to enable tPAL measurements. Integration with high-density nanopore arrays has the potential to increase the measurement throughput substantially. Further development of this strategy could accelerate progress in proteomics and facilitate new discoveries in diagnostics and therapeutics. Although our strategy is demonstrated here with peptides, the use of appropriate reactive adapters and enzymes could mean it also holds promise for applications in nanopore-based nucleic acid and glycan sequencing.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-026-10881-1.

1. Restrepo-Pérez, L., Joo, C. & Dekker, C. Paving the way to single-molecule protein sequencing. *Nat. Nanotechnol.* **13**, 786–796 (2018).
2. Dorey, A. & Howorka, S. Nanopore DNA sequencing technologies and their applications towards single-molecule proteomics. *Nat. Chem.* **16**, 314–334 (2024).
3. Lu, C., Bonini, A., Viel, J. H. & Maglia, G. Toward single-molecule protein sequencing using nanopores. *Nat. Biotechnol.* **43**, 312–322 (2025).
4. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).
5. Gregorich, Z. R. & Ge, Y. Top-down proteomics in health and disease: challenges and opportunities. *Proteomics* **14**, 1195–1210 (2014).
6. Edman, P. A method for the determination of amino acid sequence in peptides. *Arch. Biochem.* **22**, 475 (1949).
7. Shuken, S. R. An introduction to mass spectrometry-based proteomics. *J. Proteome Res.* **22**, 2151–2171 (2023).
8. Timp, W. & Timp, G. Beyond mass spectrometry, the next step in proteomics. *Sci. Adv.* **6**, eaax8978 (2020).
9. Alfaro, J. A. et al. The emerging landscape of single-molecule protein sequencing technologies. *Nat. Methods* **18**, 604–617 (2021).
10. Zhao, Y. et al. Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling. *Nat. Nanotechnol.* **9**, 466–473 (2014).
11. Ohshiro, T. et al. Detection of post-translational modifications in single peptides using electron tunnelling currents. *Nat. Nanotechnol.* **9**, 835–840 (2014).
12. Shrestha, P. et al. Single-molecule mechanical fingerprinting with DNA nanoswitch calipers. *Nat. Nanotechnol.* **16**, 1362–1370 (2021).
13. Zheng, L., Sun, Y., Hein, L. A., Eisenstein, M. & Hyongsok, T. S. Single-molecule peptide sequencing through reverse translation of peptides into DNA. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-026-03061-z (2026).
14. Swaminathan, J. et al. Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. *Nat. Biotechnol.* **36**, 1076–1082 (2018).
15. Reed, B. D. et al. Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device. *Science* **378**, 186–192 (2022).
16. Filius, M. et al. Full-length single-molecule protein fingerprinting. *Nat. Nanotechnol.* **19**, 652–659 (2024).
17. Püntener, S. & Rivera-Fuentes, P. Single-molecule peptide identification using fluorescence blinking fingerprints. *J. Am. Chem. Soc.* **145**, 1441–1447 (2023).
18. Yan, S. et al. Single molecule ratcheting motion of peptides in a *Mycobacterium smegmatis* porin A (MspA) nanopore. *Nano Lett.* **21**, 6703–6710 (2021).
19. Brinkerhoff, H., Kang, A. S. W., Liu, J., Aksimentiev, A. & Dekker, C. Multiple rereads of single proteins at single-amino acid resolution using nanopores. *Science* **374**, 1509–1513 (2021).
20. Motone, K. et al. Multi-pass, single-molecule nanopore reading of long protein strands. *Nature* **633**, 662–669 (2024).
21. Luo, R. Y. et al. Exploration of semiconductor chip-based single-molecule protein sequencing for identification of hemoglobin variants. *Anal. Chem.* **98**, 3458–3465 (2026).
22. Cherf, G. M. et al. Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nat. Biotechnol.* **30**, 344–348 (2012).
23. Manrao, E. A. et al. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* **30**, 349–353 (2012).
24. Deamer, D., Akeson, M. & Branton, D. Three decades of nanopore sequencing. *Nat. Biotechnol.* **34**, 518–524 (2016).
25. Garalde, D. R. et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* **15**, 201–206 (2018).
26. Chen, Z. et al. Controlled movement of ssDNA conjugated peptide through *Mycobacterium smegmatis* porin A (MspA) nanopore by a helicase motor for peptide sequencing application. *Chem. Sci.* **12**, 15750–15756 (2021).
27. Nova, I. C. et al. Detection of phosphorylation post-translational modifications along single peptides with nanopores. *Nat. Biotechnol.* **42**, 710–714 (2024).
28. Chen, X. et al. Resolving sulfation posttranslational modifications on a peptide hormone using nanopores. *ACS Nano* **18**, 28999–29007 (2024).
29. Nivala, J., Marks, D. B. & Akeson, M. Unfoldase-mediated protein translocation through an α-hemolysin nanopore. *Nat. Biotechnol.* **31**, 247–250 (2013).
30. Nivala, J., Mulroney, L., Li, G., Schreiber, J. & Akeson, M. Discrimination among protein variants using an unfoldase-coupled nanopore. *ACS Nano* **8**, 12365–12375 (2014).
31. Zhang, S. et al. Bottom-up fabrication of a proteasome-nanopore that unravels and processes single proteins. *Nat. Chem.* **13**, 1192–1199 (2021).
32. Martin-Baniandres, P. et al. Enzyme-less nanopore detection of post-translational modifications within long polypeptides. *Nat. Nanotechnol.* **18**, 1335–1340 (2023).
33. Yu, L. et al. Unidirectional single-file transport of full-length proteins through a nanopore. *Nat. Biotechnol.* **41**, 1130–1139 (2023).
34. Sauciuc, A., Morozzo della Rocca, B., Tadema, M. J., Chinappi, M. & Maglia, G. Translocation of linearized full-length proteins through an engineered nanopore under opposing electrophoretic force. *Nat. Biotechnol.* **42**, 1275–1281 (2024).
35. Faller, M., Niederweis, M. & Schulz, G. E. The structure of a mycobacterial outer-membrane channel. *Science* **303**, 1189–1192 (2004).
36. Cao, J. et al. Giant single molecule chemistry events observed from a tetrachloroaurate(III) embedded *Mycobacterium smegmatis* porin A nanopore. *Nat. Commun.* **10**, 5668 (2019).
37. Wang, Y. et al. Nanopore sequencing accurately identifies the mutagenic DNA lesion O^6-carboxymethyl guanine and reveals its behavior in replication. *Angew. Chem. Int. Ed.* **58**, 8432–8436 (2019).
38. Liu, Y. et al. Nanopore identification of alditol epimers and their application in rapid analysis of alditol-containing drinks and healthcare products. *J. Am. Chem. Soc.* **144**, 13717–13728 (2022).
39. Wang, Y. et al. Identification of nucleoside monophosphates and their epigenetic modifications using an engineered nanopore. *Nat. Nanotechnol.* **17**, 976–983 (2022).
40. Zhang, S. et al. A nanopore-based saccharide sensor. *Angew. Chem. Int. Ed.* **61**, e202203769 (2022).
41. Wang, K. et al. Unambiguous discrimination of all 20 proteinogenic amino acids and their modifications by nanopore. *Nat. Methods* **21**, 92–101 (2024).
42. de la Torre, D. & Chin, J. W. Reprogramming the genetic code. *Nat. Rev. Genet.* **22**, 169–184 (2021).
43. Maras, B. et al. Aminopeptidase from *Streptomyces griseus*. *Eur. J. Biochem.* **236**, 843–846 (1996).
44. David, M., Dursi, L. J., Yao, D., Boutros, P. C. & Simpson, J. T. Nanocall: an open source basecaller for Oxford Nanopore sequencing data. *Bioinformatics* **33**, 49–55 (2017).
45. Kovaka, S. et al. Uncalled4 improves nanopore DNA and RNA modification detection via fast and accurate signal alignment. *Nat. Methods* **22**, 681–691 (2025).
46. Compeau, P. E. C., Pevzner, P. A. & Tesler, G. How to apply de Bruijn graphs to genome assembly. *Nat. Biotechnol.* **29**, 987–991 (2011).
47. Bloom, S. et al. Decarboxylative alkylation for site-selective bioconjugation of native proteins via oxidation potentials. *Nat. Chem.* **10**, 205–211 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2026

Methods

Preparation of nanopores

$(N90C, N93AzK)_1(M2)_7$ is a hetero-octameric MspA, composed of seven monomeric units of M2 MspA-D16H6 and one monomeric unit of N90C, N93AzK MspA-H6. The azido group at site 93 of the N90C, N93AzK MspA-H6 monomer is introduced by GCE, in response to the amber codon (TAG). The hexa-histidine tag (H6) at the C terminus of each monomer is designed to assist purification by nickel affinity chromatography. The 16-aspartic acid tag (D16) placed on the M2 MspA-D16H6 monomer is designed to enhance the discrimination between different heterogeneous assemblies during gel electrophoresis. To prepare $(N90C, N93AzK)_1(M2)_7$, the genes of M2 MspA-D16H6 and N90C, N93TAG MspA-H6 were custom synthesized and inserted in a co-expression vector pETDuet-1. Specifically, the gene of M2 MspA-D16H6 was inserted between the restriction sites of NcoI and HindIII, and the gene encoding N90C, N93TAG MspA-H6 was inserted between the restriction sites of NdeI and BlpI. The produced plasmid DNA is referred to as pETDuet-1-$(N90C, N93TAG)_1(M2)_7$.

Experimentally, 40 ng of pETDuet-1-$(N90C, N93TAG)_1(M2)_7$ and 40 ng of pSupAR-MmPylRS for AzK insertion were co-transformed into 100 µl of *Escherichia coli* BL21 (DE3) competent cells. The transformed cells were plated onto LB agar plates containing ampicillin (50 µg ml^{-1}) and chloramphenicol (34 µg ml^{-1}) and incubated at 37 °C for 18 h. Subsequently, a single colony was picked and inoculated into 100 ml LB broth supplemented with 50 µg ml^{-1} kanamycin and 34 µg ml^{-1} chloramphenicol. The culture was grown at 37 °C with shaking at 180 rpm until the optical density at 600 nm ($OD_{600\,nm}$) reached 0.6–0.8. Protein expression was induced by adding 5 mM AzK, 0.1 mM IPTG and 0.2% (w/v) L-arabinose. After 24 h of expression at 16 °C, cells were collected by centrifugation (6,000 rpm, 15 min, 4 °C).

The collected bacterial pellet was resuspended in 24 ml lysis buffer 1 (150 mM NaCl, 100 mM NaH_2PO_4, 0.1 mM EDTA and 0.5% (w/v) Genapol X-80, pH 6.5) and heated at 60 °C for 50 min. The suspension was then cooled on ice for 10 min and centrifuged at 13,000 rpm for 50 min at 4 °C. The supernatant was collected and filtered through a 0.2-µm syringe filter. Subsequently, the filtered supernatant was loaded onto a HisTrap HP column. The column was washed with buffer A1 (0.5 M NaCl, 20 mM HEPES, 5 mM imidazole and 0.5% (w/v) Genapol X-80, pH 8.0), followed by elution with a linear gradient of imidazole from buffer A1 to buffer B1 (0.5 M NaCl, 20 mM HEPES, 500 mM imidazole and 0.5% (w/v) Genapol X-80, pH 8.0). Fractions containing heterogeneously assembled MspA were collected and concentrated using centrifugal spin concentrators (Amicon Ultra-0.5 ml, molecular weight cut-off (MWCO) 50 kDa).

To separate $(N90C, N93AzK)_1(M2)_7$ from the mixture, gel electrophoresis of the previously collected fractions was performed using a 10% SDS–PAGE gel. The electrophoresis was run at 160 V for 16 h using Tris-Glycine-SDS running buffer. The gel was stained with Coomassie blue staining solution (1.25 g Coomassie Brilliant Blue R250, 225 ml methanol, 50 ml acetic acid and 225 ml ultrapure water) for 4 h. Thereafter, the gel was de-stained with de-staining buffer (400 ml methanol, 100 ml acetic acid and 500 ml ultrapure water) until protein bands were clearly visible. The protein band corresponding to $(N90C, N93AzK)_1(M2)_7$ was excised from the gel, crushed and immersed in extraction solution (150 mM NaCl, 20 mM HEPES, 0.2% *N*-dodecyl-β-D-maltoside (DDM), 0.5% Genapol X-80 and 10 mM EDTA, pH 7.5) for 12 h. The supernatant was collected and stored at −80 °C for subsequent use.

$(D56AzK, N90C)_1(M2)_7$ is a hetero-octameric MspA comprising seven M2 MspA-D16H6 monomers and one D56AzK, N90C MspA-H6 monomer. The azido group at position 56 of the D56AzK, N90C MspA-H6 monomer was incorporated by GCE in response to the amber codon (TAG). To prepare $(D56AzK, N90C)_1(M2)_7$, the genes encoding M2 MspA-D16H6 and D56TAG, N90C MspA-H6 were custom synthesized and cloned into the co-expression vector pETDuet-1. Specifically, the M2 MspA-D16H6 gene was inserted between the NcoI and HindIII restriction sites, whereas the D56TAG, N90C MspA-H6 gene was inserted between the NdeI and BlpI restriction sites. The resulting plasmid was designated pETDuet-1-$(D56TAG, N90C)_1(M2)_7$. The expression and purification of $(D56AzK, N90C)_1(M2)_7$ followed the same protocol as that used for $(N90C, N93AzK)_1(M2)_7$.

$(M2)_8$ is a homo-octamer formed by eight identical units of M2 MspA-D16H6. The gene encoding M2 MspA-D16H6 was cloned into a pET-30a(+) vector, flanked by the NdeI and HindIII restriction sites, resulting in the construct known as pET-30a-M2 MspA. Expression of $(M2)_8$ was achieved in *E. coli* BL21 (DE3) cells, followed by purification by nickel affinity chromatography[39]. The purified $(M2)_8$ served as a reference standard in gel electrophoresis, as illustrated in Fig. 1 and Supplementary Fig. 2.

$(N90C)_1(M2)_7$ represents a hetero-octameric protein comprising one subunit of N90C MspA-H6 and seven subunits of M2 MspA-D16H6. For the synthesis of $(N90C)_1(M2)_7$, the genes encoding both M2 MspA-D16H6 and N90C MspA-H6 were co-inserted into a pETDuet-1 co-expression vector. The N90C MspA-H6 gene was positioned between the NcoI and HindIII restriction sites, and the M2 MspA-D16H6 gene was situated between the NdeI and BlpI sites. The resultant plasmid is denoted as pETDuet-1-$(N90C)_1(M2)_7$. Expression and purification of $(N90C)_1(M2)_7$ were done in *E. coli* BL21(DE3) pLysS cells, following previously established protocols[39]. This prepared $(N90C)_1(M2)_7$ was used as a reference standard in the nanopore measurements depicted in Supplementary Fig. 3.

Preparation of Mal-PEG4-MspA-NTA-Ni and Mal-PEG24-MspA-NTA-Ni

To synthesize Mal-PEG4-MspA-NTA-Ni, the $(N90C, N93AzK)_1(M2)_7$ protein was first modified with an NTA-Ni adapter at position 90. This was followed by the conjugation of DBCO-PEG4-maleimide at position 93. Experimentally, 5 µl $(N90C, N93AzK)_1(M2)_7$ was mixed with 5 µl maleimido-C3-NTA (20 mM, in 150 mM NaCl, 20 mM HEPES, pH 7.5) and vortexed (800 rpm) at 37 °C for 3 h in a shaking incubator. The resultant product is termed $(N90NTA, N90AzK)_1(M2)_7$. Subsequently, 1 µl $NiSO_4$ (100 mM, in ultrapure water) was added to initiate Ni^{2+} chelation, resulting in the chelation product $(N90NTA\text{-}Ni, N90AzK)_1(M2)_7$. Thereafter, 11 µl of $(N90NTA\text{-}Ni, N90AzK)_1(M2)_7$ solution was mixed with 0.5 µl of DBCO-PEG4-maleimide or DBCO-PEG24-maleimide (100 mM in *N*,*N*-dimethylformamide (DMF)) and incubated at 25 °C for 12 h. The resulting products were designated as Mal-PEG4-MspA-NTA-Ni and Mal-PEG24-MspA-NTA-Ni, respectively.

Preparation of 56Peptide–dT15-PEG4-MspA-NTA-Ni

The 56Peptide–dT15-MspA-NTA-Ni construct features an NTA-Ni modification at position 90 and a peptide-dT15 modification at position 56. The peptide-dT15-DBCO was custom synthesized before the nanopore modification. The cysteine-containing peptide was dissolved in PB buffer (150 mM NaCl, 0.1 M NaH_2PO_4 and 5 mM EDTA, pH 7.2) at a concentration of 10 mM. The oligonucleotide, bearing a 5′ DBCO-TEG modification and a 3′ maleimide modification, which was named DBCO-dT15-maleimide, was dissolved in PB buffer at a concentration of 100 µM. To form the peptide-dT15-DBCO conjugate, 5 µl of the peptide, 0.5 µl TCEP (100 mM, in PB buffer) and 19.5 µl PB buffer were mixed and shaken (800 rpm) at 37 °C for 1 h. After this, 10 µl DBCO-dT15-maleimide was added, thoroughly mixed and further incubated at 25 °C for 12 h.

For the 56Peptide–dT15-MspA-NTA-Ni, 5 µl $(D56AzK, N90C)_1(M2)_7$ was mixed with 5 µl maleimido-C3-NTA (20 mM, in 150 mM NaCl and 20 mM HEPES, pH 7.5) and shaken (800 rpm) at 37 °C for 3 h. Subsequently, 1 µl of $NiSO_4$ (100 mM, in ultrapure water) was added to chelate Ni^{2+}, resulting in $(D56AzK, N90NTA\text{-}Ni)_1(M2)_7$. This was then mixed with 5 µl of the peptide-dT15-DBCO and incubated at 25 °C for 12 h. The

final product, 56Peptide–dT15-MspA-NTA-Ni, was stored at −80 °C for long-term storage.

Nanopore measurements

Nanopore measurements were performed using a custom-built measurement set-up. The device comprises two chambers separated by a Teflon film with a 100-µm hole. Before each measurement, three drops of hexadecane (0.5% v/v in pentane) were applied to the hole. During measurements, the device was secured in a Faraday cage placed on an optical table. Each chamber was filled with 500 µl KCl buffer (1.5 M KCl, 10 mM CHES, pH 9). A pair of Ag–AgCl electrodes, electrically connected to a patch-clamp amplifier, were immersed in both chambers to contact the buffers. By convention, the grounded chamber was designated as *cis*, with the opposing chamber defined as *trans*. A drop of DPhPC (5 mg ml^{-1} in pentane) was added to each chamber to form a lipid bilayer across the orifice. Nanopores were introduced into the *cis* chamber to initiate pore incorporation. After insertion of a single nanopore, the *cis* chamber buffer was immediately exchanged with fresh buffer to prevent further insertions.

All single-channel recordings were done using an Axopatch 200B patch-clamp amplifier paired with a Digidata 1550B digitizer. Recordings were sampled at 25 kHz and low-pass-filtered at a corner frequency of 1 kHz. Unless specified otherwise, all measurements were performed in 1.5 M KCl, 10 mM CHES (pH 9) buffer at room temperature.

Real-time construction of peptide–PEG4-MspA-NTA-Ni and peptide–PEG24-MspA-NTA-Ni

The construction of peptide–PEG4-MspA-NTA-Ni and peptide–PEG24-MspA-NTA-Ni was monitored in real time. To initiate the conjugation of the peptide to the nanopore, after the insertion of a single Mal-PEG4-MspA-NTA-Ni or Mal-PEG24-MspA-NTA-Ni, a cysteine-containing peptide was introduced into the *cis* chamber at a final concentration of 20 µM. The observation of frequent, repetitive peptide-binding events indicates the formation of peptide–PEG4-MspA-NTA-Ni or peptide–PEG24-MspA-NTA-Ni. Afterwards, immediate exchange of the buffer in the *cis* chamber was performed to eliminate interferences caused by excess peptide.

Preparation of cholesterolized sgAP

sgAP-AviTag-H6 is an aminopeptidase that incorporates an AviTag (GLNDIFEAQKIEWHE) and a hexa-histidine (H6) tag at its C terminus. This AviTag facilitates site-specific biotinylation at the lysine residue within the tag through the action of biotin-protein ligase (BirA)[48]. To cholesterylate sgAP, the sgAP-AviTag-H6 was at first biotinylated using BirA, and then conjugated to a DNA that had been dual-modified with both biotin and cholesterol through streptavidin-mediated linkage.

Experimentally, the gene encoding sgAP-AviTag-H6 was inserted into a pET-9a vector between the NdeI and BamHI restriction sites (Supplementary Table 10). The resultant plasmid, pET-9a-sgAP-AviTag-H6, was then transformed into *E. coli* BL21 (DE3) competent cells. These transformed cells were plated on an LB agar plate containing kanamycin (25 µg ml^{-1}) and incubated overnight at 37 °C. A single colony was picked and transferred to 10 ml LB medium supplemented with kanamycin (25 µg ml^{-1}), and the culture was incubated at 37 °C with shaking at 180 rpm for 12 h. Thereafter, the culture was inoculated into 200 ml TB medium supplemented with sorbitol (1 M), betaine (2.5 mM) and kanamycin (25 µg ml^{-1}) (ref. 49). The medium was shaken at 37 °C and 180 rpm until the $OD_{600\,nm}$ reached approximately 0.6. To induce protein expression, IPTG was added to a final concentration of 0.1 mM. After 40 h of growth at 18 °C, the cells were collected by centrifugation at 6,000 rpm, 4 °C, for 15 min.

The cell pellets were resuspended in 30 ml lysis buffer 2 (30 mM HEPES, 1 mM $CaCl_2$, pH 8), subjected to sonication for 15 min and then heat-treated at 50 °C for 20 min. The supernatant was subsequently collected by centrifugation at 13,000 rpm, 4 °C, for 50 min and diluted with an equal volume of buffer A2 (100 mM NaCl, 30 mM HEPES and 20 mM imidazole, pH 8). After this, the supernatant was filtered through a 0.2-µm syringe filter and loaded onto a HisTrap HP column. The column was washed with buffer A2 and then eluted with a linear gradient of imidazole to buffer B2 (100 mM NaCl, 30 mM HEPES and 500 mM imidazole, pH 8). The fractions containing sgAP-AviTag-H6 were buffer-exchanged to storage buffer (3 mM NaCl, 25 mM CHES, 1 mM $CaCl_2$ and 0.05 mM $ZnCl_2$, pH 9) using a Micro Bio-Spin 6 column. Protein concentrations were quantified using the Bradford method.

The purified sgAP-AviTag-H6 was biotinylated using a biotin labelling kit, following the manufacturer's instructions. The biotinylated product was then exchanged into storage buffer using a Micro Bio-Spin 6 column. This biotinylated form of sgAP is denoted as sgAP–biotin. The degree of biotinylation was assessed by SDS–PAGE gel-shift analysis (Extended Data Fig. 3).

For cholesterylation of sgAP, equimolar quantities of sgAP–biotin, streptavidin and biotin–dA15–cholesterol were mixed and incubated at 25 °C for 30 min. The resultant complex, sgAP-dA15-chol, was used directly for subsequent applications without the need for further purification.

tPAL measurement by real-time enzymatic digestion

After the real-time construction of a peptide–PEG4-MspA-NTA-Ni complex, sgAP-dA15-chol, $ZnCl_2$ and $CaCl_2$ were introduced into the *trans* chamber. The final concentrations were 8 nM for sgAP-dA15-chol, 5 µM for $ZnCl_2$ and 1 mM for $CaCl_2$. sgAP-dA15-chol was used to initiate the sequential cleavage of the N-terminal amino acids during tPAL measurement. A repetitive voltage protocol (+100 mV for 8 s, −100 mV for 1.5 s and 0 mV for 1.5 s) was applied to facilitate enzymatic peptide cleavage while mitigating the risk of spontaneous pore gating during the measurement. The binding events observed at +100 mV were collected and analysed.

Extraction of event features and machine learning

Binding events were analysed using the single-channel search module in Clampfit 10.7 to identify the start and end times of each event. To eliminate interference from noise, events with a dwell time shorter than 10 ms were excluded. A custom MATLAB program was used to extract key features for each event, such as event amplitude (ΔI), standard deviation (SD), event dwell time (t_{off}), skewness (skew), kurtosis (kurt), median, first quartile (Q1), third quartile (Q3), interquartile range (IQR), mean of the top 10% blockage values (mean_max10), mean of the top 20% blockage values (mean_max20). Specifically, for the extraction of frequency domain parameters including spectral centroid and spectral entropy, the current data are transferred into spectral data by fast Fourier transform (FFT) in MATLAB[50]. Spectral centroid refers to the centre-of-gravity position of the spectrum. Spectral entropy is a quantitative index to measure the complexity and randomness of the spectrum.

Machine-learning analysis was done using the Classification Learner toolbox in MATLAB. To maximize the generality of the data, peptide events were randomly selected from three independent measurements to form the dataset. The input dataset was then randomly split into a training set (80% of the labelled dataset) and a testing set (20%) for model training and model testing. A series of models, including ensembles, decision trees, discriminant analyses, naive Bayes, SVM, *k*-nearest neighbour (KNN) and neural networks, were trained on the training set. Tenfold cross-validation was performed to generate validation accuracy for each model, with test accuracy calculated using the independent testing set. The best-performing model was subsequently applied to predict unlabelled data. Data analysis including scatter plots, frequency distribution histograms and curve fittings, were done using Origin 2021.

Sequence decoding from tPAL events

Sequence decoding was performed using a custom MATLAB algorithm incorporating a pre-trained wide neural network classifier. First, raw

tPAL data obtained from sequential cleavage of peptide were segmented into distinct stages. These stages are designated as aa1 to aan. From each event, 11 event features (ΔI, SD, skew, kurt, t_{off}, median, IQR, Q1, Q3, mean_max10 and mean_max20) were extracted to form the feature matrix. For each stage, the pre-trained model was applied to predict *k*-mer labels, according to which the frequency distribution of candidate *k*-mers in each stage was reported.

To visualize the sequence-prediction results across all stages, a position-specific amino acid probability matrix was constructed. A projection step was performed to map *k*-mer probabilities back to individual amino acid positions. For each position in the peptide, the probability contributions for the 20 standard amino acids were accumulated from all *k*-mers covering this position. The cumulated values were then normalized by the total weight to yield a valid probability distribution. The resulting matrix was visualized as a heat map, within which the amino acid exhibiting the maximum likelihood at each position was identified.

Peptide sequence reconstruction was performed using a stage-wise directed graph-based assembly approach. In this framework, each predicted *k*-mer serves as a node, positioned within its specific sequencing stage. Directed edges were established between nodes in adjacent stages exhibiting strict suffix–prefix overlaps. An exhaustive recursive search was performed to explore all valid paths connecting the initial and the terminal stage. For each path, a score was calculated as the joint probability of all constituent *k*-mers. The path with the maximum cumulative probability was identified as the optimal path, and the final peptide sequence was derived by sequentially merging the overlapping *k*-mers along this route.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The machine-learning training data are provided as Supplementary Data ('tPAL_decoder'). Datasets that support the findings of this study are available from the corresponding author upon request. Source data are provided with this paper.

Code availability

The custom machine-learning code for peptide decoding and corresponding training data is included as Supplementary Data ('tPAL_decoder').

48. Beckett, D., Kovaleva, E. & Schatz, P. J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* **8**, 921–929 (1999).
49. Fundoiano-Hershcovitz, Y. et al. Identification of the catalytic residues in the double-zinc aminopeptidase from *Streptomyces griseus*. *FEBS Lett.* **571**, 192–196 (2004).
50. Zhang, H. et al. Chiral discrimination of all proteinogenic amino acid enantiomers by nanopore sensing. *Angew. Chem. Int. Ed.* **64**, e202515531 (2025).

Acknowledgements We thank G. Maglia for discussions on nanopore preparations, and P. R. Chen, R. Xie and J. Yang for providing the UAA plasmid and guidance with the GCE experiment.

Author contributions S.H., J.C. and K.W. conceived the project. J.C., H.Z., T.L., W.S., Z.W., X.G., Z.C., Y.O., L.Y., Y.W., R.L., X.A., X.D., L.Z. and L.Q. performed the measurements. J.C., Y.X. and K.W. prepared the MspA nanopores. J.C. prepared the aminopeptidase. K.W., J.C. and H.Z. designed the machine-learning algorithms. H.Z. prepared the Supplementary Videos. P.Z. set up the instruments. S.H. and J.C. wrote the paper. S.H. supervised the project.

Funding This project was funded by the National Natural Science Foundation of China (grants 22225405 and 22534004), National Key R&D Program of China (2023YFF1205900, 2022YFA1304602), State Key Laboratory of Analytical Chemistry for Life Sciences (grant 5431ZZXM2509), Yachen Foundation of Nanjing University, National Natural Science Foundation of China (grant 223B2402 to K.W.), China National Postdoctoral Program for Innovative Talents (grant BX20250087 to K.W.), Jiangsu Funding Program for Excellent Postdoctoral Talent (grant 2025ZB212 to K.W.) and China Postdoctoral Science Foundation (grant 2025M780957 to K.W.).

Competing interests S.H. and J.C. have filed patents on the tPAL method and its applications. The remaining authors declare no competing interests.

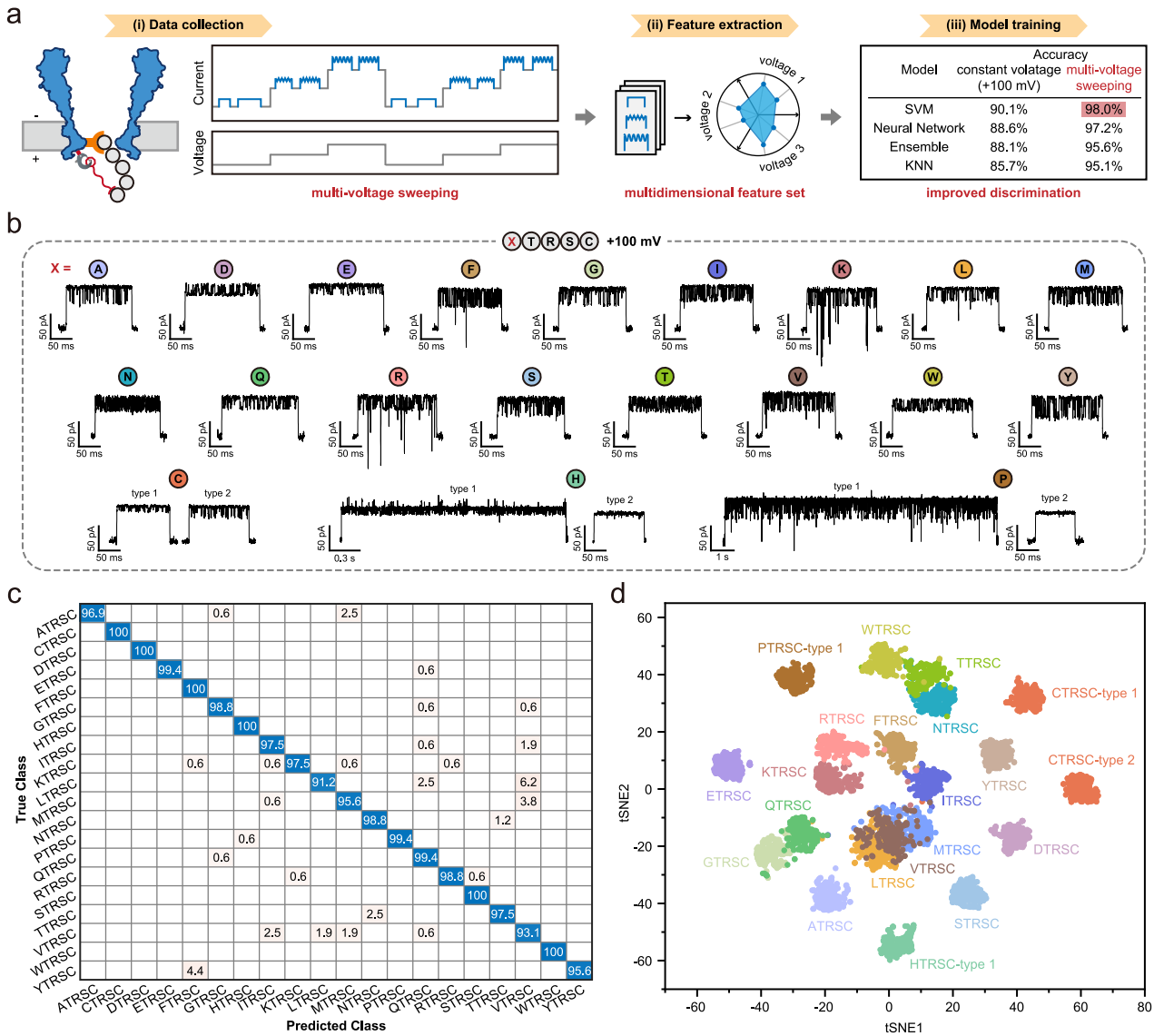
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-026-10881-1.

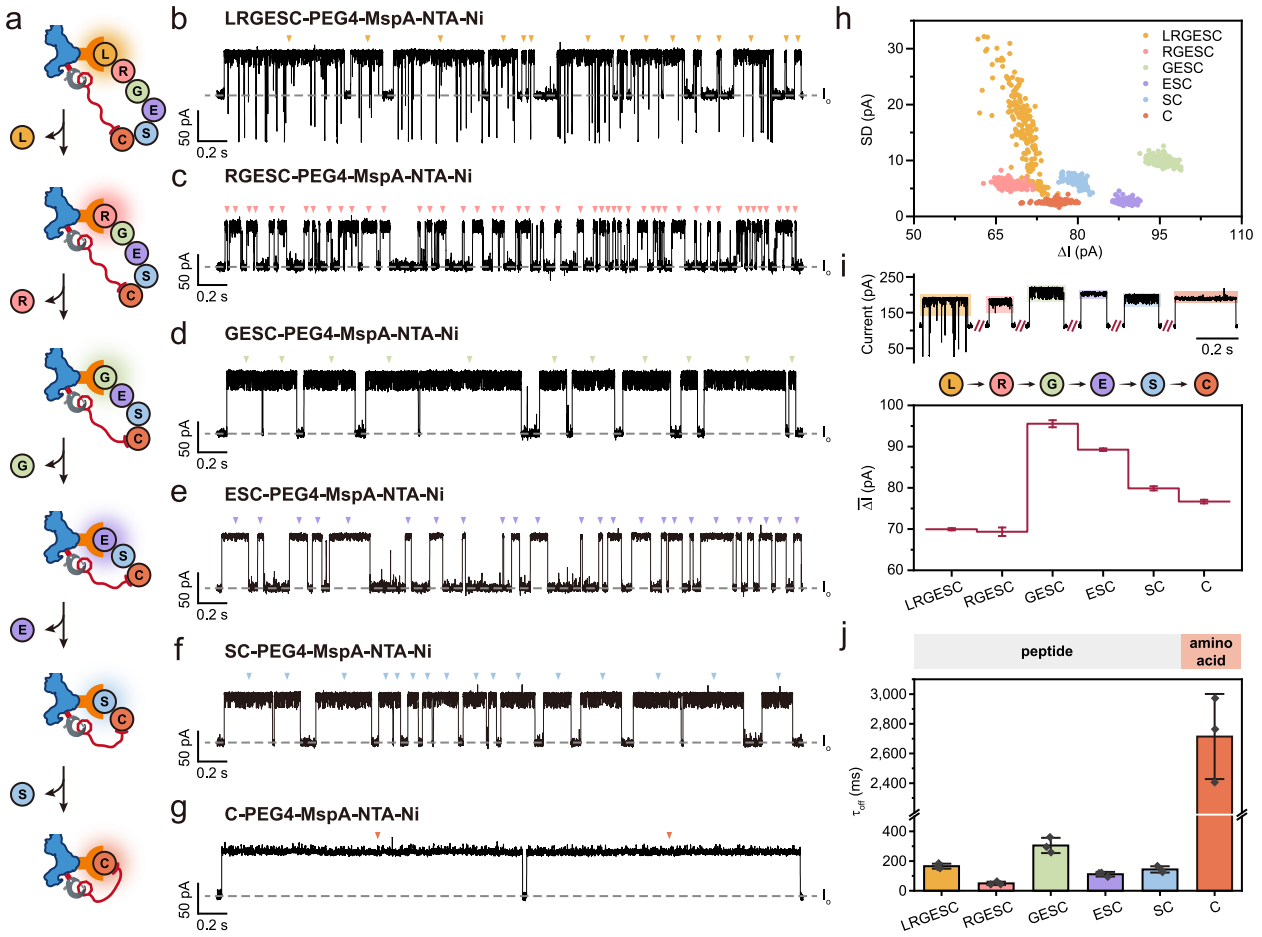
Correspondence and requests for materials should be addressed to Shuo Huang.

Peer review information *Nature* thanks Weihua Guan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

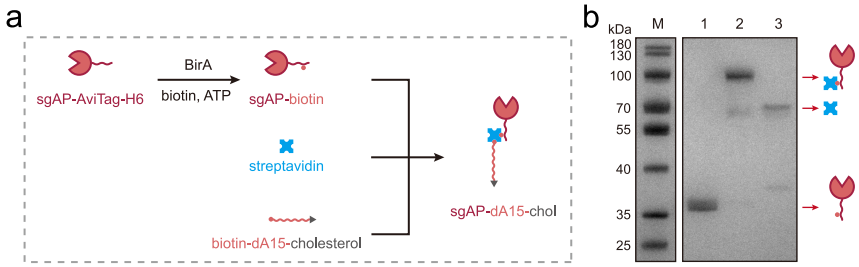
Reprints and permissions information is available at http://www.nature.com/reprints.



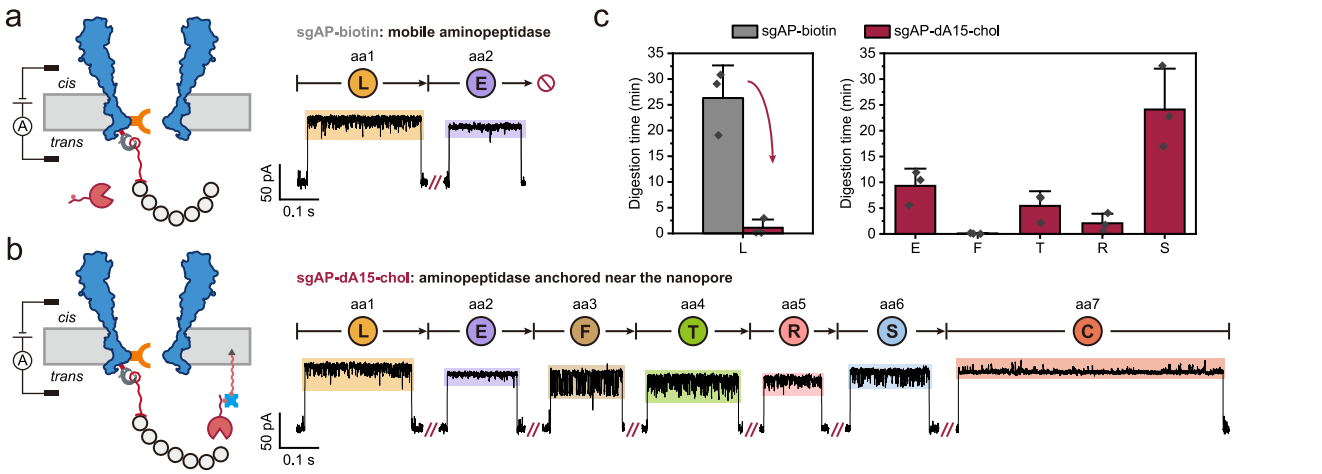
Extended Data Fig. 1 | Discriminating peptides with single-amino-acid substitutions using a multi-voltage sweeping strategy. (**a**) The workflow of multi-voltage sweeping and machine learning. In brief, each peptide was measured using a multi-voltage (+60 mV, +100 mV and +140 mV) protocol to acquire its tPAL events. For each voltage, 200 events were collected per peptide to construct the dataset. For each peptide, nine features (ΔI, SD, skew, kurt, t_{off}, median, IQR, spectral centroid and spectral entropy) were extracted from the events at each of the three voltages and subsequently concatenated to form a 27-dimensional feature matrix. Machine learning was performed with the Classification Learner toolbox of MATLAB, and model performance was evaluated by tenfold cross-validation. With the multi-voltage sweeping, the highest validation accuracy reached 98.0% with the quadratic SVM model. (**b**) Representative tPAL events of 20 XTRSC peptide variants acquired at +100 mV. Here, X represents each of the 20 proteinogenic amino acids. tPAL events of 20 XTRSC peptide variants acquired at other voltages are demonstrated in Supplementary Figs. 18–23. Specifically, CTRSC, HTRSC and PTRSC produce two types of nanopore events. The type-1 events of HTRSC and PTRSC, characterized by long dwell times, are attributed to N-terminal coordination, while the type-2 events arise from side-chain coordination. For CTRSC, the side chain of N-terminal cysteine may also participate in coordination along with the N-terminal amino group, giving rise to the secondary type of event. (**c**) The confusion matrix of peptide classification generated by the quadratic SVM model using a dataset from multi-voltage sweeping. (**d**) t-SNE visualization of 20 XTRSC peptide variants (n = 4,200). For HTRSC and PTRSC, only type-1 events were collected for plotting, owing to the infrequent occurrence of type-2 events.



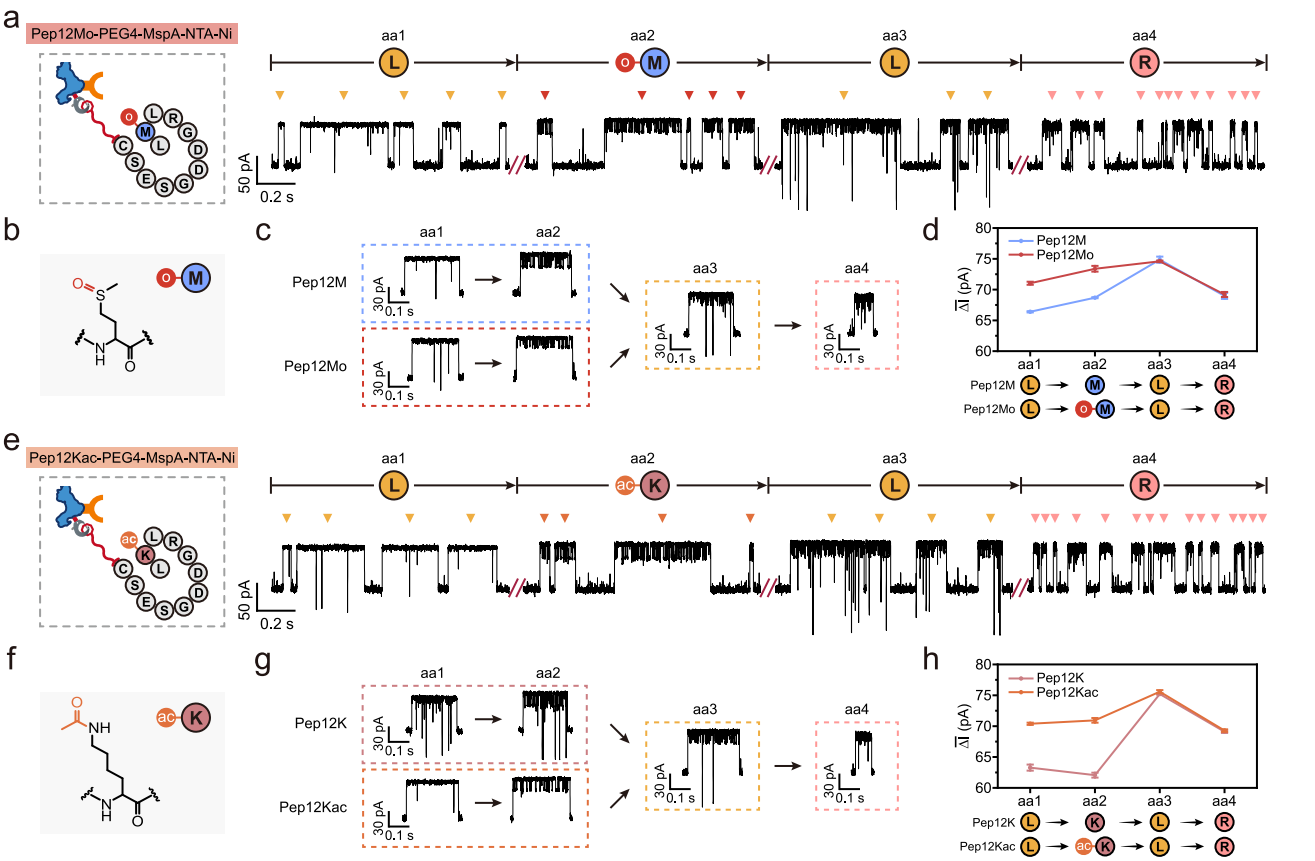
Extended Data Fig. 2 | Discriminating stepwise-shortened peptides with single-amino-acid resolution. (**a**) Schematic of nanopore measurements separately performed with different conjugated peptide. (**b**–**g**) Representative tPAL sensing traces of (**b**) LRGEDSC, (**c**) RGESC, (**d**) GESC, (**e**) ESC, (**f**) SC and (**g**) C. All measurements were performed as described in the Methods. The unbound state of each trace (I_o) is marked with a grey dashed line. The events resulting from peptide binding are labelled with corresponding coloured triangles. (**h**) The event scatter plot of Δ*I* versus SD for stepwise-shortened peptides. 200 events acquired at each condition were used to generate the plot. (**i**) Top: Representative tPAL sensing events of each peptide. Bottom: The corresponding step plot of $\overline{\Delta I}$. The cartoons above the plot demonstrate the N-terminal amino acid of each peptide. (**j**) Comparison of τ_{off} values between different peptides. Cysteine reports a significantly larger value of τ_{off}. Data are presented as means ± standard deviation values derived from results of three independent measurements.



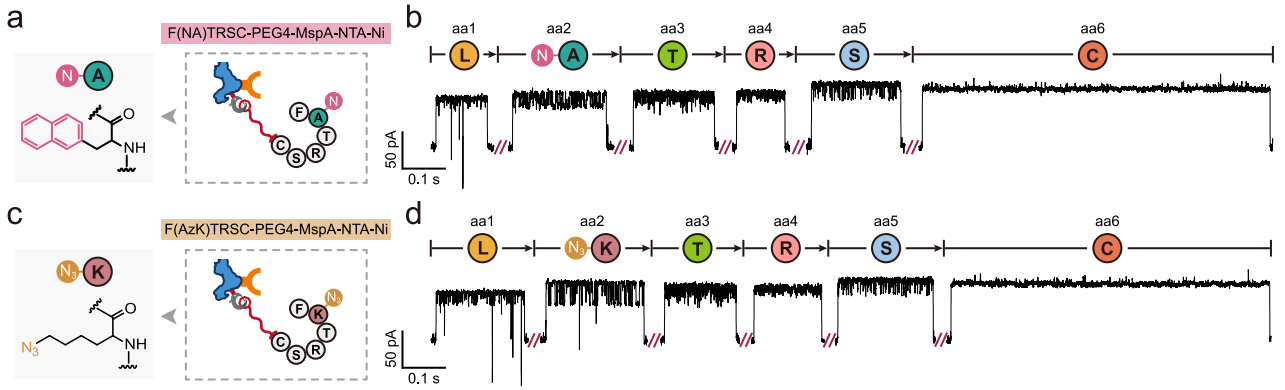
Extended Data Fig. 3 | Construction of sgAP-dA15-chol. (**a**) Schematic depicting the construction of sgAP-dA15-chol. The modification of sgAP involved the initial introduction of a biotin group to sgAP-AviTag-H6 using BirA in the presence of ATP. The biotinylated sgAP was then combined with streptavidin and biotin–dA15–cholesterol in equimolar ratios to form sgAP-dA15-chol. (**b**) Assessment of the biotinylation extent of sgAP using SDS–PAGE. The gel electrophoresis was conducted continuously for 100 min at +160 V using a 12% SDS–PAGE gel. Lane M: pre-stained protein ladder. Lane 1: sgAP–biotin, obtained by biotinylating sgAP-AviTag-H6 with BirA (Methods). Lane 2: sgAP-streptavidin, produced by conjugating sgAP–biotin and streptavidin. The sgAP–biotin and streptavidin were mixed in equimolar quantities and incubated at 25 °C for 30 min prior to gel analysis. Lane 3: streptavidin. The gel indicates that the majority of the sgAP has been effectively biotinylated. For gel source data, see Supplementary Fig. 55. Three independent experiments were performed with similar results.



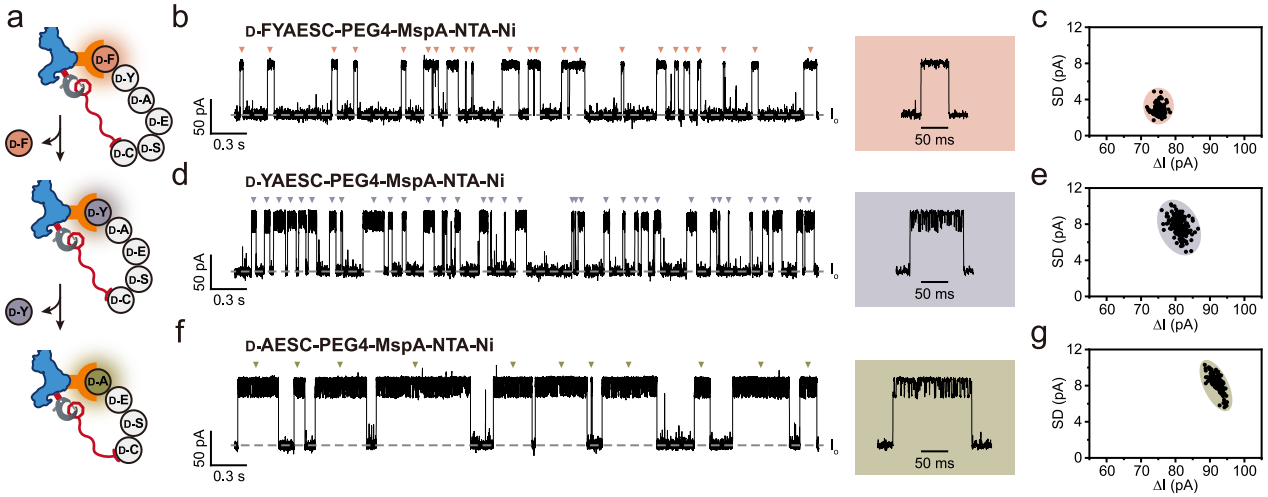
Extended Data Fig. 4 | Sequential enzymatic cleavage with sgAP–biotin or sgAP-dA15-chol. (**a**,**b**) Representative events obtained with LEFTRSC as the anchored peptide, using (**a**) sgAP–biotin or (**b**) sgAP-dA15-chol as the cleavage enzyme. Schematics depicting the corresponding strategies are positioned on the left. In the digestion strategy with sgAP–biotin, only the first amino acid was successfully cleaved, and no subsequent event changes were observed over the following four hours, owing to the inefficient accessibility of the mobile aminopeptidase to the single peptide conjugated to the nanopore. In contrast, sgAP-dA15-chol, which could be anchored near the nanopore via cholesterol–lipid bilayer interactions, facilitated the complete sequential digestion of the immobilized peptide. Six enzymatic cleavages were observed, resulting in seven stages labelled as aa1–aa7. (**c**) Left: Comparison of the time taken for the first amino acid digestion in LEFTRSC sequential cleavage using sgAP–biotin versus sgAP-dA15-chol. It is evident that sgAP-dA15-chol significantly improved the digestion efficiency. Right: Digestion times for other amino acids using sgAP-dA15-chol. All statistical data were derived from results of three independent measurements. The error bars represent standard deviation values.



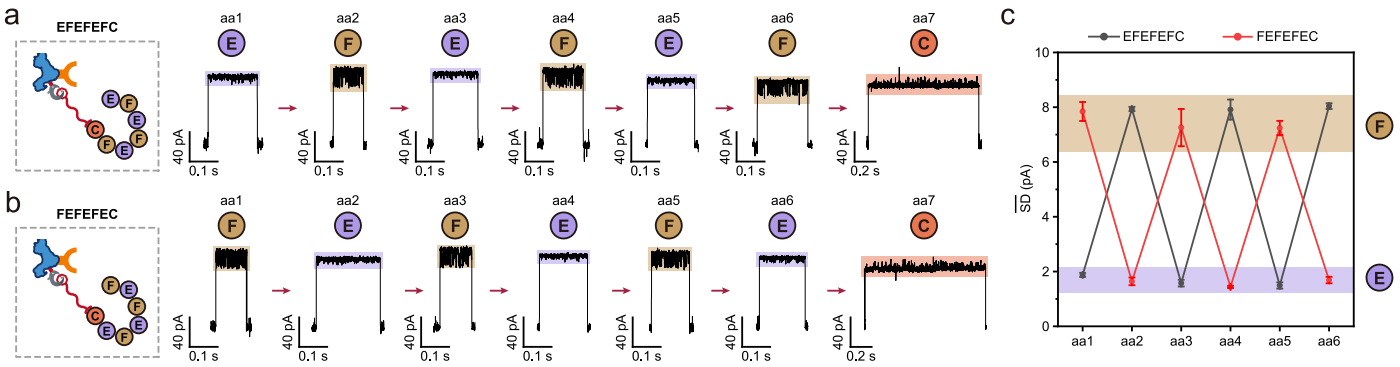
Extended Data Fig. 5 | tPAL measurements of peptides containing PTMs. (**a**,**e**) Representative traces acquired with (**a**) Pep12Mo and (**e**) Pep12Kac. The measurements were performed as described in the Methods. Three enzymatic cleavages were observed in each trial, generating four stages with distinct types of events. These four stages are referred to as aa1, aa2, aa3 and aa4. The events at each stage are marked with corresponding coloured triangles, and the corresponding N-terminal amino acid of each stage is labelled above the trace. (**b**) The chemical structure of oxidized methionine. (**c**) Comparison of representative events at each stage between Pep12Mo and Pep12M. The pattern of binding events converged after the cleavage of the second amino acid. (**d**) Comparison of $\overline{\Delta I}$ at each stage between Pep12Mo and Pep12M. The cartoons on the bottom illustrate the amino acid sequence of Pep12M and Pep12Mo. (**f**) The chemical structure of lysine acetylation. (**g**) Comparison of representative events at each stage between Pep12Kac and Pep12K. The pattern of binding events converged after the cleavage of the second amino acid. (**h**) Comparison of $\overline{\Delta I}$ at each stage between Pep12Kac and Pep12K. The cartoons on the bottom illustrate the N-terminal sequence of Pep12K and Pep12Kac. All error bars represent the standard deviations of means derived from the results of three independent measurements.



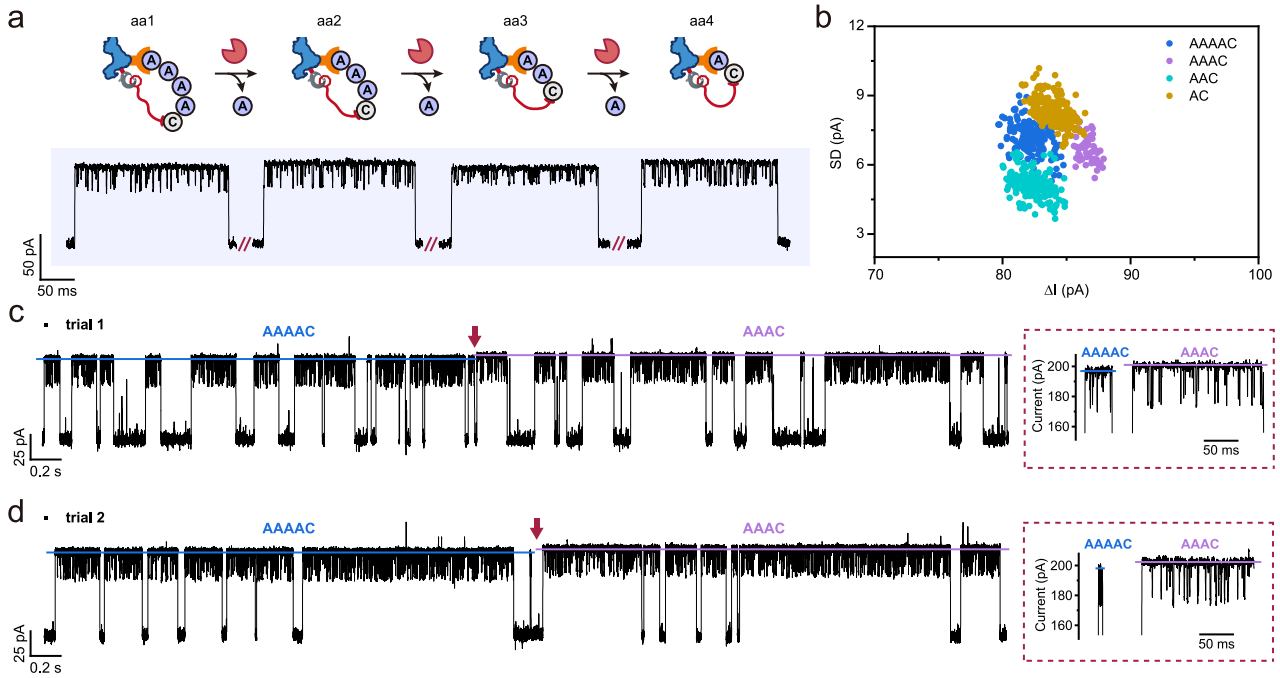
Extended Data Fig. 6 | Measurements of peptides containing UAAs. (**a**,**c**) Schematics of (**a**) F(NA)TRSC-PEG4-MspA-NTA-Ni and (**c**) F(AzK)TRSC-PEG4-MspA-NTA-Ni. The chemical structures of 2-naphthyl-L-alanine (NA) and 6-azido-L-lysine (AzK) are shown on the left. (**b**,**d**) Representative traces acquired with (**b**) F(NA)TRSC and (**d**) F(AzK)TRSC. The measurements were performed as described in the Methods. For each peptide, five enzymatic cleavages were sequentially observed, generating six stages, respectively referred to as aa1–aa6, with distinct binding events. The corresponding N-terminal amino acid of each stage is labelled above each event.



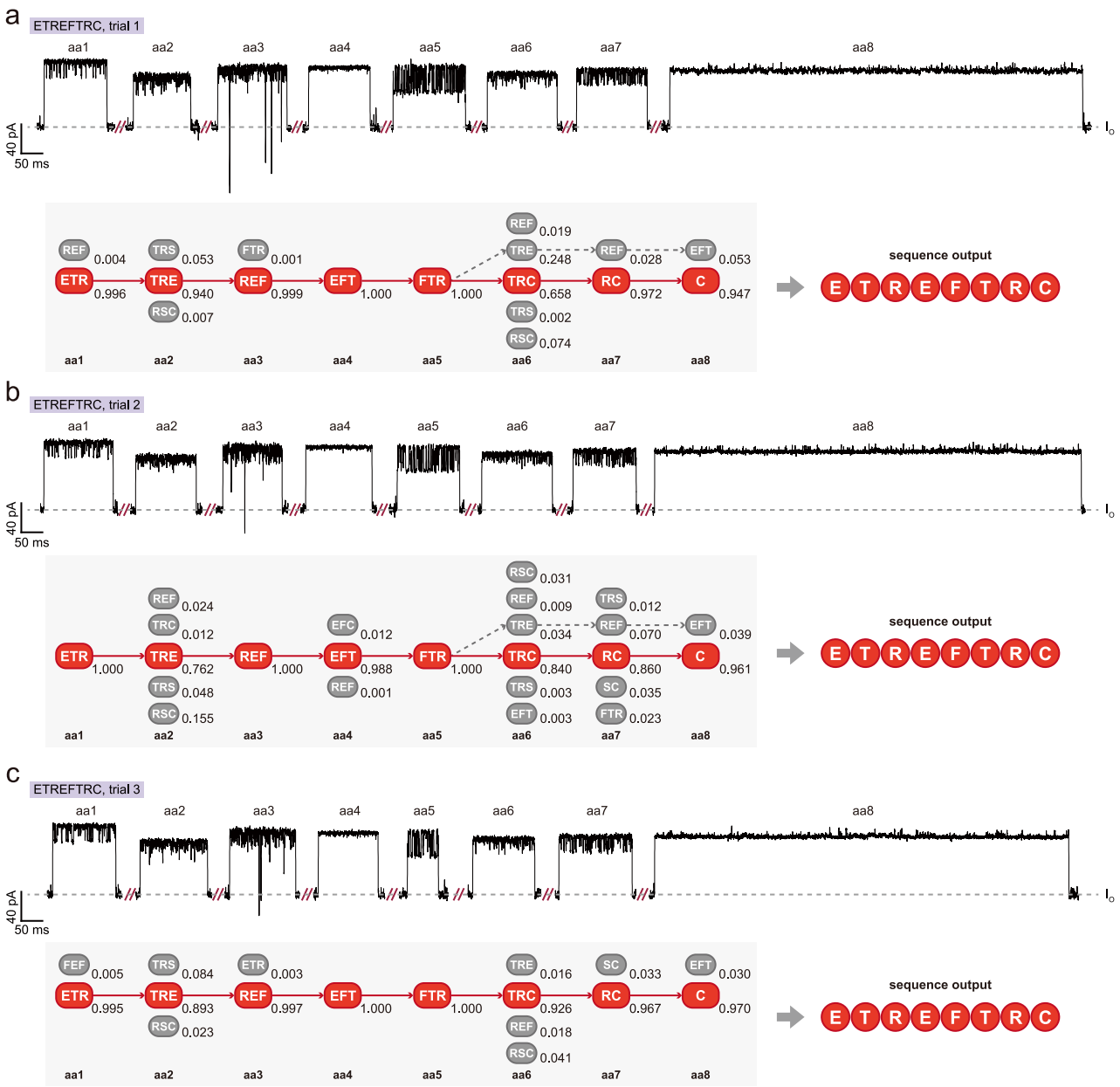
Extended Data Fig. 7 | tPAL sensing of D-peptides. (**a**) Schematic illustration of tPAL sensing with progressively shortened D-peptides, including D-FYAESC, D-YAESC and D-AESC. (**b**,**d**,**f**) Left: Representative tPAL sensing traces for (**b**) D-FYAESC, (**d**) D-YAESC, (**f**) D-AESC. All measurements were conducted in a 1.5 M KCl, 10 mM CHES buffer at pH 9.0, with a constant applied voltage of +100 mV. I_o denotes the current level in the unbound state. Right: Representative tPAL sensing events for (**b**) D-FYAESC, (**d**) D-YAESC, (**f**) D-AESC. (**c**,**e**,**g**) Scatter plots of Δ*I* versus SD obtained from (**c**) D-FYAESC-PEG4-MspA-NTA-Ni, (**e**) D-YAESC-PEG4-MspA-NTA-Ni, and (**g**) D-AESC-PEG4-MspA-NTA-Ni. Each scatter plot encompasses 200 events.



Extended Data Fig. 8 | Sequential cleavage of EFEFEFC and FEFEFEC. (**a**,**b**) Representative events acquired during the sequential cleavage of (**a**) EFEFEFC and (**b**) FEFEFEC. The measurements were conducted as described in the Methods using sgAP-dA15-chol. For each peptide, six enzymatic cleavages were observed, yielding seven distinct stages, which are respectively referred to as aa1–aa7. The corresponding N-terminal amino acid of each stage is labelled above the trace. (**c**) Plot of $\overline{SD}$ at the aa1–aa6 stages derived from EFEFEFC and FEFEFEC sequential cleavage. All statistical data were derived from results of three independent measurements. The error bars represent standard deviation values.



Extended Data Fig. 9 | Sequential cleavage of AAAAC. (**a**) Schematic (top) and representative events (bottom) acquired during AAAAC sequential cleavage. (**b**) The scatter plot of Δ*I* versus SD for tPAL events acquired from AAAAC sequential cleavage. (**c**,**d**) Left: raw trace of tPAL event changes from AAAAC to AAAC. Right: zoomed-in view of the area indicated by a red arrow in the left trace. A slight event amplitude increase was observed irreversibly during the transition from AAAAC to AAAC, marking the cleavage of a terminal A. This phenomenon was validated in independent trials (trial 1 (**c**) and trial 2 (**d**)), confirming its robustness.



Extended Data Fig. 10 | tPAL measurements of ETREFTRC and sequence assembly. a–**c**, Results were obtained from three independent experiments (trial 1 (**a**), trial 2 (**b**) and trial 3 (**c**)). All measurements were conducted according to the Methods. Top: representative traces acquired with ETREFTRC during sequential cleavage. Seven enzymatic cleavages were identified during the sequential cleavage of ETREFTRC, leading to eight distinct stages. Bottom: sequence assembly graph. Each node in the graph represents a candidate *k*-mer predicted from the sensing events at corresponding stage, with their probability annotated alongside. Directed edges represent valid connections in which the suffix of a preceding *k*-mer exactly matches the prefix of the current *k*-mer. The optimal path, highlighted in red, is defined as the route exhibiting the highest cumulative probability while traversing exactly one node per stage. All remaining nodes and suboptimal paths are depicted in grey. The final peptide sequence was derived by sequentially merging the *k*-mers along this optimal path.

nature portfolio

Corresponding author(s): Shuo Huang

Last updated by author(s): May 15, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Single channel recordings were performed by Axon 200B and the accompanied commerically available software Axonpatch 10.7
Data analysis	Data analysis was performed with Origin Pro 2021. Event identification was performed by machine learning. The code and training data used for machine learning model building were submitted as a zip file named "tPAL_decoder"

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The machine learning training data is provided as a supplementary information file, referred to as “tPAL_decoder”. Datasets that support the findings of this study are available from the corresponding authors upon request. Source data are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N.A.
Reporting on race, ethnicity, or other socially relevant groupings	N.A.
Population characteristics	N.A.
Recruitment	N.A.
Ethics oversight	N.A.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	A sample is defined as a completely independent trial carried out with a specific experiment condition. At least three independent samples were analyzed for each experiment/condition.
Data exclusions	No data were excluded from the analysis.
Replication	Replication is based on quantitative outcomes of experimental results. Replication of results can be clearly evaluated according to results presented in the manuscript.
Randomization	Not applicable to this study; does not include subjects that require allocation into experimental groups.
Blinding	Our experiment didn't involve subjective trials. Blinding is not necessary in this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	N.A.
Novel plant genotypes	N.A.
Authentication	N.A.