Research
A SCIENCE PARTNER JOURNAL

RESEARCH ARTICLE

Direct Identification of O-Glycopeptides by Low-Temperature Assisted Nanopore Technique

Jia-Hong Wang[1†], Wenjing Ma[2†], Zheng-Li Hu[1], Zhaobing Gao[3*], Yi-Tao Long[1], Tiehai Li[2*], and Yi-Lun Ying[1,4*]

[1]Molecular Sensing and Imaging Center, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210023, P. R. China. [2]State Key Laboratory of Chemical Biology, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, P. R. China. [3]State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, P. R. China. [4]Chemistry and Biomedicine Innovation Center, Nanjing University, Nanjing 210023, P. R. China.

*Address correspondence to: yilunying@nju.edu.cn (Y.-L.Y.); tiehaili@simm.ac.cn (T.L.); zbgao@simm.ac.cn (Z.G.)
†These authors contributed equally to this work.

Citation: Wang JH, Ma W, Hu ZL, Gao Z, Long YT, Li T, Ying YL. Direct Identification of O-Glycopeptides by Low-Temperature Assisted Nanopore Technique. *Research* 2025;8:Article 0850. https://doi.org/10.34133/research.0850

Submitted 19 May 2025
Revised 29 July 2025
Accepted 30 July 2025
Published 4 September 2025

Copyright © 2025 Jia-Hong Wang et al. Exclusive licensee Science and Technology Review Publishing House. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License (CC BY 4.0).

O-glycopeptides are highly expressed in various human cancers and play a key role in cancer progression and metastasis, making them promising biomarkers for early diagnostics. However, the inherent complexity and heterogeneity of glycans pose a major challenge for the simultaneous and precise analysis of multiple glycopeptides. Here, we developed a low-temperature nanopore technique capable of simultaneously discriminating 4 truncated O-glycopeptides with varied glycoforms. This method enables the direct identification and relative quantification of O-glycopeptides from a mixture, achieving a discrimination accuracy of 92.9%. This general strategy holds promise for the label-free analysis of glycopeptide biomarkers, with potential applications in cancer diagnostics.

Introduction

Protein glycosylation is a complicated posttranslational modification that mediates a diversity of cellular processes [1–3]. Mucin-type (GalNAc type) O-glycosylation is one of the most diverse glycosylation of secretory and membrane-bound proteins. Its alteration serves as a hallmark of cancer that has helped to shape the management and understanding of cancer [4]. Typically, the tumor-associated glycoprotein Mucin 1 (MUC1) is commonly overexpressed in various epithelial adenocarcinomas such as lung, liver, colon, breast, pancreatic, and ovarian cancers [5,6]. It comprises a highly variable number of glycosylated repeat tandem regions (VNTRs), a transmembrane region, and other segments [7]. Due to reduced glycosyltransferase activity with simultaneously increased sialyltransferase activity in the tumor tissue, the O-glycosylation of MUC1 tends to be altered from long branched chains to short structures and capped with terminal sialic acids [8]. This process is known as truncated O-glycosylation (Fig. 1A). A representative truncated mucin-type O-glycan is the Thomsen–Friedenreich antigen (also known as core 1, TF), which forms the core of many long and complex structures [9]. The biosynthetic pathway of TF is as follows: First, a GalNAc residue is linked to a serine or threonine residue by polypeptide N-acetylgalactosaminyltransferases (ppGalNAc-Ts) to form Tn antigen (GalNAcα-) (Fig. 1A, step i). Next, the Tn antigen is further converted to TF (Galβ1-3GalNAcα) or sialyl-Tn (Neu5Acα2-6GalNAcα-, STn), which is catalyzed by the enzyme core 1 β1,3-Gal-transferase (core 1 GalT) or sialyltransferase ST6GalNAc1, respectively (Fig. 1A, step ii). After that, TF can be further elongated into a series of glycan structures, or be sialylated to form sialyl-TF (Neu5Acα2-3Galβ1-3GalNAcα-, STF), which is catalyzed by the enzyme α2,3-sialyltransferases (ST3Gal1) (Fig. 1A, step iii) [8]. Recognition of multiple glycosylated variants of peptides is highly required for reliable cancer diagnostics and prognosis. Strategies for glycopeptide detection include enzymatic or chemical cleavage methods to release glycans from glycopeptides [10,11]. The enzymatic methods are generally selective for certain specific glycan linkages or glycoprotein substrates, whereas chemical methods (e.g., base-mediated β-elimination) may cause unintended side reactions, complicating glycan analysis [12]. Moreover, these approaches often require costly enzymes or careful handling of reactive chemicals. As an alternative, direct characterization of glycopeptides offers valuable insights into site-specific glycosylation, commonly carried out by liquid chromatography–tandem mass spectrometry (LC-MS/MS) or immunoassays [2,13]. For mass spectrometry, ionization and detection efficiency of glycopeptides vary with their chemical properties, leading to potential uncertainty. For instance, sialic acid modifications at glycan terminal can result in lower mass intensity compared to other glycopeptides at equimolar concentrations [14]. Immunoassays, on the other hand, require antibody or lectin labeling, and may be influenced by factors such as glycan binding affinity and potential cross-reactivity with structurally similar glycans [15]. Given

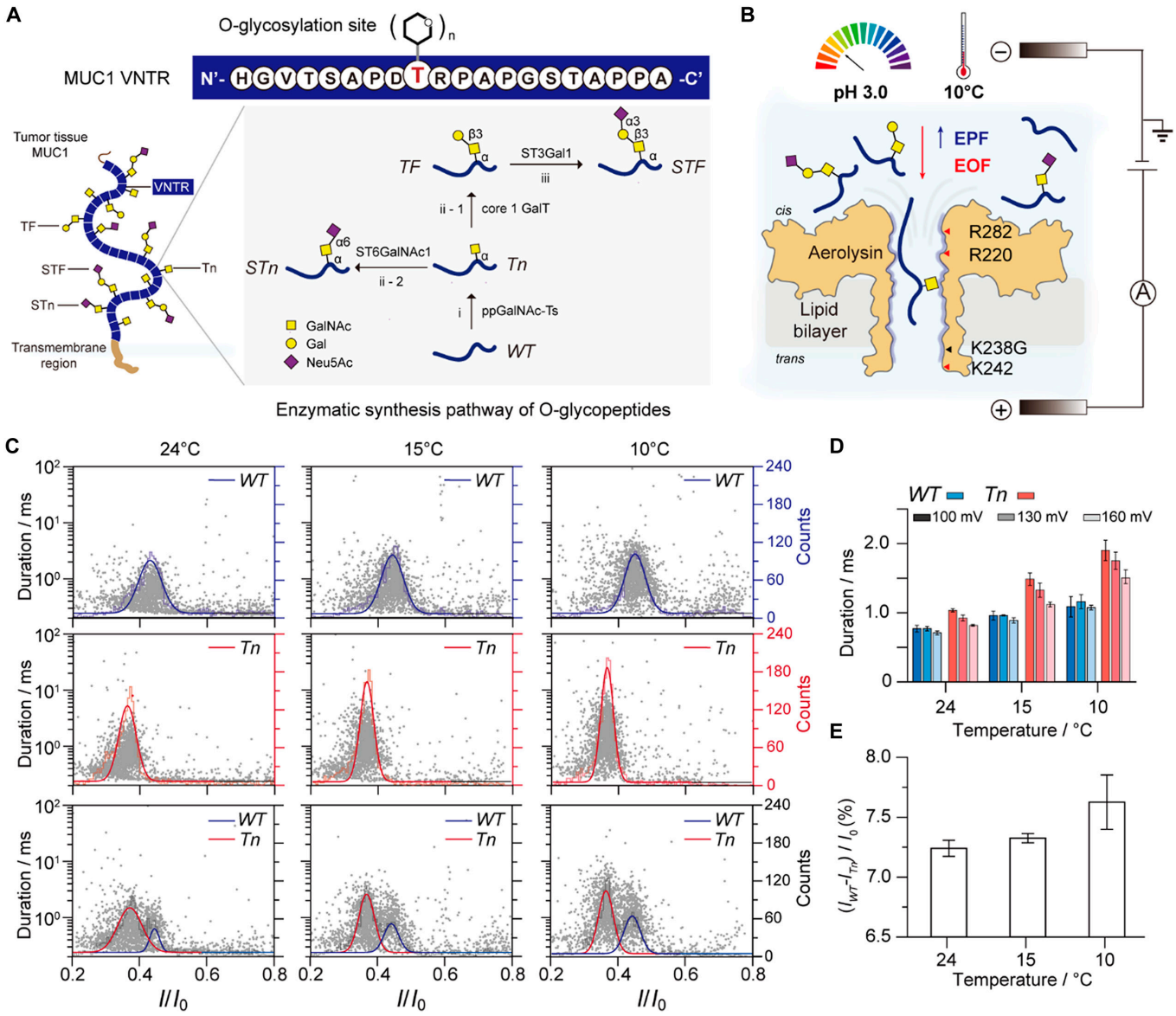


Fig. 1. A low-temperature nanopore approach for identification of multiple O-glycopeptides. (A) Schematic diagram of truncated O-glycosylation of tumor-associated glycoprotein Mucin 1 (MUC1). MUC1 comprises the glycosylated repeat tandem region (VNTR, blue) with the sequence of N′-HGVTSAPDTRPAPGSTAPPA-C′. The enzymatic synthesis pathway of core 1 glycan (TF) at MUC1 is shown at the bottom. Firstly, a GalNAc residue is linked to a threonine (T) or serine (S) residue by polypeptide N-acetylgalactosaminyltransferases (ppGalNAc-Ts) to form the Tn antigen (GalNAcα-) via step i. Next, the Tn antigen is converted to the core 1 antigen (TF, Galβ1-3GalNAcα) by core 1 β1,3-Gal-transferase (core 1 GalT) (step ii-1) or to sialyl-Tn (Neu5Acα2-6GalNAcα-, STn) (step ii-2) by sialyltransferase ST6GalNAc1. The core 1 glycan can then be elongated to other structures or sialylated to form sialyl-TF (Neu5Acα2-3Galβ1-3GalNAcα-, STF) (step iii) by α2,3-sialyltransferases (ST3Gal1). (B) Schematic representation for the detection of single glycopeptide using the K238G Aerolysin at pH 3 and 10 °C. The bias potential is applied using Ag/AgCl electrodes, and the *cis* compartment is defined as a virtual ground. The negatively charged amino acids of the nanopore are prone to be protonated at pH 3, exposing the positively charged amino acids of R282, R220, and K242 at the lumen (red triangle). The lysin (K) is mutated to glycine (G) at the 238 site (black triangle). Under a positive applied potential, the electrophoretic force (EPF) and electroosmotic flow (EOF) acting on the peptide are in opposite directions. The experimental details are shown in Materials and Methods. (C) Scatterplots of I/I_0 versus duration and histograms of I/I_0 for WT (upper row), Tn (middle row), and mixture of WT and Tn (bottom row) at temperatures of 24 °C (left), 15 °C (middle), and 10 °C (right). I and I_0 represent the blockage current and open-pore current, respectively. The blue and red curves represent the Gaussian fitting to the I/I_0 histograms for WT and Tn peptides, respectively. The light blue line represents the outline of the histogram of WT, and the light red line represents the outline of the histogram of Tn. The gray line represents the outline of the histogram of mixture. (D) Temperature dependence of the duration for WT (blue) and Tn (red) peptides at various voltages. (E) Temperature dependence on the current peak separation of $(I_{WT} - I_{Tn})/I_0$. The concentration of WT and Tn peptides is fixed at 10 μM, and the electrolyte solution was 1 M KCl, 10 mM tris, and pH 3. Error bars depict the standard deviation (SD) of 3 independent measurements.

these considerations, there remains a need for approaches that enable the direct identification of O-glycopeptides in a label-free manner.

Nanopore-based single-molecule sensing utilizes the measurement of ionic current flow caused by individual analytes to detect their structure, composition, and length. Due to the confined space, nanopores exhibit high sensitivity in analyzing DNA [16], peptides [17–23], proteins [24–27], oligosaccharides [28–33], and other small molecules [34,35]. This approach offers the advantages of being label-free and enabling rapid detection.

However, there are still challenges in detecting glycopeptides with nanopore techniques. Unlike posttranslational modifications such as phosphorylation, acetylation, and methylation, glycosylation involves larger and more complex chemical structures [36]. Recently, only few studies have reported nanopore-based glycoprotein detection [37–39], with most focusing on identifying a monosaccharide modification on peptides (e.g., glucose). Glycopeptides bearing oligosaccharide modifications (such as disaccharides or trisaccharides) face larger entropic barriers during nanopore translocation compared to other post-translational modifications (e.g., phosphorylation or acetylation), due to the greater steric hindrance and conformational restrictions imposed by the bulky glycan moieties. Additionally, glycopeptides differing by a single glycan residue often generate similar ionic current events, making them more challenging to distinguish using nanopore sensing. To address these challenges, we developed a general strategy using a low-temperature and low-pH aerolysin nanopore technique to enhance the capture capability for O-glycopeptides with monosaccharide to trisaccharide modifications. It also extends the interaction time inside the nanopore and thereby improves the discrimination ability of varied O-glycopeptides. This method enables direct, label-free identification and quantification of O-glycopeptides, achieving a 92.9% discrimination accuracy, and holds promise for potential application in cancer diagnostics.

Results and Discussion

Fine-tuning temperature and pH conditions for single O-glycopeptide detection

As a proof of concept, herein, we selected truncated O-glycosylated variants (Tn, STn, TF, STF) of MUC1 peptides (Fig. 1A). Considering the neutral net charge of the MUC1 peptides, we first employed N226Q/S228K aerolysin (AeL) due to its good performance in the capture and sensing of neutrally charged peptides [40]. All experimental details and characterization of glycopeptides and nanopores are shown in Materials and Methods and Figs. S1 to S3. The wild-type (WT) peptide can be captured by N226Q/S228K AeL. However, the WT peptide exhibited relatively short event durations, with a slight increase observed as the voltage increased (Figs. S4 and S5). Structural prediction indicated that residues 11 to 15 of the WT MUC1 peptide may form a loop structure (Fig. S6). This is likely due to proline, which introduces a kink in the backbone and promotes β-turn formation [41–43]. Since N226Q/S228K AeL shows a small diameter of around 0.8 nm at its constriction [40], the partially folded structure of the peptides may hinder their translocation through the nanopore. Compared to N226Q/S228K AeL, K238G AeL shows a relatively larger diameter of around 1.1 nm at the constricted region [44], making it potentially more suitable for O-glycopeptide detection. However, K238G AeL hardly captures the WT peptide in pH 7.4, due to its weak electroosmotic flow in the inward direction (from *cis* to *trans*) (Fig. S7). To detect the neutrally charged and partially folded MUC1 peptides, we employed the K238G AeL nanopore under acidic conditions (pH 3) to enhance electroosmotic flow (EOF). At low pH, glutamate and aspartate residues are protonated, leading to the disruption of interactions between paired acidic and basic residues (R282:D216, R220:D222, K238:E258, K242:E254) [21,45,46], which in turn strengthens EOF and facilitates peptide capture (Fig. 1B and Fig. S8). Acidic conditions may also disrupt intramolecular interactions within the folded peptide [47], making the peptide more prone to be captured by the nanopore. As shown in Fig. 1C, MUC1 peptides are effectively captured at pH 3 with the K238G AeL, due to the protonation-induced exposure of up to 21 positively charged residues in the heptameric K238G AeL under acidic conditions, thereby further enhancing EOF for peptide capture efficiency (Table S1).

To demonstrate the versatility of this approach, we further detected the Tn peptide, which was N-acetylgalactosaminylated. Due to the larger size of Tn, it causes a deeper current blockage than the unmodified WT peptide (Fig. 1C). The study on voltage dependence showed that the WT peptide could traverse the K238G nanopore at voltages above +160 mV, at pH 3 and 24 °C (Fig. 1D). Interestingly, the voltage dependence revealed that N-acetylgalactosaminylation of MUC1 peptides facilitates their passage through the nanopore. To further validate the discrimination ability of this method, a mixture of WT and Tn was analyzed. As shown in Fig. 1C, the events of WT and Tn fall into two Gaussian distributions of I/I_0, with WT located at $I/I_0 = 0.44$ and Tn at $I/I_0 = 0.37$. However, the peak separation needs to be further improved due to wide and overlapping event clusters. Previous studies showed that the prolonged duration of the nanopore blockade would lead to a narrower I/I_0 distribution with high peak separation [48,49]. Therefore, we further lowered the solution temperature from 24 °C to 10 °C to prolong the duration. The results showed that the durations of WT and Tn peptides increased by 1.5 and 1.8 times at +160 mV, respectively (Fig. 1D). As a result, at 10 °C, the peak separation between WT and Tn was improved compared to that at 24 °C (Fig. 1E and Figs. S9 and S10), proving that lowering temperature enhances the sensing ability of nanopores. Interestingly, at 10 °C, the duration of the Tn peptide showed a decrease with increasing voltage, consistent with the trend observed at 24 °C (Fig. 1D). These results suggest that Tn peptides may translocate across the nanopore even at lower temperatures. In contrast, the WT peptide demonstrated limited sensitivity to voltage changes across the examined temperature range. The temperature-dependent experiments further revealed the enthalpic and entropic changes associated with peptide–nanopore interactions (Fig. S11). Compared to the unmodified peptide, the Tn peptide modified with N-acetylgalactosamine exhibited a greater enthalpy change and a lower entropy change inside the nanopore. Moreover, the capture frequency of Tn is positively correlated with the applied voltage. The high voltage of +160 mV also results in a good signal-to-noise ratio of 17.3 (Fig. S12 and Table S2). Thus, the optimal measurement conditions for distinguishing the O-glycopeptides with K238G AeL are +160 mV, 10 °C, and pH 3. The low-pH and low-temperature activated K238G AeL not only enhances the capture capability but also strengthens the interaction with the glycopeptides. Note that the open-pore current is stable for about at least 100 min at this experimental condition, ensuring the long-time and stable sensing (Fig. S13).

Identification of various truncated O-glycopeptides

Subsequently, we employed this method to detect 4 types of truncated O-glycopeptides with different modifications, which are Tn, TF, STF, and STn. As shown in Fig. 2A and B, all the O-glycopeptides could be efficiently captured by K238G AeL and produced distinct blockage events. Since O-glycopeptides could interact with the bilayer [50] (Fig. S14), the open-pore current shows slight fluctuations following nanopore insertion and peptide addition, especially for TF and STn. Therefore, Gaussian fitting was applied to evaluate baseline noise and

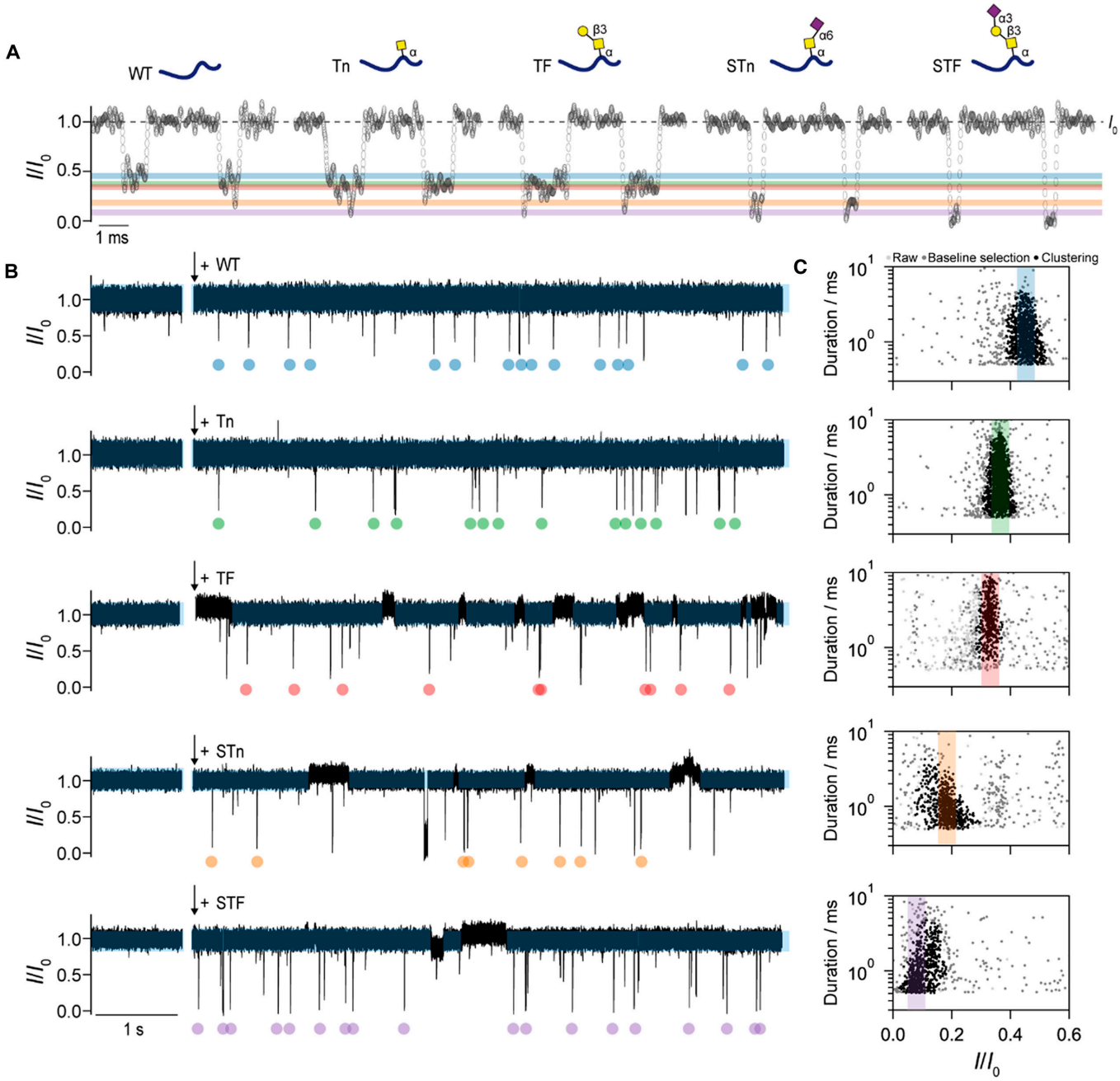


Fig. 2. Identification of truncated O-glycopeptides with K238G AeL at 10 °C. (A) Representative raw events of WT, Tn, TF, STn, and STF peptides, respectively (from left to right). (B) Typical current traces before and after addition of WT, Tn, TF, STn, and STF peptides (from top to bottom). The selected range of baseline is shown in blue band covering the current trace. The colored dot labels the characteristic event from the related truncated O-glycopeptides. (C) Scatterplots of I/I_0 versus duration. The light gray, gray, and black dots represent the raw events, events after baseline selection, and events after clustering of the truncated O-glycopeptides corresponding to (B). The center of color bands highlights the statistical peak I/I_0 values obtained from Gaussian fitting (see Fig. S16), which are 0.45 for WT (blue), 0.36 for Tn (green), 0.33 for TF (pink), 0.18 for STn (orange), and 0.10 for STF (purple), respectively. Gray scatter points may result from baseline fluctuations due to O-glycopeptide–bilayer interactions and/or trace impurities arising from the synthetic complexity of larger glycans. The experiments were performed in 1 M KCl, 10 mM tris, and pH 3 under an applied voltage of +160 mV.

minimize its interference in the statistical analysis of single-peptide events (Fig. S15). Subsequently, we clustered the events using the HDBSCAN algorithm to remove outliers and noise (Fig. S16, Materials and Methods). The results showed that STn and STF peptides showed 2 population in scatterplots, which could be attributed to the electrostatic interactions between the sialic acid carboxyl group of the peptide and arginine residues inside the nanopore, similar to the results from sialylated glycans in α-hemolysin [31]. This difference may also result from conformational changes of the glycopeptides inside the nanopore. As shown in Fig. 2C, these 5 peptide variants differ in both blockage current and duration, demonstrating the distinguishability of the activated K238G AeL. The shorter durations of sialylated glycans (STn and STF) may be attributed to their partially negative charge at pH 3, which likely enhances electrophoretic transport under +160 mV.

Discrimination of O-glycopeptides by machine learning

To further achieve automatic identification of 4 types of truncated O-glycopeptides, we employed a supervised machine learning algorithm to recognize these peptides (Materials and Methods and Fig. 3). The entire machine learning workflow includes data acquisition, feature extraction, and model training (Fig. 3A). A total of 2,400 events from 5 pure peptide samples detected by K238G AeL were collected. From these blockade events, 4 statistical features were extracted to form the feature matrix, which are I/I_0, duration, standard deviation (SD), and skewness. Subsequently, 80% of the dataset was extracted as the training set and 20% as the test set.

To prevent overfitting, 10-fold cross-validation was used during model training and validation. We trained 7 models, including support vector machine (SVM), k-nearest neighbors (KNN), Random forest, Adaboost, Naïve Bayes, Gaussian process, and multilayer perceptron (MLP), and then evaluated their classification performance based on the area under the receiver operating characteristic curve (AUC) and accuracy (Fig. 3A). Among these models, the MLP model performed the best, achieving the highest AUC (0.995) and relatively high accuracy (92.9%) on the test set. The feature contribution proved that I/I_0 contributed the most to discrimination (74.5%), while the contribution of SD, duration, and skewness are relatively slight, which are 6.3%, 15.0%, and 4.2%, respectively (Fig. 3B). Additionally, the learning curve demonstrated that the model's accuracy reached 92.6% at 570 events, proving that our data size is entirely sufficient (Fig. 3C). The calibration curves align closely with the diagonal, and both the precision–recall (average precision > 0.93) and receiver operating characteristic (AUC > 0.98) curves show good performance for all peptide classes (Fig. S18). The confusion matrix generated by the MLP model showed excellent performance in classifying all classes, and the highest true positive rate is 97.2% for WT peptide.

Relative quantification in the mixture of various O-glycopeptides

To validate the effectiveness of the machine learning model, we tested mixtures with different molar ratios. The current traces and scatterplots of the mixtures composed of 5 types of peptides are shown in Fig. 4A and B. The trained model, incorporating the capture efficiency of pure peptides, allows identification and relative quantification of the glycopeptides (Fig. 4C, Materials and Methods, and Table S3). As shown in Fig. 4C, the proportions predicted by the machine learning model were consistent to the true proportions for all mixtures. In particular, STn, a clinical cancer biomarker [51], showed recoveries of 99.4% and 93.3% in the STF:STn:TF:Tn:WT mixtures (0:1:1:0:0) and (1:1:1:0:0), respectively (Fig. 4C). The minor difference in the predicted

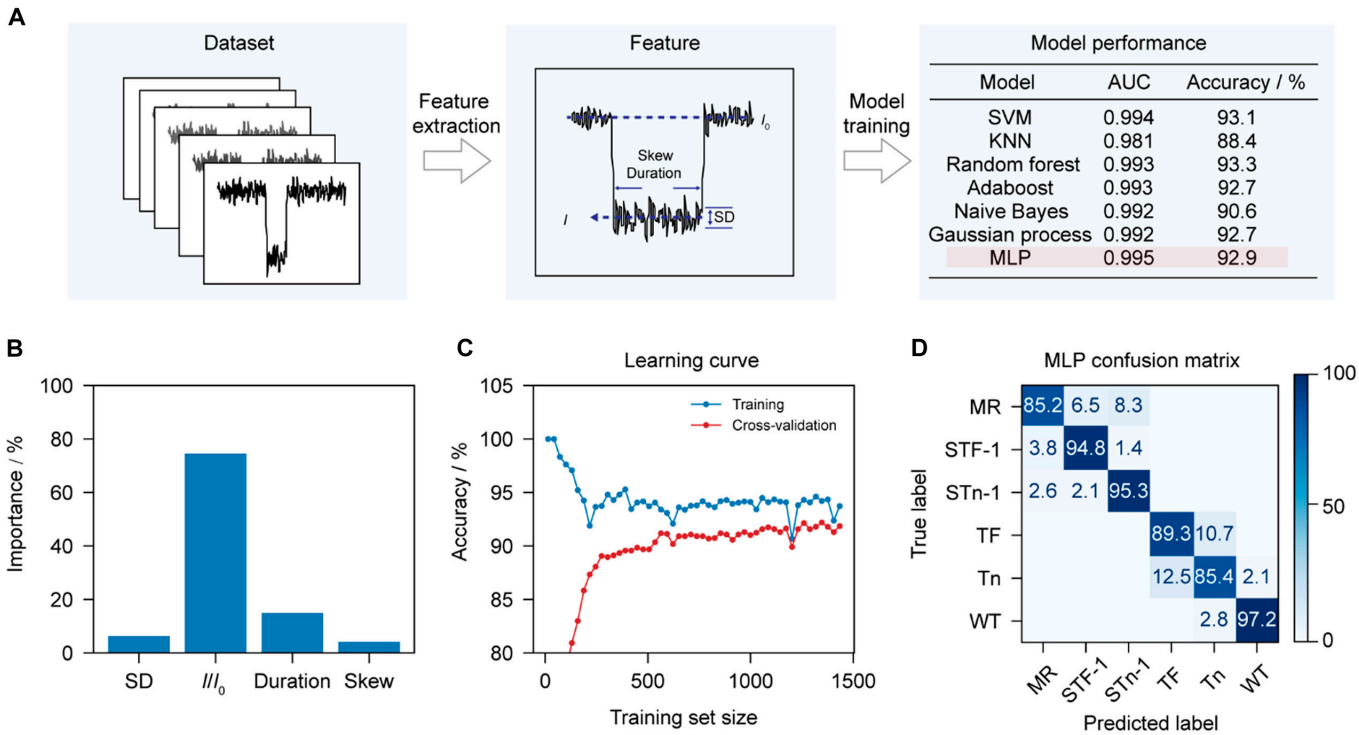


Fig. 3. Identification of various O-glycopeptides by machine learning. (A) Workflow of the machine learning. A total of 2,400 events in the dataset were used, which include WT, Tn, TF, STn-1, and STF-1 events and mix events. Because both STn and STF events showed 2 distributions in the scatterplot, the STn and STF events were clustered into 2 classes, labeled STn-1, STn-2 and STF-1, STF-2, respectively. The STn-2 and STF-2 events, which show substantial overlap, were merged into a single class called merging region (MR) events. Events from the MR class were ignored during statistics. More details of the data analysis and feature definition can be found in Materials and Methods and Fig. S17. Four features (SD, skewness, duration, and I/I_0) were extracted from ionic current events to form a feature matrix. Here, SD represents the standard deviation of the current amplitude within the indicated range of the event, while Skew refers to its skewness. Seven models were employed separately, and their corresponding AUC values and test set accuracies were reported. Among these, the multilayer perceptron (MLP) model performed the best. (B) Importance of 4 features (SD, I/I_0, duration, and skewness) in model training. (C) Learning curve of the MLP model produced by performing model training and validation using varying sample size. (D) Confusion matrix for the classification of various truncated O-glycopeptides generated by the MLP model.

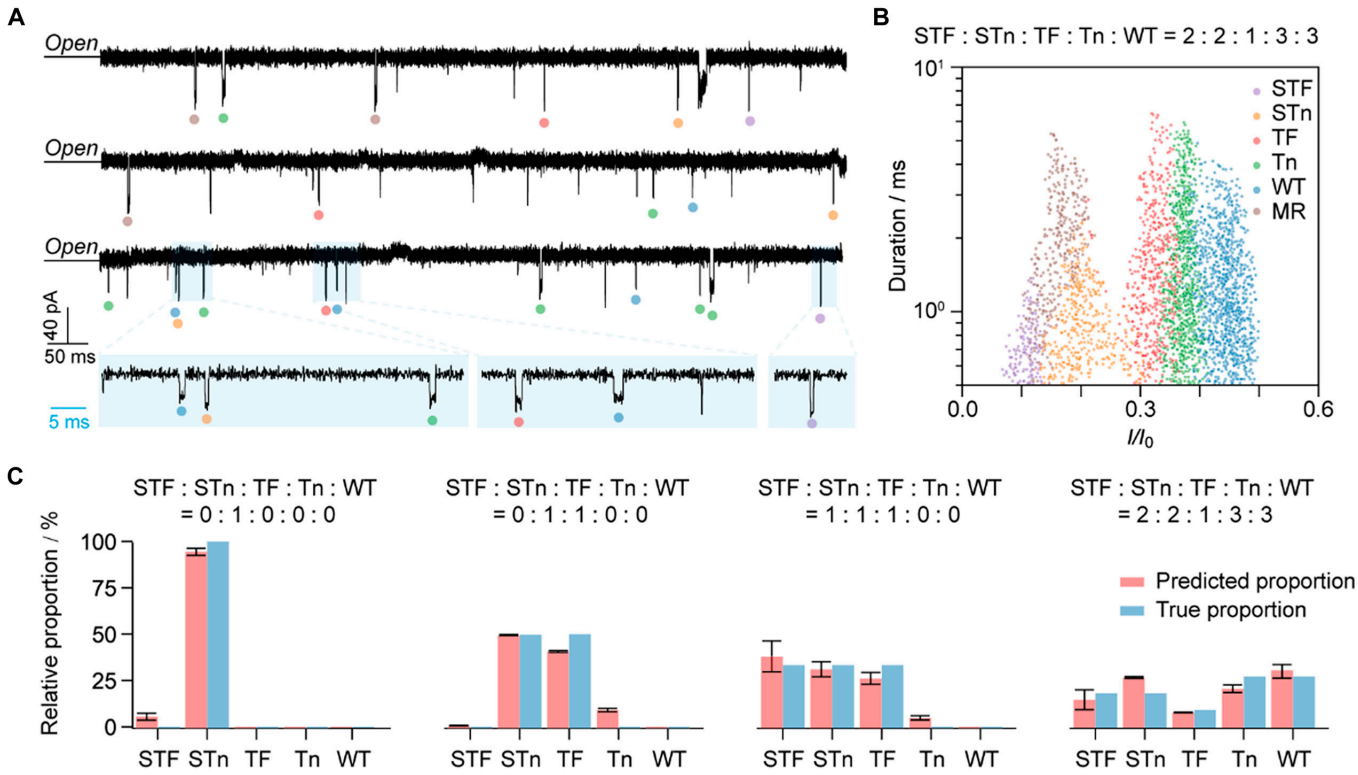


Fig. 4. Validation of the effectiveness of developed machine learning model. (A) Current traces of the mixture composed of 5 types of O-glycopeptides. The labels were automatically generated by the MLP model. The blue area represents the zoom-in view of the current trace. All current traces are obtained from the same experiment. (B) Scatterplots of I/I_0 versus duration of the mixture corresponding to (A). The label was predicted by the MLP model. The blockade ratio (I/I_0) and its statistical SD for each predicted peptide in the mixture were listed in Table S4. (C) Comparison of the predicted proportion and true proportion of the various peptides in mixtures. The molar proportions of these analytes were STF:STn:TF:Tn:WT of 0:1:0:0:0, 0:1:1:0:0, 1:1:1:0:0, and 2:2:1:3:3 (from left to right), respectively. The measurement was performed in 1 M KCl, 10 mM tris, and pH 3 under an applied voltage of +160 mV at 10 ± 1 °C.

proportion and true proportion may be attributed to the potential competitive capture among these substances present. To validate the stability of this method in a complex background, we analyzed O-glycopeptides in 0.2% cell lysate and 1% bovine serum albumin (BSA), respectively. The scatterplots from the machine learning predictions demonstrated that the method could identify the components of the mixture (Figs. S19 to S21). These glycosylated variants have been reported with varying expression levels across different types of cancers [52,53]. Our method can recognize and quantify these variants in complex mixtures and may help identify these glycopeptides in physiological samples.

Conclusion

In summary, we developed a highly sensitive and label-free nanopore strategy for the identification and relative quantification of multiple O-glycopeptides. This versatile approach utilizes low-pH and low-temperature conditions to activate K238G AeL, enhancing its capture and sensing capabilities for the direct detection of glycopeptides. Notably, glycopeptides can be identified within just 20 min of recording. The accuracy of the machine learning model reaches 92.9%, with an AUC value of 0.995. The development of nanopore array chip [54] could enable high-throughput single-molecule sensing of complex glycopeptide variants, opening up potential applications for glycosylation-based biomarker detection.

To enable practical applications, it is critical to address the complexity of sample pretreatment and ensure reproducibility in real biological samples. Moreover, the reported concentrations of truncated O-glycosylated peptide variants in serum are around the nanomolar range, posing a challenge for nanopore detection sensitivity [55]. A possible solution involves integrating MUC1-specific antibody-based enrichment prior to nanopore detection, allowing for the selective isolation of target glycopeptides from complex biological matrices. Additional optimization is likely required to integrate the system into a filter membrane-based device for reliable sample analysis. To better analyze longer and more complex glycan chains, we are exploring enzymatic or chemical cleavage methods to release and individually detect glycans and peptides using aerolysin nanopore. Future advancements, such as de novo protein design [56], may help develop elegant nanopore interfaces capable of recognizing larger glycan branches and more intricate glycopeptides. Since the glycosylation of MUC1 involves multiple glycosyltransferases, real-time monitoring of glycopeptide substrates and products by nanopore sensing may provide useful insights into enzymatic pathways. When combined with nanopipette-based single-cell analysis, this approach could potentially be applied to probing glycosyltransferase activity and recognition of the glycopeptide at the single-cell level.

Materials and Methods

Reagents and materials

Tris (hydroxymethyl) aminomethane (tris, ≥99.9%), hexadecane, hexane, trypsin-agarose, BSA, potassium chloride (≥99%), and

decane (anhydrous, ≥99.8%) were purchased from Sigma-Aldrich Co. Ltd. (St. Louis, USA). 1,2-Diphytanoyl-sn-glycero-3-phosphocholine (powder) was purchased from Avanti Polar Lipids Inc. (Alabaster, USA). Hydrochloric acid (HCl) was purchased from Nanjing Chemical Reagent Co. Ltd. (Jiangsu, China). Radioimmunoprecipitation assay (RIPA) lysis buffer (moderate) was purchased from Beyotime (Shanghai, China). Sodium dodecyl sulfate–polyacrylamide gel electrophoresis analysis (SDS-PAGE) preparation kit was purchased from Sangon Biotech Co. Ltd. (Shanghai, China). To facilitate the formation of the lipid bilayer, the preparation solution was made by mixing hexadecane and hexane with a volume ratio of 1:99. Phospholipid solution with a concentration of 30.0 mg/ml was obtained by dissolving phospholipids in decane. The truncated mucin-type O-glycopeptides were synthesized according to our previous work [57]. The LC-MS characterizations are shown in Fig. S3. The concentrations of glycopeptides in the pure sample experiment were fixed at 10 μM (WT), 10 μM (Tn), 2 μM (TF), 5 μM (STn), and 3 μM (STF), respectively. In the mixture of STF:STn:TF:Tn:WT, the glycopeptide concentrations are as follows: 2 μM STn and 2 μM TF for the 0:1:1:0:0 mixture; 2 μM (STF), 2 μM (STn), and 2 μM (TF) for the 1:1:1:0:0 mixture; 2 μM (STn), 2 μM (STF), 2 μM (STn), 1 μM (TF), 3 μM (Tn), and 3 μM (WT) for the 2:2:1:3:3 mixture. The pH of the acidic buffer solution was measured using a pH meter (LE422, Mettler-Toledo, UK; accuracy ± 0.01) after the addition of 10 μl glycopeptides solution, yielding values of 3.00 at 24 °C, 3.04 at 15 °C, and 3.07 at 10 °C under homebuilt temperature controller (accuracy ± 1 °C). Unless noted, all solutions in this work were prepared with ultrapure water (18.2 MΩ·cm at 25 °C) obtained using a Milli-Q system (Massachusetts, USA). For cell extraction, about 10^6 gliacyte cells were collected and resuspended in 500 μl of 1× phosphate-buffered saline buffer. Then, the cells were pelleted and resuspended in ice-cold RIPA lysis buffer. After that, the cell lysate was centrifuged at 10,000 rpm for 5 min to remove debris and insoluble material. The supernatant was then continuously centrifuged by Amicon 30 K Ultra Centrifugal Filter Device (Merck KGaA, Darmstadt, Germany). Therefore, the molecules with a molecular weight of less than 30 kDa were almost excluded from the lysate. Then, the lysate was ready for nanopore experiments.

Proaerolysin preparation and characterization

Proaerolysin was prepared according to our previous work [58]. We reconstructed the K238G and N226Q/S228K mutant expression plasmids by polymerase chain reaction and transformed them into *Escherichia coli* BL21 (DE3) pLysS cells. Protein expression was induced by adding isopropyl-β-D-thiogalactoside to a 0.5 mM final concentration at 16 °C for 6 h. The collected K238G and N226Q/S228K proaerolysin proteins were characterized by 12% SDS-PAGE (Fig. S1). To obtain the activated monomers, proaerolysin was pretreated with a trypsin solution at a mass ratio of 10:1 (trypsin to proaerolysin) for 6 h at room temperature.

Single-channel recording

Nanopore experiments were performed according to our previous work [58]. The homemade nanopore sensing device consists of 2 compartments that are separated by a Teflon sheet (FP301200, Goodfellow Cambridge Ltd., UK), which contains an aperture of ~50 μm in diameter for lipid membrane formation. The 2 compartments are termed as the *cis* chamber and the *trans* chamber, where the *cis* chamber is grounded. Firstly, the aperture was pretreated with hexadecane solution for 20 min, and then ~1.5 μl of phospholipid solution was added to the surface of electrolyte solutions in both chambers. After the phospholipid is evenly dispersed on the solution surface, the air–solution interface is moved below and above the aperture using a syringe device to form a lipid bilayer. When a mechanically stable lipid membrane is obtained, approximately 1.0 μl of activated proaerolysin solution is injected near the aperture into the *cis* chamber. Once the insertion of a single AeL was observed, the analytes were added into the *cis* chamber. Then, the voltage was adjusted to +100, +130, and +160 mV, respectively, and the corresponding ionic current traces were recorded. The experimental temperature was controlled by a home-made temperature device. The nanopore currents were collected via a pair of Ag/AgCl electrodes by the nanopore instrument Cube-D2 (https://zenodo.org/records/11609574), and all data were recorded with a low-pass filter of 5 kHz and a sampling frequency of 100 kHz. The K238G AeL was characterized by *I*–*V* curves (Fig. S2).

Feature extraction and statistical analysis

The single-molecule ionic current events were extracted and analyzed by the software "PyNanoLab" (https://doi.org/10.5281/zenodo.11383973). In the statistical analysis, the I/I_0 values were extracted from the Gaussian fits of the residual current blockage (I/I_0) histograms, where I is the blockage current and I_0 is the open-pore current. The events with duration time less than 0.50 ms were excluded for minimizing the evaluation errors.

The average open-pore current (I_0) and SD of baseline (σ) for each recording were determined statistically, and a threshold of $I_0 - 5\sigma$ was used to define events for the target peptides in the experiments. The statistical duration was obtained by the Gaussian fitting of the log(duration) histogram. The peak separation is defined as $(I_{WT} - I_{Tn})/I_0$. The skewness is a measure of the asymmetry of a data distribution; its calculation formula is as follows:

$$\text{Skewness} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^3}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\right)^3} \quad (1)$$

where n represent the number of samples and x_i represents the value of every sample. From each event, 5 event features, that is, I/I_0, SD, skewness and duration, were extracted using the software "PynanoLab". The feature importance was calculated using the permutation importance method, which evaluates the impact of each feature on the model's predictive performance by measuring the change in accuracy when the feature values are randomly permuted [59]

According to the Eyring equation (Eq. 2), a linear fit of $\ln(k_{off}/T)$ versus $1/T$ allows for the calculation of the enthalpy (ΔH) and entropy (ΔS) changes for the interactions between the nanopore and analytes, based on the slope and intercept of the fitted line:

$$ln\left(\frac{k_{off}}{T}\right) = -\frac{\Delta H}{R}\cdot\frac{1}{T} + ln\left(\frac{k_B}{h}\right) + \frac{\Delta S}{R} \quad (2)$$

where $k_{off} = 1/\tau_{off}$, T is the experimental temperature, k_B is the Boltzmann constant, h is the Planck constant, and R is the gas constant. As shown in Fig. S11, the Tn and WT peptides exhibit

distinct temperature-dependent behaviors. The Tn peptide shows higher enthalpy (47.74 kJ/mol) and lower entropy (−20.3 J/mol·K) changes than the WT peptide (25.89 kJ/mol and −95.2 J/mol·K), indicating that the N-acetylgalactosamine modification enhances the interaction strength with the nanopore.

Clustering analysis by HDBSCAN algorithm

To remove noises from nanopore gating and interaction between peptide and the membrane, an HDBSCAN algorithm was used for cluster analysis of the recorded events. To preserve the most clustered population in the scatterplot of the I/I_0 versus duration, the min cluster size and the min samples were adjusted to appropriate values. The data before and after cluster analysis are shown in Fig. S16. In the comparison, the distribution of the events is similar, confirming the reliability of the analysis.

Both STn and STF show 2 distributions I/I_0 (Fig. S17). Thus, the events of STn and STF are clustered into 2 types by Agglomerative Clustering algorithm, which are STn-1, STn-2 and STF-1, STF-2, respectively. Then, the events of STn-2 and STF-2 are combined into one class named "merging region" (MR). Events from this class would be ignored for further statistical analysis, because they are more dispersed and highly overlapping (Fig. S17). The HDBSCAN algorithm and Agglomerative Clustering algorithm come from scikit-learn, performed with Python 3.10.

Machine learning

The dataset contains a total of 2,400 events, composed of 6 classes: WT, Tn, TF, STn-1, STF-1, and MR. The input data were randomly split into a training set (80% of the labeled dataset) and a testing set (20%) for model training and model testing. The data in the training set were first standardized and then applied to train 7 models, including SVM, KNN, Random forest, Adaboost, Naïve bayes, Gaussian process, and MLP. According to the 10-fold cross-validation accuracy, the models with the highest AUC value and accuracy were selected. A learning curve of the best model for varying training sample size was generated to evaluate whether the amount of data is sufficient and to determine if the model is overfitting. The feature importance was calculated using the permutation importance method, which evaluates the impact of each feature on the model's predictive performance by measuring the change in accuracy when the feature values are randomly permuted. A confusion matrix was generated using the testing set for model evaluation. All models come from scikit-learn, performed with Python 3.10.

The relative quantification values obtained by machine learning will be corrected based on the capture frequency of pure peptide samples. The results are shown in Table S3.

Acknowledgments

Funding: This research was supported by the National Key R&D Program of China (2022YFA1304604), the Shanghai Municipal Science and Technology Major Project, the National Natural Science Foundation of China (22334006, 22077130, and 22377134), and Excellent Research Program of Nanjing University (ZYJH004).

Author contributions: Y.-L.Y., T.L., and Z.G. conceived the concept and supervised the project. J.-H.W. carried out the experiments and performed data analysis with the assistance from Z.-L.H.. W.M. synthesized and characterized the glycopeptides. J.-H.W. and Y.-L.Y. prepared the figures. J.-H.W., Y.-L.Y., T.L., Z.G., and Y.-T.L. prepared the manuscript. All authors discussed the findings and revised the manuscript.

Competing interests: The authors declare that they have no competing interests.

Data Availability

All data supporting the findings of this study are included within the article and its Supplementary Materials, as well as from the corresponding author upon reasonable request.

Supplementary Materials

Figs. S1 to S21
Tables S1 to S4

References

1. Reily C, Stewart TJ, Renfrow MB, Novak J. Glycosylation in health and disease. *Nat Rev Nephrol*. 2019;15(6):346–366.
2. Bagdonaite I, Malaker SA, Polasky DA, Riley NM, Schjoldager K, Vakhrushev SY, Halim A, Aoki-Kinoshita KF, Nesvizhskii AI, Bertozzi CR, et al. Glycoproteomics. *Nat Rev Methods Primers*. 2022;2:Article 48.
3. Chen C, Yi X, Liu P, Li J, Yan B, Zhang D, Zhu L, Yu P, Li L, Zhang J, et al. CD147 facilitates the pathogenesis of psoriasis through glycolysis and h3k9me3 modification in keratinocytes. *Research*. 2023;6:Article 0167.
4. Matthew RK, Tongzhong J, Jamie H-M, Richard DC. Simple sugars to complex disease—mucin-type O-glycans in cancer. In: Richard RD, Lauren EB, editors. *Glycosylation and cancer*. New York (NY): Academic Press; 2015. Vol. 126 p. 53–135.
5. Kufe DW. Mucins in cancer: Function, prognosis and therapy. *Nat Rev Cancer*. 2009;9(12):874–885.
6. Gao T, Cen Q, Lei H. A review on development of MUC1-based cancer vaccine. *Biomed Pharmacother*. 2020;132:Article 110888.
7. Taylor-Papadimitriou J, Burchell J, Miles D, Dalziel M. MUC1 and cancer. *Biochim Biophys Acta*. 1999;1455(2–3):301–313.
8. Pinho SS, Reis CA. Glycosylation in cancer: Mechanisms and clinical implications. *Nat Rev Cancer*. 2015;15(9):540–555.
9. Schjoldager KT, Narimatsu Y, Joshi HJ, Clausen H. Global view of human protein glycosylation pathways and functions. *Nat Rev Mol Cell Biol*. 2020;21(12):729–749.
10. Song X, Ju H, Lasanajak Y, Kudelka MR, Smith DF, Cummings RD. Oxidative release of natural glycans for functional glycomics. *Nat Methods*. 2016;13(6):528–534.
11. Kailemia MJ, Xu G, Wong M, Li Q, Goonatilleke E, Leon F, Lebrilla CB. Recent advances in the mass spectrometry methods for glycomics and cancer. *Anal Chem*. 2017;90(1):208–224.
12. Vos GM, Weber J, Sweet IR, Hooijschuur KC, Sastre Toraño J, Boons G-J. Oxidative release of O-glycans under neutral conditions for analysis of glycoconjugates having base-sensitive substituents. *Anal Chem*. 2023;95(23):8825–8833.
13. Ruhaak LR, Xu G, Li Q, Goonatilleke E, Lebrilla CB. Mass spectrometry approaches to glycomic and glycoproteomic analyses. *Chem Rev*. 2018;118(17):7886–7930.
14. Stavenhagen K, Hinneburg H, Thaysen-Andersen M, Hartmann L, Varon Silva D, Fuchser J, Kaspar S, Rapp E, Seeberger PH, Kolarich D. Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: An evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *J Mass Spectrom*. 2013;48(6):627–639.

15. Li Q, Li Z, Duan X, Yi W. A tandem enzymatic approach for detecting and imaging tumor-associated Thomsen-Friedenreich antigen disaccharide. *J Am Chem Soc*. 2014;136(36):12536–12539.
16. Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci USA*. 1996;93(24):13770–13773.
17. Sutherland TC, Long Y-T, Stefureac R-I, Bediako-Amoa I, Kraatz H-B, Lee JS. Structure of peptides investigated by nanopore analysis. *Nano Lett*. 2004;4(7):1273–1277.
18. Ying Y, Yang J, Meng F, Li S, Li M, Long Y-T. A nanopore phosphorylation sensor for single oligonucleotides and peptides. *Research*. 2019;8:Article 1050735.
19. Chen Z, Wang Z, Xu Y, Zhang X, Tian B, Bai J. Controlled movement of ssDNA conjugated peptide through mycobacterium smegmatis porin A (MspA) nanopore by a helicase motor for peptide sequencing application. *Chem Sci*. 2021;12(47):15750–15756.
20. Versloot RCA, Straathof SAP, Stouwie G, Tadema MJ, Maglia G. β-Barrel nanopores with an acidic-aromatic sensing region identify proteinogenic peptides at low pH. *ACS Nano*. 2022;16(5):7258–7268.
21. Stierlen A, Greive SJ, Bacri L, Manivet P, Cressiot B, Pelta J. Nanopore discrimination of coagulation biomarker derivatives and characterization of a post-translational modification. *ACS Cent Sci*. 2023;9(2):228–238.
22. Zhang Y, Yi Y, Li Z, Zhou K, Liu L, Wu H-C. Peptide sequencing based on host-guest interaction-assisted nanopore sensing. *Nat Methods*. 2024;21(1):102–109.
23. Zhang M, Tang C, Wang Z, Chen S, Zhang D, Li K, Sun K, Zhao C, Wang Y, Xu M, et al. Real-time detection of 20 amino acids and discrimination of pathologically relevant peptides with functionalized nanopore. *Nat Methods*. 2024;21:609–618.
24. Oukhaled G, Mathe J, Biance A-L, Bacri L, Betton J-M, Lairez D, Pelta J, Auvray L. Unfolding of proteins and long transient conformations detected by single nanopore recording. *Phys Rev Lett*. 2007;98(15):Article 158101.
25. Cressiot B, Oukhaled A, Patriarche G, Pastoriza-Gallego M, Betton J-M, Auvray L, Muthukumar M, Bacri L, Pelta J. Protein transport through a narrow solid-state nanopore at high voltage: Experiments and theory. *ACS Nano*. 2012;6(7):6236–6243.
26. Pastoriza-Gallego M, Rabah L, Gibrat G, Thiebot B, van der Goot FG, Auvray L, Betton J-M, Pelta J. Dynamics of unfolded protein transport through an aerolysin pore. *J Am Chem Soc*. 2011;133(9):2923–2931.
27. Oukhaled A, Bacri L, Pastoriza-Gallego M, Betton JM, Pelta J. Sensing proteins through nanopores: Fundamental to applications. *ACS Chem Biol*. 2012;7(12):1935–1949.
28. Ramsay WJ, Bayley H. Single-molecule determination of the isomers of D-glucose and D-fructose that bind to boronic acids. *Angew Chem Int Ed Engl*. 2018;57(11):2841–2845.
29. Zhang S, Cao Z, Fan P, Wang Y, Jia W, Wang L, Wang K, Liu Y, Du X, Hu C, et al. A nanopore-based saccharide sensor. *Angew Chem Int Ed Engl*. 2022;61(33):Article e202203769.
30. Li M, Xiong Y, Cao Y, Zhang C, Li Y, Ning H, Liu F, Zhou H, Li X, Ye X, et al. Identification of tagged glycans with a protein nanopore. *Nat Commun*. 2023;14(1):Article 1737.
31. Xia B, Fang J, Ma S, Ma M, Yao G, Li T, Cheng X, Wen L, Gao Z. Mapping the acetylamino and carboxyl groups on glycans by engineered α-hemolysin nanopores. *J Am Chem Soc*. 2023;145(35):18812–18824.
32. Yao G, Tian Y, Ke W, Fang J, Ma S, Li T, Cheng X, Xia B, Wen L, Gao Z. Direct identification of complex glycans via a highly sensitive engineered nanopore. *J Am Chem Soc*. 2024;146(19):13356–13366.
33. Yao G, Ke W, Xia B, Gao Z. Nanopore-based glycan sequencing: State of the art and future prospects. *Chem Sci*. 2024;15(17):6229–6243.
34. Zhu R, Qin F, Zheng X, Fang S, Ding J, Wang D, Liang L. Single-molecule lipopolysaccharides identification and the interplay with biomolecules via nanopore readout. *Biosens Bioelectron*. 2023;240:Article 115641.
35. Zhao Y, Liu L, Tu Y, Wu HC. Investigating the effect of mono-and multivalent counterions on the conformation of poly (styrenesulfonic acid) by nanopores. *Electrophoresis*. 2019;40(16–17):2180–2185.
36. Walsh G, Jefferis R. Post-translational modifications in the context of therapeutic proteins. *Nat Biotechnol*. 2006;24(10):1241–1252.
37. Fahie MA, Chen M. Electrostatic interactions between OmpG nanopore and analyte protein surface can distinguish between glycosylated isoforms. *J Phys Chem B*. 2015;119(32):10198–10206.
38. Restrepo-Pérez L, Wong CH, Maglia G, Dekker C, Joo C. Label-free detection of post-translational modifications with a nanopore. *Nano Lett*. 2019;19(11):7957–7964.
39. Versloot RCA, Lucas FLR, Yakovlieva L, Tadema MJ, Zhang Y, Wood TM, Martin NI, Marrink SJ, Walvoort MTC, Maglia G. Quantification of protein glycosylation using nanopores. *Nano Lett*. 2022;22(13):5357–5364.
40. Niu H, Li M-Y, Ying Y-L, Long Y-T. An engineered third electrostatic constriction of aerolysin to manipulate heterogeneously charged peptide transport. *Chem Sci*. 2022;13(8):2456–2461.
41. Greive SJ, Bacri L, Cressiot B, Pelta J. Identification of conformational variants for bradykinin biomarker peptides from a biofluid using a nanopore and machine learning. *ACS Nano*. 2024;18(1):539–550.
42. Ratinho L, Bacri L, Thiebot B, Cressiot B, Pelta J. Identification and detection of a peptide biomarker and its enantiomer by nanopore. *ACS Cent Sci*. 2024;10(6):1167–1178.
43. Iesu L, Sai M, Torbeev V, Kieffer B, Pelta J, Cressiot B. Single-molecule nanopore sensing of proline cis/trans amide isomers. *Chem Sci*. 2025;16(22):9730–9738.
44. Wang Y-Q, Li M-Y, Qiu H, Cao C, Wang M-B, Wu X-Y, Huang J, Ying Y-L, Long Y-T. Identification of essential sensitive regions of the aerolysin nanopore for single oligonucleotide analysis. *Anal Chem*. 2018;90(13):7790–7794.
45. Asandei A, Schiopu I, Chinappi M, Seo CH, Park Y, Luchian T. Electroosmotic trap against the electrophoretic force near a protein nanopore reveals peptide dynamics during capture and translocation. *ACS Appl Mater Interfaces*. 2016;8(20):13166–13179.
46. Boukhet M, Piguet F, Ouldali H, Pastoriza-Gallego M, Pelta J, Oukhaled A. Probing driving forces in aerolysin and α-hemolysin biological nanopores: Electrophoresis versus electroosmosis. *Nanoscale*. 2016;8(43):18352–18359.
47. O'Brien EP, Brooks BR, Thirumalai D. Effects of pH on proteins: Predictions for ensemble and single-molecule pulling experiments. *J Am Chem Soc*. 2012;134(2):979–987.
48. Li S, Wu XY, Li MY, Liu SC, Ying YL, Long YT. T232K/K238Q aerolysin nanopore for mapping adjacent phosphorylation

sites of a single tau peptide. *Small Methods*. 2020;4(11): Article 2000014.
49. Ouldali H, Sarthak K, Ensslen T, Piguet F, Manivet P, Pelta J, Behrends JC, Aksimentiev A, Oukhaled A. Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nat Biotechnol*. 2020;38(2):176–181.
50. Palian MM, Boguslavsky V, O'Brie DF, Polt R. Glycopeptide-membrane interactions: Glycosyl enkephalin analogues adopt turn conformations by NMR and CD in amphipathic media. *J Am Chem Soc*. 2003;125(19):5823–5831.
51. Ychou M, Duffour J, Kramar A, Gourgou S, Grenier J. Clinical significance and prognostic value of CA72-4 compared with CEA and CA19-9 in patients with gastric cancer. *Dis Markers*. 2000;16(3–4):105–110.
52. Rømer TB, Aasted MKM, Dabelsteen S, Groen A, Schnabel J, Tan E, Pedersen JW, Haue AD, Wandall HH. Mapping of truncated O-glycans in cancers of epithelial and non-epithelial origin. *Br J Cancer*. 2021;125(9):1239–1250.
53. Leao B, Wen X, Duarte HO, Gullo I, Goncalves G, Pontes P, Castelli C, Diniz F, Mereiter S, Gomes J, et al. Expression of thomsen-friedenreich antigen in colorectal cancer and association with microsatellite instability. *Int J Mol Sci*. 2021;22(3):Article 1340.
54. Zhang L-L, Zhong C-B, Huang T-J, Zhang L-M, Yan F, Ying Y-L. High-throughput single biomarker identification using droplet nanopore. *Chem Sci*. 2024;15(22):8355–8362.
55. Al-Alem L, Prendergast JM, Clark J, Zarrella B, Zarrella DT, Hill SJ, Growdon WB, Pooladanda V, Spriggs DR, Cramer D, et al. Sialyl-Tn serves as a potential therapeutic target for ovarian cancer. *J Ovarian Res*. 2024;17:Article 71.
56. Shimizu K, Mijiddorj B, Usami M, Mizoguchi I, Yoshida S, Akayama S, Hamada Y, Ohyama A, Usui K, Kawamura I, et al. De novo design of a nanopore for single-molecule detection that incorporates a β-hairpin peptide. *Nat Nanotechnol*. 2022;17(1):67–75.
57. Ma W, Deng Y, Xu Z, Liu X, Chapla DG, Moremen KW, Wen L, Li T. Integrated chemoenzymatic approach to streamline the assembly of complex glycopeptides in the liquid phase. *J Am Chem Soc*. 2022;144(20):9057–9065.
58. Hu Z-L, Li Z-Y, Ying Y-L, Zhang J, Cao C, Long Y-T, Tian H. Real-time and accurate identification of single oligonucleotide photoisomers via an aerolysin nanopore. *Anal Chem*. 2018;90(7):4268–4272.
59. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: A corrected feature importance measure. *Bioinformatics*. 2010;26(10):1340–1347.