

An engineered nanopore identifies saccharides, amino acids, peptides and ribonucleotides

Received: 15 October 2025

Accepted: 17 August 2026

Published online: 14 September 2026

 Check for updates

Lang Yao (姚琅), Zixuan Wang (王子璇), Jialu Chen (陈佳璐) , Wen Sun (孙雯), Kefan Wang (王可凡), Yunqi Xiao (肖云麒) , Hanhan Zhang (张含含), Wenzheng Li (李文正), Yifan Wang (王逸凡), Lulu Zhao (赵露露), Xinyi Dai (代馨怡), Lu Qian (钱璐), Panke Zhang (张盼科)  & Shuo Huang (黄硕)  

Simultaneous characterization of proteins, RNAs and glycans with a nanopore remains a substantial challenge. To advance toward this goal, we introduce a multifunctional adaptor, maleimido-C2-formylphenylboronic acid (maleimido-C2-FPBA), into the *Mycobacterium smegmatis* porin A (MspA) nanopore at the pore constriction site. Here we demonstrate that the engineered nanopore, MspA-FPBA, enables the simultaneous identification of a diverse set of analytes, including 21 proteinogenic amino acids, three post-translationally modified (PTM) amino acids, four canonical nucleoside monophosphates (NMPs), three epigenetically modified NMPs, four monosaccharides and five peptides. Combined with machine learning, this sensor achieves an overall accuracy of 98.7%. Furthermore, it generates predictable event signatures for specific analyte types, enabling the machine learning model to identify analytes without strictly relying on prearchived event features, as demonstrated by the nanopore analysis of yeast cell extract. Lastly, we apply MspA-FPBA for compositional analysis of G2 glycopeptides. Future integration of hydrolase with MspA-FPBA may further improve MspA-FPBA utility as a single system to identify diverse biomolecules.

The distinct physicochemical properties of proteins, RNAs and glycans have led to the development of distinct analytical methods to characterize them. For proteins, conventional analytical methods include mass spectrometry (MS)^{1–4} and Edman degradation⁵. RNA sequence analysis encompasses Sanger sequencing⁶, next-generation sequencing⁷ and third-generation sequencing^{8–13}. For glycans, analytical techniques involve MS¹⁴, nuclear magnetic resonance (NMR)¹⁵ and high-performance liquid chromatography (HPLC)¹⁶. A sensor that enables simultaneous analysis of proteins, RNAs and glycans remains technically challenging. Nanopores, a high-performance single-molecule sensor, enable the sequencing of DNA and RNA^{12,17}

but have yet to achieve sequencing of proteins^{18–20} and glycans^{21,22} because of their more complex sequences and structures. To further realize nanopore protein sequencing, two strategies, strand sequencing^{23–25} and exo-sequencing²⁶, have been proposed. The challenges for strand sequencing stem from the limited spatial resolution, making it unable to distinguish monomeric units free from interference by neighboring units. Specifically for glycans, the presence of branched structures^{27,28} renders strand sequencing nearly impossible. On the other side, exo-sequencing²⁶, which integrates a hydrolase above the nanopore, can sequentially chop and drop terminal monomers to the pore constriction for unambiguous monomer identification. Although

not yet realized, the exo-sequencing strategy is in principle suitable for nucleic acids, proteins and glycans, suggesting its potential for nanopore multiomics analysis. However, establishing a universal nanopore sequencing system for nucleic acids, proteins and glycans would require installation within the pore of an adaptor capable of simultaneously responding to amino acids, nucleoside monophosphates (NMPs) and monosaccharides.

Mycobacterium smegmatis porin A (MspA), a naturally occurring porin protein²⁹, is renowned for its application in nanopore DNA sequencing¹⁷. Its high resolution also enables direct observation of refined chemical reactions within the pore lumen, appearing as a nanopore reactor³⁰. To guarantee the precision of the sole reactive adaptor placement at the pore constriction, the hetero-octameric MspA was developed³¹. It demonstrates superior sensing performance for small molecules and ions, including saccharides^{31,32}, NMPs³³, amino acids³⁴ and rare earth elements³⁵, suggesting its suitability as the high-resolution pore template. Although nanopore detection of NMPs, amino acids and monosaccharides has been separately reported by different adaptors, these adaptors cannot substitute for each other's functions because of their specialized design (Extended Data Fig. 1). To the best of our knowledge, a single adaptor generally suitable for all above-described analytes has not yet been reported.

Results

Synthesis and sensing mechanism of maleimido-C2-FPBA

To endow MspA with the ability to reversibly interact with amino acids, NMPs, saccharides and peptides, a maleimido-C2-formylphenylboronic acid (maleimido-C2-FPBA) was designed and custom-synthesized. It was synthesized through a three-step reaction (Fig. 1a and Methods). Specifically, Boc-protected ethylenediamine and compound **1** undergo nucleophilic addition and intramolecular cyclization to form compound **2**. Then, under the catalysis of trifluoroacetic acid (TFA), compound **2** undergoes nucleophilic substitution to remove the Boc protecting group to generate compound **3**. Finally, under the catalysis of HATU, the amino group of compound **3** and the carboxyl group of compound **4** undergo amide condensation to form the final maleimido-C2-FPBA. The results of NMR spectroscopy, LC-MS and high-performance LC (HPLC) analyses confirmed successful synthesis of the maleimido-C2-FPBA (Supplementary Figs. 1–4).

For simplicity, the hetero-octameric MspA modified with the maleimido-C2-FPBA at its pore constriction is denoted as MspA-FPBA. It reversibly interacts with amino acids and their post-translational modifications (PTMs), saccharides, peptides, NMPs and their epigenetic modifications through reversible covalent interactions (Fig. 1b). Specifically, regarding amino acids and their PTMs, their α -amino groups can react with the FPBA adaptor to yield iminoboronate compounds^{36,37}. Likewise, the N-terminal amino groups of peptides can react with the FPBA adaptor to form iminoboronate compounds³⁸. In the case of saccharides, NMPs and their epigenetic modifications, their *cis*-diol groups can react with the FPBA adaptor to generate boronic esters^{39,40}. Given that the formation of iminoboronates and boronic esters is dynamically reversible in aqueous solution, MspA-FPBA is suitable for simultaneous identification of multiple analytes by the same sensor.

Construction of MspA-FPBA

To site-specifically modify a single FPBA adaptor at the MspA pore constriction, a hetero-octameric MspA, which had a single substitution at Cys90 (Supplementary Table 1), abbreviated as (N90C)₁(M2)₇, was first prepared (Methods)^{33,34}. The maleimide group of the synthesized maleimido-C2-FPBA adaptor reacts with the side-chain thiol group of cysteine of (N90C)₁(M2)₇, leading to site-specific modification of a single FPBA adaptor at the pore constriction to generate MspA-FPBA (Fig. 2b). MspA-FPBA can reversibly bind amino acids, NMPs and saccharides to report corresponding nanopore events (Fig. 2a). All above measurements were carried out in a custom-built measurement

chamber containing *cis* and *trans* compartments. Here, the *cis* represents the electrically grounded compartment and the *trans* represents the opposing compartment. Specifically, the *cis* compartment was filled with 1.5 M KCl and 10 mM *N*-cyclohexyl-2-aminoethanesulfonic acid (CHES) buffer at pH 9.0. The *trans* compartment was filled with 1.5 M KCl and 10 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) buffer at pH 6.0 (Methods).

Stepwise construction of MspA-FPBA could also be observed during single-channel recording (Fig. 2b). With (N90C)₁(M2)₇ inserted and a +160-mV bias continually applied, a stable open-pore current (I_o) of 490 pA was observed. Because of the unmodified Cys90 of (N90C)₁(M2)₇, some transient spike noises were also observed. At this stage, maleimido-C2-FPBA was further added to the *cis* side to a final concentration of 500 μ M. Immediately afterward, an abrupt current drop measuring approximately 160 pA was monitored and the previously observed spike noises also disappeared immediately. No further changes in the open-pore current were observed thereafter. These phenomena indicate the successful modification of maleimido-C2-FPBA to (N90C)₁(M2)₇ and the generation of MspA-FPBA. At this stage, the open-pore current of an MspA-FPBA (I_{FPBA}) measured approximately 332 pA (Fig. 2b, Supplementary Fig. 5 and Supplementary Table 2). MspA-FPBA could also be batch-prepared by preincubating (N90C)₁(M2)₇ and maleimido-C2-FPBA (Methods). The open-pore current of the batch-prepared MspA-FPBA was identical to that previously characterized by single-channel recording (Fig. 2b). Thus, unless otherwise specified, all subsequent measurements were conducted using batch-prepared MspA-FPBA.

Simultaneous sensing of amino acids, NMPs and saccharides

To validate the sensing capacity of MspA-FPBA for amino acids, NMPs and saccharides, we selected Phe, guanosine 5'-monophosphate (GMP) and D-fructose (Fru) as model analytes (Fig. 2c). Experimentally, the *cis* chamber was filled with a 1.5 M KCl and 10 mM CHES buffer at pH 9.0 and the *trans* chamber was filled with a 1.5 M KCl and 10 mM MES buffer at pH 6.0. With an MspA-FPBA inserted and the addition of Phe to the *cis* side at 400 μ M, corresponding nanopore events were immediately detected (Fig. 2c). Notably, the appearance rate of Phe events increases when there was a pH difference between the *cis* and *trans* side buffers (Supplementary Fig. 6), suggesting an improved rate of event appearance at an asymmetric buffer pH configuration. Thus, unless otherwise specified, all subsequent measurements used this measurement condition. In separate measurements, after the addition of GMP and Fru at 2 mM and 20 mM to the *cis* side, respectively (Fig. 2c), distinct types of events were then immediately observed, confirming the feasibility of simultaneous sensing of amino acids, NMPs and saccharides.

For quantitative event description, various core event parameters, including open-pore current (I_{FPBA}), event current (I_{event}), blockage amplitude ($\Delta I = I_{FPBA} - I_{event}$), noise amplitude (SD), dwell time (t_{off}), interevent duration (t_{on}), skewness (skew), kurtosis (kurt), local SD ratio, minimum, maximum and range (minimum–maximum) (Supplementary Fig. 7), were used to quantitatively describe all nanopore events. Notably, all traces presented in this study were not subjected to any further digital filtering. Short-residing events (<2 ms) were consistently excluded before the statistical analysis. Scatter plots of ΔI versus SD of Phe, GMP and Fru events were generated, with distinct event clusters for each analyte, indicating the unambiguous event discrimination (Fig. 2d). Specifically, Phe produced one event cluster, while GMP and Fru each produced two event clusters, labeled as GMP1/2 and Fru1/2, respectively. Concurrent addition of Phe, GMP and Fru to *cis* at 50 μ M, 1 mM and 5 mM, respectively, resulted in distinct nanopore events (Fig. 2e and Supplementary Video 1), consistent with results when they were measured separately. This further confirms the discrimination capacity of MspA-FPBA. However, this sensing capacity was not demonstrated in assays performed with

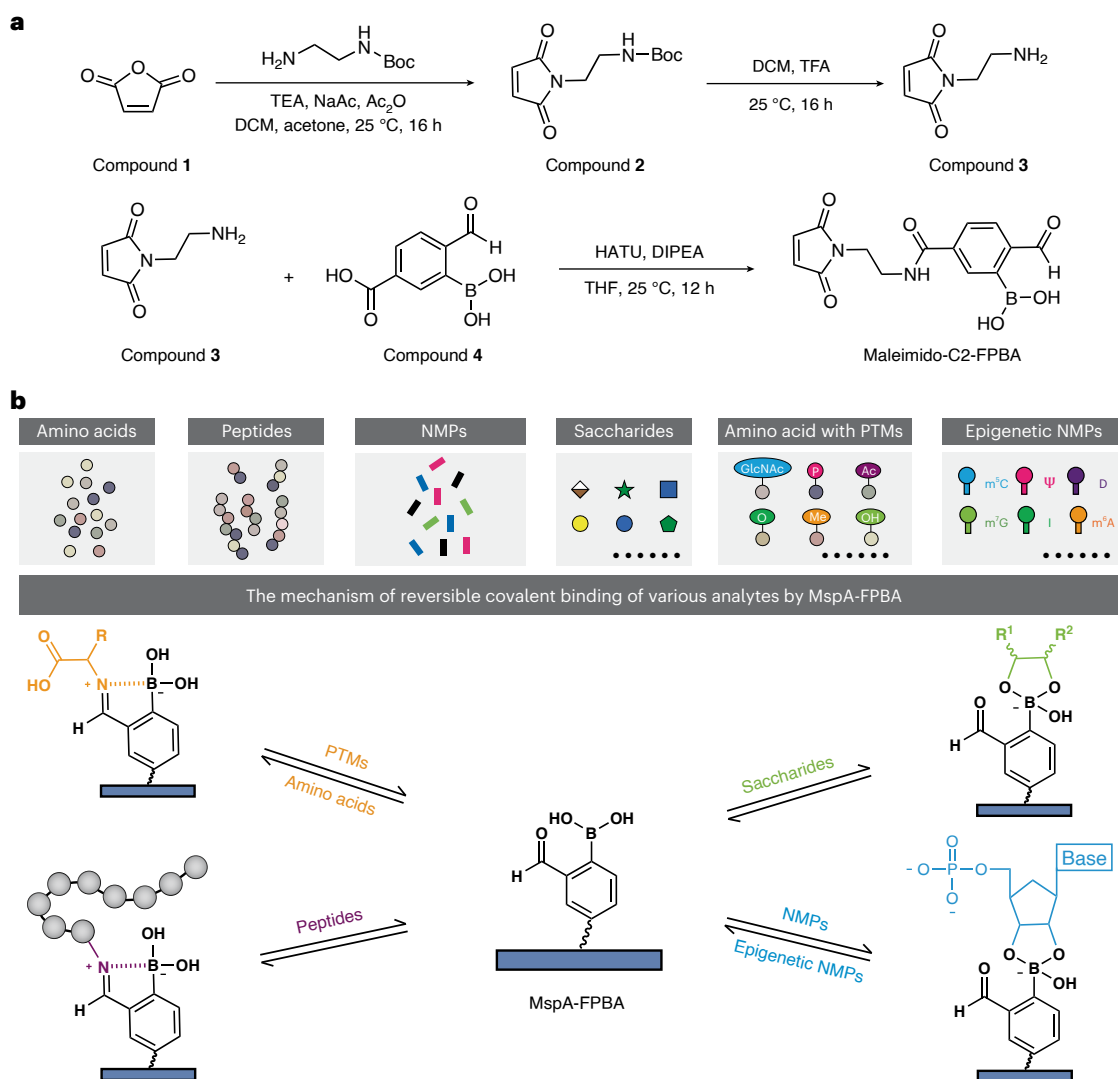


Fig. 1 | Synthesis of the maleimido-C2-FPBA adaptor and mechanism of sensing. **a**, The synthetic route of the maleimido-C2-FPBA adaptor. The compound was synthesized through a three-step reaction. Briefly, compound 1 underwent nucleophilic addition and intramolecular cyclization to form compound 2 (Supplementary Fig. 1). Then, the Boc protecting group of compound 2 was removed through nucleophilic substitution under acid catalysis, releasing the amino group to form compound 3. Finally, the amino group of compound 3 and the carboxyl group of compound 4 were condensed under the catalysis of HATU to form maleimido-C2-FPBA (Supplementary Figs. 2–4).

b, The mechanism by a single FPBA adaptor-modified MspA nanopore (MspA-FPBA). MspA-FPBA reversibly interacts with a sole analyte at a time. Left: for amino acids and their PTMs, their α -amino groups react with the FPBA adaptor to form iminoboronates. Similarly, N-terminal amino groups of peptides can react with the FPBA adaptor to form iminoboronates. Right: for saccharides, NMPs and their epigenetic modifications, their *cis*-diol structures can react with the FPBA adaptor to form boronic esters. The adduct of each analyte with the FPBA adaptor is reversible and dynamic, enabling MspA-FPBA for simultaneous identification of multiple types of analytes in the same assay.

(N90C)₁(M2)₇, MspA-PBA and MspA-NTA-Ni (Supplementary Fig. 8 and Extended Data Fig. 1), again underscoring the crucial role of the FPBA adaptor for simultaneous sensing of amino acids, saccharides and NMPs in the same assay.

Discrimination of 21 proteinogenic amino acids, four canonical NMPs and four saccharides

To illustrate the discriminative prowess of MspA-FPBA in identifying diverse amino acids, NMPs and saccharides, we performed single-channel recordings on a comprehensive panel of 21 proteinogenic amino acids, including selenocysteine, as well as four canonical NMPs (adenosine 5'-monophosphate (AMP), uridine 5'-monophosphate (UMP), cytidine 5'-monophosphate (CMP) and GMP) and four randomly chosen saccharides (iduronic acid (IdoA), L-arabinose (Ara), Fru and *N*-acetyl-D-glucosamine (GlcNAc)). Their representative events are depicted in Extended Data Fig. 2a, revealing that amino

acid events typically exhibit a single, steady level with minimal fluctuations, while certain amino acids display transient spikes. Upon using different amino acids in separate measurements, each type yielded a distinct, identifiable event profile (Supplementary Figs. 9–12). The majority of amino acids produced a single type of event, with cysteine, histidine and lysine exceptions, which generated two types each, denoted as C1/2, H1/2 and K1/2, respectively. Notably, because of the formation of a stable five-membered ring structure⁴¹ involving cysteine's α -amino group and side-chain thiol group with the FPBA adaptor, cysteine events often exhibit extremely prolonged durations (Supplementary Fig. 13). Thus, unless otherwise stated, cysteine was measured after reaction of maleimide to avoid the appearance of the prolonged events (Supplementary Fig. 14).

Each type of NMP generates two representative event types (Supplementary Fig. 15), denoted as NMP1/2. Briefly, the first type of NMP events displays short t_{off} and minimal blockage level fluctuations,

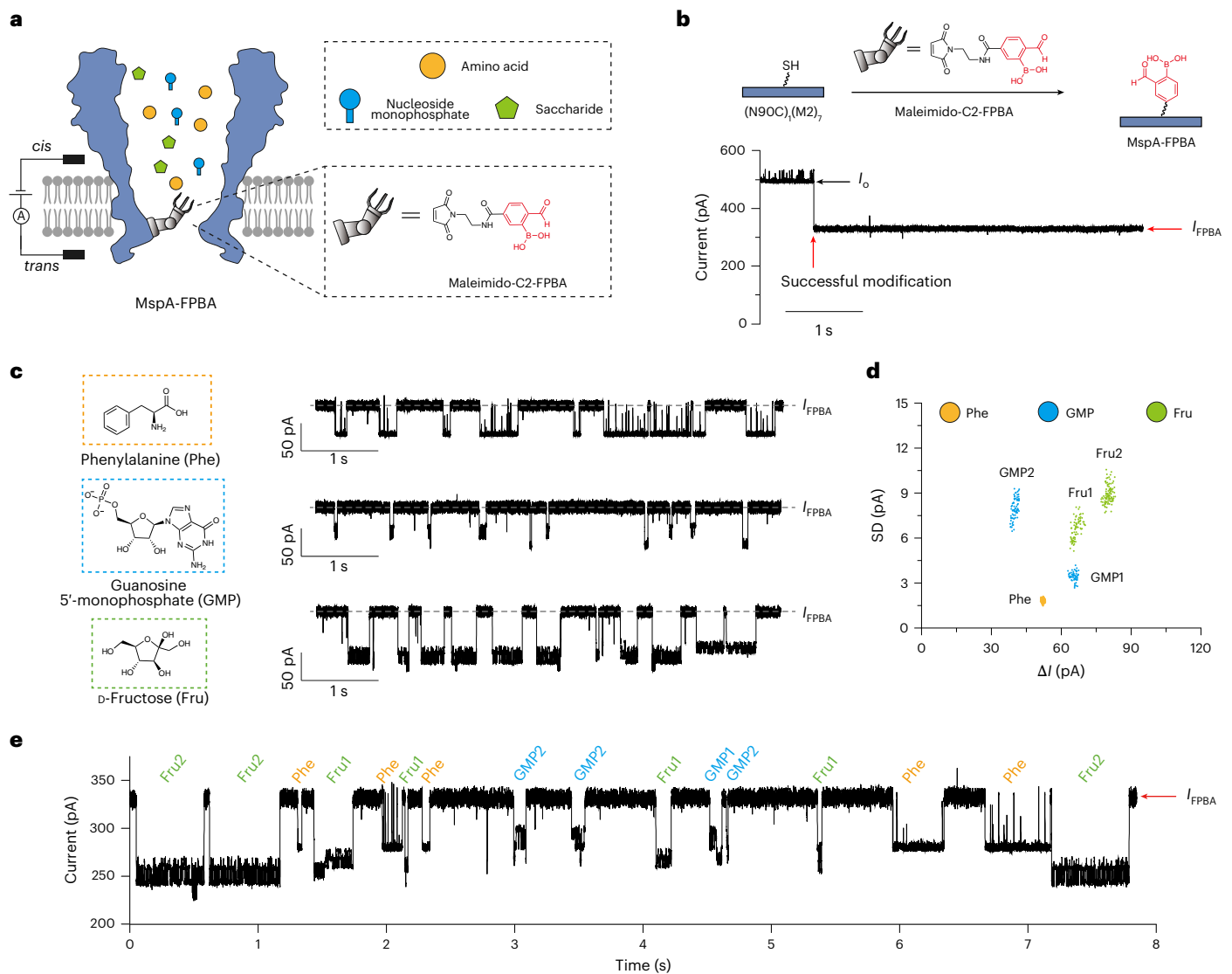


Fig. 2 | Construction of MspA-FPBA for simultaneously identifying amino acids, NMPs and saccharides. a, The site-specific modification of a single FPBA adaptor at the MspA pore constriction generates MspA-FPBA, enabling sensing of amino acids, NMPs and saccharides with the same sensor. **b**, Real-time construction of MspA-FPBA. A single maleimido-C2-FPBA adaptor (Fig. 1) reacts with Cys90 of (N90C)₁(M2)₇ to form MspA-FPBA (Methods). During single-channel recording, the sudden drop of current amplitude from I_o to I_{FPBA} indicates the success of pore modification. Here, I_o and I_{FPBA} represent the open-pore currents of (N90C)₁(M2)₇ and MspA-FPBA, respectively. **c**, Representative traces acquired with Phe, GMP and Fru using MspA-FPBA, respectively. Left: chemical structures of Phe, GMP and Fru. Under a continually applied +160-mV bias, each analyte

was separately added to the *cis* side. The final concentrations of Phe, GMP and Fru were 400 μ M, 2 mM and 20 mM, respectively. For clarity, I_{FPBA} is marked by a gray dashed line. **d**, The scatter plot of ΔI versus SD for Phe, GMP and Fru events obtained as described in **c**. Both GMP and Fru have two types of events, which are labeled as GMP1/2 and Fru1/2. At least 100 events of each analyte were used to generate the scatter plot. **e**, A representative trace acquired during simultaneous sensing of Phe, GMP and Fru using MspA-FPBA. Phe, GMP and Fru were simultaneously added to the *cis* side at final concentrations of 50 μ M, 1 mM and 5 mM, respectively. Different nanopore events were identified and labeled on the basis of their event features as described in **c**.

while the second type involves intense telegraphic switching. On the other side, saccharide events exhibit intense telegraphic switching across multiple levels. Specifically, IdoA generates a single type of event. However, Ara produces three types of events, designated as Ara1/2/3, while Fru and GlcNAc each yield two types of events, labeled as Fru1/2 and GlcNAc1/2, respectively. Individual saccharide measurements revealed characteristic event shapes for each type (Supplementary Fig. 16). The unique event characteristics of each analyte underscore the critical role of the FPBA adaptor in discerning among different amino acids, NMPs and saccharides with the same sensor and measurement condition. To visualize nanopore sensing results, scatter plots based on different parameter combinations were constructed using amino acids Ser, Ala, Gln, Thr, Val, Met, Ile,

Leu and Trp as model systems (Supplementary Fig. 17). Among them, the ΔI versus SD plot clearly resolved nine distinct event clusters corresponding to different analytes. Therefore, ΔI and SD were used for two-dimensional (2D) data visualization throughout this study.

Results of three independent measurements for each amino acid demonstrated notable reproducibility (Supplementary Table 3). A scatter plot of ΔI versus SD was plotted for results acquired with 21 proteinogenic amino acids (Extended Data Fig. 2b). The scatter plot also revealed that the 21 proteinogenic amino acids were effectively differentiated (Extended Data Fig. 2b). To elucidate the event distribution, the dense region of the scatter plot was further magnified (Extended Data Fig. 3). While there was slight overlap between Met and Pro events, their actual source events are visually different

(Extended Data Fig. 2a) and their events could be efficiently differentiated by simultaneously considering the maximum parameter (Supplementary Fig. 18). Notably, MspA-FPBA could also unambiguously differentiate Leu from its isomer Ile, a challenge for conventional MS, further emphasizing the amino acid resolution capabilities of MspA-FPBA (Supplementary Fig. 19).

For the NMPs, three independent measurements yielded highly reproducible events (Supplementary Table 4). The type 1 events (NMP1) generated by the four NMPs were quantified and a scatter plot of ΔI versus SD was generated (Extended Data Fig. 2c). The scatter plot indicated that all four canonical NMPs were well separated. Although the type 2 events exhibited notable overlap and are not suitable for distinction (Extended Data Fig. 4), they can be easily recognized and avoided in further quantitative analysis. The events of the four saccharides were quantified and a scatter plot of ΔI versus SD was generated (Extended Data Fig. 2d), demonstrating that the four saccharides were effectively distinguished by MspA-FPBA. On the basis of three independent measurements for each saccharide, the resultant sensing events also produced highly reproducible data (Supplementary Table 5).

Identification of amino acids, NMPs and saccharides by machine learning

To mitigate the potential bias inherent in manual event identification and to establish an automated sensing event identification process, an automatic identification algorithm for amino acids, NMPs and saccharides based on machine learning was developed (Methods). The machine learning workflow encompassed the assembly of a training dataset, the construction of a feature matrix and the training of various models (Supplementary Fig. 20). Initially, MspA-FPBA was used to detect each amino acid, NMP and saccharide individually, yielding a total of 5,400 representative events with known labels. These events were then fed to a custom MATLAB code to extract nine characteristic parameters (ΔI , SD, local SD ratio, minimum, maximum, range, t_{off} , skew and kurt) for each event, which were used to construct a nine-dimensional (9D) feature matrix. Each event in the matrix was labeled with the analyte applied for data acquisition. Using the 'classification learner' toolbox of MATLAB, the performance of seven classification models, namely *k*-nearest neighbor (KNN), support vector machine (SVM), naive Bayes, decision trees, neural network, discriminant analysis and ensemble, was assessed using tenfold cross-validation. The bagged trees model demonstrated the highest validation accuracy, reaching an impressive 98.9% (Supplementary Table 6).

The 9D confusion matrix produced by the bagged trees model revealed that over half of the amino acids, including Ala, Asp, Glu, Phe, Gly, Lys, Leu, Arg, Ser, Sec, Trp, Tyr and Cys, achieved a perfect true-positive rate of 100%. Despite a minor overlap between methionine and proline events in the ΔI versus SD scatter plot (Extended Data Fig. 2b), simultaneous consideration of all nine event features by machine learning enabled the precise identification of these two amino acids, with validation accuracies of 99.4% and 98.8%, respectively (Fig. 3a). To visualize the feature distribution of various amino acid, NMP and saccharide events, a parallel coordinate plot based on the 9D feature matrix was also generated (Supplementary Fig. 21). Furthermore, a learning curve was plotted to assess the efficiency of model training (Supplementary Fig. 21), revealing that the model's accuracy reached 98.1% with the inclusion of at least 1,368 input events. It also confirmed that overfitting was not happening.

To further assess the capability of MspA-FPBA in distinguishing various amino acids, NMPs and saccharides in the same assay, a panel of ten amino acids (Arg, Asn, Tyr, Phe, Trp, Gly, Val, Gln, Met and Ile), two NMPs (UMP and GMP) and one saccharide (Fru) were chosen to demonstrate concurrent sensing in a mixture. Experimentally, with a continually applied +160-mV bias, these analytes were introduced simultaneously to the *cis* side and various events were immediately

observed afterward (Fig. 3b–d). Subsequently, all events were automatically predicted and classified using the trained bagged trees model (Fig. 3b–d and Supplementary Figs. 22–24). A scatter plot comparing ΔI versus SD for events before and after prediction was also generated (Supplementary Fig. 25). These findings suggest that, with the aid of machine learning, MspA-FPBA is capable of simultaneously identifying multiple types of amino acids, NMPs and saccharides from a mixture, all relying on the same sensor.

Identification of PTM amino acids, epigenetic NMPs and peptides

Three representative PTM amino acids (N^6 -acetyl-Lys, Ac-Lys; *O*-phospho-Tyr, P-Tyr; N^3,N^4 -dimethyl-Arg, Me-Arg) and three epigenetic NMPs (pseudo-UMP, ψ ; N^6 -methyl-AMP, m^6A ; inosine-5'-monophosphate, IMP) were selected as model analytes. The modifications in their chemical structures were highlighted with red boxes (Extended Data Fig. 5a). Experimentally, a +160-mV bias was continually applied and Ac-Lys, P-Tyr, Me-Arg, ψ , m^6A and IMP were separately added to the *cis* side. Subsequently, corresponding nanopore events were immediately observed (Extended Data Fig. 5b,d). The events of PTM amino acids were extracted and a scatter plot of ΔI versus SD was generated (Extended Data Fig. 5c), showing three completely separated event clusters. Compared to proteinogenic amino acids, the events of PTM amino acids exhibited distinct event characteristics in the scatter plot of ΔI versus SD (Supplementary Fig. 26). Additionally, the epigenetic NMP events were also extracted and shown in the scatter plot of ΔI versus SD (Extended Data Fig. 5e), appearing as three completely separated event clusters, distinct from their nonmodified counterparts (Supplementary Fig. 27). Although demonstrated with a few representative analytes, this technique is in principle also applicable to other types of modifications of amino acids and NMPs.

On the other hand, under identical conditions, five representative peptides (Ile-Lys, Leu-Arg-Gly, Ser-Leu-Arg, Gly-Tyr-Ile-Lys and Leu-Arg-Gly-Tyr) were also analyzed using MspA-FPBA (Supplementary Fig. 28). Longer peptides may create greater steric hindrance, impeding the ability of the peptide N-terminal amino group to be captured by the FPBA adaptor. Nevertheless, longer peptides may still be detectable through resistive-pulse sensing. All data obtained from PTM amino acids, epigenetic NMPs and peptides were added to the existing machine learning training dataset (Extended Data Fig. 6a). Nine event features were extracted for analysis and the model's performance was evaluated using tenfold cross-validation (Supplementary Table 7). Within the ensemble model, the bagged trees model achieved the highest validation accuracy of 98.7%, a value only 0.2% lower than the 98.9% observed for the original set of 21 amino acids, four NMPs and four saccharides, suggesting the high resolution of the sensor and the great expandability of the method. The corresponding confusion matrix demonstrated that all 40 analytes achieved an accuracy of over 92.5%, with 18 of these analytes attaining 100.0% accuracy (Extended Data Fig. 6b). These results underscore the high resolution of MspA-FPBA and confirm that additional data from PTM amino acids, epigenetic NMPs and peptides do not impair performance of the machine learning algorithm. This indicates that nanopore technology holds scalability for an even broader range of biological analytes in future applications.

For all analytical predictions, a comprehensive set of nine parameters (ΔI , SD, local SD ratio, minimum, maximum, range, t_{off} , skew and kurt) was simultaneously used for event identification. To systematically assess the contribution of distinct features, the permutation feature importance characterization was applied^{42,43}. Specifically, higher importance mean values directly correlate with greater feature importance. As shown in Supplementary Table 8 and Supplementary Fig. 29, ΔI and SD made the greatest contribution to the model's discrimination capacity, followed by local SD ratio and t_{off} . Other parameters, including maximum, minimum, range, kurt and skew, also showed their overall

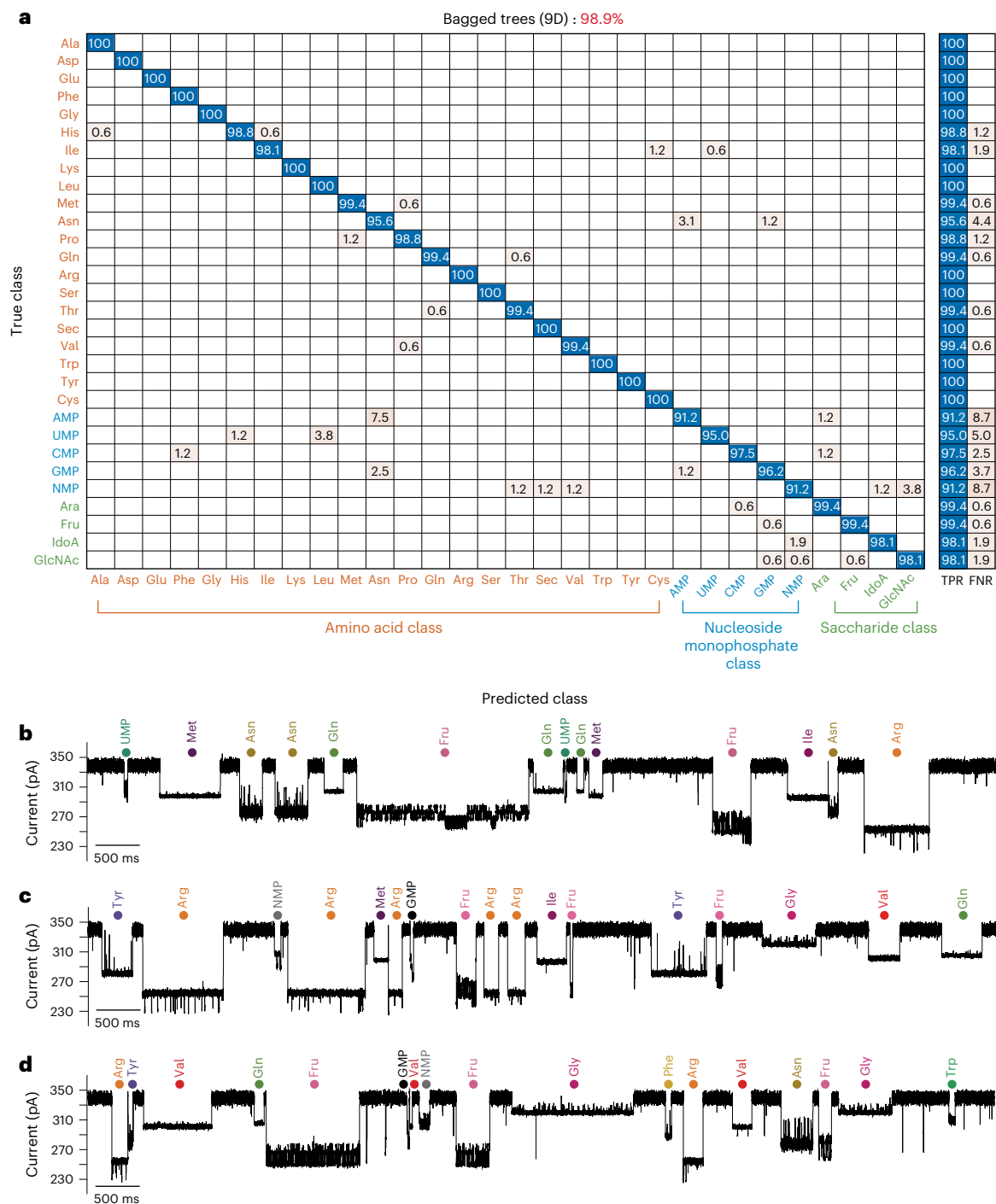


Fig. 3 | Identification of 21 proteinogenic amino acids, four canonical NMPs and four saccharides using machine learning. **a**, The confusion matrix for amino acid (orange), NMP (blue) and saccharide (green) events generated by the bagged trees model using a 9D feature matrix. Each amino acid class contained 200 events, while that of lysine contained 100 events. Each of AMP, UMP, CMP and GMP contained 100 type 1 events and the four NMPs comprised 100 type 2 events. Each saccharide class contained 200 events. Model training was performed with the classification learner toolbox in MATLAB using tenfold cross-validation. Among these models, the bagged trees model reported the highest validation accuracy of 98.9% (Supplementary Table 6). TPR, true-positive rate; FNR, false-negative rate. **b–d**, Representative traces acquired during simultaneous sensing of ten amino acids (Arg, Asn, Tyr, Phe, Trp, Gly, Val, Gln,

Met and Ile), two NMPs (UMP and GMP) and one saccharide (Fru) using MspA-FPBA (Methods). Representative trace segments from the same original trace are shown in **b–d** to clearly illustrate event details. With a +160-mV bias applied, ten amino acids, two NMPs and Fru were added to the *cis* side simultaneously. The final concentrations of each amino acid, each NMP and Fru were 100 μ M, 200 μ M and 2 mM, respectively. Zoomed-in views of these representative current traces are shown (Supplementary Figs. 22–24). The representative events were automatically predicted by the trained model and labeled with color-coded dots (Arg, orange; Asn, brown; Tyr, light purple; Phe, yellow; Trp, green; Gly, dark red; Val, red; Gln, light green; Met, dark purple; Ile, purple; UMP, dark green; GMP, black; NMP, gray; Fru, pink). The NMP events marked with gray dots represent the type 2 events of the four NMPs.

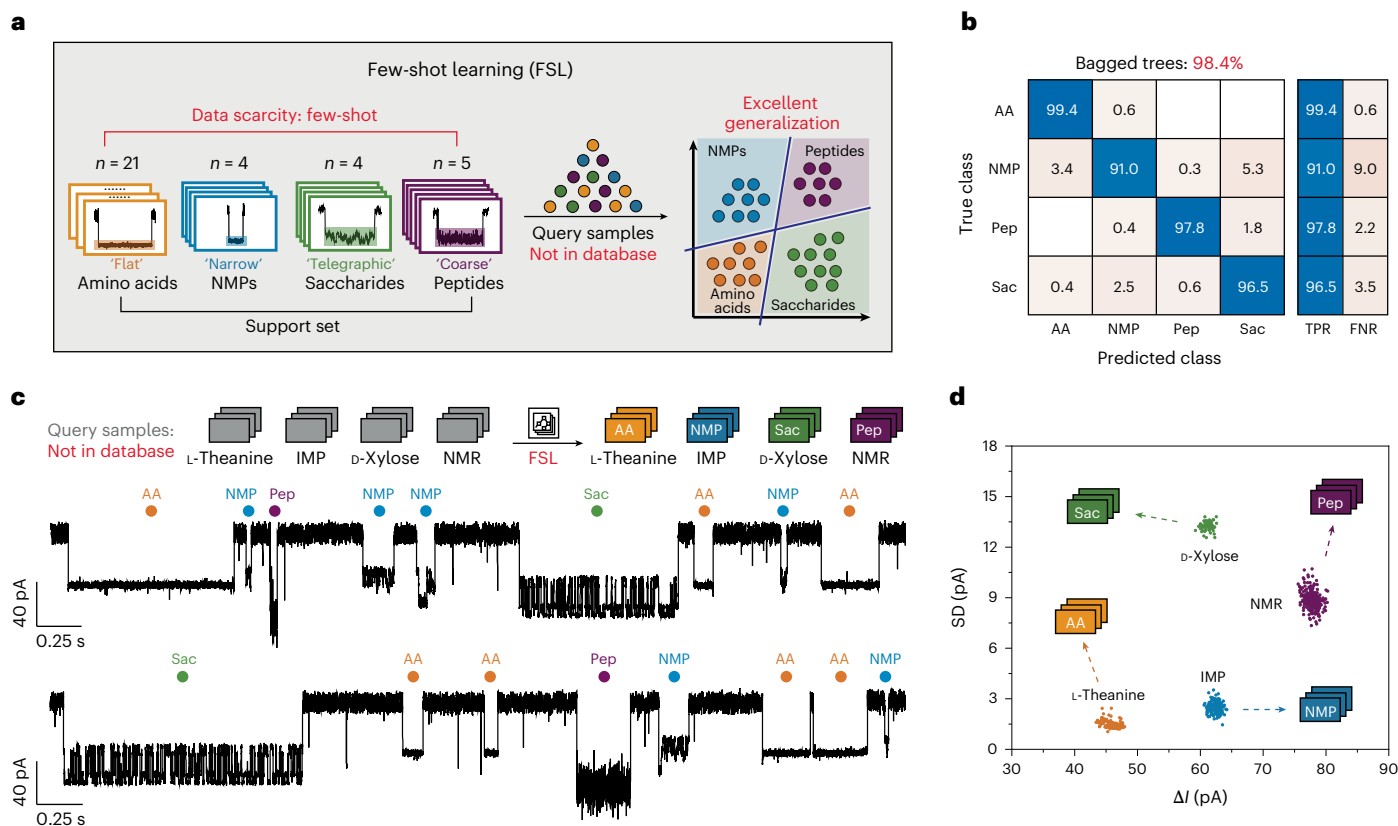


Fig. 4 | Classification of out-of-database analytes by few-shot learning.

a, The workflow of few-shot learning for classification of out-of-database analytes. The previously separately labeled events of 21 proteinogenic amino acids, four canonical NMPs, four saccharides and five peptides were respectively integrated into the classes of amino acids, NMPs, saccharides and peptides, thereby forming the support set for machine learning. The model trained on these few labeled samples, each with its own event features, achieved classification for query samples outside the database. **b**, The confusion matrix for the amino acid, NMP, peptide and saccharide event set generated by the bagged trees model. **c**, Representative traces acquired during simultaneous sensing of query samples outside the database (theanine, IMP, D-xylose and

tripeptide Asn-Met-Arg) using Mspa-FPBA. With a +160-mV bias applied, theanine, IMP, D-xylose and tripeptide Asn-Met-Arg were added to the *cis* side simultaneously. The final concentrations of theanine, IMP, D-xylose and tripeptide Asn-Met-Arg were 100 μ M, 1 mM, 20 mM and 100 μ M, respectively. The events were automatically classified by the constructed few-shot learning and labeled with color-coded dots (amino acids, orange; NMPs, blue; saccharides, light green; peptides, purple). **d**, The scatter plot of ΔI versus SD for the amino acids, NMP, peptides and saccharides events obtained from **c**. The scatter plot was generated using a 30-min continually recorded trace. The four clusters corresponding to amino acids, NMP, peptides and saccharides were identified by few-shot learning.

contribution. All nine event parameters simultaneously contributed to the model's identification capacity.

The four-feature model trained on the minimal feature subset (ΔI , SD, local SD ratio and t_{off}) exhibited performance comparable to the original nine-feature model (Supplementary Table 9). This highlights the potential for a lightweight model configuration. However, this equivalence in performance was observed specifically in the context of differentiating 29 analytes (Fig. 3a). When the analytical scope was expanded to 40 analytes, including amino acids with PTMs, epigenetic NMPs and peptides (Extended Data Fig. 6b), the performance gap became more evident (Supplementary Table 10). To maintain consistency and in anticipation of future broadening of the analytical scope, we retained the nine-feature model throughout the paper. While not demonstrated here, the underlying principle is also theoretically applicable to other molecular classes, such as D-amino acids, nucleotide sugars, oligosaccharides and oligonucleotides.

Quantification of amino acids, NMPs, saccharides and peptides

During the detection of amino acids, NMPs, saccharides and peptides, within a certain analyte concentration range, the count of nanopore events generated per min by a specific species (F) is linearly correlated with the species concentration, forming the basis

for quantitative analysis (Supplementary Tables 11 and 12 and Supplementary Figs. 30–36). To demonstrate that, four amino acids (Met, Asn, Tyr and Arg), UMP, Fru and the tripeptide Leu-Arg-Gly were mixed at known concentrations (Methods). The mixture was subsequently subjected to nanopore analysis (Supplementary Fig. 37). According to the quantitative results, the nanopore results were generally consistent with the true values (Methods and Supplementary Fig. 37). These results not only demonstrate its quantitative capability for mixed samples but also validate the machine learning model's capability for mixture analysis.

To demonstrate the identification performance of the model trained on the single-analyte dataset for multianalyte mixtures, we performed nanopore sensing on a defined multianalyte mixture (Supplementary Fig. 38) and compared the produced events to those from corresponding single-analyte measurements (Supplementary Fig. 39). Quantitative analysis through 2D scatter plots of ΔI versus SD demonstrated that the event signatures in multianalyte mixtures were identical to those collected during single-analyte measurements, with all events falling within the 99% confidence ellipses derived from the training data (Supplementary Fig. 39), confirming the model's consistency across different sample matrices. Notably, our quantitative evaluation of a reference mixture featuring a 100-fold concentration gradient (0.5–50 mM) (Supplementary Fig. 37) revealed no loss in analytical accuracy,

demonstrating the method's robustness to concentration imbalances in multispecies environments.

Although nanopore sensing theoretically offers single-molecule detection limits, this would require infinitely long measurement durations. To establish practical benchmarks, we defined the limit of detection (LOD) as the minimum analyte concentration capable of generating at least five binding events within a 10-min continuous acquisition period. On the basis of this criterion, the LODs for amino acids, NMPs, saccharides and peptides are summarized in Supplementary Tables 13–16. Notably, the LOD for each concentration was confirmed only when the event count criterion was consistently met across all three independent replicate measurements. Future sensitivity improvements could be achieved through the integration of ultra-flat, miniaturized fluidic chambers with high-density sensor arrays, combined with extended measurement durations similar to a MinION sequencing device.

Nanopore identification of amino acids, NMPs, saccharides or peptides has been previously performed. For amino acid, phenanthroline ring-modified⁴⁴ and cyclodextrin-embedded⁴⁵ α -hemolysin (α -HL) identified amino acid enantiomers in the presence of Cu²⁺. Single Ni²⁺-modified³⁴ and Cu²⁺-modified⁴⁶ MspA distinguished all 20 proteinogenic amino acids. For nucleotide, cyclodextrin-embedded α -HL detected four nucleoside diphosphates⁴⁷. A single phenylboronic acid-modified MspA distinguished 11 NMPs³³. For saccharide, single phenylboronic acid-modified α -HL distinguished three saccharides⁴⁸, while similarly modified MspA differentiated major monosaccharides and disaccharides with different glycosidic linkages^{31,32}. Mutant α -HL recognized saccharides bearing acetamido and carboxyl groups²². For peptide, Aerolysin⁴⁹ and cucurbituril-assisted α -HL⁵⁰ distinguished peptides differing by a single amino acid. Wild-type aerolysin⁵¹, the *Nocardia farcinica* channel⁵² and the bacteriophage T7 connector⁵³ differentiated arginine peptides of different lengths. Translocon EXP2 distinguished polylysine peptides with different molecular weights⁵⁴. Fragaceatoxin C differentiated various angiotensin peptides⁵⁵. Aerolysin distinguished vasopressin enantiomers⁵⁶. α -HL identified multiple peptide PTMs⁵⁷. However, all the above studies were generally limited to a specific analyte class for each pore type. In contrast, MspA-FPBA enables the simultaneous identification of amino acids, NMPs, saccharides and peptides, allowing their classification in the same platform. Assisted by machine learning, it can also perform analyte type classification without strictly relying on events generated by identical standard compounds.

Identification of amino acids and glycosylation by enzymatic digestion

Given that MspA-FPBA can simultaneously identify amino acids, peptides and saccharides, it is anticipated that the established sensing strategy could be used to analyze the amino acid composition and glycosylation modifications on peptides or proteins. To validate this, the sensing strategy was extended to analyze glycosylated peptides following enzymatic digestion (Extended Data Fig. 7a). Tentatively, a glycosylated heptapeptide with the sequence Gly-Met-Gln-Arg-Asn-(GlcNAc)-Ile-Ser was synthesized. Experimentally, this glycosylated heptapeptide was incubated with leucine aminopeptidase (LAP) at 37 °C for 12 h to ensure complete enzymatic hydrolysis. Following hydrolysis, the reaction mixture was ultrafiltered to eliminate LAP. Under a continually applied +160-mV bias, 25 μ l of the ultrafiltrate was introduced to the *cis* side of the MspA-FPBA for single-channel recording (Methods and Supplementary Fig. 40), immediately after which corresponding events were observed (Extended Data Fig. 7b and Supplementary Video 2). A scatter plot of ΔI versus SD was constructed using events from a 10-min continuous recording trace (Extended Data Fig. 7c). A density-based spatial clustering of applications with noise (DBSCAN) algorithm was applied to eliminate nonclustered background noise (Supplementary Fig. 41) and seven distinct event clusters were

identified. With the previously trained bagged trees model, it was confirmed that four clusters were generated by Gly, Met, Gln and Arg (Extended Data Fig. 7c). This aligns perfectly with the amino acid composition preceding the glycosylation site of the Gly-Met-Gln-Arg-Asn-(GlcNAc)-Ile-Ser peptide. Furthermore, three types of previously unidentified events were detected and labeled GP1, GP2 and GP3. The corresponding MS results confirmed the generation of Gly, Met, Gln, Arg and Asn-(GlcNAc)-Ile-Ser (Supplementary Fig. 42). Nanopore measurements on the enzymatic digestion filtrate of the nonglycosylated peptide were also performed and the detected events were identified as Gly, Met, Gln, Arg, Asn, Ile and Ser (Supplementary Figs. 43 and 44). The three previously unidentified events were not detected. Together with the MS results and the results acquired with the nonglycosylated peptide, it was confirmed that the three unidentified events indeed originated from the Asn-(GlcNAc)-Ile-Ser fragment generated during the glycopeptide enzymatic digestion.

In a separate measurement, the untreated Gly-Met-Gln-Arg-Asn-(GlcNAc)-Ile-Ser peptide was added to the *cis* side of MspA-FPBA for single-channel measurement, resulting only in a single type of event (Supplementary Fig. 45). This event was attributed to the binding of the peptide's N-terminal amino group to the FPBA adaptor. The absence of enzymatic treatment and the resulting larger steric hindrance prevented the glycosyl group from binding to the FPBA adaptor, thus prohibited the generation of additional event types. These findings successfully demonstrate that MspA-FPBA can efficiently analyze the amino acid composition and glycosylation modification of glycosylated peptides following enzymatic digestion.

Identification of out-of-database analytes using few-shot learning

The MspA-FPBA sensor exhibits differentiated event characteristics for amino acids, NMPs, saccharides and peptides. Therefore, it is necessary to classify the events to enable comprehensive analysis of complex biological samples using MspA-FPBA. Meanwhile, the diversity of amino acids, NMPs, saccharides and peptides in biological samples poses a data scarcity challenge to the currently constructed database. To address this issue, we applied few-shot learning⁵⁸ to classify analytes that were not included in the database. Here, because of data scarcity, we use a few labeled amino acids, NMPs, saccharides and peptides as the support set for machine learning. As amino acid, NMP, saccharide and peptide events respectively exhibit flat, narrow, telegraphic and coarse characteristics, the trained bagged trees model (Supplementary Table 17) could rapidly classify query samples (analytes not included in the database) into amino acids, NMPs, peptides or saccharides (Fig. 4a,b).

To evaluate the classification performance of the few-shot learning, four types of analytes not included in the dataset, specifically theanine, IMP, D-xylose and tripeptide Asn-Met-Arg, were applied for nanopore sensing. Afterward, event classification was conducted using few-shot learning. Experimentally, with a constantly applied +160-mV bias, these analytes were introduced simultaneously to the *cis* side of the device. Distinct nanopore events were observed shortly thereafter. Subsequently, all events were automatically classified into the categories of amino acids, NMP, peptides and saccharides (Fig. 4c) and their corresponding data were presented in a scatter plot of ΔI versus SD (Fig. 4d). The classification results were consistent with their true values (Supplementary Fig. 46), confirming the technique's ability to recognize out-of-database analytes and its generality in sensing.

Rapid identification of amino acids, NMPs, saccharides and peptides in yeast cell extract

Yeast cell extract, generated by autolysis or enzymatic hydrolysis of yeast cells, is rich in various types of nutrients, including amino acids, peptides, NMPs, saccharides and minerals⁵⁹. Therefore, it is an ideal sample for testifying the performance of MspA-FPBA in the

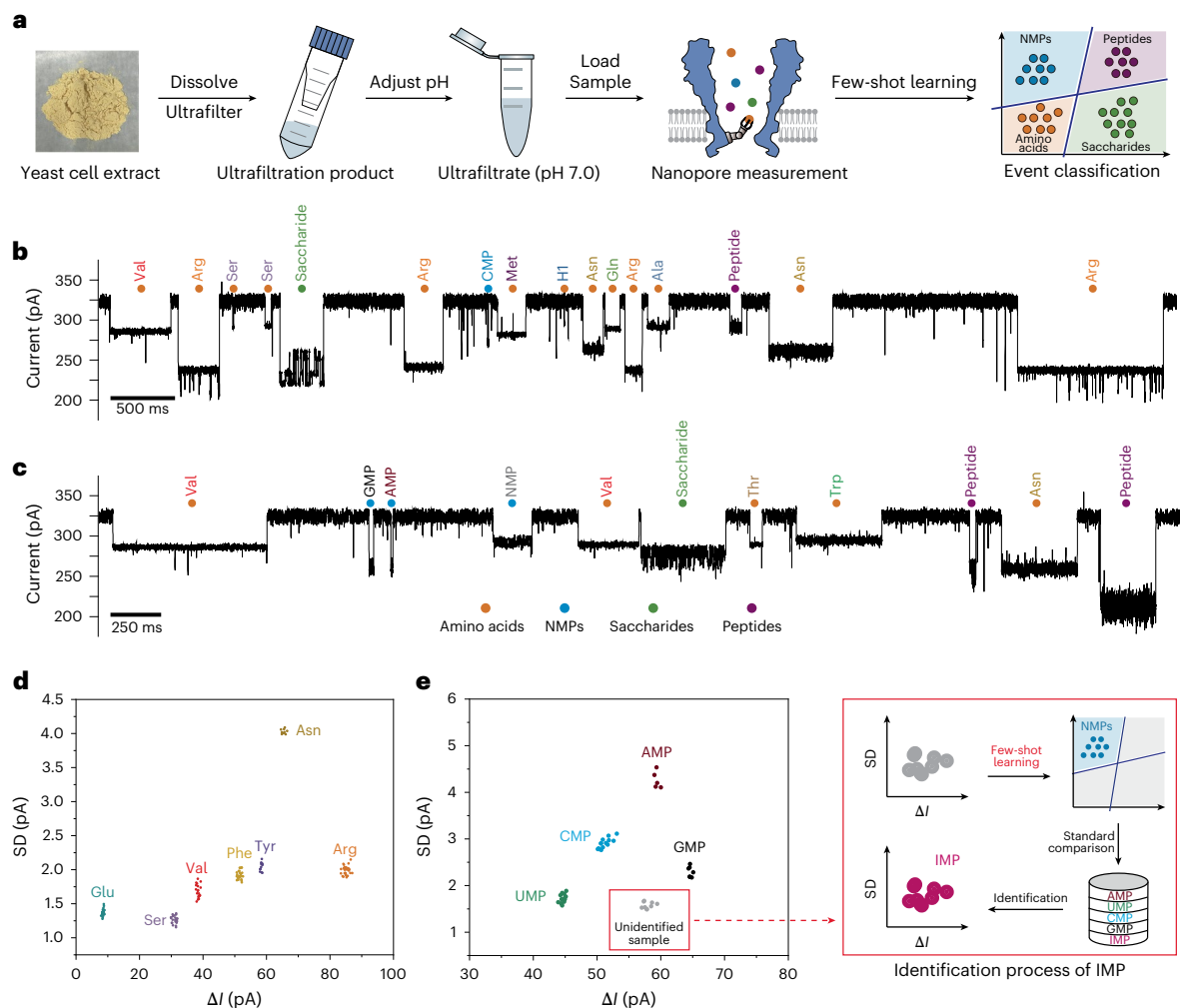


Fig. 5 | Identification of amino acid, NMP, saccharide and peptide events in yeast cell extract. **a**, Nanopore analysis of yeast cell extract and event classification using few-shot learning (Methods). Experimentally, 0.1 g of yeast cell extract was dissolved in 1 ml of ultrapure water and the resulting solution was ultrafiltered. Then, the pH of the ultrafiltrate was adjusted to 7.0 by titration with KOH. With MspA-FPBA, 5 μ l of the above filtrate was added to the *cis* side and a +160-mV bias was continually applied. Lastly, various nanopore events were acquired and classified by few-shot learning. **b, c**, Representative traces acquired with yeast cell extract. Representative trace segments from the same original trace are shown in **b, c** to clearly illustrate event details. The trace was split into

two panels (**b, c**) due to space limitations. Events were automatically classified using few-shot learning (Fig. 4) and labeled with color-coded dots (amino acids, orange; NMPs, blue; saccharides, green; peptides, purple). With the trained bagged trees model (Fig. 3a), different amino acids and NMPs were further identified. **d**, The scatter plot of ΔI versus SD for amino acid events produced by the yeast cell extract. Seven clusters respectively corresponding to Glu, Ser, Val, Phe, Tyr, Asn and Arg were identified. **e**, The scatter plot of ΔI versus SD for NMPs events produced by yeast cell extract and the process of identifying the unidentified sample as IMP. All measurements in **b, c** were conducted under a bias of +160 mV, with 5 μ l of ultrafiltrate added to the *cis* side.

rapid identification of amino acids, NMPs, saccharides and peptides in real biological samples. First, commercially available yeast cell extract was dissolved in ultrapure water and subjected to ultrafiltration. Subsequently, the pH of the resulting filtrate was adjusted to neutral. Under a continually applied +160-mV bias, the filtrate was added to the *cis* side of the device. With the few-shot learning (Fig. 4), the acquired events were automatically classified (Methods, Fig. 5a and Supplementary Fig. 47).

Ultrafiltration pretreatment effectively removes macromolecular analytes from yeast cell extract (Supplementary Fig. 48). Specifically, the extract solution was loaded into an ultrafiltration tube with a 3-kDa molecular weight cutoff and ultrafiltered at 6,000 rpm for 60 min at 4 °C. Although nanopore analysis of untreated yeast cell extract is also feasible, macromolecular clogging or trapping events can more or less interrupt data acquisition (Supplementary Fig. 49). However, these interference events exhibit signature profiles distinct from those of small molecular analytes and can be excluded from statistical analysis after data acquisition.

Representative traces acquired with yeast cell extract are shown in Fig. 5b,c and Supplementary Video 3. The events of amino acids, NMPs, saccharides and peptides were automatically classified by the few-shot learning and marked with orange, blue, green and purple dots, respectively (Fig. 5b,c and Supplementary Figs. 50 and 51). For the classified amino acid and NMP events, they were further predicted by the trained bagged trees model (Fig. 3a). Corresponding scatter plots of ΔI versus SD were generated using the amino acid events and NMP events extracted from a 20-min continually recorded trace (Fig. 5d,e). With the trained bagged trees model, the seven amino acid clusters shown were identified to be Glu, Ser, Val, Phe, Tyr, Asn and Arg (Fig. 5d). Additionally, the presence of other amino acids such as Met, His, Gln, Ala, Thr and Trp was also detected by machine learning (Fig. 5b,c). Moreover, five NMP clusters were observed. In addition to all four canonical NMP events, an unidentified sample was also detected and its events are annotated with gray circles (Fig. 5e). The identification workflow of the unidentified sample is shown in the red box. First, using few-shot learning, the unidentified sample was first classified to be a type of

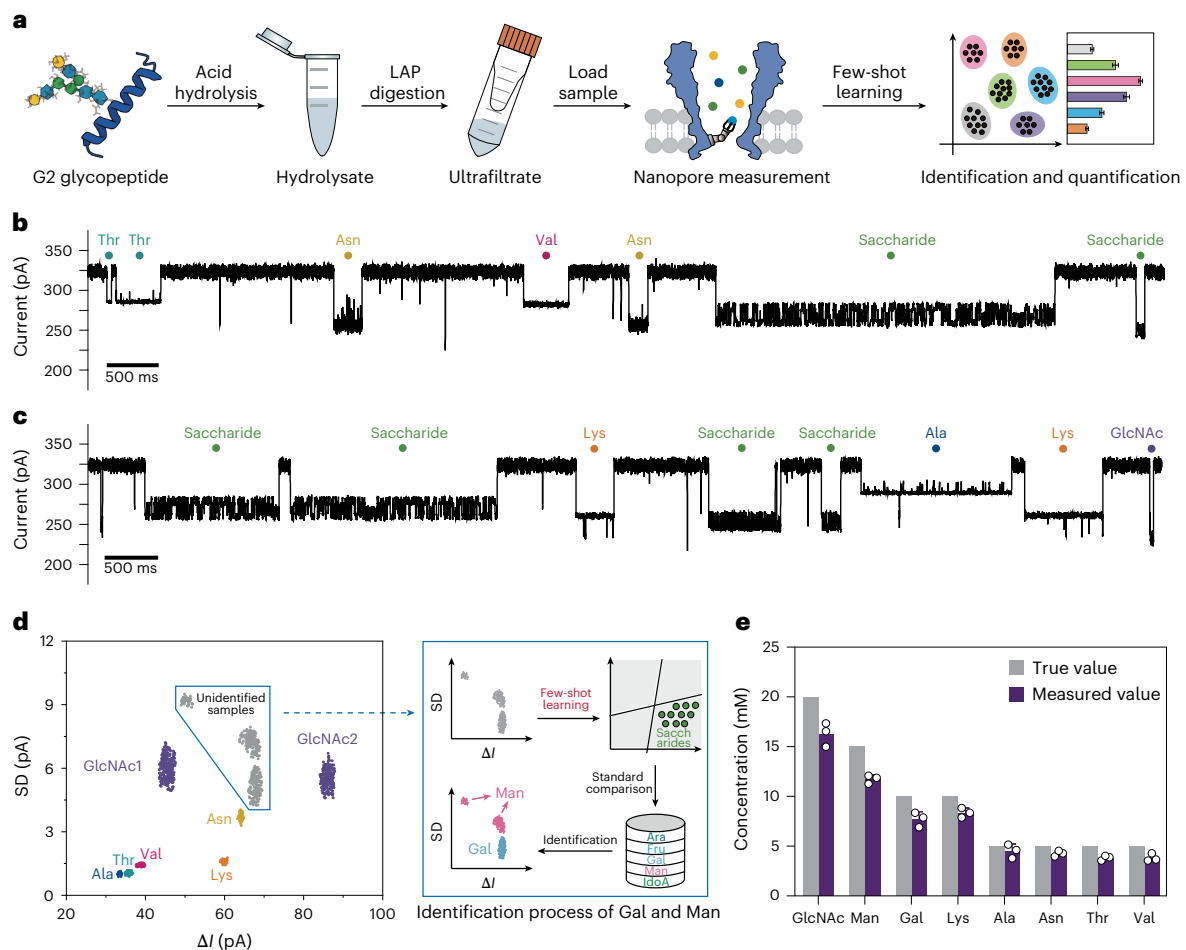


Fig. 6 | Quantitative analysis of amino acid and saccharide constituents in the G2 glycopeptide. **a**, Nanopore profiling of hydrolyzed G2 glycopeptide and quantitation of its components. Experimentally, 3 mg of G2 glycopeptide was dissolved in 250 μ l of 0.5 M TFA and incubated at 80 $^{\circ}$ C for 12 h for hydrolysis. Afterward, sodium phosphate was added to the solution to achieve 50 mM final concentration with pH adjusted to 8.0. The sample was subsequently enzymatically digested using LAP and subjected to ultrafiltration (Methods). With an MspA-FPBA, 10 μ l of filtrate was applied to the *cis* side with a +160-mV bias. Component identification and quantification were performed using nanopore analysis (Methods). **b,c**, Representative traces acquired with G2 glycopeptide hydrolysates. Representative trace segments from the same original trace are shown in **b,c** to clearly illustrate event details. The trace

was split into two panels (**b, c**) due to space limitations. Event classification was performed by few-shot learning. A trained bagged trees model enabled differential identification of amino acids and GlcNAc constituents. **d**, Scatter plot of ΔI versus SD for amino acid and saccharide events derived from G2 glycopeptide hydrolysates. In addition to the five identified amino acids and GlcNAc, three unidentified event clusters were observed, marked with gray circles. Few-shot learning first classified these samples as saccharides. Comparison to standard Gal and Man events (Supplementary Fig. 53) confirmed their identities through complete agreement of event features. **e**, Comparative analysis of component concentrations measured by nanopore versus true values. Data represent triplicate measurements ($n = 3$) presented as the mean $\pm$ s.d. Error bars indicate the s.d.

NMP. It is suggested that IMP, a flavoring nucleotide widely present in yeast cell extract⁶⁰, may have generated these events. Subsequently, the event features of the unidentified sample were compared to the standard events obtained for IMP (Supplementary Fig. 27) and its identity was confirmed on the basis of a complete match of all event features. The versatility of MspA-FPBA enables concurrent detection of diverse analytes within a single assay, offering a key advantage for real-world sample analysis, as demonstrated in the yeast extract analysis. Additionally, its high-resolution sensing ensures accurate and unambiguous identification, preventing misinterpretation errors from analytical cross-talk or quantification inaccuracies.

Components analysis of egg yolk-derived G2 glycopeptide

Because of the presence of structural isomers, current MS-based approaches encounter notable challenges in achieving direct compositional identification and quantification of native glycopeptides (Supplementary Fig. 52). In contrast, the MspA-FPBA platform demonstrates distinct advantages in isomer discrimination

(Supplementary Figs. 19 and 53). To demonstrate the technical potential for native glycopeptide analysis, a naturally occurring complex-type *N*-glycopeptide (G2) isolated from egg yolk was selected as a model system and subjected to hydrolysis for nanopore characterization (Methods and Fig. 6a). The raw traces are presented in Fig. 6b,c, accompanied by corresponding scatter plot analysis in Fig. 6d and Supplementary Fig. 54. The unidentified events were classified as saccharides by few-shot learning, suggesting that they might be galactose (Gal) and mannose (Man), two core monosaccharide components in mammalian *N*-linked glycans⁶¹. This assignment was further validated through direct comparison to reference MS profiles of standard Gal and Man (Supplementary Fig. 53). As shown in the quantification results (Fig. 6e and Supplementary Table 18), despite minor experimental discrepancies, the nanopore-derived measurements maintained a high degree of consistency with reference values. These findings confirm the suitability of this nanopore platform for both compositional profiling and quantitative analysis of native glycopeptides.

Beyond glycopeptides, MspA-FPBA demonstrates theoretical applicability for compositional profiling and quantitative analysis of other hybrid biomolecules, including but not limited to glycoRNAs and RNA-binding peptides, using the versatility of the maleimido-C2-FPBA adaptor. This nanopore-based method achieves the necessary versatility and resolution suitable for complex biomolecular characterization. While its proof-of-concept demonstration with a representative glycopeptide validates the technical feasibility of this methodology, the platform represents a promising breakthrough for integrated single-molecule analysis of coexisting biomolecular entities, an analytical challenge that remains unattainable through individual nanopore technology in the field. Looking forward, this approach could revolutionize structural characterization of distinct molecular domains and offer novel mechanistic insights into cross-omics interactions at the glycome–transcriptome–proteome interface. Moreover, the platform’s sensing mechanism suggests its potential to discover previously undetected PTM types and epigenetic modifications through synergistic advancements in corresponding databases and machine learning algorithms.

Discussion

We conjugated a custom-designed adaptor Maleimido-C2-FPBA to the pore constriction of MspA, forming MspA-FPBA and demonstrated the identification of amino acids, NMPs, saccharides and peptides all in the same assay. This study highlights the capability of MspA-FPBA to analyze a diverse array of analytes, including but not limited to 21 proteinogenic amino acids, three amino acids with PTMs, four canonical NMPs, three epigenetically modified NMPs, four monosaccharides and five peptides. Unlike conventional nanopore-based amino acid detection that relies on metal ion incorporation, MspA-FPBA offers a metal-free alternative, which is compatible with applications requiring metal ion presence. The integration of machine learning achieved a general accuracy of 98.7%, underscoring the robustness of this approach during simultaneous analysis of a wide variety of analytes within the same assay. Although not demonstrated here, this technique may be applicable to the analysis of D-amino acids, nucleotide sugars, oligosaccharides and oligonucleotides. We also demonstrated the identification of glycosylation modifications in glycosylated peptides, highlighting the potential of this technique in multiomics analysis. Furthermore, a few-shot learning strategy was implemented, enabling the classification of the general composition of biological samples such as yeast cell extract, without strictly relying on events generated by the identical standard compound. This technique effectively characterized native G2 glycopeptide composition and may be applied to glycoRNA and RNA-binding peptide composition analysis following a similar principle. Looking ahead, inspired by exopeptidase protein sequencing²⁶, the integration of proteases, ribonucleases and glycosidases with MspA-FPBA for exo-sequencing is expected to unify diverse sequencing technologies into a single and versatile system.

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/s41587-026-03308-9>.

References

- Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).
- Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
- Cravatt, B. F., Simon, G. M. & Yates III, J. R. The biological impact of mass-spectrometry-based proteomics. *Nature* **450**, 991–1000 (2007).
- Nesvizhskii, A. I. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J. Proteomics* **73**, 2092–2123 (2010).
- Deol, H. et al. After 75 years, an alternative to Edman degradation: a mechanistic and efficiency study of a base-induced method for N-terminal peptide sequencing. *J. Am. Chem. Soc.* **147**, 13973–13982 (2025).
- Holley, R. W. et al. Structure of a ribonucleic acid. *Science* **147**, 1462–1465 (1965).
- Sedlazeck, F. J., Lee, H., Darby, C. A. & Schatz, M. C. Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nat. Rev. Genet.* **19**, 329–346 (2018).
- Abdel-Ghany, S. E. et al. A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* **7**, 11706 (2016).
- Zhao, H. et al. Chromosome-level reference genome and alternative splicing atlas of moso bamboo (*Phyllostachys edulis*). *GigaScience* **7**, giy115 (2018).
- Li, S., Yamada, M., Han, X., Ohler, U. & Benfey, P. N. High-resolution expression map of the *Arabidopsis* root reveals alternative splicing and lincRNA regulation. *Dev. Cell* **39**, 508–522 (2016).
- Weirather, J. L. et al. Characterization of fusion genes and the significantly expressed fusion isoforms in breast cancer by hybrid sequencing. *Nucleic Acids Res.* **43**, e116 (2015).
- Garalde, D. R. et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* **15**, 201–206 (2018).
- Zhao, L. et al. Analysis of transcriptome and epitranscriptome in plants using PacBio Iso-Seq and nanopore-based direct RNA sequencing. *Front. Genet.* **10**, 253 (2019).
- Urban, J. et al. Predicting glycan structure from tandem mass spectrometry via deep learning. *Nat. Methods* **21**, 1206–1215 (2024).
- Yao, H.-Y., Wang, J.-Q., Yin, J.-Y., Nie, S.-P. & Xie, M.-Y. A review of NMR analysis in polysaccharide structure and conformation: progress, challenge and perspective. *Food Res. Int.* **143**, 110290 (2021).
- Krishnamoorthy, L. & Mahal, L. K. Glycomic analysis: an array of technologies. *ACS Chem. Biol.* **4**, 715–732 (2009).
- 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).
- Nivala, J., Marks, D. B. & Akesson, M. Unfoldase-mediated protein translocation through an α -hemolysin nanopore. *Nat. Biotechnol.* **31**, 247–250 (2013).
- Yu, L. et al. Unidirectional single-file transport of full-length proteins through a nanopore. *Nat. Biotechnol.* **41**, 1130–1139 (2023).
- 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).
- Li, M. et al. Identification of tagged glycans with a protein nanopore. *Nat. Commun.* **14**, 1737 (2023).
- Xia, B. et al. Mapping the acetylamino and carboxyl groups on glycans by engineered α -hemolysin nanopores. *J. Am. Chem. Soc.* **145**, 18812–18824 (2023).
- 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).
- Yan, S. et al. Single molecule ratcheting motion of peptides in a *Mycobacterium smegmatis* porin A (MspA) nanopore. *Nano Lett.* **21**, 6703–6710 (2021).
- Motone, K. et al. Multi-pass, single-molecule nanopore reading of long protein strands. *Nature* **633**, 662–669 (2024).
- Bonini, A., Sauciuc, A. & Maglia, G. Engineered nanopores for exopeptidase protein sequencing. *Nat. Methods* **21**, 16–17 (2024).

27. Haworth, W. N., Peat, S. & Bourne, E. J. Synthesis of amylopectin. *Nature* **154**, 236–236 (1944).
28. Meyer, K. H., Gürtler, P. & Bernfeld, P. Structure of amylopectin. *Nature* **160**, 900–901 (1947).
29. Faller, M., Niederweis, M. & Schulz, G. E. The structure of a mycobacterial outer-membrane channel. *Science* **303**, 1189–1192 (2004).
30. 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).
31. Zhang, S. et al. A nanopore-based saccharide sensor. *Angew. Chem. Int. Ed. Engl.* **61**, e202203769 (2022).
32. Zhang, S. et al. Discrimination of disaccharide isomers of different glycosidic linkages using a modified MspA nanopore. *Angew. Chem. Int. Ed. Engl.* **63**, e202316766 (2024).
33. Wang, Y. et al. Identification of nucleoside monophosphates and their epigenetic modifications using an engineered nanopore. *Nat. Nanotechnol.* **17**, 976–983 (2022).
34. Wang, K. et al. Unambiguous discrimination of all 20 proteinogenic amino acids and their modifications by nanopore. *Nat. Methods* **21**, 92–101 (2024).
35. Sun, W. et al. Nanopore discrimination of rare earth elements. *Nat. Nanotechnol.* **20**, 523–531 (2025).
36. Cal, P. M. S. D. et al. Iminoboronates: a new strategy for reversible protein modification. *J. Am. Chem. Soc.* **134**, 10299–10305 (2012).
37. Cambray, S. & Gao, J. Versatile bioconjugation chemistries of ortho-boronyl aryl ketones and aldehydes. *Acc. Chem. Res.* **51**, 2198–2206 (2018).
38. Haggett, J. G. & Domaille, D. W. ortho-Boronic acid carbonyl compounds and their applications in chemical biology. *Chemistry* **30**, e202302485 (2024).
39. Arnal-Hérault, C. et al. Functional G-quartet macroscopic membrane films. *Angew. Chem. Int. Ed. Engl.* **46**, 8409–8413 (2007).
40. Hutin, M., Bernardinelli, G. & Nitschke, J. R. An iminoboronate construction set for subcomponent self-assembly. *Chemistry* **14**, 4585–4593 (2008).
41. Akgun, B. & Hall, D. G. Boronic acids as bioorthogonal probes for site-selective labeling of proteins. *Angew. Chem. Int. Ed. Engl.* **57**, 13028–13044 (2018).
42. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
43. Fisher, A., Rudin, C. & Dominici, F. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res.* **20**, 177 (2019).
44. Boersma, A. J. & Bayley, H. Continuous stochastic detection of amino acid enantiomers with a protein nanopore. *Angew. Chem. Int. Ed. Engl.* **51**, 9606–9609 (2012).
45. Guo, Y. et al. Metal-organic complex-functionalized protein nanopore sensor for aromatic amino acids chiral recognition. *Analyst* **142**, 1048–1053 (2017).
46. Zhang, M. et al. Real-time detection of 20 amino acids and discrimination of pathologically relevant peptides with functionalized nanopore. *Nat. Methods* **21**, 609–618 (2024).
47. Ayub, M., Hardwick, S. W., Luisi, B. F. & Bayley, H. Nanopore-based identification of individual nucleotides for direct RNA sequencing. *Nano Lett.* **13**, 6144–6150 (2013).
48. Ramsay, W. J. & Bayley, H. Single-molecule determination of the isomers of D-glucose and D-fructose that bind to boronic acids. *Angew. Chem. Int. Ed. Engl.* **57**, 2841–2845 (2018).
49. Ouldali, H. et al. Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nat. Biotechnol.* **38**, 176–181 (2020).
50. Zhang, Y. et al. Peptide sequencing based on host-guest interaction-assisted nanopore sensing. *Nat. Methods* **21**, 102–109 (2024).
51. Piguet, F. et al. Identification of single amino acid differences in uniformly charged homopolymeric peptides with aerolysin nanopore. *Nat. Commun.* **9**, 966 (2018).
52. Singh, P. R. et al. Pulling peptides across nanochannels: resolving peptide binding and translocation through the hetero-oligomeric channel from *Nocardia farcinica*. *ACS Nano* **6**, 10699–10707 (2012).
53. Ji, Z., Kang, X., Wang, S. & Guo, P. Nano-channel of viral DNA packaging motor as single pore to differentiate peptides with single amino acid difference. *Biomaterials* **182**, 227–233 (2018).
54. Miyagi, M., Takiguchi, S., Hakamada, K., Yohda, M. & Kawano, R. Single polypeptide detection using a translocon EXP2 nanopore. *Proteomics* **22**, e2100070 (2022).
55. Huang, G., Voet, A. & Maglia, G. FraC nanopores with adjustable diameter identify the mass of opposite-charge peptides with 44 dalton resolution. *Nat. Commun.* **10**, 835 (2019).
56. 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.* **10**, 1167–1178 (2024).
57. Wei, X., Wen, J., Wu, H., Qu, Z. & Huang, G. Obtaining narrow distributions of single-molecule peptide signals enables sensitive peptide discrimination with α -hemolysin nanopores. *J. Am. Chem. Soc.* **147**, 9304–9315 (2025).
58. Wang, Y., Yao, Q., Kwok, J. T. & Ni, L. M. Generalizing from a few examples: a survey on few-shot learning. *ACM Comput. Surv.* **53**, 63 (2020).
59. Tomé, D. Yeast extracts: nutritional and flavoring food ingredients. *ACS Food Sci. Technol.* **1**, 487–494 (2021).
60. Tao, Z. et al. Yeast extract: characteristics, production, applications and future perspectives. *J. Microbiol. Biotechnol.* **33**, 151–166 (2023).
61. He, M., Zhou, X. & Wang, X. Glycosylation: mechanisms, biological functions and clinical implications. *Signal Transduct. Target. Ther.* **9**, 194 (2024).

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 America, Inc. 2026

Methods

The synthesis of maleimido-C2-FPBA

Maleic anhydride (compound **1**; 13.50 g, 137.68 mmol, 1.0 eq) and *tert*-butyl *N*-(2-aminoethyl)carbamate (24.26 g, 151.44 mmol, 1.1 eq) were dissolved in dichloromethane (DCM; 500 ml) in a 2-L single-neck round-bottom flask. The reaction mixture was stirred for 16 h at room temperature and then concentrated under reduced pressure to remove the solvent. Acetone (850 ml) was added to the residue, followed by sequential addition of sodium acetate (12.42 g, 151.44 mmol, 1.1 eq), acetic anhydride (15.46 g, 151.44 mmol, 1.1 eq) and triethylamine (27.86 g, 275.35 mmol, 2.0 eq). The resulting mixture was stirred at 25 °C for an additional 16 h. The reaction mixture was concentrated, loaded onto silica gel and purified by flash column chromatography using a gradient of ethyl acetate and *n*-heptane (v/v, 1:10). A white solid was obtained (compound **2**; 15.00 g, 62.43 mmol, 45.35% yield). ¹H NMR (400 MHz, CDCl₃, δ): 7.00 (dd, 2H), 6.89 (m, 1H), 3.42–3.45 (m, 2H), 3.03–3.08 (m, 2H), 1.34–1.37 (m, 9H).

Tert-butyl (2-(2,5-dioxo-2,5-dihydro-1H-pyrrol-1-yl)ethyl)carbamate (compound **2**; 15.00 g, 62.43 mmol, 1.0 eq) was dissolved in TFA and DCM (30 ml each) and stirred in ambient air for 16 h. A white solid was obtained (compound **3**; 8.74 g, 62.43 mmol), which was used for next step directly. A 50-ml single-neck round-bottom flask was charged with 3-borono-4-formylbenzoic acid (compound **4**; 100.00 mg, 515.60 μmol, 1.0 eq), HATU (196.05 mg, 515.60 μmol, 1.0 eq) and DIPEA (133.27 mg, 1.03 mmol, 2.0 eq) under nitrogen atmosphere at 0 °C. THF (10 ml) was added and the system was purged with nitrogen three times before stirring at 25 °C for 30 min. Subsequently, compound **3** (72.26 mg, 515.60 μmol, 1.0 eq) was added, and the reaction was continued for 11.5 h. The obtained crude product was purified by HPLC to obtain the target product as a white solid (maleimido-C2-FPBA, 41.00 mg, 129.72 μmol, 25.16% yield). ¹H NMR (400 MHz, CDCl₃, δ): 10.03 (s, 1H), 8.56 (s, 1H), 8.20–8.18 (d, 1H), 8.05–8.03 (d, 1H), 6.78 (s, 2H), 3.91–3.88 (m, 2H), 3.75–3.71 (m, 2H), 2.85 (s, 2H). LC–MS (*m/z*) for C₁₄H₁₃BN₂O₆ [M + H]⁺: calculated, 317.09; found, 317.06.

All reactions were carried out in air/oxygen atmosphere using standard Schlenk techniques unless stated otherwise. The glassware was dried in an oven (110 °C) and heated under reduced pressure before use. Thin-layer chromatography and column chromatography analyses were performed using Anhui Liangchen silica gel (200–300) with distilled solvents. NMR spectra were recorded on a Bruker Avance 400 spectrometer operating at 400 MHz (¹H NMR) in CDCl₃. All ¹H NMR chemical shifts were reported in ppm relative to internal references of CDCl₃ at 7.26 ppm. The following abbreviations are used to describe peak patterns where appropriate: singlet (s), doublet (d), triplet (t), multiplet (m) and doublet of doublets (dd).

The preparation of hetero-octameric MspA

The hetero-octameric MspA, referred to as (N90C)₁(M2)₇, is assembled from seven identical M2 MspA-D16H6 subunits and a single N90C MspA-H6 subunit. The M2 MspA-D16H6 and N90C MspA-H6 are both mutants of M2 MspA, engineered with a C-terminal H6 tag. M2 MspA-D16H6 additionally carries a D16 tag, which is essential for separating different hetero-octameric assemblies of MspAs. In contrast, Asn90Cys MspA-H6 contains a single-point substitution of asparagine to cysteine. Both genes were custom-synthesized and coexpressed by cloning them into the same pETDuet-1 vector, followed by transformation into *Escherichia coli* BL21 (DE3) pLysS competent cells (Sangon Biotech) for protein production.

E. coli BL21 (DE3) pLysS cells harboring the recombinant plasmids were plated on Luria–Bertani (LB) agar supplemented with ampicillin (50 μg ml⁻¹) and chloramphenicol (34 μg ml⁻¹) and incubated at 37 °C for 15 h. A single colony was picked and cultured in 100 ml of LB broth containing the same antibiotics at 37 °C overnight with shaking. The overnight culture was then transferred to 1 L of LB broth with antibiotics and incubated at 37 °C, 3g for 5–6 h until the optical density

at 600 nm reached 0.7. Protein expression was induced by adding dioxane-free IPTG to a final concentration of 0.1 mM, followed by incubation at 16 °C, 3g for 24 h. The bacterial pellet was harvested by centrifuging the medium at 1,500g and 4 °C for 20 min and the resulting pellet was stored at –80 °C for long-term use.

The bacterial pellet was resuspended with 150 ml of lysis buffer (100 mM Na₂HPO₄/NaH₂PO₄, 0.1 mM EDTA, 150 mM NaCl and 0.5% (w/v) Genapol X-80, pH 6.5) and heated at 60 °C for 50 min. The obtained lysate was centrifuged at 16,000g and 4 °C for 40 min. The supernatant, which contained the target protein, was collected. Then, MspA was purified by nickel affinity chromatography and eluted with a linear gradient of imidazole (5–500 mM) by mixing buffer A (0.5 M NaCl, 20 mM HEPES, 5 mM imidazole, 2 mM TCEP and 0.5% (w/v) Genapol X-80, pH 8.0) with buffer B (0.5 M NaCl, 20 mM HEPES, 500 mM imidazole, 2 mM TCEP and 0.5% (w/v) Genapol X-80, pH 8.0). The elution fractions were characterized by 4–15% SDS–PAGE. Different assemblies of hetero-octameric MspAs were separated by 10% SDS–PAGE for 16 h. The protein band corresponding to (N90C)₁(M2)₇ was excised, with which the (N90C)₁(M2)₇ was recovered by incubation in an extraction solution (150 mM NaCl, 15 mM Tris-HCl, 0.2% DDM, 0.5% Genapol X-80, 5 mM TCEP and 10 mM EDTA, pH 7.5) for 12 h. The collected (N90C)₁(M2)₇ was either used immediately or stored at –80 °C for long-term use.

Nanopore modification

To modify (N90C)₁(M2)₇ with FPBA, the previously prepared (N90C)₁(M2)₇ and maleimido-C2-FPBA (20 mM) were mixed and coincubated for 30 min at 4 °C at a volume ratio of 1:8. The resulting product, which is referred to as MspA-FPBA, was immediately used or stored at –80 °C.

Electrolyte buffer preparations

The KCl buffers used in single-channel measurement (1.5 M KCl, 10 mM CHES pH 9.0, 1.5 M KCl, 10 mM MES pH 6.0, 1.5 M KCl and 10 mM MOPS pH 7.0) were prepared with Milli-Q water. Then, the buffer was filtered through a membrane (0.2 μm) under reduced pressure to remove insoluble substances. Lastly, the pH of the electrolyte buffers was adjusted to the desired values before use.

Nanopore measurements

Nanopore single-channel measurements were conducted inside a custom-built Faraday cage mounted on an optical table (Jiangxi Liansheng Technology). The measurement setup consisted of two custom-fabricated chambers separated by a Teflon film containing a drilled aperture (~100 μm) at the center. This drilled aperture was pre-treated with 0.5% (v/v) hexadecane in pentane before each experiment. By convention, the electrically grounded chamber is designated as the *cis* side, while the opposite chamber is referred to as the *trans* side. The *cis* and *trans* chambers were each filled with 0.5 ml of 1.5 M KCl buffer (*cis*: 1.5 M KCl and 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl and 10 mM MES, pH 6.0) and connected through Ag/AgCl electrodes to a patch-clamp amplifier. A lipid bilayer was formed across the aperture by adding a drop of diphytanoylphosphatidylcholine (5 mg ml⁻¹ in pentane) to both chambers. Nanopores were then introduced to the *cis* chamber to initiate insertion. Upon detection of a single nanopore insertion, the *cis* chamber was promptly replaced with fresh 1.5 M KCl buffer to suppress additional nanopore insertions. All electrophysiological measurements were carried out with an Axonpatch 200B patch-clamp amplifier paired with a Digidata 1550B digitizer.

Data analysis

Nanopore events were detected from raw single-channel recording traces using the single-channel search function in Clampfit 10.7 (Molecular Devices), with events shorter than 2 ms in *t*_{off} excluded from analysis. For each detected event, nine event features, that is, ΔI, SD, local SD ratio, minimum, maximum, range, *t*_{off}, skew and kurt,

were extracted using a custom MATLAB program. All subsequent data analysis was conducted in Origin 2023b (Origin Lab).

Machine learning analysis was conducted using MATLAB's classification learner toolbox. A total of 200 events acquired with each amino acid (except lysine, 100 events) were included. A total of 100 type 1 events acquired with each NMP and 200 events with each saccharide were included. All events were collected to construct a labeled dataset, with each event labeled according to the corresponding amino acid, NMP and saccharide used during data acquisition. The extracted event features, that is, ΔI , SD, local SD ratio, minimum, maximum, range, t_{off} , skew and kurt, were assembled into a feature matrix, which was randomly divided into a training set (80%) and a testing set (20%). The training and testing sets were used by MATLAB's classification learner toolbox for model training and testing. Multiple built-in classification algorithms, including KNN, SVM, naive Bayes, decision trees, neural network, discriminant analysis and ensemble, were evaluated. To reduce the risk of overfitting, tenfold cross-validation was applied and both validation accuracy and test accuracy were recorded. The tenfold cross-validation was carried out by randomly dividing the training set into ten equal subsets. Each subset was used once as the validation set, while the remaining nine subsets were used for model training. This process was repeated ten times and the average validation accuracy was calculated to evaluate model performance. The classifier with the highest cross-validation accuracy was selected as the optimal model and subsequently applied to predict unlabeled data. A confusion matrix was generated to visualize the results of the model. A learning curve based on varying sample sizes was constructed to evaluate the efficiency of model training. A DBSCAN analysis was performed using Python, with the epsilon parameter set to 0.5 and the minimum number of points set to 5.

Nanopore measurements of the mixed sample

About 1.7 mg of the tripeptide Leu-Arg-Gly, 3.5 mg of Arg, 4.0 mg of Asn, 6.0 mg of Met, 11.3 mg of Tyr, 55.2 mg of UMP and 90.1 mg of Fru were dissolved together in 10 ml of ultrapure water to achieve final concentrations of 0.5, 2, 3, 4, 5, 15 and 50 mM, respectively. To initiate the nanopore measurement, 10 μl of the standard sample was added to the *cis* chamber. The measurements were carried out in a 1.5 M KCl buffer (*cis*: 1.5 M KCl and 10 mM CHES pH 9.0; *trans*: 1.5 M KCl and 10 mM MES pH 6.0). A transmembrane voltage of +160 mV was continually applied during the measurements.

Quantitative analysis of the standard sample

By using an experimentally measurable parameter F , defined as the number of nanopore events per min generated by a specific species when MspA-FPBA is not occupied by other species, nanopore measurements were conducted to investigate the concentration-dependent behavior of various species (Supplementary Figs. 30–36). The results indicate that F strongly correlates with the applied concentration of species.

For a specific analyte, the F value versus its concentration (C) can be described by a linear relationship:

$$F_i = a_i + b_i C_i \quad (1)$$

Here, i (from 1 to 7) represents the identity of different analytes, such as Met, Asn, Tyr, Arg, UMP, Fru and Leu-Arg-Gly, respectively. The parameters a_i and b_i are coefficients for the specific analytes, they were derived from linear fitting of the concentration-dependent measurements (Supplementary Figs. 30–36).

Subsequently, the concentration of each component in the standard sample can be calculated using the experimentally measured F_i values (Supplementary Table 11) and the previously determined coefficients (Supplementary Table 12) according to the following equation, which is derived from Eq. (1):

$$C_i = \frac{F_i - a_i}{b_i} \quad (2)$$

Glycosylated peptide digestion

The Gly-Met-Gln-Arg-Asn-(GlcNAc)-Ile-Ser peptide was custom-synthesized and used as a model glycosylated peptide for the demonstration. Gly-Met-Gln-Arg-Asn-(GlcNAc)-Ile-Ser peptide was dissolved in a 50 mM sodium phosphate pH 8.0 buffer with a concentration of 10 mg ml⁻¹. LAP was prepared in a 50 mM sodium phosphate pH 8.0 buffer with a concentration of 50 mg ml⁻¹ (>7 U per mg). To initiate the hydrolysis reaction, 10 μl of LAP solution and 20 μl of 10 mM MgCl₂ were added to 70 μl of peptide solution and incubated at 37 °C for 12 h in a dry block incubator. The reaction was stopped by heating the mixture to 80 °C for 5 min to inactivate the LAP. Afterward, another 100 μl of ultrapure water was added to the mixture and loaded into an ultracentrifuge tube with a 10-kDa molecular weight cutoff. The filtration was then performed at 8,000 rpm for 60 min at 4 °C. The digestion ultrafiltrate was collected and stored at -20 °C for subsequent single-channel measurement.

Nanopore analysis of amino acids, NMPs, saccharides and peptides in yeast cell extract

The yeast cell extract, a commercially available nutrient-rich additive, was purchased from Aladdin (China) and dissolved in ultrapure water with a concentration of 100 mg ml⁻¹. Afterward, the yeast cell extract solution was loaded into an ultracentrifuge tube with a 3-kDa molecular weight cutoff. The filtration was then performed at 6,000 rpm for 60 min at 4 °C. Then, the pH of the ultrafiltrate was adjusted to 7.0 by titration with KOH, forming a stock solution. To initiate the nanopore measurement, the 1.5 M KCl buffers (*cis*: 1.5 M KCl and 10 mM CHES pH 9.0; *trans*: 1.5 M KCl and 10 mM MES pH 6.0) were used. Next, 5 μl of the stock solution was added to the *cis* chamber and a +160-mV bias was continuously applied. Lastly, amino acid, NMP, saccharide and peptide events were classified and identified using the developed machine learning algorithm.

G2 glycopeptide digestion

The G2 glycopeptide (3 mg) was dissolved in 250 μl of 0.5 M TFA and incubated at 80 °C for 12 h. Then, sodium phosphate was added to the acid hydrolysis product to a final concentration of 50 mM and the pH was adjusted to 8.0. Subsequently, the solution was digested with LAP and ultrafiltered according to the previously described method. Lastly, the digestion ultrafiltrate was collected and stored at -20 °C for subsequent single-channel measurement.

G2 glycopeptide composition quantification

During nanopore analysis of G2 glycopeptide hydrolysates, the target amino acid and saccharide concentrations (C_i) were evaluated according to the following equation:

$$C_i = E_i / (\delta_i \times t)$$

Here, the annotation i (from 1 to 8) stands for parameters relevant to amino acids (Lys, Ala, Asn, Thr and Val) and saccharides (Gal, Man and GlcNAc). E_i is the number of corresponding analyte binding events detected during a continuous sensing of G2 glycopeptide hydrolysates. δ_i is the calibration coefficient, which is defined as the number of analyte binding events occurring per unit concentration (μM) per min. The values of δ_i were acquired during measurements with a sole analyte at a known concentration at +160 mV. δ_i values are also summarized in Supplementary Table 18. t is the recording time.

Statistics and reproducibility

All abf files recorded with Clampfit 10.7 were imported to MATLAB (MathWorks). For results acquired using MspA-FPBA (Fig. 3a), 200

events acquired with each amino acid (except lysine, 100 events) were included, whereas 100 type 1 events acquired with each NMP and 200 events with each saccharide were included. All events were collected to form a training set for model building. The trained model was then applied for predictions of unlabeled data. Unless indicated differently, the statistical results were derived from three independent measurements and the mean values of three measurements are reported, with error bars representing s.d. values. No statistical method was used to predetermine sample size. No data were excluded from the analyses. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

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

Data availability

The data supporting the results of this study are available within the article and its Supplementary Information. The training data are provided in Supplementary Information, referred to as 'general classifier'. Source data are provided with this paper.

Code availability

The custom machine learning codes and training data for the identification of amino acids, NMPs, saccharides and peptides, as well as for few-shot learning, are provided in Supplementary Information.

Author contributions

S.H. and L.Y. conceptualized the project. L.Y. performed pore engineering. L.Y., Z.W., J.C., Y.W., W.L., L.Z., X.D. and L.Q. performed the nanopore measurements. L.Y., Y.X., W.S., H.Z. and K.W. designed the machine learning algorithms. L.Y., Y.X. and W.S. prepared the supplementary videos. P.Z. set up the instruments. S.H. and L.Y. wrote the paper. S.H. supervised the project.

Funding

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

Competing interests

S.H. and L.Y. have filed patents describing the preparation of MspA containing a single FPBA adaptor and its applications thereof. The other authors declare no other competing interests.

Additional information

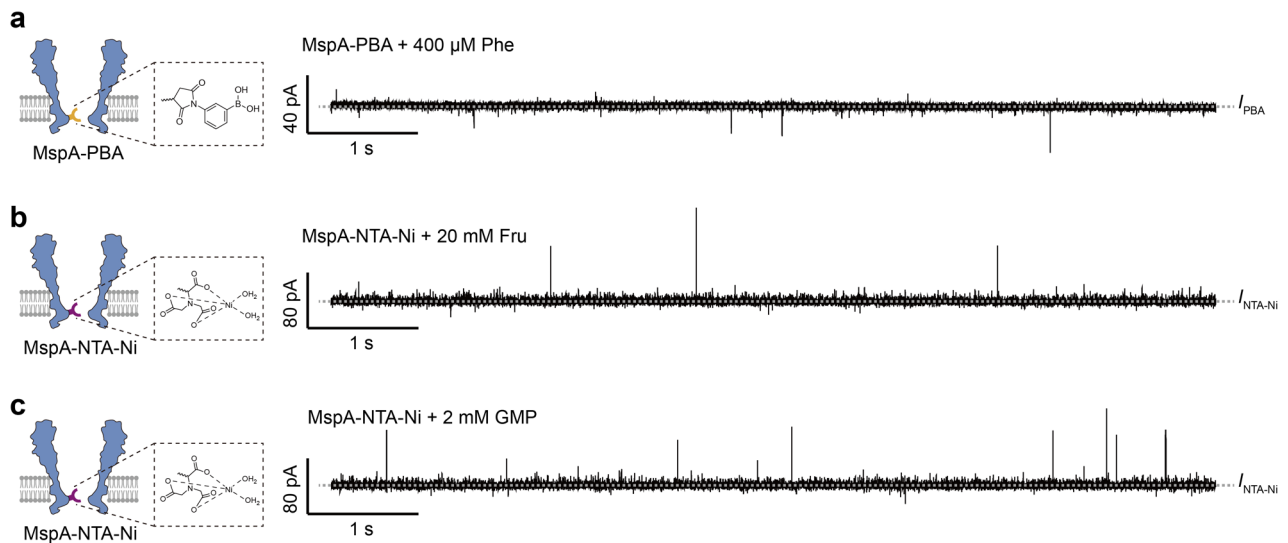
Extended data is available for this paper at <https://doi.org/10.1038/s41587-026-03308-9>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41587-026-03308-9>.

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

Peer review information *Nature Biotechnology* thanks Manish Kumar who co-reviewed with Priyanshu R. Gupta, Meni Wanunu and the other, anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

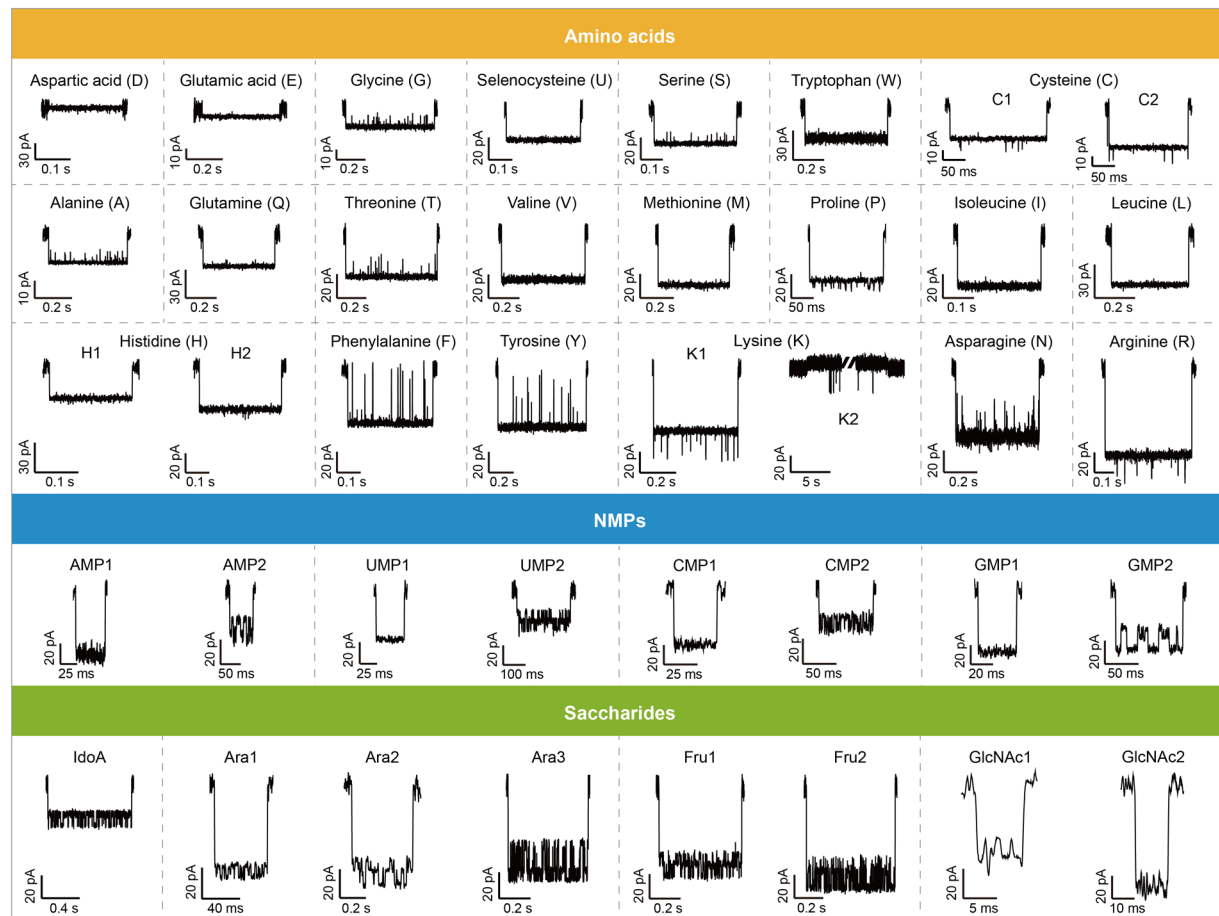
Reprints and permissions information is available at www.nature.com/reprints.



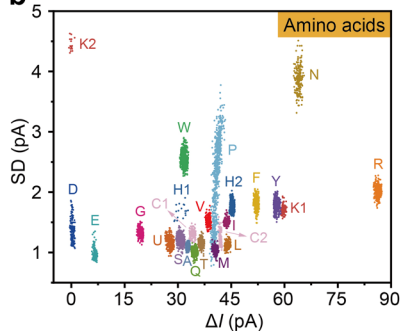
Extended Data Fig. 1 | Amino acid, saccharide and NMP sensing performed by different types of nanopores. (a) Sensing of phenylalanine (Phe) with MspA-PBA. Left: The structure of MspA-PBA. Right: A representative trace of Phe sensing performed with MspA-PBA. **(b)** Sensing of D-fructose (Fru) with MspA-NTA-Ni. Left: The structure of MspA-NTA-Ni. Right: A representative trace of Fru sensing performed with MspA-NTA-Ni. **(c)** Sensing of guanosine 5'-monophosphate (GMP) with MspA-NTA-Ni. Left: The structure of MspA-NTA-Ni.

Right: A representative trace of GMP sensing performed with MspA-NTA-Ni. The final concentrations of Phe, Fru and GMP were 400 μ M, 20 mM and 2 mM, respectively. No nanopore events were observed in all above measurements. The open pore current of MspA-PBA (+160 mV bias) and MspA-NTA-Ni (+100 mV bias) were defined as I_{PBA} and $I_{\text{NTA-Ni}}$, respectively. The corresponding open pore current was marked by a gray dashed line.

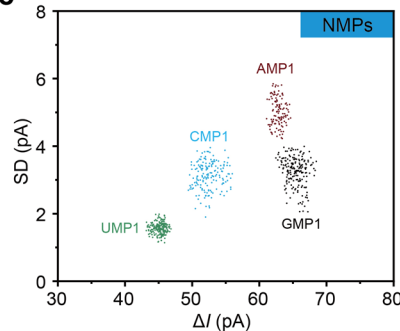
a



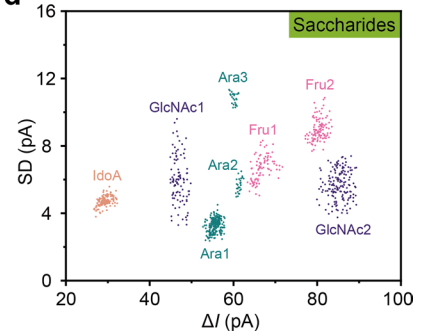
b



c

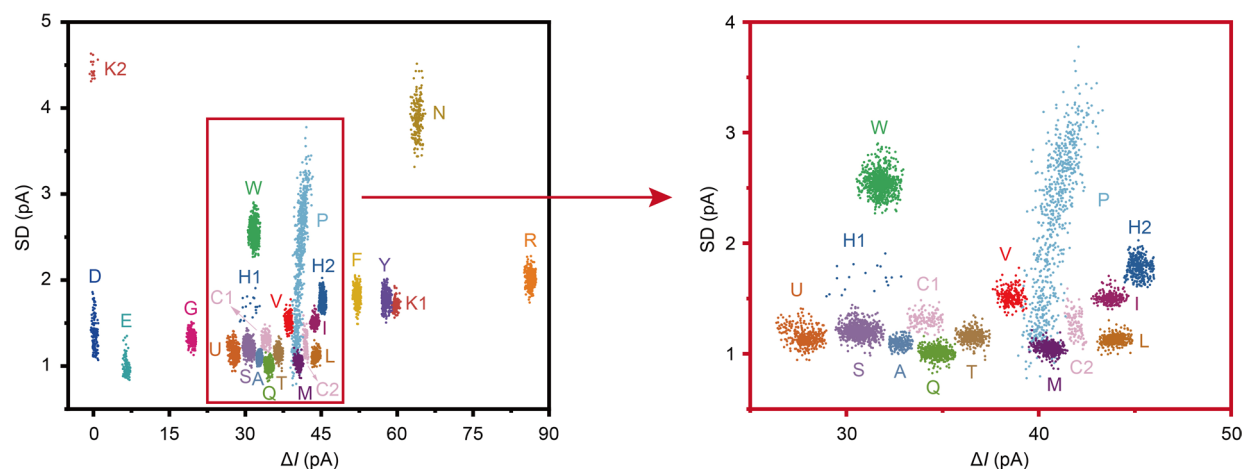


d



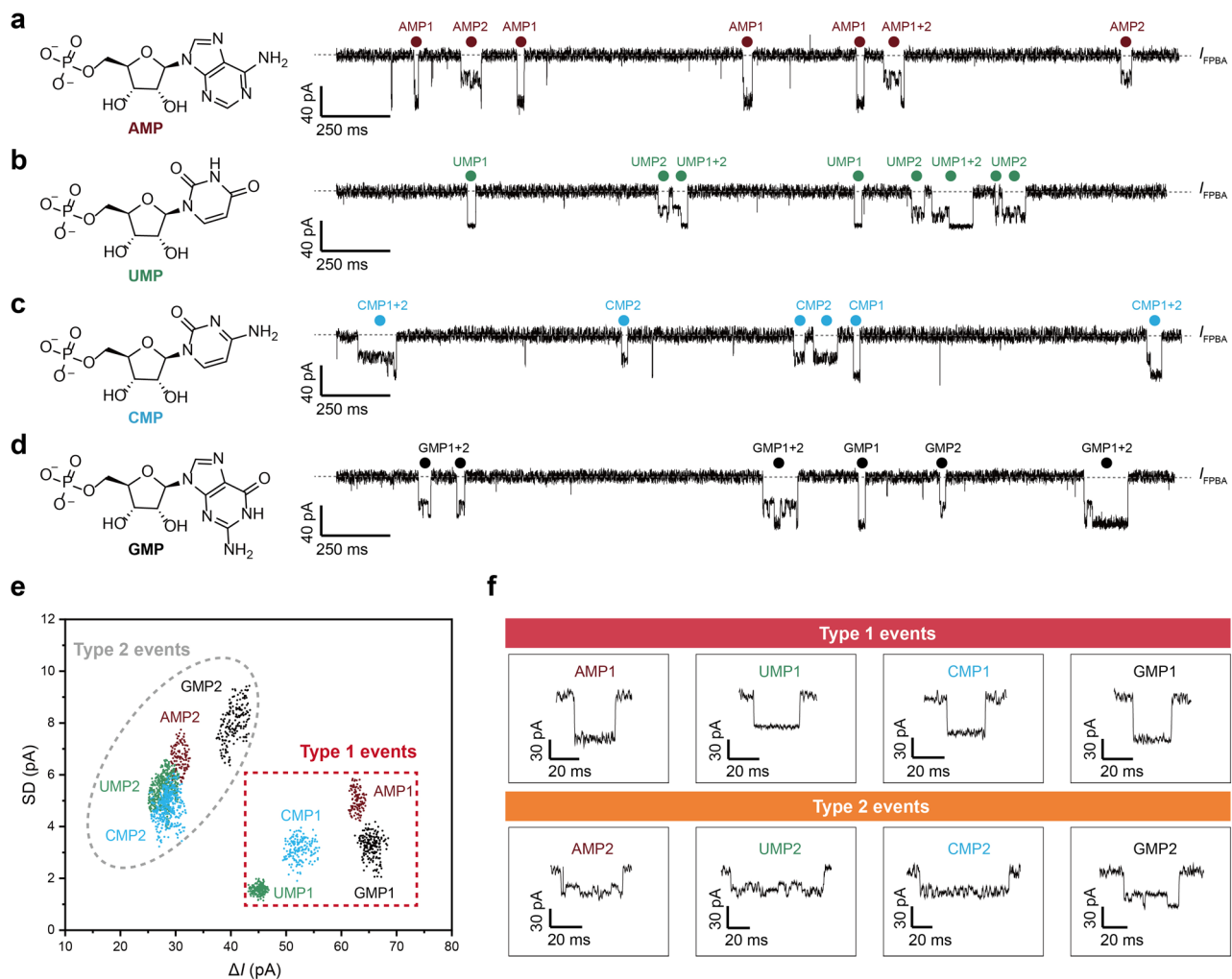
Extended Data Fig. 2 | Distinguishing 21 proteinogenic amino acids, 4 canonical NMPs and 4 saccharides using MspA-FPBA. (a) Representative events of different amino acids, NMPs and saccharides acquired by MspA-FPBA (Methods). 21 amino acids (including 20 common proteinogenic amino acids and selenocysteine), 4 NMPs (adenosine 5'-monophosphate, AMP; uridine 5'-monophosphate, UMP; cytidine 5'-monophosphate, CMP; guanosine 5'-monophosphate, GMP), and 4 saccharides (iduronic acid, IdoA; L-arabinose, Ara; D-fructose, Fru; N-acetyl-D-glucosamine, GlcNAc) were respectively added to the *cis* side during the measurement. Cysteine, histidine and lysine each have two representative types of events, respectively labeled as C1/2, H1/2 and K1/2. Notably, cysteine was blocked with maleimide prior to the measurement to avoid

the appearance of long-residing events (Supplementary Fig. 14). Each NMPs has two representative types of events, respectively labeled as AMP1/2, UMP1/2, CMP1/2, and GMP1/2. Ara has three types of representative events, labeled as Ara1/2/3. Fru and GlcNAc both have two types of representative events, labeled as Fru1/2 and GlcNAc1/2, respectively. (b) The scatter plot of ΔI versus SD for events of the 21 amino acids. The scatter plot was generated using at least 100 events for each amino acid. (c) The scatter plot of ΔI versus SD for the type 1 events (A/U/C/GMP1) of 4 NMPs. At least 100 type 1 events of each NMP were used to generate the plot. (d) The scatter plot of ΔI versus SD for the representative events of the 4 saccharides. The plot was generated using at least 100 events for each saccharide.



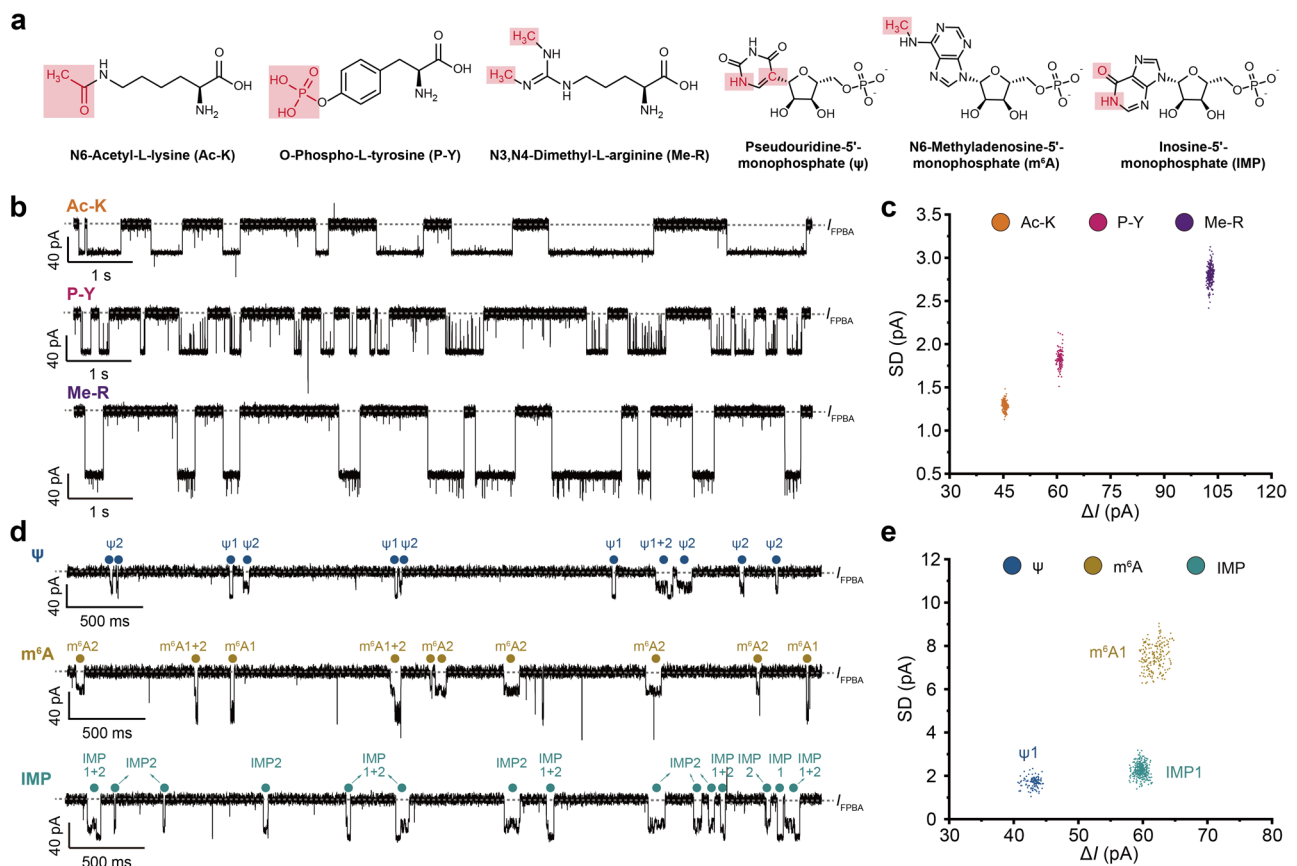
Extended Data Fig. 3 | The scatter plot of ΔI versus SD of events acquired with different amino acids. At least 100 events acquired with each amino acid were used to generate the plot, according to which, most amino acid events are fully distinguishable. To clarify the detail, the events inside the red box are further zoomed-in and shown on the right. Although the events corresponding

to P and M appear to overlap in the plot, their event characteristics are visually different and can be discriminated in the 3D event scatter plot of ΔI , SD, and max (Supplementary Fig. 18). High-precision identification of P and M events can be achieved by machine learning, with validation accuracies of 98.8% and 99.4% respectively (Fig. 3a).



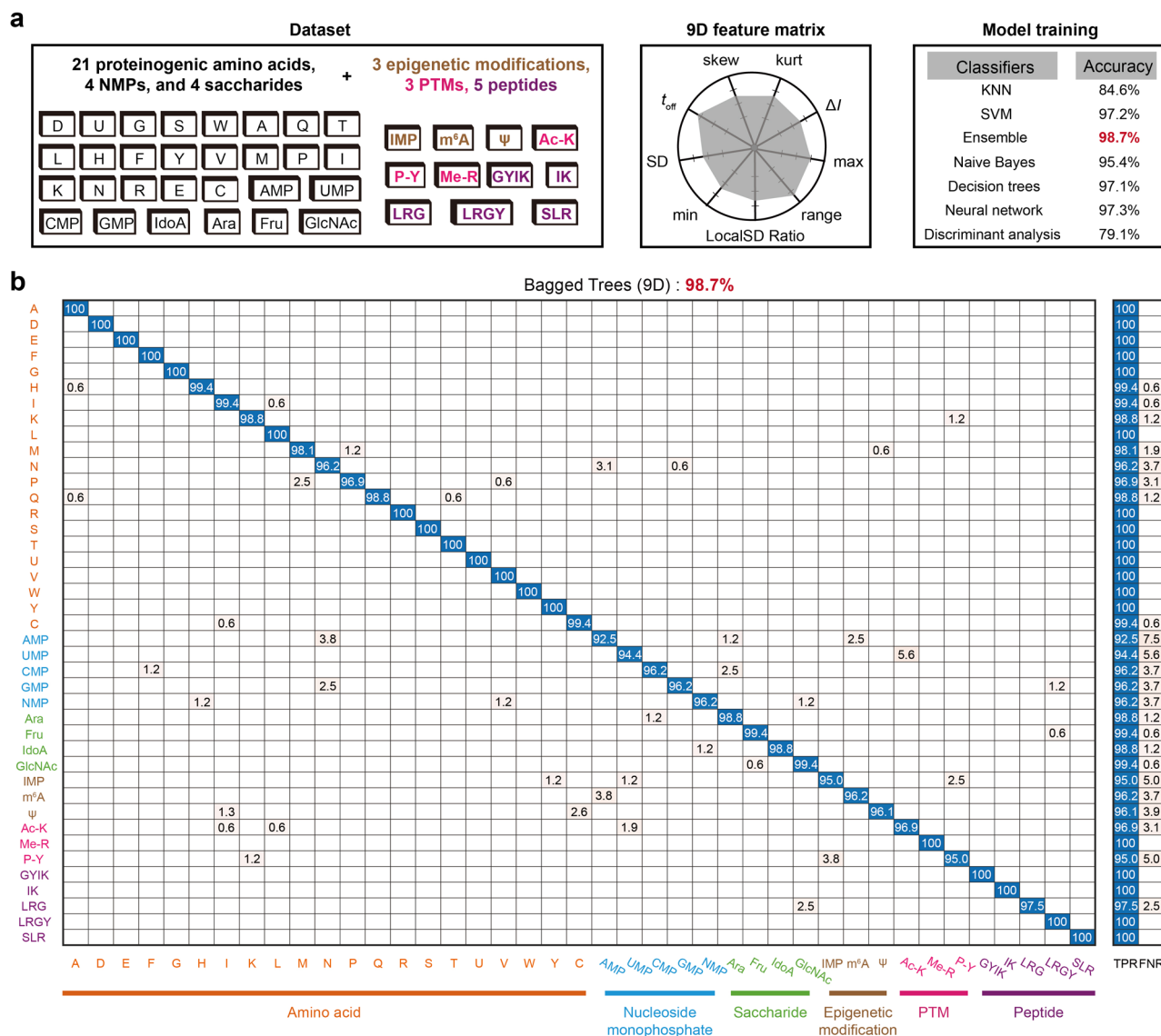
Extended Data Fig. 4 | Type 1 and type 2 events of NMPs. The measurements were carried out as described in Methods. A 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0) was used. A transmembrane voltage of +160 mV was continually applied. **(a–d)** Left: Chemical structures of nucleoside monophosphates. Right: Representative traces containing events caused by AMP **(a)**, UMP **(b)**, CMP **(c)** and GMP **(d)** binding. Each NMP was separately added to *cis* with a final concentration of 2 mM. I_{FPBA} represents the open pore current of MspA-FPBA. Each nucleoside monophosphate generates type 1 events (AMP1, UMP1, CMP1 and GMP1) and type 2 events (AMP2, UMP2,

CMP2 and GMP2). A/U/C/GMP1 + 2 represents the combination of type 1 events and type 2 events. **(e)** The scatter plot of ΔI versus SD generated by type 1 events (marked with a red dashed box) and type 2 events (marked with a gray dashed ellipse) of nucleoside monophosphates. Type 1 events (AMP1, $n = 113$; UMP1, $n = 207$; CMP1, $n = 169$; GMP1, $n = 190$) and type 2 events (AMP2, $n = 125$; UMP2, $n = 437$; CMP2, $n = 388$; GMP2, $n = 178$) were employed to generate the statistics. **(f)** Top: Representative type 1 events of the corresponding nucleoside monophosphates. Bottom: Representative type 2 events of the corresponding nucleoside monophosphates.



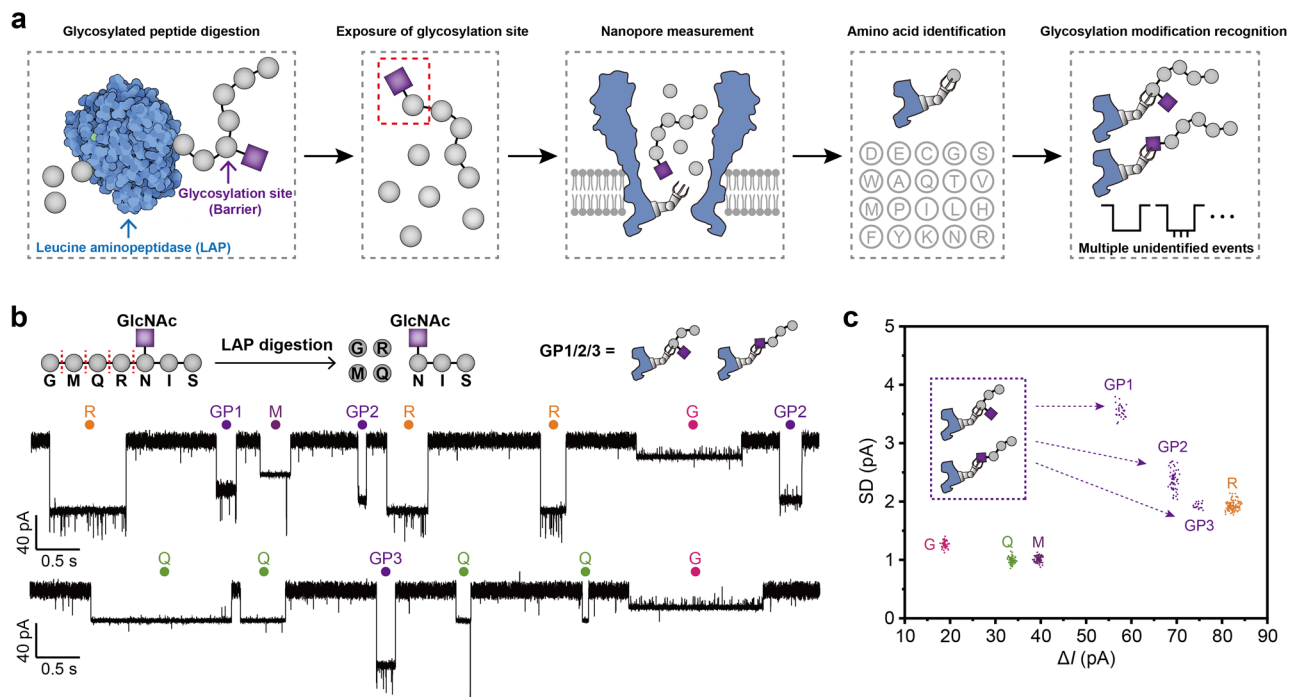
Extended Data Fig. 5 | Identification of amino acids with PTMs and epigenetic NMPs by MspA-FPBA. (a) The chemical structures of amino acids and NMPs with modifications. N6-Acetyl-L-lysine (Ac-K), O-Phospho-L-tyrosine (P-Y), and N3,N4-Dimethyl-L-arginine (Me-R) represent the acetylation, phosphorylation and methylation modifications of amino acids, respectively. Three epigenetic NMPs, namely pseudouridine-5'-monophosphate (ψ), N6-Methyladenosine-5'-monophosphate (m⁶A), and inosine-5'-monophosphate (IMP) were also investigated. (b) Representative traces acquired during MspA-FPBA sensing of Ac-K, P-Y, and Me-R, respectively. Each amino acid was added to the *cis* side,

reaching a final concentration of 400 μM. The I_{FPBA} was marked with a gray dashed line. (c) The scatter plot of ΔI versus SD for the Ac-K, P-Y, and Me-R events obtained as described in b. At least 100 events of each analyte were used to generate the scatter plot. (d) Representative traces obtained during MspA-FPBA sensing of ψ, m⁶A and IMP. Each analyte was added to the *cis* side at a final concentration of 2 mM. I_{FPBA} was marked with a gray dashed line. (e) The scatter plot of ΔI versus SD for the ψ, m⁶A and IMP events obtained from d. At least 100 events of each analyte were used to generate the scatter plot.



Extended Data Fig. 6 | Machine learning identification of 24 amino acids, 7 nucleoside monophosphates, 4 saccharides, and 5 peptides. (a) The workflow of machine learning. Nanopore events acquired with 24 amino acids (21 proteinogenic amino acids and 3 amino acids with post translational modifications), 7 nucleoside monophosphates (4 canonical nucleoside monophosphates and 3 nucleoside monophosphates with epigenetic modifications), 4 saccharides, and 5 peptides (IK, LRG, SLR, GYIK, and LRGY) were collected to form the training dataset. Nine event features, including ΔI , SD, LocalSD Ratio, min, max, range, t_{off} , skew and kurt, were employed to form a

nine-dimensional feature matrix. All models were evaluated with the 10-fold cross-validation accuracies, according to which Ensemble model (bagged trees) has performed the best with a validation accuracy of 98.7% (Supplementary Table 7). **(b)** The confusion matrix for amino acids (orange), nucleoside monophosphates (blue), saccharides (green), nucleoside monophosphates with epigenetic modifications (brown), amino acids with post translational modifications (pink), and peptides (purple) generated by the bagged trees model employing a nine-dimensional feature matrix. The row of the matrix represents the true class and the column represents the predicted class.



Extended Data Fig. 7 | Identification of amino acids and glycosylation modifications by protease digestion of glycosylated peptides. (a) Schematic diagram of the identification of amino acids and glycosylation modifications following the enzymatic cleavage of glycosylated peptides using MspA-FPBA. Leucine aminopeptidase (LAP) was employed to cleave the N-terminal amino acids of glycosylated peptides until reaching the glycosylation site that impeded enzymatic digestion, thereby exposing the glycosylation site. Then, MspA-FPBA was used to identify the generated amino acids, thereby confirming the amino acid components before the glycosylation site. For the short peptide containing the glycosylation site, both the N-terminal amino group and the glycosyl group could be recognized by MspA-FPBA, resulting in multiple types of events. (b) Representative traces of the glycosylated heptapeptide with the sequence GMQRN(GlcNAc)IS following LAP enzymatic digestion

(Methods). Representative events corresponding to different amino acids were automatically predicted via machine learning and labeled with color-coded dots (G: dark red; M: dark purple; Q: light green; R: orange). Three unidentified nanopore events (GP1/2/3) were attributed to the N(GlcNAc)IS peptide generated by enzymatic digestion, with the glycosylation site being exposed. (c) The scatter plot of ΔI versus SD for the amino acid and glycosylated tripeptide events obtained from (b). The scatter plot was generated using a 10 min continually recorded trace. The four clusters of amino acids corresponding to G, M, Q, and R were identified by machine learning. The three distinct clusters corresponding to the N(GlcNAc)IS peptide were respectively labeled as GP1, GP2 and GP3 based on the blockage amplitude and noise amplitude. All measurements (b-c) were conducted under a continuous bias of +160 mV, with 25 μ L of the enzymatic hydrolysate added to the *cis* side.

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 (<i>n</i>) 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 <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☒	☐ 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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ 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 <i>d</i> , Pearson's <i>r</i>), 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 commercially available software Axonpatch 10.7
Data analysis	Data analysis was performed by machine learning. The code and training data used for machine learning model building were provided in the submission system as a supplementary file.

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](#)

A set of demo events were accompanied for code validation. All other data presented in this work can be provided by the corresponding authors upon reasonable requests.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	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 necessary in this study.
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

Methods

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