BIOPHYSICS

Nanopore profiling and structure determination of oligosaccharides by fragmentation

Yunqi Xiao[1,2]†, Shanyu Zhang[1,2]†, Xinmeng Gao[1,2], Tian Li[1,2], Hanhan Zhang[1,2], Panke Zhang[1], Shuo Huang[1,2]*

Oligosaccharides, polymers of covalently linked monosaccharides, are essential biomolecules mediating diverse biological and physiological processes. Despite their critical roles, the structural complexity of many oligosaccharides poses a notable challenge for comprehensive characterization, limiting our understanding of their biological functions. Nanopores as emerging single-molecule sensors offer promise for oligosaccharide structural analysis. Here, we demonstrate direct discrimination of diverse mono- and oligosaccharides using a heterooctameric *Mycobacterium smegmatis* porin A (MspA) nanopore integrated with a single phenylboronic acid (PBA) adapter (MspA-90PBA). The high-resolution event signatures generated by MspA-90PBA enable structural analysis of complex oligosaccharides via hydrolysis into smaller fragments, which can be analyzed for either structure profiling or determination. Using a minimized reference database, we achieve structural determination of raffinose, stachyose, and verbascose through nanopore characterization of its hydrolyzed fragments followed by fragment assembly. This work establishes a framework for nanopore-based oligosaccharide analysis, combining single-molecule sensing with a fragment assembly strategy to address the long-standing challenge of glycan structural elucidation.

INTRODUCTION

Oligosaccharides are biopolymers of monosaccharides sequentially linked together by glycosidic linkages (*1*). Oligosaccharides are widely found in nature, in the form of either isolated entities or conjugates covalently appended to RNAs (*2*), proteins (*3*), lipids (*4*), or secondary metabolites (*5*). Oligosaccharides play pivotal roles in biological activities, including energy storage (*6*), protein secretion (*7*), cell communication (*8*), pathogen invasion (*9*), and immune regulation (*10*). The diverse roles of oligosaccharides are immediately associated with their complex structures. Monosaccharides are linked via a glycosidic bond between the anomeric carbon of one monosaccharide and the hydroxyl group of the other (*11*). Unlike nucleic acids or peptides, which are linear, each monosaccharide unit has multiple glycosidic hydroxyl groups, and the orientation of the glycosidic bond at the anomeric carbon (α versus β) also affects the structure, generating complex and branched oligosaccharide structures (*1*, *11*), which makes it extremely difficult to analyze the structure of oligosaccharides.

Conventionally, the structural elucidation of oligosaccharides can be carried out via nuclear magnetic resonance (NMR) (*12*), mass spectrometry (MS) (*13*), liquid chromatography (LC) (*14*), infrared (IR) spectroscopy (*15*), or combinations of these techniques. NMR spectroscopy can resolve the oligosaccharide structure at atomic resolution, in which different chemical groups of an oligosaccharide can be recognized by the corresponding chemical shifts (*16*). However, this method requires milligrams of analyte, which is usually impossible for oligosaccharide samples from biological resources (*12*). MS, which is more sensitive, is also widely applied in oligosaccharide profiling and sequencing (*17*, *18*). However, single-stage MS provides insufficient structural details for oligosaccharide sequencing and fails to discriminate between oligosaccharide isomers (*19*). When equipped with activation methods such as collision-induced dissociation (*20*) or electron transfer dissociation (*21*), tandem MS (*22*) can provide more information on oligosaccharide branching and linkage information (*23*–*25*). However, tandem MS is low throughput (*26*) and often requires the availability of oligosaccharide profile databases for structural analysis (*27*–*29*). Various LC separation methods coupled with MS have been used to separate and analyze oligosaccharide structures, with an emphasis on oligosaccharide isomer identification (*14*). The LC separation conditions, however, require complex optimization for different types of samples (*30*, *31*). IR spectroscopy, which is rapid, nondestructive and accessible (*32*), is used in the identification of saccharide types, substitution groups, and anomeric carbon configurations (*33*). However, the complexity of IR spectrum interpretation has limited its wide application (*33*). All of the above-described conventional analytical methods require bulky apparatus and complex data interpretation for oligosaccharide sequence elucidation.

Nanopores are emerging single-molecule sensors widely used in the sequencing of DNA (*34*, *35*) and RNA (*36*). Nanopore nucleic acid sequencing offers technical advantages, including, but not limited, to a long read length (*37*), a portable device (*38*), fast speed (*39*), and resolution to resolve a variety of epigenetic modifications (*40*). Nanopore nucleic acid sequencing is achieved by the ratcheting motion of the nucleic acid strand driven by a motor protein (*34*, *35*). Motivated by the success of nanopore nucleic acid sequencing, it is widely anticipated that proteins (*41*) or polysaccharides (*42*) may also be sequenced by nanopores in a similar manner. Despite recent advances in nanopore protein/peptide sequencing (*43*–*45*), possibly owing to the structural complexity of polysaccharides, the developments aimed at nanopore oligosaccharide sequencing clearly lag behind.

With solid-state nanopores, polysaccharides of large molecular weights, such as glycosaminoglycans (*46*–*50*), xylans (*51*), and starch

[1]State Key Laboratory of Analytical Chemistry for Life Sciences, School of Chemistry and Chemical Engineering, Nanjing University, 210023 Nanjing, China. [2]Chemistry and Biomedicine Innovation Center (ChemBIC), ChemBioMed Interdisciplinary Research Center, Nanjing University, 210023 Nanjing, China.
*Corresponding author. Email: shuo.huang@nju.edu.cn
†These authors contributed equally to this work.

(*52*), can be sensed. On the other hand, by observing pore rectification via phenylboronic acid (PBA)–modified solid-state nanopores (*53*–*55*), the presence of monosaccharides such as glucose or fructose can also be detected. Although the sensing of diverse types of saccharides has been performed via solid-state nanopores, solid-state nanopores generally lack sufficient resolution to discriminate between saccharides with subtle sequence differences or saccharides of small sizes such as monosaccharides or low molecular weight oligosaccharides.

Saccharide sensing has also been performed with biological nanopores. For example, when equipped with a saccharide-binding protein (*56*, *57*), a cytolysin A nanopore is immediately turned into a glucose/maltose sensor. Aerolysin (*58*) and mutated α-hemolysin (α-HL) (*59*) can directly sense saccharides with tags (*58*) or specific functional groups (*59*) by saccharide-nanopore interactions. However, these techniques are limited with either a tag molecule that must be appended to the saccharide analyte before measurement (*58*) or few types of suitable saccharides (*59*), demonstrating limited generality. Nanopore discrimination of oligosaccharides of different length was also demonstrated (*60*). However, the nanopore sensor used is only suitable for a very limited range of saccharide sequences.

With a hetero-heptameric α-HL nanopore modified with a sole PBA adapter (*61*), nanopore sensing of D-glucose, D-fructose, and D-maltose was achieved by PBA-diol interaction. However, owing to the limited resolution of α-HL, the simultaneous discrimination of these saccharides was only barely demonstrated, and the sensing of other saccharides has not been reported. Given the high resolution offered by a conical lumen geometry, a hetero-octameric MspA conjugated with a PBA adapter (MspA-90PBA) was reported to be able to discriminate among all major monosaccharides (*62*). In a later study, direct determination of the glycosidic linkages of a variety of disaccharide isomers was demonstrated (*63*). The challenge of unambiguous oligosaccharide identification requires a saccharide-responsive nanopore that simultaneously offers high generality and discrimination resolution, without the need for analyte tagging. To date, MspA-90PBA is the only nanopore that fully meets these requirements, highlighting its suitability as a sensor for nanopore-based oligosaccharide structure analysis.

The workflow is outlined in Fig. 1. First, a collection of monosaccharides and oligosaccharides is analyzed using MspA-90PBA, generating unique nanopore event signatures for each saccharide. These signatures are compiled into a reference database. For an unidentified oligosaccharide, direct measurement with MspA-90PBA is first attempted: If its nanopore events exactly match those of a saccharide in the database, then it is directly identified. If not, then the oligosaccharide is hydrolyzed into smaller fragments, which are then measured by MspA-90PBA. The collective pattern of events from these fragments forms a distinct nanopore profile, a structural "fingerprint" of the parent oligosaccharide. If the fragment identities match entries in the database, then the original oligosaccharide's structure can be reconstructed through fragment assembly, enabling structural determination. Even the unhydrolyzed oligosaccharide generates characteristic nanopore events. Thus, once its structure is elucidated via fragment assembly, its identity and corresponding nanopore event features are incorporated into the database. This continuous integration expands the database's coverage, enabling analysis of increasingly complex oligosaccharides.

RESULTS

Discrimination of oligosaccharides by MspA-90PBA

A hetero-octameric MspA containing a sole cysteine at its pore constriction was first prepared (Materials and Methods). This hetero-octameric MspA, also referred to as $(N90C)_1(M2)_7$, was assembled from one monomer of the N90C mutant and seven monomers of the M2 mutant (*64*, *65*). To append a PBA adapter to the pore construction, $(N90C)_1(M2)_7$ was mixed with 3-(maleimide) PBA (MPBA), generating a PBA-modified $(N90C)_1(M2)_7$, referred to as MspA-90PBA. To evaluate the sensing generality of MspA-90PBA for oligosaccharides, a variety of disaccharides and trisaccharides were respectively sensed (Fig. 2A), following the mechanism that PBA reversibly react with 1,2- or 1,3-*cis*-diol on the saccharide to form five- or six-membered boronate ester (Fig. 2B) (*62*). Nanopore measurements were carried out with a MspA-90PBA in a 1.5 M KCl buffer [1.5M KCl and 100 mM 3-morpholinopropanesulfoinc acid (Mops; pH 7.0)], and a +160-mV potential was continually applied (Materials and Methods). Under these conditions, a stable open pore current (I_0) was reported.

In separate measurements, sucrose (Suc) (20 mM), lactulose (10 mM), lactose (Lac) (10 mM), melibiose (Mel) (10 mM), or galactoseα1-6galactose (G-G) (5 mM) (table S1), which differ in either their monosaccharide subunits or their glycosidic linkages, was respectively added to the cis chamber, immediately after which corresponding nanopore events were observed. To evaluate the results quantitatively, the mean blockage amplitude (ΔI) and the noise amplitude of the blockage level (*SD*) were extracted from all acquired events (fig. S1), according to which all five disaccharides were clearly discriminable. When probed by MspA-90PBA, Suc reported only a single type of event, whereas the other four disaccharides reported multiple event types (Fig. 2C). The saccharide analyte, which contains multiple hydroxyl groups, may have multiple possible configurations when bound to the PBA adapter, thus reporting multiple types of events. In the corresponding event scatter plots of $\Delta I/I_0$ versus *SD*, simultaneous reports of multiple event types appear as clearly discriminable patterns of event distribution (figs. S2 to S6). According to the results of three independent measurements ($N = 3$) (figs. S2 to S6), this observation is consistently reported, demonstrating high data reproducibility.

During nanopore saccharide sensing, noise events, which appear as nonclustering events in the scatter plot, are inevitable. These noise events might result from either impurity in the sample or nonspecific binding of the target saccharide to the PBA adapter. To minimize the interference caused by noise events, a density-based spatial clustering of applications with noise (DBSCAN) algorithm was developed (figs. S2 to S6), with which the noise events can be automatically removed. Generally, all five types of disaccharides report distinct event features and patterns of event distribution. Notably, Lac and Mel, which differ only in their glycosidic linkages, are clearly distinguished by MspA-90PBA. In addition, the pair of lactulose and Lac and the pair of Mel and G-G differ by only a monosaccharide subunit. However, their event features are completely discriminable when probed by MspA-90PBA.

The same measurements were also performed with a set of five trisaccharides (table S2). Under identical experimental conditions, kestose (Kes) (10 mM), maltotriose (Mal) (20 mM), lactosyl fructoside (L-F) (10 mM), 2′-fucosyllactose (2F-L) (20 mM), and

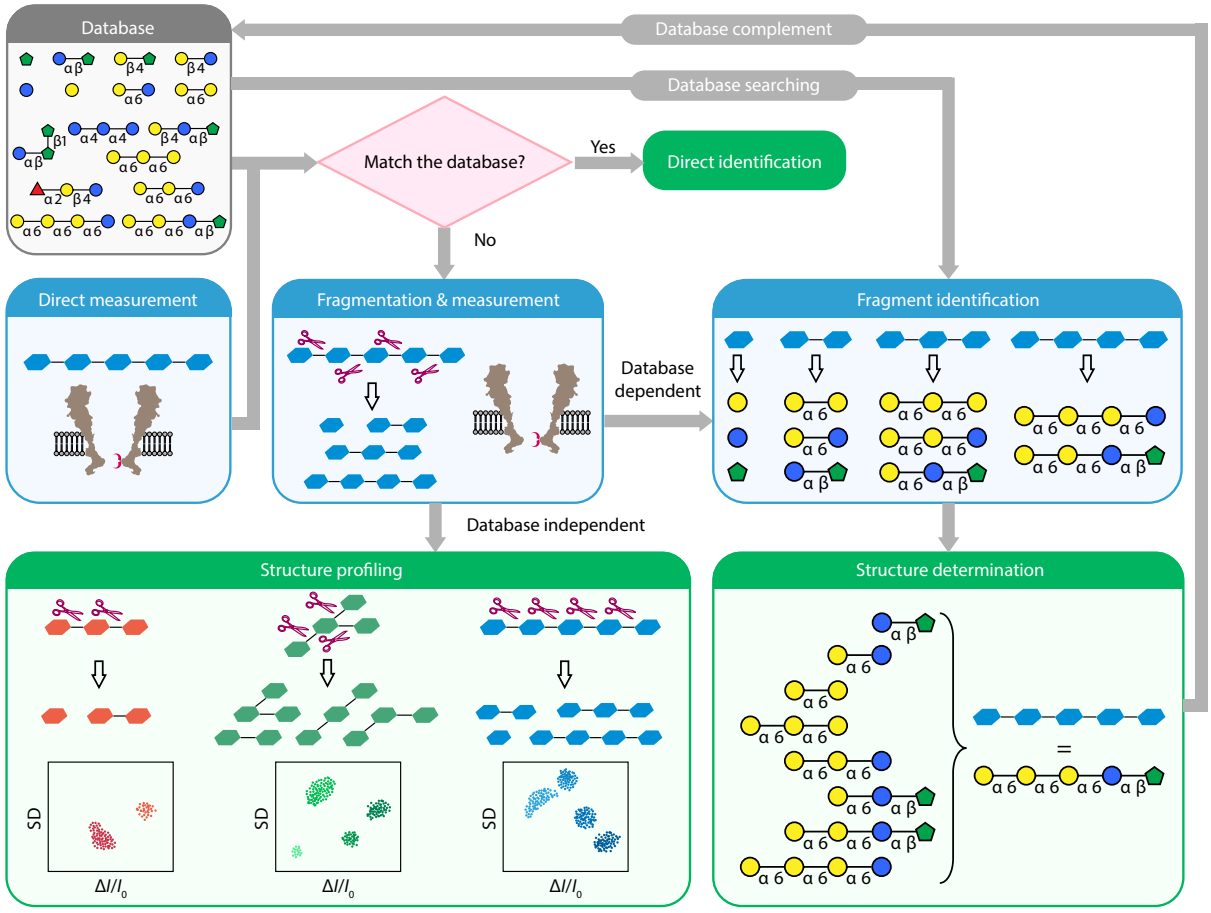

Fig. 1. Schematic of nanopore oligosaccharide analysis using MspA-90PBA. In this framework, saccharides can be analyzed via three sensing modes: direct identification, structural profiling, and structural determination. First, a saccharide database is established containing model saccharides with known identities and their corresponding nanopore event features. For direct identification, if the nanopore events generated by an unknown saccharide exactly match those of a reference saccharide in the database, then its structure is directly confirmed. For structural profiling, the oligosaccharide is hydrolyzed into fragments, which are then measured by MspA-90PBA to obtain structural signatures. This mode operates without relying on preexisting database information. For structural determination, if hydrolyzed fragments match entries in the database, then the oligosaccharide's full structure can be reconstructed through fragment assembly. Newly identified saccharides and their nanopore features are continuously integrated into the database to expand its coverage.

manninotriose (10 mM) were respectively added to the cis chamber in separate measurements. Each type of trisaccharide has a unique pattern of event distribution ($N = 3$), demonstrating clear analyte discrimination (Fig. 2D and figs. S7 to S11). All noise events were removed by DBSCAN. The event differences among Lac, L-F, and 2F-L demonstrate that the addition of a single monosaccharide at the reducing end or nonreducing end of the oligosaccharide notably changes the event features. To this end, the event features of 10 oligosaccharides probed by MspA-90PBA were well acquired. By demonstrating high resolution in the discrimination of oligosaccharides with only subtle structural differences, the ability of MspA-90PBA to sense a variety of oligosaccharides has been confirmed. Generally, the rate of event appearance is linear correlated to the saccharide concentration (fig. S12). The limit of detection can reach μM level (table S3). The high concentration used is to efficiently gain more data for following saccharide database construction. While demonstrated in a limited number of cases, this sensing mode represents the direct identification approach described in Fig. 1.

Construction of the saccharide database by machine learning

To enable accurate identification of saccharides in mixture of hydrolysates, we established a nanopore saccharide database tailored for machine learning–based saccharide identification (Fig. 3A). First, a reference collection was compiled containing 10 oligosaccharides (Fig. 2) and nine monosaccharides (*62*): D-glucose (Glc), D-mannose (Man), D-galactose (Gal), *N*-acetyl-D-galactosamine (GalNAc), D-fructose (Fru), L-sorbose (Sor), L-rhamnose (Rha), D-xylose (Xyl), and D-ribose (Rib). These nanopore events were separately acquired with each of the 19 pure saccharide samples and compiled as standard reference events. From each event, eight features, including $\Delta I/I_0$, *SD* (fig. S1), event dwell time (*time*), $\Delta max/I_0$, $\Delta min/I_0$, $\Delta median/I_0$, interquartile range $(IQR)/I_0$, and $range/I_0$, were extracted (fig. S13). Detailed definitions of these event features are provided in fig. S13. The detailed feature calculation method is described in Materials and Methods.

These event features were collected to form a feature matrix, with which machine learning operations were performed for model building and identity prediction. Briefly, all events recorded in the

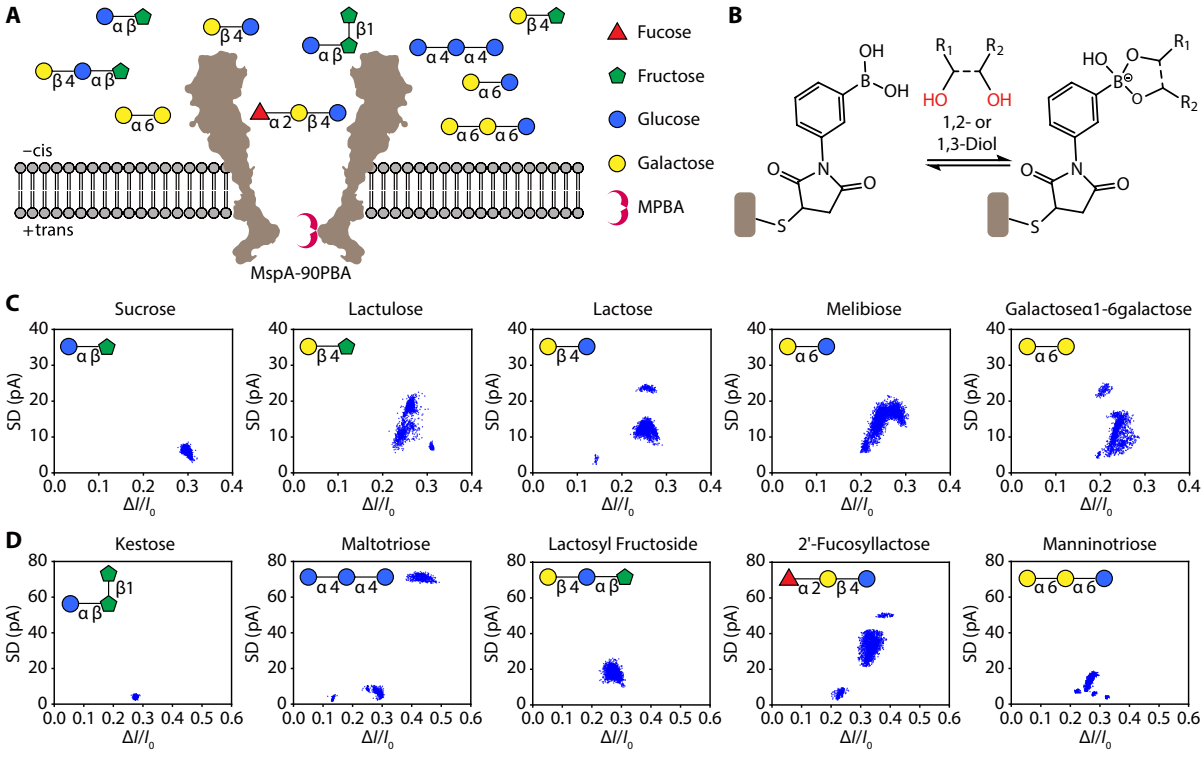

Fig. 2. Oligosaccharide identification by MspA-90PBA. (**A**) Schematic diagram of oligosaccharide identification by MspA-90PBA. MspA-90PBA is generally suitable for the sensing of oligosaccharides with a wide range of lengths and sequences. (**B**) The mechanism of oligosaccharide sensing by MspA-90PBA. Via the reversible formation of a boronate ester between PBA and 1,2- or 1,3-*cis*-diol on the target saccharide, corresponding nanopore events are generated. (**C**) The event scatter plot of $\Delta I/I_0$ versus *SD* of events respectively acquired with five disaccharides, including sucrose (Suc) ($n = 2175$), lactulose ($n = 1514$), lactose (Lac) ($n = 2229$), melibiose (Mel) ($n = 3509$), and galactoseα1-6galactose (G-G) ($n = 1767$). (**D**) The event scatter plot of $\Delta I/I_0$ versus *SD* of events respectively acquired with five trisaccharides, including kestose (Kes) ($n = 1394$), maltotriose (Mal) ($n = 1352$), lactosyl fructoside (L-F) ($n = 2533$), 2′-fucosyllactose (2F-L) ($n = 2597$), and manninotriose ($n = 1600$). Events of each type of saccharide (C and D) are from the results of three independent measurements after cluster analysis treatment (figs. S2 to S11). Experimentally, all the oligosaccharides (C and D) were separately added to the cis chamber. The final concentrations of Suc, Mal, and 2F-L were 20 mM. The final concentrations of lactulose, Lac, Mel, Kes, L-F, and manninotriose were 10 mM. The final concentration of G-G was 5 mM. All nanopore measurements were performed as described in Materials and Methods in a 1.5 M KCl buffer, and a potential of +160 mV was continuously applied.

database were first subjected to noise reduction via DBSCAN. Afterward, 1000 events acquired with each type of saccharide were collected and split into a training set (80%, each saccharide 800 events) and a testing set (20%, each saccharide 200 events) (Materials and Methods). The training set was used for model training and validation, the results of which were applied for model selection. The testing set was used to verify the model performance on the nontraining data. Cross-validation was performed to obtain validation accuracy for model performance evaluation. Among the variety of models being trained, the bagged tree model had the highest validation accuracy of 94.6% (table S4). To evaluate the model robustness, a sensitivity analysis on feature reduction is applied. With the number of features reduced, both validation and testing accuracy decreased slightly (table S5), confirming the suitability of the eight-feature model.

The testing result of the bagged tree model is shown in the confusion matrix (Fig. 3B). The total testing accuracy, precision, recall and F1 score were 94.9, 94.7, 94.9, and 94.8%, respectively (table S6). In addition, the *t*-distributed stochastic neighbor embedding (*t*-SNE) algorithm was applied to reduce the dimensions of eight event features to the visual two-dimensional (2D) space. The events in the training set that belong to different saccharides were well distinguished from each other (Fig. 3C), although the events of these saccharides partially overlapped in the dimensions of $\Delta I/I_0$ versus *SD* (Fig. 2).

Database searching algorithm

When probed by MspA-90PBA, both the reported event features and corresponding event distributions are useful for confirming the presence of a specific type of saccharide analyte in the sample. Technically, with events acquired by MspA-90PBA, each event is first predicted by the previously trained bagged tree model, according to which each event is assigned a predicted class along with a predictive score. A predictive score threshold is set so that events with a low predictive score can be removed. The wrongly predicted events can be excluded by this step.

Among all events, the events assigned with the same predicted class are separately grouped. For each group, the corresponding event distribution (fig. S14 and Materials and Methods) is then compared with the standard event distribution of its predicted class, which is previously recorded in the database. The comparison would reveal a matching degree (fig. S15). A sufficiently high matching degree would confirm the correct prediction of this group of events, and the presence of this specific type of saccharide analyte is thus

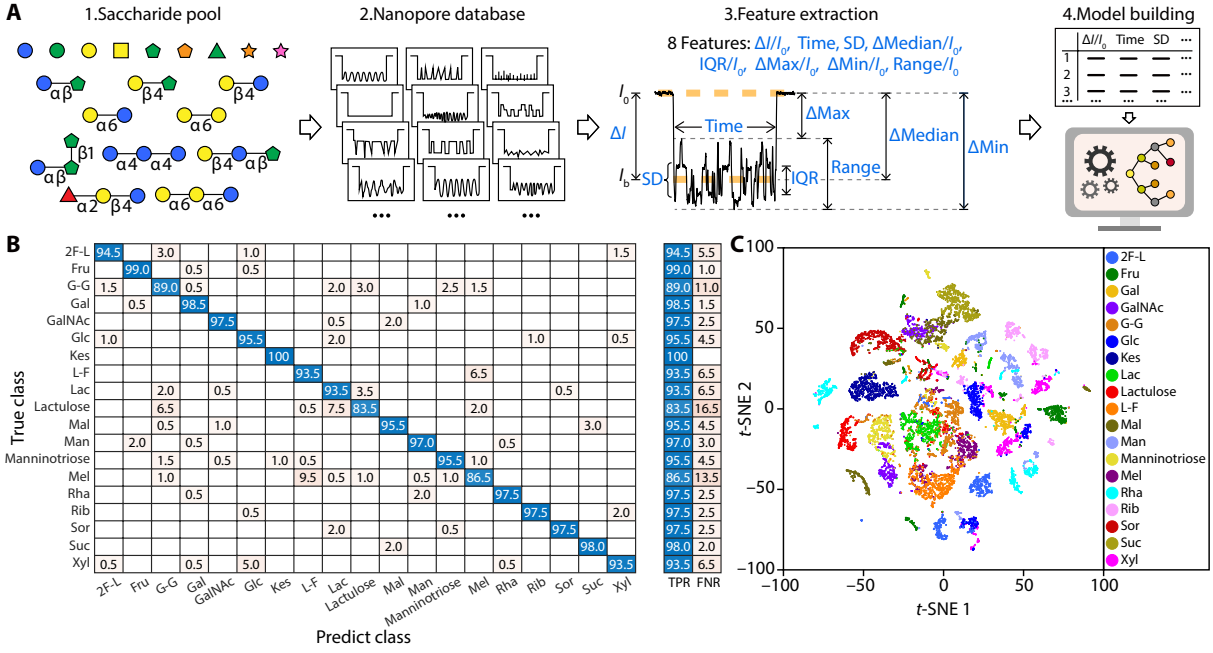

Fig. 3. Machine learning–based saccharide identification. (**A**) Flow diagram of machine learning model building. A saccharide pool containing 10 oligosaccharides (Suc, lactulose, Lac, Mel, G-G, Kes, Mal, L-F, 2F-L, and manninotriose) (Fig. 2) and nine monosaccharides (*62*) (Glc, Man, Gal, GalNAc, Fru, Sor, Rha, Xyl, and Rib) is first established. All 19 standard saccharides are respectively sensed by MspA-90PBA to acquire corresponding event features. For each event, eight features, including $\Delta I/I_0$, *SD*, *time*, $\Delta max/I_0$, $\Delta min/I_0$, $\Delta median/I_0$, IQR/I_0, and $range/I_0$, were extracted (fig. S13), from which the saccharide database is established. One thousand events of each saccharide are selected for model building (800 events) and testing (200 events). According to the validation results, the best performing model is selected for all the prediction tasks. (**B**) Confusion matrix generated by the bagged tree model using the testing set. Here, the bagged tree model is the best-performing model, with a 94.9% testing accuracy. TPR and FNR represent the true and false classification rates for each true class, respectively. (**C**) *t*-SNE plot generated by dimension reduction of eight features to visual 2D space ($n = 15{,}200$), according to which all 19 saccharides are fully separated in the plot.

validated (Fig. 4A). Here, only event distributions with a matching degree over a set threshold are considered correct. Otherwise, this event distribution is considered to be generated by a certain saccharide type that is not previously recorded in the saccharide database.

To validate the above-described algorithm, a mixed oligosaccharide sample containing lactulose, Lac, Mel, 2F-L, and L-F, which differ in their monosaccharide composition, glycosidic linkages, and count of monosaccharide subunits (fig. S16), was used for nanopore measurement. Experimentally, lactulose, Lac, Mel, 2F-L, and L-F were simultaneously added to the cis chamber. The final concentration of lactulose was 2 mM, and the concentrations of Lac, Mel, 2F-L, and L-F were 5 mM for each component. All events acquired with MspA-90PBA were subjected to cluster analysis to remove noise events (fig. S17B). All remaining events were then predicted by the bagged tree model, and some events were wrongly predicted as other saccharides (fig. S17C). To mitigate the interference of incorrect predictions, a predictive score threshold was selected. The selection principle is to minimize the proportion of incorrect events while preserving as many correct events as feasible. Among different predictive score thresholds, a value of 0.8 yields the largest difference between the proportions of correct and incorrect events, consistent with the aforementioned selection principle (fig. S18). Therefore, a predictive score threshold of 0.8 was set for data processing. Only events with a predictive score of ≥0.8 were retained for further analysis. All remaining events, which were mainly predicted to be lactulose, Lac, Mel, 2F-L, and L-F, are shown in the scatter plot of $\Delta I/I_0$ versus *SD* (Fig. 4B). The discrimination of these five oligosaccharides is also shown by dimensionality reduction in the *t*-SNE plot (Fig. 4C).

According to their predicted class, these events were then split into different scatter plots according to the predict labels (fig. S19). The patterns of event distribution corresponding to lactulose, Lac, Mel, 2F-L, and L-F are consistent with those shown in Fig. 2, indicating high measurement consistency. The derivation of matching degree for events of lactulose, Lac, Mel, 2F-L, and L-F to that of the standard events recorded in the database is shown in fig. S20, according to which the matching degrees of lactulose, Lac, Mel, 2F-L, and L-F are 0.84625, 0.8225, 0.9025, 0.905, and 0.9175, respectively. The matching degrees of all the predicted classes are shown in Fig. 4D.

The matching threshold is set sufficiently high to minimize false predictions while preserving valid results. By setting the matching degree threshold to 80%, the predictions of lactulose, Lac, Mel, 2F-L, and L-F were confirmed to be correct. Although some events were predicted and a sufficiently high predictive score was reported, these events failed to report a high matching degree. Therefore, the presence of these saccharide compounds was not validated (Fig. 4D). Matching degree threshold of 80% is used for following analysis. To this end, the database searching algorithm of event prediction followed by matching degree confirmation was confirmed by the mixture of complex saccharide samples.

Moreover, to verify whether this method can determine the saccharide that did not exist in the database, raffinose (Raf) and stachyose (Sta) are used for MspA-90PBA measurement. Raf (20 mM) was

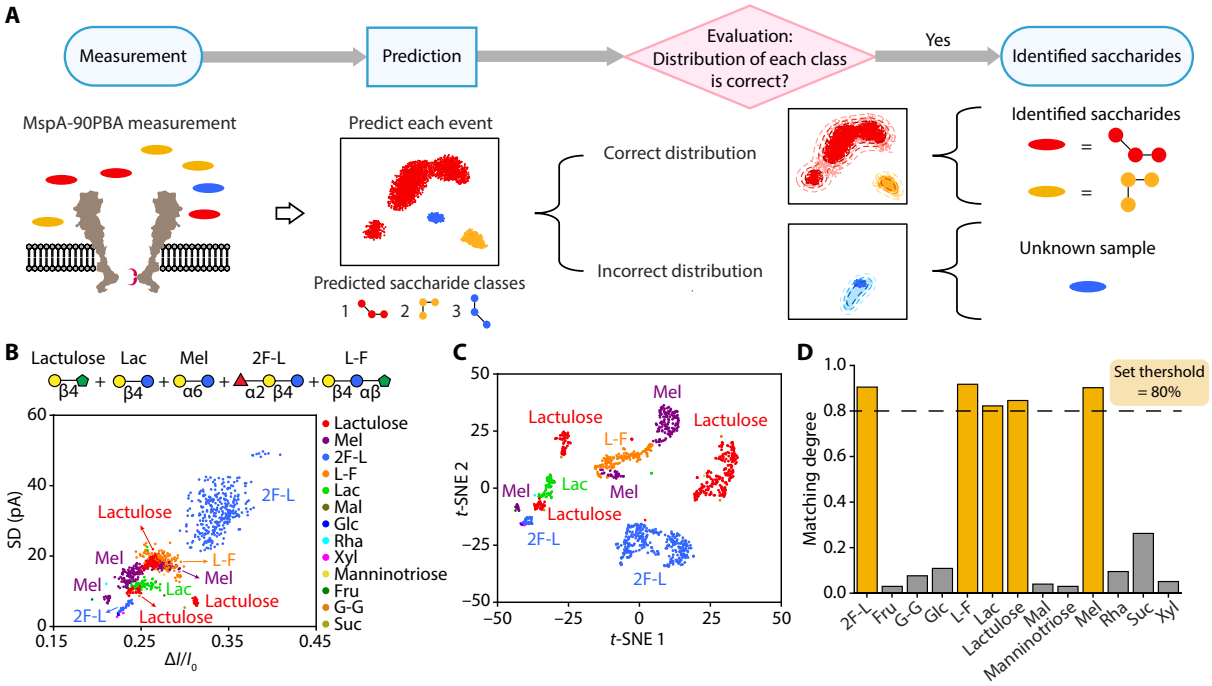

Fig. 4. Event distribution prediction for saccharide events. (**A**) The workflow of nanopore identification of saccharides. A sample containing unknown saccharides is measured by MspA-90PBA, which generated corresponding events. Each event is then predicted by the previously trained bagged tree model (Fig. 3B). Events predicted to be of the same class are then evaluated as an event distribution. By evaluating the matching degree compared with standard event distributions previously obtained with model saccharides (fig. S15 and Materials and Methods), correct and incorrect distributions are identified. Here, the incorrect distributions represent events generated by saccharides not belonging to any standard samples in the database. (**B**) The event scatter plot of $\Delta I/I_0$ versus *SD* of events acquired with a mixture of lactulose, Lac, Mel, 2F-L, and L-F. Only events with a predictive score of ≥0.8 are shown in the plot ($n = 1087$). All the nanopore measurements were performed in a 1.5 M KCl buffer, and a +160-mV potential was continuously applied (Materials and Methods). Lactulose, Lac, Mel, 2F-L, and L-F were simultaneously added to the cis chamber. The final concentration of lactulose was 2 mM, and the final concentrations of Lac, Mel, 2F-L, and L-F were 5 mM. (**C**) *t*-SNE plot generated by dimension reduction of eight features (fig. S13) to visual 2D space for events in (B) ($n = 1087$). The events corresponding to lactulose, Lac, Mel, 2F-L, and L-F are well separated in the plot. (**D**) The matching degrees of the predictions for the results shown in (B). Matching degree threshold is set to 80%. Lactulose, Lac, Mel, 2F-L, and L-F, which report a matching degree ≥80% for the corresponding event distribution, are confirmed to be the correct prediction.

measured three times by MspA-90PBA, and noise events were removed by DBSCAN (fig. S21). Raf reported consistent event patterns in three independent measurements. The events were predicted by the bagged tree model. A large fraction of events received a low predictive score and were thus removed before further analysis. All remaining events reported low matching degrees, confirming that Raf did not equal to any saccharide in the current database (fig. S22). Also, Sta (20 mM) was measured by MspA-90PBA, which generated corresponding nanopore events (fig. S23). The event distribution of Sta failed to match any standard event distribution recorded in the saccharide database (fig. S24), confirming that Sta is not an existing saccharide in the database. The identification mode of Raf and Sta represents the direct identification with database searching approach described in Fig. 1.

Nanopore identification of the base of monosaccharide composition

The above saccharide sensing principle can be applied for identifying the base of monosaccharide composition in an oligosaccharide (Fig. 5A). Briefly, an oligosaccharide that is unidentified can be completely hydrolyzed to monosaccharides by acid. The generated monosaccharide mixture would then be measured by MspA-90PBA. If the standard event features and the corresponding event distributions of all detected monosaccharides were previously recorded in the saccharide database, then they can be directly identified by the matching degree integrated machine learning algorithm via database searching (Fig. 4).

To validate its feasibility, manninotriose, Raf, and Sta were selected as proof-of-concept model saccharides. To ensure that unhydrolyzed Raf and Sta could also be recognized by the bagged tree model, these saccharides were incorporated into the database. Updating the previously trained model yielded a confusion matrix with a total testing accuracy, precision, recall, and F1 score of 95.0, 95.0, 95.0, and 95.0%, respectively (fig. S25, table S6), indicating that Raf and Sta generated distinct events from existing database entries and their inclusion did not compromise the database's sensing performance.

Manninotriose, Raf, and Sta were then fully hydrolyzed into monosaccharides via acid treatment. Experimentally, 80 mM solutions of each oligosaccharide were incubated in 0.5 M trifluoroacetic acid (TFA) at 80°C for 12 hours (Materials and Methods). The reaction was quenched by adjusting the pH to 7, followed by ultrafiltration to remove high–molecular-weight by-products. Using MspA-90PBA, 40 μl of the manninotriose ultrafiltrate was added to the cis chamber to initiate the measurements. Three independent trials ($N = 3$) showed consistent results in scatter plots (fig. S26). Events

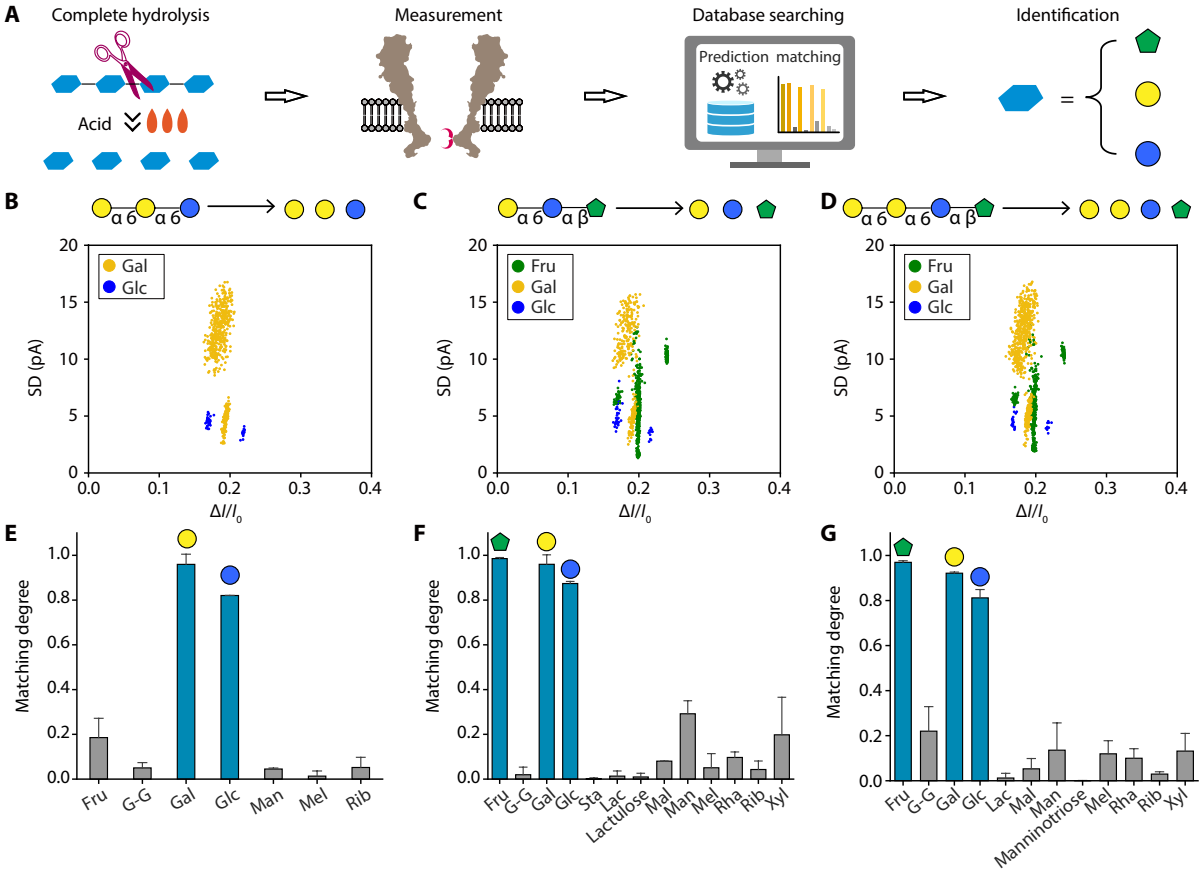

Fig. 5. Monosaccharide analysis by nanopore. (**A**) A flow diagram of nanopore monosaccharide analysis. Completely hydrolyzed by acid, the oligosaccharide would be transformed into monosaccharides (Materials and Methods). The hydrolysates are respectively measured by MspA-90PBA to produce corresponding nanopore events. By database searching (Fig. 4), the base of the monosaccharide composition can be identified. (**B**) The event scatter plot of $\Delta I/I_0$ versus *SD* of events acquired with the acid hydrolysate of manninotriose ($n = 594$). (**C**) The event scatter plot of $\Delta I/I_0$ versus *SD* of events acquired with the acid hydrolysate of Raf ($n = 858$). (**D**) The event scatter plot of $\Delta I/I_0$ versus *SD* of events acquired with the acid hydrolysate of Sta ($n = 1140$). The results in (B) to (D) were predicted by the bagged tree model (fig. S25). Only predicted event distributions with a matching degree of ≥80% are considered to be correct distributions and only events with a predictive score of ≥0.8 were shown. Other events were shown in figs. S27, S29, and S31. (**E**) The matching degree of the saccharides predicted from the acid hydrolysate of manninotriose ($N = 3$). Gal and Glc are recognized as the correct predictions. (**F**) The matching degree of the predicted saccharides of the acid hydrolysate of Raf ($N = 3$). Fru, Gal, and Glc are recognized as the correct predictions. (**G**) The matching degree of the predicted saccharides of the acid hydrolysate of Sta ($N = 3$). Fru, Gal, and Glc are recognized as the correct predictions. The nanopore measurements were performed as described in Materials and Methods in a 1.5 M KCl buffer, and a +160-mV potential was continually applied. Acid hydrolysate (manninotriose, 40 μl; Raf, 20 μl; Sta, and 20 μl) was added to the cis chamber.

were predicted by the bagged tree model and filtered using a predictive score threshold of 0.8 (fig. S27), identifying Gal and Glc in the event scatter plot (Fig. 5B). Gal and Glc exhibited matching degrees of >80%, whereas other saccharides had matching degrees of <20% (Fig. 5E), confirming that manninotriose consists of Gal and Glc.

Similarly, ultrafiltrate of hydrolysis products of Raf and Sta were analyzed by MspA-90PBA (figs. S28 to S31), with the bagged tree model predicting events for Fru, Gal, and Glc (Fig. 5, C and D). Measured event distributions matched standard reference profiles for Fru, Gal, and Glc with matching degrees of >80% (Fig. 5, F and G), while other saccharides showed <40% matching degrees. Both Raf and Sta were determined to contain Fru, Gal, and Glc as monosaccharide components. To this end, the method's versatility in identifying monosaccharide composites of oligosaccharide has been well confirmed.

To verify whether the monosaccharide results were influenced by potential side reactions or degradation during acid hydrolysis, a simulated sample, formulated to match the monosaccharide ratio of Raf (Gal:Glc:Fru = 1:1:1), was measured using MspA-90PBA. The predicted distributions of Gal, Glc, and Fru, as well as the final matching degree of the simulated sample, were consistent with the results obtained from the acid-hydrolyzed Raf (fig. S32). This suggests that potential side reactions or degradation caused by acid hydrolysis did not notably interfere with the identified monosaccharide information.

Oligosaccharide structure determination by MspA-90PBA

In principle, oligosaccharides can be partially hydrolyzed into saccharide fragments using specific glycosidases. Elucidating the entire oligosaccharide structure involves identifying and assembling fragments containing characteristic glycosidic linkages. Here, Raf and Sta were used to demonstrate oligosaccharide structure determination via fragment assembly. Guided by their known base of monosaccharide compositions, glycosidases targeting Fru, Gal, and Glc

were selected for partial hydrolysis (Materials and Methods). Experimentally, 10 μl of the α-galactosidase–hydrolyzed Raf ultrafiltrate was added to the cis chamber for MspA-90PBA measurements (fig. S33). Events generated were analyzed using the pretrained bagged tree model, identifying Gal, Suc, and unhydrolyzed Raf (Fig. 6A and fig. S34). Matching degrees for Gal, Suc, and Raf all exceeded 80%, corroborating the identification (Fig. 6E). This result indicates partial hydrolysis of Raf into Gal and Suc fragments. The identified unhydrolyzed Raf suggests that the glycosidase hydrolysis was incomplete. However, these events were ignored and will not interfere with downstream fragment assembly operations.

Conversely, β-D-fructofuranosidase–hydrolyzed Raf ultrafiltrate (10 μl) yielded events identified as Fru and Mel via machine learning and matching degree screening (Fig. 6, B and F, and figs. S35 and S36). No Raf clusters were detected, suggesting complete hydrolysis of Raf. In contrast, β-galactosidase–, α-glucosidase–, and β-glucosidase–treated Raf measurements only detected unhydrolyzed Raf (figs. S37 to S42), implying that these enzymes could not hydrolyze Raf, likely hindered by glycosidic bond orientation or nonterminal Glc residues.

For Sta, 10 μl of α-galactosidase–hydrolyzed Sta ultrafiltrate was analyzed by MspA-90PBA (fig. S43), with the bagged tree model identifying four major saccharides: Gal, Raf, Sta, and Suc (Fig. 6C). Matching degree evaluations confirmed these results (Fig. 6G and fig. S44), indicating that Sta was hydrolyzed into Gal and Raf, with some Raf further degraded into Gal and Suc. Additionally, β-D-fructofuranosidase–hydrolyzed Sta ultrafiltrate (10 μl) yielded events predicted as Fru and manninotriose (Fig. 6D), supported by matching degree analysis (Fig. 6H and figs. S45 and S46), suggesting that β-D-fructofuranosidase cleaved Sta into Gal and manninotriose. Similar to Raf, Sta remained unhydrolyzed by β-galactosidase, α-glucosidase, and β-glucosidase (figs. S47 to S52).

Collectively, these results indicate that hydrolysis of Raf yields Mel (Galα-6Glc) and Suc (Glcα-βFru) fragments. Structural reconstruction assembled the Raf sequence as Galα-6Glcα-βFru (Fig. 6I). For Sta, hydrolysis produced Raf (Galα-6Glcα-βFru), Suc (Glcα-βFru), and manninotriose (Galα-6Galα-6Glc) fragments, enabling the full Sta sequence to be determined as Galα-6Galα-6Glcα-βFru via fragment assembly (Fig. 6J). Notably, the structural reconstruction of Raf (Fig. 6I) revealed the nanopore event features of Raf, which

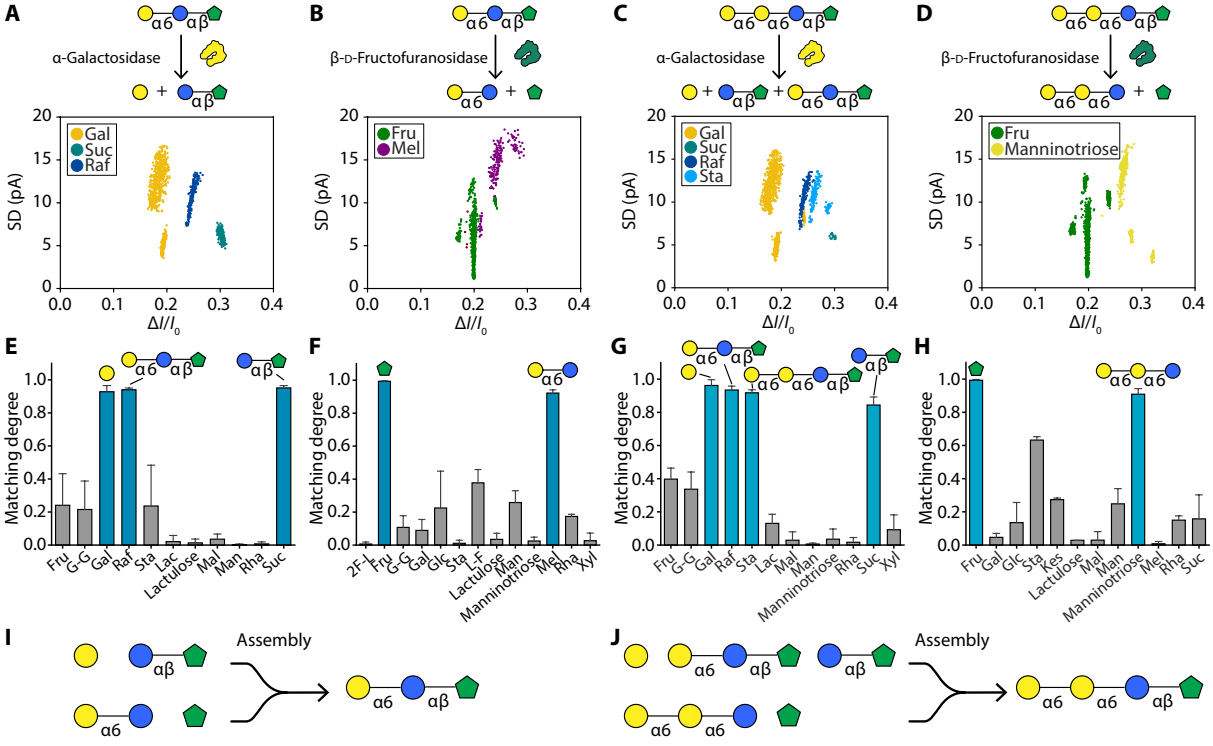

Fig. 6. Nanopore structural determination of oligosaccharides. (**A**) Event scatter plot of $\Delta I/I_0$ versus *SD* for events acquired with the α-galactosidase hydrolysate of Raf ($n = 1083$). (**B**) Event scatter plot of $\Delta I/I_0$ versus *SD* for events acquired with the β-D-fructofuranosidase hydrolysate of Raf ($n = 1379$). (**C**) Event scatter plot of $\Delta I/I_0$ versus *SD* for events acquired with the α-galactosidase hydrolysate of Sta ($n = 1160$). (**D**) Event scatter plot of $\Delta I/I_0$ versus *SD* for events acquired with the β-D-fructofuranosidase hydrolysate of Sta ($n = 1796$). The results in (A) to (D) were analyzed using a bagged tree model (fig. S25). Only events with a matching degree of ≥80% and a predictive score of ≥0.8 are displayed. Additional events are presented in figs. S34, S36, S44, and S46. (**E**) Matching degrees for predicted saccharides in the α-galactosidase hydrolysate of Raf ($N = 3$), with Gal, Raf, and Suc identified as correct predictions. (**F**) Matching degrees for predicted saccharides in the β-D-fructofuranosidase hydrolysate of Raf ($N = 3$), with Fru and Mel identified as correct predictions. (**G**) Matching degrees for predicted saccharides in the α-galactosidase hydrolysate of Sta ($N = 3$), with Gal, Raf, Sta, and Suc identified as correct predictions. (**H**) Matching degrees for predicted saccharides in the β-D-fructofuranosidase hydrolysate of Sta ($N = 3$), with Fru and manninotriose identified as correct predictions. (**I**) Structure assembly of Raf from the identified fragments, according to which the sequence of Raf was determined to be Galα-6Glcα-βFru. (**J**) Structure assembly of Sta from the fragments, according to which the sequence of Sta was determined to be Galα-6Galα-6Glcα-βFru. Experimentally, all nanopore measurements were conducted in 1.5 M KCl buffer with the samples added to cis chamber under a constant applied potential of +160 mV (Materials and Methods).

were critical for determining Sta's structure. Because Sta's hydrolysate contains Raf, this newly acquired knowledge facilitated Sta's structural elucidation. These findings demonstrate how database expansion enhances sensing capacity, enabling analysis of increasingly complex oligosaccharides.

Compatibility with more complex oligosaccharides

Because branched and larger glycans play a central role in glycobiology, the compatibility of MspA-90PBA and the saccharide database with more complex oligosaccharides is of great importance. Blood type tetrasaccharide A type 6 (A tetra 6) and blood type tetrasaccharide B type 6 (B tetra 6) are branched structures derived from 2F-L (Fig. 7A). In separate measurements, A tetra 6 (2 mM) and B tetra 6 (2 mM) were respectively introduced into the cis chamber for MspA-90PBA measurements. Despite their subtle structural difference, the scatter plots of A tetra 6 and B tetra 6 exhibit distinct distributions ($N = 3$) (Fig. 7, B and C, and figs. S53 and S54). These results demonstrate that MspA-90PBA is capable of discrimination of branched oligosaccharides with subtle structural variations. Verbascose (Ver) is a pentasaccharide that contains one additional Gal compared to Sta (Fig. 7D). Ver (20 mM) was added to the cis chamber for MspA-90PBA measurements ($N = 3$) (fig. S55). Ver produces three distinct event clusters (Fig. 7E), which are highly different from those of Sta (fig. S23).

To quantitatively assess the resolution of branched and large oligosaccharides, event data for A tetra 6 and B tetra 6 were incorporated into the saccharide database to support machine learning–based evaluation. With the inclusion of these three saccharides, the bagged tree model achieved an overall test accuracy of 94.1%, demonstrating no notable interference to the model performance (Fig. 7F). These results confirm the compatibility with and resolution of MspA-90PBA for branched and larger saccharides, thereby further supporting the development of a larger and more comprehensive saccharide database. To further demonstrate structure determination of more complex oligosaccharides, we applied the expanded database for Ver determination. Briefly, Ver was hydrolyzed by α-galactosidase and β-D-fructofuranosidase. The generation of Sta, verbascotetraose, Raf, Suc, Gal, and Fru fragments has confirmed its hydrolysis, and the structure of Ver was determined as Galα-6Galα-6Galα-6Glcα-βFru through fragment assembly (fig. S56).

Nanopore profiling of oligosaccharides by MspA-90PBA

As a database independent approach, the oligosaccharide can be hydrolyzed into smaller fragments and analyzed by MspA-90PBA to generate its unique structural profile (Fig. 8A), a mode termed structural profiling (Fig. 1). To demonstrate this approach, four model oligosaccharides, including Mal, manninotriose, 2F-L, and Ver, which differ in monosaccharide composition and molecular size, were used to generate structural profiles. Experimentally, Mal (80 mM), manninotriose (20 mM), 2F-L (80 mM), and Ver (20 mM) were each hydrolyzed with 0.5 M TFA at 80°C for 4 hours (see Materials and Methods). The ultrafiltered hydrolysate (40 μl) was added

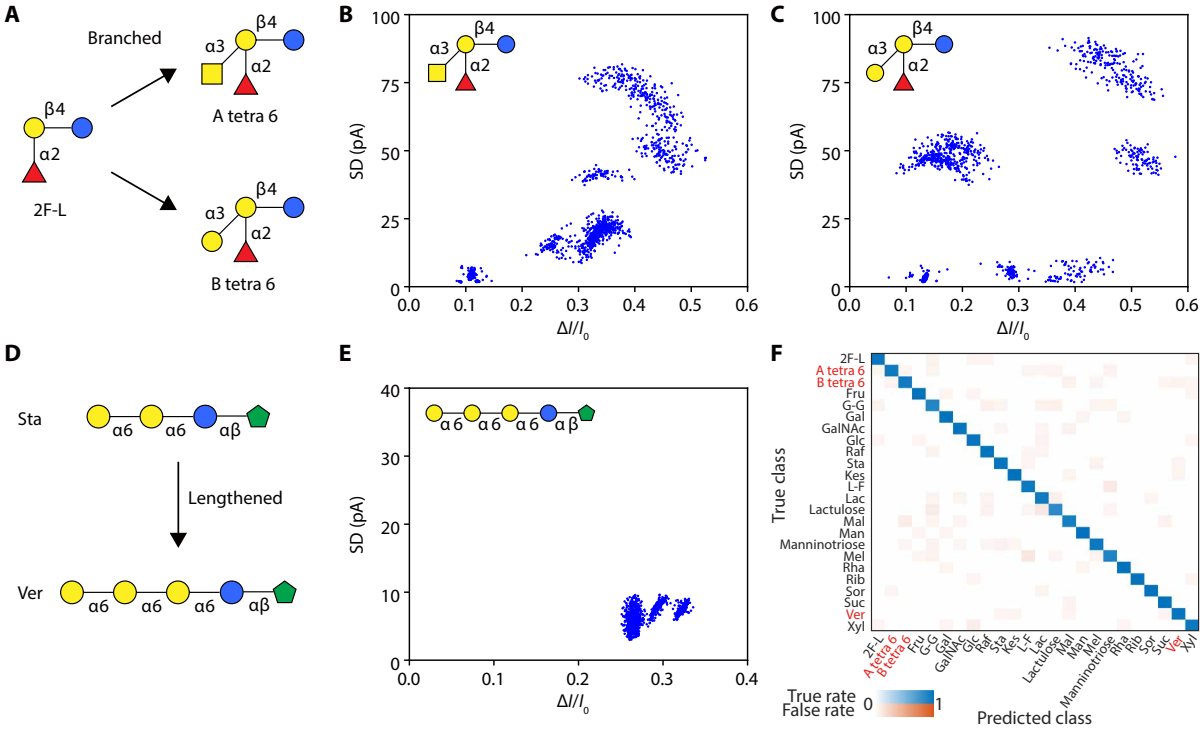

Fig. 7. Nanopore identification of branched and larger oligosaccharides. (**A**) Structural relationship of branched A tetra 6, B tetra 6, and 2F-L. (**B** and **C**) Event scatter plot of $\Delta I/I_0$ versus SD of events acquired with A tetra 6 (n = 1663) and B tetra 6 (n = 1090). (**D**) Structural relationship of Ver and Sta. (**E**) Event scatter plot of $\Delta I/I_0$ versus *SD* of events acquired with Ver (n = 1443). (**F**) Confusion matrix generated by the enlarged bagged tree model using the testing set. A tetra 6, B tetra 6, and 2F-L are further added to the saccharide database. Eight hundred events of each saccharide were used for model training. Two hundred events of each saccharide were used for model testing. The testing accuracy is 94.1%. All nanopore measurements were performed in 1.5 M KCl buffer under a constant applied potential of +160 mV (Materials and Methods). The final concentrations of A tetra 6 and B tetra 6 were 2 mM. The final concentration of Ver was 20 mM.

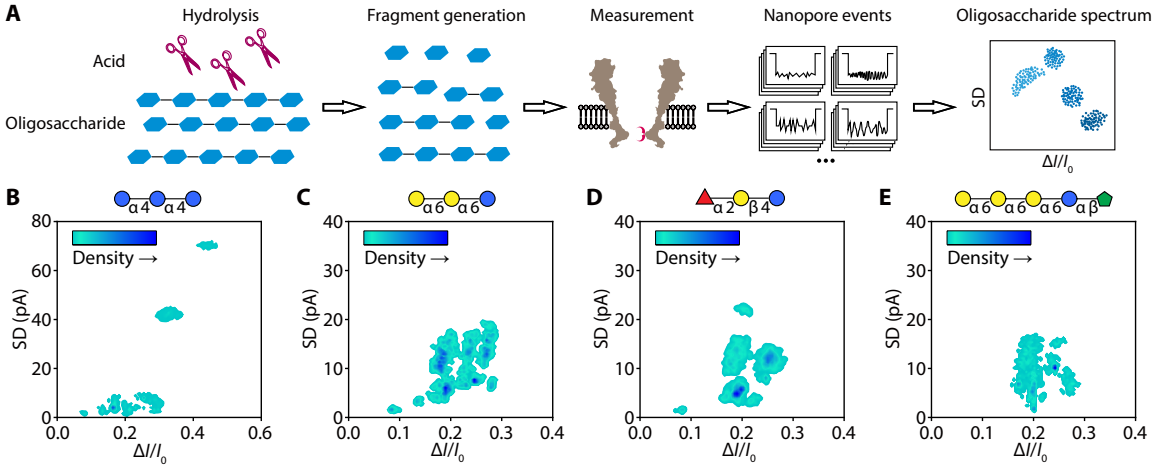

Fig. 8. Nanopore oligosaccharide profiling. (**A**) Flow diagram of nanopore oligosaccharide profiling using MspA-90PBA. Oligosaccharides are first acid-hydrolyzed to generate saccharide fragments. To characterize structural differences between oligosaccharides, MspA-90PBA is applied to characterize these fragments, producing distinct nanopore events that discriminate between different structures. The resulting profiles, which originate from different parent oligosaccharides, are reflected in a kernel density plot of these events. (**B** to **E**) Nanopore profiles of Mal, manninotriose, 2F-L, and Ver. Oligosaccharides with different compositions and chain lengths yield distinct profiles as kernel density plots. Experimentally, these four oligosaccharides were each hydrolyzed with 0.5 M TFA for 4 hours. In separate measurements, 40 μl of each hydrolysate was added to the cis chamber to initiate the measurement. All nanopore measurements were performed in 1.5 M KCl buffer under a constant applied potential of +160 mV (Materials and Methods).

to the cis chamber for nanopore measurements. Each oligosaccharide yielded distinct nanopore events with unique distributions on scatter plots (fig. S57). Unclustered noise events were filtered out using the DBSCAN algorithm, and the remaining events were visualized as kernel density plots (Fig. 8, B to E). Each oligosaccharide exhibited a distinct pattern in these plots, highlighting the specificity of structural profiling.

The nanopore structural profiling mode for oligosaccharides operates independently of a reference database. Even structurally similar oligosaccharides, those that might generate indistinguishable nanopore events when measured intact, can be uniquely identified through analysis of their hydrolyzed fragment mixtures, as the combined event patterns provide more distinct structural signatures. While demonstrated here using acid hydrolysis, oligosaccharide structural profiling can also be achieved through enzymatic hydrolysis with individual or mixed glycosidases.

DISCUSSION

A hetero-octameric MspA nanopore integrated with a single PBA adapter (MspA-90PBA), exhibiting high discrimination capability toward diverse mono- and oligosaccharides, serves as a robust sensor for nanopore-based oligosaccharide direct identification, structural determination and profiling (fig. S58). A series of oligosaccharide can be directly measured by MspA-90PBA and were used to construct a reference database. If the structure of the unhydrolyzed oligosaccharide can be covered by fragments in the database, then the full structure can be determined via hydrolyzed fragment assembly, a mode termed structural determination. To demonstrate this workflow, we constructed a minimized reference database containing 19 characterized saccharides. A database-searching algorithm was developed to identify saccharides in complex samples, enabling MspA-90PBA to characterize the base of the monosaccharide compositions and oligosaccharide fragments in hydrolysates (fig. S59). This approach successfully discriminated oligosaccharides varying in monosaccharide types and chain lengths. As a proof of concept, we demonstrated fragment assembly and structural determination for Raf. Critically, newly acquired nanopore signatures of previously unknown oligosaccharides can be integrated into the reference database, expanding its capacity to analyze increasingly complex oligosaccharides, exemplified by using Raf's nanopore signature to determine the structure of Sta. While MspA-90PBA can identify many intact oligosaccharides via its direct identification mode or fragmented oligosaccharides via its structural determination mode, oligosaccharides generating indistinguishable event features or not covered by the database may require characterization through fragmentation and structural profiling which is free of database. To exemplify this, four model oligosaccharides were acid hydrolyzed and analyzed by MspA-90PBA, yielding distinct nanopore event signatures correlated with their fragment mixtures, enabling straightforward identification of the four oligosaccharides. Notably, structural profiling does not require prior knowledge of the nanopore signatures of hydrolyzed fragments.

The saccharide detection system based on MspA-90PBA exhibits exceptional resolution, generating distinct event signatures for isomers and structurally similar compounds. To date, nanopore technologies have not yet established a systematic saccharide analysis framework comparable to MS glycomics. Compared with LC methods, which often require optimized separation conditions for different analytes, nanopores can discriminate diverse saccharides directly and simultaneously. Unlike common spectral techniques such as NMR and IR, nanopore offers a single-molecule resolution. Our nanopore method can fill a critical gap in the glycomics workflow. Nanopore glycan analysis offers a cost-effective, miniaturized, and portable platform, enabling direct and rapid analysis of complex mixtures with high resolution without the need for prior separation (table S7). Furthermore, it provides exceptional resolution for

isomer discrimination, making it highly valuable for distinguishing similar glycan structures. Although this technology is still in its early stages of development, its complementary nature with other analytical approaches and ongoing advancements underscore its great potential. In prospects, through the integration of nanopore arrays or reduction of measurement volume, it is envisioned that the measurement throughput and sensitivity can be notably improved. With a detection limit in the micromolar range, reducing the measurement volume to the microliter scale can enhance the sensitivity to the picomolar to nanomolar level, which is comparable to that of LC/MS methods.

In addition, the nanopore platform compatibility of branched and large saccharides has been confirmed, which showed the potential of developing a larger and more comprehensive database. While the current database size is limited, largely due to the inaccessibility of diverse model oligosaccharides to academic groups like ours at this stage, the examples above robustly validate the concept of nanopore-based oligosaccharide characterization. For database expansion, commercially available, biologically derived, and chemically synthesized oligosaccharides can be measured using MspA-90PBA, with the results subsequently updated into the database. When further combined with a panel of highly specific exoglycosidases, a ladder of well-defined, sequential digestion products can also be generated. By analyzing each digestion product, we can systematically and efficiently populate the database with the nanopore signatures of a series of structurally related glycans, creating a hierarchically organized library. If feasible, we can integrate the nanopore glycan database with existing glycoinformatic resources, leveraging established enzymatic and bioinformatic tools. Annotating our reference saccharides and data with accession numbers or codes would enable direct cross-referencing with major community databases. We also envision that our current results could contribute to a community-maintained and continuously updated nanopore glycan database, seeded from the current database. With sufficient data accumulation and artificial intelligence integration, de novo sequencing of unknown glycans could eventually be achieved, paralleling advanced tandem MS strategies (*27*) but offering unique advantages, including structural isomer resolution, single-molecule sensitivity, label-free analysis, and portability.

MATERIALS AND METHODS

Materials

Pentane, hexadecane, tris(2-carboxyethyl) phosphine hydrochloride (TCEP), EDTA, Genapol X-80, ammonium persulfate, SDS, *N*,*N*,*N*′,*N*′-tetramethylethylenediamine, and 30% acrylamide/bisacrylamide solution were from Sigma-Aldrich (USA). 1,2-Diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPC) was from Avanti Polar Lipids (USA). Potassium chloride (KCl), sodium chloride (NaCl), sodium hydroxide, potassium hydroxide (KOH), Mops, dimethyl sulfoxide, sodium hydrogen phosphate (Na_2HPO_4), sodium acetate (CH_3COONa), TFA, Coomassie brilliant blue R250, methanol, acetic acid, *n*-dodecyl-β-D-maltoside (DDM), D-(+)-Raf pentahydrate, D-(+)-Mel monohydrate, Sta hydrate, Mal hydrate, Ver, and α-galactosidase (2000 U/g) were from Aladdin (China). Sodium dihydrogen phosphate (NaH_2PO_4) and potassium dihydrogen phosphate (KH_2PO_4) were from Sinopharm (China). Hepes, Suc, lactulose, 2F-L, and α-glucosidase (50 U/mg) were from Macklin (China). Lac, manninotriose, β-glucosidase (12 U/mg), and β-galactosidase (110 U/mg) were from Shanghai Yuanye Bio-Technology (China). G-G (6-O-α-D-galactopyranosyl-D-galactopyranose) was from SHANGHAI ZZBIO Co. Ltd. Kes was from Meryer (China). Blood type A tetra 6 and blood type B tetra 6 were from Wuhan GLYCOGENE Pharmaceutical Co. Ltd. (China). Lactosyl fructoside was from Glycarbo. β-D-Fructofuranosidase (invertase) (300 U/mg) was from Sigma-Aldrich (USA). Dioxane-free isopropyl-β-D-thiogalactopyranoside (IPTG), ampicillin, imidazole, and tris (hydroxymethyl)aminomethane (tris) were from Solarbio (China). SDS–polyacrylamide gel electrophoresis (PAGE) sample loading buffer and SDS-PAGE electrophoresis buffer powder were from Beyotime (China). Precision Plus Protein Dual color Standards, TGX FastCastTM Acylamide Kit (4 to 15%), stacking gel buffer (0.5 M tris-HCl buffer, pH 6.8), and resolving gel buffer (1.5 M tris-HCl buffer, pH 8.8) were from Bio-Rad (USA). Luria-Bertani (LB) broth and LB agar were from Hopebio (China). *Escherichia coli* BL21(DE3) pLysS competent cells and chloramphenicol were from Sangon Biotech (China). MPBA was from Santa Cruz Biotechnology (Shanghai, China).

KCl buffer [1.5 M; 1.5 M KCl and 100 mM Mops (pH 7.0)], lysis buffer [100 mM Na_2HPO_4/NaH_2PO_4, 0.1 mM EDTA, 150 mM NaCl, and 0.5% (v/v) Genapol X-80 (pH 6.5)], buffer A [0.5 M NaCl, 20 mM Hepes, 5 mM imidazole, and 0.5% (v/v) Genapol X-80 (pH 8.0)], and buffer B [0.5 M NaCl, 20 mM Hepes, 500 mM imidazole, and 0.5% (v/v) Genapol X-80 (pH 8.0)] were prepared with Milli-Q water and membrane (0.2 μm, Whatman, UK) filtered before use.

Preparation of nanopore

The hetero-octameric MspA, which is also referred to as $(N90C)_1(M2)_7$, is composed of one monomer of N90C MspA-H6 and seven monomers of M2 MspA-D16H6 (*64*, *65*). To prepare for $(N90C)_1(M2)_7$, the genes respectively coding for N90C MspA-H6 and M2 MspA-D16H6 were custom synthesized and simultaneously inserted in the same coexpression vector pETDuet-1. A hexahistidine tag (H6) was added to the C terminus of both genes to assist nickel affinity chromatography purification. Besides, a 16 aspartate tag (D16) was added to the C terminus of the M2 MspA gene to assist the separation of different pore assemblies during gel electrophoresis.

Experimentally, 100 ng of plasmid was added to 100 μl of *E. coli* BL21 (DE3) pLysS competent cells, and the mixture was incubated on ice for 30 min. After heat shock transformation at 42°C for 90 s, the mixture was incubated on ice for another 3 min. Then, 800 μl of LB medium was added to the mixture and cultured at 37°C and 175 rpm for 50 min. Subsequently, the medium was spread onto an LB agar plate containing ampicillin (50 μg/ml) and chloramphenicol (34 μg/ml) and cultured at 37°C for 18 hours. Afterward, a single colony was taken and added to a 10 ml of LB broth containing ampicillin (50 μg/ml) and chloramphenicol (34 μg/ml). The mixture was cultured at 37°C and 175 rpm until optical density at 600 nm (OD_{600}) reached 0.6 to 0.8. The mixture was then added to 1 liter of LB broth containing ampicillin (50 μg/ml) and chloramphenicol (34 μg/ml). The mixture was shaken at 37°C and 175 rpm until OD_{600} reached 0.6 to 0.8. To induce protein expression, IPTG was then added to the medium with a final concentration of 0.1 mM. The medium was shaken at 16°C and 175 rpm for 24 hours. Last, the medium was centrifuged (4500 rpm, 20 min, 4°C) to collect the bacterial pellet.

The bacterial pellet was resuspended in a 160-ml lysis buffer [100 mM Na_2HPO_4/NaH_2PO_4, 0.1 mM EDTA, 150 mM NaCl, and 0.5% (v/v)

Genapol X-80 (pH 6.5)] and heated at 65°C for 50 min. After cooling to room temperature, the suspension was centrifuged at 13,000 rpm for 1 hour at 4°C. Then, the supernatant was filtered by a 0.2-μm syringe filter (Pall, USA), and the filtrate was loaded to a HisTrap HP nickel affinity column (GE Healthcare, Sweden). The loaded column was first eluted with buffer A [0.5 M NaCl, 20 mM Hepes, 5 mM imidazole, 2 mM TCEP, and 0.5% (v/v) Genapol X-80 (pH 8.0)] until the ultraviolet absorbance readout was stabilized. It was then eluted using a linear gradient of buffer B [0.5 M NaCl, 20 mM Hepes, 500 mM imidazole, 2 mM TCEP, and 0.5% (v/v) Genapol X-80 (pH 8.0)]. The elution fractions were collected and characterized by SDS-PAGE. The fractions corresponding to different assemblies of MspA were collected for further purification.

Purification of $(N90C)_1(M2)_7$ was performed on a 10% SDS-PAGE. Electrophoresis was continually run for 16 hours with a +160-V applied voltage. The gel was then stained by Coomassie brilliant blue (1.25 g of Coomassie brilliant blue R250, 225 ml of methanol, 50 ml of acetic acid, and 225 ml of ultrapure water) for 4 hours. Subsequently, it was decolorized by the destaining buffer (400 ml of methanol, 100 ml of glacial acetic acid, and 500 ml of ultrapure water). The protein band corresponding to $(N90C)_1(M2)_7$ was then excised from the gel and immersed in an extraction solution [150 mM NaCl, 15 mM tris-HCl (pH 7.5), 0.2% DDM, 0.5% Genapol X-80, 5 mM TCEP, and10 mM EDTA] for 12 hours. The mixture was centrifuged, and the supernatant, which contains $(N90C)_1(M2)_7$, was collected for future uses. The prepared $(N90C)_1(M2)_7$ was either immediately used or stored at −80°C for long-term storages.

To prepare MspA-90PBA, 1 μl of 500 mM MPBA, 2.5 μl of $(N90C)_1(M2)_7$, and 21.5 μl of 1.5 M KCl buffer were mixed and incubated at room temperature for 10 min. The prepared MspA-90PBA was immediately used in nanopore measurements.

Nanopore measurements

To avoid interferences caused by environment noises, the measurement device was fixed in a Faraday cage mounted on a floating optical table (Jiangxi Liansheng Technology Co. Ltd). The measurement device consists of two chambers separated by a Teflon film containing an orifice (~100-μm diameter). Before each measurement, the orifice was first treated with 2% (v/v) hexadecane in pentane. Both chambers were then filled with 500 μl of 1.5 M KCl buffer [1.5 M KCl and 100 mM Mops (pH 7.0)]. A pair of Ag/AgCl electrodes, which were electrically connected to a patch-clamp amplifier, were respectively placed in both chambers, in contact with the buffer. Conventionally, the chamber that is electrically grounded is defined as cis, and its opposing chamber is defined as trans. By adding a drop of DPhPC (5 mg/ml) in pentane to each chamber and pipetting the buffer in either chamber up and down for several times, a lipid bilayer was spontaneously formed at the orifice. Then, nanopores were added to the cis chamber to trigger pore insertion. Upon single nanopore insertion, the buffer in cis was immediately exchanged with fresh buffer to avoid further nanopore insertions.

All nanopore measurements were carried out by single-channel recording with an Axopatch 200B patch-clamp amplifier coupled with a Digidata 1550B digitizer. The sampling rate is 25 kHz, and the acquired trace is further digitally low-pass filtered with a corner frequency of 1 kHz. All measurements were performed in a 1.5 M KCl buffer with a continually applied +160-mV potential. If not otherwise stated, all analytes were added to the cis chamber and measured for technical replicates.

Data analysis

All sensing events were detected by the "Single-Channel Search" function in Clampfit 10.7, and the "Ignore duration" was set to 10 ms. Raw Axon abf files were then imported into MATLAB R2021a using the "abfload" function downloaded from www.mathworks.com/matlabcentral/fileexchange/6190-abfload. The event start time (t_{start}) and event end time (t_{end}) were from the result of Single-Channel Search. Each event was extracted from the current trace on the basis of the time stamp. The baseline current (I_0) is calculated from the average of 10 ms current before event start time and 10 ms current after event end time. The event current is defined as the current between t_{start} and t_{end}, defined as blockage level. To avoid the impact of sudden current changes at the beginning and end of the event on the current statistics, the 0.4-ms current at both the start and end of the event was ignored in the calculation. I_b is the average blockage level current. The mean blockage amplitude (ΔI) is the difference between I_0 and I_b ($\Delta I = I_0 - I_b$). SD is the standard deviation of the blockage level. Time is the event dwell time ($t_{end} - t_{start}$). The minimum (*min*), maximum (*max*), median (*med*), first quartile (*Q1*), and third quartile (*Q3*) of the blockage level are calculated. These statistics are then used to compute Δmin ($I_0 - min$), Δmax ($I_0 - max$), $\Delta median$ ($I_0 - med$), *IQR* ($Q3 - Q1$), and *range* ($max - min$). Eight event features, including $\Delta I/I_0$, *time*, *SD*, $\Delta median/I_0$, IQR/I_0, $\Delta max/I_0$, $\Delta min/I_0$, and $range/I_0$, were extracted subsequently. To remove noises appearing as nonclustering events, cluster analysis was performed by a DBSCAN algorithm. The dimensionality reduction was performed by the *t*-SNE algorithm using Seuclidean distance.

Database and machine learning

A saccharide pool containing 10 oligosaccharides (Suc, lactulose, Lac, Mel, G-G, Kes, Mal, L-F, 2F-L, and manninotriose; tables S1 and S2) and nine monosaccharides (*62*) (Glc, Man, Gal, GalNAc, Fru, Sor, Rha, Xyl, and Rib) is established. Events acquired with these saccharides were collected to build the saccharide database. For each saccharide, all events from three independent measurements are recorded in the database. Machine learning was performed by MATLAB R2021a. In the database, 1500 events acquired with each saccharide type were randomly selected for machine learning model building. To remove noise events, all events were cluster analysis treated by DBSCAN algorithm. After the treatment, for each saccharide type, 1000 events were randomly selected from the remained events. The selected data were randomly split into a training set (80%, 800 events of each saccharide) and a testing set (20%, 200 events of each saccharide), respectively, for model training and testing. Model training was performed with the training set using default models in the Classification Learner toolbox of MATLAB R2021a. Ten-fold cross-validation was used to evaluate the model performance. Among all models being tested, the bagged tree model reported the best performance, which is used for all predictions. The confusion matrix of testing set was used to show the accuracy of each saccharide.

The machine learning code, training data, and saccharide database are publicly available on Figshare. Please follow the link: https://doi.org/10.6084/m9.figshare.29539016.v2 for download.

The analysis of oligosaccharide events

All events acquired with the mixed oligosaccharides, unidentified oligosaccharides, and oligosaccharide hydrolysates are first cluster analyzed by DBSCAN and then predicted by the trained bagged tree

model. According to machine learning prediction, a predictive score is given to each event. Only events with predictive scores ≥0.8 are retained. Events predicted to be of the same class are classified as a distribution. Otherwise, this distribution is considered to be generated by an unidentified oligosaccharide and the events of which are thus not recorded in the database.

All retained events were then separately grouped to different scatter plots according to their prediction classes. For each class of events, their event frequency distribution on the dimensions of $\Delta I/I_0$ versus *SD*, which is in the form of a 2D histogram, is generated as described in fig. S14. $\Delta I/I_0$ between 0 and 0.6 is divided into 200 bins (0.003 per bins), and SD between 0 and 100 pA is divided into 200 bins (0.5 pA per bins). The event counts in each bin are normalized to event frequency. The event distribution is then compared with the standard event distribution of its predicted class, which is generated from the training set recorded in the database. By overlapping the histograms of the measured distribution and the standard event distribution in the database, a matching degree value is generated (fig. S15). The overlapping principle is, if the bin in the standard distribution and the same bin in the measured distribution are nonzero, then the bin in the standard distribution is specified to be overlapped. The matching degree equals to the sum of standard frequency in all overlapped bins. Only distributions that show a ≥80% matching degree with the event distribution generated by corresponding standard events in the database are considered the correct distributions (fig. S15).

Acid hydrolysis of oligosaccharides

To paritially hydrolyze the oligosaccharides, Mal (80 μmol), manninotriose (20 μmol), 2F-L (80 μmol), and Ver (20 μmol) were respectively added to 1 ml of 0.5 M TFA and incubated at 80°C for 4 hours. To completely hydrolyze the oligosaccharides into monosaccharides, manninotriose (80 μmol), Raf (80 μmol), and Sta (80 μmol) were respectively added to 1 ml of 0.5 M TFA and incubated at 80°C for 12 hours. The pH of the samples was adjusted to 7.0 by KOH titration. The samples were then ultrafiltration treated (3 kDa, 6000 rpm, 1 hour), and the filtrate was collected.

Enzymatic hydrolysis of oligosaccharides

Raf, Sta, and Ver were enzymatically hydrolyzed at different conditions, as described below:

1) Raf (300 μmol) and Sta (300 μmol) were respectively added to 1 ml of pH 5.0 (100 mM CH_3COONa) buffer dissolved with β-D-fructofuranosidase (300 U) and incubated at 55°C for 1 hour.

2) Raf (300 μmol) and Sta (300 μmol) were respectively added to 1 ml of pH 7.0 (100 mM NaH_2PO_4) buffer dissolved with α-glucosidase (50 U) and incubated at 37°C for 1 hour.

3) Raf (50 μmol) and Sta (50 μmol) were respectively added to 1 ml of pH 5.0 (100 mM CH_3COONa) buffer dissolved with β-glucosidase (10 U) and incubated at 37°C for 1 hour.

4) Raf (90 μmol) and Sta (90 μmol) were respectively added to 300 μl of pH 6.5 (100 mM KH_2PO_4) buffer dissolved with α-galactosidase (2 U) and incubated at 25°C for 2 hours.

5) Raf (50 μmol) and Sta (50 μmol) were respectively added to 1 ml of pH 5.0 (100 mM CH_3COONa) buffer dissolved with β-galactosidase (10 U) and incubated at 37°C for 1 hour.

6) Ver (75 mM) was added to 200 μl of pH 6.5 (100 mM KH_2PO_4) buffer dissolved with α-galactosidase (0.4 U) and incubated at 25°C for 2 hours.

7) Ver (75 mM) was added to 200 μl of pH 5.0 (100 mM CH_3COONa) buffer dissolved with β-D-fructofuranosidase (60 U) and incubated at 55°C for 1 hour.

To remove enzymes, all hydrolysis products were ultrafiltered (3 kDa, 6000 rpm, 1 hour), and the filtrates were collected.

Supplementary Materials

This PDF file includes:

Figs. S1 to S59
Tables S1 to S7
References

REFERENCES

1. M. Guberman, P. H. Seeberger, Automated glycan assembly: A perspective. *J. Am. Chem. Soc.* **141**, 5581–5592 (2019).
2. R. A. Flynn, K. Pedram, S. A. Malaker, P. J. Batista, B. A. H. Smith, A. G. Johnson, B. M. George, K. Majzoub, P. W. Villalta, J. E. Carette, C. R. Bertozzi, Small RNAs are modified with N-glycans and displayed on the surface of living cells. *Cell* **184**, 3109–3124.e22 (2021).
3. R. Apweiler, H. Hermjakob, N. Sharon, On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* **1473**, 4–8 (1999).
4. H. Du, H. Yu, F. Yang, Z. Li, Comprehensive analysis of glycosphingolipid glycans by lectin microarrays and MALDI-TOF mass spectrometry. *Nat. Protoc.* **16**, 3470–3491 (2021).
5. N. J. Agard, C. R. Bertozzi, Chemical approaches to perturb, profile, and perceive glycans. *Acc. Chem. Res.* **42**, 788–797 (2009).
6. J. V. Gomes, S. Singh-Bhagania, M. Cenci, C. C. Cordon, M. Singh, J. A. Butterwick, The molecular basis of sugar detection by an insect taste receptor. *Nature* **629**, 228–234 (2024).
7. K. W. Moremen, M. Tiemeyer, A. V. Nairn, Vertebrate protein glycosylation: Diversity, synthesis and function. *Nat. Rev. Mol. Cell. Biol.* **13**, 448–462 (2012).
8. F. Tang, M. Zhou, K. Qin, W. Shi, A. Yashinov, Y. Yang, L. Yang, D. Guan, L. Zhao, Y. Tang, Y. Chang, L. Zhao, H. Yang, H. Zhou, R. Huang, W. Huang, Selective *N*-glycan editing on living cell surfaces to probe glycoconjugate function. *Nat. Chem. Biol.* **16**, 766–775 (2020).
9. J. Poole, C. J. Day, M. von Itzstein, J. C. Paton, M. P. Jennings, Glycointeractions in bacterial pathogenesis. *Nat. Rev. Microbiol.* **16**, 440–452 (2018).
10. P. Li, Z. Liu, Glycan-specific molecularly imprinted polymers towards cancer diagnostics: Merits, applications, and future perspectives. *Chem. Soc. Rev.* **53**, 1870–1891 (2024).
11. B. A. H. Smith, C. R. Bertozzi, The clinical impact of glycobiology: Targeting selectins, Siglecs and mammalian glycans. *Nat. Rev. Drug. Discov.* **20**, 217–243 (2021).
12. C. Fontana, G. Widmalm, Primary structure of glycans by NMR spectroscopy. *Chem. Rev.* **123**, 1040–1102 (2023).
13. L. R. Ruhaak, G. Xu, Q. Li, E. Goonatilleke, C. B. Lebrilla, Mass spectrometry approaches to glycomic and glycoproteomic analyses. *Chem. Rev.* **118**, 7886–7930 (2018).
14. L. Veillon, Y. Huang, W. Peng, X. Dong, B. G. Cho, Y. Mechref, Characterization of isomeric glycan structures by LC-MS/MS. *Electrophoresis* **38**, 2100–2114 (2017).
15. K. Greis, C. Kirschbaum, G. von Helden, K. Pagel, Gas-phase infrared spectroscopy of glycans and glycoconjugates. *Curr. Opin. Struct. Biol.* **72**, 194–202 (2022).
16. H.-Y. Yao, J.-Q. Wang, J.-Y. Yin, S.-P. Nie, M.-Y. Xie, A review of NMR analysis in polysaccharide structure and conformation: Progress, challenge and perspective. *Food Res. Int.* **143**, 110290 (2021).
17. M. Grabarics, M. Lettow, C. Kirschbaum, K. Greis, C. Manz, K. Pagel, Mass spectrometry-based techniques to elucidate the sugar code. *Chem. Rev.* **122**, 7840–7908 (2022).
18. J. Wang, J. Zhao, S. Nie, M. Xie, S. Li, Mass spectrometry for structural elucidation and sequencing of carbohydrates. *TrAC Trends Anal. Chem.* **144**, 116436 (2021).
19. W. Peng, C. D. Gutierrez Reyes, S. Gautam, A. Yu, B. G. Cho, M. Goli, K. Donohoo, S. Mondello, F. Kobeissy, Y. Mechref, MS-based glycomics and glycoproteomics methods enabling isomeric characterization. *Mass Spectrom. Rev.* **42**, 577–616 (2023).
20. S. A. McLuckey, Principles of collisional activation in analytical mass spectrometry. *J. Am. Soc. Mass Spectrom.* **3**, 599–614 (1992).
21. J. E. P. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz, D. F. Hunt, Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9528–9533 (2004).
22. F. W. McLafferty, Tandem mass spectrometry (MS/MS): A promising new analytical technique for specific component determination in complex mixtures. *Acc. Chem. Res.* **13**, 33–39 (1980).
23. P. Bansal, A. Ben Faleh, S. Warnke, T. R. Rizzo, Multistage ion mobility spectrometry combined with infrared spectroscopy for glycan analysis. *J. Am. Soc. Mass Spectrom.* **34**, 695–700 (2023).

24. C. Y. Liew, C.-C. Yen, J.-L. Chen, S.-T. Tsai, S. Pawar, C.-Y. Wu, C.-K. Ni, Structural identification of *N*-glycan isomers using logically derived sequence tandem mass spectrometry. *Commun. Chem.* **4**, 92 (2021).
25. L. Han, C. E. Costello, Electron transfer dissociation of milk oligosaccharides. *J. Am. Soc. Mass Spectrom.* **22**, 997–1013 (2011).
26. C. Ashwood, C.-H. Lin, M. Thaysen-Andersen, N. H. Packer, Discrimination of isomers of released *N*- and *O*-glycans using diagnostic product ions in negative ion PGC-LC-ESI-MS/MS. *J. Am. Soc. Mass Spectrom.* **29**, 1194–1209 (2018).
27. J. Urban, C. Jin, K. A. Thomsson, N. G. Karlsson, C. M. Ives, E. Fadda, D. Bojar, Predicting glycan structure from tandem mass spectrometry via deep learning. *Nat. Methods* **21**, 1206–1215 (2024).
28. Y. Watanabe, K. F. Aoki-Kinoshita, Y. Ishihama, S. Okuda, GlycoPOST realizes FAIR principles for glycomics mass spectrometry data. *Nucleic Acids Res.* **49**, D1523–D1528 (2021).
29. C. A. Hayes, N. G. Karlsson, W. B. Struwe, F. Lisacek, P. M. Rudd, N. H. Packer, M. P. Campbell, UniCarb-DB: A database resource for glycomic discovery. *Bioinformatics* **27**, 1343–1344 (2011).
30. G. Zauner, A. M. Deelder, M. Wuhrer, Recent advances in hydrophilic interaction liquid chromatography (HILIC) for structural glycomics. *Electrophoresis* **32**, 3456–3466 (2011).
31. S. C. Churms, Recent progress in carbohydrate separation by high-performance liquid chromatography based on hydrophilic interaction. *J. Chromatogr. A* **720**, 75–91 (1996).
32. E. Wiercigroch, E. Szafraniec, K. Czamara, M. Z. Pacia, K. Majzner, K. Kochan, A. Kaczor, M. Baranska, K. Malek, Raman and infrared spectroscopy of carbohydrates: A review. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **185**, 317–335 (2017).
33. T. Hong, J.-Y. Yin, S.-P. Nie, M.-Y. Xie, Applications of infrared spectroscopy in polysaccharide structural analysis: Progress, challenge and perspective. *Food Chem. X* **12**, 100168 (2021).
34. G. M. Cherf, K. R. Lieberman, H. Rashid, C. E. Lam, K. Karplus, M. Akeson, Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nat. Biotechnol.* **30**, 344–348 (2012).
35. E. A. Manrao, I. M. Derrington, A. H. Laszlo, K. W. Langford, M. K. Hopper, N. Gillgren, M. Pavlenok, M. Niederweis, J. H. Gundlach, Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* **30**, 349–353 (2012).
36. D. R. Garalde, E. A. Snell, D. Jachimowicz, B. Sipos, J. H. Lloyd, M. Bruce, N. Pantic, T. Admassu, P. James, A. Warland, M. Jordan, J. Ciccone, S. Serra, J. Keenan, S. Martin, L. McNeill, E. J. Wallace, L. Jayasinghe, C. Wright, J. Blasco, S. Young, D. Brocklebank, S. Juul, J. Clarke, A. J. Heron, D. J. Turner, Turner, highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* **15**, 201–206 (2018).
37. Method of the year 2022: Long-read sequencing. *Nat. Methods* **20**, 1 (2023).
38. J. Quick, N. J. Loman, S. Duraffour, J. T. Simpson, E. Severi, L. Cowley, J. A. Bore, R. Koundouno, G. Dudas, A. Mikhail, N. Ouédraogo, B. Afrough, A. Bah, J. H. Baum, B. Becker-Ziaja, J. P. Boettcher, M. Cabeza-Cabrerizo, Á. Camino-Sánchez, L. L. Carter, J. Doerrbecker, T. Enkirch, I. G. Dorival, N. Hetzelt, J. Hinzmann, T. Holm, L. E. Kafetzopoulou, M. Koropogui, A. Kosgey, E. Kuisma, C. H. Logue, A. Mazzarelli, S. Meisel, M. Mertens, J. Michel, D. Ngabo, K. Nitzsche, E. Pallash, L. V. Patrono, J. Portmann, J. G. Repits, N. Y. Rickett, A. Sachse, K. Singethan, I. Vitoriano, R. L. Yemanaberhan, E. G. Zekeng, R. Trina, A. Bello, A. A. Sall, O. Faye, O. Faye, N.'F. Magassouba, C. V. Williams, V. Amburgey, L. Winona, E. Davis, J. Gerlach, F. Washington, V. Monteil, M. Jourdain, M. Bererd, A. Camara, H. Somlare, A. Camara, M. Gerard, G. Bado, B. Baillet, D. Delaune, K. Y. Nebie, A. Diarra, Y. Savane, R. B. Pallawo, G. J. Gutierrez, N. Milhano, I. Roger, C. J. Williams, F. Yattara, K. Lewandowski, J. Taylor, P. Rachwal, D. J. Turner, G. Pollakis, J. A. Hiscox, D. A. Matthews, M. K. O'Shea, A. M. Johnston, D. Wilson, E. Hutley, E. Smit, A. D. Caro, R. Wölfel, K. Stoecker, E. Fleischmann, M. Gabriel, S. A. Weller, L. Koivogui, B. Diallo, S. Keïta, A. Rambaut, P. Formenty, S. Günther, M. W. Carroll, Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).
39. R. A. Bull, T. N. Adikari, J. M. Ferguson, J. M. Hammond, I. Stevanovski, A. G. Beukers, Z. Naing, M. Yeang, A. Verich, H. Gamaarachchi, K. W. Kim, F. Luciani, S. Stelzer-Braid, J.-S. Eden, W. D. Rawlinson, S. J. van Hal, I. W. Deveson, Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. *Nat. Commun.* **11**, 6272 (2020).
40. M. C. Lucas, L. P. Pryszcz, R. Medina, I. Milenkovic, N. Camacho, V. Marchand, Y. Motorin, L. R. de Pouplana, E. M. Novoa, Quantitative analysis of tRNA abundance and modifications by nanopore RNA sequencing. *Nat. Biotechnol.* **42**, 72–86 (2024).
41. Y.-L. Ying, Z.-L. Hu, S. Zhang, Y. Qing, A. Fragasso, G. Maglia, A. Meller, H. Bayley, C. Dekker, Y.-T. Long, Nanopore-based technologies beyond DNA sequencing. *Nat. Nanotechnol.* **17**, 1136–1146 (2022).
42. G. Yao, W. Ke, B. Xia, Z. Gao, Nanopore-based glycan sequencing: State of the art and future prospects. *Chem. Sci.* **15**, 6229–6243 (2024).
43. S. Yan, J. Zhang, Y. Wang, W. Guo, S. Zhang, Y. Liu, J. Cao, Y. Wang, L. Wang, F. Ma, P. Zhang, H.-Y. Chen, S. Huang, Single molecule ratcheting motion of peptides in a *Mycobacterium smegmatis* porin a (MspA) nanopore. *Nano Lett.* **21**, 6703–6710 (2021).
44. H. Brinkerhoff, A. S. W. Kang, J. Liu, A. Aksimentiev, C. Dekker, Multiple rereads of single proteins at single–amino acid resolution using nanopores. *Science* **374**, 1509–1513 (2021).
45. I. C. Nova, J. Ritmejeris, H. Brinkerhoff, T. J. R. Koenig, J. H. Gundlach, C. Dekker, Detection of phosphorylation post-translational modifications along single peptides with nanopores. *Nat. Biotechnol.* **42**, 710–714 (2024).
46. B. I. Karawdeniya, Y. M. N. D. Y. Bandara, J. W. Nichols, R. B. Chevalier, J. R. Dwyer, Surveying silicon nitride nanopores for glycomics and heparin quality assurance. *Nat. Commun.* **9**, 3278 (2018).
47. F. Rivas, O. K. Zahid, H. L. Reesink, B. T. Peal, A. J. Nixon, P. L. DeAngelis, A. Skardal, E. Rahbar, A. R. Hall, Label-free analysis of physiological hyaluronan size distribution with a solid-state nanopore sensor. *Nat. Commun.* **9**, 1037 (2018).
48. K. Xia, J. T. Hagan, L. Fu, B. S. Sheetz, S. Bhattacharya, F. Zhang, J. R. Dwyer, R. J. Linhardt, Synthetic heparan sulfate standards and machine learning facilitate the development of solid-state nanopore analysis. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2022806118 (2021).
49. F. Rivas, P. L. DeAngelis, E. Rahbar, A. R. Hall, Optimizing the sensitivity and resolution of hyaluronan analysis with solid-state nanopores. *Sci. Rep.* **12**, 4469 (2022).
50. J. Im, S. Lindsay, X. Wang, P. Zhang, Single molecule identification and quantification of glycosaminoglycans using solid-state nanopores. *ACS Nano* **13**, 6308–6318 (2019).
51. Y. Cai, B. Zhang, L. Liang, S. Wang, L. Zhang, L. Wang, H.-L. Cui, Y. Zhou, D. Wang, A solid-state nanopore-based single-molecule approach for label-free characterization of plant polysaccharides. *Plant Commun.* **2**, 100106 (2021).
52. W. Xie, S. He, S. Fang, B. Yin, R. Tian, Y. Wang, D. Wang, Analysis of starch dissolved in ionic liquid by glass nanopore at single molecular level. *Int. J. Biol. Macromol.* **239**, 124271 (2023).
53. Q. H. Nguyen, M. Ali, R. Neumann, W. Ensinger, Saccharide/glycoprotein recognition inside synthetic ion channels modified with boronic acid. *Sens. Actuators B Chem.* **162**, 216–222 (2012).
54. Z. Sun, C. Han, L. Wen, D. Tian, H. Li, L. Jiang, pH gated glucose responsive biomimetic single nanochannels. *Chem. Commun. (Camb)* **48**, 3282–3284 (2012).
55. S. Zhao, Y.-B. Zheng, S.-L. Cai, Y.-H. Weng, S.-H. Cao, J.-L. Yang, Y.-Q. Li, Sugar-stimulated robust nanodevice: 4-Carboxyphenylboronic acid modified single glass conical nanopores. *Electrochem. Commun.* **36**, 71–74 (2013).
56. X. Li, K. H. Lee, S. Shorkey, J. Chen, M. Chen, Different anomeric sugar bound states of maltose binding protein resolved by a cytolysin a nanopore tweezer. *ACS Nano* **14**, 1727–1737 (2020).
57. N. S. Galenkamp, M. Soskine, J. Hermans, C. Wloka, G. Maglia, Direct electrical quantification of glucose and asparagine from bodily fluids using nanopores. *Nat. Commun.* **9**, 4085 (2018).
58. M. Li, Y. Xiong, Y. Cao, C. Zhang, Y. Li, H. Ning, F. Liu, H. Zhou, X. Li, X. Ye, Y. Pang, J. Zhang, X. Liang, G. Qing, Identification of tagged glycans with a protein nanopore. *Nat. Commun.* **14**, 1737 (2023).
59. B. Xia, J. Fang, S. Ma, M. Ma, G. Yao, T. Li, X. Cheng, L. Wen, Z. Gao, Mapping the acetylamino and carboxyl groups on glycans by engineered alpha-hemolysin nanopores. *J. Am. Chem. Soc.* **145**, 18812–18824 (2023).
60. G. Yao, B. Xia, F. Wei, J. Wang, Y. Yang, S. Ma, W. Ke, T. Li, X. Cheng, L. Wen, Y.-T. Long, Z. Gao, Glycan sequencing based on glycosidase-assisted nanopore sensing. *J. Am. Chem. Soc.* **147**, 1721–1731 (2025).
61. W. J. Ramsay, H. Bayley, 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).
62. S. Zhang, Z. Cao, P. Fan, Y. Wang, W. Jia, L. Wang, K. Wang, Y. Liu, X. Du, C. Hu, P. Zhang, H.-Y. Chen, S. Huang, A nanopore-based saccharide sensor. *Angew. Chem. Int. Ed. Engl.* **61**, e202203769 (2022).
63. S. Zhang, Z. Cao, P. Fan, W. Sun, Y. Xiao, P. Zhang, Y. Wang, S. Huang, Discrimination of disaccharide isomers of different glycosidic linkages using a modified MspA nanopore. *Angew. Chem. Int. Ed. Engl.* **63**, e202316766 (2024).
64. T. Z. Butler, M. Pavlenok, I. M. Derrington, M. Niederweis, J. H. Gundlach, Single-molecule DNA detection with an engineered MspA protein nanopore. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20647–20652 (2008).
65. Y. Wang, S. Zhang, W. Jia, P. Fan, L. Wang, X. Li, J. Chen, Z. Cao, X. Du, Y. Liu, K. Wang, C. Hu, J. Zhang, J. Hu, P. Zhang, H.-Y. Chen, S. Huang, Identification of nucleoside monophosphates and their epigenetic modifications using an engineered nanopore. *Nat. Nanotechnol.* **17**, 976–983 (2022).
66. R. R. Kapaev, P. V. Toukach, GRASS: Semi-automated NMR-based structure elucidation of saccharides. *Bioinformatics* **34**, 957–963 (2018).

Acknowledgments

Funding: This work was supported by the National Key R&D Program of China (grant no. 2023YFF1205900, to S.H.), the National Natural Science Foundation of China (grant nos. 22225405 and 22534004, to S.H.), the Fundamental Research Funds for the Central Universities (grant no. 020514380336, to S.H.), and the State Key Laboratory of Analytical Chemistry for Life Science (grant no.5431ZZXM2509, to S.H.). **Author contributions:** Conceptualization: S.H.,

S.Z., and Y.X. Methodology: Y.X. and S.Z. Investigation: Y.X., S.Z., X.G., T.L., and H.Z. Resources: P.Z. Visualization: Y.X. and X.G. Supervision: S.H. Writing—original draft: Y.X. and S.H. Writing—review and editing: Y.X. and S.H. **Competing interests:** S.H. and S.Z. have filed patents (WO/2023/056960, 13.04.2023) held by Nanjing University that covers the heterogeneous MspA and its applications thereof with national phase entries in United States of America, Japan, China, and European Patent Office. All other authors declare that they have no competing interests. **Data and materials availability:** All data and code needed to evaluate and reproduce the results in the paper are present in the paper and/or the Supplementary Materials. The code and training data used for machine learning model building and the saccharide database are submitted as a zip file named "oligosaccharide classifier" and can be found at https://doi.org/10.6084/m9.figshare.29539016.v2.

Submitted 22 July 2025
Accepted 25 November 2025
Published 1 January 2026
10.1126/sciadv.aea8462