nature biotechnology

Supplementary information

https://doi.org/10.1038/s41587-026-03308-9

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

In the format provided by the authors and unedited

Table of contents

Materials

Maleic anhydride, tert-butyl N-(2-aminoethyl)carbamate, dichloromethane (DCM), acetone, sodium acetate (NaAc), acetic anhydride (Ac_2O), triethylamine (TEA), ethyl acetate, n-heptane, trifluoroacetic acid (TFA), 2-(7-Azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HATU), N,N-Diisopropylethylamine (DIPEA) and tetrahydrofuran (THF) were from Selectoprobe (China). 3-borono-4-formylbenzoic acid was from Bidepharm Co., Ltd (China).

Hexadecane, pentane and Genapol X-80 were from Sigma-Aldrich. 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPC) was from Avanti Polar Lipids. Precision Plus Protein™ Dual Xtra Standards, TGX™ FastCastTM Acylamide Kit (4-15%), stacking gel buffer (0.5 M Tris-HCl buffer, pH 6.8), resolving gel buffer (1.5 M Tris-HCl buffer, pH 8.8) were from Bio-Rad (USA). SDS-PAGE electrophoresis buffer powder was from Beyotime. *E. coli* BL21 (DE3) was from TransGen Biotech. *E. coli* BL21 (DE3) pLysS was from Sangon Biotech. Luria-Bertani (LB) agar and LB broth were from Hopebio. Potassium chloride (KCl), 2-morpholinoethanesulfonic acid (MES), 3-morpholino propionic acid (MOPS), 2-(N-cyclohexylamino)ethane sulfonic acid (CHES) and potassium hydroxide (KOH) were supplied by Aladdin (China).

Alanine, cysteine, aspartic acid, glutamic acid, phenylalanine, glycine, histidine, isoleucine, lysine, leucine, methionine, asparagine, proline, glutamine, arginine, serine, threonine, valine, tryptophan and tyrosine were from BBI life sciences corporation. N6-Acetyl-L-lysine (Ac-K), adenosine 5'-monophosphate (AMP), uridine 5'-monophosphate (UMP), cytidine 5'-monophosphate (CMP), guanosine 5'-monophosphate (GMP), inosine-5'-monophosphate (I), D-fructose (Fru), N-acetyl-D-glucosamine (GlcNAc) and yeast cell extract were from Aladdin (China). O-Phospho-L-tyrosine (P-Y) was from Bidepharm Co., Ltd (China). N3,N4-Dimethyl-L-arginine (Me-R) was from Shanghai Acmec Biochemical Co., Ltd. Iduronic acid (IdoA) was from Glycarbo Co., Ltd (Japan). L-arabinose (Ara) was from Macklin (China). N6-methyladenosine-5'-monophosphate (m^6A) was from TargetMol (USA). Pseudouridine-5'-monophosphate (ψ) was from Wuxi AppTec. Maleimide and leucine aminopeptidase (LAP) were from Shanghai Yuanye Bio-technology (China). The glycosylated heptapeptide with the sequence GMQRN(GlcNAc)IS was synthesized by Nanjing Jietai Biotechnology Co., Ltd (China). Ile-Lys (IK), Ser-Leu-Arg (SLR), Leu-Arg-Gly (LRG), Leu-Arg-Gly-Tyr (LRGY) and Gly-Tyr-Ile-Lys (GYIK) were synthesized by DGpeptides Co., Ltd. (China).

All buffers were prepared with Milli-Q water and membrane (0.2 μm, Whatman) filtered prior to use. The potassium chloride buffer (1.5 M KCl, 10 mM MOPS) was treated with Chelex 100 resin (Bio-Rad) overnight and adjusted to pH 7.0 prior to use. The *cis* side buffer (1.5 M KCl, 10 mM CHES) was adjusted to pH 9.0 and the *trans* side buffer (1.5 M KCl, 10 mM MES) was adjusted to pH 6.0 prior to use.

Supplementary Tables

Supplementary Table 1. The protein sequences.

Source Plasmid	Protein sequence
M2 MspA-D16H6	MGLDNELSLVDGQDRTLTVQQWDTFLNGVFPLDRNRLTREWFHSGRAKYIVAGPG ADEFEGTLELGYQIGFPWSLGVGINFSYTTPNILINNGNITAPPFGLNSVITPNLFPGV SISARLGNGPGIQEVATFSVRVSGAKGGVAVSNAHGTVTGAAGGVLLRPFARLIAST GDSVTTYGEPWNMNDDDDDDDDDDDDDDDHHHHHH*
N90C MspA-H6	MGLDNELSLVDGQDRTLTVQQWDTFLNGVFPLDRNRLTREWFHSGRAKYIVAGPG ADEFEGTLELGYQIGFPWSLGVGINFSYTTPNILICNGNITAPPFGLNSVITPNLFPGV SISARLGNGPGIQEVATFSVRVSGAKGGVAVSNAHGTVTGAAGGVLLRPFARLIAST GDSVTTYGEPWNMNHHHHHH*

Footnotes:

1. The red characters indicate the amino acid mutation introduced at site 90 in each gene expression product.
2. The His-tag is highlighted with purple characters within the sequence.
3. The poly-aspartic acid tag (D16) is highlighted with orange characters within the sequence.

Supplementary Table 2. The Current-Voltage (I-V) curves. The measurements were carried out as described in **Supplementary Fig. 5**. The *cis* compartment was filled with 1.5 M KCl buffer (1.5 M KCl, 10 mM CHES, pH 9.0), and the *trans* compartment was filled with 1.5 M KCl buffer (1.5 M KCl, 10 mM MES, pH 6.0). The open pore current of $(N90C)_1(M2)_7$ and MspA-FPBA were measured respectively when the applied voltage was ramped from -200 mV to +200 mV in +10 mV intervals. Three independent measurements (N = 3) were performed for each condition to form the statistics.

Voltage (mV)	Current (pA)	
	$(N90C)_1(M2)_7$	**MspA-FPBA**
-200	-805.8±1	-689.2±2
-190	-762.6±0.5	-650.6±0.9
-180	-718.9±0.8	-609.8±1.8
-170	-677.4±1	-570.7±0.8
-160	-634.8±0.1	-530.1±0.6
-150	-592.0±0.8	-490.3±1.7
-140	-549.7±1	-452.4±1.5
-130	-507.1±0.7	-413.9±0.2
-120	-466.1±1.4	-374.1±2
-110	-424.7±1	-337.8±1
-100	-383.7±0.2	-301.5±0.7
-90	-342.8±0.3	-265.7±0.4
-80	-303.2±0.5	-230.9±0.2
-70	-263.4±0.5	-196.8±0.3
-60	-224.6±0.5	-164.7±0.2
-50	-185.8±0.2	-133.1±0.1
-40	-147.9±0.4	-103.7±0.1
-30	-110.7±0.2	-75.2±0.1
-20	-73.7±0.3	-48.7±0.2
-10	-38.0±0.2	-23.5±0.1
0	-2.4±0.2	-0.3±0.1
10	32.5±0.1	21.9±0.2
20	66.2±0.2	43.8±0.2
30	99.8±0.2	64.0±0.6
40	132.2±0.1	84.0±0.2
50	165.2±0.4	103.7±0.3
60	197.0±0.3	123.5±0.6
70	228.5±0.5	143.2±0.5
80	259.7±0.2	163.5±0.9
90	291.0±0.4	183.4±0.2
100	323.4±0.8	204.5±0.6
110	354.5±0.5	226.7±1.2
120	387.1±1.8	248.0±0.4

130	420.8±2.2	272.6±0.3
140	453.4±1	296.2±0.2
150	487.2±1.6	320.4±0.8
160	522.8±1.5	346.1±0.9
170	556.8±1.3	372.6±0.1
180	597.5±4.4	401.2±0.3
190	635.8±1.4	430.5±0.8
200	675.6±2.4	461.7±0.2

Supplementary Table 3. The statistics of core parameters of events acquired with several proteinogenic amino acids. Nanopore measurements were carried out as described in **Supplementary Figs. 9-12** in 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). A transmembrane voltage of +160 mV was continually applied during the measurements. All amino acids were separately added to *cis* with a final concentration of 400 μM. The definition and derivation of the characteristic parameters were summarized in **Supplementary Fig. 7**. Three independent measurements (*N* = 3) were performed for each analyte type to obtain the statistics.

Amino Acid	$\overline{\Delta \boldsymbol{I}}$ (pA)	$\overline{\mathbf{SD}}$ (pA)	$\overline{\boldsymbol{\tau}_{\mathbf{off}}}$ (s)	$\overline{\boldsymbol{\tau}_{\mathbf{on}}}$ (s)
Alanine (A)	33.12±0.34	1.08±0.06	0.37±0.11	1.99±0.21
Glutamic acid (E)	6.7±0.14	0.92±0.05	0.71±0.34	1.71±0.57
Glycine (G)	20.14±0.04	1.44±0.06	0.47±0.02	1.0±0.02
Isoleucine (I)	43.66±0.14	1.5±0.01	0.47±0.29	1.83±0.51
Leucine (L)	43.55±0.41	1.12±0.02	0.34±0.23	0.77±0.22
Methionine (M)	40.3±0.76	1.03±0.03	0.22±0.13	0.65±0.35
Asparagine (N)	63.74±0.33	3.8±0.13	0.16±0.01	0.76±0.14
Glutamine (Q)	34.96±0.42	1.03±0.05	0.3±0.07	0.49±0.07
Arginine (R)	85.7±0.61	2.11±0.1	0.5±0.05	0.55±0.15
Serine (S)	31.6±0.14	1.09±0.03	0.23±0.06	0.81±0.2
Valine (V)	38.81±0.12	1.47±0.02	0.31±0.13	0.65±0.23
Tyrosine (Y)	57.16±0.52	1.77±0.01	0.31±0.01	0.72±0.08

Supplementary Table 4. The statistics of core parameters of events acquired with four canonical nucleoside monophosphates. Nanopore measurements were carried out as described in **Supplementary Fig. 15** in 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). A transmembrane voltage of +160 mV was continually applied during the measurements. All nucleoside monophosphates were separately added to *cis* with a final concentration of 2 mM. The definition and derivation of the characteristic parameters were summarized in **Supplementary Fig. 7**. Binding of all nucleoside monophosphates produces two types of events (**Extended Data Fig. 4**) so that the $\overline{\Delta I}$, $\overline{\mathrm{SD}}$ and $\overline{\tau_{\mathrm{off}}}$ were derived separately. Three independent measurements (*N* = 3) were performed for each analyte type to obtain the statistics.

Nucleoside monophosphate	$\overline{\Delta I}$ (pA)	$\overline{\mathrm{SD}}$ (pA)	$\overline{\tau_{\mathrm{off}}}$ (s)	$\overline{\tau_{\mathrm{on}}}$ (s)
Adenosine 5'-monophosphate (AMP)	62.52±0.75 (AMP1)	4.57±0.14 (AMP1)	0.01±0.01 (AMP1)	0.31±0.1
	31.78±0.19 (AMP2)	6.11±0.03 (AMP2)	0.05±0.01 (AMP2)	
Uridine 5'-monophosphate (UMP)	45.34±0.07 (UMP1)	1.57±0.08 (UMP1)	0.01±0.01 (UMP1)	0.03±0.02
	28.45±0.14 (UMP2)	5.47±0.1 (UMP2)	0.04±0.02 (UMP2)	
Cytidine 5'-monophosphate (CMP)	52.71±0.82 (CMP1)	3.14±0.05 (CMP1)	0.01±0.01 (CMP1)	0.08±0.02
	28.43±0.24 (CMP2)	4.86±0.04 (CMP2)	0.03±0.01 (CMP2)	
Guanosine 5'-monophosphate (GMP)	65.31±0.23 (GMP1)	3.36±0.06 (GMP1)	0.01±0.01 (GMP1)	0.06±0.02
	38.73±2.5 (GMP2)	7.88±0.1 (GMP2)	0.05±0.02 (GMP2)	

Supplementary Table 5. The statistics of core parameters of events acquired with four saccharides. Nanopore measurements were carried out as described in **Supplementary Fig. 16** in 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). A transmembrane voltage of +160 mV was continually applied during the measurements. All saccharides were separately added to *cis* with a final concentration of 4 mM (IdoA), 20 mM (Fru), and 60 mM (Ara and GlcNAc). The final concentration of Ara and GlcNAc was set higher as a result of the poor event appearance rate specifically for the two saccharide types. The definition and derivation of the characteristic parameters were summarized in **Supplementary Fig. 7**. Binding of Fru, Ara and GlcNAc both produce multiple types of events (**Supplementary Fig. 16**) so that the $\overline{\Delta I}$, $\overline{\mathrm{SD}}$ and $\overline{\tau_{\mathrm{off}}}$ of different event types were derived separately. Three independent measurements (N = 3) were performed for each analyte type to obtain the statistics.

Saccharide	$\overline{\Delta I}$ (pA)	$\overline{\mathbf{SD}}$ (pA)	$\overline{\tau_{\mathbf{off}}}$ (s)	$\overline{\tau_{\mathbf{on}}}$ (s)
Iduronic acid (IdoA)	29.24±0.43	4.72±0.03	0.5±0.13	1.35±0.09
D-fructose (Fru)	66.04±0.68 (Fru1)	6.6±0.61 (Fru1)	0.36±0.49 (Fru1)	0.42±0.17
	80.27±0.29 (Fru2)	8.71±0.22 (Fru2)	0.43±0.1 (Fru2)	
N-acetyl-D-glucosamine (GlcNAc)	47.32±1.34 (GlcNAc1)	5.76±0.29 (GlcNAc1)	0.003±0.002 (GlcNAc1)	0.67±0.09
	85.25±0.15 (GlcNAc2)	5.75±0.22 (GlcNAc2)	0.003±0.001 (GlcNAc2)	
L-arabinose (Ara)	55.21±0.75 (Ara1)	3.11±0.3 (Ara1)	0.01±0.01 (Ara1)	0.19±0.02
	60.46±0.25 (Ara2)	5.21±0.23 (Ara2)	0.24±0.02 (Ara2)	
	56.51±5.31 (Ara3)	10.79±0.45 (Ara3)	0.26±0.11 (Ara3)	

Supplementary Table 6. The accuracy of different models for twenty-one proteinogenic amino acids, four canonical nucleoside monophosphates and four saccharides classification performed with multi-parametric machine learning. The procedure of machine learning was demonstrated in **Supplementary Fig. 20**. Sensing events of twenty-one proteinogenic amino acids (including 20 commonly seen proteinogenic amino acids and selenocysteine), four nucleoside monophosphates and four saccharides were collected to form a labelled dataset. The dataset was split into a training set (80%) and a testing set (20%) for model training and model testing. Nine parameters including ΔI, SD, LocalSD Ratio, min, max, range, t_{off}, skew and kurt were employed as event features. All models were evaluated with ten-fold cross-validation and the corresponding validation accuracy and test accuracy were reported. The bagged trees model scored the best validation accuracy of 98.9% (marked in red), and was selected for further use.

Model		Validation Accuracy	Test Accuracy
Ensemble	Boosted Trees	93.5	93.1
	Bagged Trees	98.9	99.1
	Subspace Discriminant	84.6	83.5
	Subspace KNN	82.8	80.0
	RUSBoost Trees	80.9	77.5
Support Vector Machine (SVM)	Linear SVM	97.3	96.4
	Cubic SVM	97.7	97.0
	Fine Gaussian SVM	78.5	74.9
	Medium Gaussian SVM	95.9	95.4
	Quadratic SVM	97.9	97.1
	Coarse Gaussian SVM	94.6	93.7
Decision Trees	Fine Tree	98.5	98.2
	Medium Tree	59.1	58.2
	Coarse Tree	18.5	18.5
Naïve Bayes	Gaussian Naïve Bayes	94.8	94.5
	Kernel Naïve Bayes	96.3	95.6
Neural Network	Narrow Neural work	97.7	97.2
	Medium Neural work	96.8	96.1
	Wide Neural work	96.9	96.0
	Bilayered Neural work	97.6	96.6
	Trilayered Neural work	97.2	96.5
Discriminant Analysis	Linear Discriminant	83.4	83.4
	Quadratic Discriminant	Fail	Fail
K Nearest Neighbor (KNN)	Fine KNN	89.7	87.9
	Medium KNN	86.8	86.0
	Coarse KNN	75.3	75.0
	Cosine KNN	81.3	80.8
	Cubic KNN	83.9	84.2
	Weighted KNN	89.4	87.2

Supplementary Table 7. The accuracy of machine learning for 24 amino acids, 7 nucleoside monophosphates, 4 saccharides, and 5 peptides. The workflow of machine learning was demonstrated in **Extended Data Fig. 6a** and **Methods**. A feature matrix containing sensing events of 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) was collected to form an input dataset. The dataset was split into a training set (80%) and a testing set (20%) for model training and model testing. Nine parameters including ΔI, SD, LocalSD Ratio, min, max, range, t_{off}, skew and kurt were employed as event features (**Supplementary Fig. 7**). All models were evaluated with the 10-fold cross-validation accuracies, in which bagged trees model performed best with a validation accuracy of 98.7% (marked in red).

Model		Validation Accuracy	Test Accuracy
Ensemble	Boosted Trees	84.0	84.2
	Bagged Trees	98.7	96.7
	Subspace Discriminant	79.6	78.1
	Subspace KNN	82.9	83.8
	RUSBoost Trees	86.8	83.3
Support Vector Machine (SVM)	Linear SVM	96.6	95.3
	Cubic SVM	97.0	96.0
	Fine Gaussian SVM	74.3	69.6
	Medium Gaussian SVM	95.0	93.7
	Quadratic SVM	97.2	96.0
	Coarse Gaussian SVM	93.3	92.5
Decision Trees	Fine Tree	97.1	96.0
	Medium Tree	54.5	53.3
	Coarse Tree	15.3	15.3
Naïve Bayes	Gaussian Naïve Bayes	93.1	92.5
	Kernel Naïve Bayes	95.4	94.4
Neural Network	Narrow Neural work	97.3	94.9
	Medium Neural work	96.5	95.6
	Wide Neural work	96.2	95.0
	Bilayered Neural work	96.8	95.3
	Trilayered Neural work	95.9	96.3
Discriminant Analysis	Linear Discriminant	79.1	77.5
	Quadratic Discriminant	Fail	Fail
K Nearest Neighbor (KNN)	Fine KNN	84.2	81.6
	Medium KNN	81.6	81.1
	Coarse KNN	70.2	70.5
	Cosine KNN	75.3	75.4
	Cubic KNN	78.2	77.7
	Weighted KNN	84.6	83.0

Supplementary Table 8. The ImportanceMean values of different event parameters. The ImportanceMean values of nine event parameters (Δ*I*, SD, LocalSD Ratio, min, max, range, t_{off}, skew and kurt) were calculated for the Bagged Trees model (**Extended Data Fig. 6b** and **Supplementary Table 7**) used to distinguish all 40 analytes, using the Permutation Feature Importance method. A higher ImportanceMean value indicates that a parameter plays a more critical role in the model's ability to distinguish analytes. The ImportanceStandardDeviation quantifies the variability in the importance of event parameters.

Parameters	ImportanceMean	ImportanceStandardDeviation
Δ*I*	0.8684	0.0045
SD	0.5270	0.0044
LocalSD Ratio	0.1163	0.0020
t_{off}	0.0617	0.0028
max	0.0115	0.0012
min	0.0064	0.0010
range	0.0040	8.2033×10^{-4}
kurt	0.0029	5.8286×10^{-4}
skew	0.0013	2.5816×10^{-4}

Supplementary Table 9. Accuracy of distinguishing 29 analytes using the 4-feature model and the 9-feature model. Sensing events of twenty-one proteinogenic amino acids (including 20 commonly seen proteinogenic amino acids and selenocysteine), four nucleoside monophosphates and four saccharides were collected to form a labelled dataset.

Bagged Trees Model	Validation Accuracy	Test Accuracy
4-feature model	0.9882	0.9843
9-feature model	0.9887	0.9898

Supplementary Table 10. Accuracy of distinguishing 40 analytes using the 4-feature model and the 9-feature model. 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 a labelled dataset.

Bagged Trees Model	Validation Accuracy	Test Accuracy
4-feature model	0.9828	0.9585
9-feature model	0.9870	0.9677

Supplementary Table 11. The average *F* values of the analytes measured at different concentrations. All measurements were performed using MspA-FPBA. The analytes were added to *cis* with the desired concentrations. A +160 mV bias was continually applied. All statistical results were derived from results of three independent measurements (*N* = 3).

Analytes	Concentration (μM)	*F* (min^{-1})
Met	30	4.8±0.43
	60	11.75±0.58
	120	19±0.22
	250	39.84±1.82
	400	72.17±4.08
	600	96.68±3.84
Asn	30	8.35±0.51
	60	16.85±0.42
	120	28.94±1.65
	250	57.55±3.82
	400	97.27±2.9
	600	122.08±3.14
Tyr	30	4.06±0.82
	60	8.84±1.41
	120	12.41±2.13
	250	27.64±0.61
	400	42.59±3.41
	600	59.7±2.27
Arg	30	10.58±2.36
	60	13.26±0.94
	120	22.08±1.8
	250	45.87±0.78
	400	64.18±1.24
	600	88.41±2.13
UMP	120	8.62±0.62
	250	12.32±1.18
	500	17.43±0.93
	800	23.61±0.55
	1100	29.48±0.63
	1400	38.03±0.73
Fru	600	7.3±0.21
	1200	13.53±0.4
	2400	18.24±0.52
	4800	35.07±1.53
	7500	46.79±1.02
	10000	58.56±2.71

LRG	5	4.06±0.49
	10	10.19±1.33
	15	14.59±1.08
	20	23.00±1.05
	25	28.29±0.97
	30	36.46±0.72

Supplementary Table 12. Coefficients (*a* and *b*) of different analytes. The value of *a* and *b* were derived from results of linear fitting of the concentration dependence assay as shown in **Supplementary Figs. 30-36**. All measurements were carried out in 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). A transmembrane voltage of +160 mV was continually applied during the measurements.

Analytes	***a***	***b***
Met	0.36	0.16
Asn	3.2	0.21
Tyr	1.76	0.1
Arg	8.99	0.14
UMP	5.97	0.02
Fru	4.37	0.01
LRG	-2.6	1.28

Supplementary Table 13. The limit of detection (LOD) for amino acids. The limit of detection (LOD) is defined as the minimum concentration required to detect 5 events within 10 min of continuous measurement. In separate measurements, different amino acids were respectively added to *cis*. Nanopore measurements were carried out in 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). A transmembrane voltage of +160 mV was continually applied during the measurements.

Amino acids	LOD (μM)
A	20
D	10
E	25
F	15
G	10
H	5
I	5
K	15
L	10
M	15
N	10
P	20
Q	10
R	5
S	20
T	10
U	10
V	15
W	10
Y	5
C	10
Ac-K	20
Me-R	20
P-Y	15

Supplementary Table 14. The limit of detection (LOD) for NMPs. The limit of detection (LOD) is defined as the minimum concentration required to detect 5 events within 10 min of continuous measurement. In separate measurements, different NMPs were respectively added to *cis*. Nanopore measurements were carried out in 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). A transmembrane voltage of +160 mV was continually applied during the measurements.

NMPs	LOD (μM)
AMP	10
UMP	15
CMP	20
GMP	30
ψ	20
m^6A	20
IMP	20

Supplementary Table 15. The limit of detection (LOD) for monosaccharides. The limit of detection (LOD) is defined as the minimum concentration required to detect 5 events within 10 min of continuous measurement. In separate measurements, different monosaccharides were respectively added to *cis*. Nanopore measurements were carried out in 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). A transmembrane voltage of +160 mV was continually applied during the measurements.

Monosaccharides	LOD (μM)
Ara	250
Fru	150
IdoA	100
GlcNAc	200

Supplementary Table 16. The limit of detection (LOD) for peptides. The limit of detection (LOD) is defined as the minimum concentration required to detect 5 events within 10 min of continuous measurement. In separate measurements, different peptides were respectively added to *cis*. Nanopore measurements were carried out in 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). A transmembrane voltage of +160 mV was continually applied during the measurements.

Peptides	LOD (μM)
IK	4
LRG	2
SLR	6
GYIK	5
LRGY	1

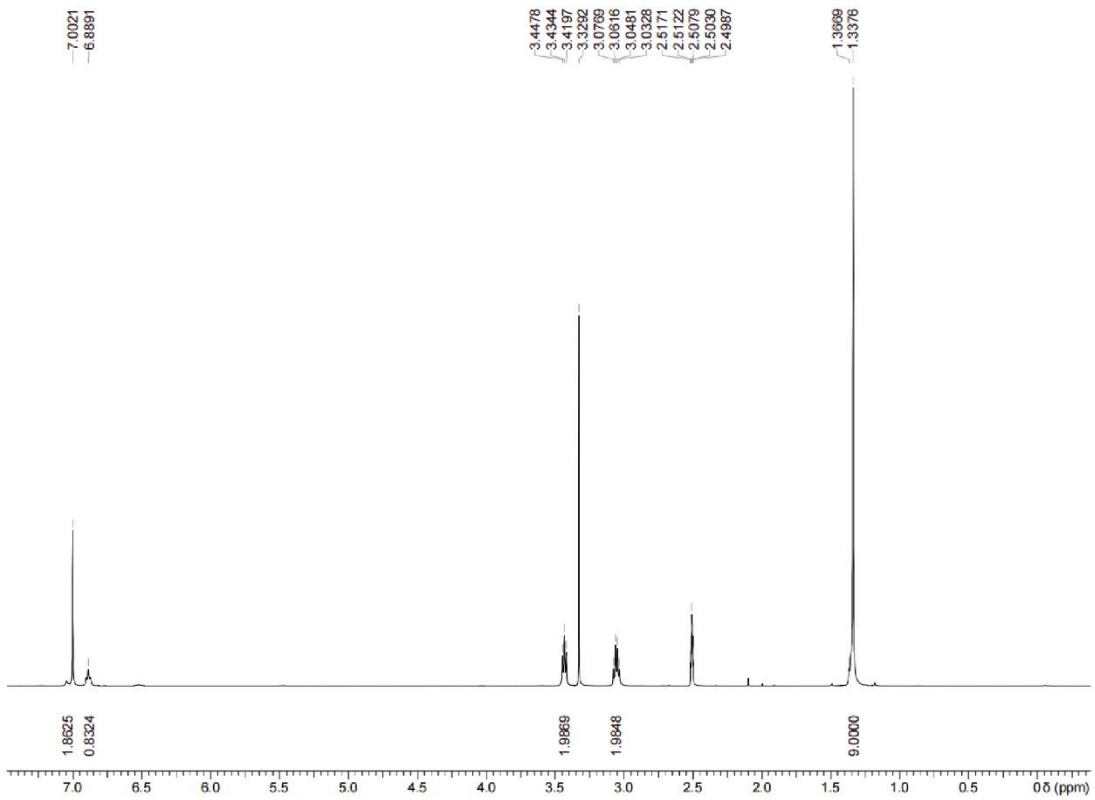
Supplementary Table 17. Construction of the prediction model for amino acids, nucleoside monophosphates, saccharides, and peptides. The workflow of few-shot learning was demonstrated in **Fig. 4a**. The previously sensed events of 21 proteinogenic amino acids, 4 canonical NMPs, 4 saccharides and 5 peptides are respectively integrated, thereby forming the support set for machine learning. The dataset was split into a training set (80%) and a testing set (20%) for model training and model testing. Nine parameters including ΔI, SD, LocalSD Ratio, min, max, range, t_{off}, skew and kurt were employed as event features (**Supplementary Fig. 7**). All models were evaluated with the 10-fold cross-validation accuracies, in which bagged trees model performed best with a validation accuracy of 98.4% (marked in red).

Model		Validation Accuracy	Test Accuracy
Ensemble	Boosted Trees	98.2	98.4
	Bagged Trees	98.4	97.9
	Subspace Discriminant	85.2	85.4
	Subspace KNN	93.9	94.6
	RUSBoost Trees	96.9	97.3
Support Vector Machine (SVM)	Linear SVM	95.0	93.9
	Cubic SVM	96.8	96.5
	Fine Gaussian SVM	93.7	94.9
	Medium Gaussian SVM	96.3	96.3
	Quadratic SVM	96.1	96.0
	Coarse Gaussian SVM	93.0	92.3
Decision Trees	Fine Tree	97.6	97.1
	Medium Tree	95.8	96.8
	Coarse Tree	87.4	87.9
Naïve Bayes	Gaussian Naïve Bayes	86.3	84.9
	Kernel Naïve Bayes	91.2	90.8
Neural Network	Narrow Neural work	97.2	96.3
	Medium Neural work	97.8	97.7
	Wide Neural work	97.9	98.2
	Bilayered Neural work	97.9	97.7
	Trilayered Neural work	97.7	97.3
Discriminant Analysis	Linear Discriminant	85.3	86.6
	Quadratic Discriminant	Fail	Fail
K Nearest Neighbor (KNN)	Fine KNN	96.5	97.0
	Medium KNN	95.3	95.6
	Coarse KNN	91.3	90.8
	Cosine KNN	95.0	94.3
	Cubic KNN	94.8	94.6
	Weighted KNN	96.0	96.8

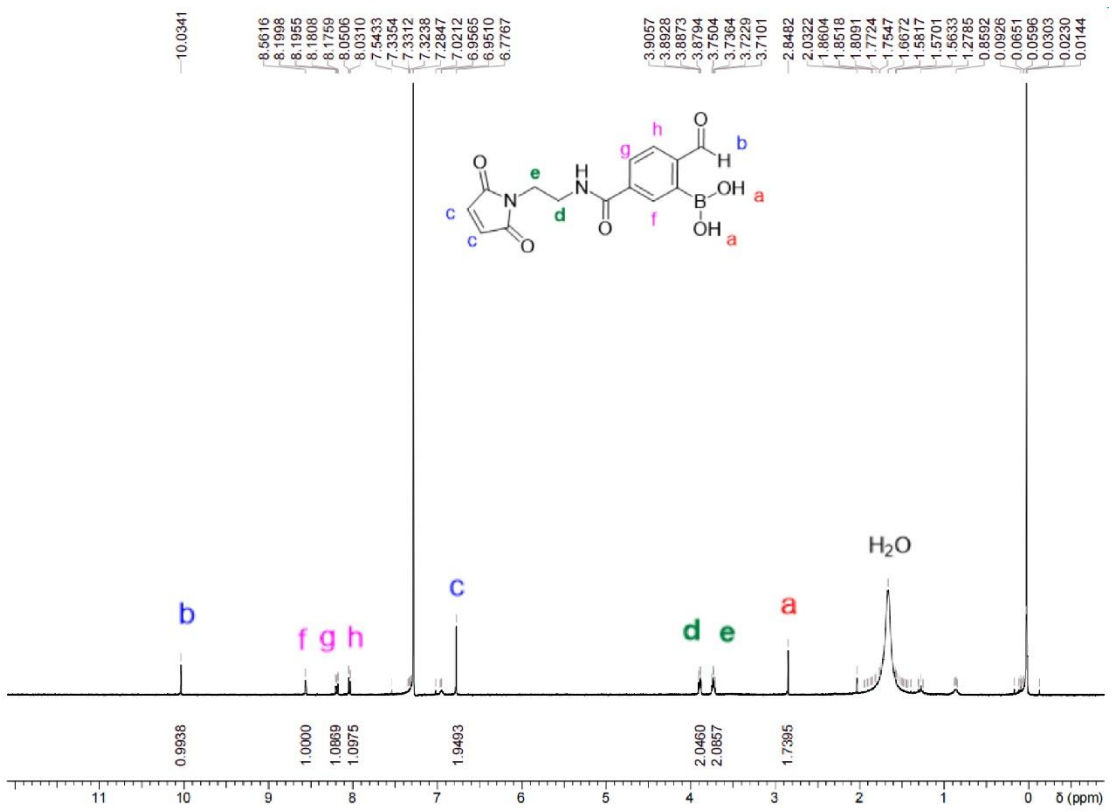
Supplementary Table 18. Calibration coefficients of different amino acids and saccharides. The calibration coefficient (δ) is defined as the number of analyte binding events occurring per unit concentration per min. The value was acquired during measurements with a sole analyte. The final concentration of each amino acid in *cis* was 400 μM. The final concentration of each saccharide in *cis* was 10 mM. A + 160 mV voltage was continually applied during the measurements. $\bar{\delta}$ is the mean value of δ from three independent measurements.

Amino acids and saccharides	$\bar{\delta}$ ($\mu M^{-1} * min^{-1}$)
K	0.0625±0.0112
V	0.0828±0.0107
A	0.0722±0.0095
N	0.1249±0.0144
T	0.1508±0.0271
Gal	0.0204±0.0075
Man	0.0481±0.0059
GlcNAc	0.0127±0.0083

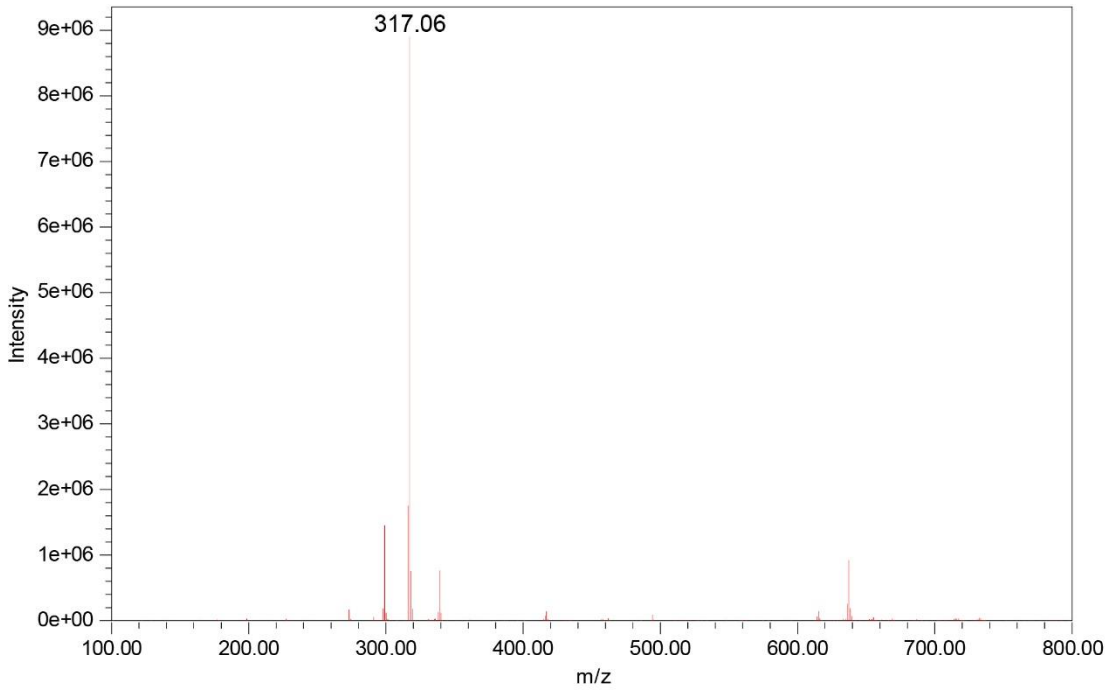
Supplementary Figures



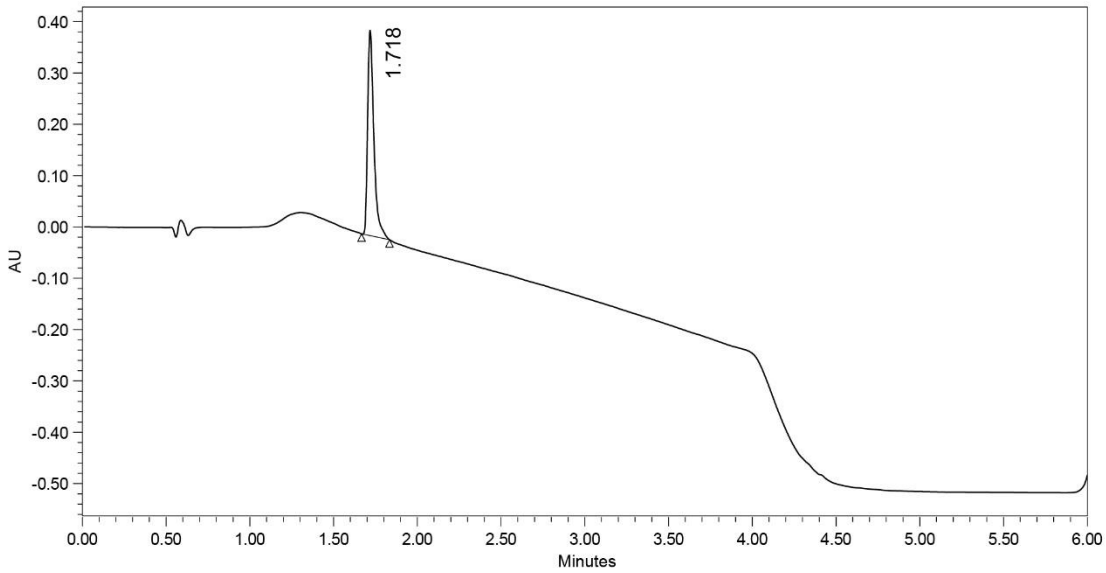
Supplementary Fig. 1: ^{1}H NMR spectrum (400 MHz) of compound 2 in $CDCl_3$.



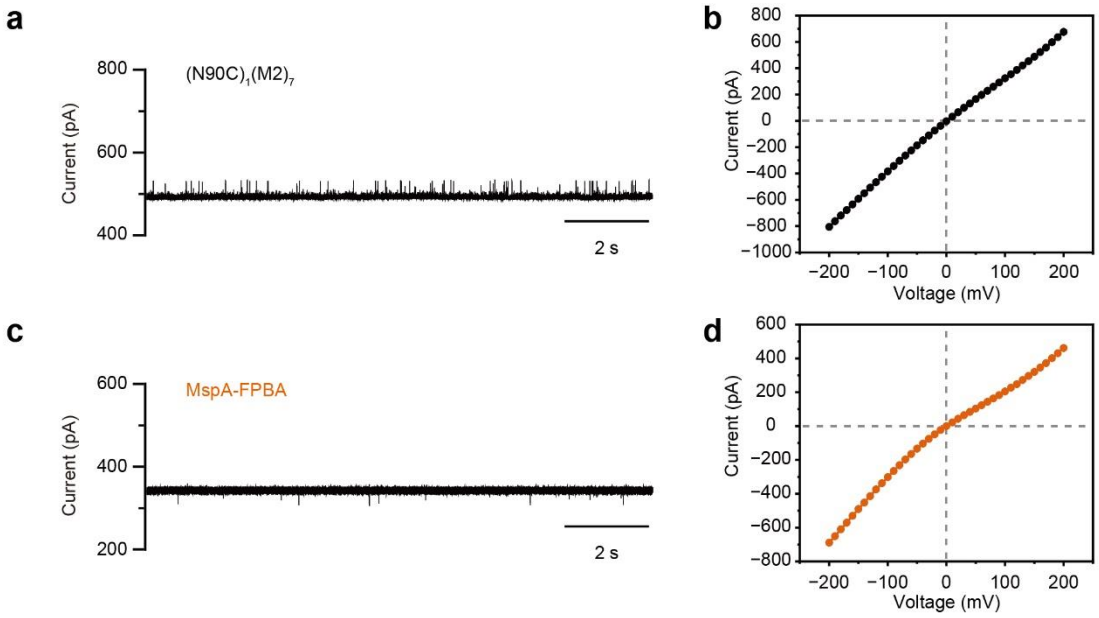
Supplementary Fig. 2: ^{1}H NMR spectrum (400 MHz) of maleimido-C2-FPBA in $CDCl_3$.



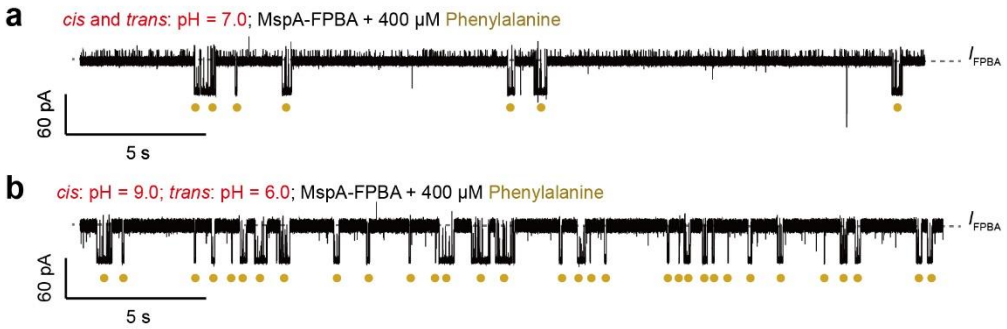
Supplementary Fig. 3: LCMS spectrum of maleimido-C2-FPBA.



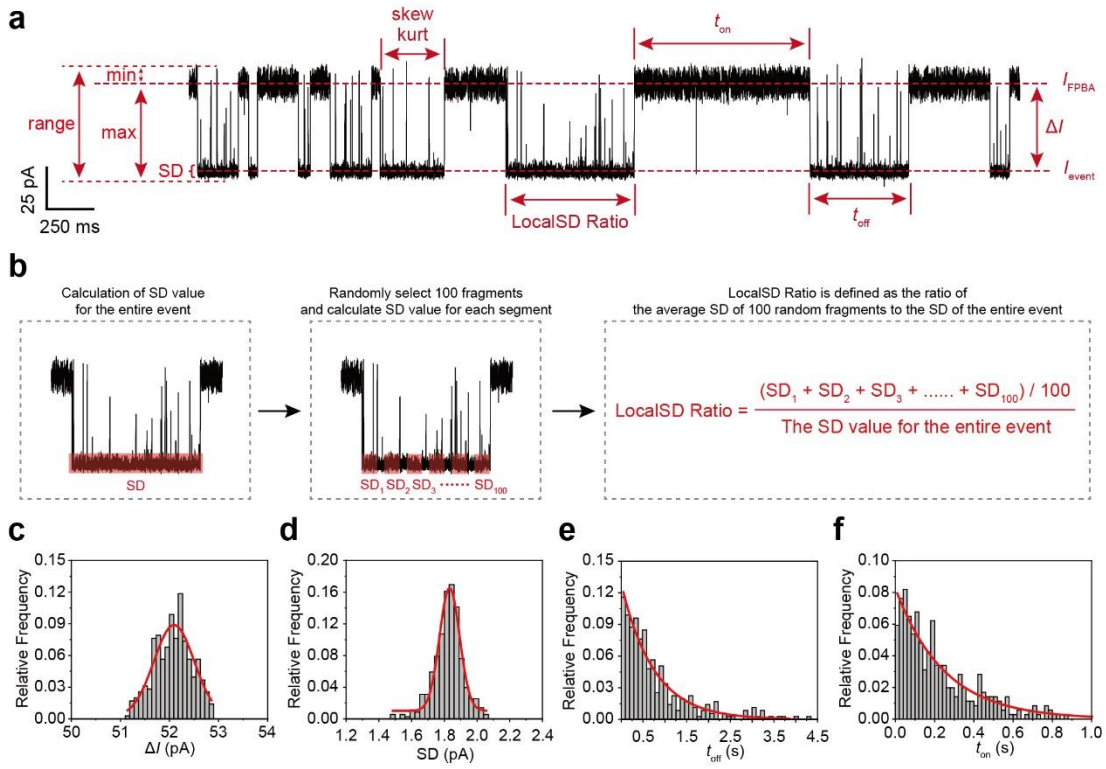
Supplementary Fig. 4: HPLC spectrum of maleimido-C2-FPBA.



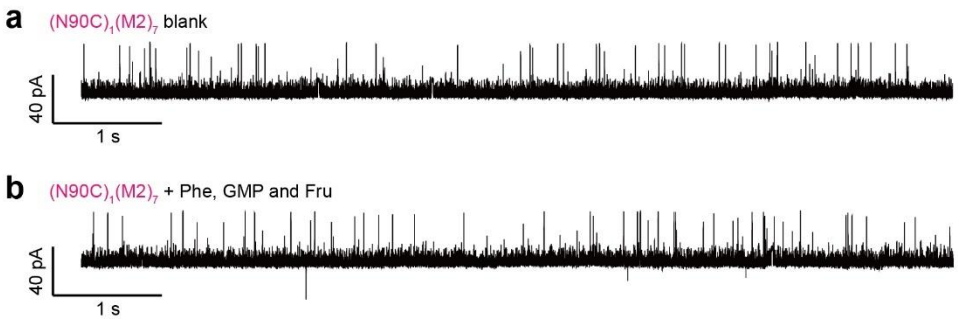
Supplementary Fig. 5: Single molecule characterization of $(N90C)_1(M2)_7$ and MspA-FPBA. The measurements were carried out as described in **Methods** in a 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0). A potential of +160 mV was continually applied. (**a, c**) Representative trace segments respectively acquired with (**a**) $(N90C)_1(M2)_7$ and (**c**) MspA-FPBA were demonstrated on the left. (**b, d**) The I-V curves acquired with (**b**) $(N90C)_1(M2)_7$ and (**d**) MspA-FPBA were demonstrated to the right of each corresponding trace. The error bars of the I-V curves were derived from three independent measurements ($N = 3$). Detailed results were listed in **Supplementary Table 2**.



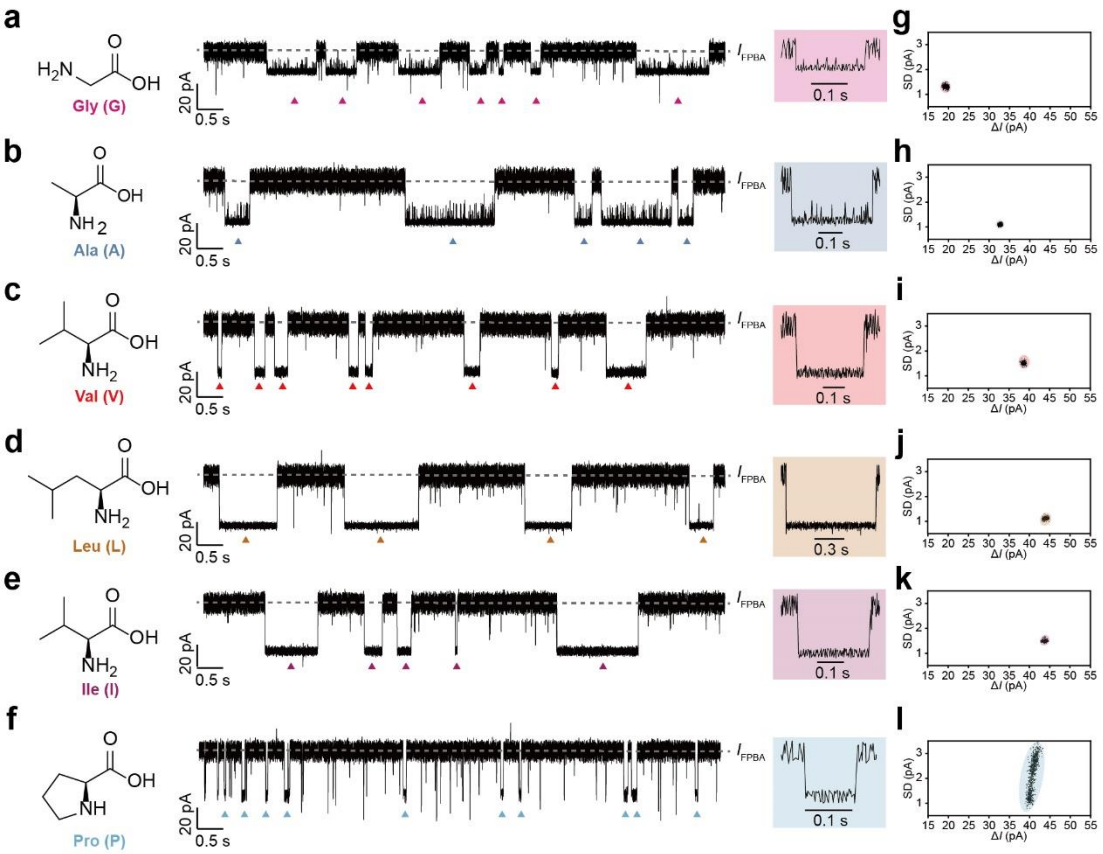
Supplementary Fig. 6: The event capture frequency and the environment pH. The measurements were carried out as described in **Methods**. A transmembrane voltage of +160 mV was continually applied. Phenylalanine was used as a model analyte and added to *cis* with a final concentration of 400 μM. (**a**) The representative trace of phenylalanine sensing performed at the same pH on both the *cis* and *trans* sides (*cis* and *trans*: 1.5 M KCl, 10 mM MOPS, pH 7.0). Phenylalanine event identities were labeled with yellow-coded dots. (**b**) The representative trace of phenylalanine sensing performed when there was a pH difference between the *cis* and *trans* sides (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0). Phenylalanine event identities were labeled with yellow-coded dots. I_{FPBA} represents the open pore current of MspA-FPBA. The capture frequency of amino acid events increased when a pH difference existed between the *cis* and *trans* sides.



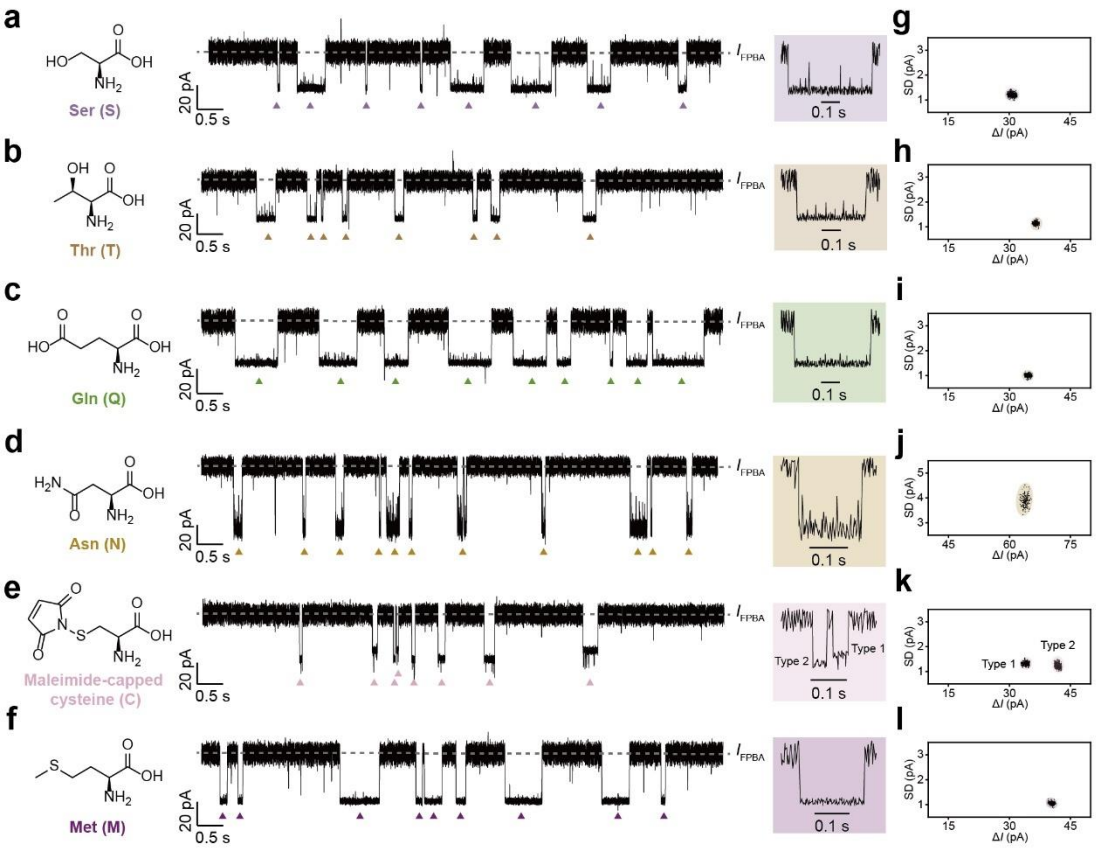
Supplementary Fig. 7: The definition of event parameters. (**a**) A representative trace. The trace was acquired with MspA-FPBA and phenylalanine was treated as a model analyte. I_{FPBA} stands for the open pore current of MspA-FPBA and I_{event} stands for the blockage level when a single amino acid is bound to nanopore. ΔI is defined as $\Delta I = I_{\mathrm{FPBA}} - I_{\mathrm{event}}$. SD is the standard deviation of the event noise. t_{off} represents the dwell time of the event and t_{on} represents the inter-event duration. For demonstration, all above parameters are respectively marked on the trace. (**b**) The definition of LocalSD Ratio. The entire event is randomly divided into 100 segments with a t_{off} of 2 ms and SD values are calculated separately. LocalSD Ratio is defined as the ratio of the average SD of 100 random fragments to the SD of the entire event. The fluctuation of different events can be evaluated through LocalSD Ratio. (**c-d**) The histogram of Δ*I* (**c**) and SD (**d**) generated by results from a continually recorded trace as demonstrated in (**a**). Gaussian fittings were performed to derive the mean Δ*I* ($\overline{\Delta I}$) and the mean SD ($\overline{\mathrm{SD}}$). (**e-f**) The histograms of t_{off} (**e**) and t_{on} (**f**) generated by results from a continually recorded trace as demonstrated in (**a**). The mean dwell time (τ_{off}) and the mean inter-event duration (τ_{on}) were derived from single-exponential fitting results, following the equation $y = a * \exp(-x/\tau)$.



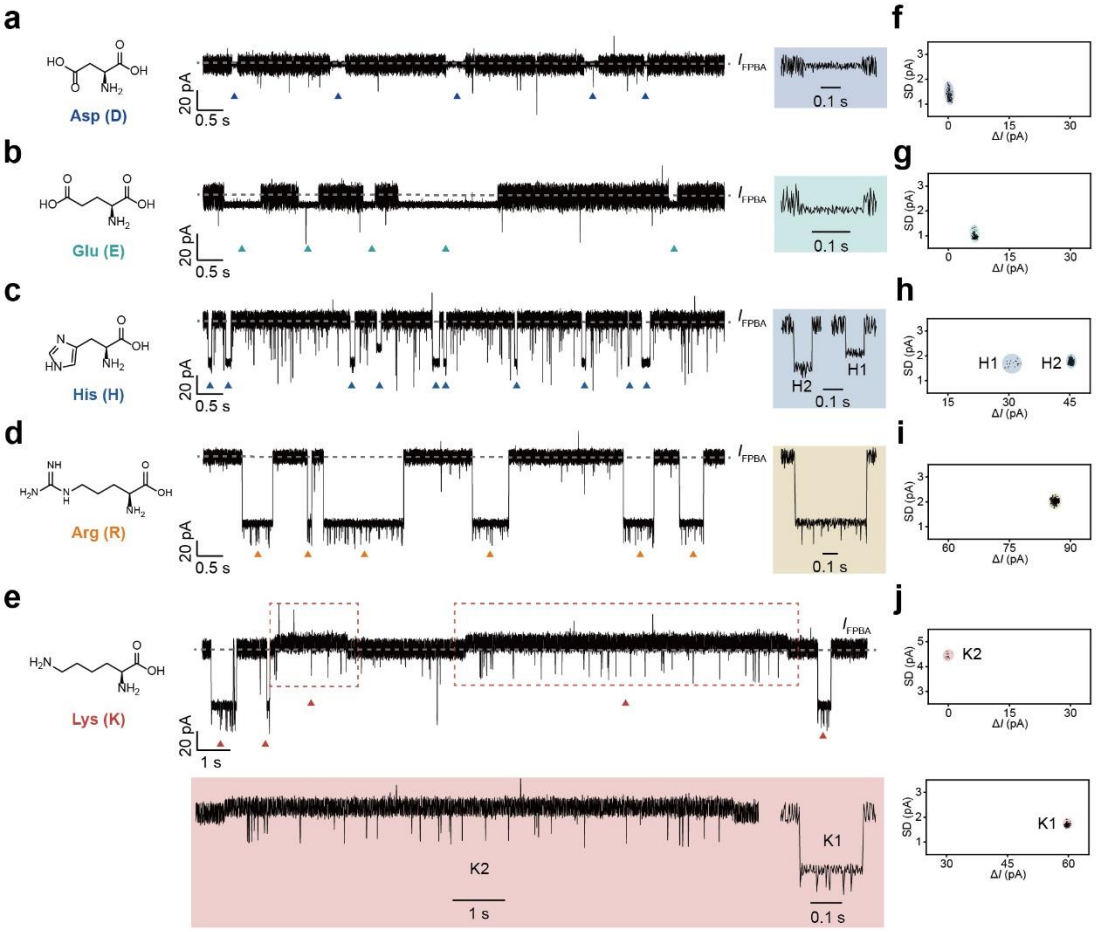
Supplementary Fig. 8: Amino acid, nucleoside monophosphate and saccharide sensing measured by $(N90C)_1(M2)_7$. The measurements were carried out in a 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0). A voltage of +160 mV was continually applied. Phenylalanine (Phe), guanosine 5'-monophosphate (GMP), and D-fructose (Fru) are respectively employed as the model amino acid, nucleoside monophosphate and saccharide. Phe, GMP and Fru were simultaneously added to the *cis* side at final concentrations of 50 μM, 1 mM and 5 mM, respectively. (**a-b**) Representative traces acquired with $(N90C)_1(M2)_7$ in the presence of no analyte (**a**) or a mixture of Phe, GMP, and Fru (**b**). No binding events were observed. Only background noises were observed with $(N90C)_1(M2)_7$, as also demonstrated in **Supplementary Fig. 5**. The addition of Phe, GMP, and Fru did not lead to reports of any events. These results confirm that the absence of adapter modification at the pore constriction prevents the generation of amino acid, nucleoside monophosphate and saccharide sensing events.



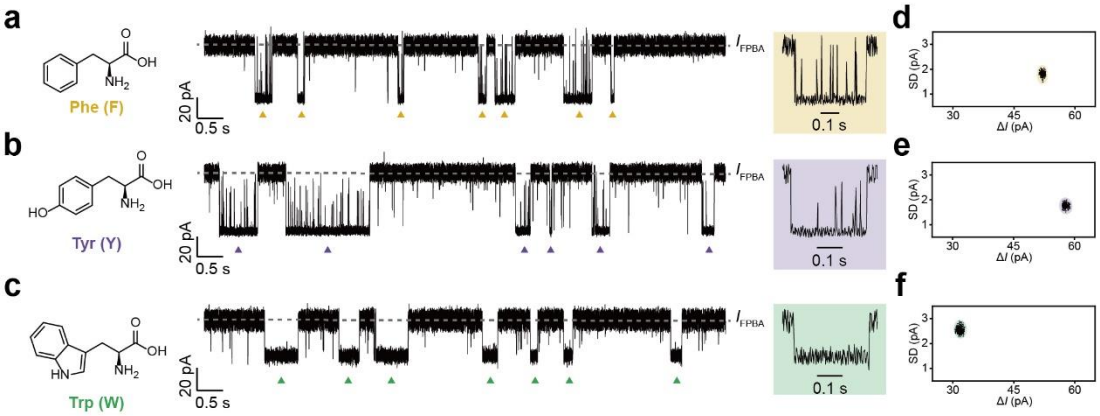
Supplementary Fig. 9: Events of nonpolar aliphatic amino acids. The measurements were carried out as described in **Methods**. A 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0) was used. A transmembrane voltage of +160 mV was continually applied. (**a-f**) Left: Chemical structures of nonpolar aliphatic amino acids. Middle: Representative traces containing events caused by G (**a**), A (**b**), V (**c**), L (**d**), I (**e**) and P (**f**) binding. Each amino acid was separately added to the *cis* chamber with a final concentration of 400 μM. I_{FPBA} represents the open pore current of MspA-FPBA. Right: Representative events of the corresponding amino acids. (**g-l**) The scatter plot of Δ*I* versus SD respectively derived from sensing results of G (**g**, n = 424), A (**h**, n = 202), V (**i**, n = 223), L (**j**, n = 371), I (**k**, n = 237) or P (**l**, n = 722).



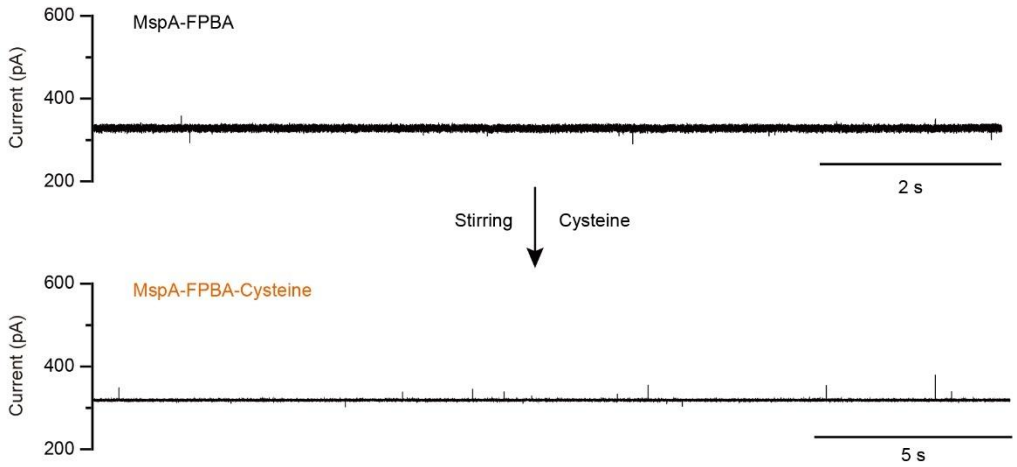
Supplementary Fig. 10: Events of polar uncharged amino acids. The measurements were carried out as described in **Methods**. A 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0) was used. A transmembrane voltage of +160 mV was continually applied during the measurements. (**a-f**) Left: Chemical structures of polar uncharged amino acids. Middle: Representative traces containing events caused by S (**a**), T (**b**), Q (**c**), N (**d**), C (**e**) and M (**f**) binding. Each amino acid was separately added to the *cis* chamber with a final concentration of 400 μM. I_{FPBA} represents the open pore current of MspA-FPBA. Cysteine was blocked with maleimide in the experiment to prevent ring formation and avoid low event occurrence rate (**Supplementary Fig. 14**). Right: Representative events of the corresponding amino acids events. (**g-l**) The scatter plot of ΔI versus SD respectively derived from sensing events of S (**g**, n = 856), T (**h**, n = 377), Q (**i**, n = 650), N (**j**, n = 201), C (**k**, n = 718) and M (**l**, n = 388).



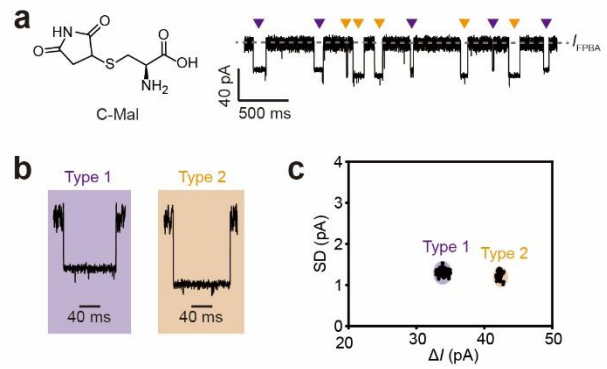
Supplementary Fig. 11: Events of charged amino acids. The measurements were carried out as described in **Methods**. A 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0) was used. A transmembrane voltage of +160 mV was continually used. (**a-b**) Left: Chemical structures of acidic amino acids. Middle: Representative current traces containing events caused by D (**a**) or E (**b**) binding. Each amino acid was separately added to *cis* with a final concentration of 400 μM. I_{FPBA} represents the open pore current of MspA-FPBA. Right: Representative events of the corresponding amino acids. (**c-e**) Left: Chemical structures of basic amino acids. Middle: Representative current traces acquired with H (**c**), R (**d**) and K (**e**) as the sole analyte. Each amino acid was separately added to *cis* with a final concentration of 400 μM. Right: Representative events of the corresponding amino acids. (**f-j**) The scatter plot of Δ*I* versus SD respectively generated from sensing results of D (**f**, n = 131), E (**g**, n = 122), H (**h**, n = 320), R (**i**, n = 467) and K (**j**, n = 155).



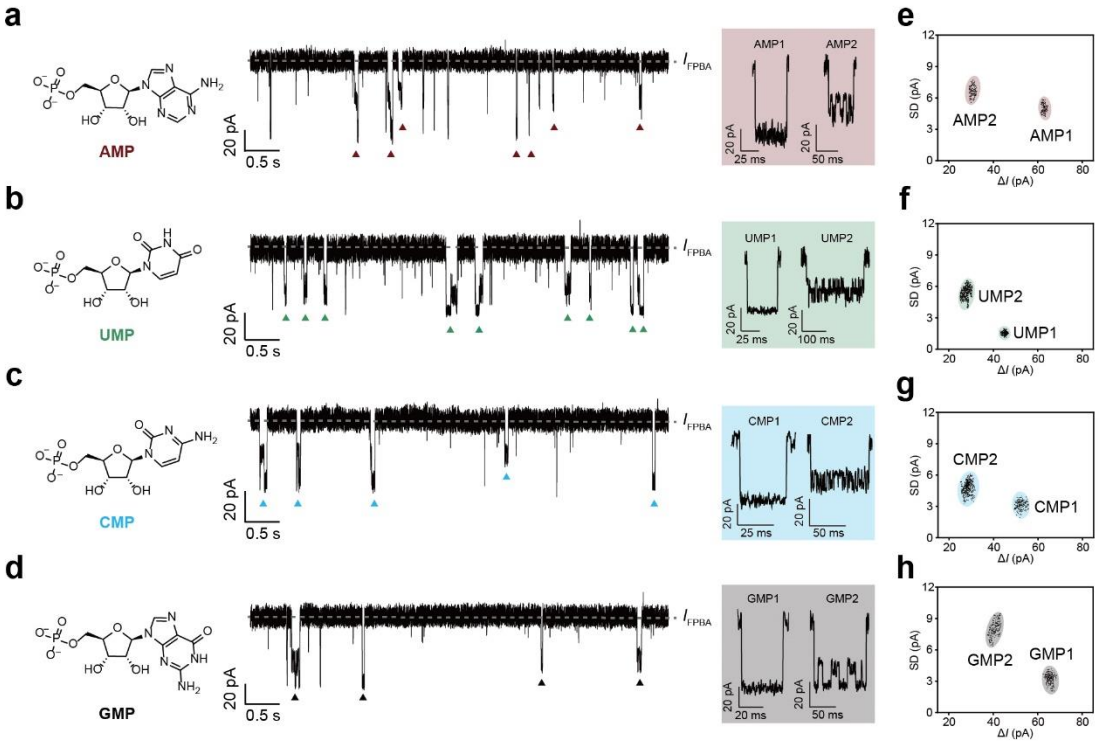
Supplementary Fig. 12: Events of aromatic amino acids. The measurements were carried out as described in **Methods**. A 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0) was used. A transmembrane voltage of +160 mV was continually applied. (**a-c**) Left: Chemical structures of aromatic amino acids. Middle: Representative current traces containing events caused by F (**a**), Y (**b**) and W (**c**) binding. Each amino acid was separately added to *cis* with a final concentration of 400 μM. I_{FPBA} represents the open pore current of MspA-FPBA. Right: Representative events of the corresponding amino acids. (**d-f**) The scatter plot of ΔI versus SD derived from sensing results of F (**d**, n = 354), Y (**e**, n = 413) and W (**f**, n = 743).



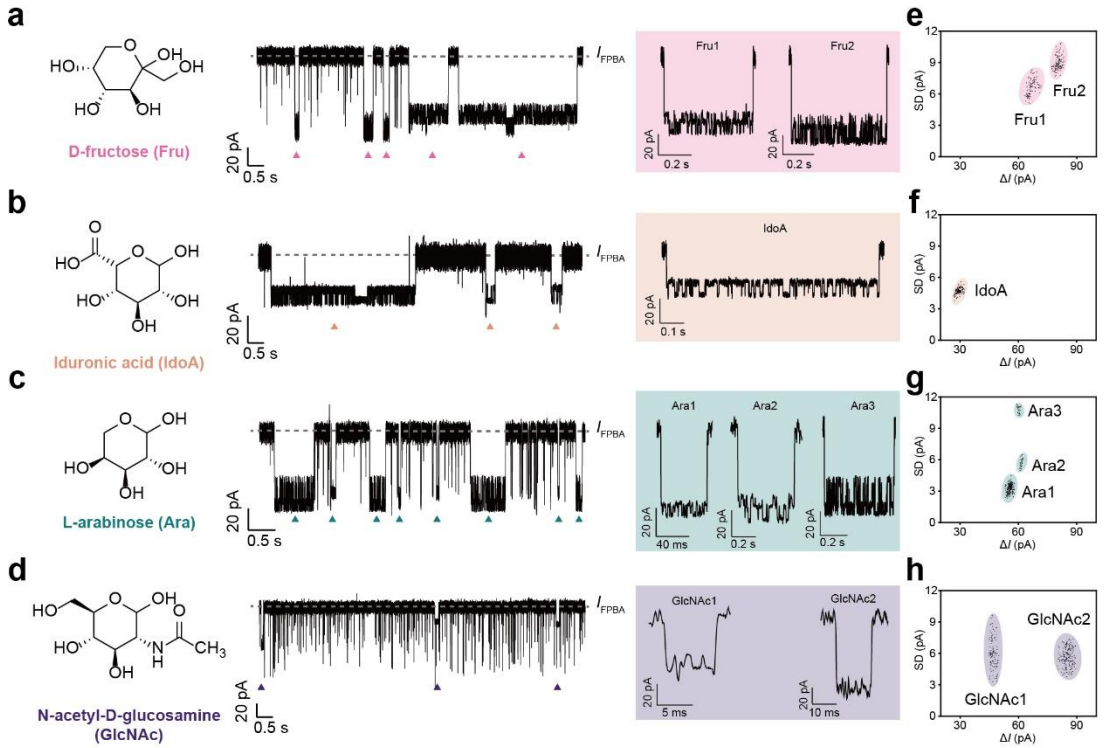
Supplementary Fig. 13: The extremely prolonged durations of cysteine events. The measurements were carried out as described in **Methods**. A 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0) was used. A transmembrane voltage of +160 mV was continually applied. Representative trace segments respectively acquired with (**above**) MspA-FPBA and (**bottom**) MspA-FPBA-Cysteine were demonstrated. Cysteine was added to the *cis* with a final concentration of 400 μM, and the extremely prolonged durations of cysteine events were observed immediately after stirring.



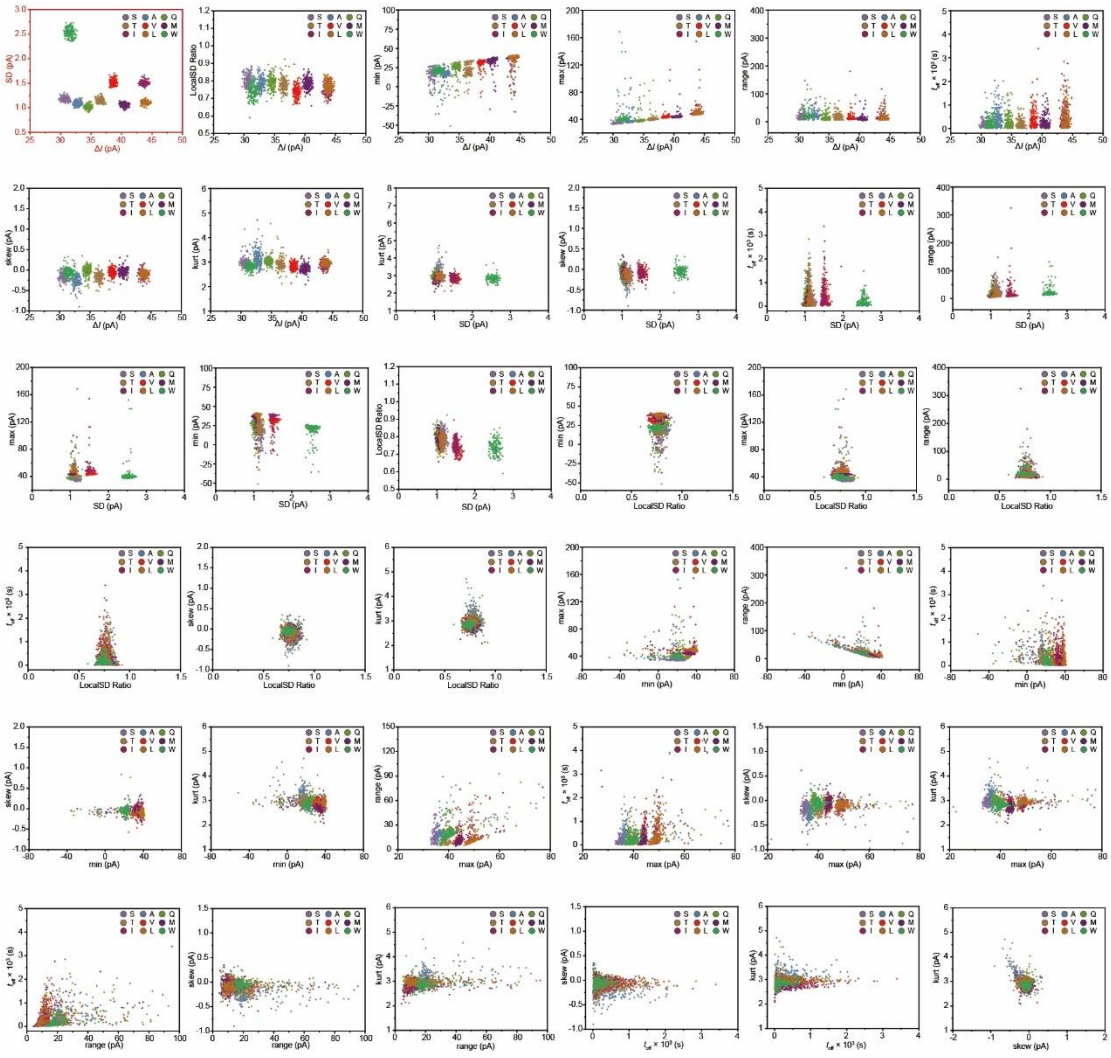
Supplementary Fig. 14: Events of maleimide-capped cysteine. The measurements were carried out as described in **Methods**. A 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0) was used. A transmembrane voltage of +160 mV was continually applied. (**a**) Left: Chemical structures of the maleimide-capped cysteine (C-Mal). Right: The representative current trace containing events caused by C-Mal binding. Type 1 and type 2 events with different blockage amplitudes were labeled with purple and orange coded triangles, respectively. C-Mal was added to *cis* with a final concentration of 400 μM. I_{FPBA} represents the open pore current of MspA-FPBA. (**b**) Representative type 1 and type 2 events of C-Mal. Type 1 and type 2 events with different blockage amplitudes were labeled with purple and orange rectangles, respectively. (**c**) The scatter plot of Δ*I* versus SD derived from sensing results of C-Mal (n = 200).



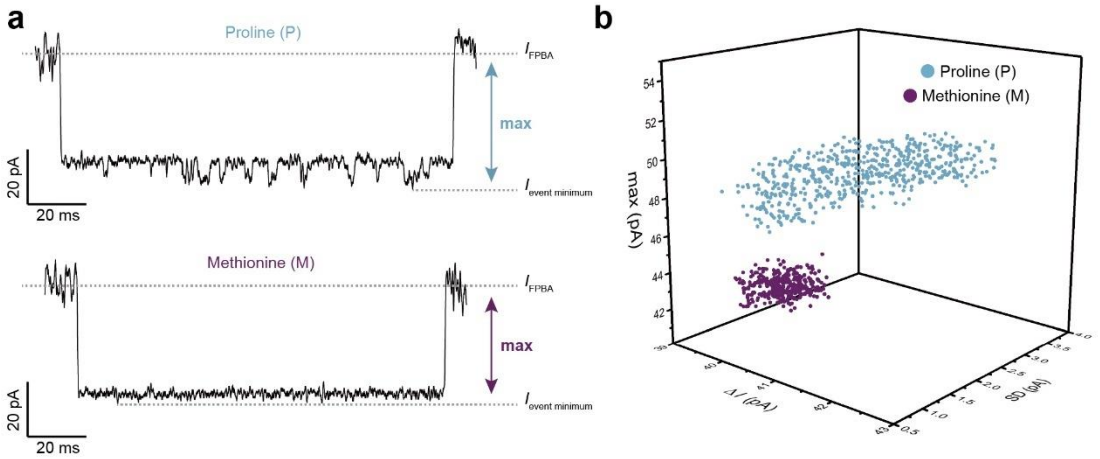
Supplementary Fig. 15: Events of nucleoside monophosphates. The measurements were carried out as described in **Methods**. A 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0) was used. A transmembrane voltage of +160 mV was continually applied. (**a-d**) Left: Chemical structures of nucleoside monophosphates. Middle: Representative current traces containing events caused by AMP (**a**), UMP (**b**), CMP (**c**) and GMP (**d**) binding. Each nucleoside monophosphate was separately added to *cis* with a final concentration of 2 mM. I_{FPBA} represents the open pore current of MspA-FPBA. Right: Representative events of the corresponding nucleoside monophosphates. (**e-h**) The scatter plot of Δ*I* versus SD derived from sensing results of AMP (**e**, n = 238), UMP (**f**, n = 644), CMP (**g**, n = 557) and GMP (**h**, n = 368).



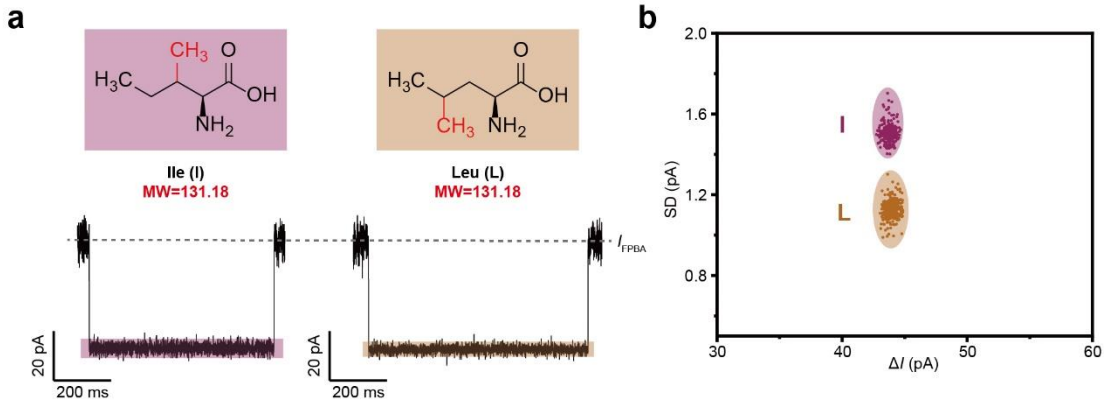
Supplementary Fig. 16: Events of saccharides. The measurements were carried out as described in **Methods**. A 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0) was used. A transmembrane voltage of +160 mV was continually applied. (**a-d**) Left: Chemical structures of saccharides. Middle: Representative current traces containing events caused by Fru (**a**), IdoA (**b**), Ara (**c**) and GlcNAc (**d**) binding. The final concentrations of saccharides were 4 mM (IdoA), 20 mM (Fru), and 60 mM (Ara and GlcNAc). I_{FPBA} represents the open pore current of MspA-FPBA. Right: Representative events of the corresponding saccharides. (**e-h**) The scatter plot of Δ*I* versus SD derived from sensing results of Fru (**e**, n = 197), IdoA (**f**, n = 134), Ara (**g**, n = 249) and GlcNAc (**h**, n = 253).



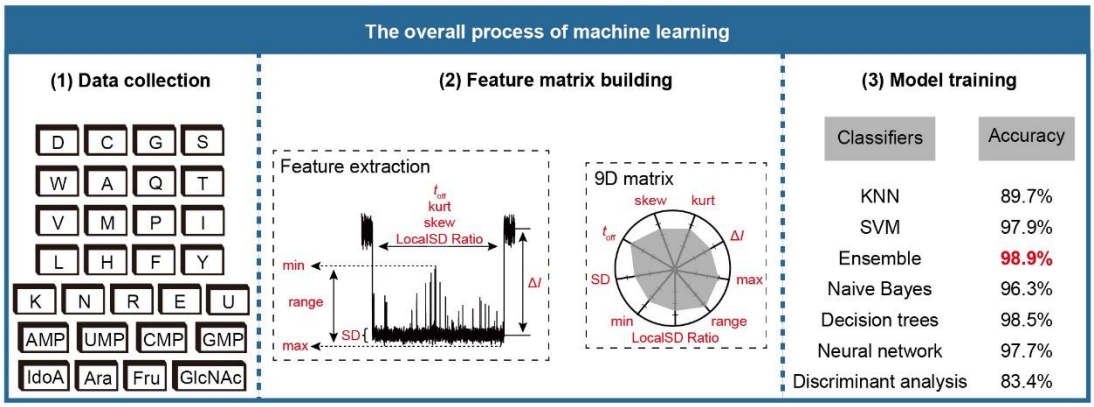
Supplementary Fig. 17: 2D scatter plots of different event parameters. Single-molecule events of the nine amino acids (S, A, Q, T, V, M, I, L, and W) were analyzed using event parameters including Δ*I*, SD, LocalSD Ratio, min, max, range, t_{off}, skew, and kurt. In the event scatter plot of Δ*I* versus SD (marked by a red box), the 2D visualization distinctly separates nine event clusters, with each cluster corresponding to a specific amino acid type. In contrast, scatter plots using other parameter combinations demonstrate significant cluster overlap between different amino acid types. These results indicate that the Δ*I* and SD parameter pair provides optimal 2D discrimination capability for single-molecule sensing analysis. Each visualization was generated from 1396 events to ensure statistical significance.



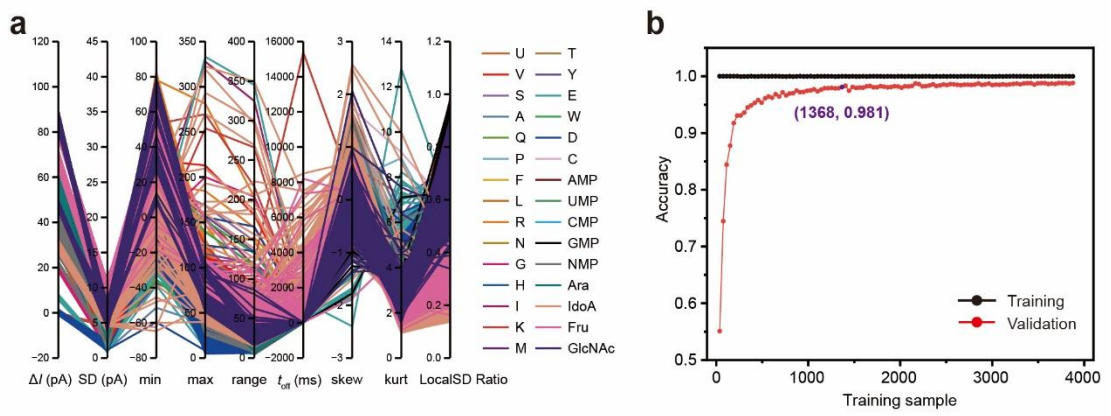
Supplementary Fig. 18: A detailed comparison between P and M events. (**a**) Representative events of P (top) and M (bottom). Though P and M events have some overlap in the two-dimension scatter plot of ΔI versus SD (**Extended Data Fig. 3**), their max ($max = I_{\mathrm{FPBA}} - I_{\mathrm{event\ minimum}}$) are significantly different. Compared to M events, the max value of P events is significantly higher. (**b**) The 3D event scatter plot of ΔI, SD, and max generated from the P and M events. According to this scatter plot, P and M events were well distinguished.



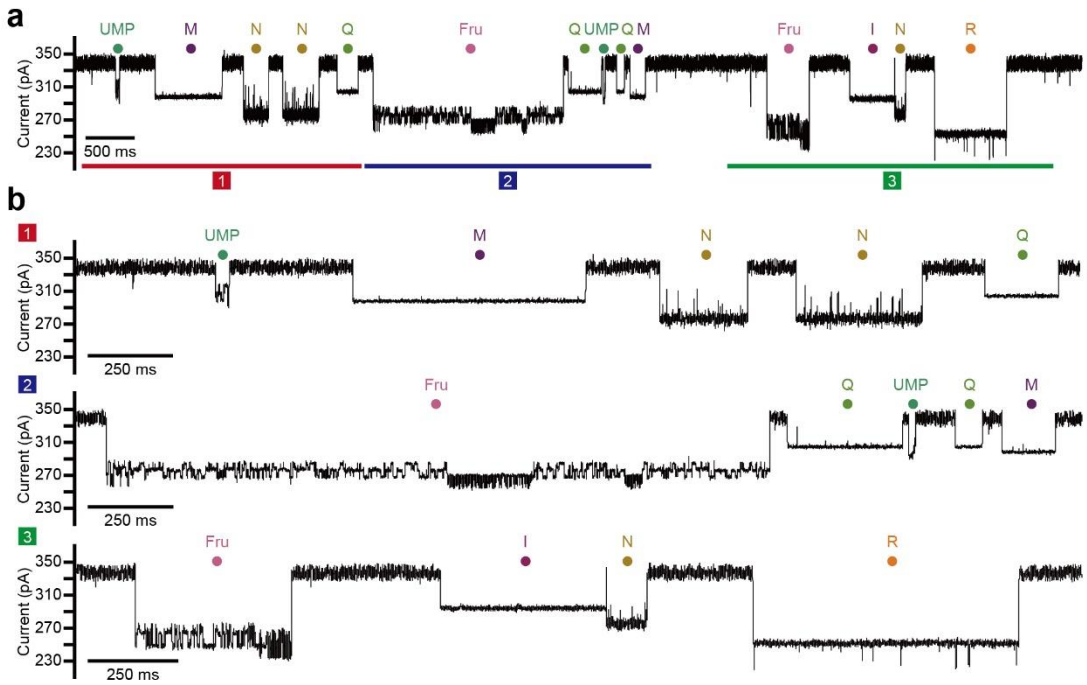
Supplementary Fig. 19: A detailed comparison between isoleucine and leucine events. The measurements were carried out as described in **Methods**. A 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0) was used. A transmembrane voltage of +160 mV was continually applied. (**a**) Top: The chemical structures of isoleucine (Ile, I) and leucine (Leu, L). Isoleucine and leucine are isomers with identical mass. Bottom: Representative events of isoleucine and leucine. I_{FPBA} represents the open pore current of MspA-FPBA. Events caused by isoleucine and leucine are easily identifiable. (**b**) The event scatter plot of ΔI versus SD generated from isoleucine and leucine events. I (n = 237) and L (n = 371) events were employed to generate the statistics. Though isoleucine and leucine has indistinguishable MW, they are fully discriminated by nanopore.



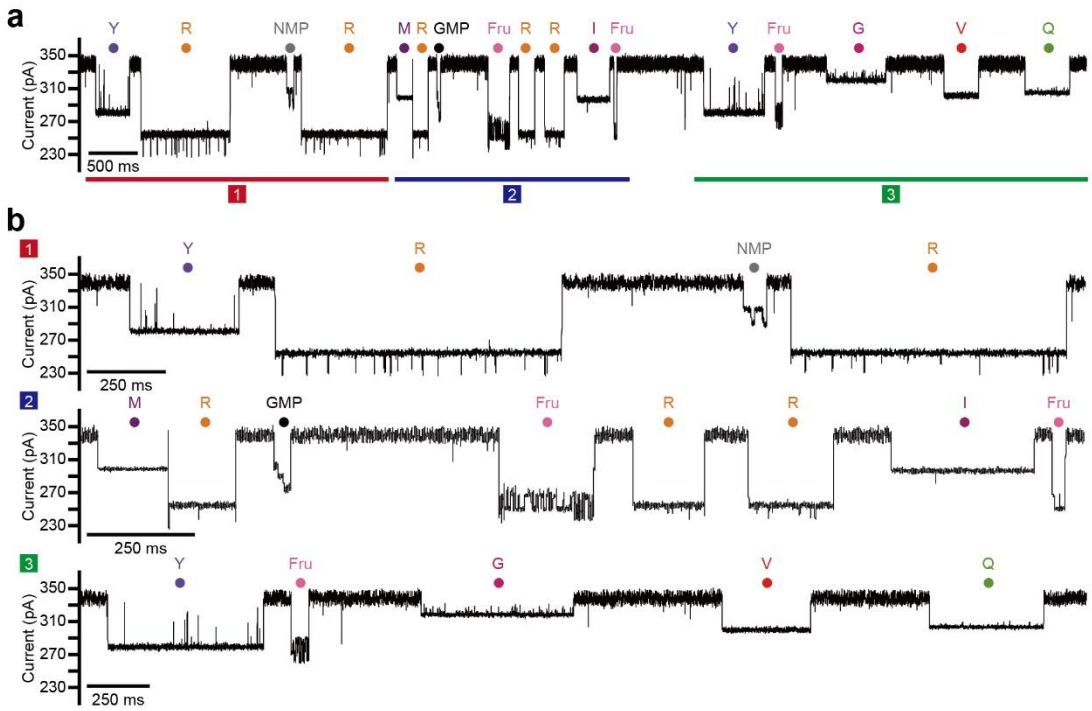
Supplementary Fig. 20: The workflow of machine learning. In brief, sensing events separately acquired with 21 amino acids, 4 nucleoside monophosphates and 4 saccharides were collected to form a dataset. Nine event features (Δ*I*, SD, LocalSD Ratio, min, max, range, t_{off}, skewness (skew) and kurtosis (kurt)) were extracted from each event to form a feature matrix. A 9D feature matrix was built for machine learning. Machine learning was performed with the Classification Learner toolbox of MATLAB. Seven classifiers were evaluated with 10-fold cross-validation to screen the best-performing model. For the 9D matrix, the highest validation accuracy reaches 98.9%, achieved by the bagged trees model (**Supplementary Table 6**).



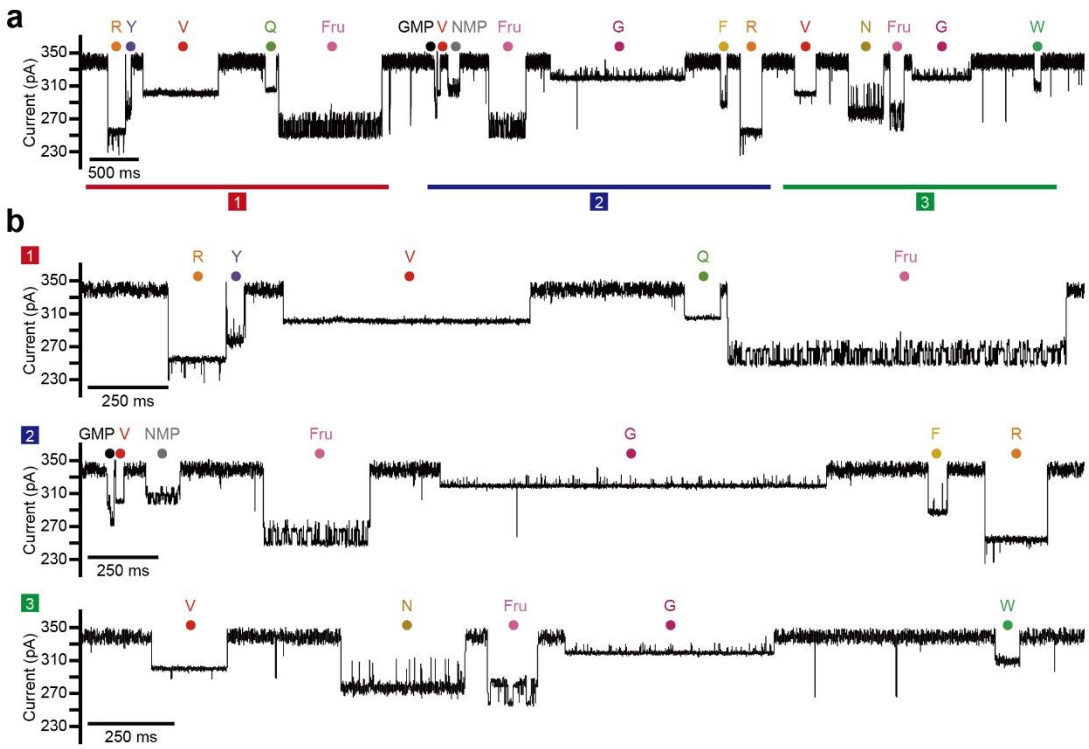
Supplementary Fig. 21: The parallel coordinate plot and the learning curve. (**a**) The parallel coordinate plots generated from the 9D feature matrix. (**b**) The learning curve generated by nine-feature machine learning. With the previously trained bagged trees model however with varying sizes of the training set, corresponding accuracy scores were reported to generate the learning curve. The training and the validation score were derived from the 10-fold cross-validation results. According to the learning curve, when the sample of the training set exceeds 1368, the accuracy of validation could reach 0.981.



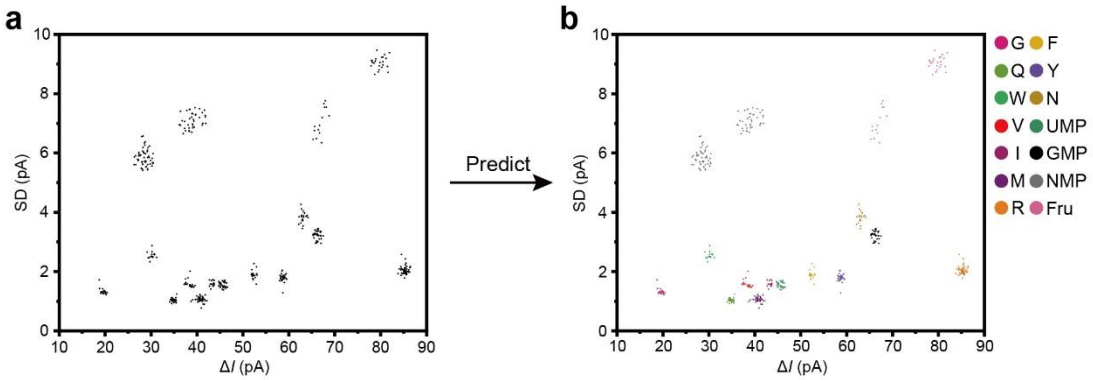
Supplementary Fig. 22: Zoomed-in view of Fig. 3b. (**a**) The raw current trace presented in **Fig. 3b**. All events were identified with the custom machine learning program and marked with corresponding identities. For a detailed demonstration, the trace is separated into three segments marked with 1 to 3. (**b**) The zoomed in view of the corresponding trace segments demonstrated in (**a**).



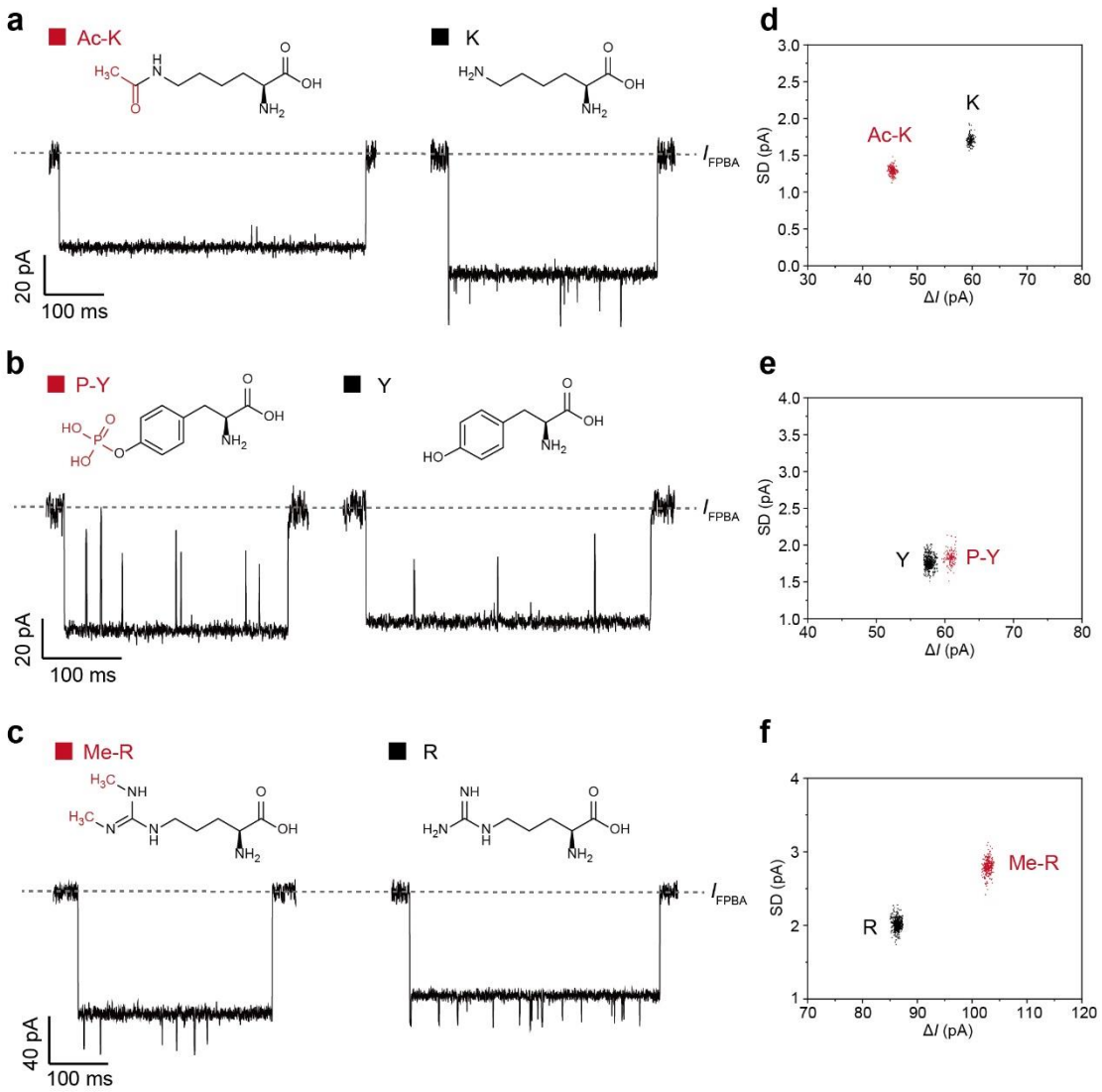
Supplementary Fig. 23: Zoomed-in view of Fig. 3c. (**a**) The raw current trace presented in **Fig. 3c**. All events were identified with the custom machine learning program and marked with corresponding identities. For a detailed demonstration, the trace is separated into three segments marked with 1 to 3. (**b**) The zoomed in view of the corresponding trace segments demonstrated in (**a**).



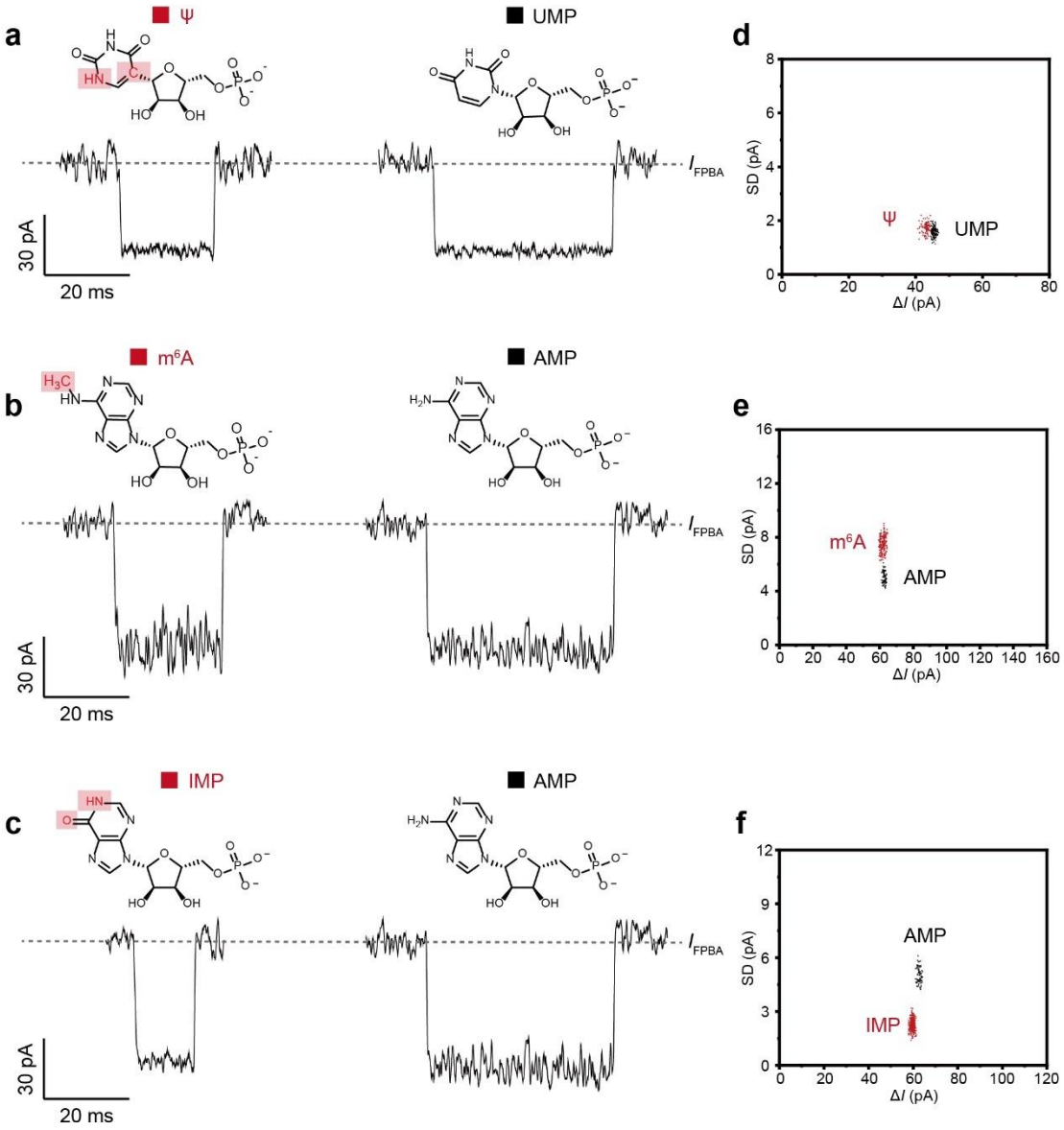
Supplementary Fig. 24: Zoomed-in view of Fig. 3d. (**a**) The raw current trace presented in **Fig. 3d**. All events were identified with the custom machine learning program and marked with corresponding identities. For a detailed demonstration, the trace is separated into three segments marked with 1 to 3. (**b**) The zoomed in view of the corresponding trace segments demonstrated in (**a**).



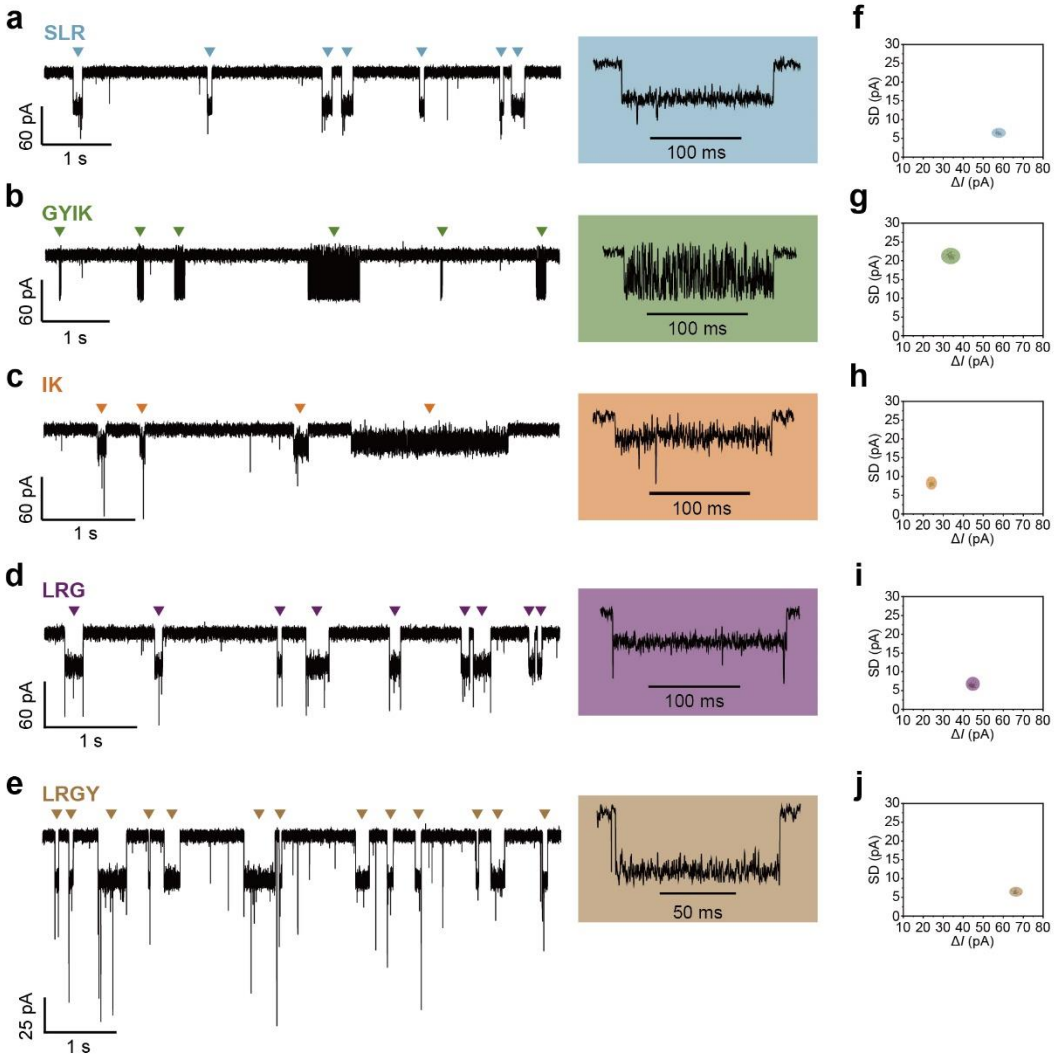
Supplementary Fig. 25: Identification of ten amino acids, two nucleoside monophosphates, and one saccharide using machine learning. (**a**) The scatter plot of Δ*I* versus SD generated by results acquired by simultaneous sensing of ten amino acids, two nucleoside monophosphates, and one saccharide (n = 449) as demonstrated in **Fig. 3b-d**. (**b**) The corresponding scatter plot of Δ*I* versus SD however with machine learning predicted labels. All event identities were predicted by the trained bagged trees model and labeled with color-coded dots.



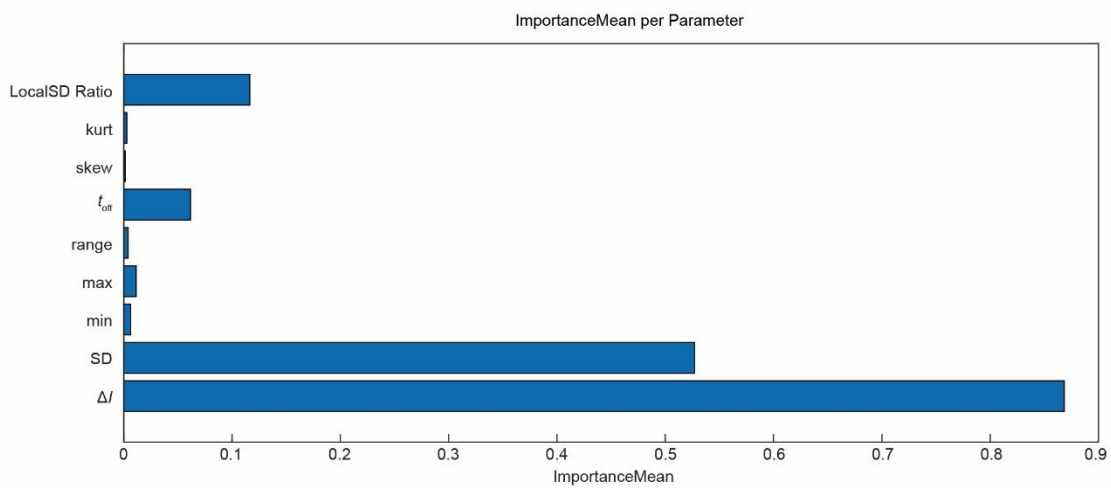
Supplementary Fig. 26: Events generated by amino acids with post-translational modifications (PTMs). The measurements were carried out as described in **Methods** in a 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0). A transmembrane voltage of +160 mV was continually applied during the measurements. (**a-c**) Left: The chemical structures of amino acids with PTMs and their representative sensing events. Right: The chemical structures of the corresponding unmodified amino acids and their representative sensing events. Each amino acid was separately added to *cis* with a final concentration of 400 μM. I_{FPBA} represents the open pore current of MspA-FPBA. (**d-f**) The event scatter plot of Δ*I* versus SD generated from sensing events of Ac-K/K (**d**, n = 350), P-Y/Y (**e**, n = 536) and Me-R/R (**f**, n = 754).



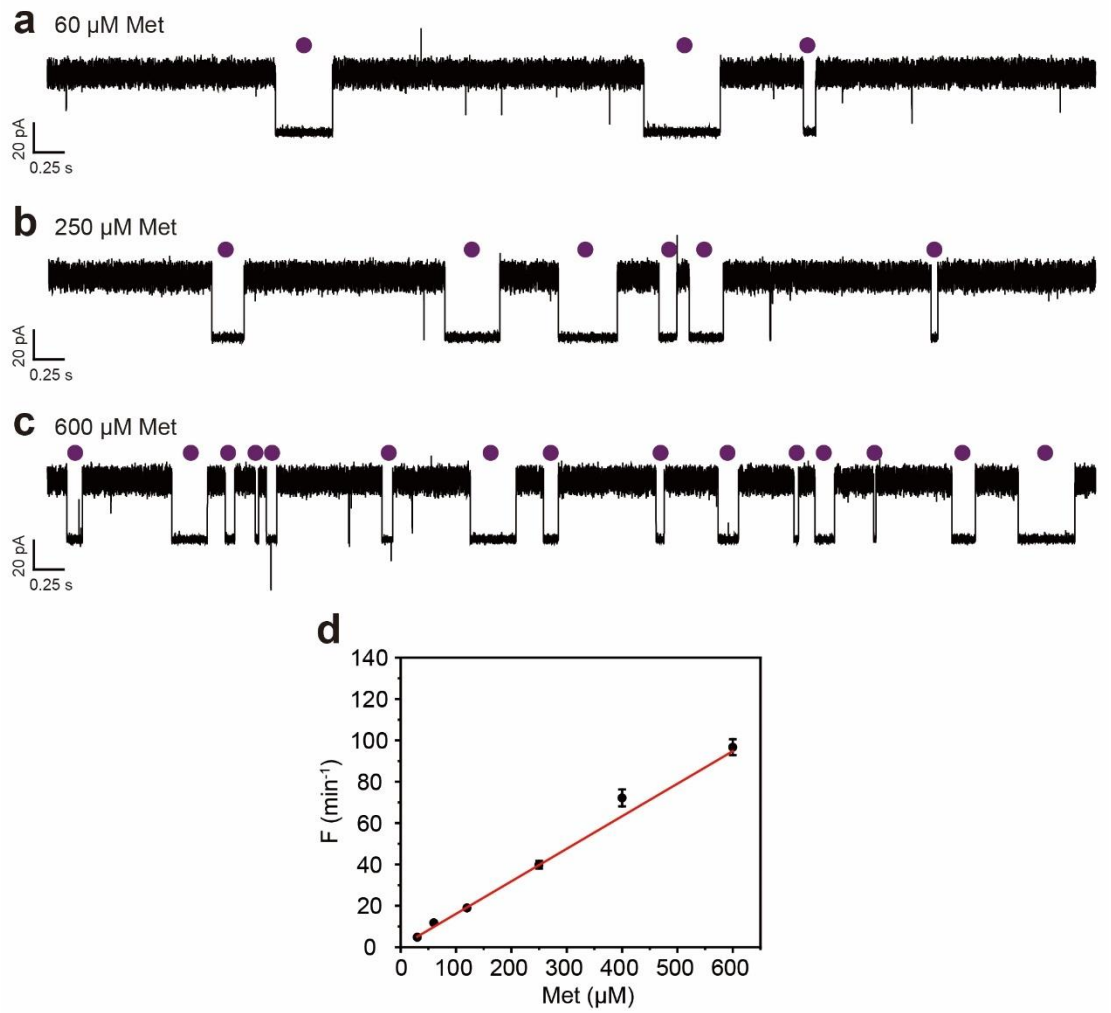
Supplementary Fig. 27: Events generated by nucleoside monophosphates with epigenetic modifications. The measurements were carried out as described in **Methods** in a 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0). A transmembrane voltage of +160 mV was continually applied during the measurements. (**a-c**) Left: The chemical structures of nucleoside monophosphates with epigenetic modifications and their representative sensing events. Right: The chemical structures of the corresponding unmodified nucleoside monophosphates and their representative sensing events. Each nucleoside monophosphate was separately added to *cis* with a final concentration of 2 mM. I_{FPBA} represents the open pore current of MspA-FPBA. (**d-f**) The event scatter plot of Δ*I* versus SD generated from sensing events of ψ/UMP (**d**, n = 238), m^6A/AMP (**e**, n = 260) and IMP/AMP (**f**, n = 415).



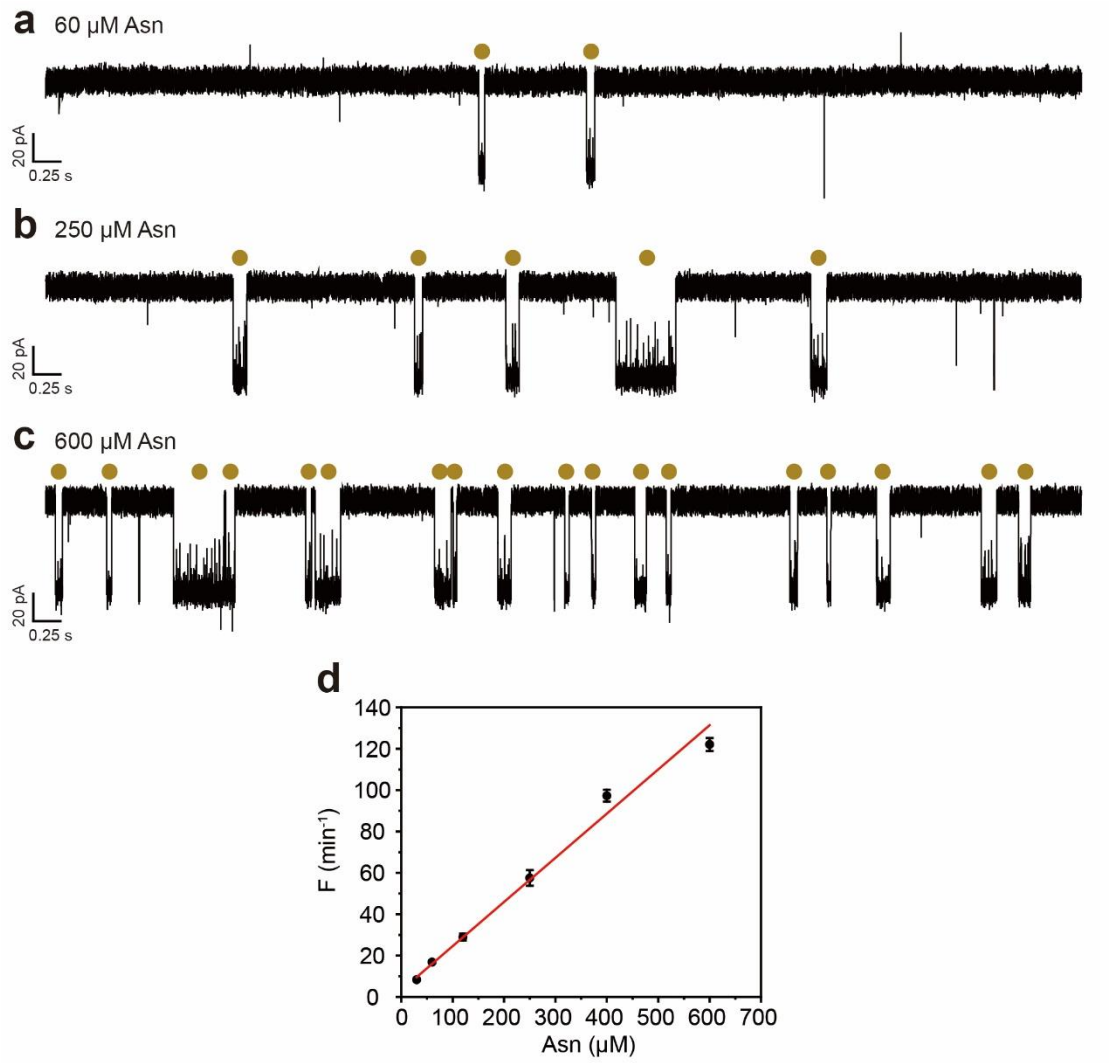
Supplementary Fig. 28: Events generated by peptides. The measurements were carried out as described in **Methods**. A 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0) was used. A transmembrane voltage of +160 mV was continually applied during the measurements. (**a**-**e**) Left: Representative traces containing events caused by SLR (**a**), GYIK (**b**), IK (**c**), LRG (**d**) and LRGY (**e**) binding. Peptide events were labeled with color-coded triangles. Each peptide was separately added to the *cis* chamber with a final concentration of 100 μM. Right: Representative events of the corresponding peptides events. (**f-j**) The scatter plot of Δ*I* versus SD respectively derived from sensing events of SLR (**f**, n = 45), GYIK (**g**, n = 84), IK (**h**, n = 35), LRG (**i**, n = 169) and LRGY (**j**, n = 53).



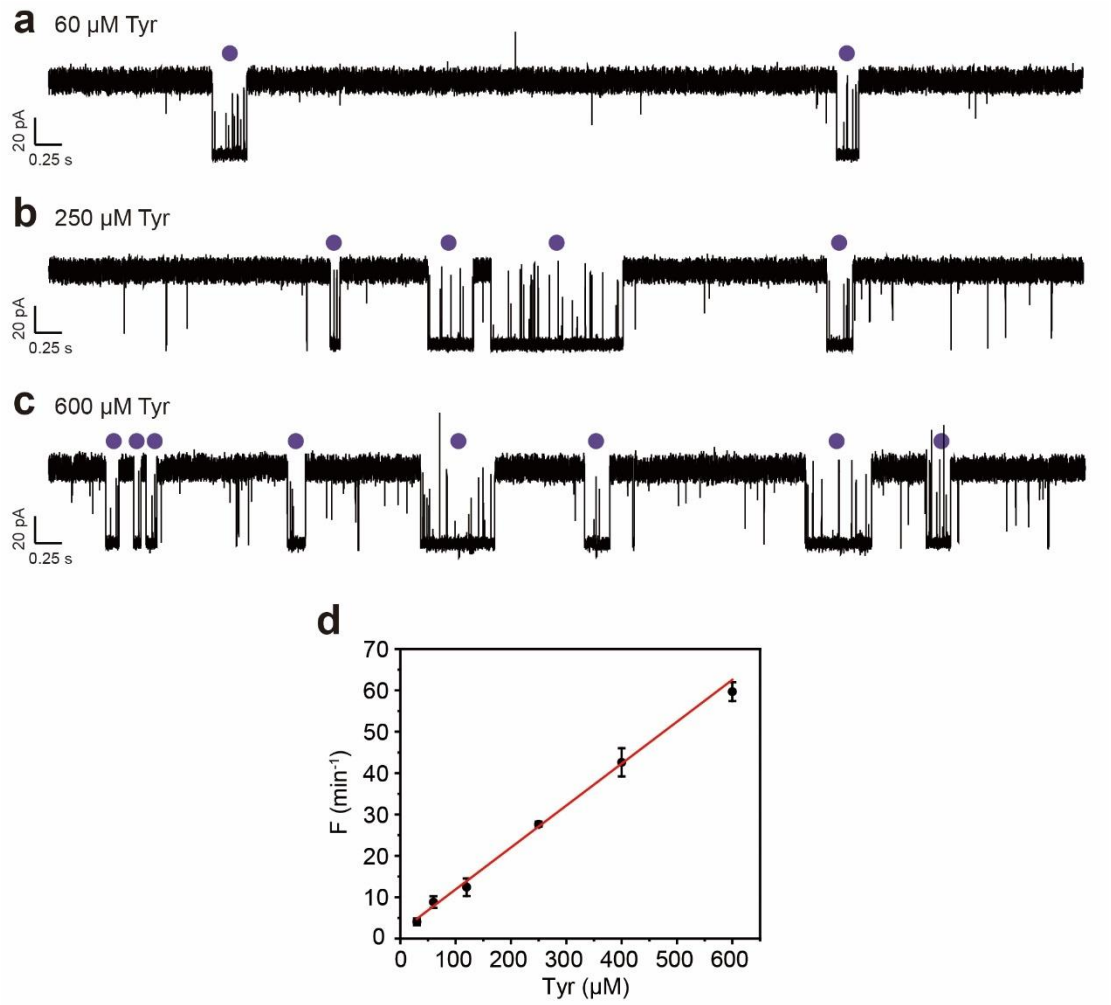
Supplementary Fig. 29: Visualization of the ImportanceMean values for different event parameters. A larger ImportanceMean value indicates that the parameter is more important in the model's differentiation of all analytes and has a greater overall contribution. The ImportanceMean values of nine event parameters (Δ*I*, SD, LocalSD Ratio, min, max, range, t_{off}, skew and kurt) were calculated for the Bagged Trees model used to distinguish all 40 analytes, using the Permutation Feature Importance method.



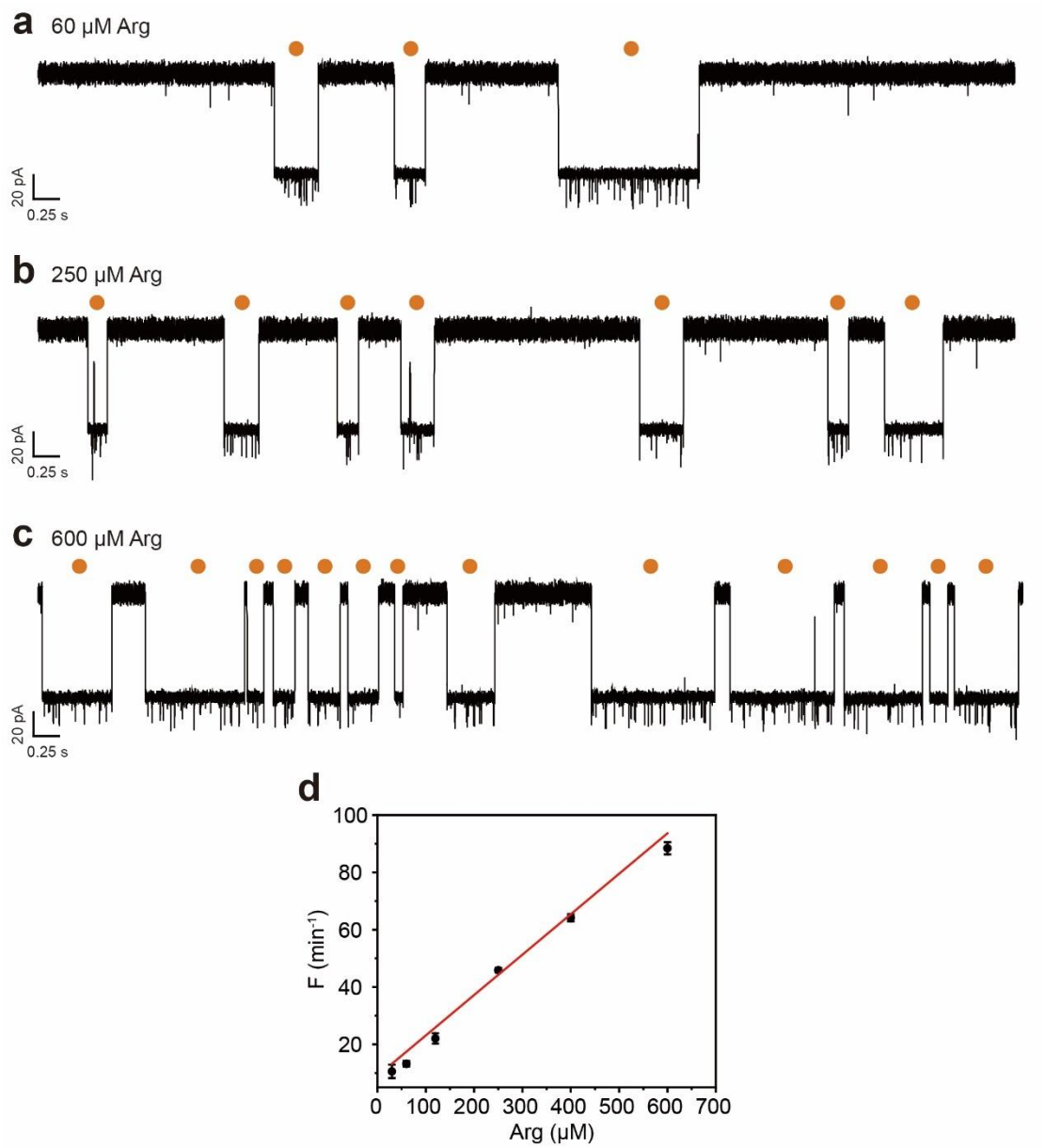
Supplementary Fig. 30: The concentration dependence of Met sensing using MspA-FPBA. The measurements were carried out in a 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0). A transmembrane voltage of +160 mV was continually applied during the measurements. Met was added to *cis* with a final concentration of 30-600 μM. (**a-c**) Representative traces acquired with varying Met concentrations. (**d**) The plot of *F* versus the Met concentration. The plot is overlaid with the corresponding linear fitting result. Three independent measurements were performed to show data consistency (*N* = 3). All statistical results are also summarized in **Supplementary Table 11**. The data in (**d**) demonstrates mean ± standard deviations derived from results of three independent measurements (*N* = 3). The error bars represent standard deviation values.



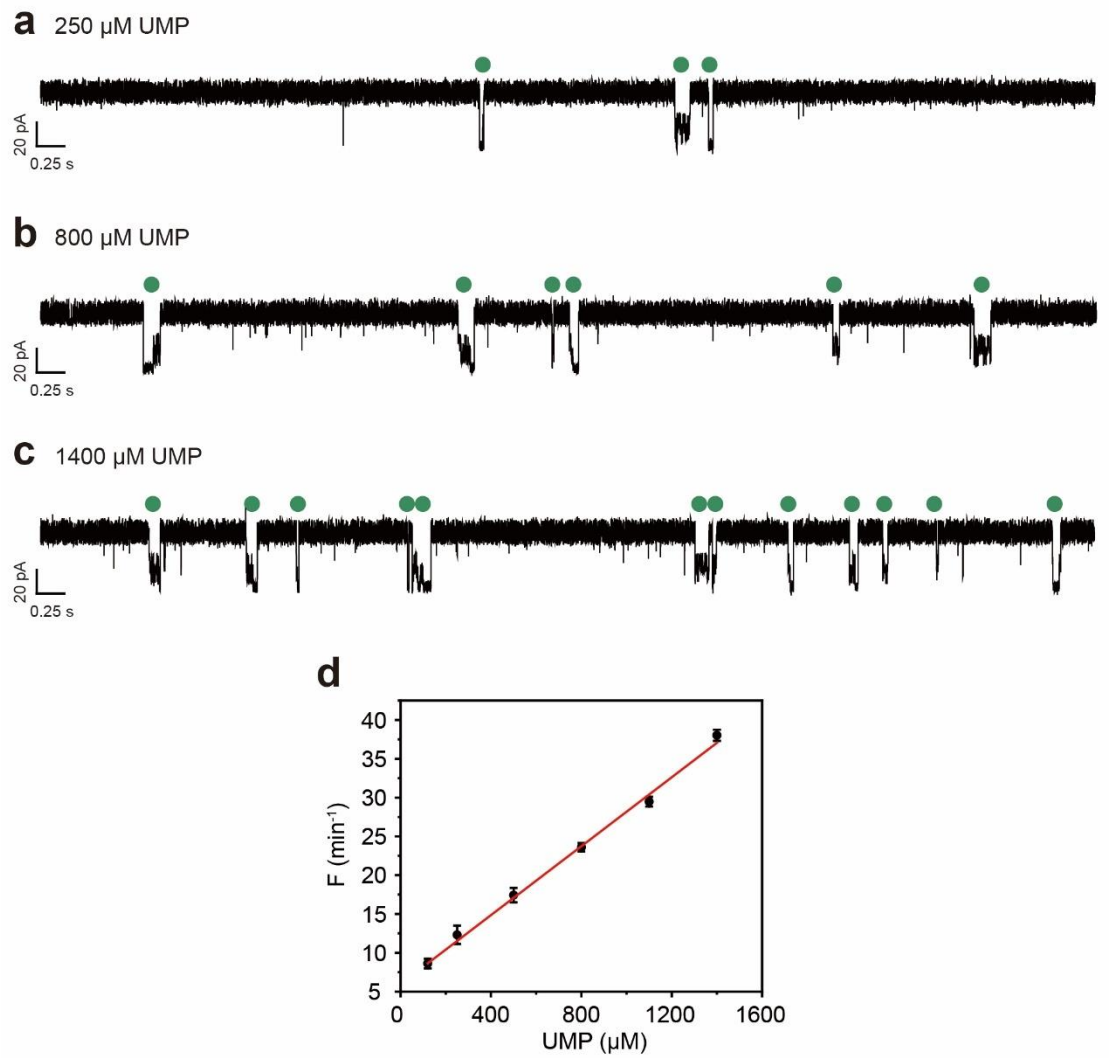
Supplementary Fig. 31: The concentration dependence of Asn sensing using MspA-FPBA. The measurements were carried out in a 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0). A transmembrane voltage of +160 mV was continually applied during the measurements. Asn was added to *cis* with a final concentration of 30-600 μM. (**a-c**) Representative traces acquired with varying Asn concentrations. (**d**) The plot of *F* versus the Asn concentration. The plot is overlaid with the corresponding linear fitting result. Three independent measurements were performed to show data consistency (*N* = 3). All statistical results are also summarized in **Supplementary Table 11**. The data in (**d**) demonstrates mean ± standard deviations derived from results of three independent measurements (*N* = 3). The error bars represent standard deviation values.



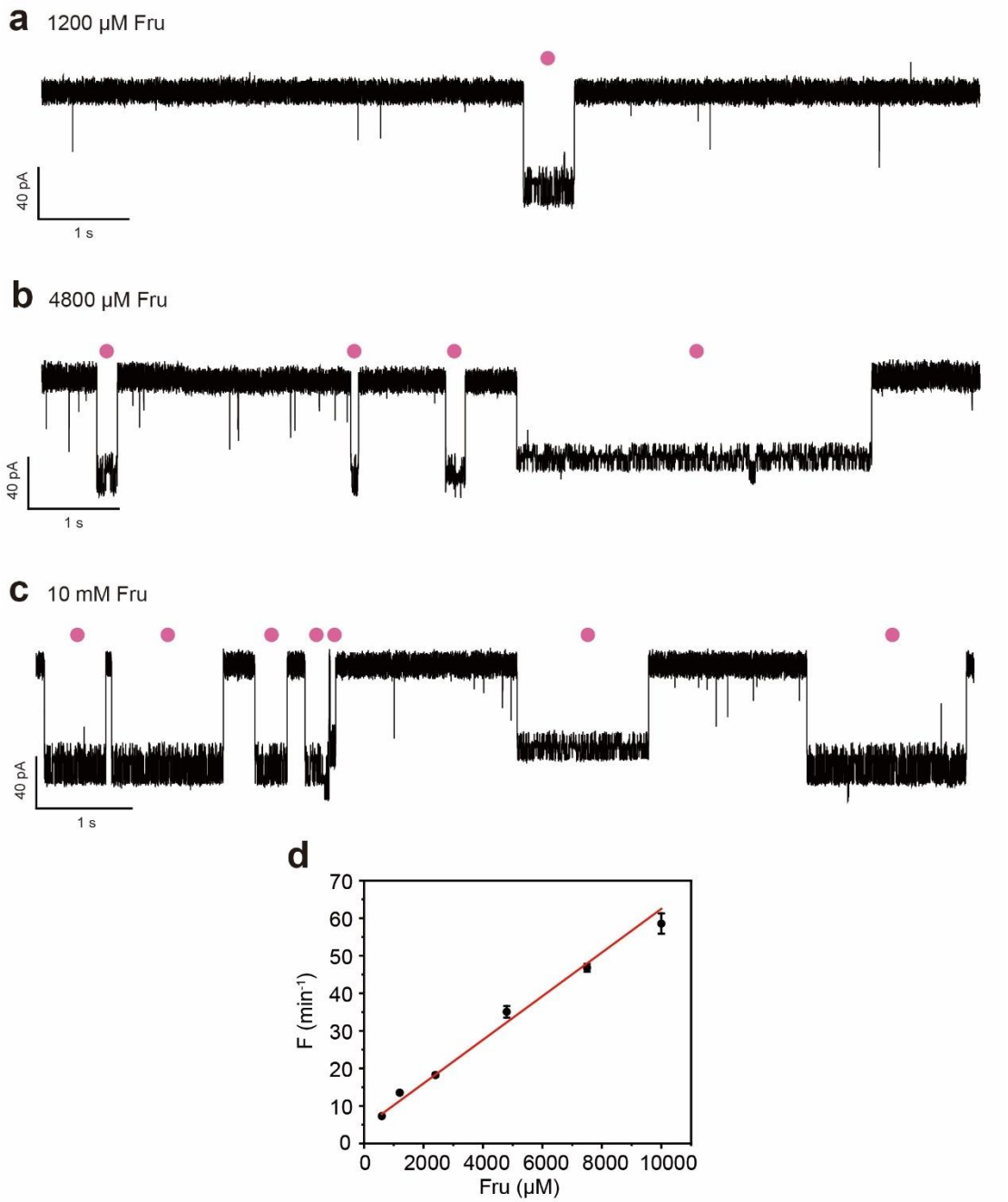
Supplementary Fig. 32: The concentration dependence of Tyr sensing using MspA-FPBA. The measurements were carried out in a 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0). A transmembrane voltage of +160 mV was continually applied during the measurements. Tyr was added to *cis* with a final concentration of 30-600 μM. (**a-c**) Representative traces acquired with varying Tyr concentrations. (**d**) The plot of *F* versus the Tyr concentration. The plot is overlaid with the corresponding linear fitting result. Three independent measurements were performed to show data consistency (*N* = 3). All statistical results are also summarized in **Supplementary Table 11**. The data in (**d**) demonstrates mean ± standard deviations derived from results of three independent measurements (*N* = 3). The error bars represent standard deviation values.



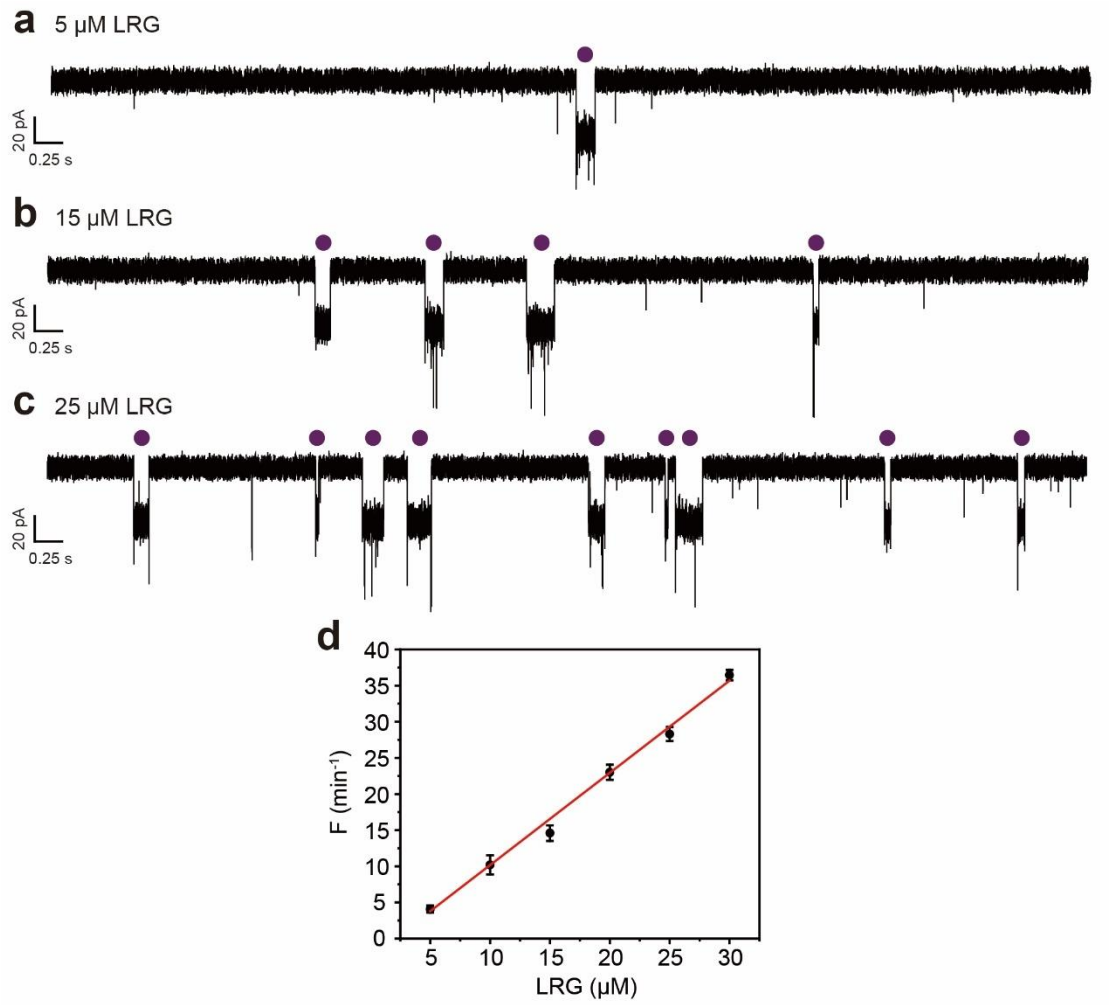
Supplementary Fig. 33: The concentration dependence of Arg sensing using MspA-FPBA. The measurements were carried out in a 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0). A transmembrane voltage of +160 mV was continually applied during the measurements. Arg was added to *cis* with a final concentration of 30-600 μM. (**a-c**) Representative traces acquired with varying Arg concentrations. (**d**) The plot of *F* versus the Arg concentration. The plot is overlaid with the corresponding linear fitting result. Three independent measurements were performed to show data consistency (*N* = 3). All statistical results are also summarized in **Supplementary Table 11**. The data in (**d**) demonstrates mean ± standard deviations derived from results of three independent measurements (*N* = 3). The error bars represent standard deviation values.



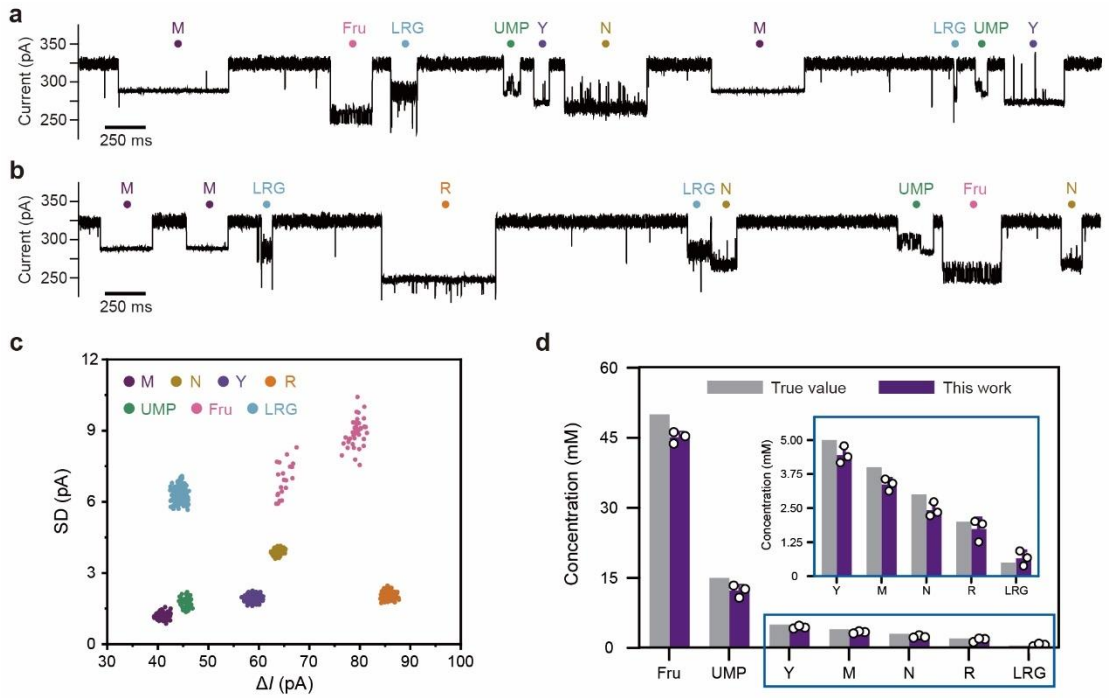
Supplementary Fig. 34: The concentration dependence of UMP sensing using MspA-FPBA. The measurements were carried out in a 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0). A transmembrane voltage of +160 mV was continually applied during the measurements. UMP was added to *cis* with a final concentration of 120-1400 μM. (**a-c**) Representative traces acquired with varying UMP concentrations. (**d**) The plot of *F* versus the UMP concentration. The plot is overlaid with the corresponding linear fitting result. Three independent measurements were performed to show data consistency (*N* = 3). All statistical results are also summarized in **Supplementary Table 11**. The data in (**d**) demonstrates mean ± standard deviations derived from results of three independent measurements (*N* = 3). The error bars represent standard deviation values.



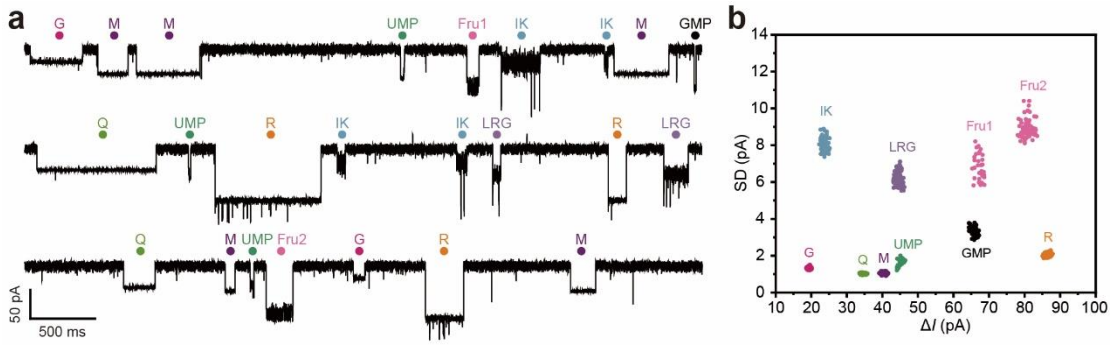
Supplementary Fig. 35: The concentration dependence of Fru sensing using MspA-FPBA. The measurements were carried out in a 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0). A transmembrane voltage of +160 mV was continually applied during the measurements. Fru was added to *cis* with a final concentration of 600-10000 μM. (**a-c**) Representative traces acquired with varying Fru concentrations. (**d**) The plot of *F* versus the Fru concentration. The plot is overlaid with the corresponding linear fitting result. Three independent measurements were performed to show data consistency (*N* = 3). All statistical results are also summarized in **Supplementary Table 11**. The data in (**d**) demonstrates mean ± standard deviations derived from results of three independent measurements (*N* = 3). The error bars represent standard deviation values.



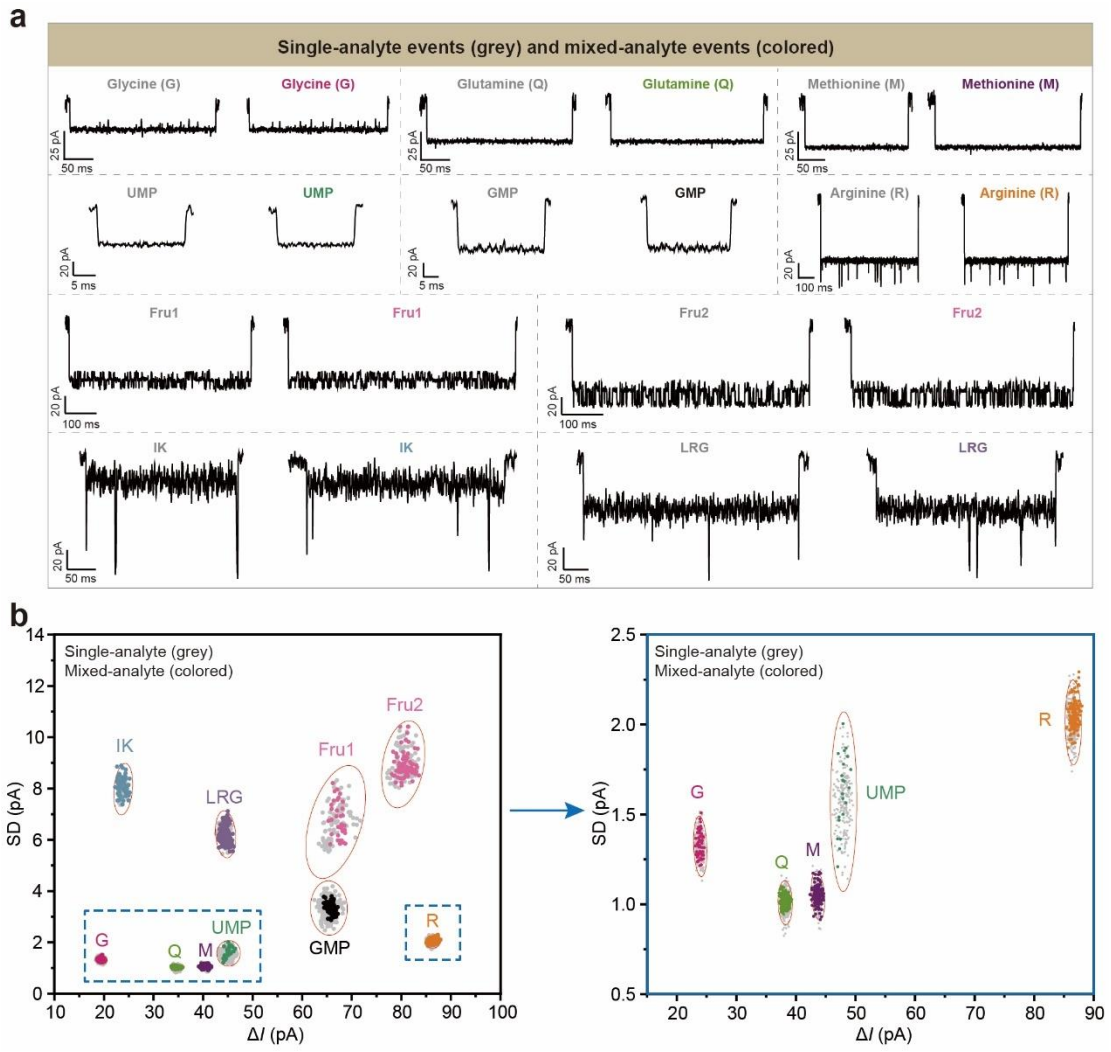
Supplementary Fig. 36: The concentration dependence of LRG sensing using MspA-FPBA. The measurements were carried out in a 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0). A transmembrane voltage of +160 mV was continually applied during the measurements. LRG was added to *cis* with a final concentration of 5-30 μM. (**a-c**) Representative traces acquired with varying LRG concentrations. (**d**) The plot of *F* versus the LRG concentration. The plot is overlaid with the corresponding linear fitting result. Three independent measurements were performed to show data consistency (*N* = 3). All statistical results are also summarized in **Supplementary Table 11**. The data in (**d**) demonstrates mean ± standard deviations derived from results of three independent measurements (*N* = 3). The error bars represent standard deviation values.



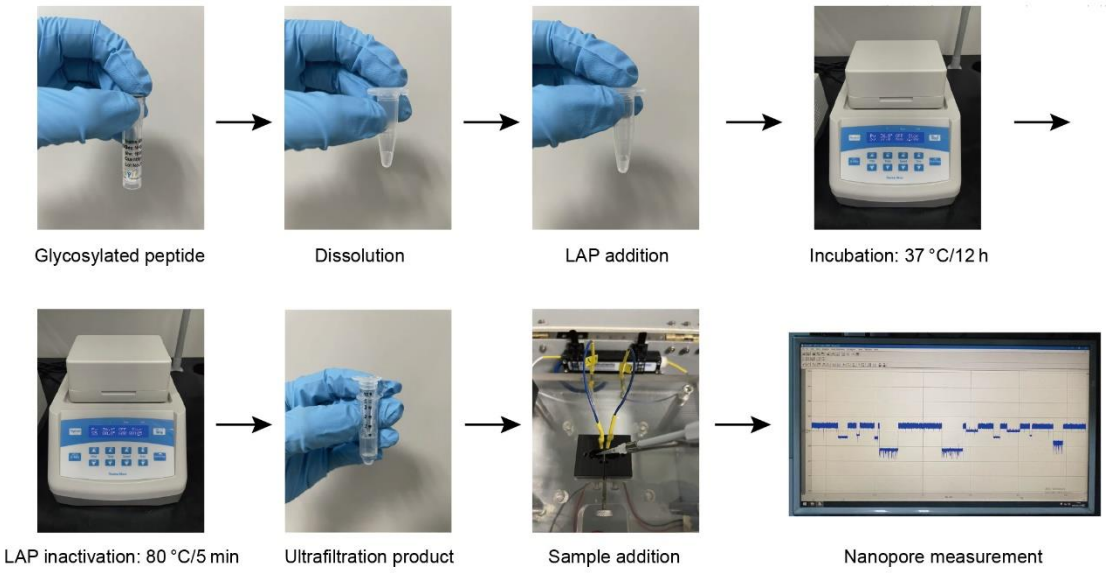
Supplementary Fig. 37: Quantitative analysis of a mixed sample containing Met (M), Asn (N), Tyr (Y), Arg (R), UMP, Fru, and the tripeptide LRG. (**a-b**) Continuous traces acquired with a mxied sample containing M, N, Y, R, UMP, Fru, and LRG. 1.7 mg LRG, 3.5 mg R, 4.0 mg N, 6.0 mg M, 11.3 mg Y, 55.2 mg UMP, and 90.1 mg Fru were dissolved in 10 mL ultrapure water to achieve final concentrations of 0.5, 2, 3, 4, 5, 15, and 50 mM, respectively. With an MspA-FPBA, 10 μL of the above-described solution was added to the *cis* chamber. A transmembrane voltage of +160 mV was continually applied during the measurements. Events were automatically predicted by the previously trained model and labeled accordingly. (**c**) The scatter plot of Δ*I* versus SD for the M, N, Y, R, UMP, Fru, and LRG events obtained from **a-b**. The scatter plot was generated using a 30 min continually recorded trace. 1072 successive events were employed to generate the statistics. (**d**) Comparison of the concentrations of Fru, UMP, Y, M, N, R, and LRG reported by nanopore analysis with the true values. The values measured in this work were determined and calibrated as described in **Methods**. For the quantitative analysis, three independent measurements (*N* = 3) were conducted to generate statistics. Data are presented as mean ± standard deviations, with error bars representing standard deviation values.



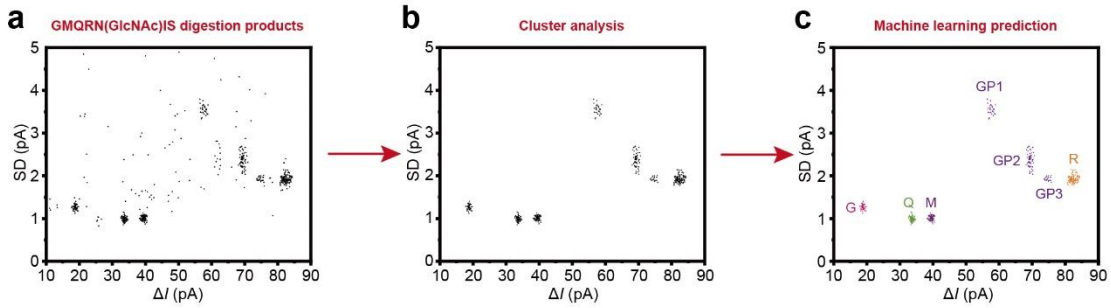
Supplementary Fig. 38: Nanopore sensing of a multi-analyte mixture containing Gly (G), Met (M), Gln (Q), Arg (R), UMP, GMP, Fru, the dipeptide IK, and the tripeptide LRG. (**a**) Continuous current traces acquired from the multi-analyte mixture. The mixture was added to the *cis* chamber to achieve final concentrations of G (85 µM), M (80 µM), Q (100 µM), R (60 µM), UMP (300 µM), GMP (400 µM), Fru (1 mM), IK (15 µM), and LRG (10 µM). A transmembrane voltage of +160 mV was continually applied during the measurements. Events were automatically predicted by the previously trained bagged trees model and labeled accordingly. (**b**) The scatter plot of Δ*I* versus SD for the G, M, Q, R, UMP, GMP, Fru, IK, and LRG events obtained from **a**. The scatter plot was generated using a 30 min continually recorded trace. 947 successive events were employed to generate the statistics.



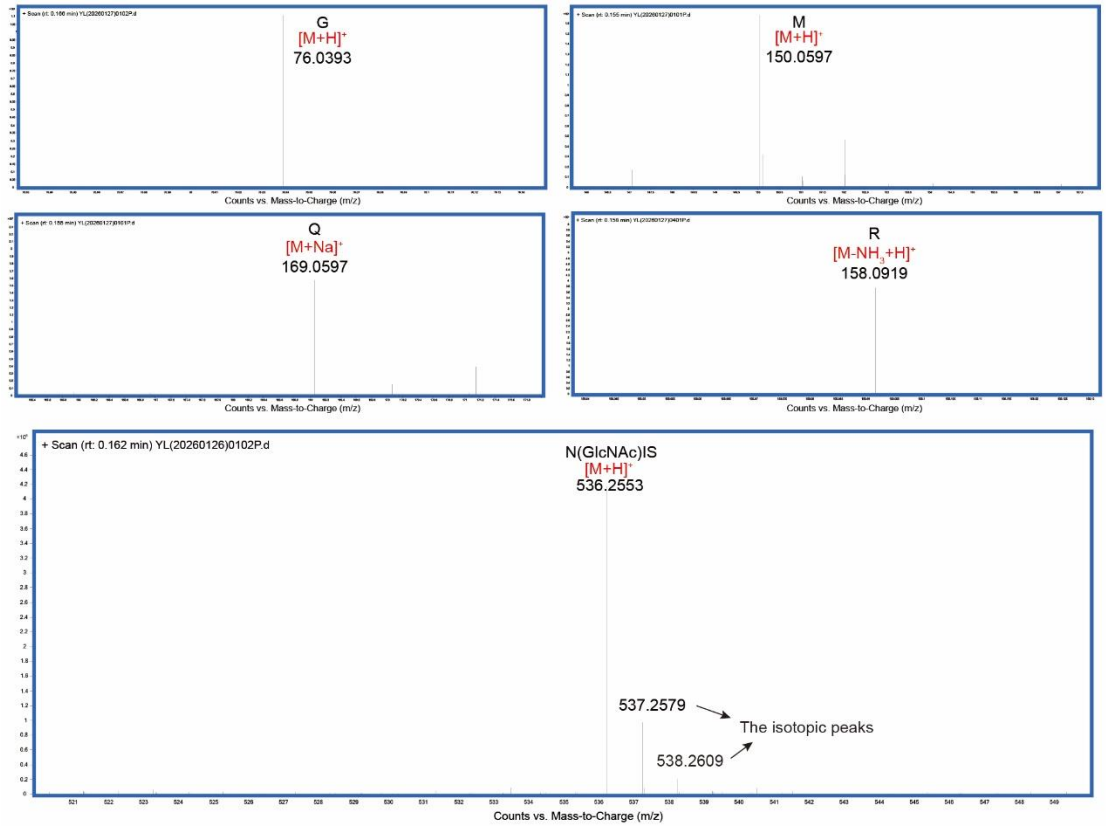
Supplementary Fig. 39: Comparison of nanopore events obtained from single-analyte and mixture experiments. (**a**) Visual comparison of representative events of analytes (G, M, Q, R, UMP, GMP, Fru, IK, and LRG) obtained from single-analyte experiments and from the mixture experiment presented in **Supplementary Fig. 38**. Single-analyte events and mixed-analyte events are indicated in gray and color, respectively. No significant differences in the event characteristics of each analyte were observed. (**b**) Overlap of event distributions between events acquired with single-analyte and mixed analytes. Gray points represent events from single-analyte experiments, with red ellipses indicating the 99% confidence regions. Colored points correspond to events recorded from mixture experiments, which largely fall within the same regions. To more clearly show the details, the events inside the blue box are further zoomed-in and shown on the right.



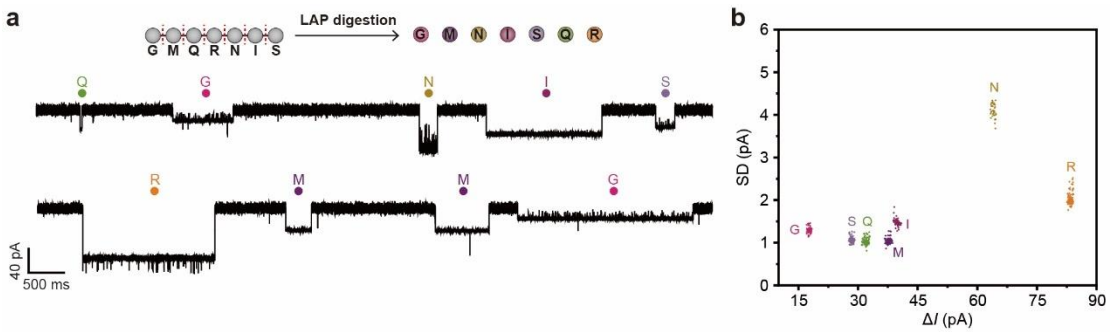
Supplementary Fig. 40: Analysis of digestion products of glycosylated peptides by MspA-FPBA. The GMQRN(GlcNAc)IS peptide, a customized synthetic glycosylated peptide, was dissolved in a 50 mM sodium phosphate, pH 8.0 buffer with a concentration of 10 mg mL^{-1}. LAP was prepared in a 50 mM sodium phosphate, pH 8.0 buffer with a concentration of 50 mg mL^{-1} (>7 U mg^{-1}). To initiate the hydrolysis reaction, 10 µL LAP solution and 20 µL 10 mM $MgCl_2$ were added to 70 µL 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 inactive the LAP. Afterwards, another 100 µL ultrapure water was added to the mixture and loaded into an ultracentrifuge tube with a 10 kDa molecular weight cut-off. The filtration was then performed at 8,000 r.p.m. for 60 min at 4 °C. To initiate nanopore measurement, 25 µL ultrafiltrate was added to the *cis* side. A +160 mV bias was continually applied.



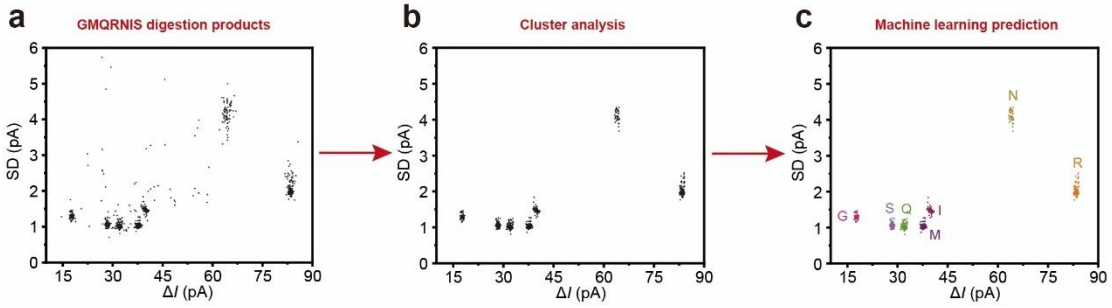
Supplementary Fig. 41: Background reduction and event prediction. (**a**) The raw scatter plots of Δ*I* versus SD of events acquired with GMQRN(GlcNAc)IS digestion products. (**b**) The scatter plot of events acquired with GMQRN(GlcNAc)IS digestion products however after background reduction treatment. Here, the background reduction treatment was performed with the DBSCAN algorithm. The epsilon was set to 0.5 and the minimum points was set to 5. Scatter dots that don't form a clear cluster were removed after the treatment. (**c**) Prediction of events performed by machine learning. Amino acid event identities were predicted by the previously trained bagged trees model and labeled with color-coded dots. N(GlcNAc)IS peptide event identities were respectively labeled as GP1, GP2 and GP3 based on the blockage amplitude and noise amplitude of traces in **Extended Data Fig. 7b**.



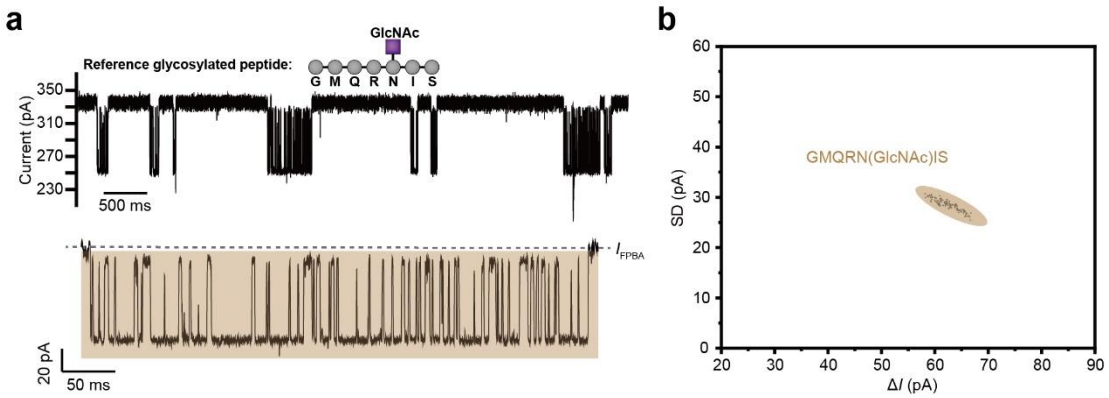
Supplementary Fig. 42: Mass spectrometric characterization of the enzymatic digest filtrate of the glycosylated peptide GMQRN(GlcNAc)IS. Molecular ion peaks corresponding to G (m/z = 76.0393), M (m/z = 150.0597), Q (m/z = 169.0597), R (m/z = 158.0919), and N(GlcNAc)IS (m/z = 536.2553) were observed, consistent with the nanopore measurement results in **Extended Data Fig. 7**.



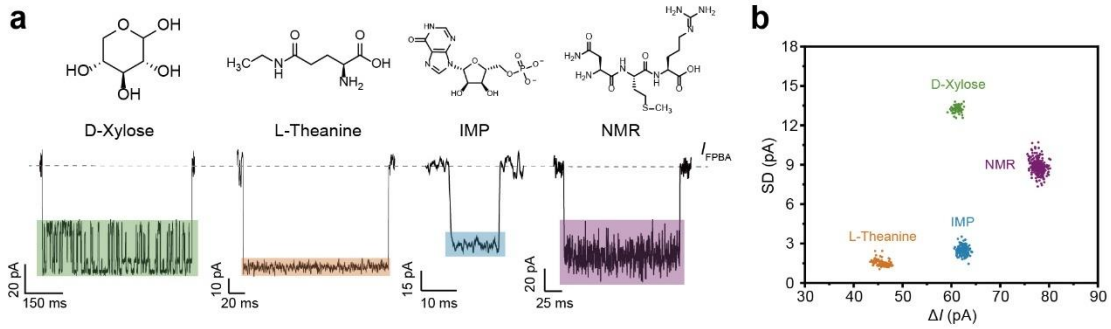
Supplementary Fig. 43: Identification of amino acids by protease digestion of GMQRNIS without glycosylation modification. a, Representative traces of the heptapeptide with the sequence GMQRNIS following LAP enzymatic digestion (**Methods**). Representative events corresponding to different amino acids were automatically predicted via machine learning and labeled with color-coded dots (G: dark red; M: dark purple; Q: light green; R: orange; N: brown; I: deep rose; S: muted mauve). **b**, The scatter plot of Δ*I* versus SD for the amino acid events obtained from **a**. The scatter plot was generated using a 20 min continually recorded trace. The seven clusters of amino acids corresponding to G, M, Q, R, N, I and S were identified by machine learning. All measurements were conducted under a continuous bias of +160 mV, with 25 μL of the enzymatic hydrolysate added to the *cis* side.



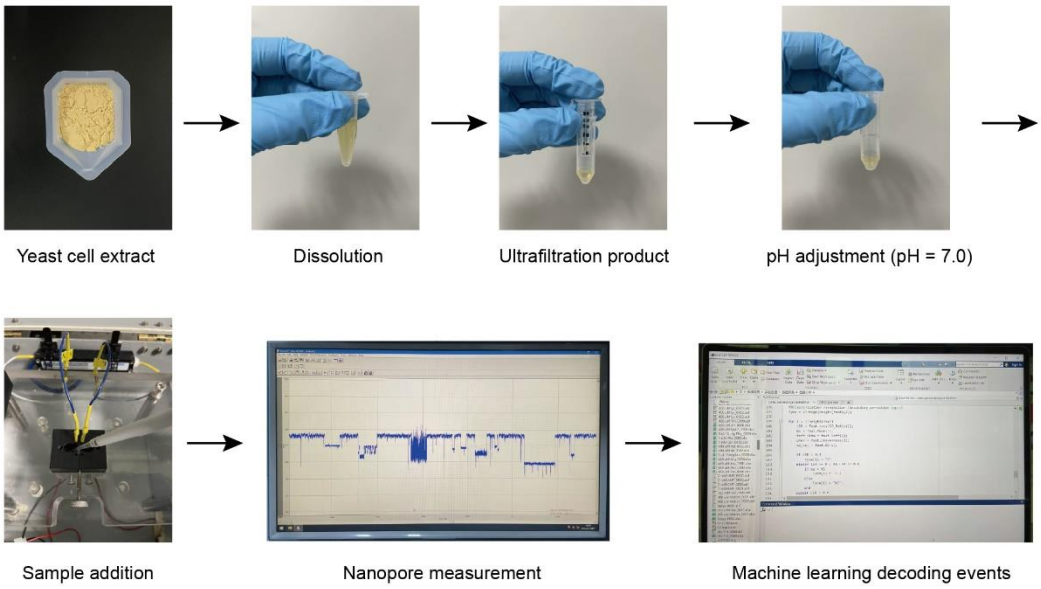
Supplementary Fig. 44: Background reduction and event prediction. (**a**) The raw scatter plots of Δ*I* versus SD of events acquired with GMQRNIS digestion products. (**b**) The scatter plot of events acquired with GMQRNIS digestion products however after background reduction treatment. Here, the background reduction treatment was performed with the DBSCAN algorithm. The epsilon was set to 0.5 and the minimum points was set to 5. Scatter dots that don't form a clear cluster were removed after the treatment. (**c**) Prediction of events by machine learning. Amino acid event identities were predicted by the previously trained bagged trees model and labeled with color-coded dots.



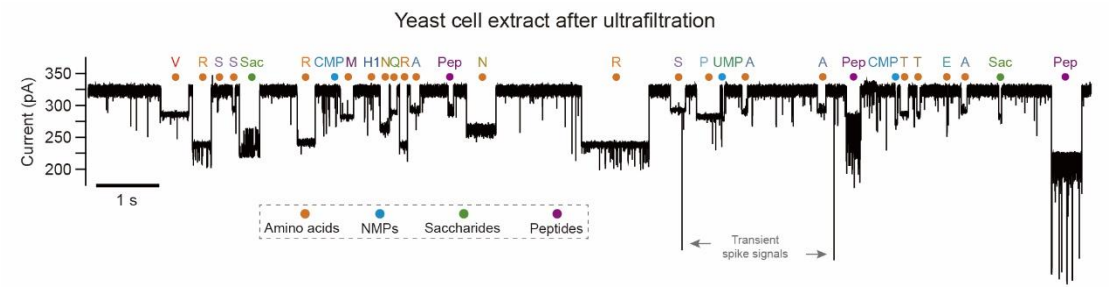
Supplementary Fig. 45: Nanopore sensing of GMQRN(GlcNAc)IS peptide. The measurements were carried out as described in **Methods**. A 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0) was used. A transmembrane voltage of +160 mV was continually applied. (**a**) Top: The representative trace of GMQRN(GlcNAc)IS peptide. Bottom: The representative event of GMQRN(GlcNAc)IS peptide. I_{FPBA} represents the open pore current of MspA-FPBA. (**b**) The event scatter plot of Δ*I* versus SD generated from GMQRN(GlcNAc)IS peptide events. 118 successive events were employed to generate the statistics.



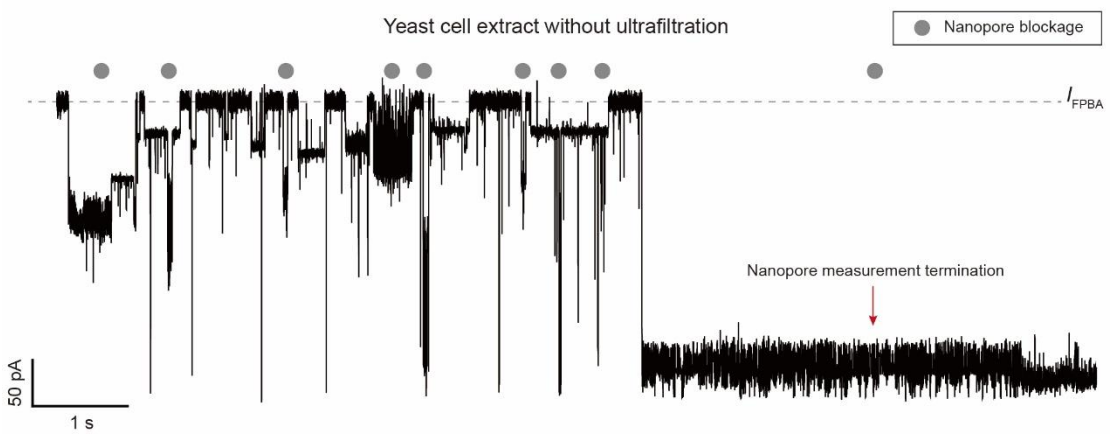
Supplementary Fig. 46: Simultaneous sensing of theanine, IMP, D-xylose and tripeptide NMR. A 1.5 M KCl buffer (*cis*: 1.5 M KCl, 10 mM CHES, pH 9.0; *trans*: 1.5 M KCl, 10 mM MES, pH 6.0) was used. A transmembrane voltage of +160 mV was continually applied. (**a**) Top: The chemical structures of theanine, IMP, D-xylose and tripeptide NMR. Bottom: Representative events of theanine, IMP, D-xylose and tripeptide NMR. I_{FPBA} represents the open pore current of MspA-FPBA. (**b**) The event scatter plot of ΔI versus SD generated by simultaneous sensing of theanine, IMP, D-xylose and tripeptide NMR. 760 events were employed to generate the statistics.



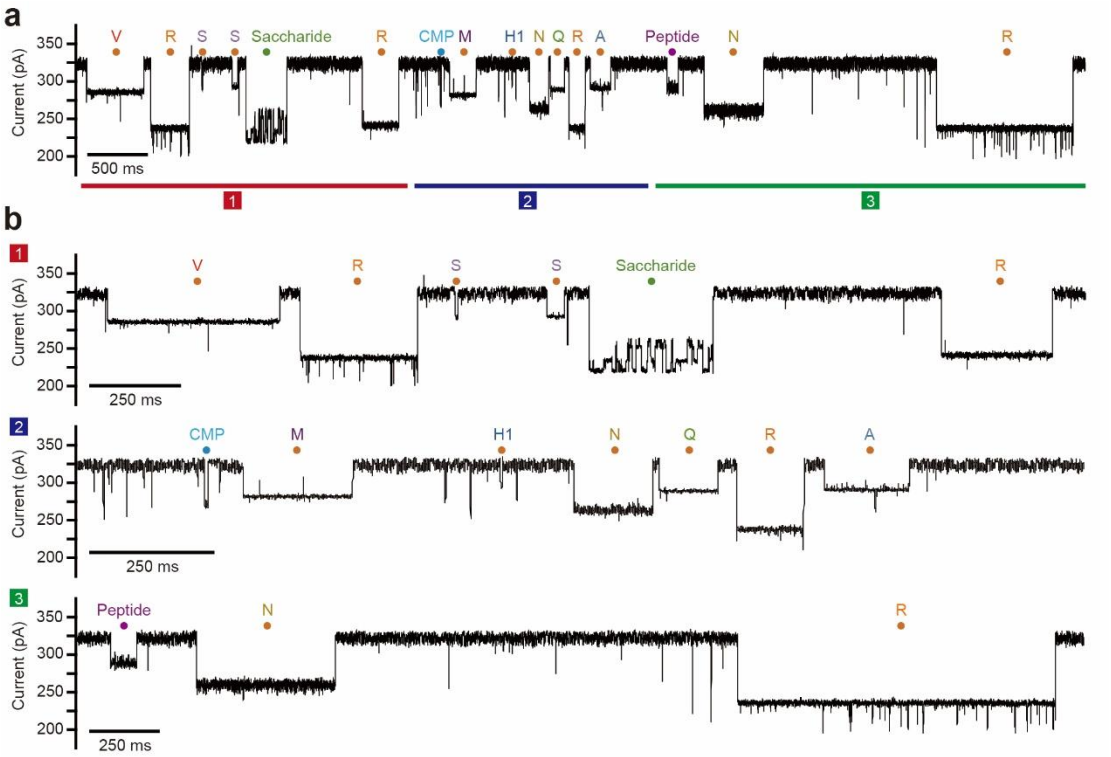
Supplementary Fig. 47: Analysis of yeast cell extract by MspA-FPBA. The yeast cell extract, a commercially available nutrient-rich additive, was dissolved in ultrapure water with a concentration of 100 mg mL^{-1}. Afterwards, the yeast cell extract solution was loaded into an ultracentrifuge tube with a 3 kDa molecular weight cut-off. The filtration was then performed at 6,000 r.p.m. for 60 min at 4 °C. Then, the pH of the ultrafiltrate was adjusted to 7.0 by titration with KOH. To initiate the nanopore measurement, 5 μL of the above filtrate was added to the *cis* side and a +160 mV bias was continuously applied. Finally, various nanopore events were classified and identified using machine learning.



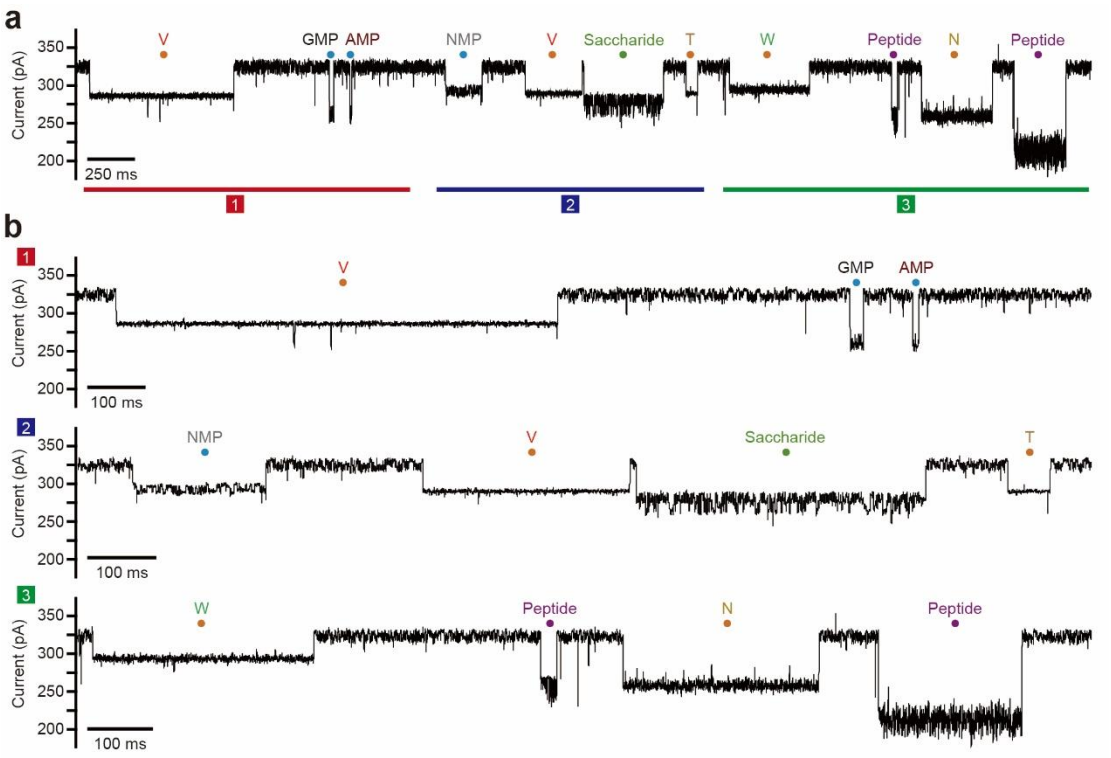
Supplementary Fig. 48: The extended representative current trace of yeast cell extract after ultrafiltration. The yeast cell extract was dissolved in ultrapure water with a concentration of 100 mg mL^{-1}. Afterwards, the yeast cell extract solution was loaded into an ultracentrifuge tube with a 3 kDa molecular weight cut-off. The filtration was then performed at 6,000 r.p.m. for 60 min at 4 °C. Then, the pH of the ultrafiltrate was adjusted to 7.0 by titration with KOH. To initiate the nanopore measurement, 5 µL of the above filtrate was added to the *cis* side and a +160 mV bias was continuously applied. Finally, various nanopore events were classified and identified using machine learning. For clarity, I_{FPBA} was marked by a gray dashed line. The transient spike signals similar to nanopore blockage were marked with gray arrows. Events were automatically classified via the 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.



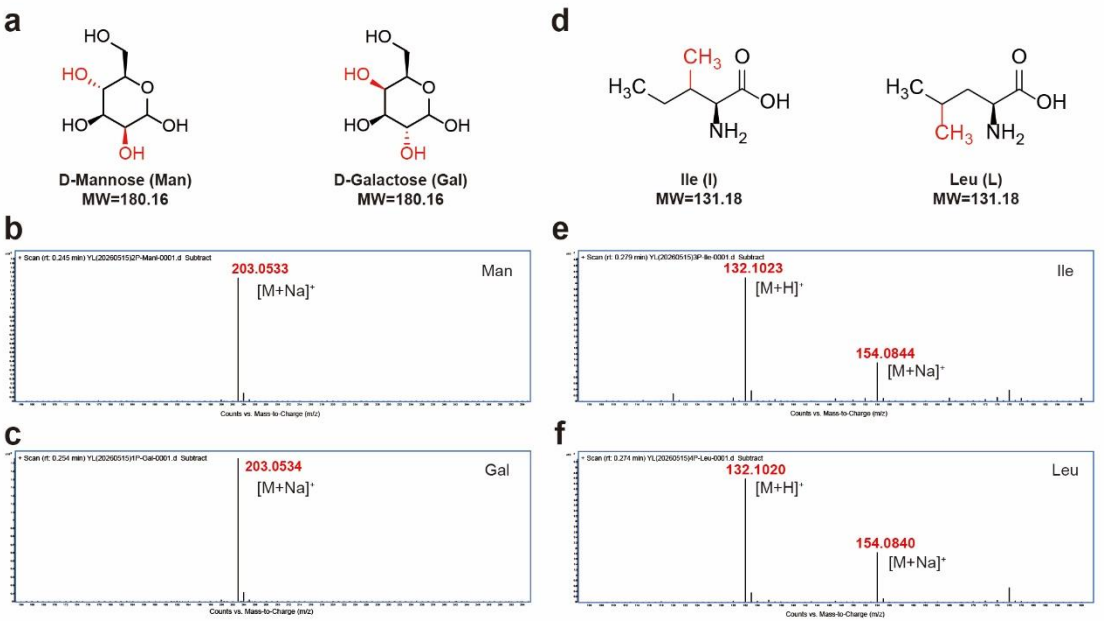
Supplementary Fig. 49: Analysis of yeast cell extract without ultrafiltration. The yeast cell extract was dissolved in ultrapure water to a concentration of 100 mg mL^{-1}, without being subjected to ultrafiltration. Then, the pH of the solution was adjusted to 7.0 by titration with KOH. To initiate the nanopore measurement, 5 µL of the above solution without ultrafiltration was added to the *cis* side and a +160 mV bias was continuously applied. Subsequently, frequent nanopore blockages were observed. The measurement was eventually terminated due to severe nanopore clogging. Nanopore clogging can be effectively removed by flipping of the applied voltage. For clarity, I_{FPBA} was marked by a gray dashed line, and the nanopore blockages were labeled with gray dots.



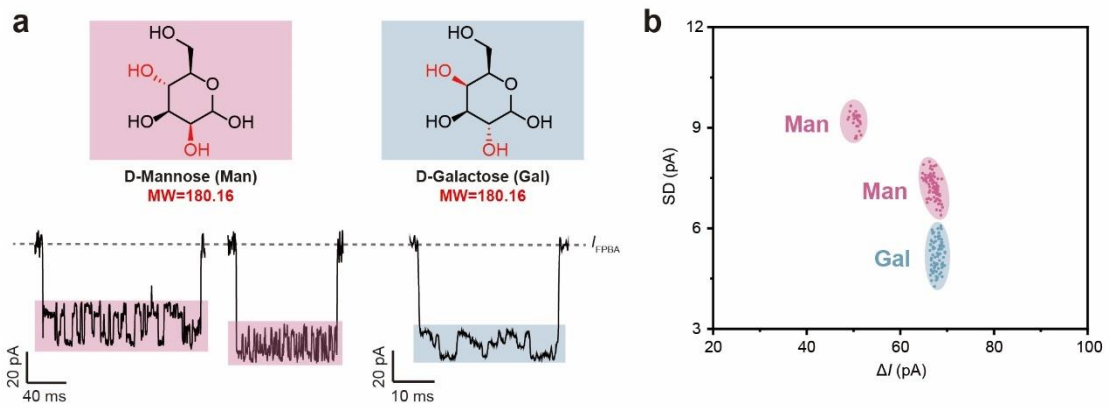
Supplementary Fig. 50: Zoomed-in view of Fig. 5b. (**a**) The raw current trace presented in **Fig. 5b**. All events were identified with the custom machine learning program and marked with corresponding identities. For a detailed demonstration, the trace is separated into three segments marked with 1 to 3. (**b**) The zoomed in view of the corresponding trace segments demonstrated in (**a**).



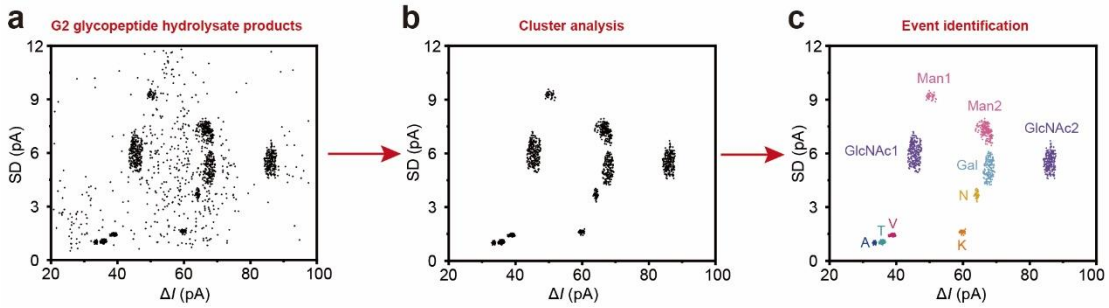
Supplementary Fig. 51: Zoomed-in view of Fig. 5c. (**a**) The raw current trace presented in **Fig. 5c**. All events were identified with the custom machine learning program and marked with corresponding identities. For a detailed demonstration, the trace is separated into three segments marked with 1 to 3. (**b**) The zoomed in view of the corresponding trace segments demonstrated in (**a**).



Supplementary Fig. 52: Saccharide and amino acid isomers resolved with mass spectrum. (**a**) The chemical structures of D-Mannose (Man) and D-Galactose (Gal). Man and Gal are isomers. (**b-c**) MS spectra of (**b**) Man and (**c**) Gal. Their molecular ion peaks are exactly the same. (**d**) The chemical structures of isoleucine (Ile, I) and leucine (Leu, L). Ile and Leu are isomers with identical mass. (**e-f**) MS spectra of (**e**) Ile and (**f**) Leu. Their molecular ion peaks are exactly the same.



Supplementary Fig. 53: A detailed comparison between D-Mannose and D-Galactose events. 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**) Top: The chemical structures of D-Mannose (Man) and D-Galactose (Gal). Man and Gal are isomers with identical mass. Bottom: Representative events of Man and Gal. I_{FPBA} represents the open pore current of MspA-FPBA. Events caused by Man and Gal are easily discriminable. (**b**) The event scatter plot of Δ*I* versus SD generated from Man and Gal events. Man (n = 103) and Gal (n = 82) events were employed to generate the statistics. Though Man and Gal have identical MW, they are fully discriminated by nanopore.



Supplementary Fig. 54: Background reduction and event prediction. (**a**) The raw scatter plots of Δ*I* versus SD of events acquired with G2 glycopeptide hydrolysate products. (**b**) The scatter plot of events acquired with G2 glycopeptide hydrolysate products however after background reduction treatment. Here, the background reduction treatment was performed with the DBSCAN algorithm. The epsilon was set to 0.5 and the minimum points was set to 5. Scatter dots that don't form a clear cluster were removed after the treatment. (**c**) Prediction of events performed by machine learning. Amino acid and GlcNAc event identities were predicted by the previously trained bagged trees model and labeled with color-coded dots. Man and Gal event identities were identified based on the characteristics of their representative events in **Supplementary Fig. 53**.