Supporting Information

Site-Specific Post-Translational Modification Detection by Polar Charged Engineered MspA Nanopores

Shijun Lin[a], Yakun Yi[b], Yiheng Liu[a], Minglun Li[c], Fan Xia[a*], Hai-Chen Wu[b*], Xiaoding Lou[a*]

a State Key Laboratory of Geomicrobiology and Environmental Changes, Faculty of Materials Science and Chemistry, China University of Geosciences, Wuhan 430074, China. louxiaoding@cug.edu.cn; xiafan@cug.edu.cn

b Beijing National Laboratory for Molecular Sciences, Key Laboratory of Analytical Chemistry for Living Biosystems, Institute of Chemistry, Chinese Academy of Sciences, Beijing 100190, China. haichenwu@iccas.ac.cn

c State Key Laboratory of Polymer Science and Technology, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun 130022, China

Table of Contents

Experimental Procedures

1. Materials and reagents

1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC) was supplied by Avanti Polar Lipids. E. ColiBL21(DE3) were purchased form Sangon Biotech. Guanidine hydrochloride and agarose were purchased form Biofoxx. Urea, ethylenediaminetetraacetic acid (EDTA), sodium acetate, disodium hydrogen phosphate (Na_2HPO_4), sodium dihydrogen phosphate (NaH_2PO_4), hydrochloric acid (HCl), sodium chloride (NaCl) and sodium hydroxide (NaOH) were purchased form Sinopharm (China). Imidazole, nickel sulfate, isopropyl-β-D-thiogalactopyranoside (IPTG) and Genapol X-80 were purchased form Aladdin (China). Yeast extract, tryptone, 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid (HEPES) were purchased from Shanghai Yuanye Bio-Technology (China). Peptides were purchased from Nanjing Yuanpeptide Biotechnology Co., Ltd, QYAOBIO (ChinaPeptides Co., Ltd.) and GL Biochem (Shanghai) Ltd. All the mass spectrometry of the peptides was shown in FigureS53-S78. The MspA nanopore with different mutations at site 91 were constructed and purified according to reference 1.

2. Nanopore preparations

All engineered MspA were generated based on the WT MspA sequence, and the specific mutation sites are shown in Table S1. The WT MspA sequence is as follows:

MGLDNELSLVDGQDRTLTVQQWDTFLNGVFPLDRNRLTREWFHSGRAKYIVAGPGADEFEGTLELGYQIGFP
WSLGVGINFSYTTPNILINNGNITAPPFGLNSVITPNLFPGVSISARLGNGPGIQEVATFSVRVSGAKGGVAVSN
AHGTVTGAAGGVLLRPFARLIASTGDSVTTYGEPWNMNHHHHHH*

The amino acid at the narrowest restriction of an octameric MspA nanopore was engineered. The gene coding for M2 MspA, MspA-Q, MspA-D, MspA-E, MspA-H, MspA-R, MspA-K, MspA-F and MspA-W were custom synthesized and constructed in a pet 30a(+) plasmid (Genscript, New Jersey) respectively. Briefly, the Q, D, E, H, R, K, F, and W character in the name of each gene stands for the glutamine, aspartic acid, glutamic acid, histidine, arginine, lysine, phenylalanine and tryptophan at 91 site. The constructed plasmid gene was heat-shock transformed into E. ColiBL21(DE3). Then, the cells were grown in LB medium(contain 50 μg/ml Kanamycin) at 37°C and 225 rpm until the optical density at 600 nm (OD_{600}) reach 0.7. After cooling the medium to 16°C, IPTG with a final concentration of 1 mM was added to induce protein expression, and the culture

was shaken at 225 rpm overnight. Afterwards, the cells were harvested by centrifugation (6000 rpm, 10 min, 4 °C).

The bacterial pellet was resuspended in a 100 mL lysis buffer (100 mM Na_2HPO_4/NaH_2PO_4, 0.1 mM EDTA, 150 mM NaCl, 0.5% (w/v) Genapol X-80, pH=6.5) and heated at 60 °C for 10 min. The suspension was cooled on ice for 10 min and centrifuged at 4 °C for 30 min at 12000 rpm. Then, the protein mixture was purified using nickel affinity chromatography and eluted with a linear gradient of imidazole (5 mM-500 mM) by mixing buffer A (0.5 M NaCl, 20 mM HEPES, 5 mM imidazole, 0.5% (w/v) Genapol X-80, pH=8.0) with buffer B (0.5 M NaCl, 20 mM HEPES, 500 mM imidazole, 2 mM TCEP, 0.5% (w/v) Genapol X-80, pH=8.0). The eluted fractions were further characterized by 4-20% SDS-PAGE gel to identify the octameric MspA. Subsequently, the protein solution was subjected to multiple rounds of dilution and ultrafiltration using ultrafiltration centrifugal units with a molecular weight cutoff of 50 kDa to reduce the imidazole concentration in the protein buffer. Then, the proteins were stored at -80 ℃ for long term storage.

3. Nanopore measurements

Nanopore recordings were performed using the Orbit16 device (Nanion Technologies, Munich, Germany) and MECA16 microelectrode cavity arrays (Ionera Technologies, Freiburg, Germany) as previously detailed. The self-assembled lipid bilayer was formed with DPhPC (5 mg/mL, dissolved in octane). The measurement buffer was prepared with 1 M NaCl, 10 mM HEPES at pH=7.4. During the experimental process, 200 µL of measurement buffer was added to the MECA16 microelectrode cavity array. A magnetic stir bar was used to uniformly coat DPhPC onto the microcavities. Current detection was performed to validate the chip's conductivity and confirm the formation of the lipid bilayer. Subsequently, 1 µL of MspA nanopore solution was titrated onto the top of the microcavities, and insertion of a single nanopore was awaited. Upon insertion, excess nanopores were removed by exchanging the buffer in the compartment.

All electrophysiology measurements were performed with a continuously applied +50 mV potential. All experiments were conducted at room temperature (25 ± 2 °C). Except for the glycosylated peptide, which was used at a concentration of 50 µM, all other peptides were used at a concentration of 5 µM. The ionic current through a single nanopore was acquired with a 10 kHz sampling rate and low-pass filtered with a 1 kHz corner frequency. To accurately resolve fast

kinetics, higher sampling bandwidth and more optimized filtering settings can also be employed. All channel binding events were detected by performing “single-channel search” using Clampfit 10.6 (Molecular Devices) and further analyzed (histogram generation, curve fitting and plotting) using Origin2021.

4. Machine learning

To differentiate distinct post-translational modification (PTM) groups based on extracted features, MATLAB’s built-in Classification Learner application was utilized. The preprocessed feature data (eg. relative residual current, dwell time, standard deviation, and instantaneous frequency of the signal), formatted as a table with each row representing a sample and each column corresponding to a specific feature, were imported into MATLAB, where classification labels representing different experimental groups were assigned as the response variable. Multiple classification models were systematically evaluated using a 10-fold cross-validation strategy to ensure robust model assessment. The optimal model was selected based on classification accuracy and generalization performance on unseen data, prioritizing the classifier with the highest predictive reliability.

5. Maldi TOF and HPLC analysis

The MspA were heated at 100°C for 10 min to generate protein monomers. The desalted sample was analyzed and quantified (detection range: 10,000–30,000 Da) prior to Maldi TOF analysis. Prior to injection into the mass analyzer, the desalted samples were separated through an ultraperformance liquid chromatography system employing a YMC C4 column with H_2O/0.1% formic acid as mobile (phase A), followed by gradient elution using acetonitrile/0.1% formic acid (phase B) at a flow rate of 0.5 mL/min.

6. MD simulations

All molecular dynamics (MD) simulations were performed using the Gromacs package[2]. The system consists of peptides (HY, HY-1p, HY-2p, and HY-3p), the protein nanopore (MspA)[3], DPhPC phospholipids, water, and ions. The peptide HY was constructed by Discovery Studio (BIOVIA,

Dassault Systèmes), and the phosphorylation modifications of Tyrosine on HY-1p, HY-2p, and HY-3p were achieved by CHARMM-GUI[4]. Wild-type (WT) MspA was obtained from Protein Data Bank (PDB) with the entry ID: 1UUN[5]. Both the MspA-K mutant (D90N/D91K/D93N/R96A/D118R/D134R/E139K) and the DPhPC structure were constructed from the WT MspA utilizing CHARMM-GUI[6]. A detailed description of the mutation protocol can be found in our earlier publications. [J. Chem. Phys. 160, 084905 (2024)] The components were placed in box with initial dimensions of 10.01383×10.01383×12.50000 nm^3. The salt (NaCl) concentrations are 1M in all simulations. To investigate the impact of phosphorylation modifications on the translocation behavior of peptides, two initial configurations for the peptide chains were chosen: one placing Y2 on the peptide near the N91K of MspA and the other placing Y8@Y10 near N91K.

The CHARMM36 force field was applied for the system[7], with the TIP3P model[8]. Electrostatic interactions are evaluated using a particle-mesh Ewald sum, with a real-space cutoff of 1.2 nm. Van der Waals interactions were evaluated using a smooth 1.0-1.2 nm cutoff. Langevin dynamics were used to maintain the temperature at 300.15 K.

During the whole simulation process, each carbon atom of the MspA protein was harmonically restrained (kSPRING = 1 kcal mol^{-1} nm^{-2}) to its initial coordinate. Each translocation system was minimized in 2,000 steps using the conjugate gradient method and then equilibrated for 45 ns at a constant number of atoms, pressure, and temperature (NPT) ensemble performed while keeping the ratio of the systems size along the place of the bilayer constant. The systems were then simulated for 50 ns in a constant number of particles, volume and temperature (NVT) ensemble. Both the timesteps for NPT and NVT above are 0.001 ps. The systems were then simulated another 300 ns in NVT ensemble under a constant electric field $E = -V/L_z$ applied along the z-axis (normal to the membrane) to produce a transmembrane bias V = 50 mV; where L_z is the dimension of the simulated system in the direction of the applied electric field. For the second NVT ensemble the time step is 0.002 ps, the data in the last 100 ns were used for analysis. Each simulation case was repeated three times with different seeds. The MD simulation results were analyzed using tools from the GROMACS 2022.6 package, MDAnalysis[9], gmxMMPBSA10, Pymol, and ChimeraX 1.9[11].

Instantaneous ionic current was followed by the studies from Aleksei Aksimentiev's group[12] where $[z_j(t+\Delta t) - z_i(t)]$ is the displacement of ion j along the z direction during the time interval Δt =

50 ps and qj is the charge of ion j. To minimize thermal noise, the current was calculated within an lz = 20 Å thick slab centered at the nanopore constriction, spanning the entire simulation system in the x-y plane. The instantaneous ionic current values were recorded along with the center of mass of the amino acid substitution's backbone. For each substitution, the data from all trajectories were sorted by the z-coordinate of the substitution in ascending order.

$$I(t) = \frac{1}{\Delta t l_z} \sum_{j=1}^{N} q_j \left[z_j(t + \Delta t) - z_i(t) \right]$$

The Molecular docking calculations were performed to validate the interaction patterns between the sensing region (site N91K) and the peptides PD, PD-cr, and PD-p. The structural files for docking were prepared using AutoDock Tools 1.5.7. Molecular docking simulations were conducted with AutoDock Vina, employing a cubic search box (edge length: 30 Å) and a grid spacing of 0.375 Å. The maximum number of binding modes was set to 100, and conformational sampling was executed via a genetic algorithm. The resulting poses were ranked by docking scores, followed by structural validation to select the most energetically and sterically favorable conformation.

Results and Discussion

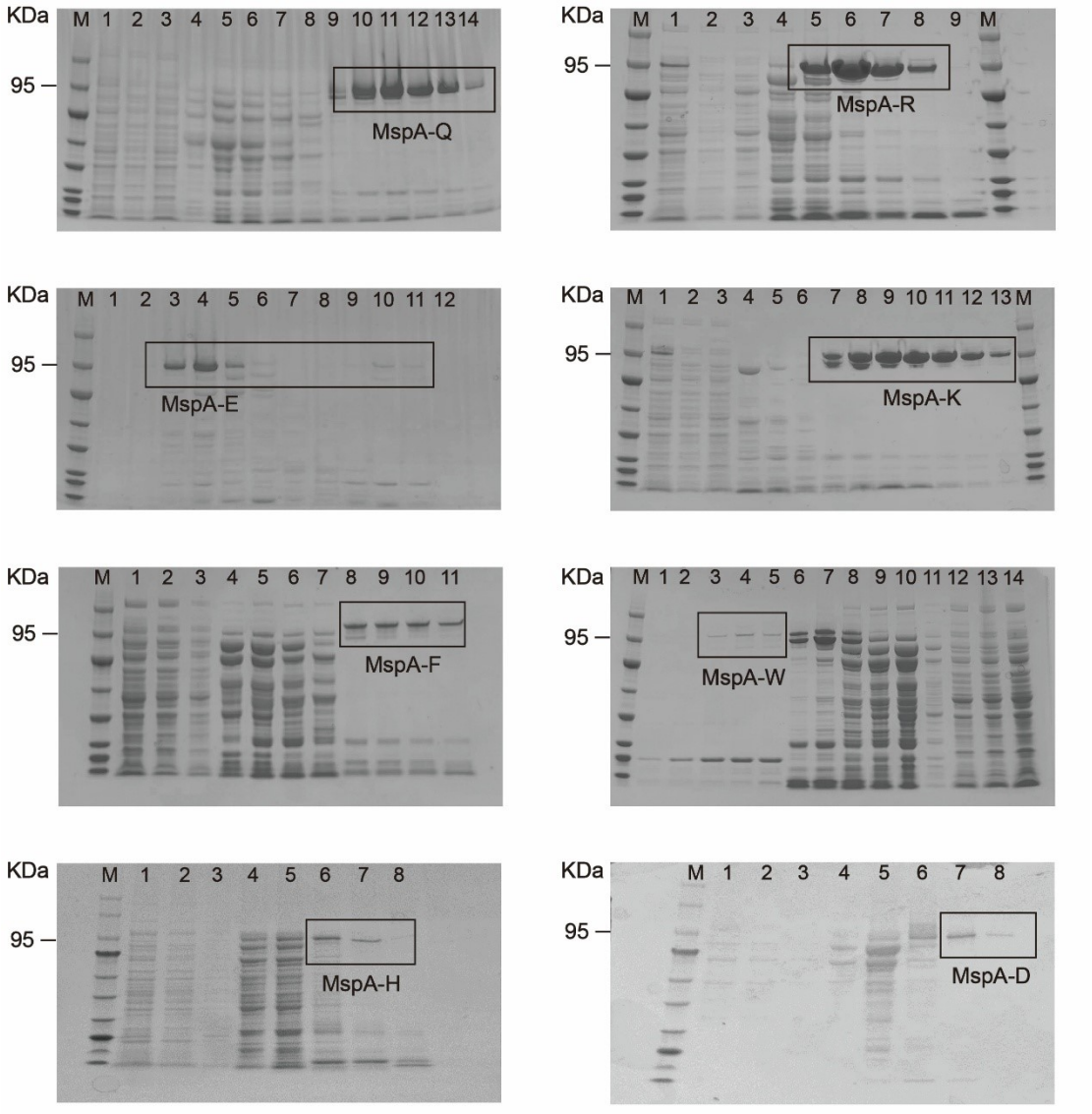

Figure S1. The preparation of MspA mutant nanopores. SDS-polyacrylamide gel electrophoresis (4-20% gradient gel) of fractions collected during MspA purification. Gel electrophoresis was continuously run for 50 min with a +140 V applied potential. Lane M: MCE 3-Color Prestained Protein Marker (10-190 kDa). Other lanes: the eluent of the bacterial lysate after column loading.

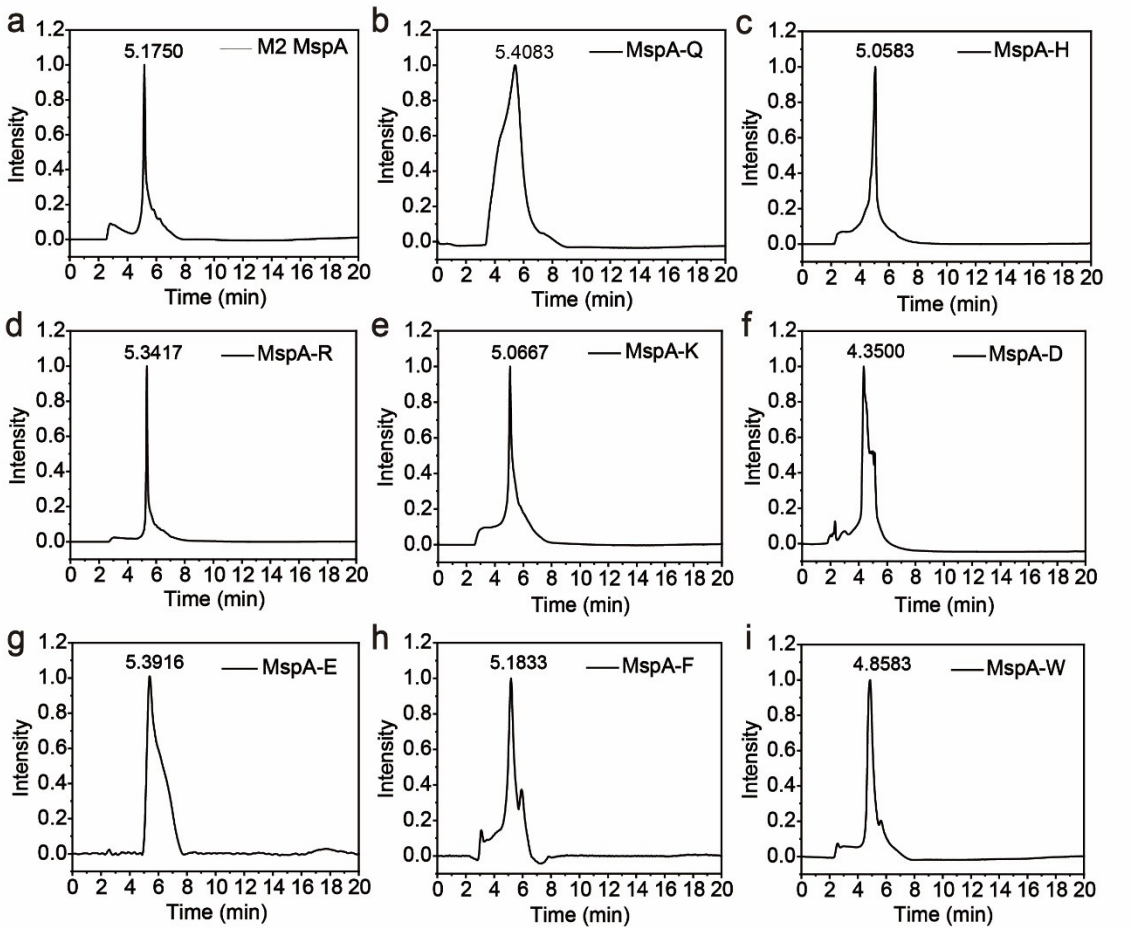

Figure S2. HPLC analysis of MSPA mutants. (a) The elution peak of M2 MspA at a retention time of 5.18 min. (b) The elution peak of MspA-Q at a retention time of 5.41 min. (c) The elution peak of MspA-H at a retention time of 5.06 min. (d) The elution peak of MspA-R at a retention time of 5.34 min. (e) The elution peak of MspA-K at a retention time of 5.07 min. (f) The elution peak of MspA-D at a retention time of 4.35 min. (g) The elution peak of MspA-E at a retention time of 5.39 min. (h) The elution peak of MspA-F at a retention time of 5.18 min. (i) The elution peak of MspA-Q at a retention time of 4.86 min. The mobile phase is water/acetonitrile.

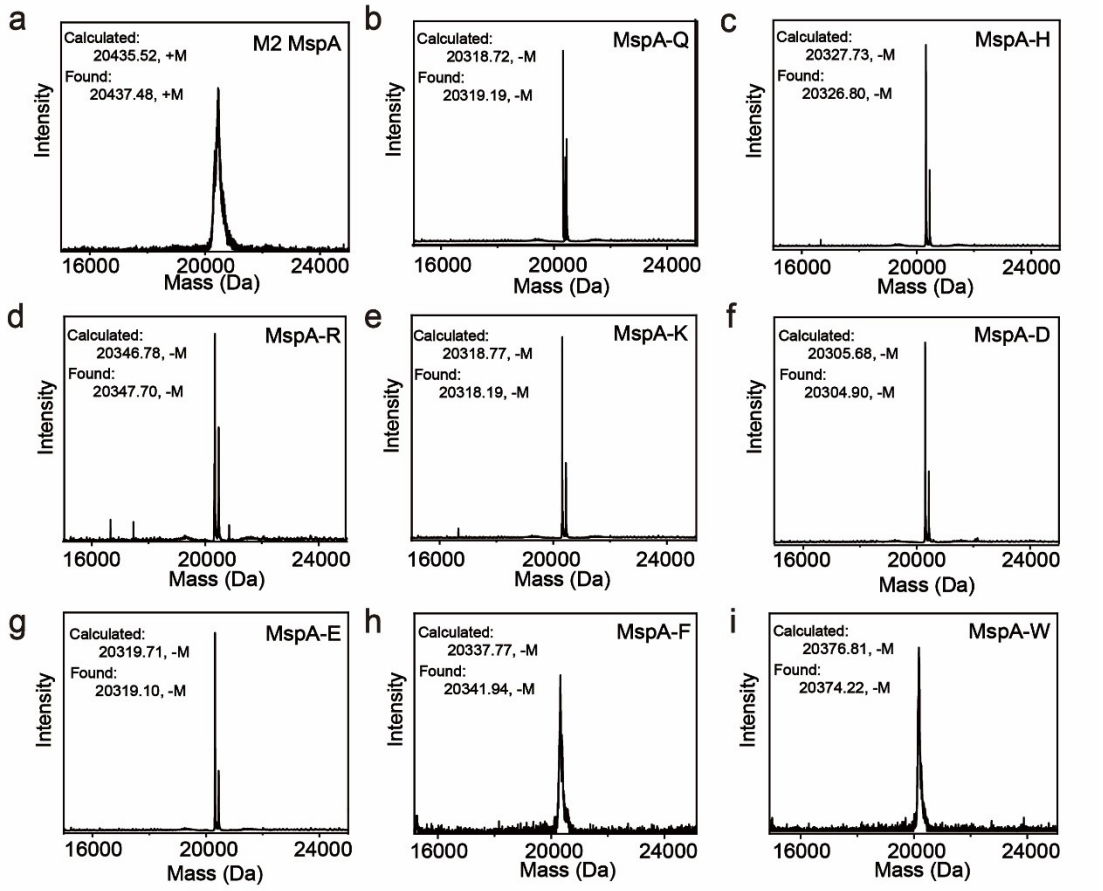

Figure S3. MALDI-TOF analysis of MspA mutants. (a) M2 MspA. Calculated: 20435.52 Da, +M. Found: 20437.48 Da, +M. (b) MspA-Q. Calculated: 20318.72 Da, -M. Found: 20319.19 Da, -M. (c) MspA-H. Calculated: 20327.73 Da, -M. Found: 20326.80 Da, -M. (d) MspA-R. Calculated: 20346.78 Da, -M. Found: 20347.70 Da, -M. (e) MspA-K. Calculated: 20318.77 Da, -M. Found: 20318.19 Da, -M. (f) MspA-D. Calculated: 20305.68 Da, -M. Found: 20304.90 Da, -M. (g) MspA-E. Calculated: 20319.71 Da, -M. Found: 20319.10 Da, -M. (h) MspA-F. Calculated: 20337.77 Da, -M. Found: 20341.94 Da, -M. (i) MspA-W. Calculated: 20376.81 Da, -M. Found: 20374.22 Da, -M. The "+M" and "-M" indicate the mass of the protein monomer with or without the initiating methionine.

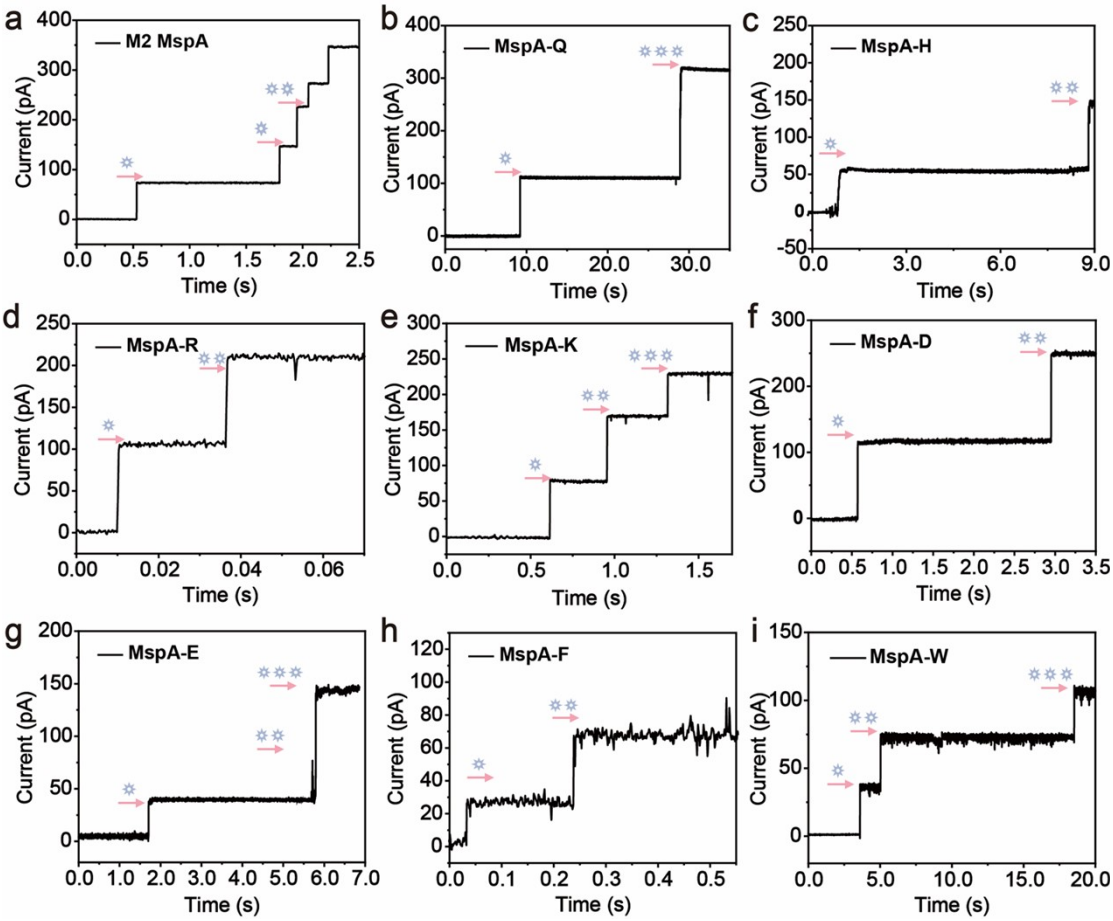

Figure S4. Successive insertions of MspA mutants (a: M2 MspA; b: MspA-Q; c: MspA-H; d: MspA-R; e: MspA-K; f: MspA-D; g: MspA-E; h: MspA-F; i: MspA-W) observed during current recording. Each current step corresponds to the sequential insertion of a nanopore into the lipid bilayer. Measurements were performed under a constant +50 mV applied voltage (1.0 M NaCl, 10.0 mM HEPES, pH 7.4).

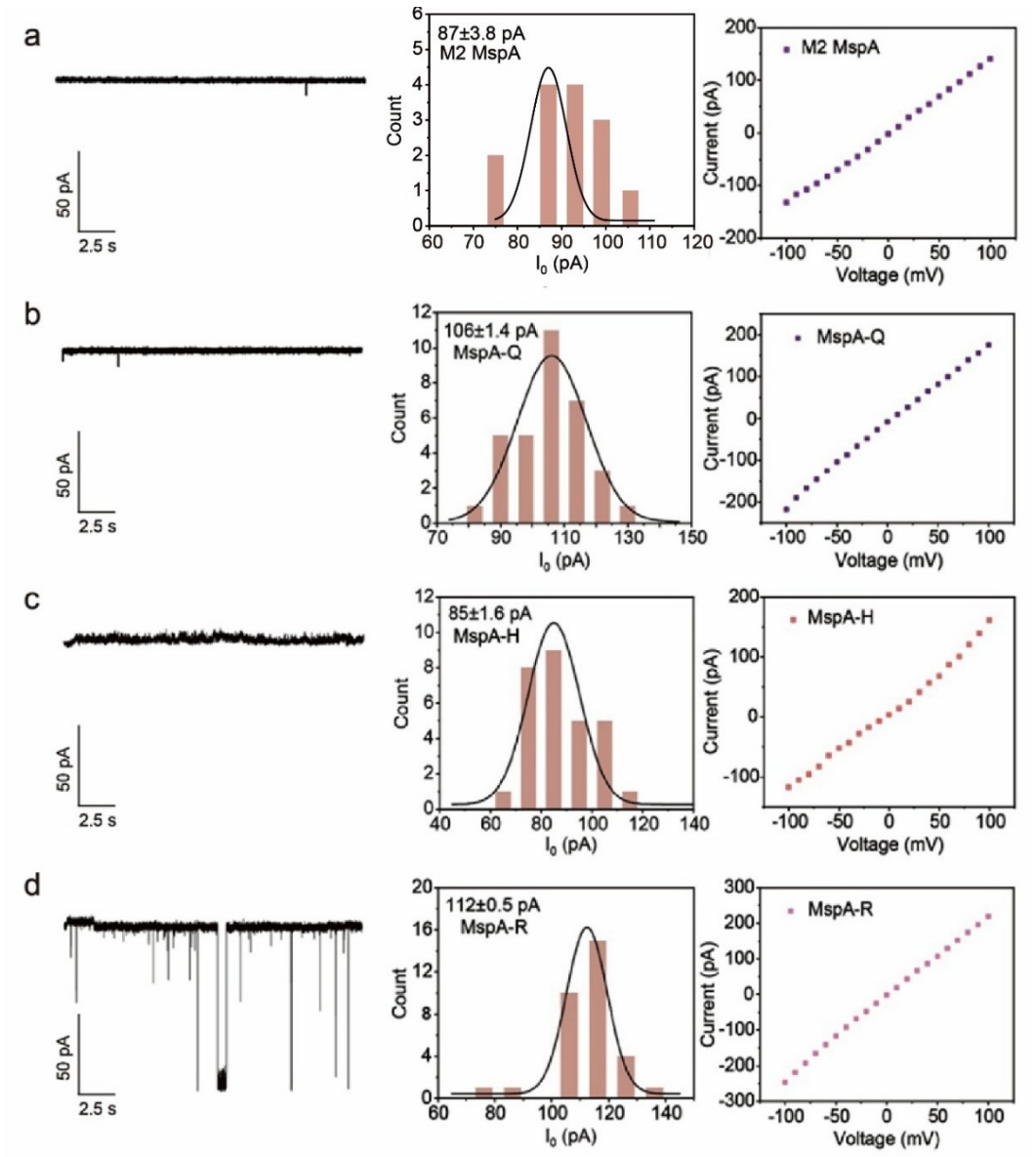

Figure S5. Representative current traces (Left), histogram of the open nanopore (I_0) (Middle) and I-V curves (Right) of the MspA nanopore (a: M2 MspA; b: MspA-Q; c: MspA-H; d: MspA-R). Representative current traces recorded and histogram of the open current at an applied potential of +50 mV. I-V curves measured by stepping the voltage from -100 mV to +100 mV. Owing to the pronounced baseline noise issue, MspA-R was excluded from further testing. The nanopore measurements were performed in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4.

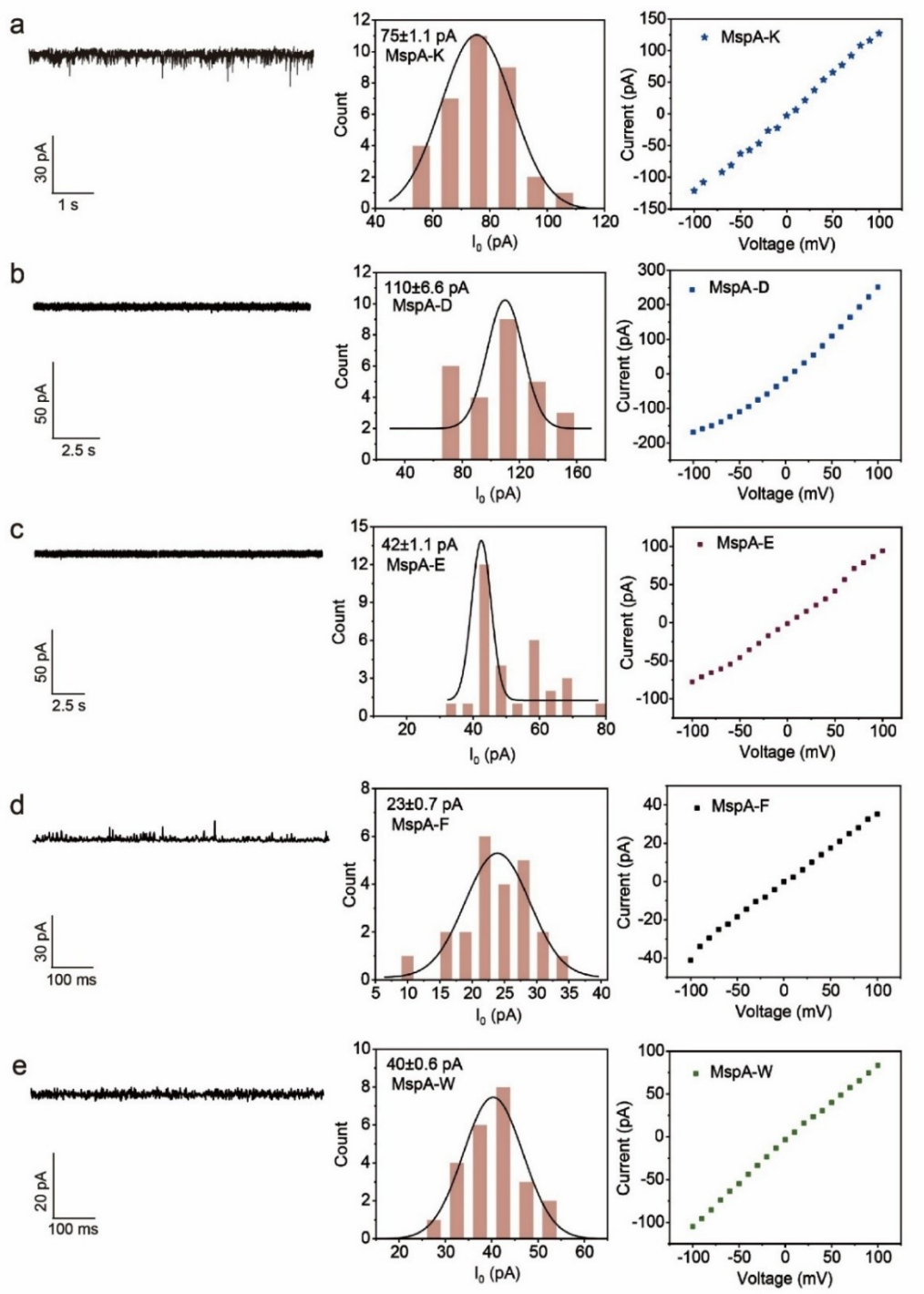

Figure S6. Representative current traces (Left), histogram of the open nanopore (I_0) (Middle) and I-V curves (Right) of the MspA nanopore (a: MspA-K; b: MspA-D; c: MspA-E; d: MspA-F; e: MspA-W). Representative current traces recorded and histogram of the open current at an applied potential of +50 mV. I-V curves measured by stepping the voltage from -100 mV to +100 mV. Owing to the pronounced baseline noise issue, MspA-R was excluded from further testing. The nanopore measurements were performed in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4.

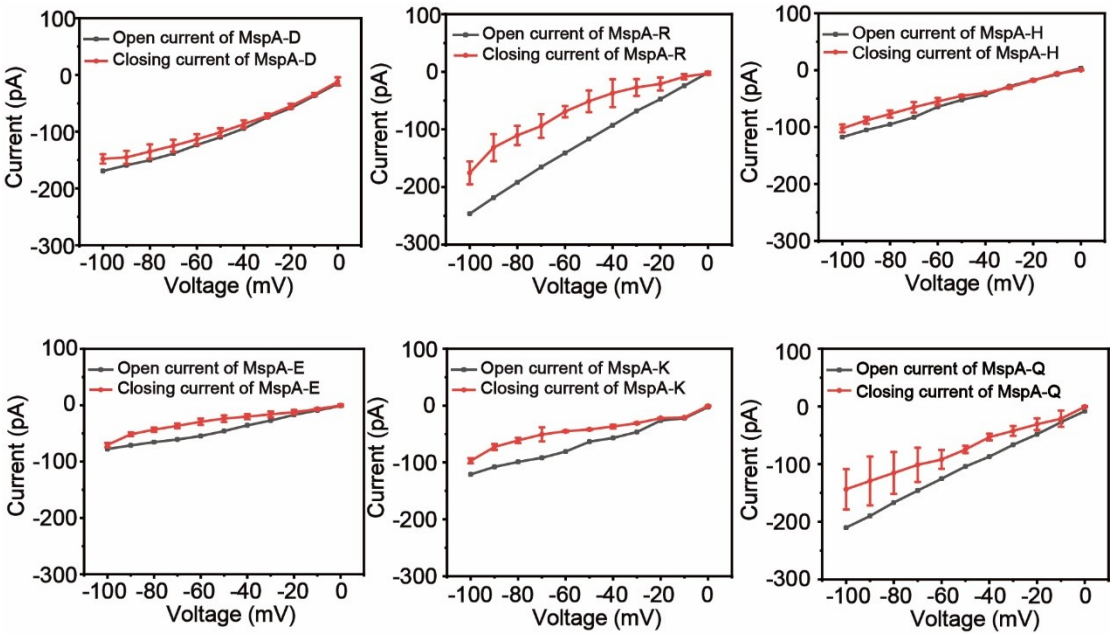

Figure S7. Gating behavior of MspA mutants under negative voltage. The nanopore measurements were performed in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4.

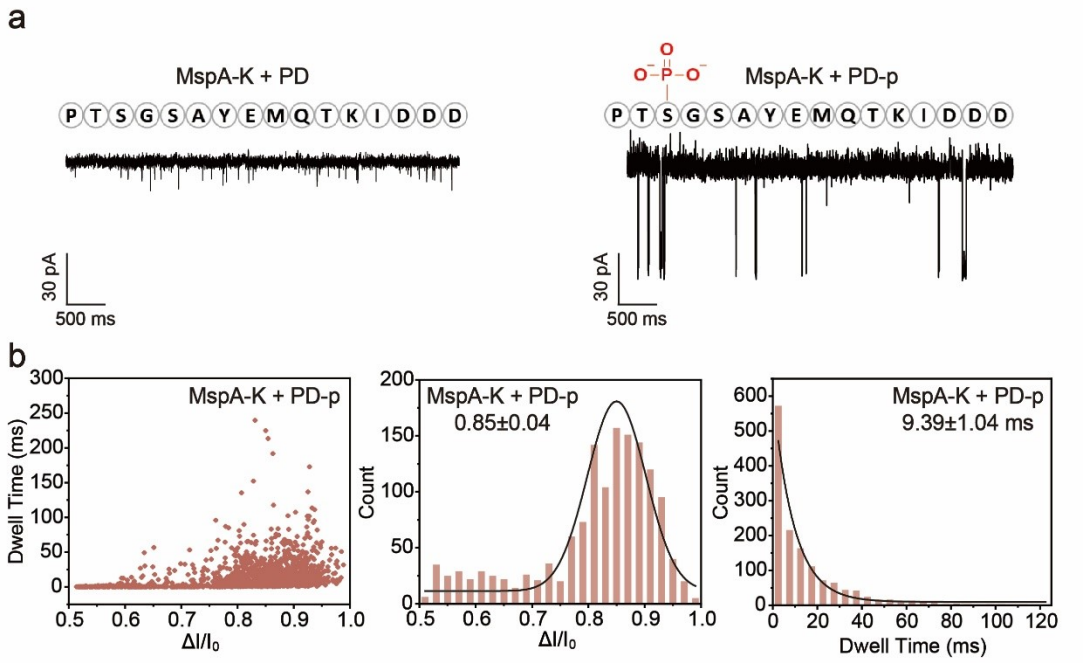

Figure S8. (a) Representative current traces of PD and PD-p with MspA-K. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4. All peptides were added to a final concentration of 5.0 μM. (b) Scatter plot (left) and histogram of $\Delta I/I_0$ (middle), and dwell time histogram (right) for PD-p interactions with MspA-K at an applied potential of +50 mV.

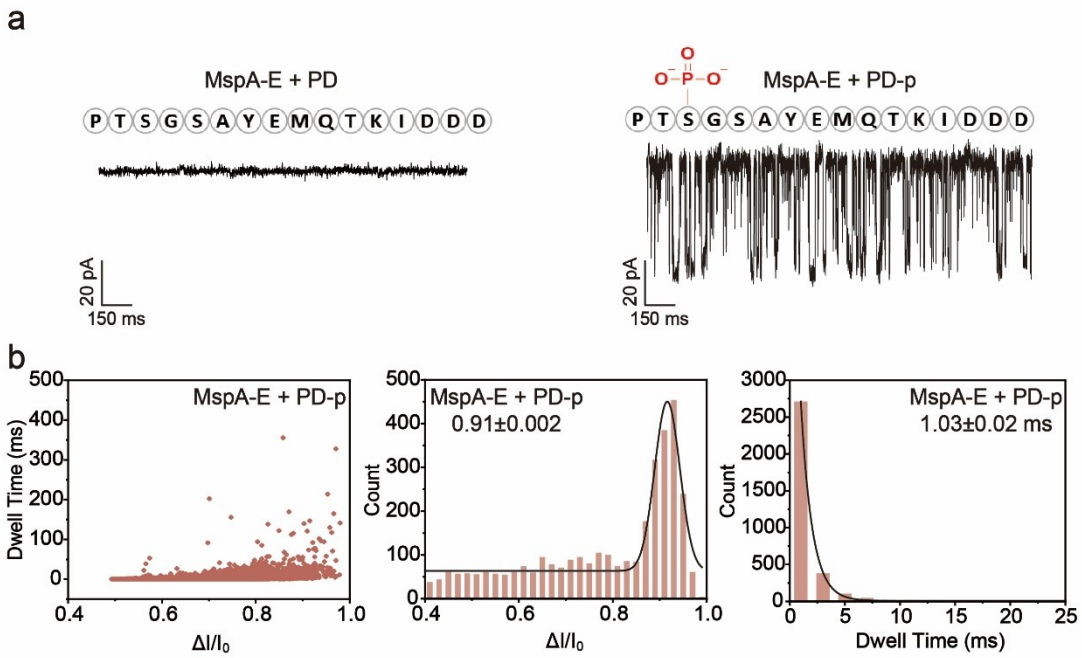

Figure S9. (a) Representative current traces of PD and PD-p with MspA-E. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4. All peptides were added to a final concentration of 5.0 μM. (b) Scatter plot (left) and histogram of $\Delta I/I_0$ (middle), and dwell time histogram (right) for PD-p interactions with MspA-E at an applied potential of +50 mV.

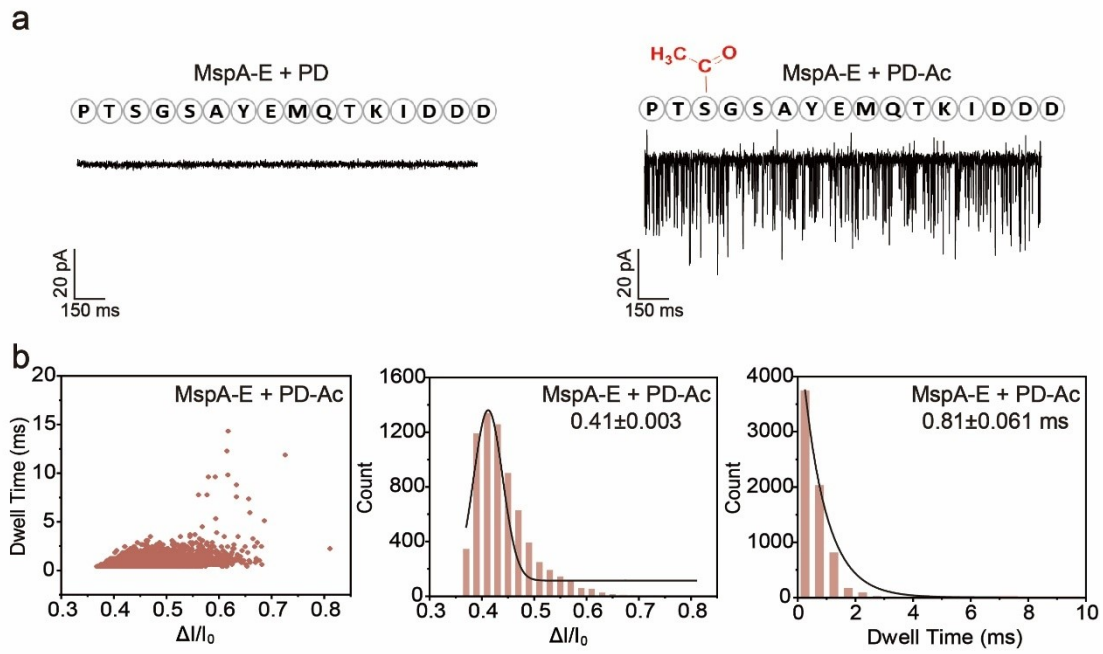

Figure S10. (a) Representative current traces of PD and PD-Ac with MspA-E. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4. All peptides were added with a final concentration of 5 μM for each analyte. (b) Scatter plot (left) and histogram of $\Delta I/I_0$ (middle), and dwell time histogram (right) for PD-Ac interactions with MspA-E at an applied potential of +50 mV.

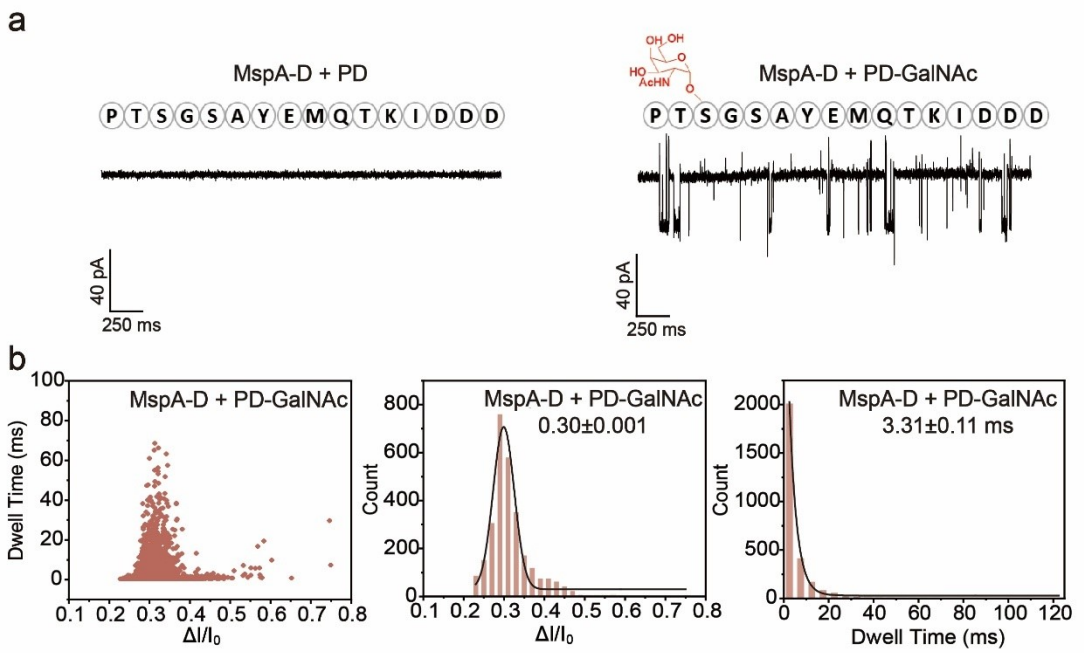

Figure S11. (a) Representative current traces of PD and PD-GalNAc with MspA-D. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4. All peptides were added with a final concentration of 50.0 μM for each analyte. (b) Scatter plot (left) and histogram of $\Delta I/I_0$ (middle), and dwell time histogram (right) for PD-GalNAc interactions with MspA-D at an applied potential of +50 mV.

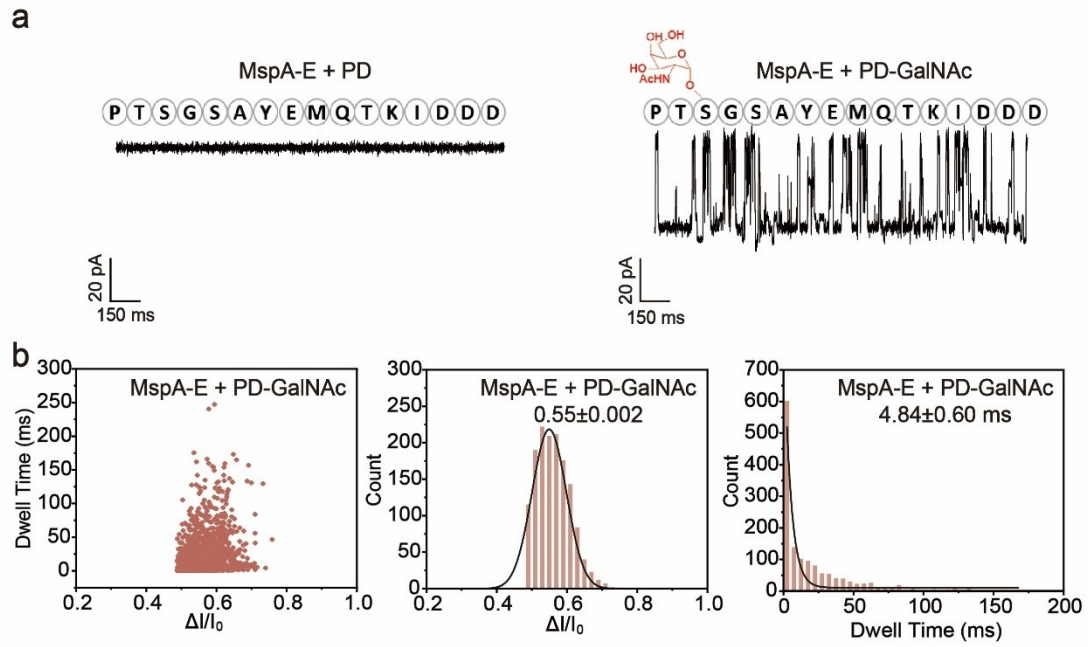

Figure S12. (a) Representative current traces of PD and PD-GalNAc with MspA-E. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4. All peptides were added with a final concentration of 50.0 μM for each analyte. (b) Scatter plot (left) and histogram of $\Delta I/I_0$ (middle), and dwell time histogram (right) for PD-GalNAc interactions with MspA-E at an applied potential of +50 mV.

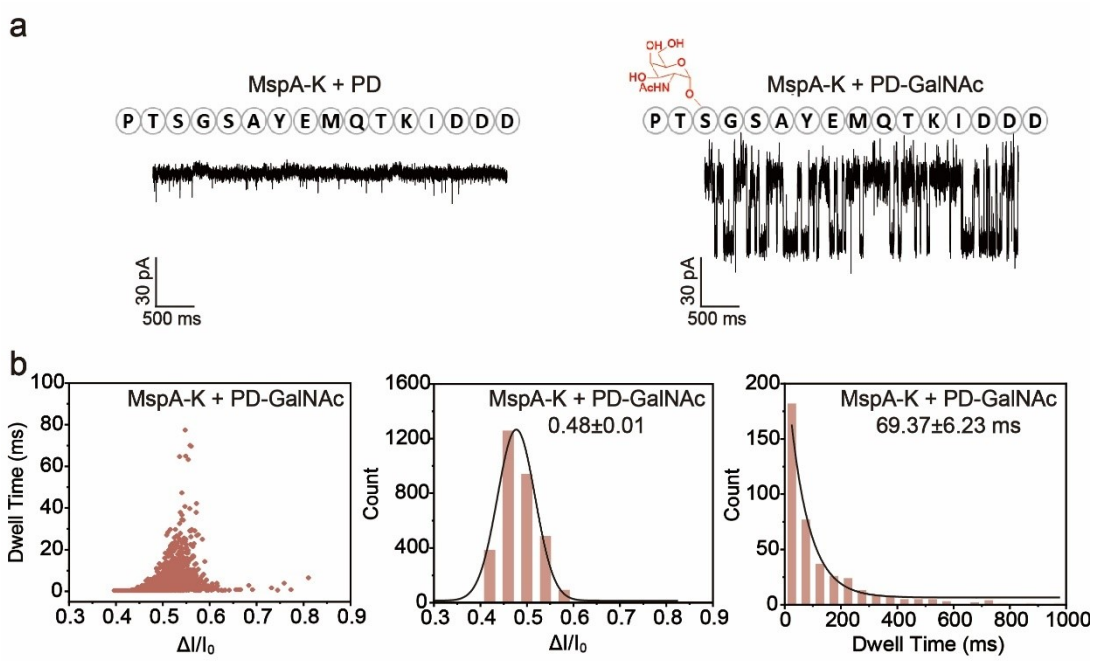

Figure S13. (a) Representative current traces of PD and PD-GalNAc with MspA-K. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4. All peptides were added with a final concentration of 50 μM for each analyte. (b) Scatter plot (left) and histogram of $\Delta I/I_0$ (middle), and dwell time histogram (right) for PD-GalNAc interactions with MspA-K at an applied potential of +50 mV.

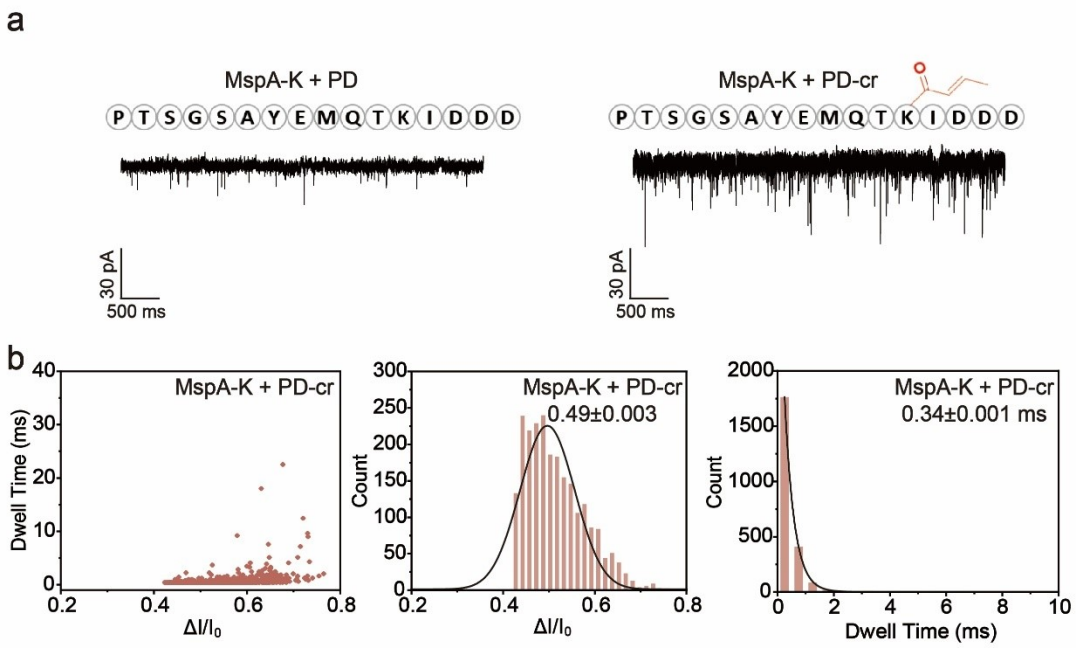

Figure S14. (a) Representative current traces of PD and PD-cr with MspA-K. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4. All peptides were added with a final concentration of 5 μM for each analyte. (b) Scatter plot (left) and histogram of $\Delta I/I_0$ (middle), and dwell time histogram (right) for PD-cr interactions with MspA-K at an applied potential of +50 mV.

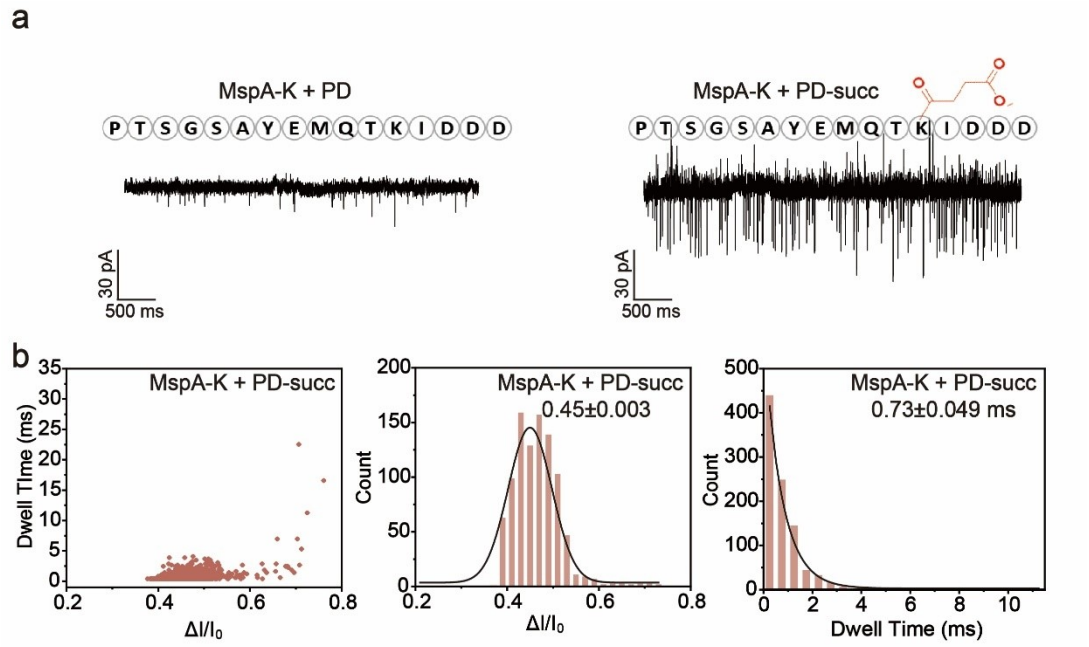

Figure S15. (a) Representative current traces of PD and PD-succ with MspA-K. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4. All peptides were added with a final concentration of 5 µM for each analyte. (b) Scatter plot (left) and histogram of $\Delta I/I_0$ (middle), and dwell time histogram (right) for PD-succ interactions with MspA-K at an applied potential of +50 mV.

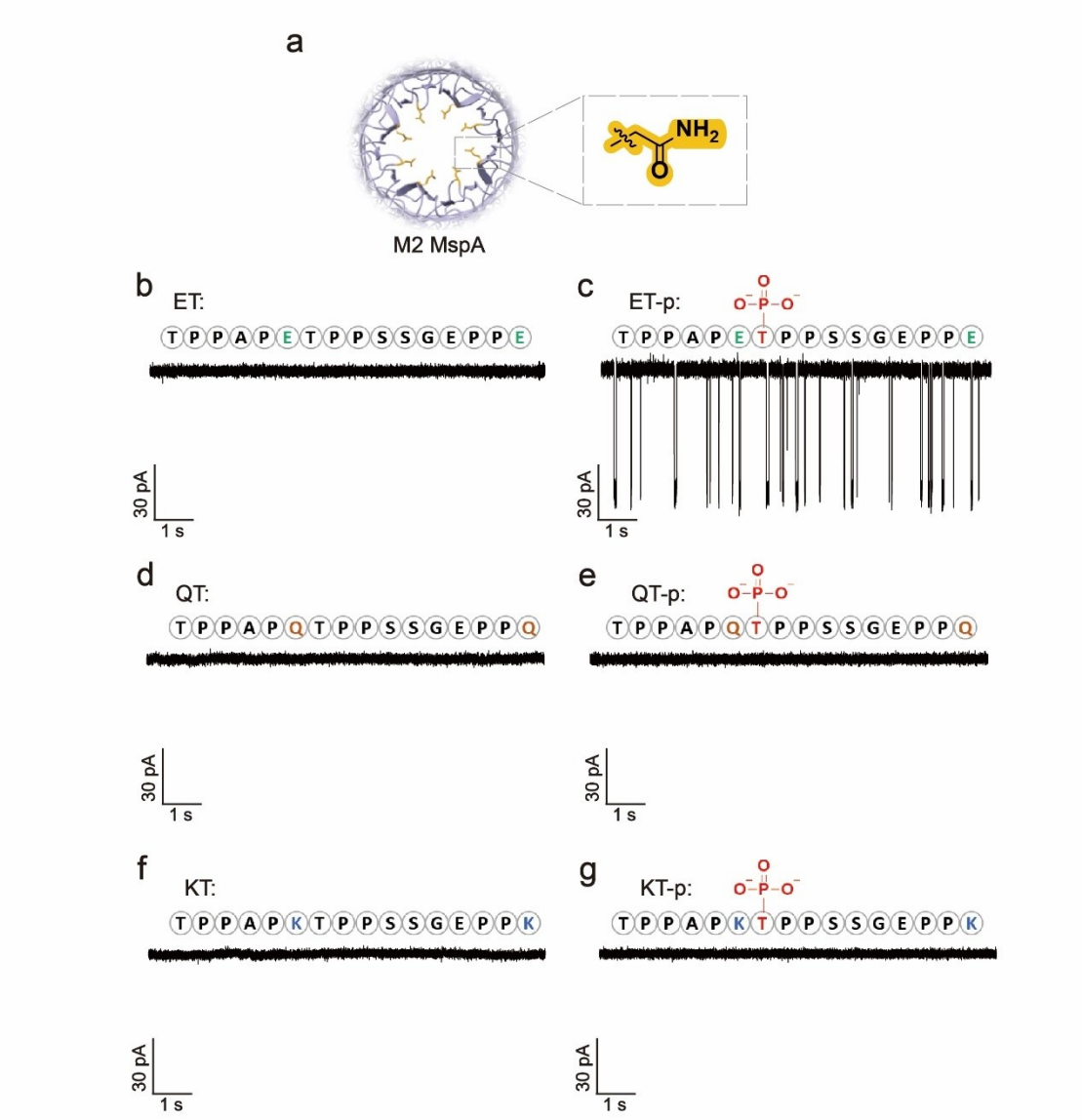

Figure S16. (a) Schematic diagram of the amino acid residue at site 91 within the M2 MspA nanopore. Representative current traces of the M2 MspA nanopore following the addition of ET (b), ET-p (c), QT (d), QT-p (e), KT (f) or KT-p (g), respectively as sole analytes. Successive sensing events were observed exclusively for ET-p, while no detectable events occurred with other peptides. Measurements were performed in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4, at an applied potential of +50 mV. Each peptide was tested separately at a final concentration of 5.0 μM.

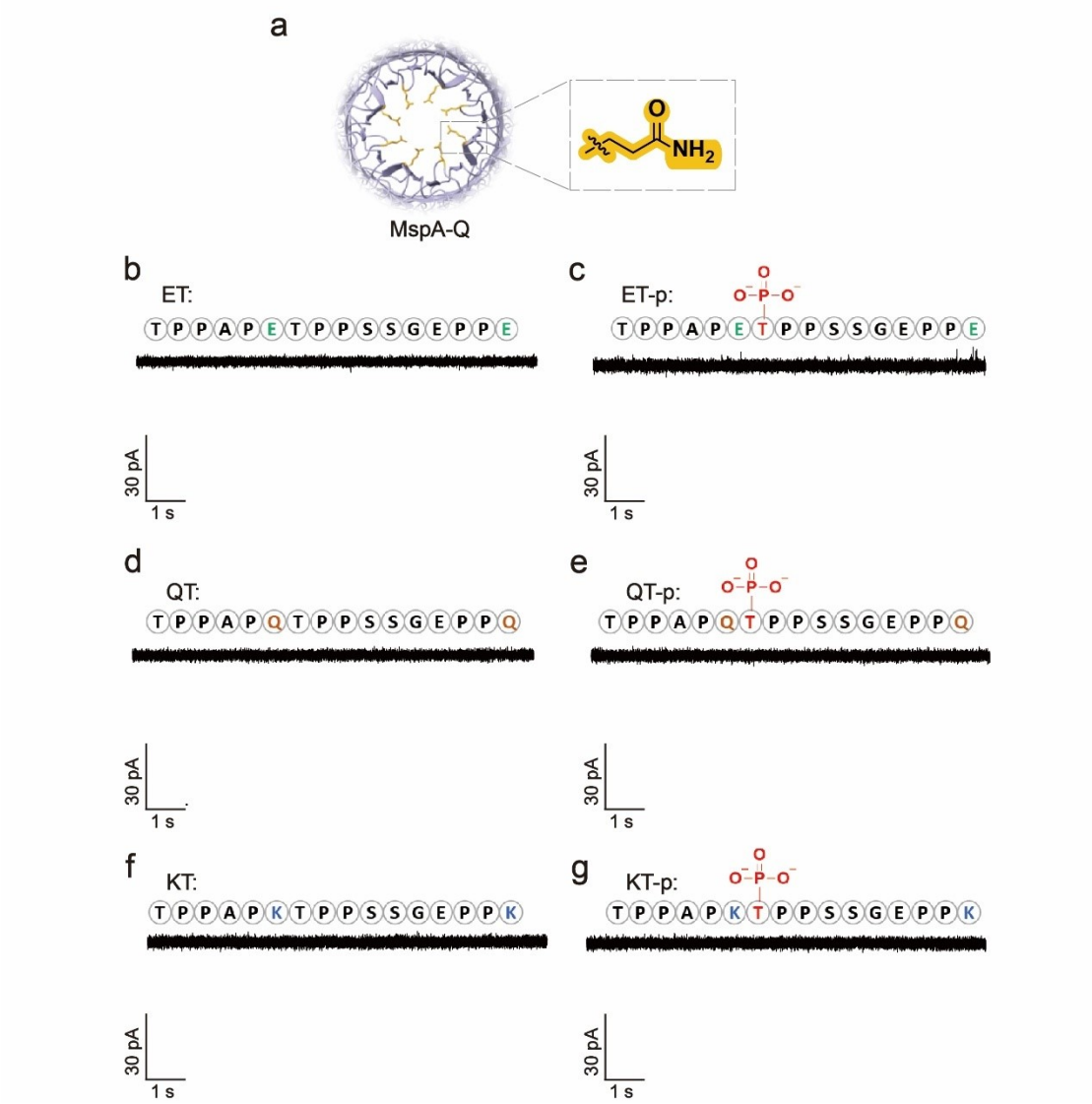

Figure S17. Schematic diagram of the acid residue at site 91 within the MspA-Q nanopore. Representative current traces of the MspA-Q nanopore following the addition of ET (b), ET-p (c), QT (d), QT-p (e), KT (f), or KT-p (g) as sole analytes. However, none of the peptides produced detectable events. Measurements were performed in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4, at an applied potential of +50 mV. Each peptide was tested separately at a final concentration of 5.0 μM.

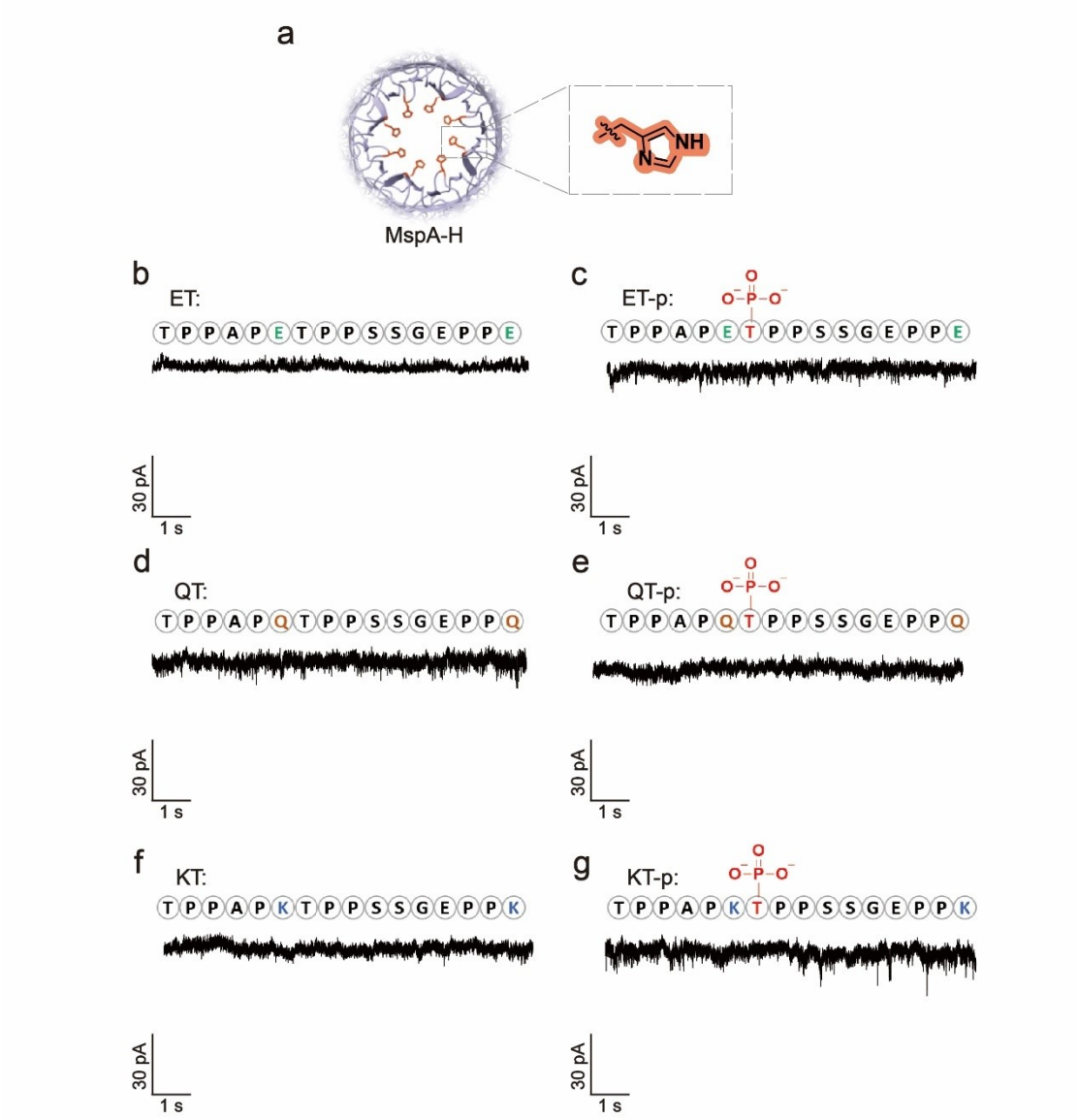

Figure S18. Schematic diagram of the acid residue at site 91 within the MspA-H nanopore. Representative current traces of the MspA-H nanopore following the addition of ET (b), ET-p (c), QT (d), QT-p (e), KT (f), or KT-p (g) as sole analytes. However, none of the peptides produced detectable events. Measurements were performed in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4 at an applied potential of +50 mV. Each peptide was tested separately at a final concentration of 5.0 μM.

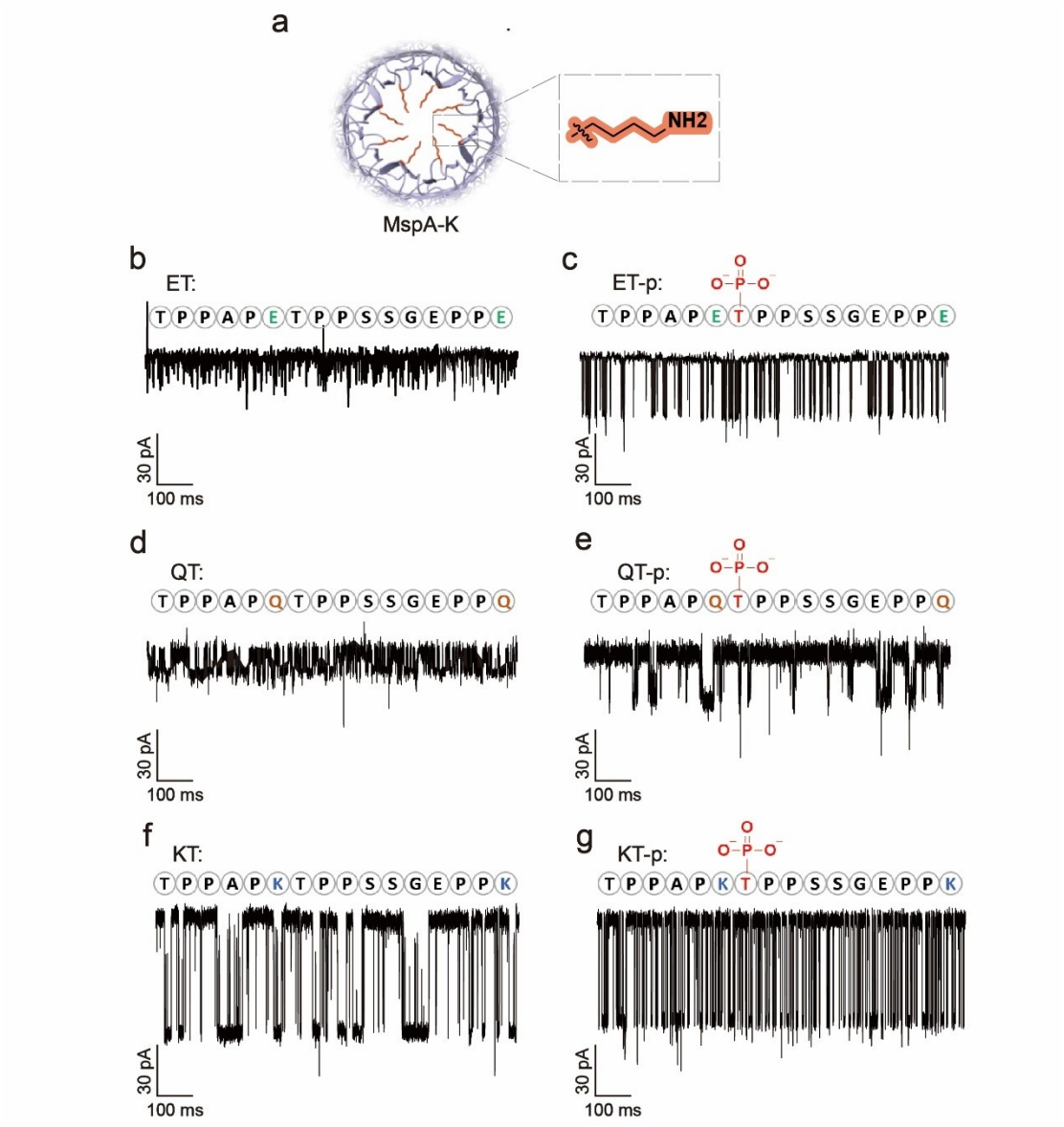

Figure S19. Schematic diagram of the acid residue at site 91 within the MspA-K nanopore. Representative current traces of the MspA-K nanopore following the addition of ET (b), ET-p (c), QT (d), QT-p (e), KT (f), or KT-p (g) as sole analytes. Measurements were performed in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4 at an applied potential of +50 mV. Each peptide was tested separately at a final concentration of 5.0 μM.

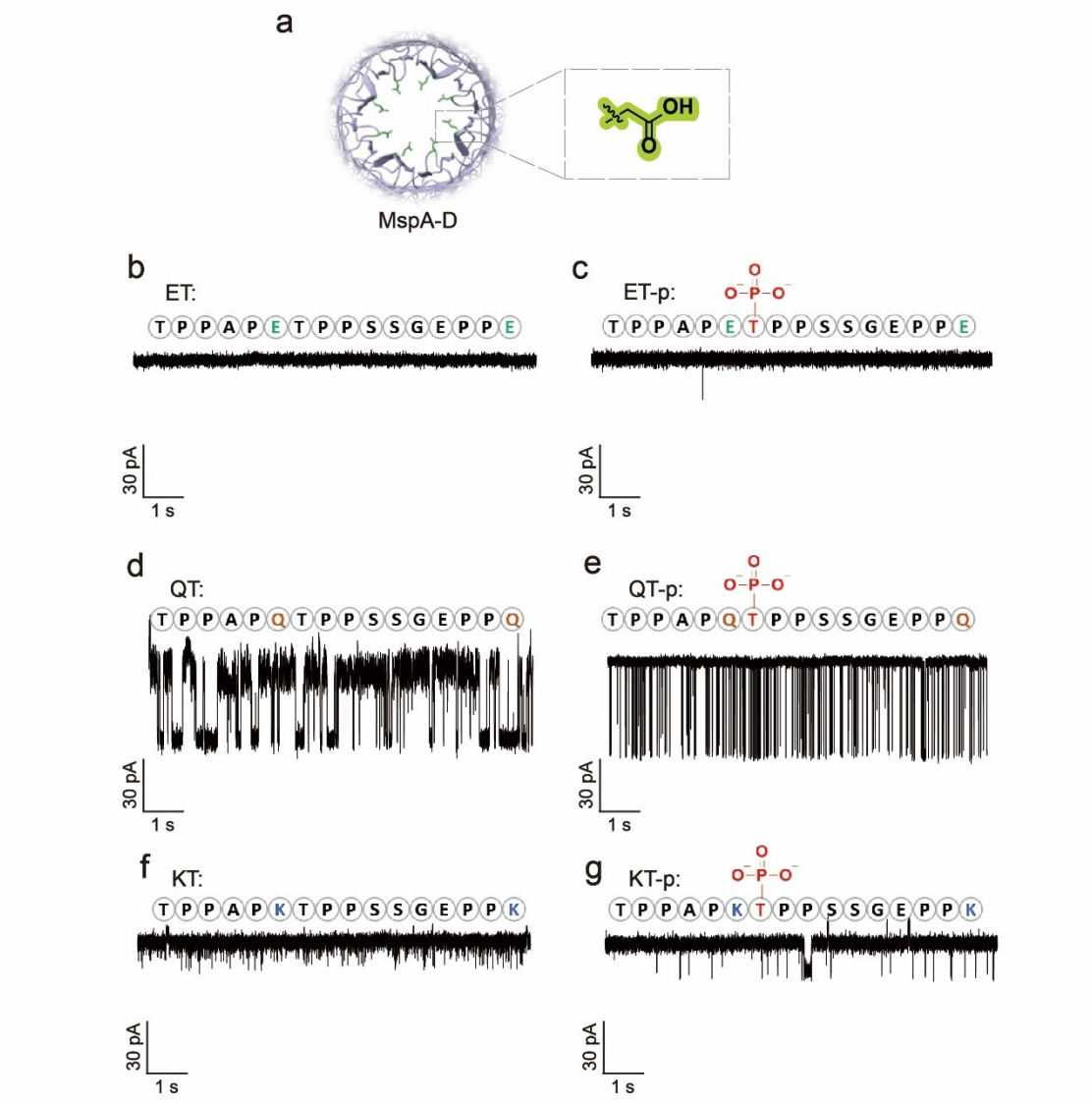

Figure S20. Schematic diagram of the acid residue at site 91 within the MspA-D nanopore. Representative current traces of the MspA-D nanopore following the addition of ET (b), ET-p (c), QT (d), QT-p (e), KT (f), or KT-p (g) as sole analytes. Measurements were performed in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4 at an applied potential of +50 mV. Each peptide was tested separately at a final concentration of 5.0 μM.

.

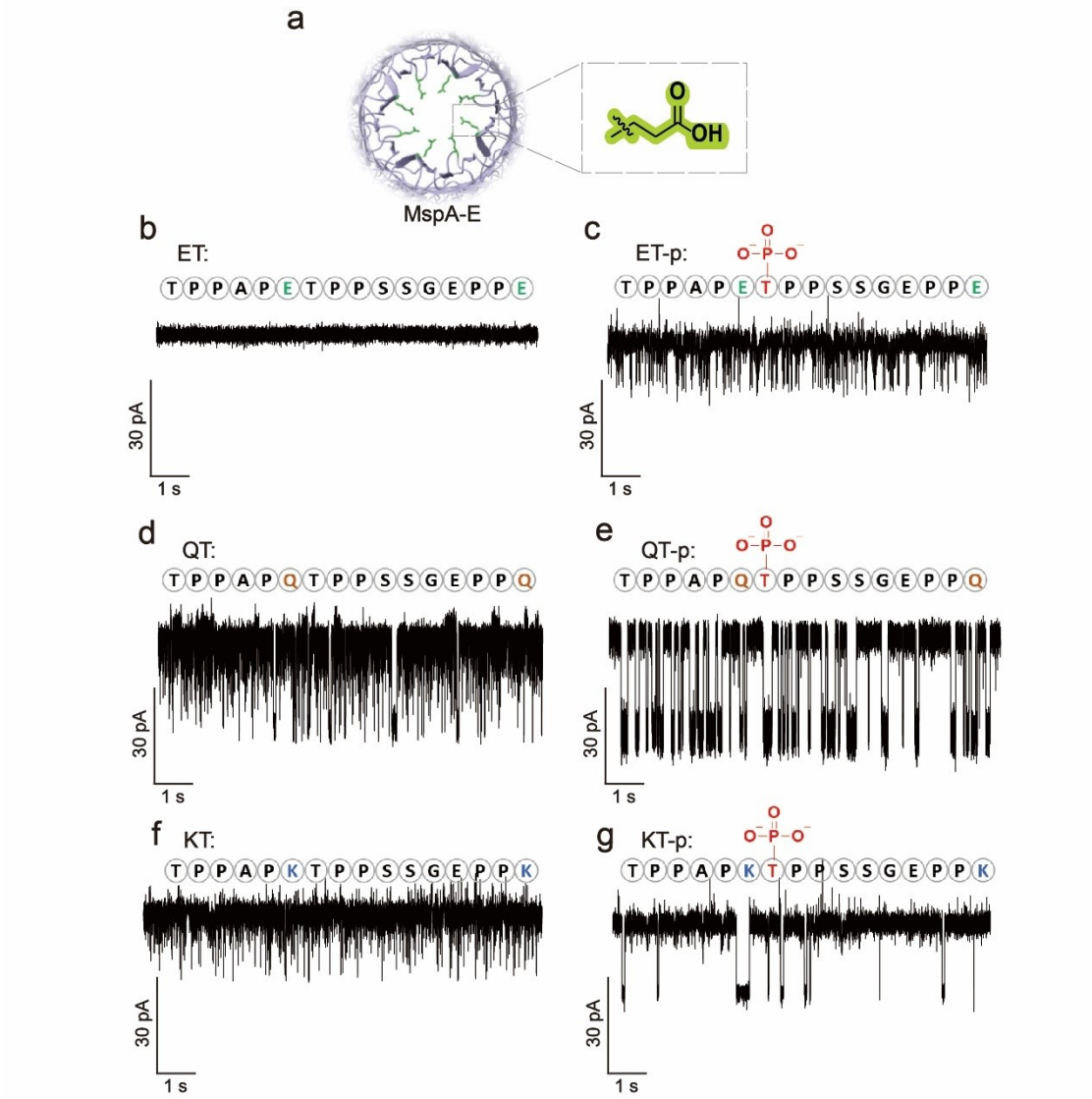

Figure S21. Schematic diagram of the acid residue at site 91 within the MspA-E nanopore. Representative current traces of the MspA-E nanopore following the addition of ET (b), ET-p (c), QT (d), QT-p (e), KT (f), or KT-p (g) as sole analytes. Measurements were performed in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4 at an applied potential of +50 mV. Each peptide was tested separately at a final concentration of 5.0 μM.

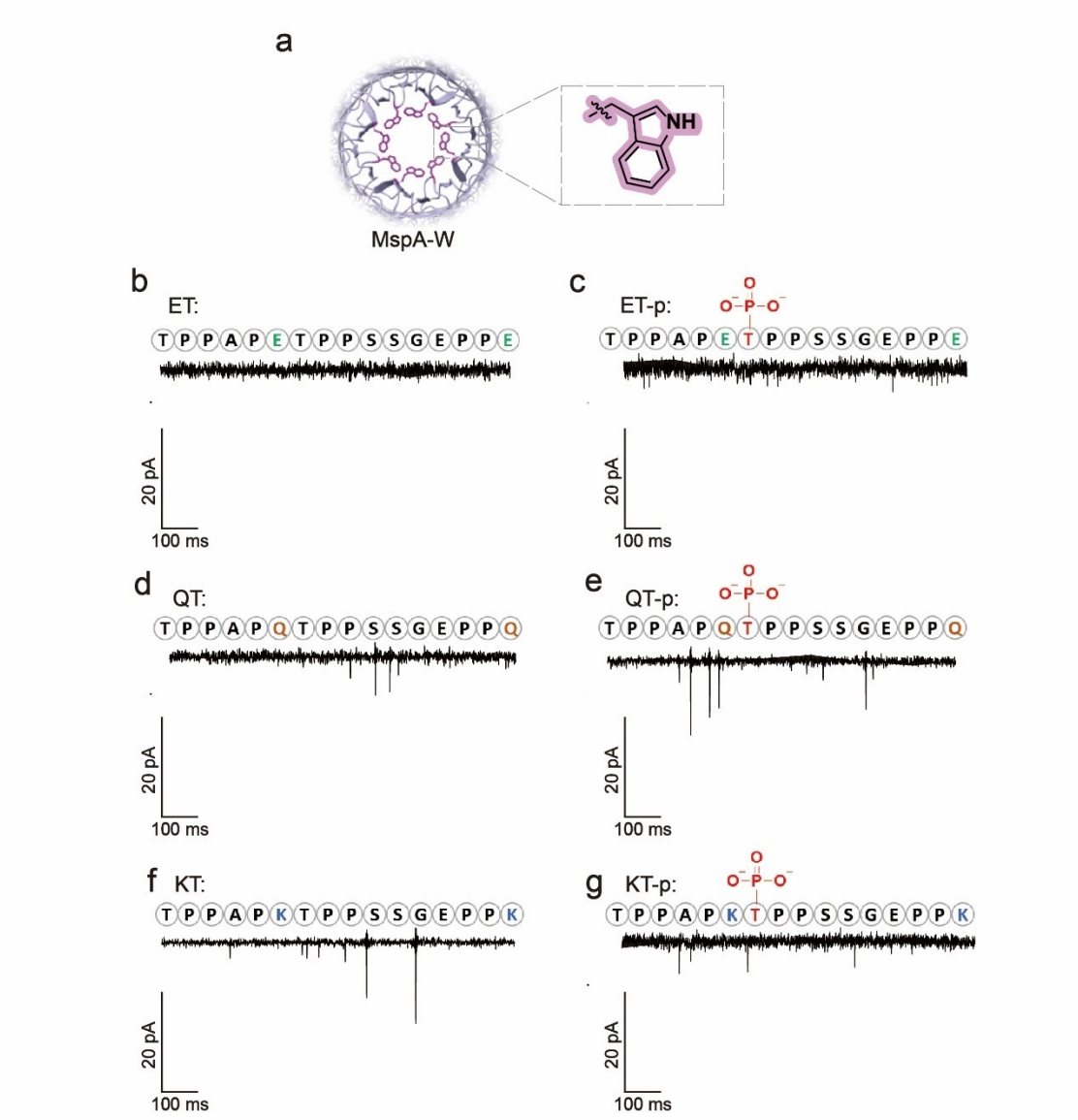

Figure S22. Schematic diagram of the acid residue at site 91 within the MspA-W nanopore. Representative current traces of the MspA-W nanopore following the addition of ET (b), ET-p (c), QT (d), QT-p (e), KT (f), or KT-p (g) as sole analytes. Measurements were performed in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4 at an applied potential of +50 mV. Each peptide was tested separately at a final concentration of 5.0 μM.

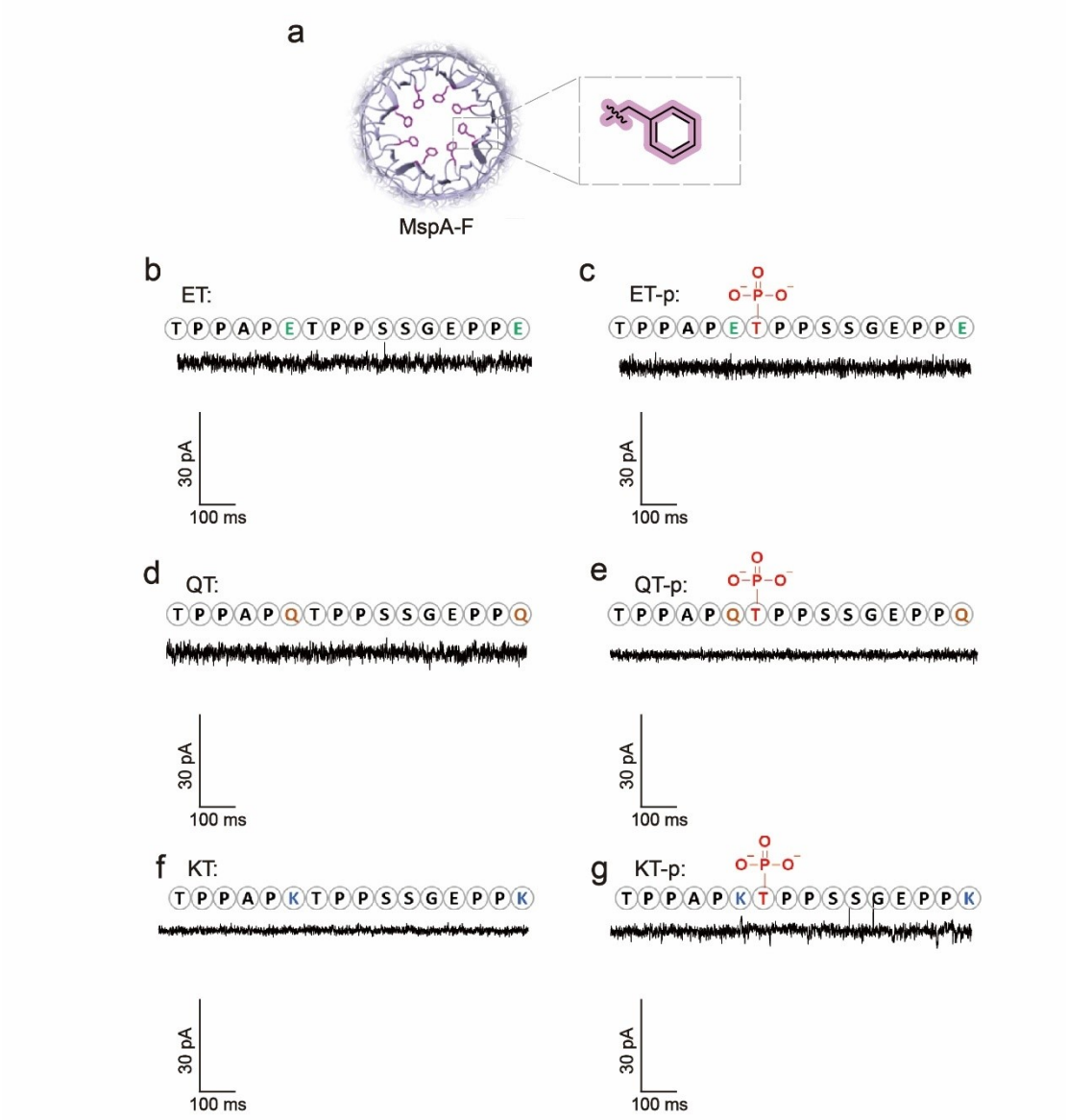

Figure S23. Schematic diagram of the acid residue at site 91 within the MspA-F nanopore. Representative current traces of the MspA-F nanopore upon addition of ET (b), ET-p (c), QT (d), QT-p (e), KT (f), or KT-p (g) as sole analytes. Measurements were performed in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4 at an applied potential of +50 mV. Each peptide was tested separately at a final concentration of 5.0 μM.

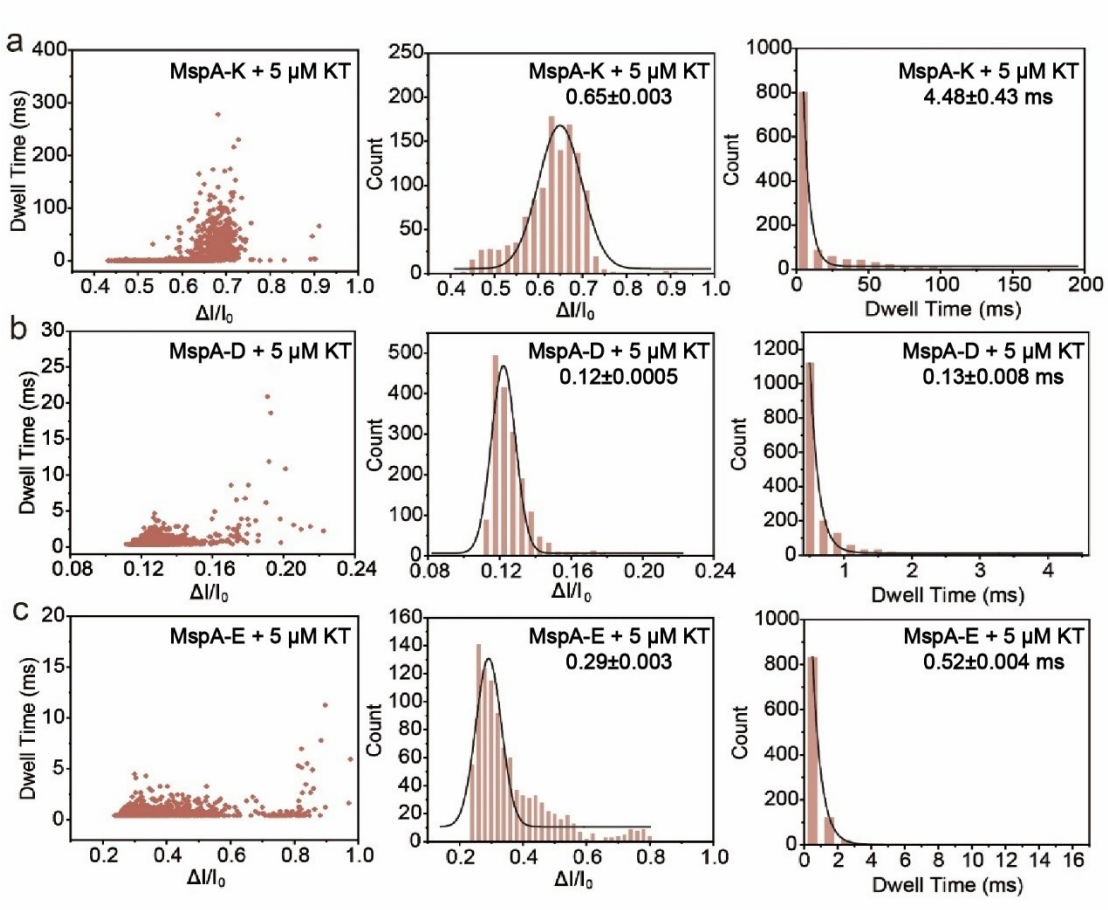

Figure S24. Scatter plot of $\Delta I/I_0$ (left), histogram of $\Delta I/I_0$ (middle) and dwell time histogram (right) for KT interactions with MspA nanopores (a: MspA-K; b: MspA-D; c: MspA-E) at an applied potential of +50 mV.

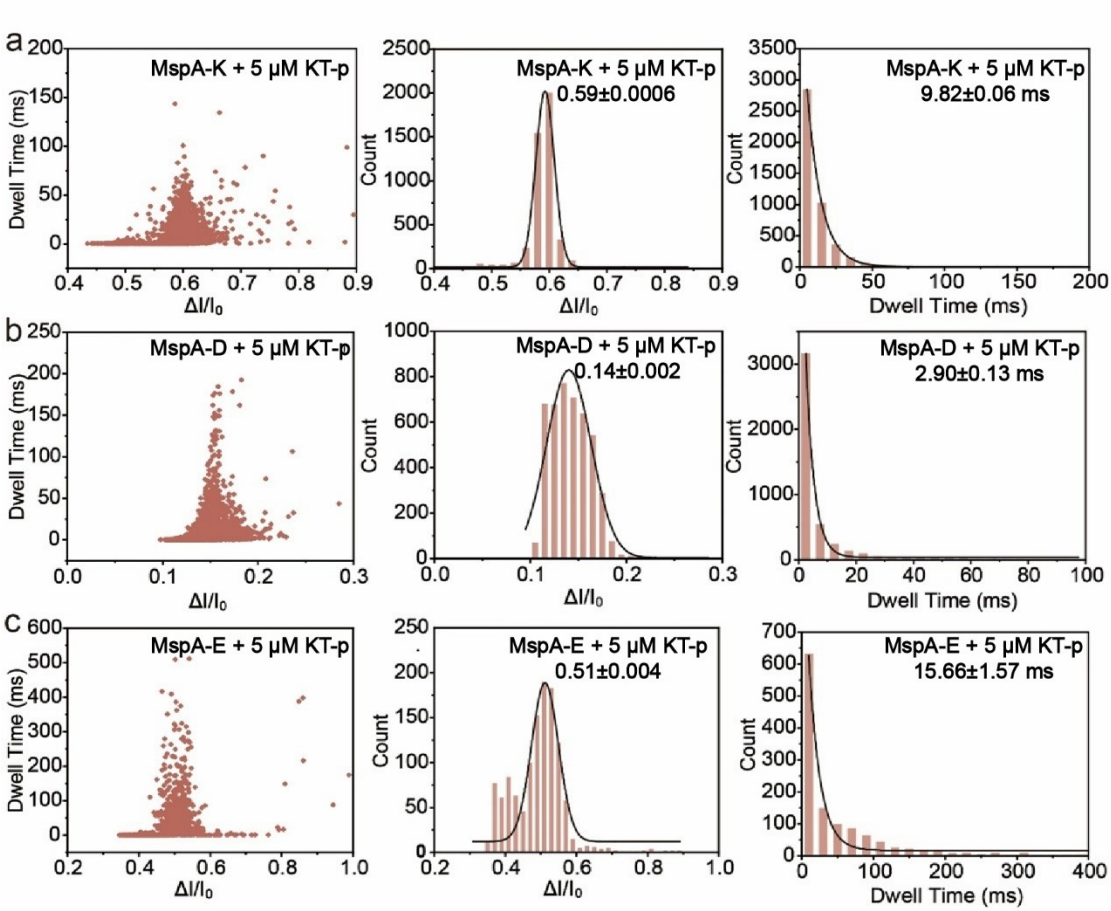

Figure S25. Scatter plot of $\Delta I/I_0$ (left), histogram of $\Delta I/I_0$ (middle) and dwell time histogram (right) for KT-p interactions with MspA nanopores (a: MspA-K; b: MspA-D; c: MspA-E) at an applied potential of +50 mV.

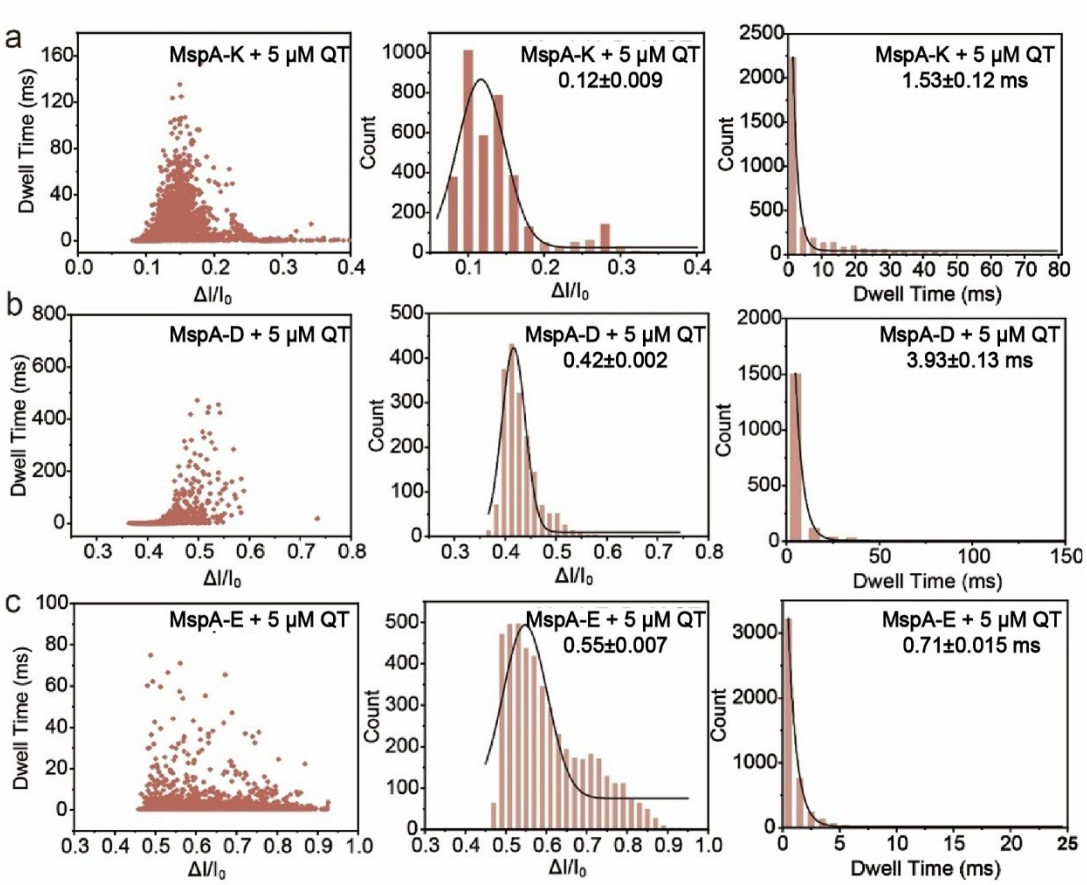

Figure S26. Scatter plot of $\Delta I/I_0$ (left), histogram of $\Delta I/I_0$ (middle) and dwell time histogram (right) for QT interactions with MspA nanopores (a: MspA-K; b: MspA-D; c: MspA-E) at an applied potential of +50 mV.

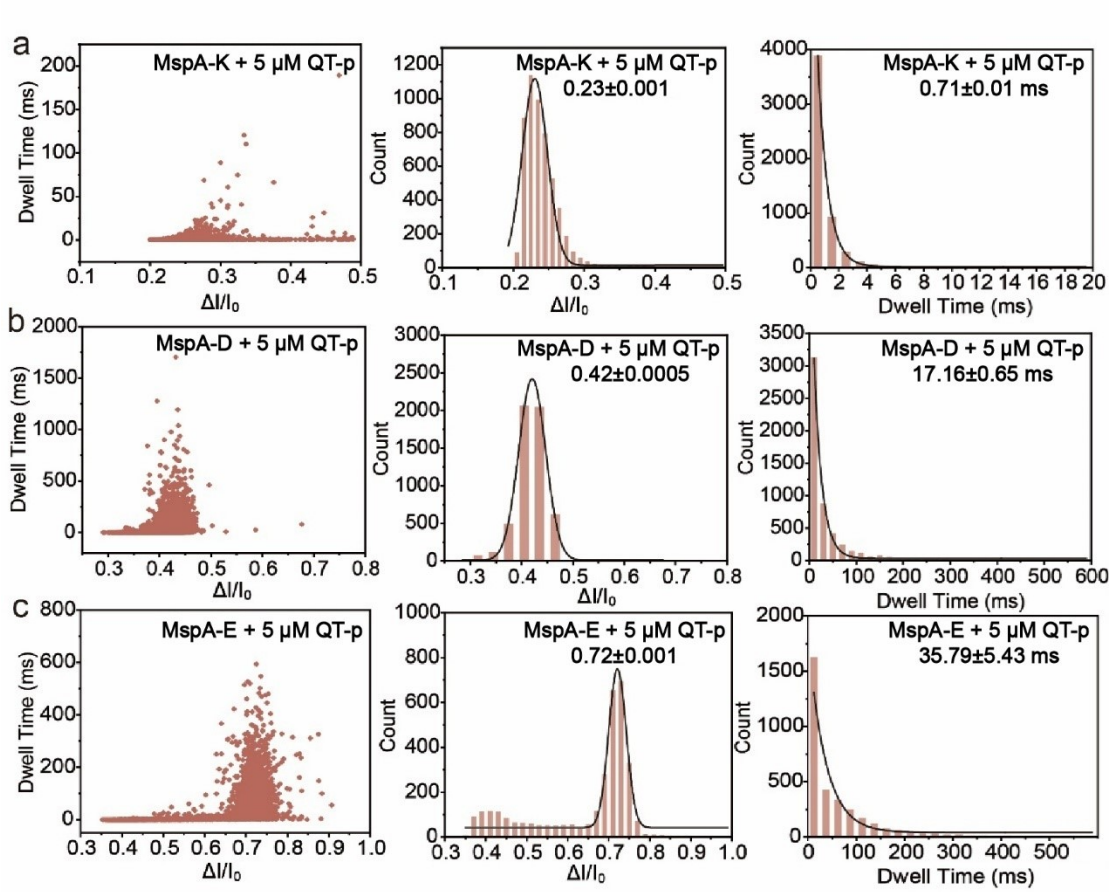

Figure S27. Scatter plot of $\Delta I/I_0$ (left), histogram of $\Delta I/I_0$ (middle) and dwell time histogram (right) for QT-p interactions with MspA nanopores (a: MspA-K; b: MspA-D; c: MspA-E) at an applied potential of +50 mV.

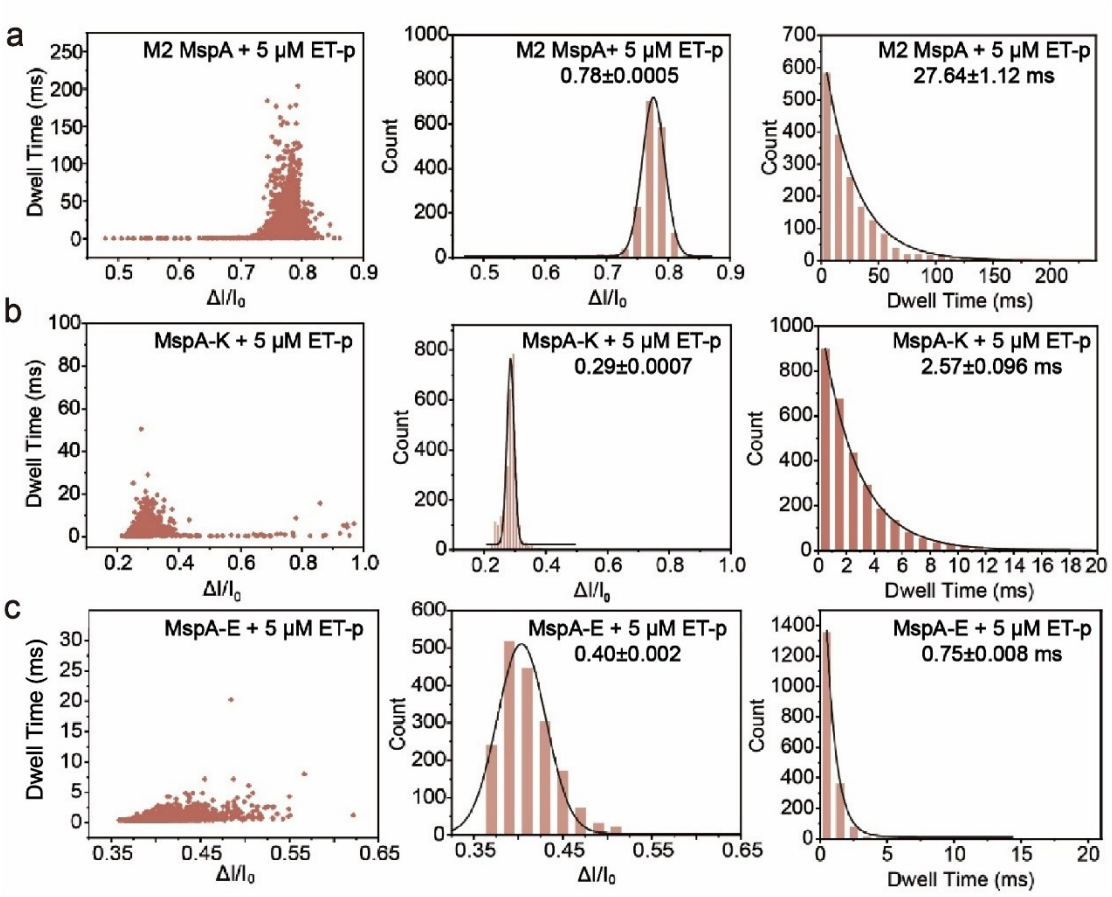

Figure S28. Scatter plot of $\Delta I/I_0$ (left), histogram of $\Delta I/I_0$ (middle) and dwell time histogram (right) for ET-p interactions with MspA nanopores (a: M2 MspA; b: MspA-K; c: MspA-E) at an applied potential of +50 mV.

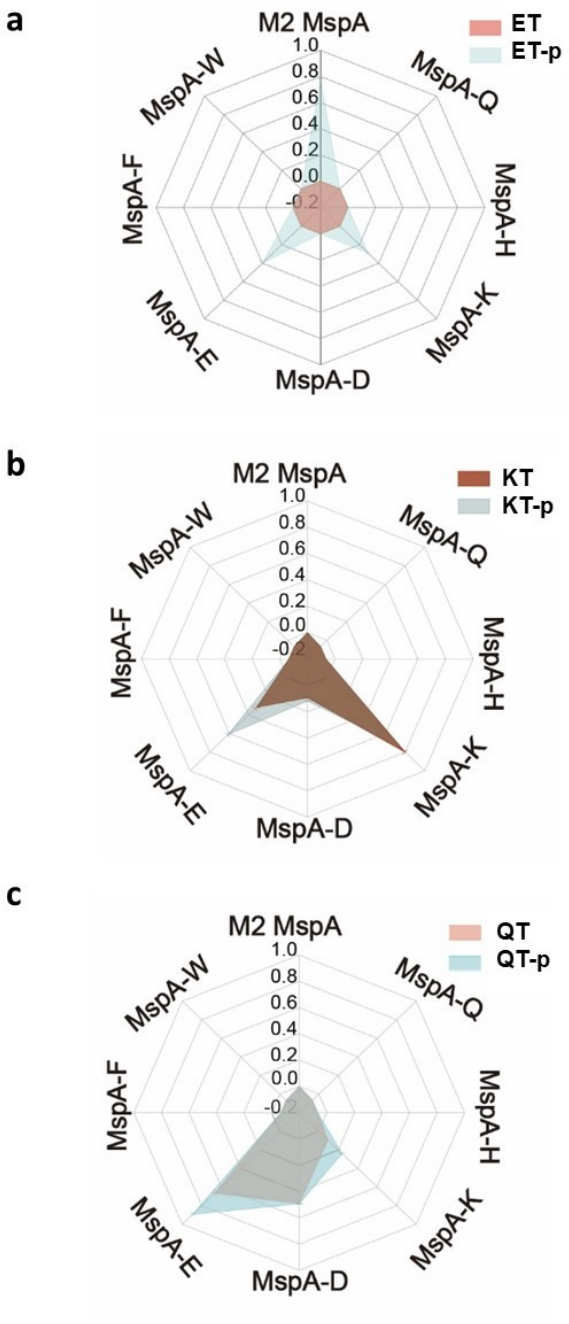

Figure S29. The radar diagrams of $\Delta I/I_0$ for phosphorylated peptides.

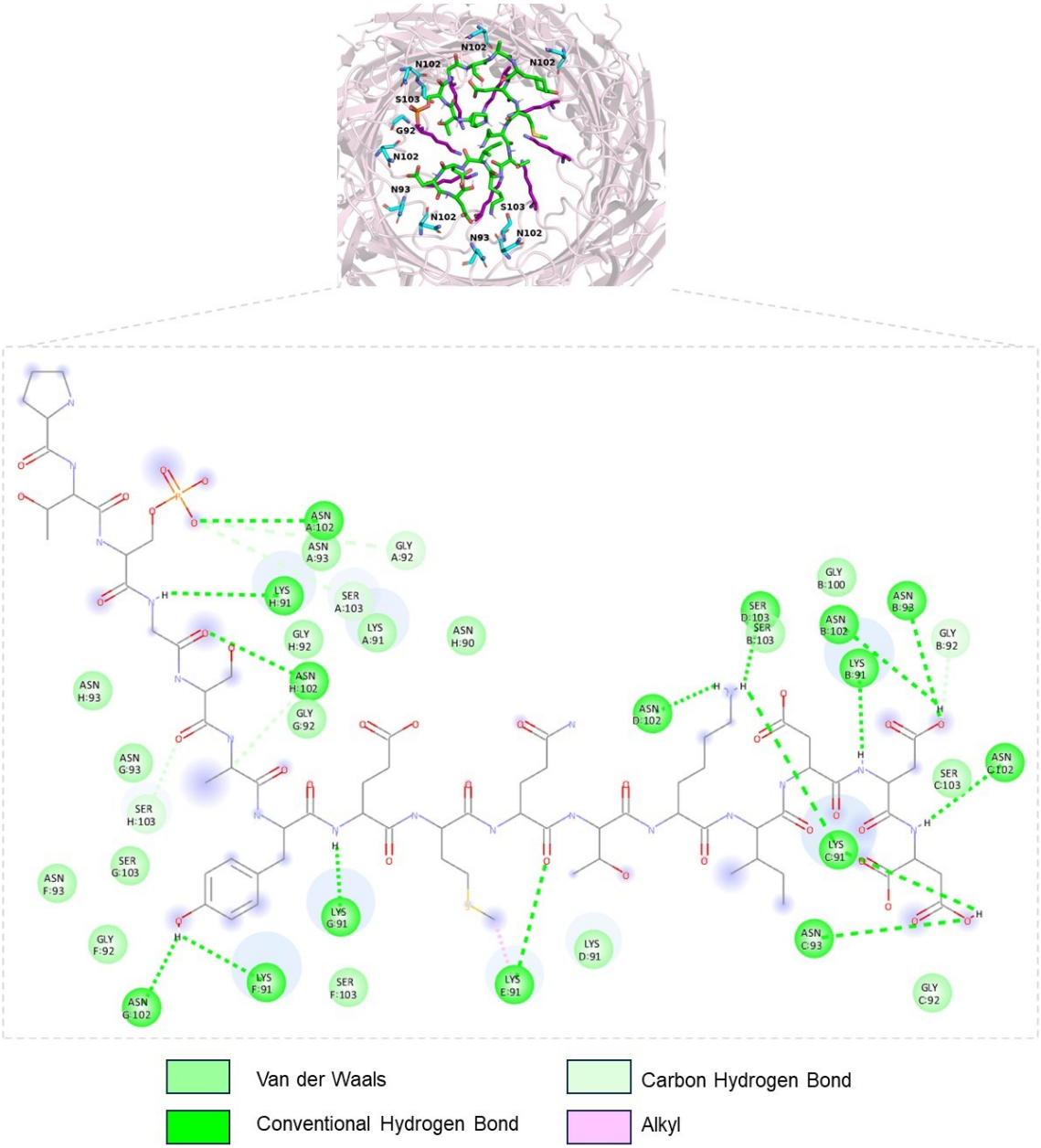

Figure S30. Characterization of PD-p peptide interactions with the microenvironment in MspA-K nanopores, including molecular docking analysis.

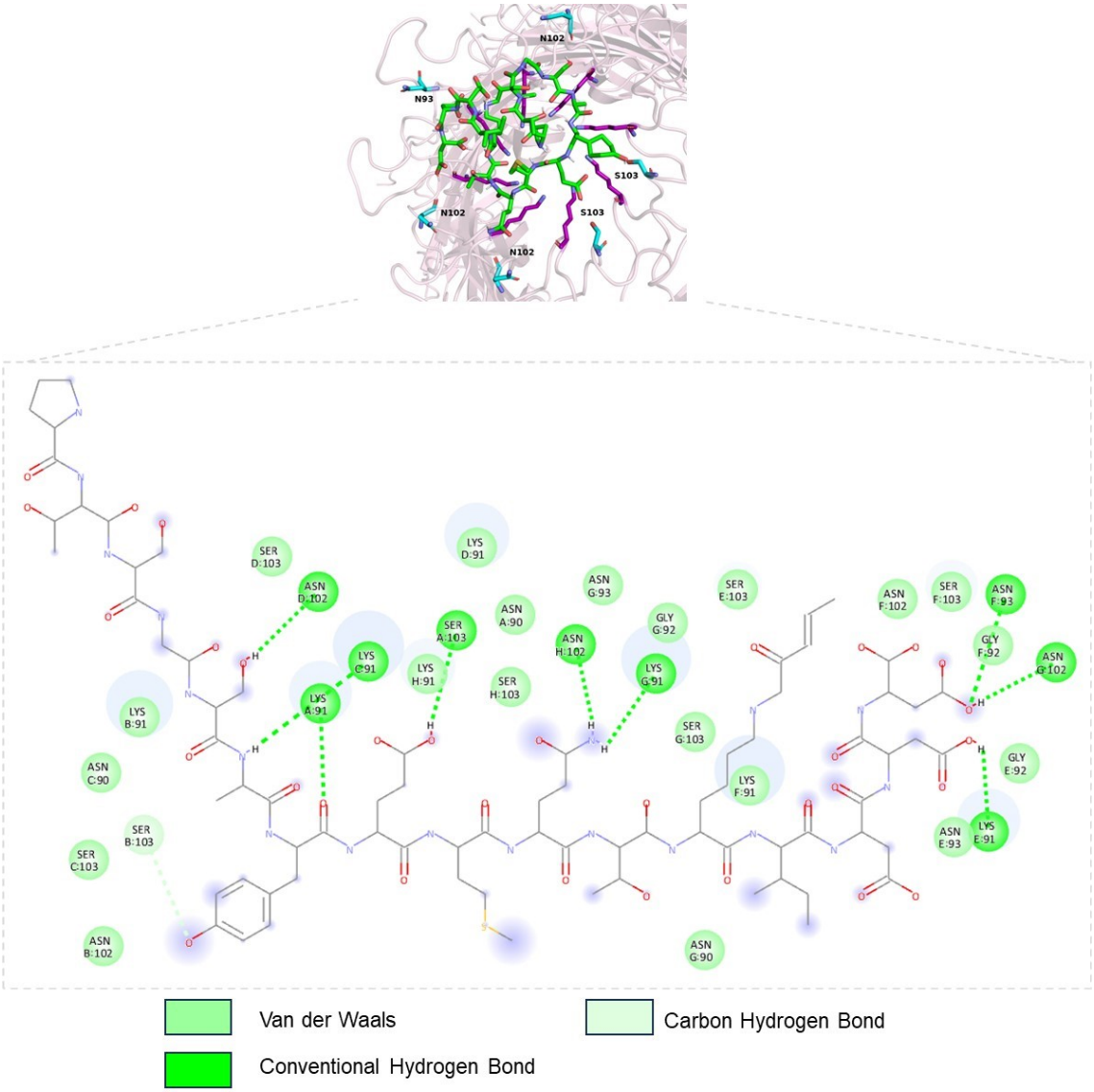

Figure S31. Characterization of PD-cr peptide interactions with the microenvironment in MspA-K nanopores, including molecular docking analysis.

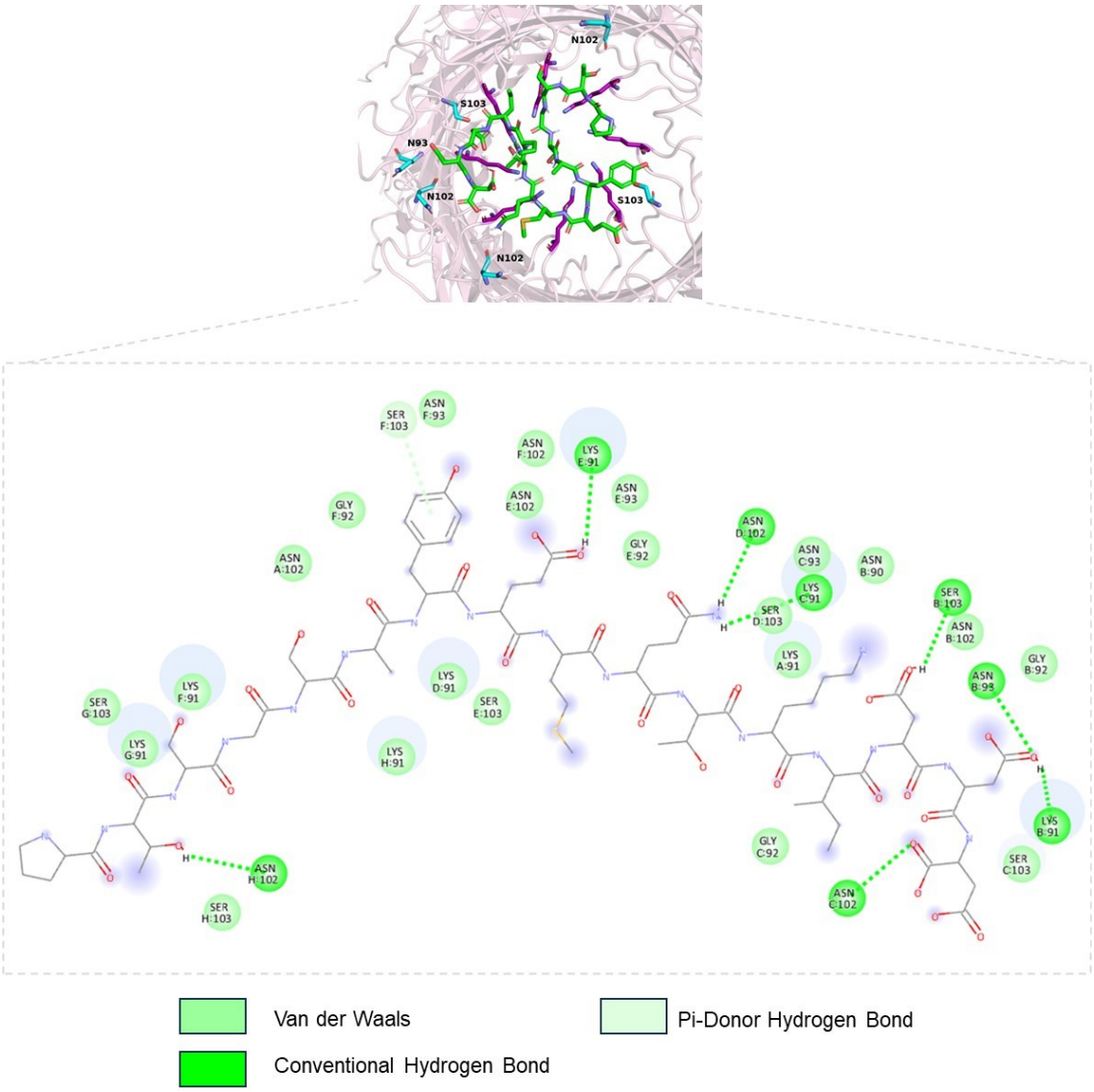

Figure S32. Characterization of PD peptide interactions with the microenvironment in MspA-K nanopores, including molecular docking analysis.

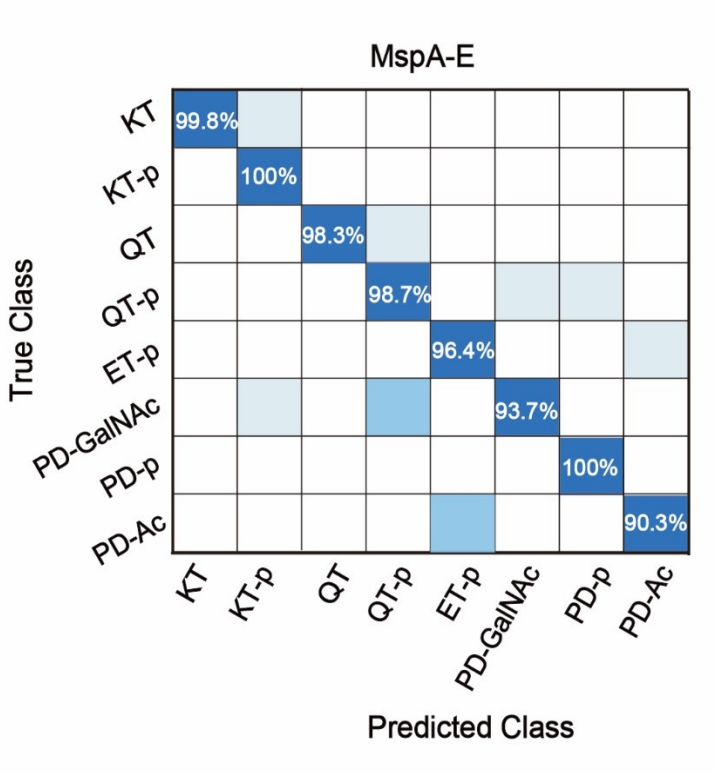

Figure S33. PTMs identification by machine learning. The confusion matrix of 8 types of peptides classification in MspA-E generated using the bagged trees model. The validation accuracy reaches a score of 0.979, and the test accuracy is 0.982.

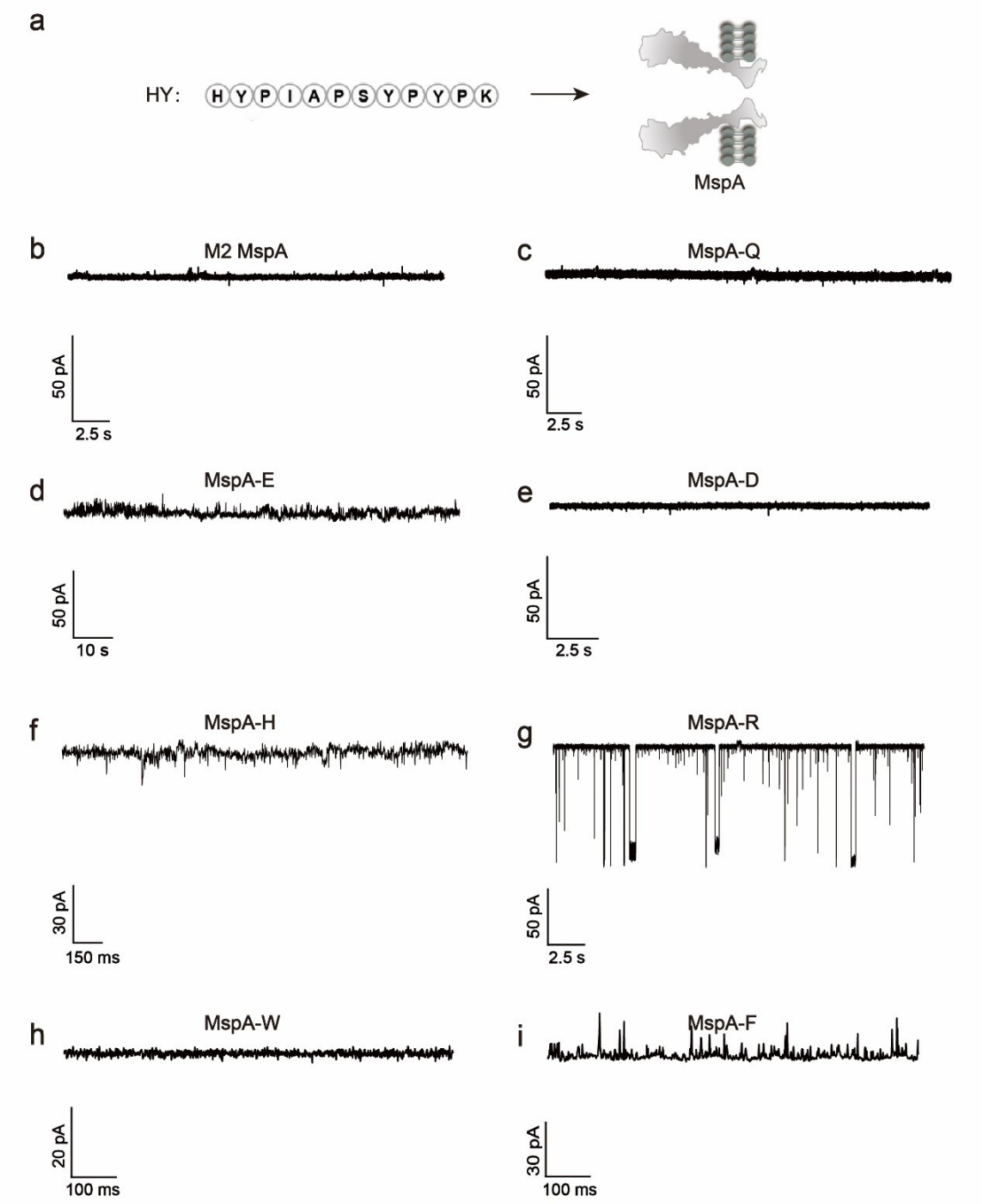

Figure S34. (a) Schematic diagram of the detection of HY with MspA nanopores. **(**b-i**)** Representative current traces of the MspA mutant nanopores following the addition of HY: M2 MspA (b); MspA-Q (c); MspA-E (d); MspA-D (e); MspA-H (f); MspA-R (g); MspA-W (h); MspA-F (i). Measurements were performed in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4 at an applied potential of +50 mV. Each peptide was tested separately at a final concentration of 5.0 μM.

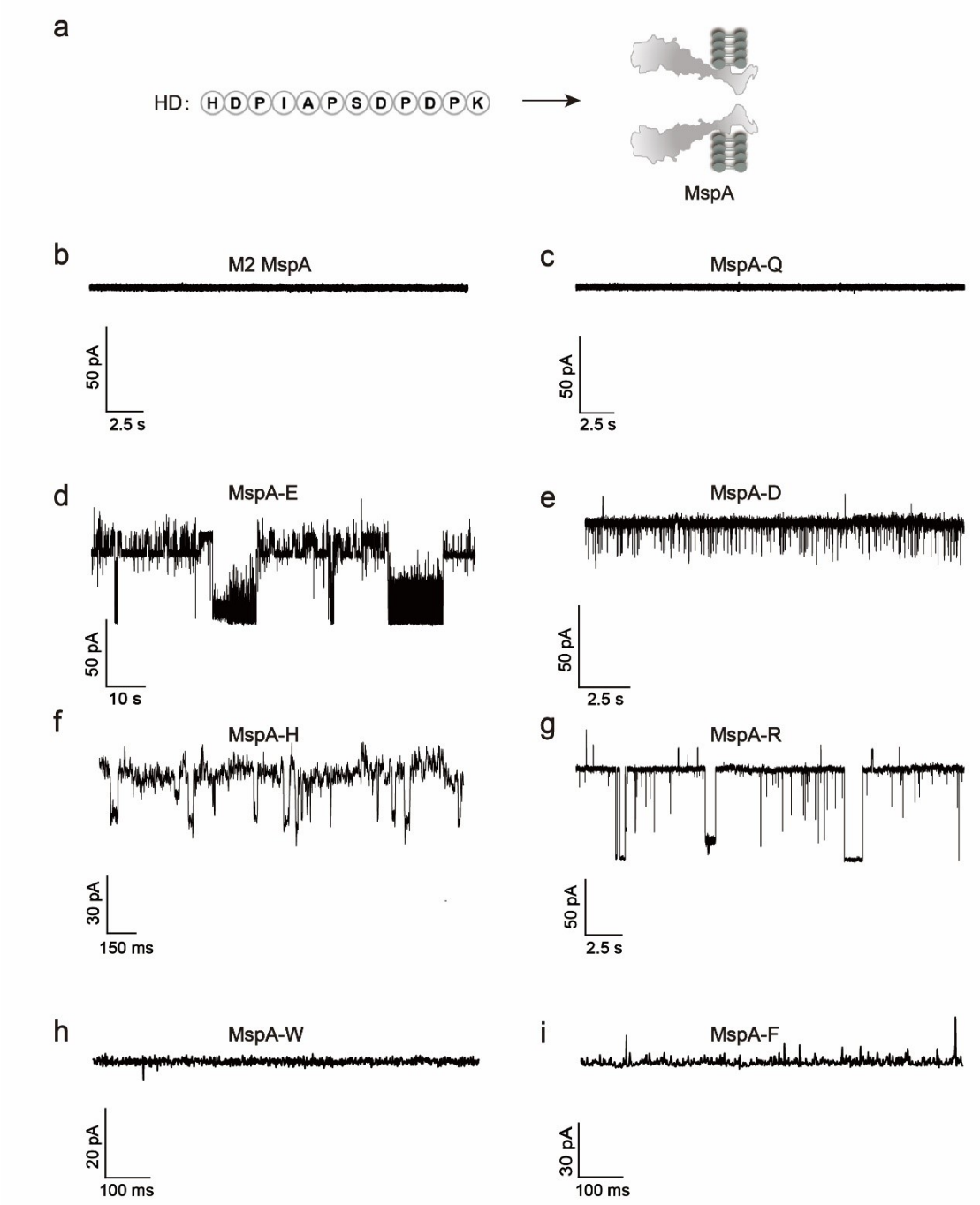

Figure S35. (a) Schematic diagram of the MspA nanopores detection for HD. **(**b-i**)** Representative current traces of the MspA mutant nanopores following the addition of HD: M2 MspA (b); MspA-Q (c); MspA-E (d); MspA-D (e); MspA-H (f); MspA-R (g); MspA-W (h); MspA-F (i). Measurements were performed in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4 at an applied potential of +50 mV. Each peptide was tested separately at a final concentration of 5.0 μM.

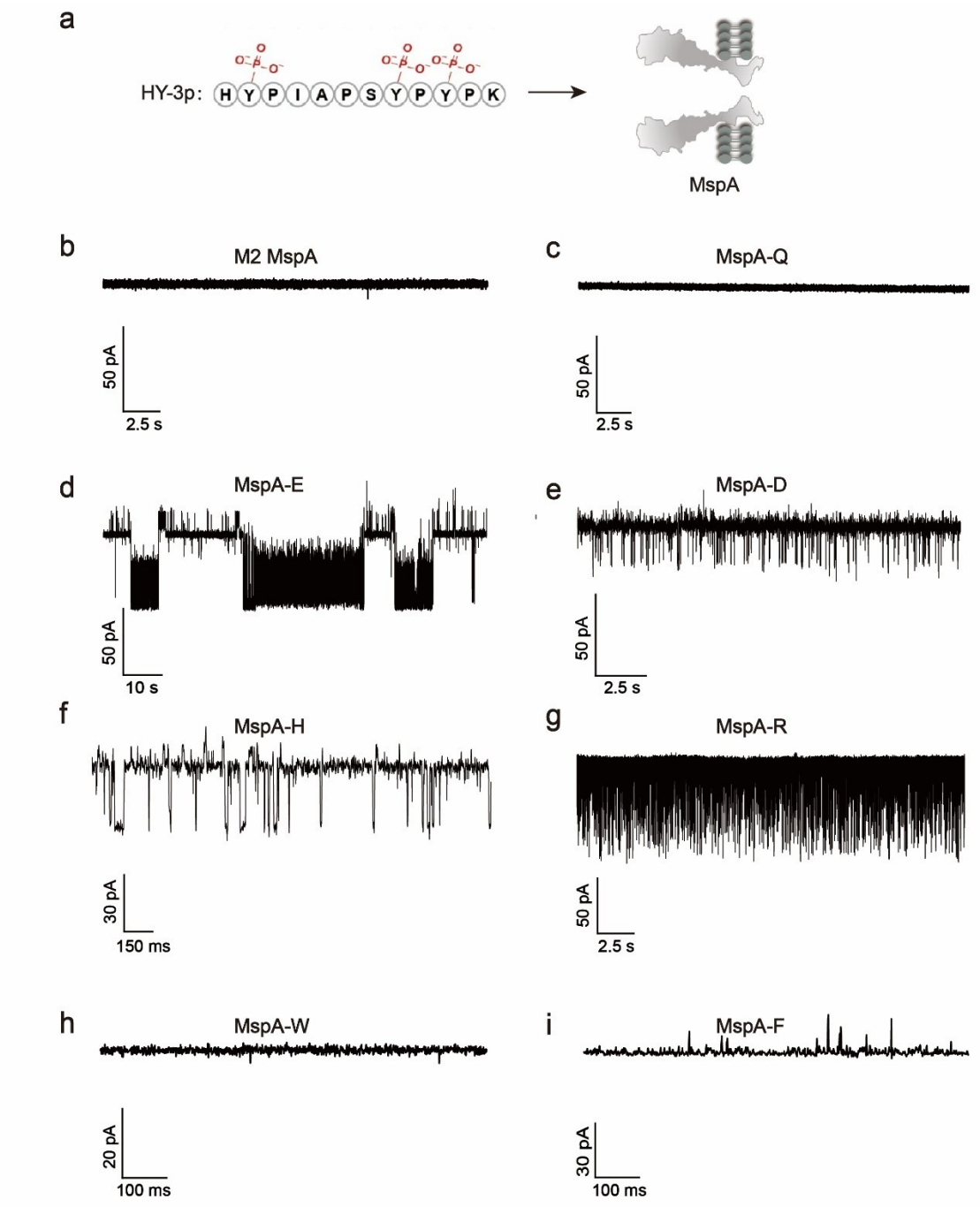

Figure S36. (a) Schematic diagram of the MspA nanopores detection for HY-p. **(**b-i**)** Representative current traces of the MspA mutant nanopores following the addition of HY: M2 MspA (b); MspA-Q (c); MspA-E (d); MspA-D (e); MspA-H (f); MspA-R (g); MspA-W (h); MspA-F (i). Measurements were performed in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4 at an applied potential of +50 mV. Each peptide was tested separately at a final concentration of 5.0 μM.

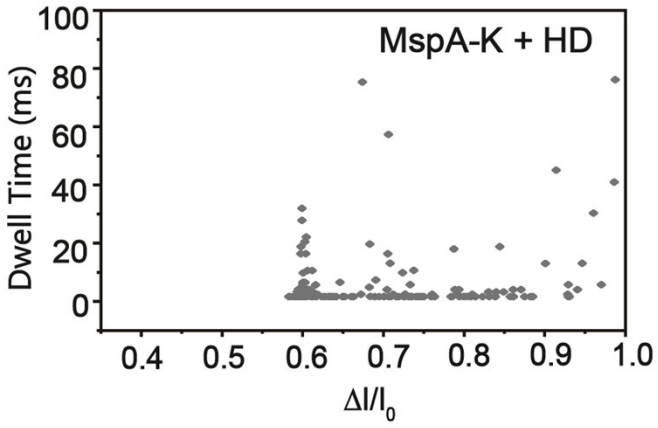

Figure S37. Scatter-plot of $\Delta I/I_0$ for HD using MspA-K. Measurements were performed in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4 at an applied potential of +50 mV.

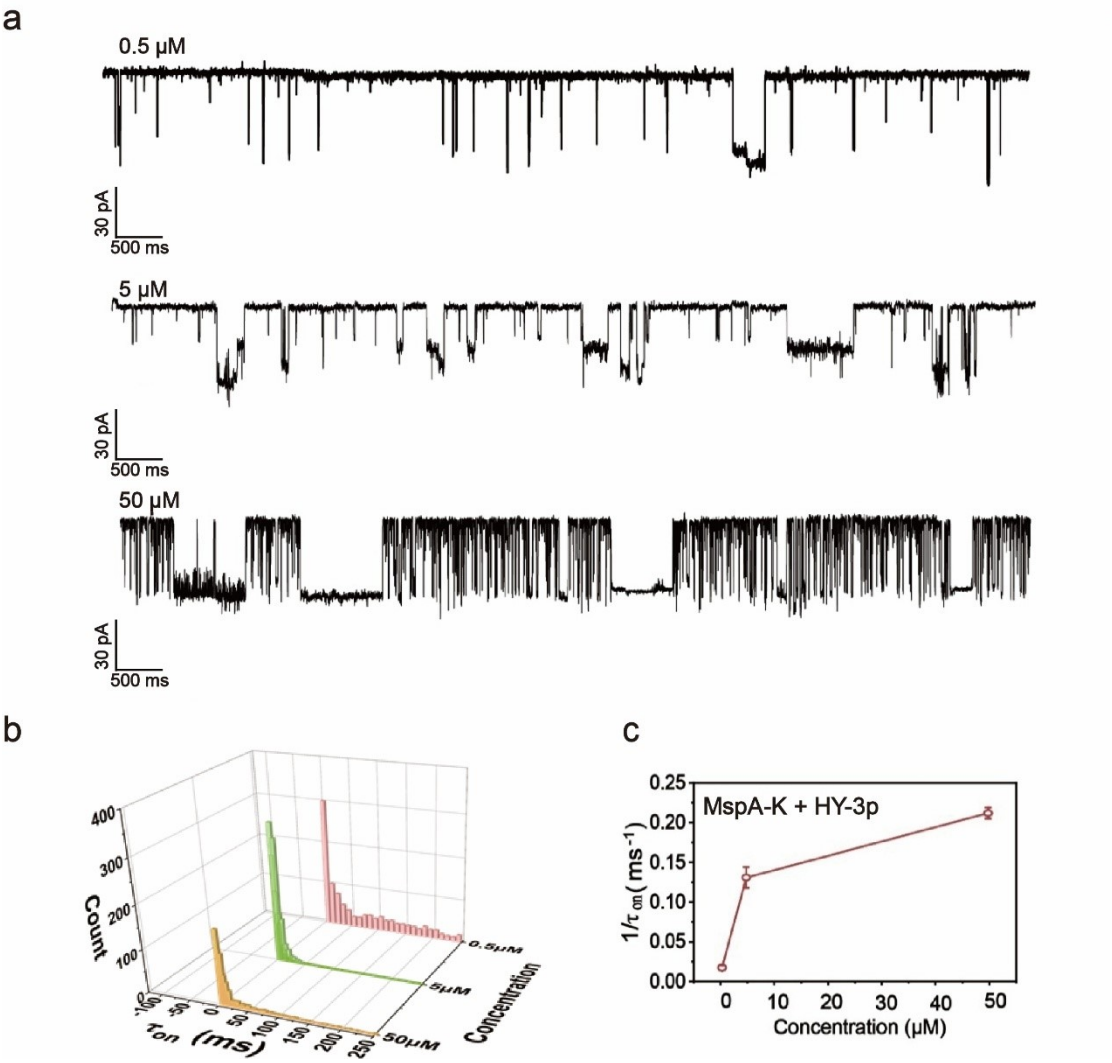

Figure S38. (a) Representative current traces acquired with varying HY-3p concentrations. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4 at an applied potential of +50 mV. A potential of +50 mV was constantly applied. (b) HY-3p was added with a final concentration of 0.5-50 μM. (c) The histogram of τ_{on} and plot of $1/\tau_{on}$ versus HY-3p concentration. Whereas $1/\tau_{on}$ increases when the HY-3p concentration is increased.

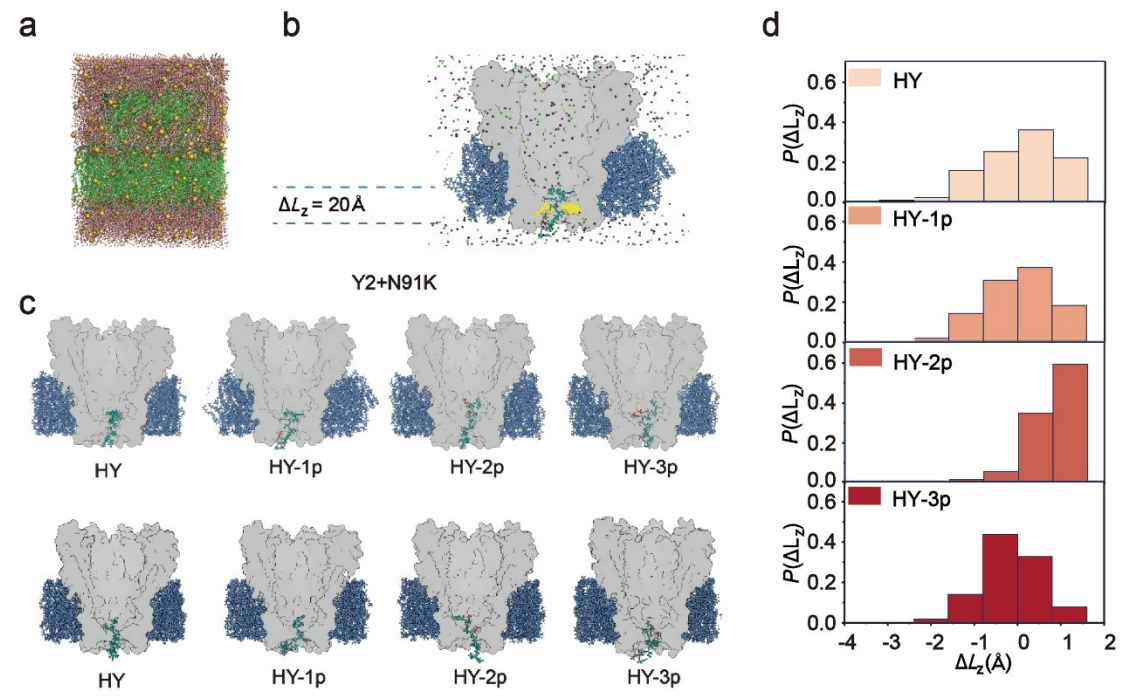

Figure S39. (a) Snapshot for the simulation box. (b) The current was calculated within an ΔL_z=20 Å thickness slab centered at the 91 site of MspA-K nanopore. (c) During the simulation, Y2 and Y8@Y10 of peptides were respectively positioned in close proximity to N91K. (d) Probability distribution of peptides displacement under electric field driving during simulated observation time (Y8@Y10 + N91K).

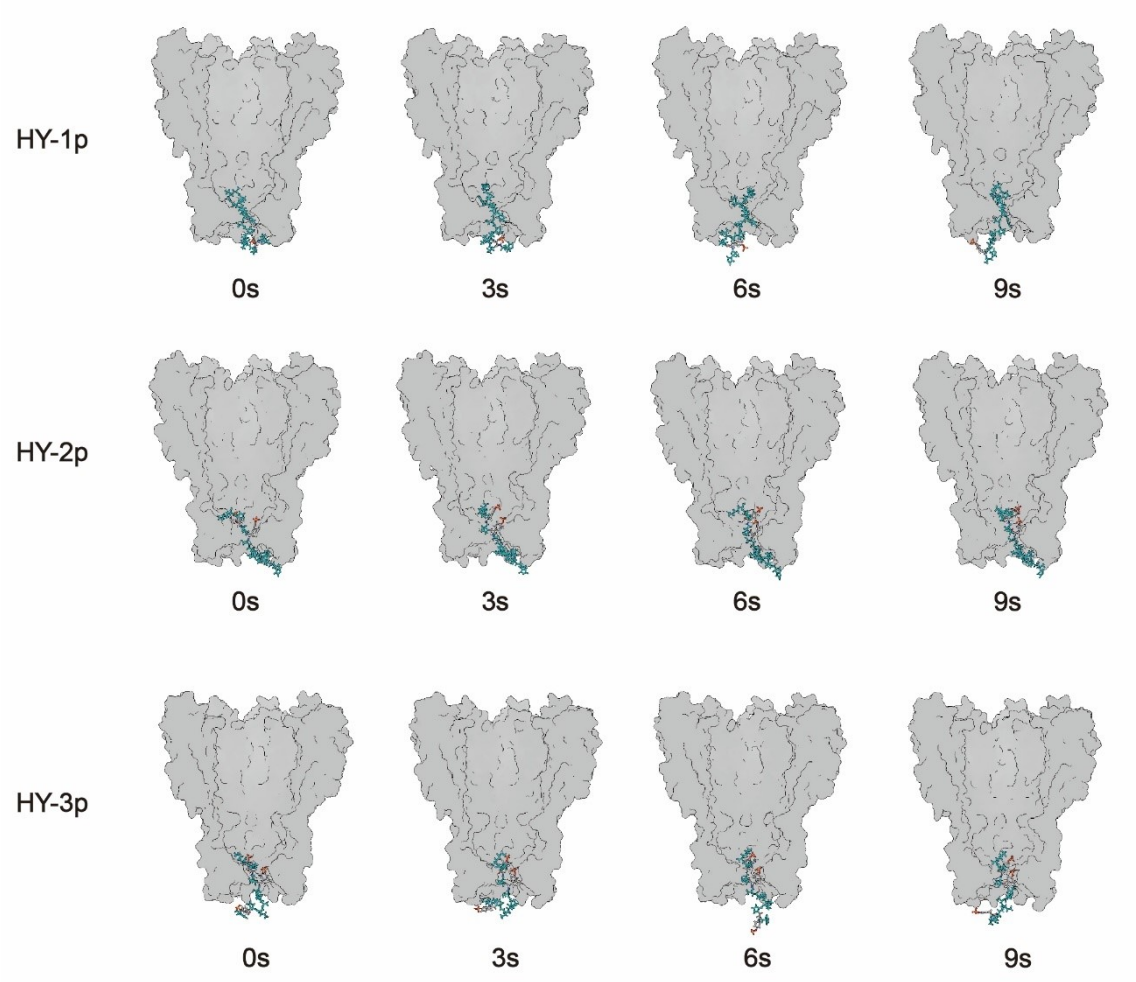

Figure S40. Time-based snapshots for the simulation scenarios studied.

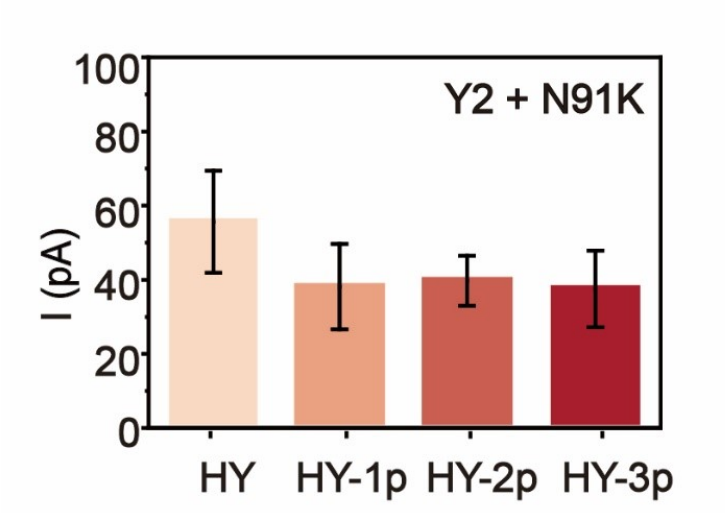

Figure S41. Simulated current blockade levels (Y2 + N91K).

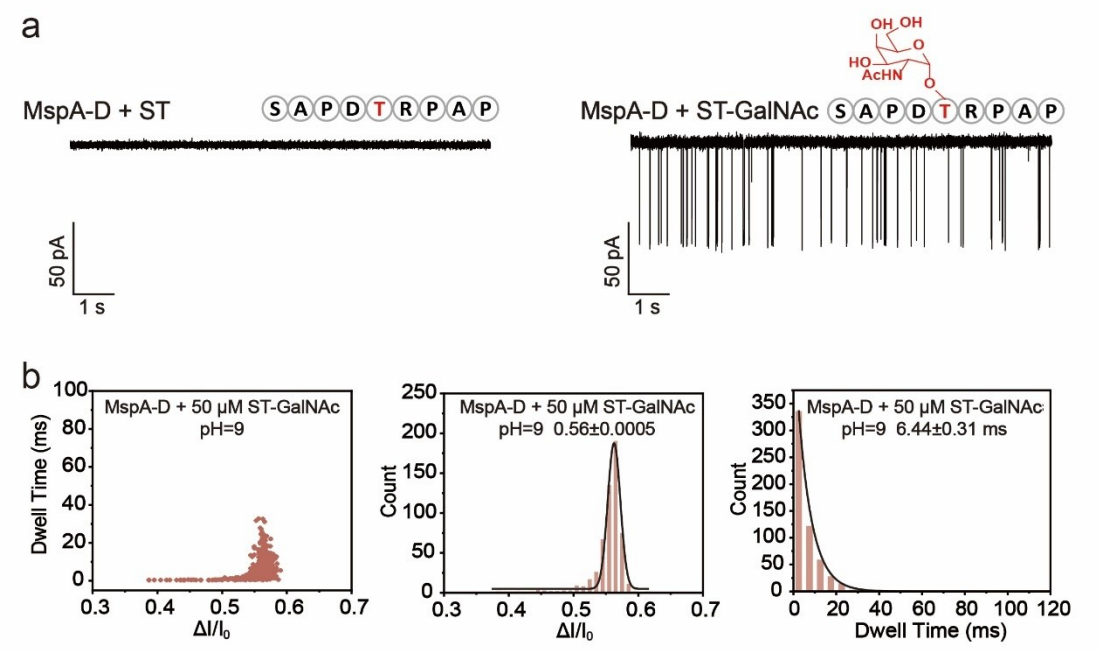

Figure S42. (a) Representative current traces of ST and ST-GalNAc with MspA-D. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 9.0. All peptides were added with a final concentration of 50.0 μM for each analyte. (b) Scatter plot of $\Delta I/I_0$, histogram of $\Delta I/I_0$ and dwell time histogram for ST-GalNAc interactions with MspA-D at an applied potential of +50 mV.

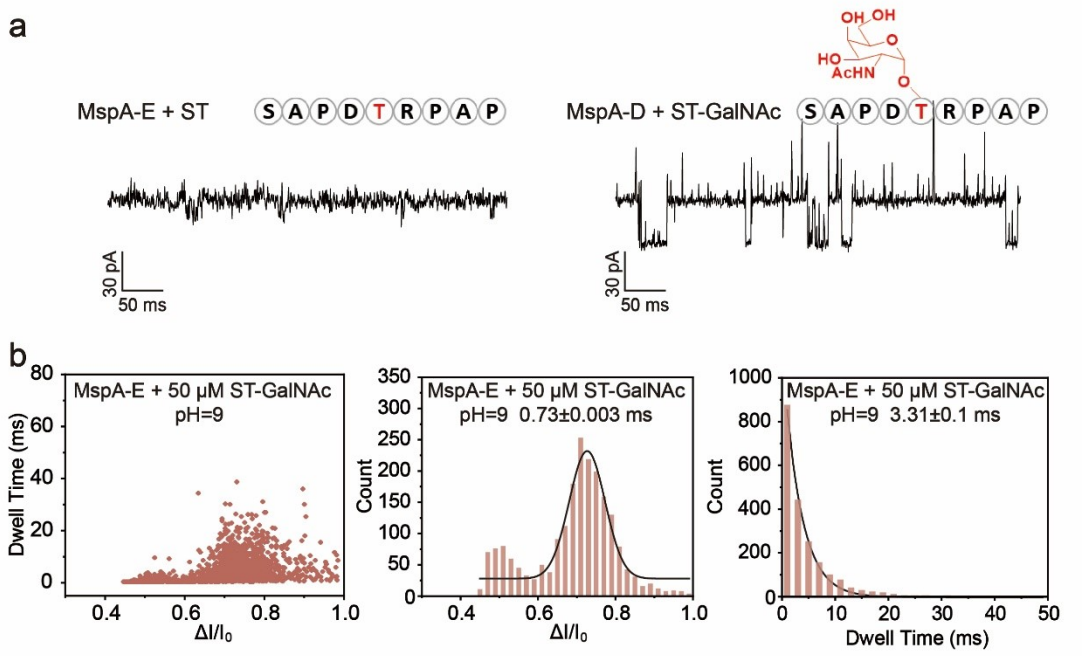

Figure S43. (a) Representative current traces of ST and ST-GalNAc with MspA-E. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 9.0. All peptides were added to a final concentration of 50.0 μM. (b) Scatter plot of $\Delta I/I_0$, histogram of $\Delta I/I_0$ and dwell time histogram for ST-GalNAc interactions with MspA-E at an applied potential of +50 mV.

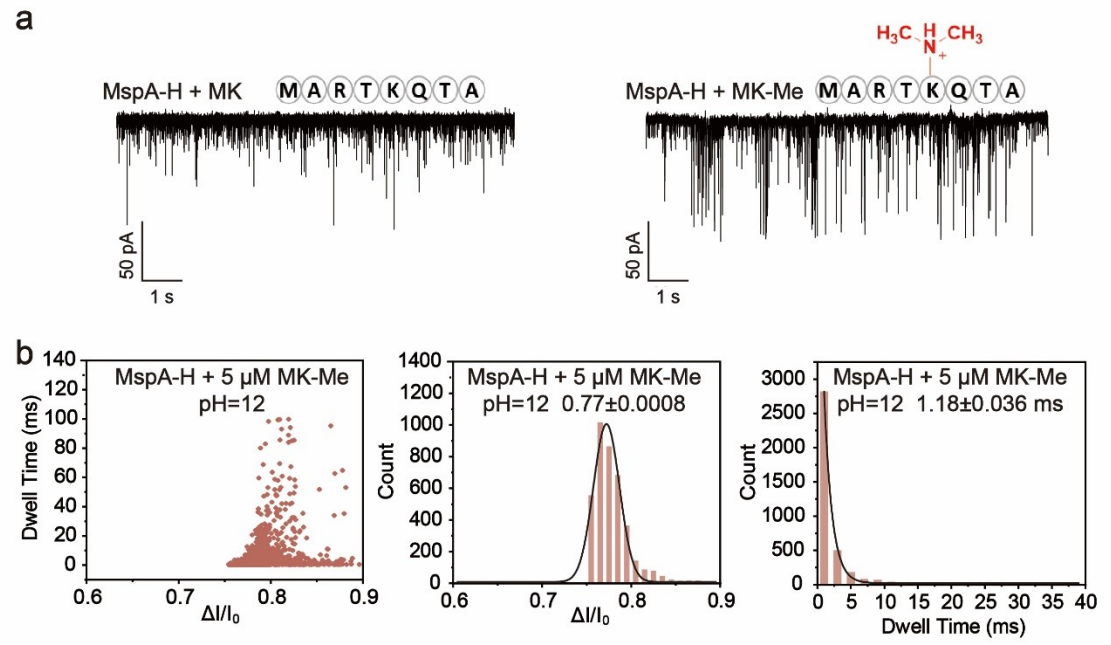

Figure S44. (a) Representative current traces of MK and MK-Me with MspA-H. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 12.0. All peptides were added to a final concentration of 5.0 μM. (b) Scatter plot of $\Delta I/I_0$, histogram of $\Delta I/I_0$ and dwell time histogram for MK-Me interactions with MspA-H at an applied potential of +50 mV.

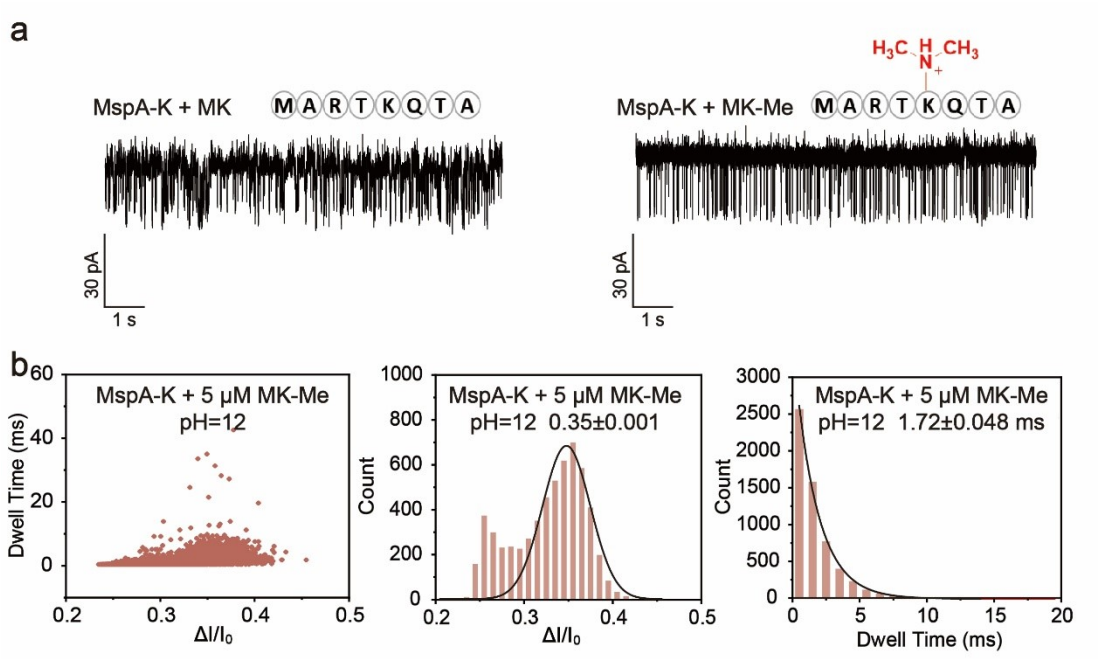

Figure S45. (a) Representative current traces of MK and MK-Me with MspA-K. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 12.0. All peptides were added to a final concentration of 5.0 μM. (b) Scatter plot of $\Delta I/I_0$, histogram of $\Delta I/I_0$ and dwell time histogram for MK-Me interactions with MspA-K at an applied potential of +50 mV.

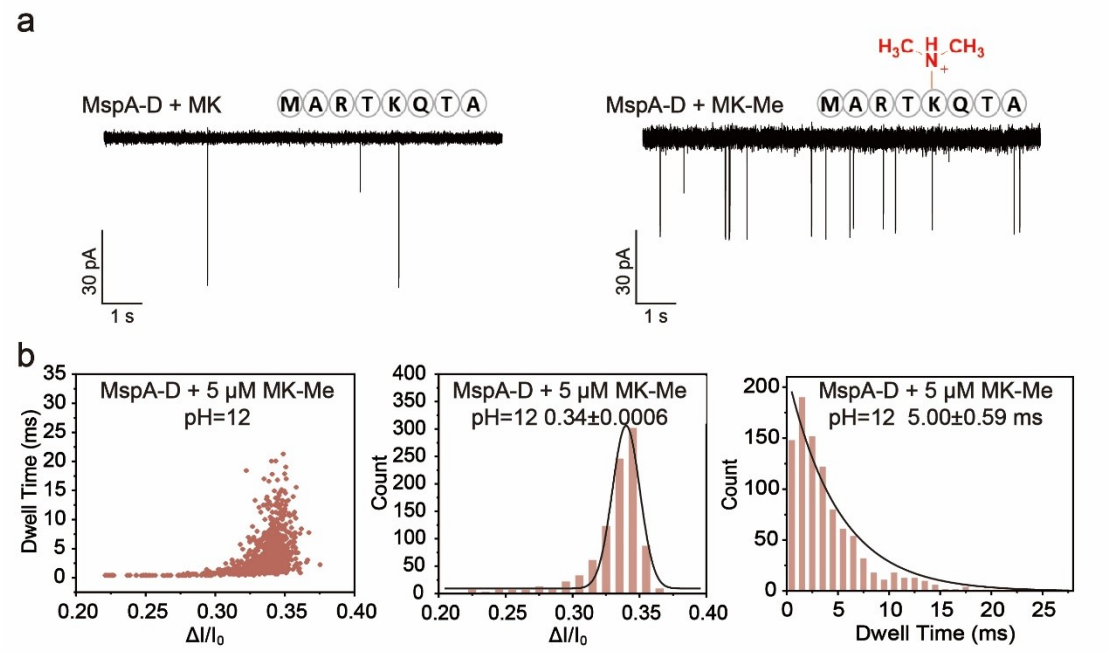

Figure S46. (a) Representative current traces of MK and MK-Me with MspA-D. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 12.0. All peptides were added to a final concentration of 5.0 μM. (b) Scatter plot of $\Delta I/I_0$, histogram of $\Delta I/I_0$ and dwell time histogram for MA-Me interactions with MspA-D at an applied potential of +50 mV.

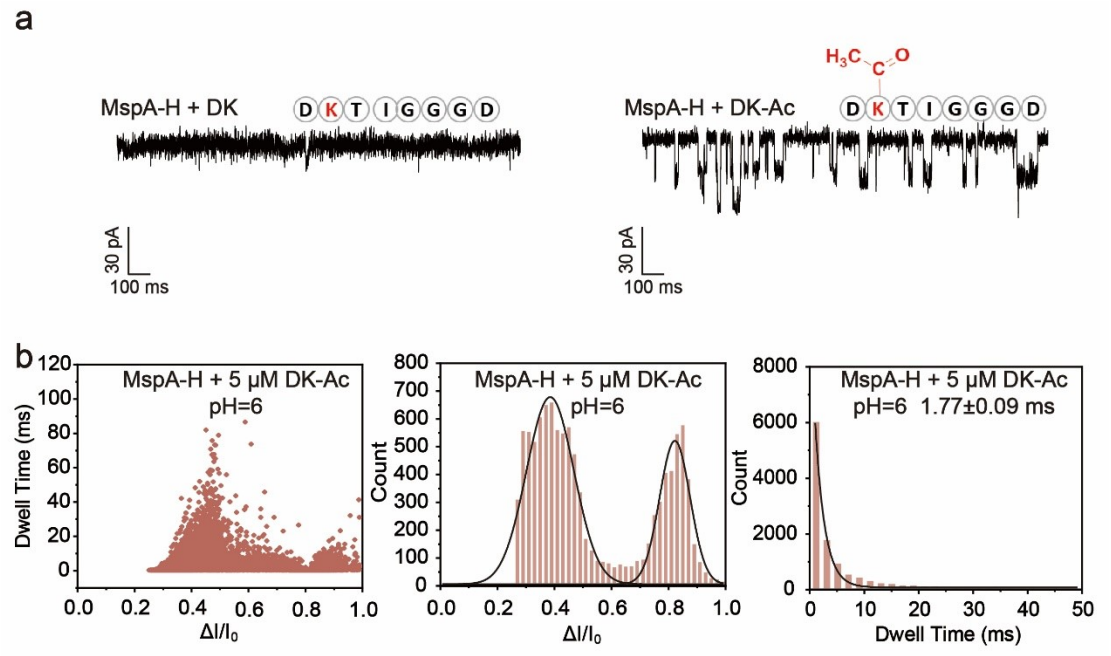

Figure S47. (a) Representative current traces of DK and DK-Ac with MspA-H. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 6.0. All peptides were added to a final concentration of 5.0 μM. (b) Scatter plot of $\Delta I/I_0$, histogram of $\Delta I/I_0$ and dwell time histogram for DK-Ac interactions with MspA-H at an applied potential of +50 mV.

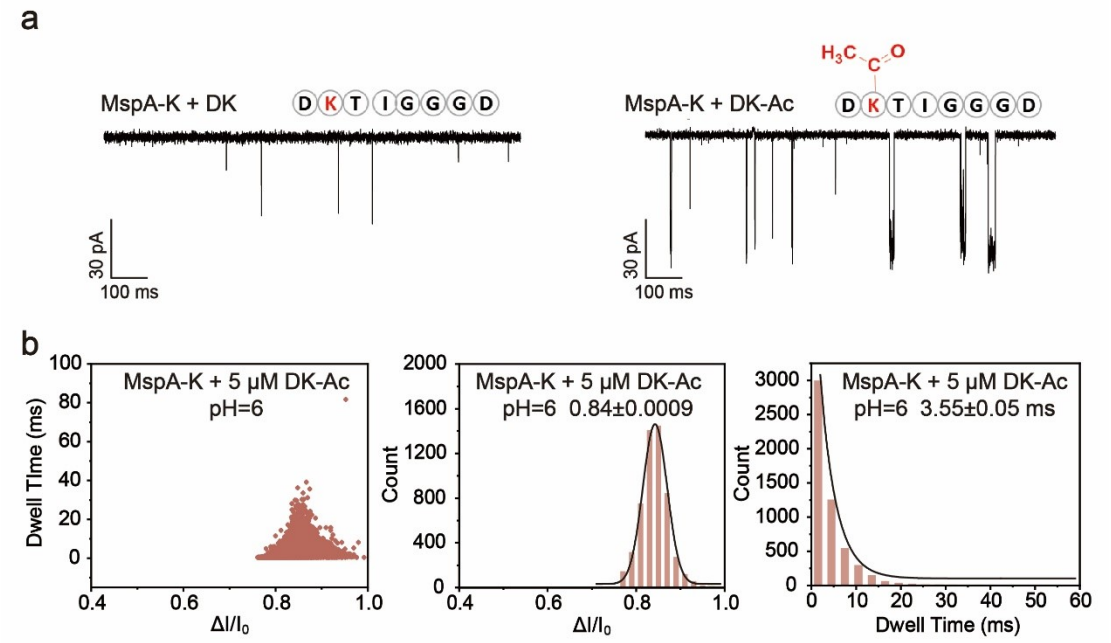

Figure S48. (a) Representative current traces of DK and DK-Ac with MspA-K. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 6.0. All peptides were added to a final concentration of 5.0 μM. (b) Scatter plot of $\Delta I/I_0$, histogram of $\Delta I/I_0$ and dwell time histogram for DK-Ac interactions with MspA-K at an applied potential of +50 mV.

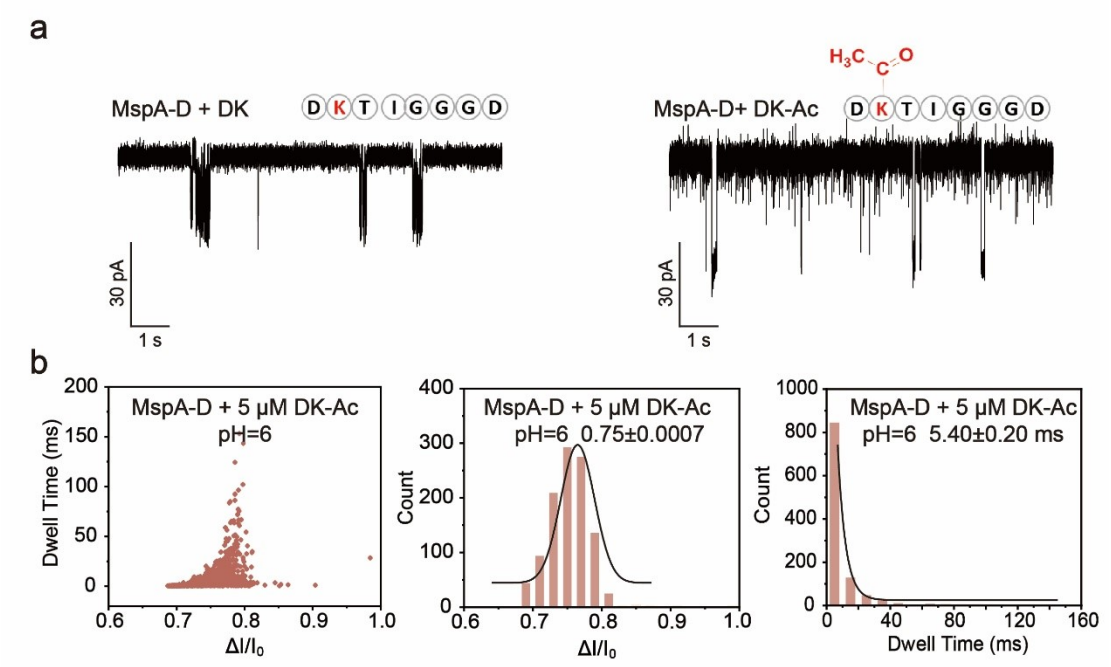

Figure S49. (a) Representative current traces of DK and DK-Ac with MspA-D. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 6.0. All peptides were added to a final concentration of 5.0 μM. (b) Scatter plot of $\Delta I/I_0$, histogram of $\Delta I/I_0$ and dwell time histogram for DK-Ac interactions with MspA-D at an applied potential of +50 mV.

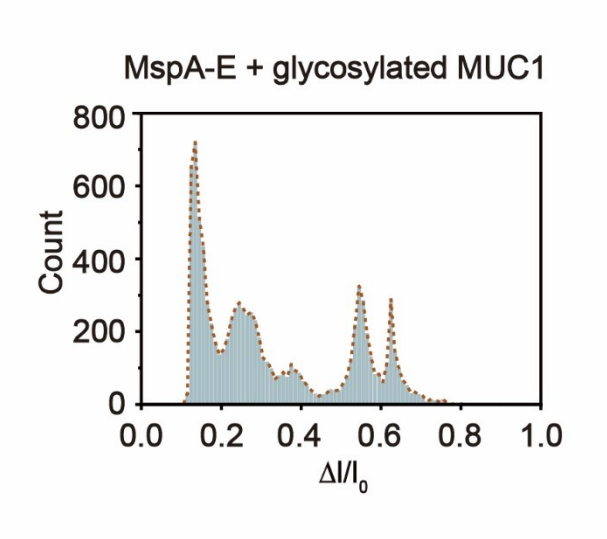

Figure S50. The histogram of $\Delta I/I_0$ for glycosylated MUC1 following tryptic digestion, as detected with the MspA-E nanopore. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 9.0, at an applied potential of +50 mV.

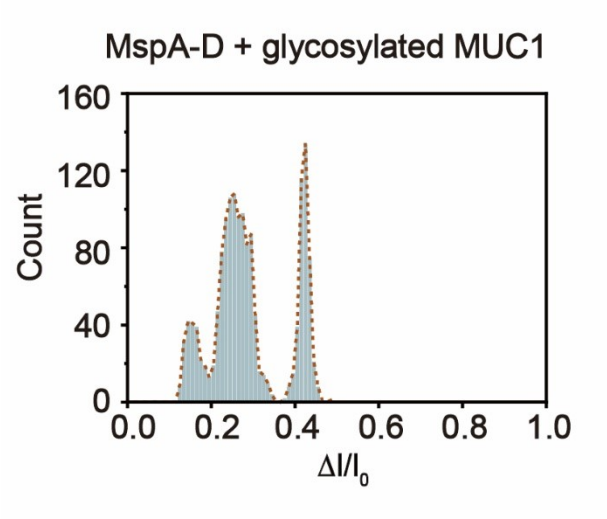

Figure S51. The histogram of $\Delta I/I_0$ for glycosylated MUC1 following tryptic digestion, as detected with the MspA-D nanopore. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 9.0, at an applied potential of +50 mV.

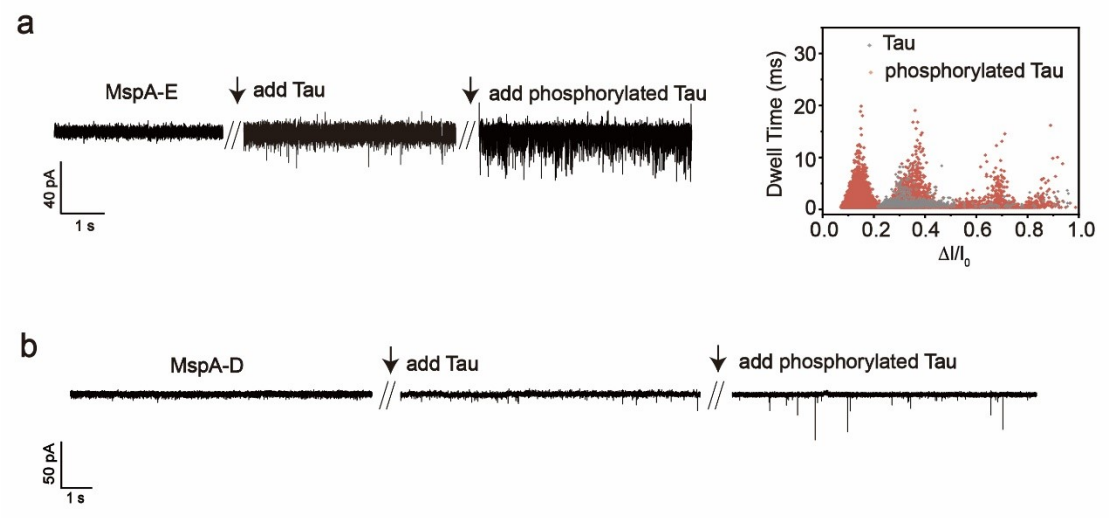

Figure S52. (a) Representative current traces of tryptic digested Tau and phosphorylated Tau protein with the MspA-E nanopore(left). Scatter plot of $\Delta I/I_0$ with the addition of tryptic digested Tau and phosphorylated Tau protein (right). (b) Representative current traces of tryptic digested Tau and phosphorylated Tau protein with the MspA-D nanopore. The measurements were carried out in the buffer of 1.0 M NaCl, 10.0 mM HEPES, pH 7.4, at an applied potential of +50 mV.

Supplemental Tables

Table S1. Mutation sites of the MspA protein.

Name	Mutation Site
M2 MspA	D93N/D91N/D90N/D118R/D134R/E139K
MspA-Q	D93N/N91Q/D90N/D118R/D134R/E139K
MspA-H	D93N/N91H/D90N/D118R/D134R/E139K
MspA-K	D93N/N91K/D90N/D118R/D134R/E139K
MspA-R	D93N/N91R/D90N/D118R/D134R/E139K
MspA-D	D93N/N91D/D90N/D118R/D134R/E139K
MspA-E	D93N/N91E/D90N/D118R/D134R/E139K
MspA-F	D93N/N91F/D90N/D118R/D134R/E139K
MspA-W	D93N/N91W/D90N/D118R/D134R/E139K

Table S2. Peptide sequence and pH value of electrolyte. Peptide sequence displayed from the N-terminus (left) to the C-terminus (right). Amino acid residues harboring post-translational modifications (PTMs) and their specific modifying groups were marked in red. Phosphorylation, methylation, acetylation, glycosylation, crotonylation, succinylation, nitration and sulfonation were abbreviated as p, Me, Ac, GalNAc, cr, succ, n and s. The nanopore measurements were performed in 1 M NaCl and 10 mM HEPES buffer.

	Name	Peptide sequences	The pH of measurement buffer
1	HY	HYPIAPSYPYPK	7.4
2	HY-3p	H(Y-p)PIAPS(Y-p)P(Y-p)PK	7.4
3	HY-2p	HYPIAPS(Y-p)P(Y-p)PK	7.4
4	HY-1p	H(Y-p)PIAPSYPYPK	7.4
5	HD	HDPIAPSDPDPK	7.4
6	ET	TPPAPETPPSSGEPPE	7.4
7	QT	TPPAPQTPPSSGEPPQ	7.4
8	KT	TPPAPKTPPSSGEPPK	7.4
9	ET-p	TPPAPE(T-p)PPSSGEPPE	7.4
10	QT-p	TPPAPQ(T-p)PPSSGEPPQ	7.4
11	KT-p	TPPAPK(T-p)PPSSGEPPK	7.4
12	MK	MARTKQTA	12
13	MK-Me	MART(K-Me)QTA	12
14	DK	DKTIGGGD	6
15	DK-Ac	D(K-Ac)TIGGGD	6
16	ST	SAPDTRPAP	9
17	ST-GalNAc	SAPD(T-α-D-GalNAc)RPAP	9
18	PD	PTSGSAYEMQTKIDDD	7.4
19	PD-p	PT(S-p)GSAYEMQTKIDDD	7.4
20	PD-Me	PT(S-Me)GSAYEMQTKIDDD	7.4
21	PD-Ac	PT(S-Ac)GSAYEMQTKIDDD	7.4

22	PD- GalNAc	PT(S-α-D-GalNAc)GSAYEMQTKIDDD	7.4
23	PD-cr	PTSGSAYEMQT(K-cr)IDDD	7.4
24	PD-succ	PTSGSAYEMQT(K-succ)IDDD	7.4
25	PD-n	PTSGSA(Y-n)EMQTKIDDD	7.4
26	PD-s	PTSGSA(Y-s)EMQTKIDDD	7.4

Table S3. $\Delta I/I_0$ and dwell time of KT, KT-p, QT, QT-p, ET and ET-p.

		KT	KY-p	QT	QT-p	ET	ET-p
M2 MspA	$\Delta I/I_0$	/	/	/	/	/	0.78±0.0005
	dwell time (ms)	/	/	/	/	/	27.64±1.12
MspA-Q	$\Delta I/I_0$	/	/	/	/	/	/
	dwell time (ms)	/	/	/	/	/	/
MspA-K	$\Delta I/I_0$	0.65±0.003	0.59±0.0006	0.12±0.009	0.23±0.001		0.29±0.0007
	dwell time (ms)	4.48±0.43	9.82±0.06	1.53±0.12	0.71±0.01	/	2.57±0.096
MspA-D	$\Delta I/I_0$	0.12±0.0005	0.14±0.002	0.42±0.002	0.42±0.0005	/	/
	dwell time (ms)	0.13±0.008	2.90±0.13	3.93±0.13	17.16±0.65	/	/
MspA-E	$\Delta I/I_0$	0.29±0.003	0.51±0.004	0.55±0.007	0.72±0.001	/	0.40±0.002
	dwell time (ms)	0.52±0.004	15.66±1.57	0.71±0.015	35.79±5.43	/	0.75±0.008
MspA-H	$\Delta I/I_0$	/	/	/	/	/	/
	dwell time (ms)	/	/	/	/	/	/
MspA-W	$\Delta I/I_0$	/	/	/	/	/	/
	dwell time (ms)	/	/	/	/	/	/
MspA-F	$\Delta I/I_0$	/	/	/	/	/	/
	dwell time (ms)	/	/	/	/	/	/

Table S4. Validation and testing accuracies of different models for MspA-K. The dataset was then split into a training set (90%) and a testing set (10%) for model training and testing. All models were trained using the Classification Learner toolbox in MATLAB. The validation accuracies were derived from the 10-fold cross-validation results. The Bagged Trees model reported the best score, which is marked in red.

	Model Name	Model Type	Accuracy (Validation)	Accuracy (Test)
1	Fine Tree	Tree	97.8 %	97.8 %
2	Medium Tree	Tree	97.4 %	97.8 %
3	Coarse Tree	Tree	81.9 %	81.9 %
4	Linear Discriminant	Discriminant	95.2 %	96.2 %
5	Quadratic Discriminant	Discriminant	96.6 %	97.3 %
6	Gaussian Naive Bayes	Naive Bayes	96.1 %	96.7 %
7	Kernel Naive Bayes	Naive Bayes	96.0 %	96.1 %
8	Linear SVM	SVM	97.4 %	98.2 %
9	Quadratic SVM	SVM	97.9 %	98.5 %
10	Fine Gaussian SVM	SVM	97.0 %	97.6 %
11	Medium Gaussian SVM	SVM	97.9 %	98.3 %
12	Coarse Gaussian SVM	SVM	97.0 %	97.9 %
13	Fine KNN	KNN	95.8 %	96.4 %
14	Medium KNN	KNN	95.4 %	95.8 %
15	Coarse KNN	KNN	89.5 %	90.8 %
16	Cosine KNN	KNN	86.9 %	88.0 %
17	Cubic KNN	KNN	95.2 %	95.4 %
18	Weighted KNN	KNN	96.1 %	96.3 %
19	Boosted Trees	Ensemble	97.7 %	98.0 %
20	Bagged Trees	Ensemble	98.2 %	98.3 %
21	Subspace Discriminant	Ensemble	92.3 %	94.0 %
22	Subspace KNN	Ensemble	90.0 %	90.7 %
23	RUSBoosted Trees	Ensemble	97.6 %	98.1 %
24	Narrow Neural Network	Neural Network	97.8 %	98.1 %
25	Medium Neural	Neural Network	97.9 %	98.4 %
26	Wide Neural Network	Neural Network	97.8 %	98.4 %
27	Bilayered Neural	Neural Network	98.0 %	98.6 %
28	Trilayered Neural	Neural Network	97.9 %	98.2 %
29	SVM Kernel	Kernel	95.3 %	95.6 %
30	Logistic Regeression	Kernel	94.3 %	94.8 %

Table S5. Validation and testing accuracies of different models for MspA-E. The dataset was then split into a training set (90%) and a testing set (10%) for model training and testing. All models were trained using the Classification Learner toolbox in MATLAB. The validation accuracies were derived from the 10-fold cross-validation results. The Bagged Trees model reported the best score, which is marked in red.

	Model Name	Model Type	Accuracy (Validation)	Accuracy (Test)
1	Fine Tree	Tree	97.6 %	97.9 %
2	Medium Tree	Tree	96.4 %	97.1 %
3	Coarse Tree	Tree	81.6 %	83.5 %
4	Linear Discriminant	Discriminant	92.0 %	92.6 %
5	Quadratic Discriminant	Discriminant	94.0 %	94.5 %
6	Gaussian Naive Bayes	Naive Bayes	93.2 %	93.9 %
7	Kernel Naive Bayes	Naive Bayes	96.7 %	96.7 %
8	Linear SVM	SVM	95.3 %	96.6 %
9	Quadratic SVM	SVM	96.1 %	96.5 %
10	Fine Gaussian SVM	SVM	95.8 %	95.8 %
11	Medium Gaussian SVM	SVM	96.2 %	96.6 %
12	Coarse Gaussian SVM	SVM	95.4 %	96.3 %
13	Fine KNN	KNN	95.5 %	95.8 %
14	Medium KNN	KNN	95.2 %	96.0 %
15	Coarse KNN	KNN	92.3 %	93.0 %
16	Cosine KNN	KNN	93.8 %	94.3 %
17	Cubic KNN	KNN	95.2 %	96.0 %
18	Weighted KNN	KNN	95.8 %	96.5 %
19	Boosted Trees	Ensemble	97.0 %	98.0 %
20	Bagged Trees	Ensemble	97.9 %	98.2 %
21	Subspace Discriminant	Ensemble	86.3 %	85.1 %
22	Subspace KNN	Ensemble	92.0 %	90.6 %
23	RUSBoosted Trees	Ensemble	97.2 %	97.3 %
24	Narrow Neural Network	Neural Network	96.9 %	98.2 %
25	Medium Neural	Neural Network	97.0 %	97.9 %
26	Wide Neural Network	Neural Network	96.9 %	97.4 %
27	Bilayered Neural	Neural Network	97.1 %	98.0 %
28	Trilayered Neural	Neural Network	96.8 %	97.9 %
29	SVM Kernel	Kernel	94.6 %	94.7 %
30	Logistic Regeression	Kernel	93.9 %	93.3 %

Table S6. Analysis of the interaction energy between PD (the peptide sequence is PTSGSAYEMQTKIDDD) and MspA-K using molecular docking.

Name	Distance	Categroy	Ypype
B:ASN93:N - ligand:O137	3.03416	Hydrogen Bond	Conventional Hydrogen Bond
C:ASN102:ND2 - ligand:O144	3.36078	Hydrogen Bond	Conventional Hydrogen Bond
ligand:H17 - H:ASN102:O	2.35263	Hydrogen Bond	Conventional Hydrogen Bond
ligand:H69 - E:LYS91:O	2.76622	Hydrogen Bond	Conventional Hydrogen Bond
ligand:H89 - D:ASN102:O	2.49718	Hydrogen Bond	Conventional Hydrogen Bond
ligand:H90 - C:LYS91:O	2.7119	Hydrogen Bond	Conventional Hydrogen Bond
ligand:H130 - B:SER103:OG	2.11737	Hydrogen Bond	Conventional Hydrogen Bond
ligand:H140 - B:LYS91:O	2.67342	Hydrogen Bond	Conventional Hydrogen Bond
F:SER103:OG - ligand	4.09166	Hydrogen Bond	Pi-Donor Hydrogen Bond

Table S7. Analysis of the interaction energy between PD-p (the peptide sequence is PT(S-p)GSAYEMQTKIDDD) and MspA-K using molecular docking.

Name	Distance	Categroy	Type
B:ASN93:N - d:ligand:O21	3.19943	Hydrogen Bond	Conventional Hydrogen Bond
C:ASN93:N - d:ligand:O22	3.27526	Hydrogen Bond	Conventional Hydrogen Bond
E:LYS91:NZ - d:ligand:O12	3.36267	Hydrogen Bond	Conventional Hydrogen Bond
H:ASN102:ND2 - d:ligand:O43	3.20311	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:O23 - A:ASN102:OD1	3.25526	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:O50 - H:ASN102:O	2.92943	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:O61 - G:LYS91:O	3.36143	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:O61 - H:ASN102:O	3.208	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:2H17 - D:ASN102:O	2.26411	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:3H17 - C:LYS91:O	2.48761	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:3H17 - D:SER103:OG	2.79887	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:4H21 - B:LYS91:O	2.57307	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:8H21 - B:ASN102:OD1	2.25343	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:8H22 - C:ASN102:O	2.85048	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:2H23 - C:LYS91:O	2.13196	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:H44 - H:LYS91:O	2.30103	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:H88 - G:ASN102:O	2.54972	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:H88 - F:LYS91:O	2.20945	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:H98 - G:LYS91:O	2.56875	Hydrogen Bond	Conventional Hydrogen Bond
A:GLY92:CA - d:ligand:O23	3.43596	Hydrogen Bond	Carbon Hydrogen Bond
A:SER103:CB - d:ligand:O23	3.73074	Hydrogen Bond	Carbon Hydrogen Bond
B:GLY92:CA - d:ligand:O21	3.32017	Hydrogen Bond	Carbon Hydrogen Bond
H:SER103:CB - d:ligand:O50	3.59117	Hydrogen Bond	Carbon Hydrogen Bond
d:ligand:C59 - H:ASN102:O	3.39058	Hydrogen Bond	Carbon Hydrogen Bond
d:ligand:C11 - E:LYS91	3.86796	Hydrophobic	Alkyl

Table S8. Analysis of the interaction energy between PD-cr (the peptide sequence is PTSGSAYEMQT(K-cr)IDDD) and MspA-K using molecular docking.

Name	Distance	Categroy	Type
A:LYS91:NZ - d:ligand:O72	3.31371	Hydrogen Bond	Conventional Hydrogen Bond
F:ASN93:N - d:ligand:O22	3.15841	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:5H10 - A:SER103:OG	2.97859	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:8H13 - H:ASN102:O	2.02827	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:9H13 - G:LYS91:O	2.19619	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:7H21 - E:LYS91:O	2.09235	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:1H23 - G:ASN102:O	2.36592	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:H58 - D:ASN102:O	2.31379	Hydrogen Bond	Conventional Hydrogen Bond
d:ligand:H64 - C:LYS91:O	2.78346	Hydrogen Bond	Conventional Hydrogen Bond
B:SER103:CB - d:ligand:O80	3.48453	Hydrogen Bond	Carbon Hydrogen Bond

Table S9. The sequence of MUC1. Due to the restriction of trypsin activity by the proline (Pro) heterocyclic structure in the MUC1 sequence (especially at KP or RP sites), cleavage occurs at only three specific sites (indicated in red).

The sequence of MUC1
MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSMTSSVLSSHSPGSGSSTTQGQDVTLAPATEPAS GSAATWGQDVTSVPVTRPALGSTTPPAHDVTSAPDNKPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAP PAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGV TSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT RPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSGSHHHHHH*

Supplemental Data

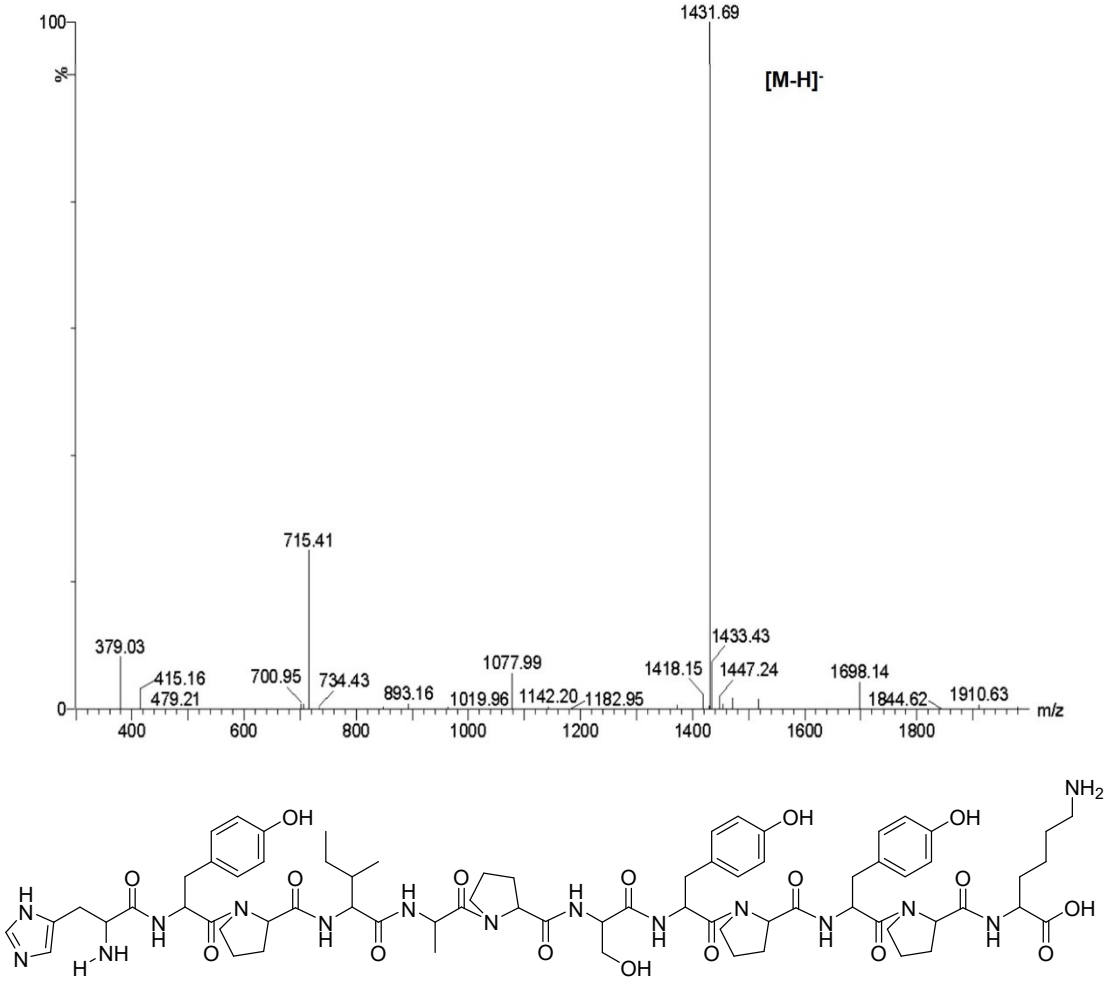

Figure S53. MS analysis of the peptide HYPIAPSYPYPK (termed HY). m/z: 1432.65 (100.0%). MS (ESI) m/z: $[M-H]^-$ calcd, 1431.65; found,1431.69. $[M-2H]^{2-}$ calcd, 715.33; found, 715.41.

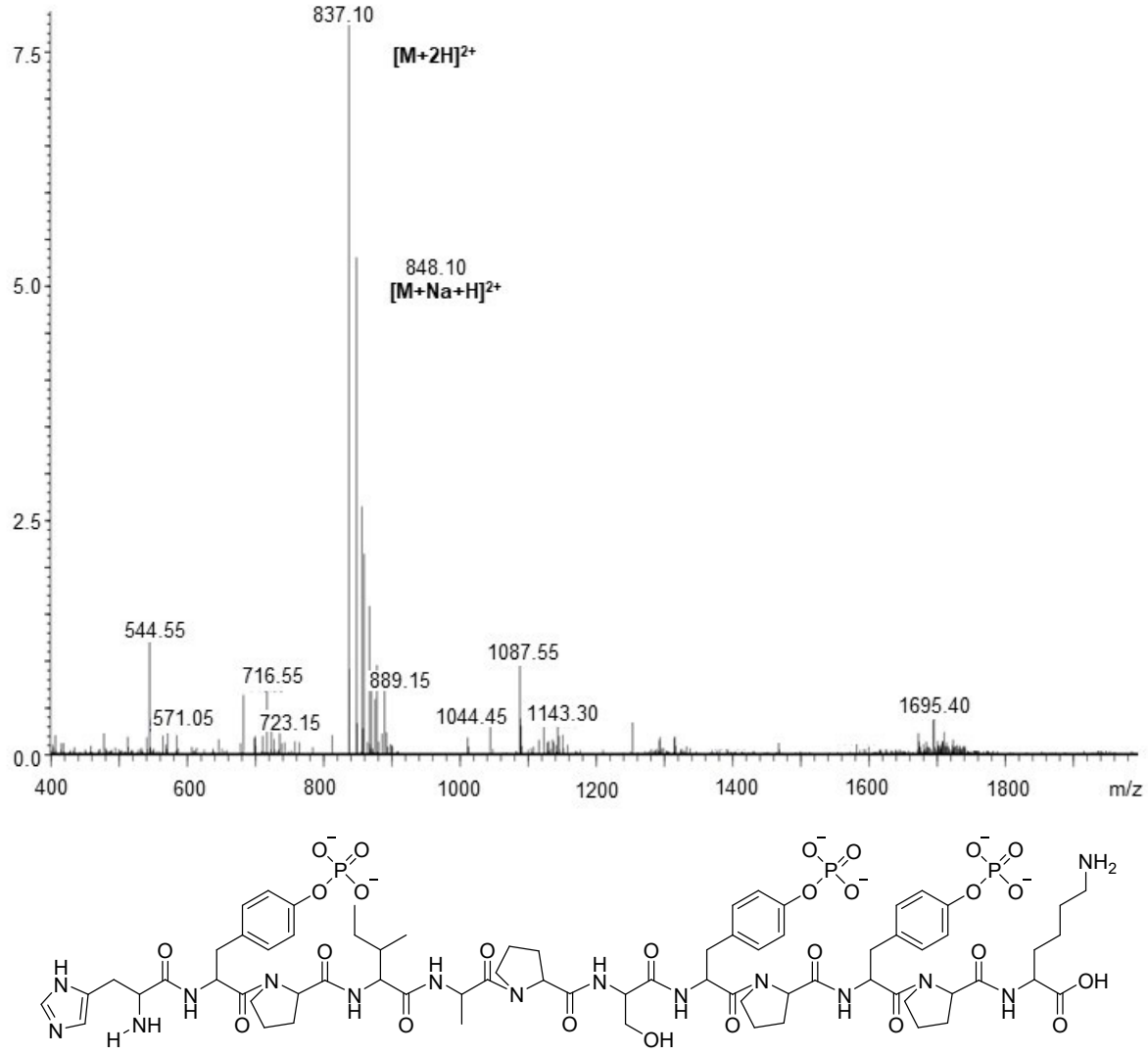

Figure S54. Mass spectrometry of the peptide H(Y-p)PIAPS(Y-p)P(Y-p)PK (termed HY-3p). m/z: 1672.55 (100.0%). MS (ESI) m/z: $[M+2H]^{2+}$ calcd, 837.28; found, 837.10.

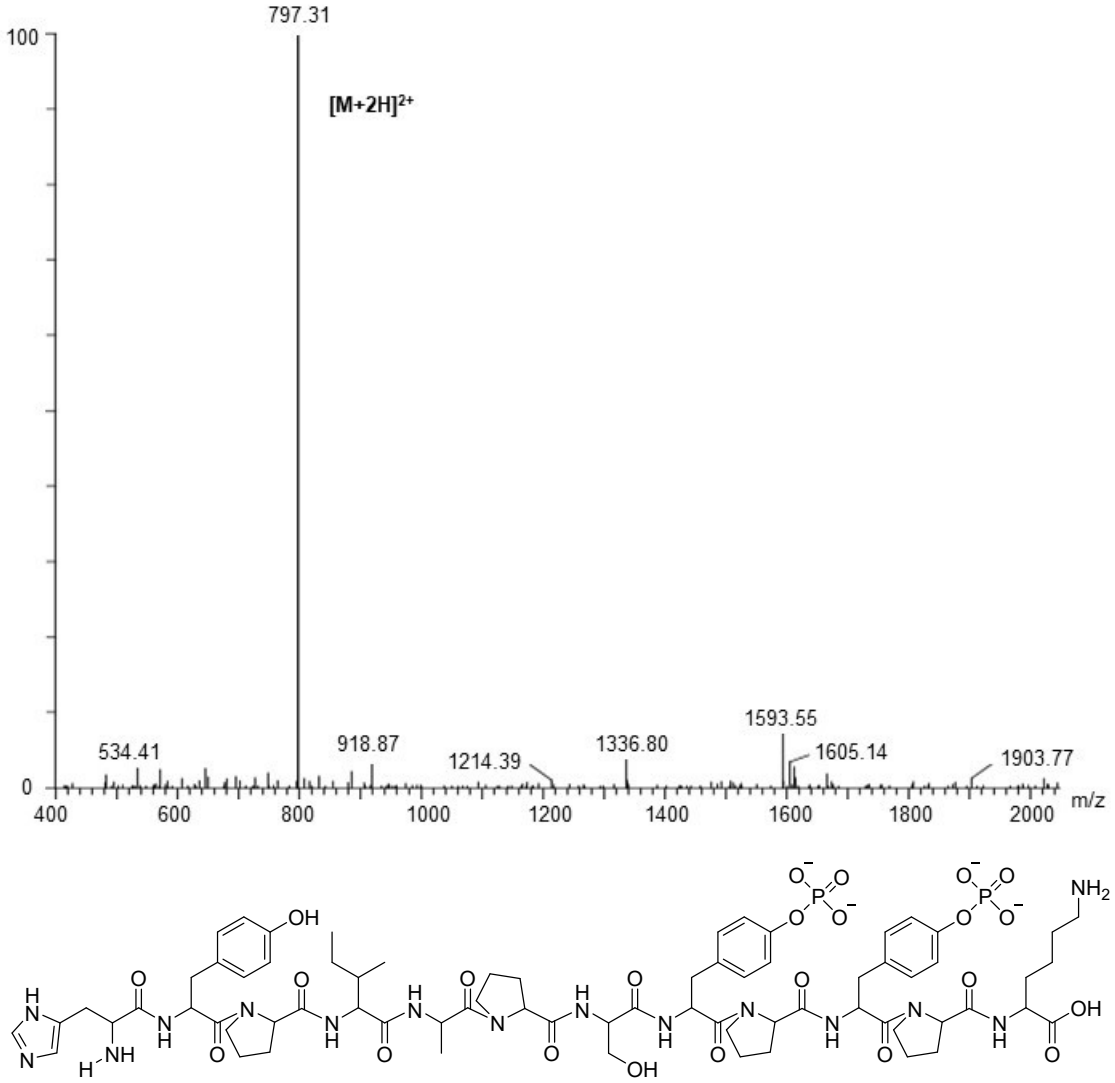

Figure S55. Mass spectrometry of the peptide HYPIAPS(Y-p)P(Y-p)PK (termed HY-2p). m/z: 1593.60 (100.0%). MS (ESI) m/z: $[M+2H]^{2+}$ calcd, 797.80; found, 797.31.

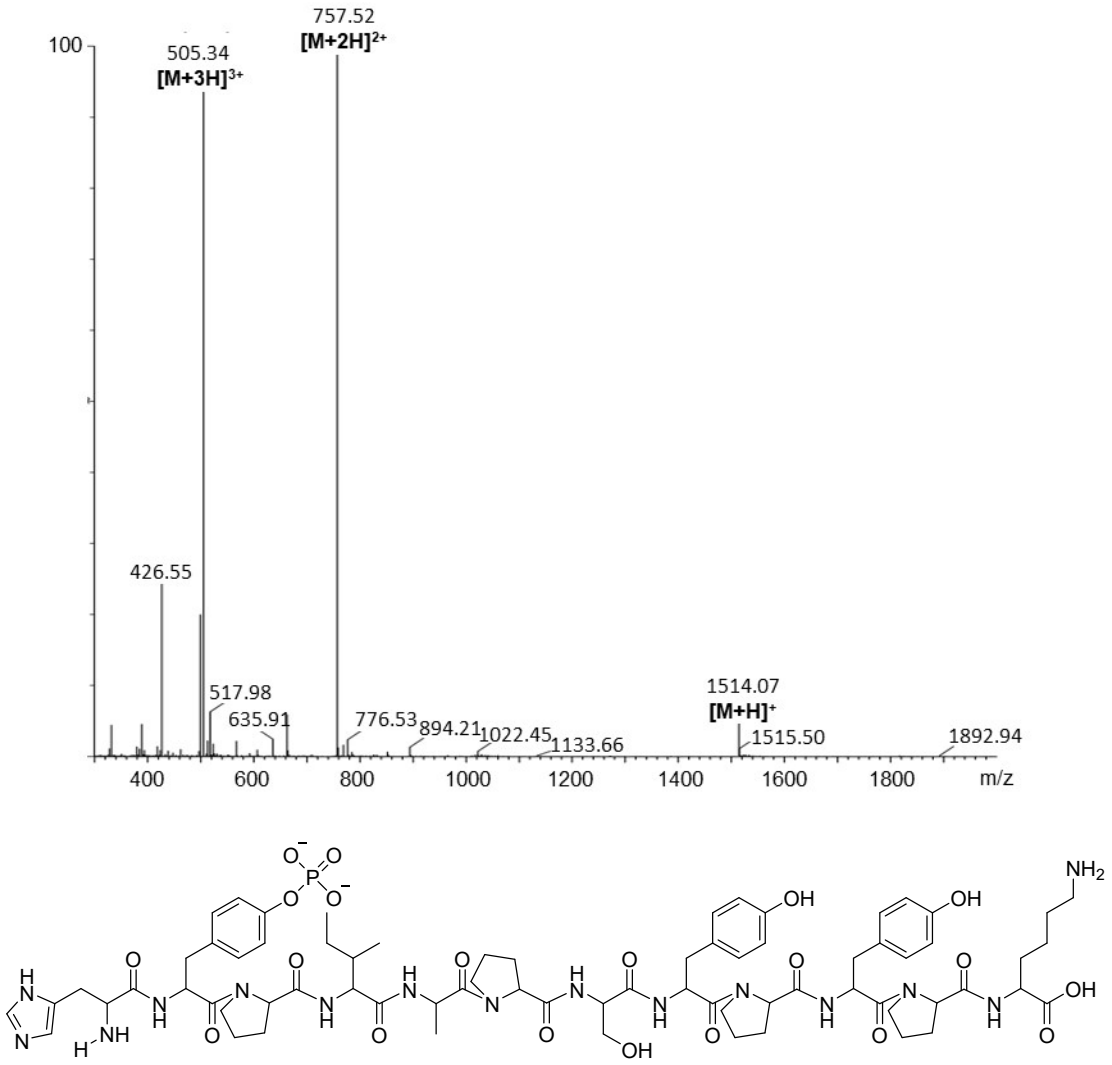

Figure S56. Mass spectrometry of the peptide H(Y-p)PIAPSYPYPK (termed HY-1p). m/z: 1512.64 (100.0%). MS (ESI) m/z: $[M+H]^+$ calcd, 1513.64; found, 1514.07. $[M+2H]^{2+}$ calcd, 757.32; found, 757.52. $[M+3H]^{3+}$ calcd, 505.21; found, 505.34.

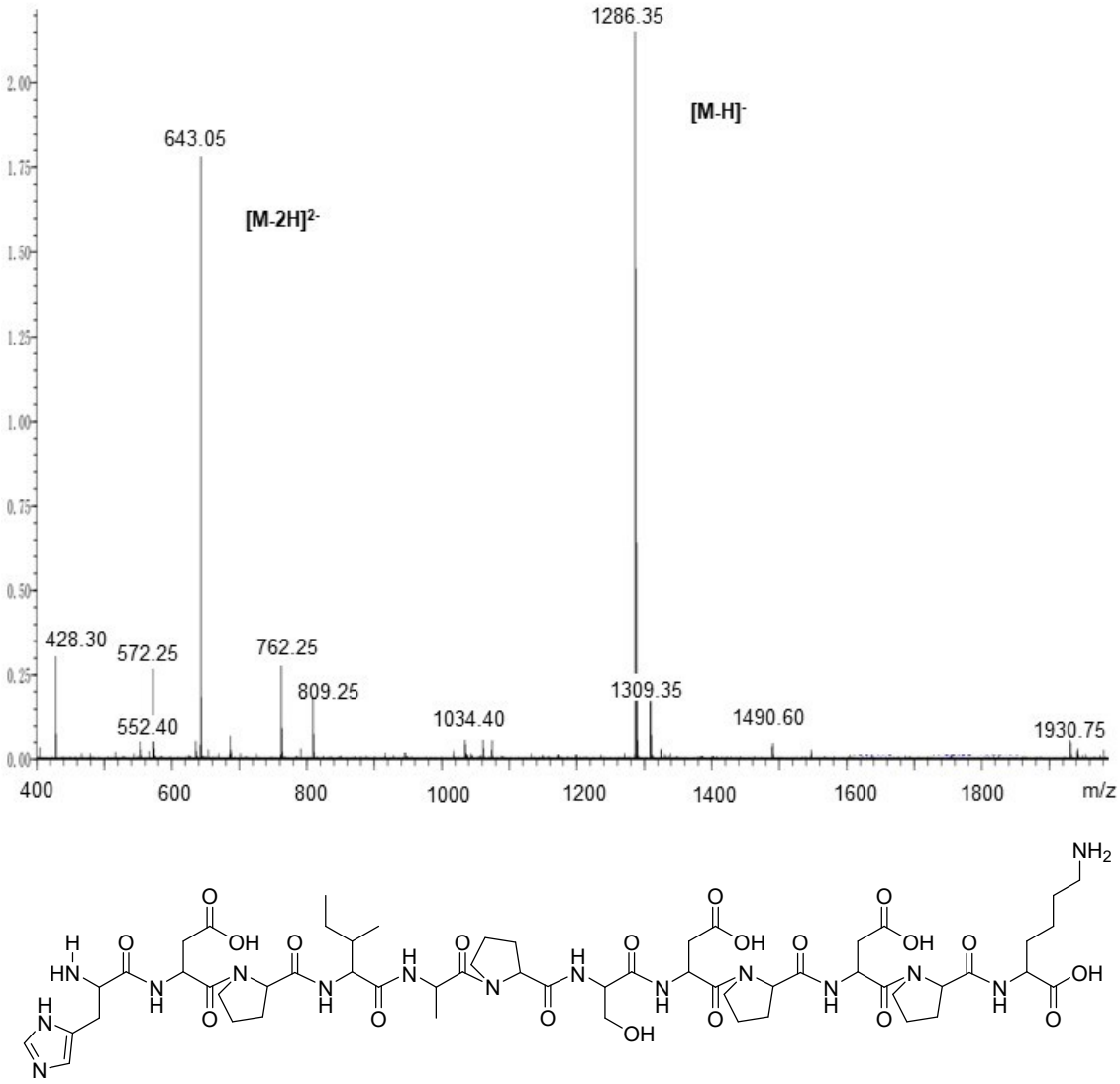

Figure S57. Mass spectrometry of the peptide HDPIAPSDPDPK (termed HD). m/z: 1288.38 (100.0%). MS (ESI) m/z: $[M-H]^-$ calcd, 1287.38; found, 1286.35. $[M-2H]^{2-}$ calcd, 643.19; found, 643.05.

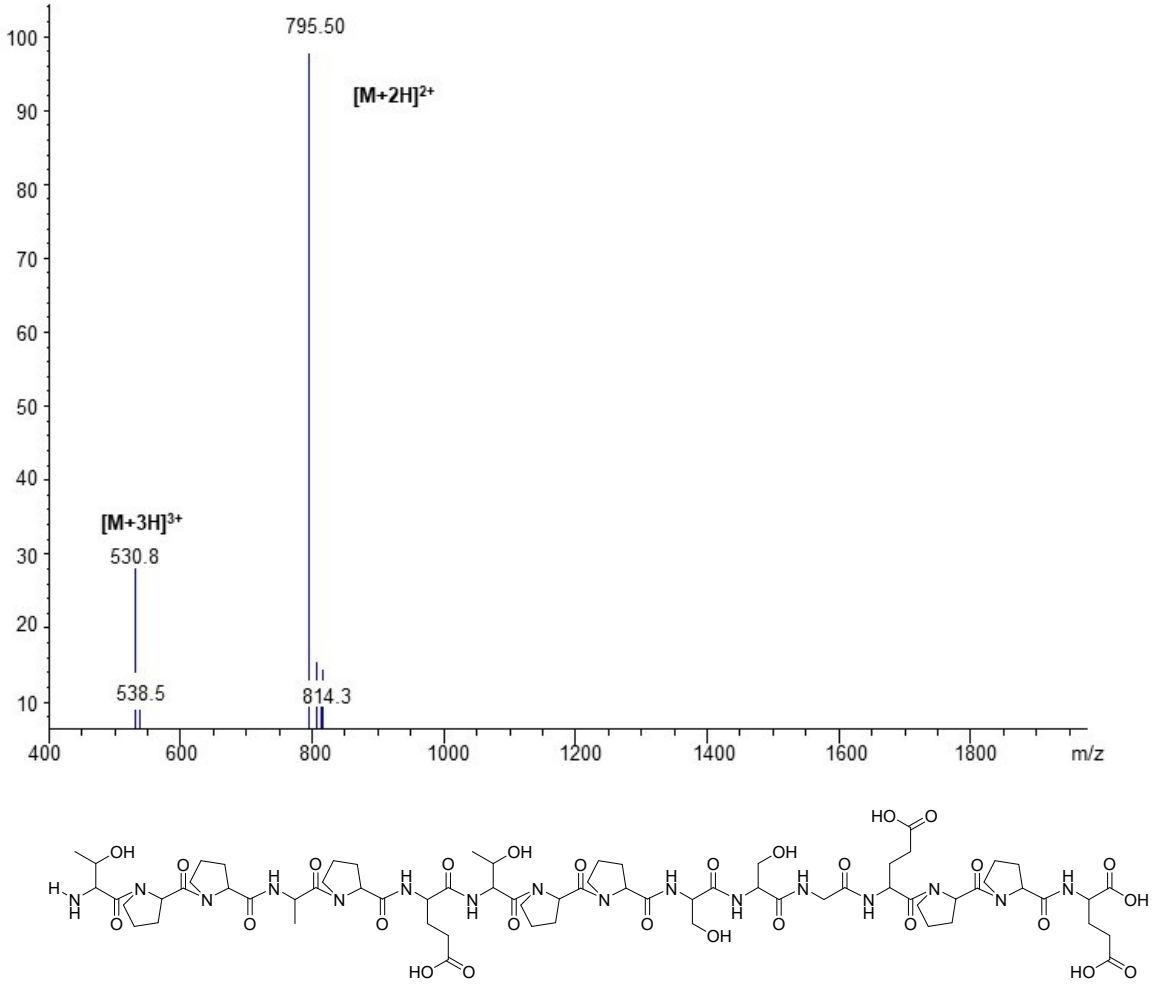

Figure S58. Mass spectrometry of the peptide TPPAPETPPSSGEPPE (termed ET). m/z: 1589.65 (100.0%). MS (ESI) m/z: $[M+2H]^{2+}$ calcd, 795.83; found, 795.50. $[M+3H]^{3+}$ calcd, 530.87; found, 530.80.

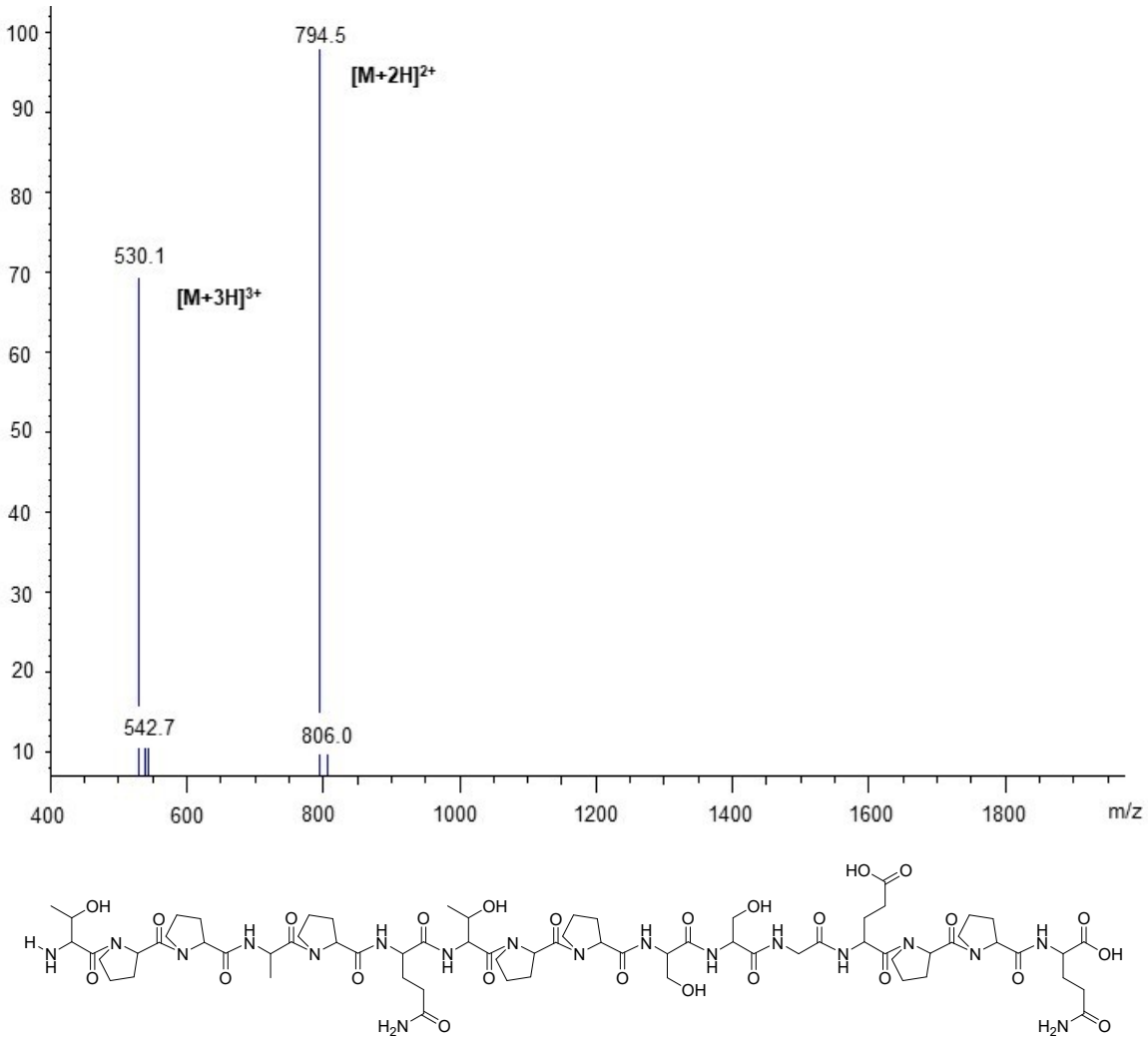

Figure S59. Mass spectrometry of the peptide TPPAPQTPPSSGEPPQ (termed QT). m/z: 1587.68 (100.0%). MS (ESI) m/z: $[M+2H]^{2+}$ calcd, 794.84; found, 794.50. $[M+3H]^{3+}$ calcd, 530.23; found, 530.10.

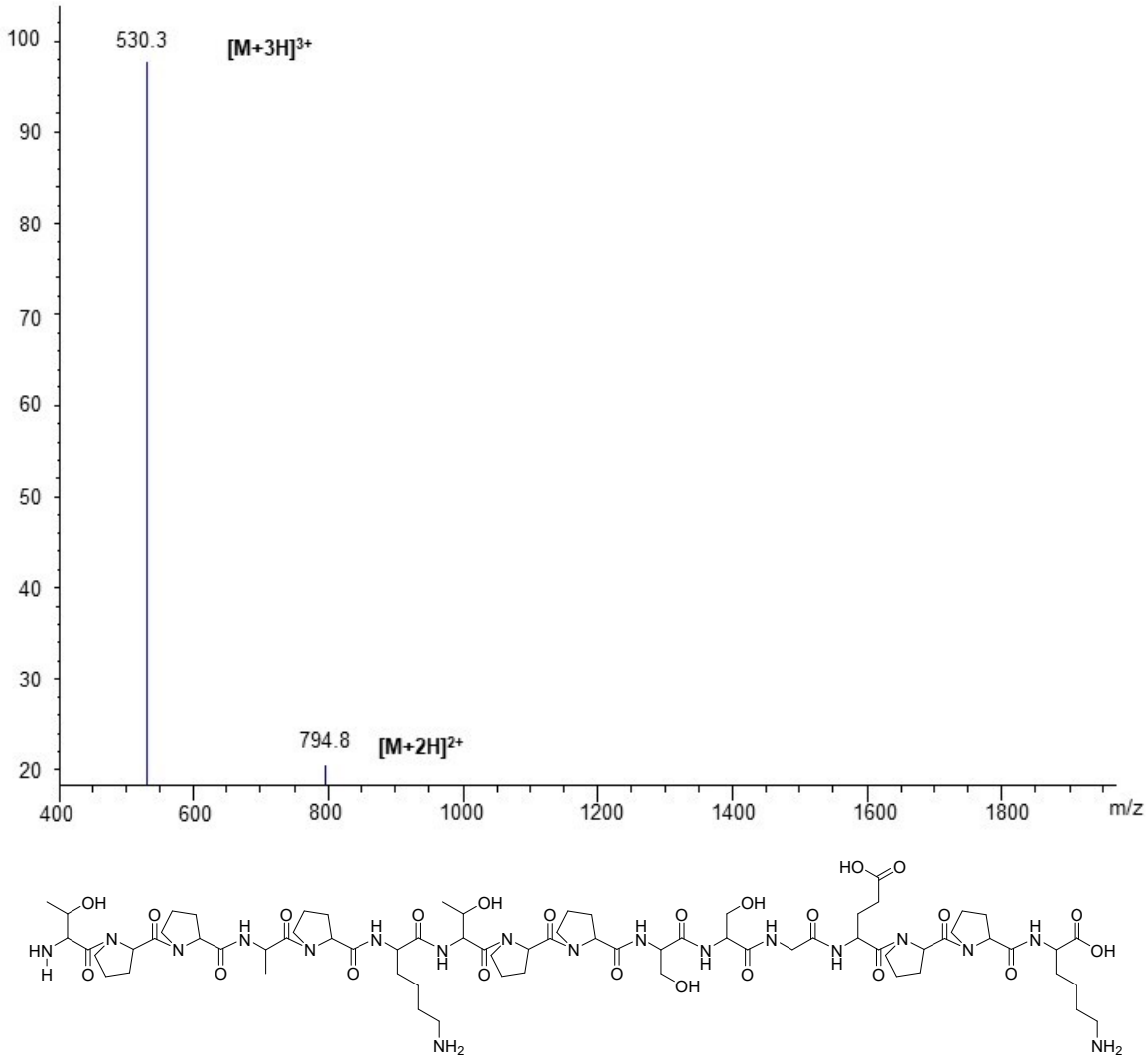

Figure S60. Mass spectrometry of the peptide TPPAPKTPPSSGEPPK (termed KT). m/z: 1587.77 (100.0%). MS (ESI) m/z: $[M+2H]^{2+}$ calcd, 794.89; found, 794.80. $[M+3H]^{3+}$ calcd, 530.26; found, 530.30.

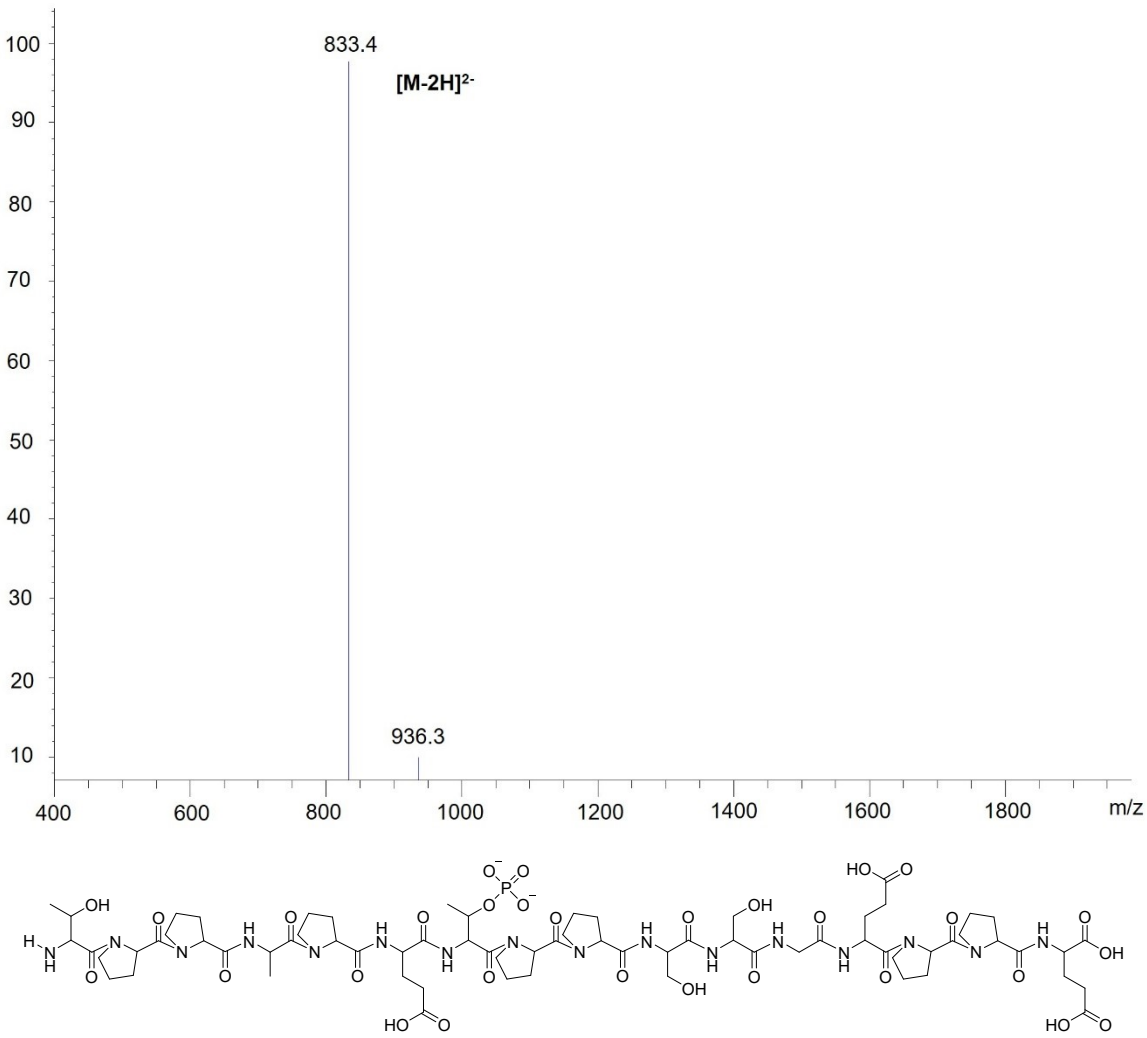

Figure S61. Mass spectrometry of the peptide TPPAPE(T-p)PPSSGEPPE. (termed ET-p). m/z: 1669.63 (100.0%). MS (ESI) m/z: $[M-2H]^{2-}$ calcd, 833.82; found, 833.40.

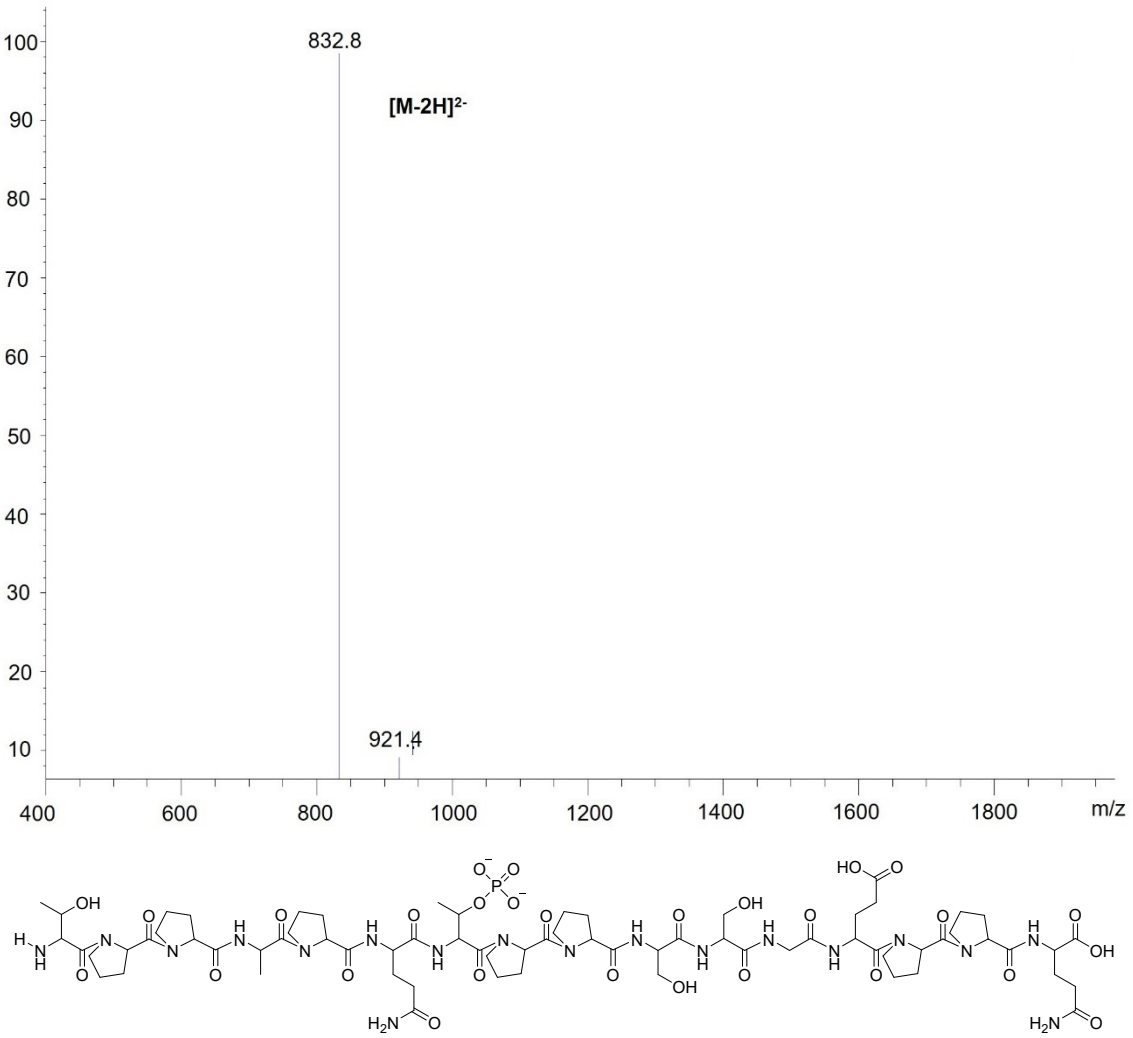

Figure S62. Mass spectrometry of the peptide TPPAPQ(T-p)PPSSGEPPQ (termed QT-p). m/z: 1667.66 (100.0%). MS (ESI) m/z: $[M-2H]^{2-}$ calcd, 832.83; found, 832.80.

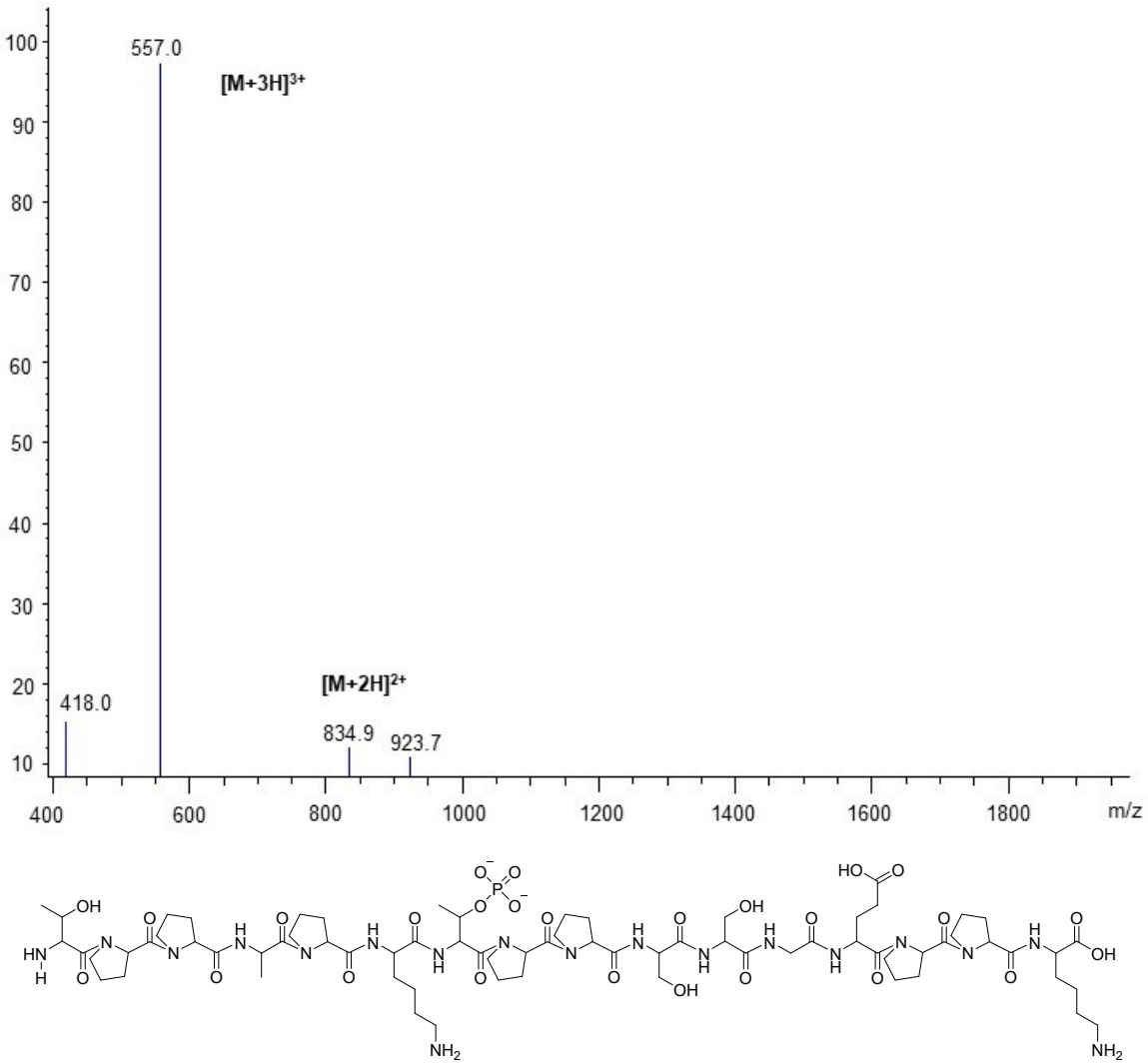

Figure S63. Mass spectrometry of the peptide TPPAPK(T-p)PPSSGEPPK (termed KT-p). m/z: 1667.75 (100.0%). MS (ESI) m/z: $[M+2H]^{2+}$ calcd, 834.88; found, 834.90. $[M+3H]^{3+}$ calcd, 556.92; found, 557.00.

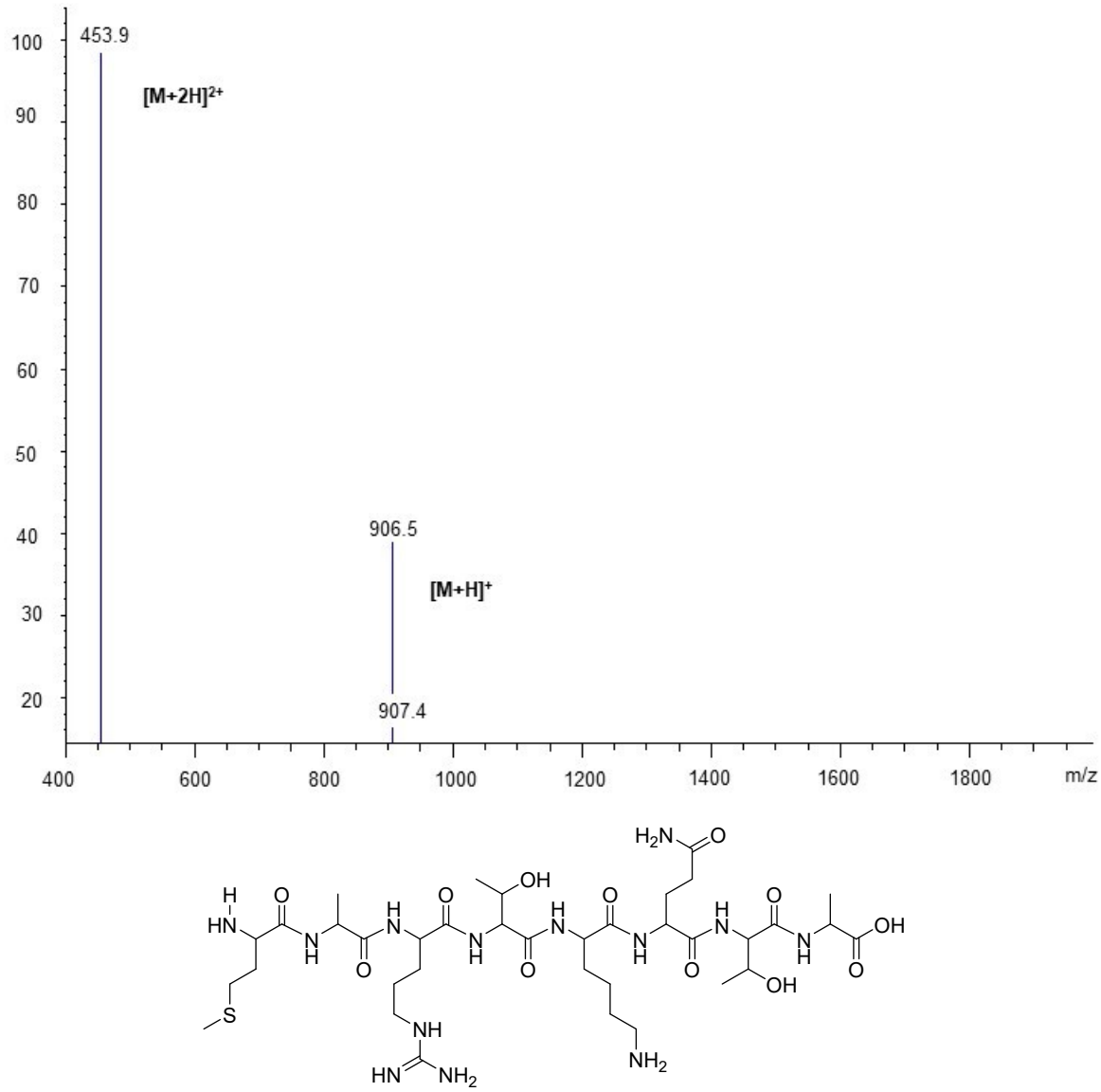

Figure S64. Mass spectrometry of the peptide MARTKQTA (termed MA). m/z: 906.06 (100.0%). MS (ESI) m/z: $[M+H]^+$ calcd, 907.50; found, 906.50. $[M+2H]^{2+}$ calcd, 454.03; found, 453.90.

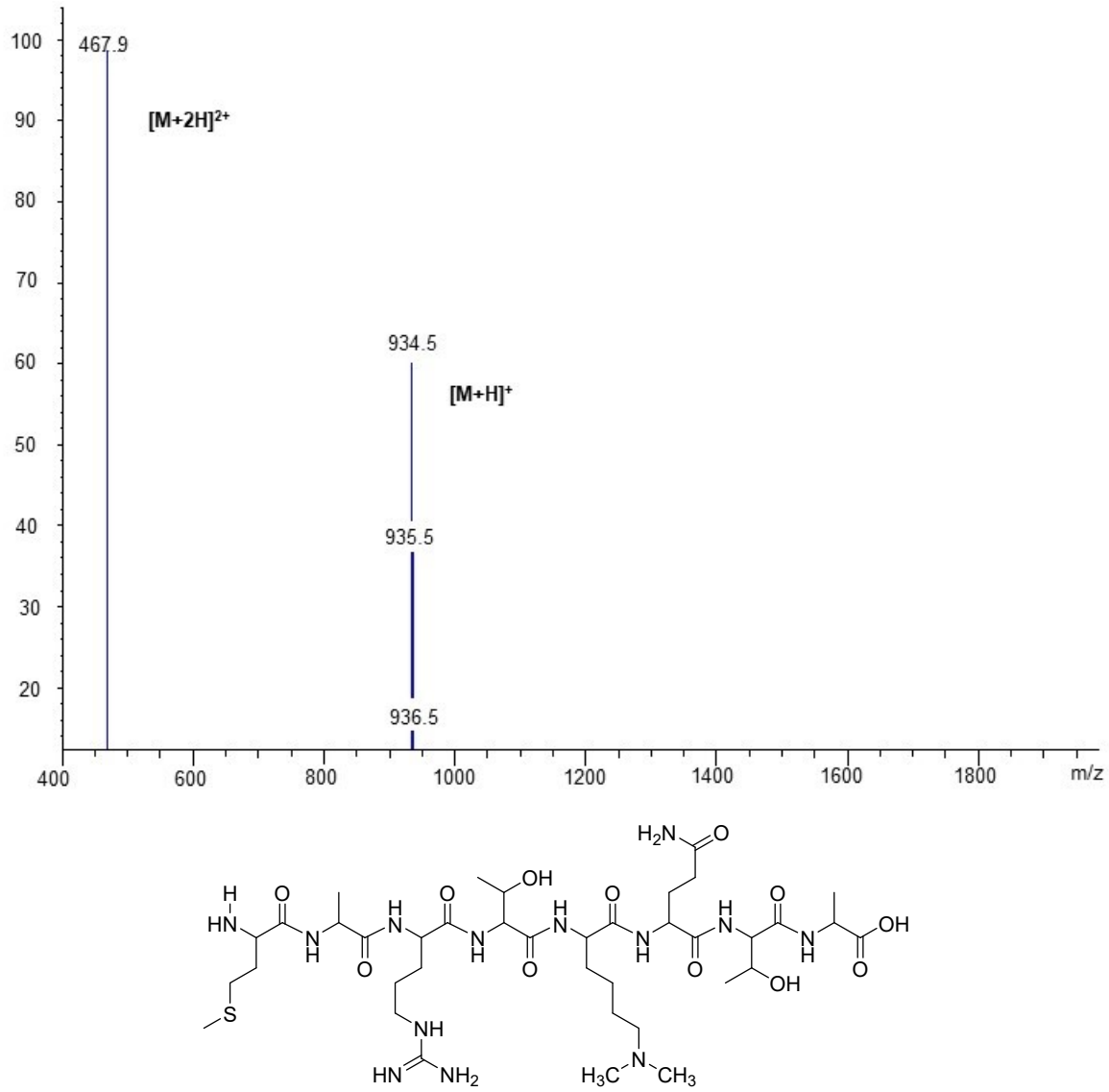

Figure S65. Mass spectrometry of the peptide MART(K-Me)QTA (termed MA-Me). m/z: 934.06 (100.0%). MS (ESI) m/z: $[M+H]^+$ calcd, 935.06; found, 934.50. $[M+2H]^{2+}$ calcd, 468.03; found, 467.90.

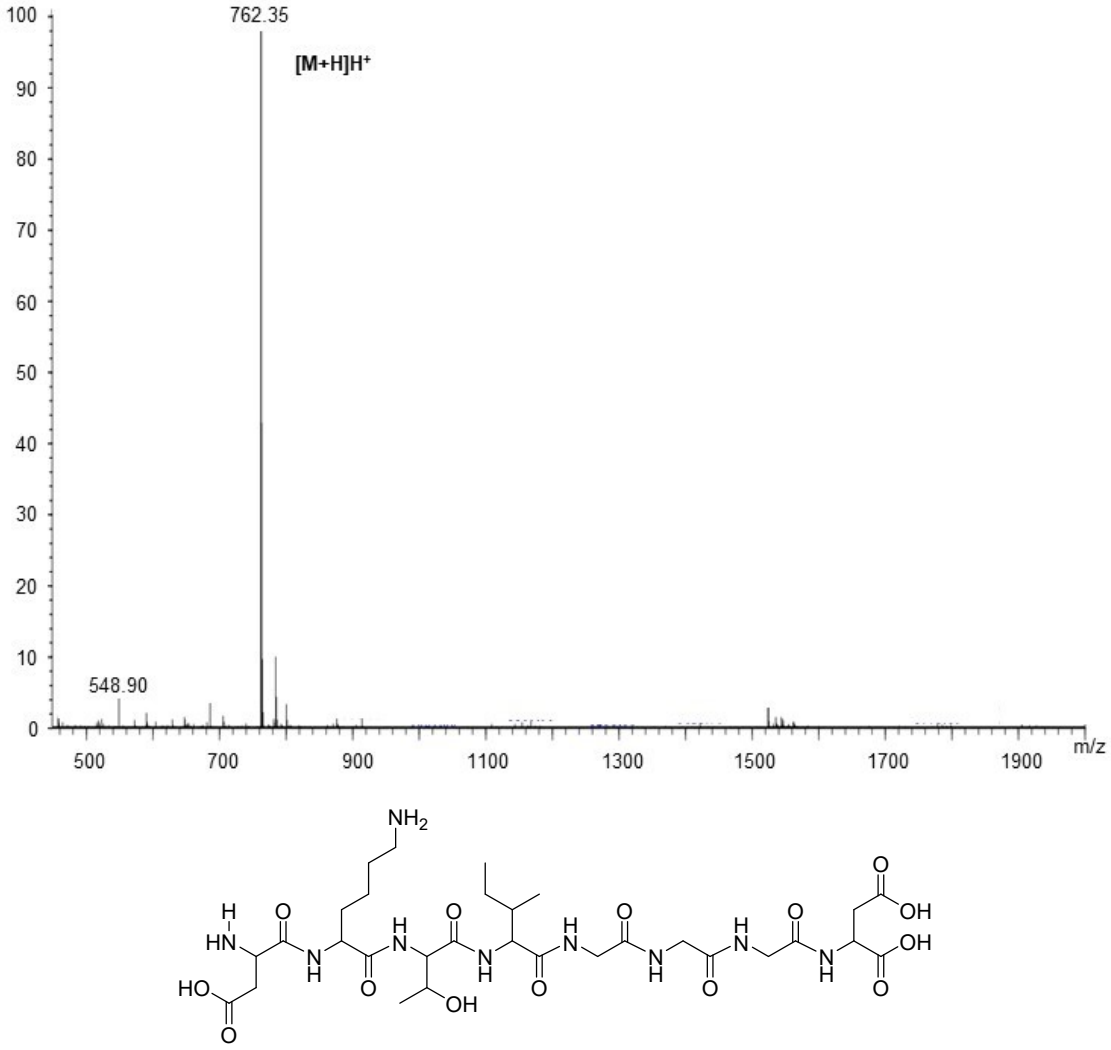

Figure S66. Mass spectrometry of the peptide DKTIGGGD (termed DK). m/z: 761.77 (100.0%). MS (ESI) m/z: $[M+H]^+$ calcd, 762.77; found, 762.35.

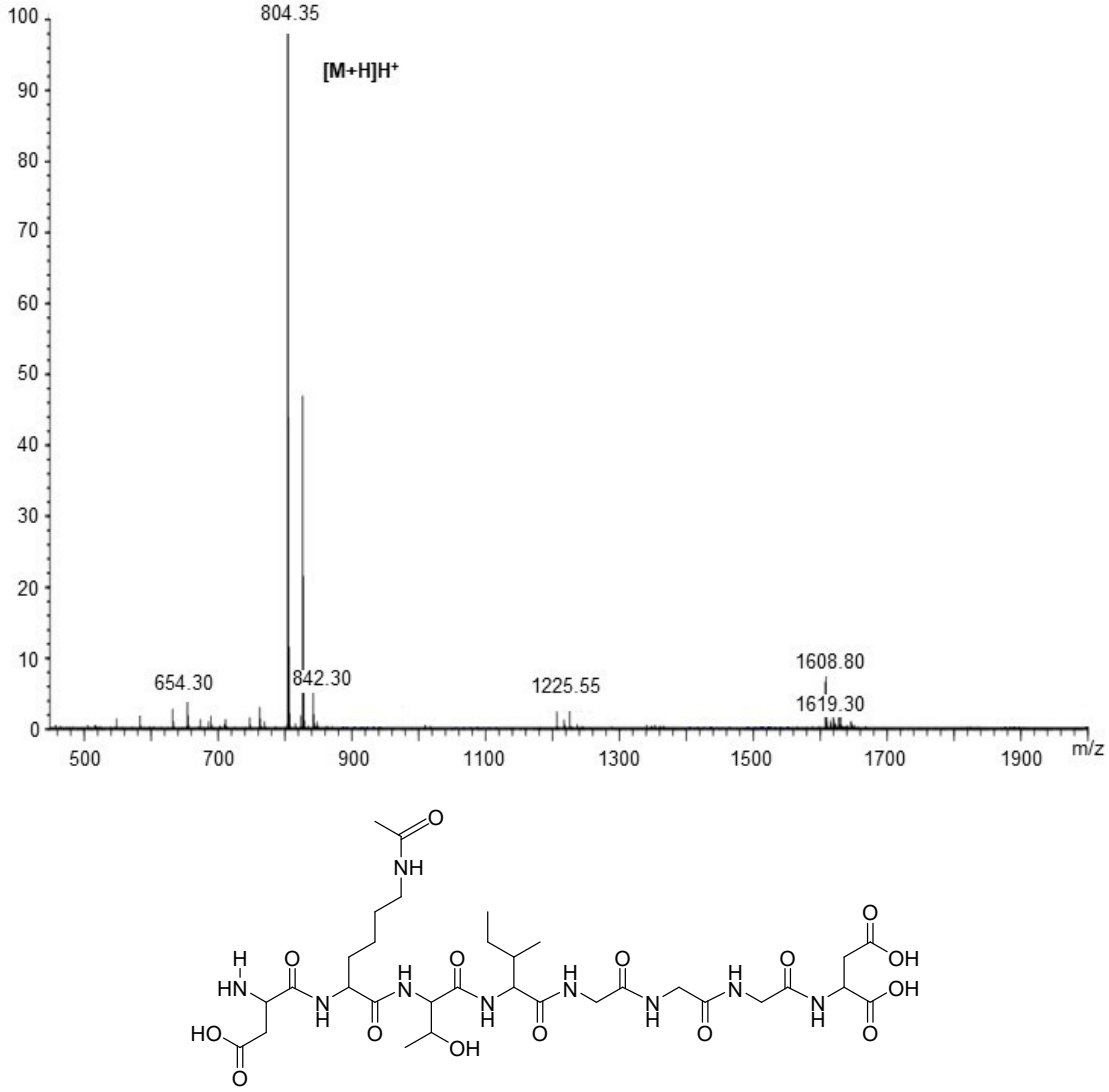

Figure S67. Mass spectrometry of the peptide D(K-Ac)TIGGGD (termed DK-Ac). m/z: 803.81 (100.0%). MS (ESI) m/z: $[M+H]^+$ calcd, 804.81; found, 804.35.

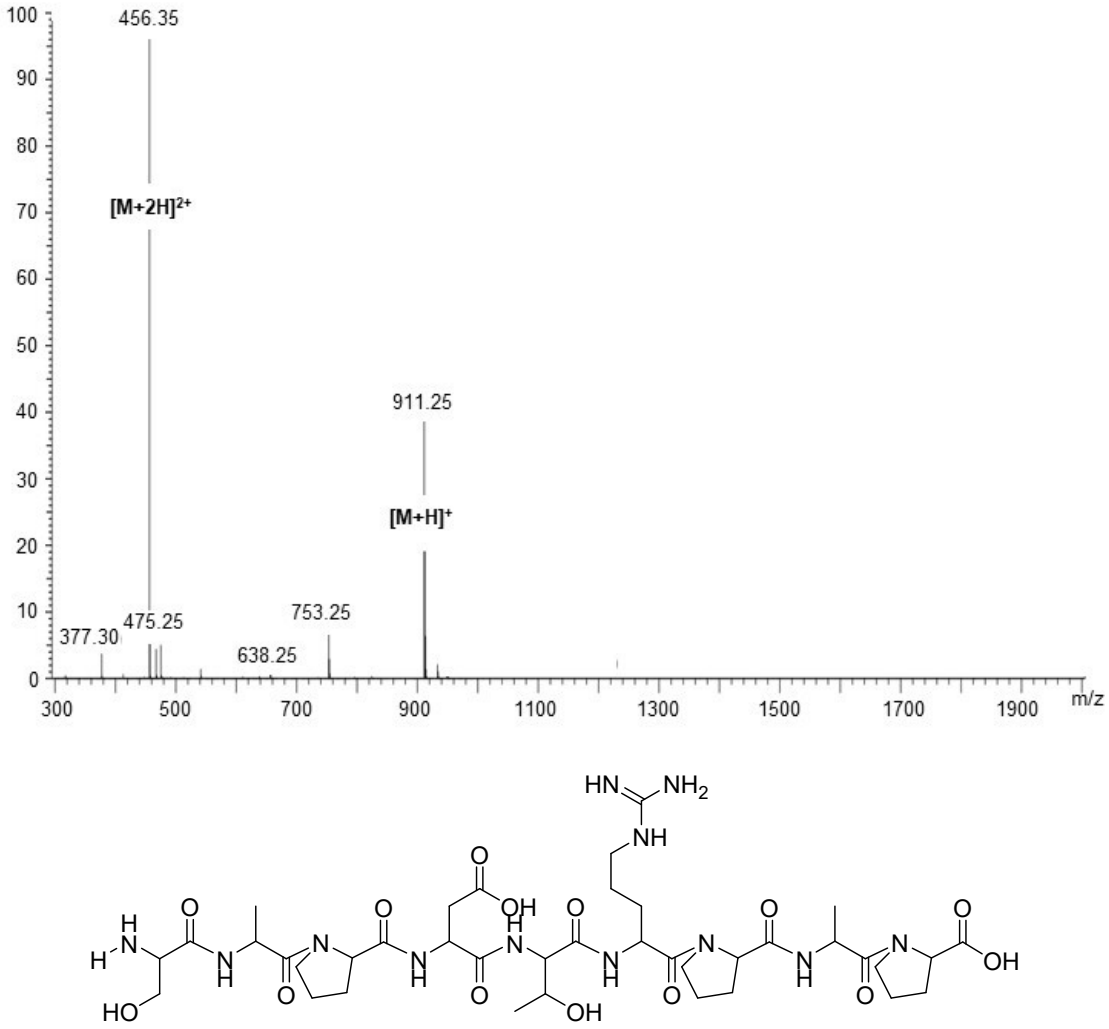

Figure S68. Mass spectrometry of the peptide SAPDTRPAP (termed ST). m/z: 910.95 (100.0%). MS (ESI) m/z: $[M+H]^+$ calcd, 911.95; found, 911.25. $[M+2H]^{2+}$ calcd, 456.48; found, 456.35.

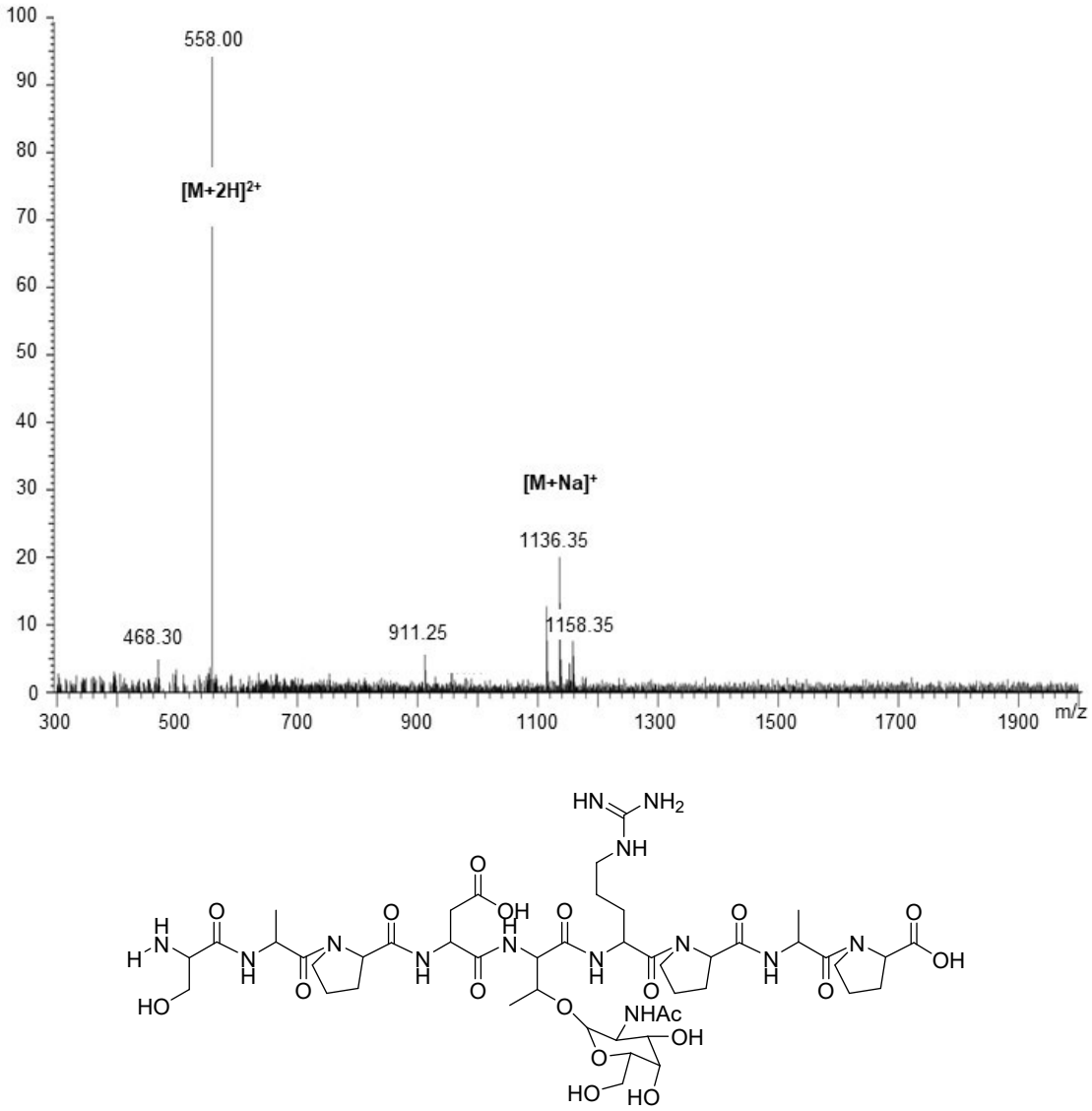

Figure S69. Mass spectrometry of the peptide SAPD(T-α-D-GalNAc)RPAP (termed ST- GalNAc). m/z: 1114.21 (100.0%). MS (ESI) m/z: $[M+Na]^+$ calcd, 1137.20; found, 1136.35. $[M+2H]^{2+}$ calcd, 558.11; found, 558.00.

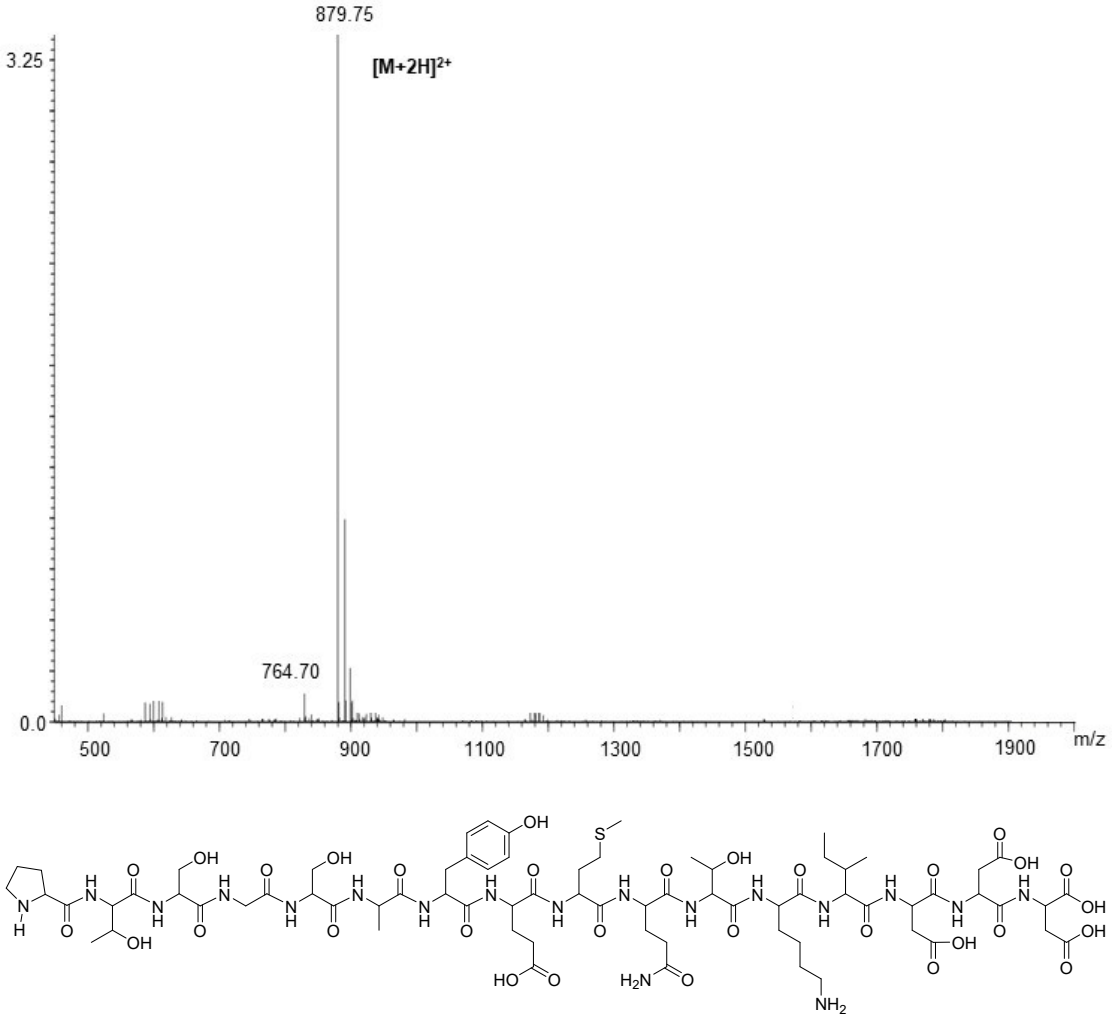

Figure S70. Mass spectrometry of the peptide PTSGSAYEMQTKIDDD (termed PD). m/z: 1757.82 (100.0%). MS (ESI) m/z: $[M+2H]^{2+}$ calcd, 879.91; found, 879.75.

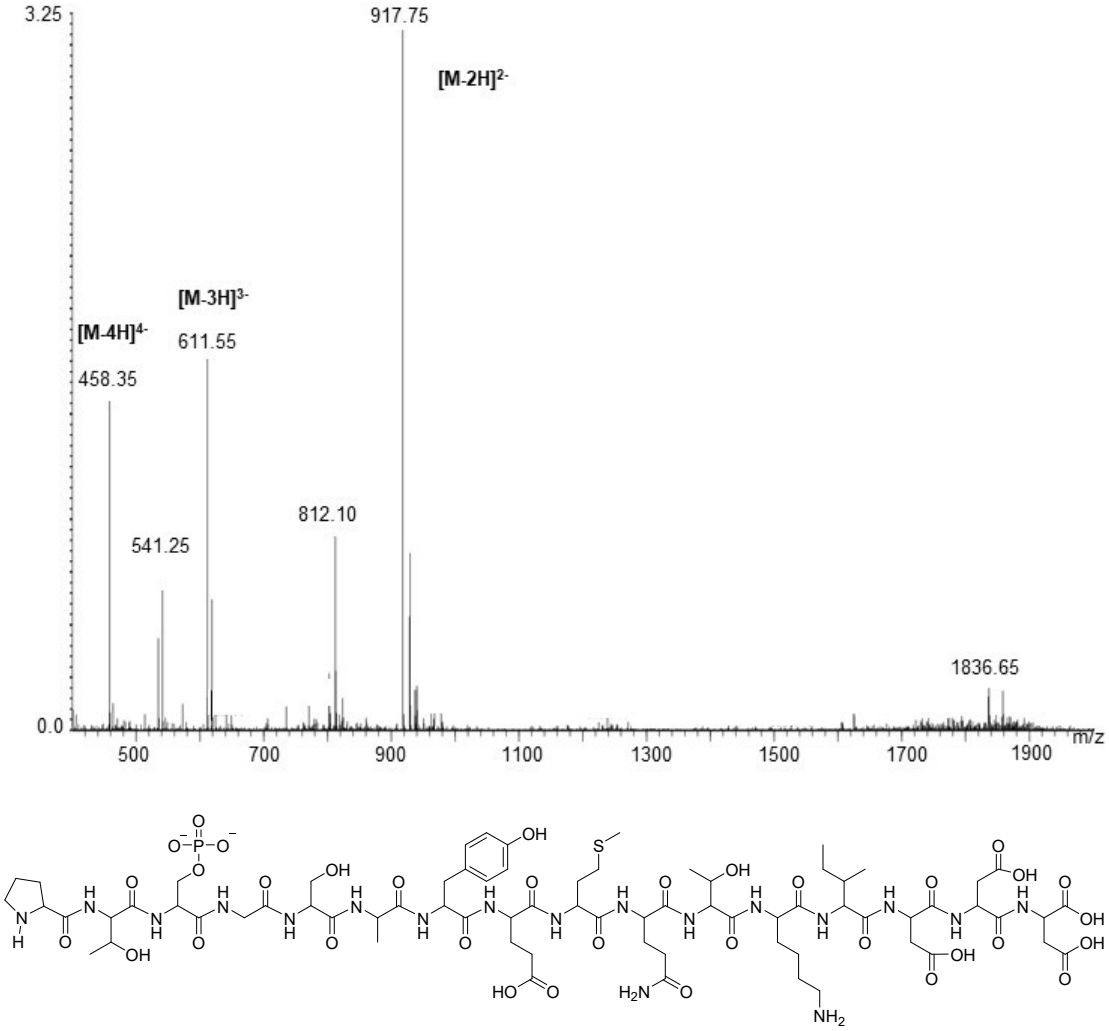

Figure S71. Mass spectrometry of the peptide PT(S-p)GSAYEMQTKIDDD (termed PD-p). m/z: 1837.80 (100.0%). MS (ESI) m/z: $[M-2H]^{2-}$ calcd, 917.90; found, 917.75. $[M-3H]^{3-}$ calcd, 611.60; found, 611.55. $[M-4H]^{4-}$ calcd, 458.45; found, 458.35.

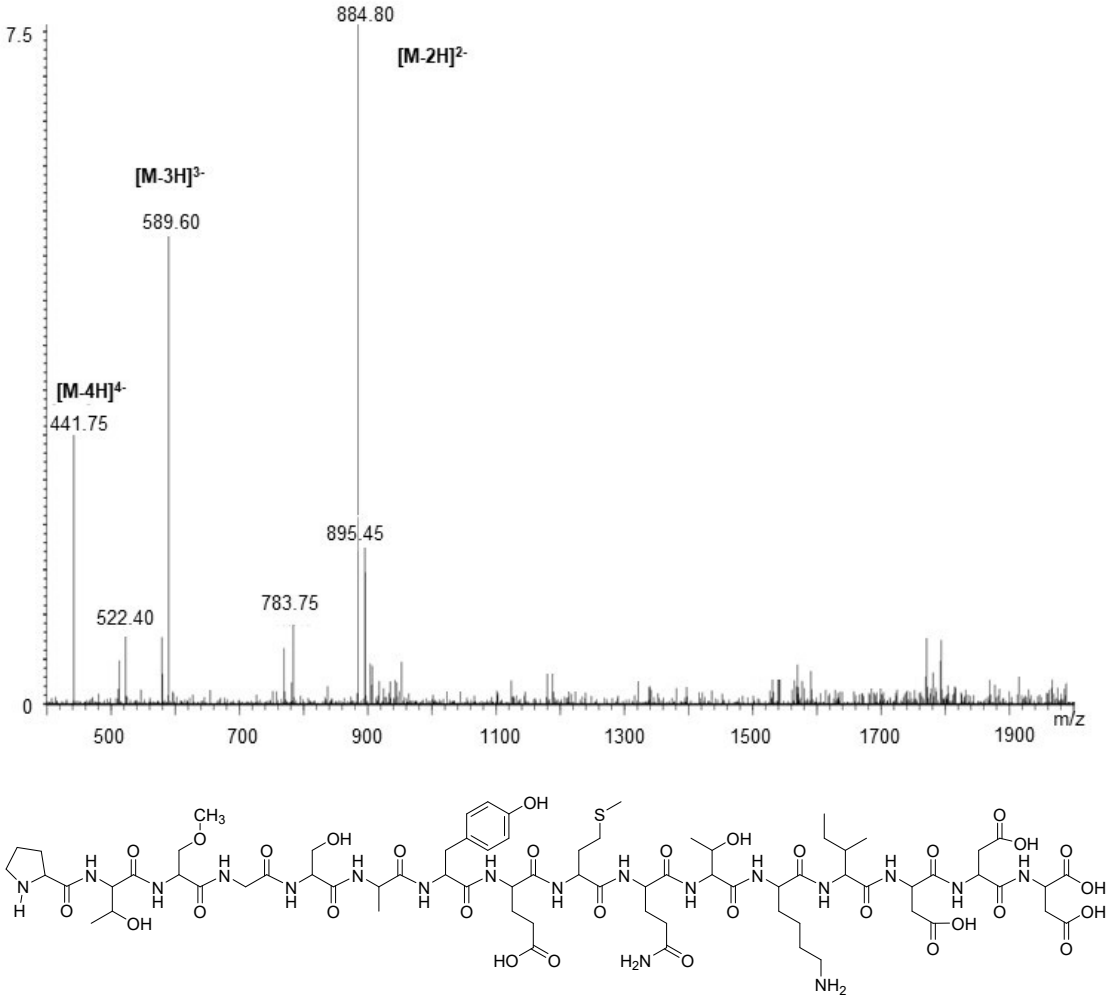

Figure S72. Mass spectrometry of the peptide PT(S-Me)GSAYEMQTKIDDD (termed PD-Me). m/z: 1771.85 (100.0%). MS (ESI) m/z: $[M-2H]^{2-}$ calcd, 884.93; found, 884.80. $[M-3H]^{3-}$ calcd, 589.62; found, 589.60. $[M-4H]^{4-}$ calcd, 441.96; found, 441.75.

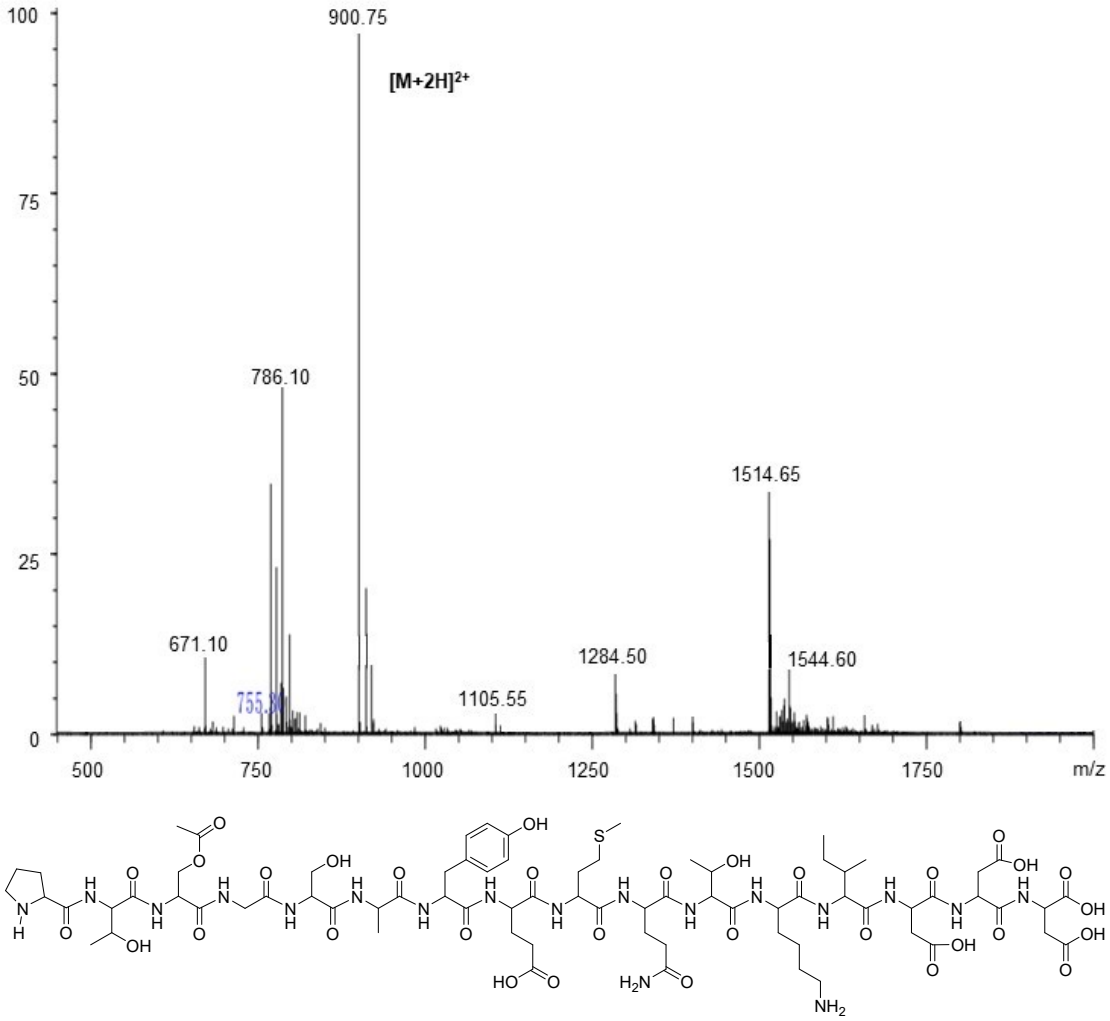

Figure S73. Mass spectrometry of the peptide PT(S-Ac)GSAYEMQTKIDDD (termed PD-Ac). m/z: 1799.86 (100.0%). MS (ESI) m/z: $[M+2H]^{2+}$ calcd, 900.93; found, 900.75.

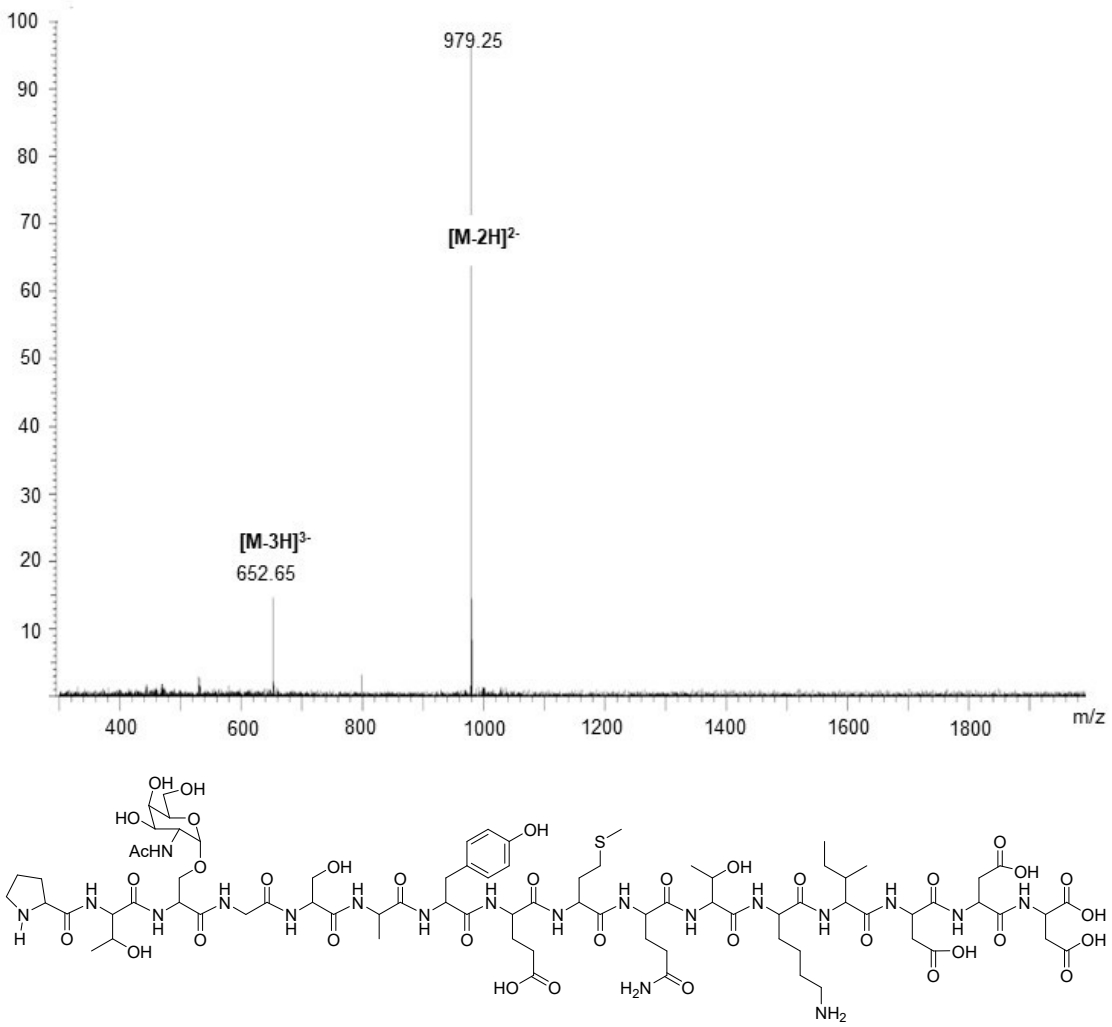

Figure S74. Mass spectrometry of the peptide PT(S-α-D-GalNAc)GSAYEMQTKIDDD (termed PD-GalNAc). m/z: 1961.02 (100.0%). MS (ESI) m/z: $[M-2H]^{2-}$ calcd, 979.51; found, 979.25. $[M-3H]^{3-}$ calcd, 652.67; found, 652.65.

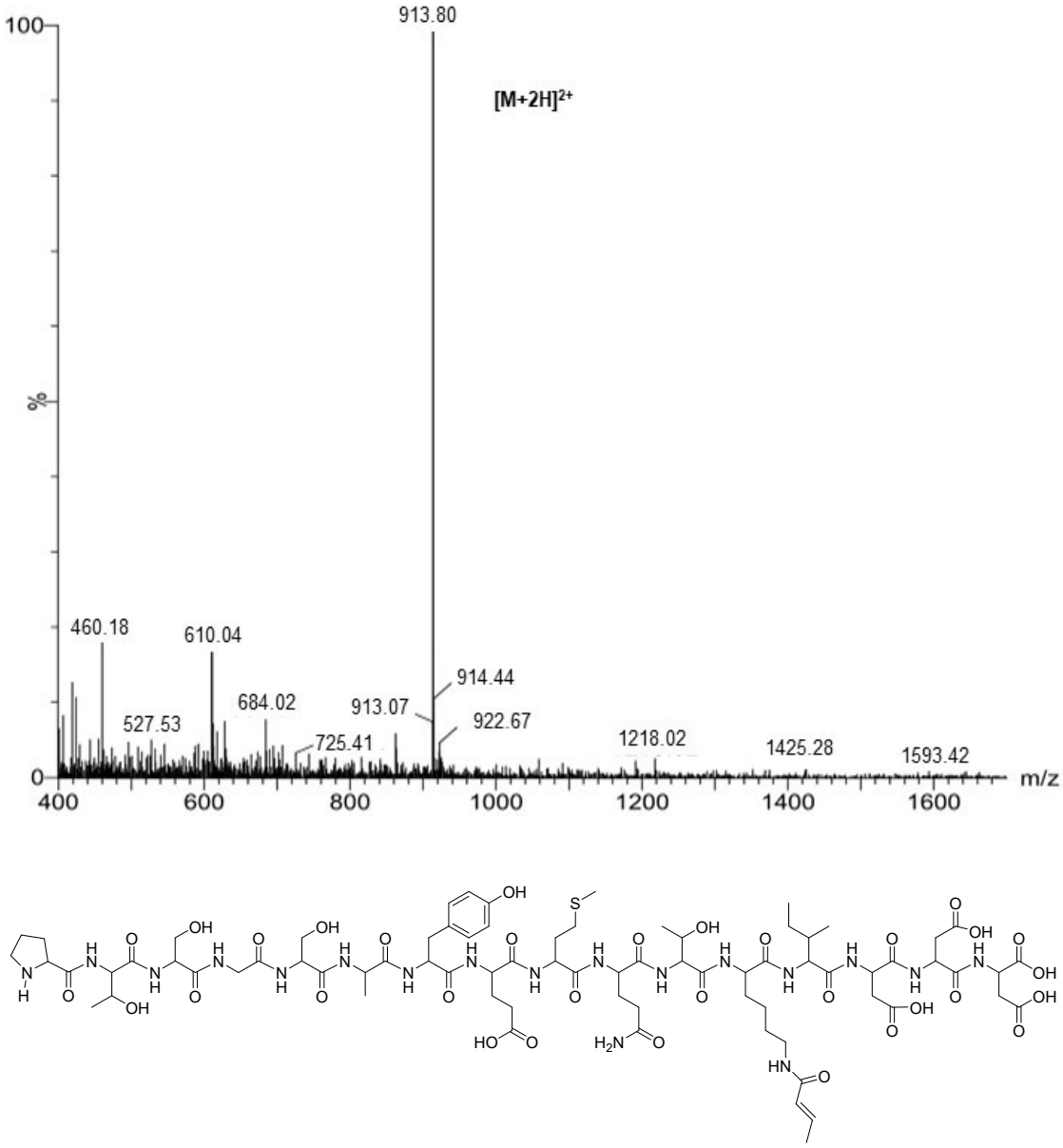

Figure S75. Mass spectrometry of the peptide PTSGSAYEMQT(K-cr)IDDD (termed PD-cr). m/z: 1825.90 (100.0%). MS (ESI) m/z: $[M+2H]^{2+}$ calcd, 913.95; found, 913.80.

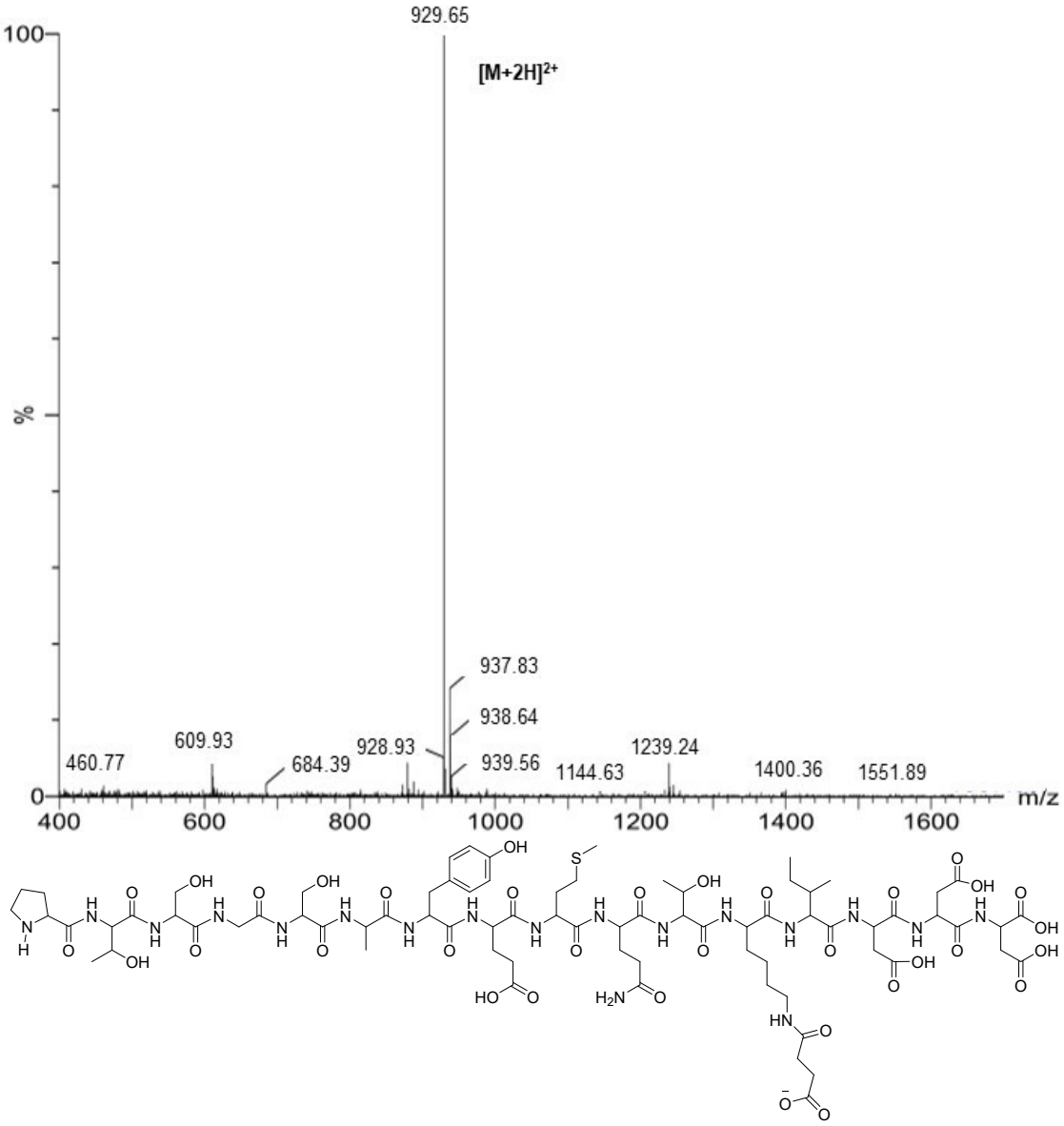

Figure S76. Mass spectrometry of the peptide PTSGSAYEMQT(K-succ)IDDD (termed PD-succ). m/z: 1857.9 (100.0%). MS (ESI) m/z: $[M+2H]^{2+}$ calcd, 929.95; found, 929.65.

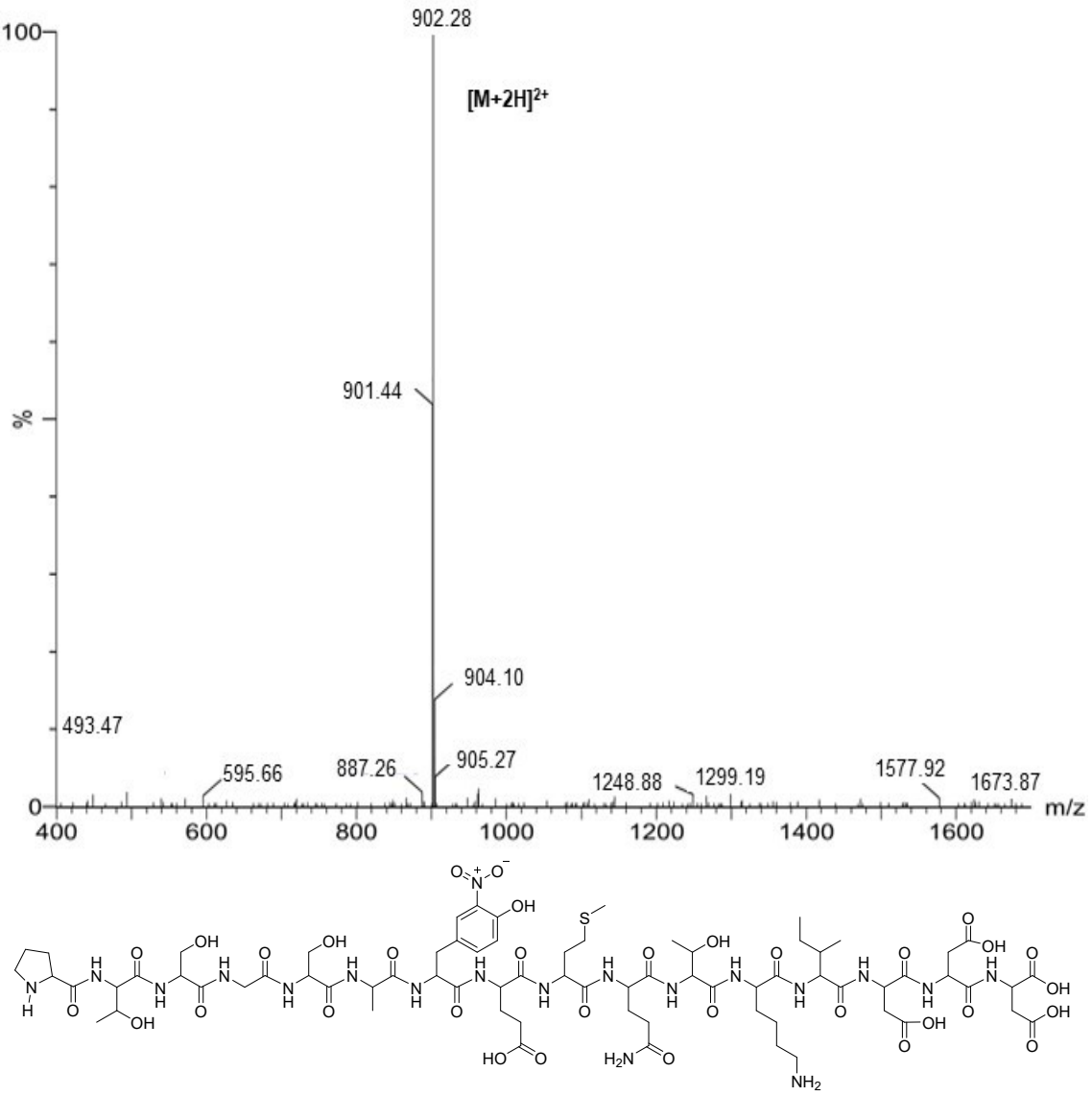

Figure S77. Mass spectrometry of the peptide PTSGSA(Y-n)EMQTKIDDD (termed PD-n). m/z: 1802.80 (100.0%). MS (ESI) m/z: $[M+2H]^{2+}$ calcd, 902.40; found, 902.28.

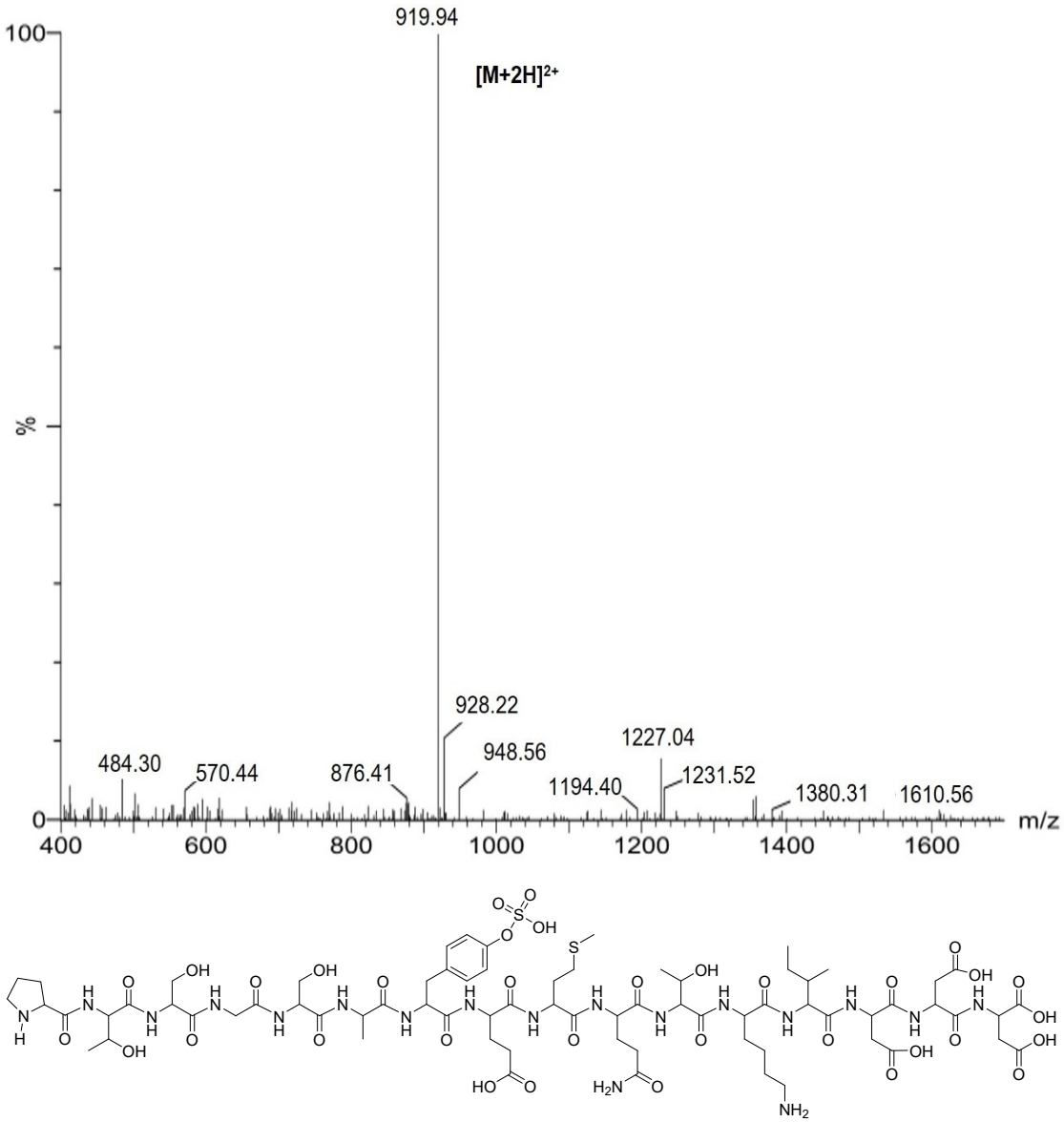

Figure S78. Mass spectrometry of the peptide PTSGSA(Y-s)EMQTKIDDD (termed PD-s). m/z: 1837.90 (100.0%). MS (ESI) m/z: $[M+2H]^{2+}$ calcd, 919.95; found, 919.94.

References

1. S. Wang, J. Cao, W. D. Jia, W. M. Guo, S. H. Yan, Y. Q. Wang, P. K. Zhang, H. Y. Chen, S. Huang, *Chem. Sci.* 2020, 11, 879-887.
2. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX* 2015, 1-2, 19-25.
3. M. Faller, M. Niederweis and G. E. Schulz, *Science* 2004, 303, 1189-1192.
4. J. Lee, X. Cheng, S. Jo, A. D. MacKerell, J. B. Klauda and W. Im, *J. Chem. Theory. Comput.* 2016, 12, 405-413.
5. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.* 2000, 28, 235-242.
6. J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell, Jr. and R. W. Pastor, *J. Phys. Chem. B*, 2010, 114, 7830-7843.
7. K. Hart, N. Foloppe, C. M. Baker, E. J. Denning, L. Nilsson and A. D. MacKerell, Jr., *J. Chem. Theory. Comput.* 2012, 8, 348-362.
8. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.* 1983, 79, 926-935.
9. N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, *J. Comput. Chem.* 2011, 32, 2319-2327.
10. B. R. Miller, III, T. D. McGee, Jr., J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *J. Chem. Theory. Comput.* 2012, 8, 3314-3321..
11. E. C. Meng, T. D. Goddard, E. F. Pettersen, G. S. Couch, Z. J. Pearson, J. H. Morris and T. E. Ferrin, *Protein Sci.* 2023, 32, e4792.
12. A. Aksimentiev and K. Schulten, *Biophys. J.* 2005,88, 3745-3761.