Supplementary Information for

Sub-dalton Discrimination of Citrullination by a Semi-rationally Designed Aerolysin Nanopore

Xinjia Zhao,[1§] Jiarong Zhang,[1,3§] Juanjuan Zhao,[2§] Wei Yong,[1] Zeyuan Zhao,[2] Minmin Li,[1] Yuchen Cao,[1,3] Wenjing Sun,[1] Jing Wang,[1,3] Dongdong Wang,[1] Jiaqi Li,[1] Kouxi Xu,[4] Yangyang Bian,[2] Guangyan Qing[1]**

1. State Key Laboratory of Phytochemistry and Natural Medicines, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, P. R. China

2. Key Laboratory of Resource Biology and Modern Biotechnology in Western China, College of Life Science, Northwest University, Xi'an 710069, P. R. China

3. University of Chinese Academy of Sciences, Beijing 100049, P. R. China

4. College of Chemistry and Molecular Sciences, Henan University, Kaifeng 475000, P. R. China

§ *These authors contribute equally to this work.*

*E-mail: bianyy@nwu.edu.cn

qinggy@dicp.ac.cn

Table of Contents

1. PIToN model

1.1 Semi-rational theoretical model based on the first principles of physical chemistry — the PIToN model

1.1.1 Conceptual formulation of PIToN model

PIToN is introduced here as a physically motivated, semi-quantitative reduced model for organizing how multiple analyte–pore interaction descriptors may contribute to measurable nanopore responses. Rather than serving as a universal predictive model, the model is used in the present study as a comparative tool for mutation prioritization within the tested aerolysin system.

We propose that the total sensing response (S) generated when a molecule translocates through a nanopore (typically correlated with the blockade current or dwell time), as a weighted linear combination of signal contributions from multiple physical interactions, with its core expression formulated as:

$$S = \sum w_i \cdot \Psi_i$$

Here, Ψ_i represents the contribution of the i-th interaction, and w_i denotes its corresponding weighting factor, which is dependent on the specific nanopore structure and experimental conditions. Taken possible interactions into accounts, the formula could be:

$$S = w_1 \cdot \Psi_{Steric} + w_2 \cdot \Psi_{Bi} + w_3 \cdot \Psi_{VDW} + w_4 \cdot \Psi_{Electrostatic} + w_5 \cdot \Psi_{Hydrophobic} + w_6 \cdot \Psi_{Electrophoreti} + w_7 \cdot \Psi_{Electroosmotic}$$

In this formulation, the seven terms represent steric exclusion, local residue–analyte interaction propensity, van der Waals (VDW) interaction, electrostatic interaction,

hydrophobic contribution, electrophoretic contribution, and electroosmotic contribution, respectively.

In the present implementation, equal weighting was adopted as a transparent first-order approximation to avoid over-parameterization and to preserve model interpretability, given the limited number of pore variants examined in this study. We do not claim that the physically optimal weighting scheme has been established here. Instead, equal weighting is used as an operational starting point for comparative mutation screening, while more data-rich calibration across multiple pores, analytes, and conditions would be required before mutation-specific or system-specific weighting schemes could be meaningfully inferred.

In addition, the present implementation adopts a reduced local approximation. We focused on K238 as an engineering-relevant position within the constriction region, while recognizing that analyte discrimination in aerolysin likely reflects multiple simultaneous interactions across a broader local sensing zone. In this approximation, a single protomer and a local residue-centered analyte interaction are used to represent one dominant component of the broader pore environment, rather than the full many-body interaction network of the heptameric pore.

1.1.2 Parameterization of the seven descriptors using WT nanopore as an example

To illustrate how the PIToN descriptors were parameterized in the present aerolysin system, the WT nanopore is used here as a reference example. In this implementation, each descriptor is assigned an operational expression intended to capture one dominant

aspect of analyte – pore interaction, rather than a complete many-body physical treatment. The seven terms are parameterized below in the same order as in Eq. (2), using simplified local assumptions chosen for interpretability and tractability within the present mutation-screening framework.

First term, steric hindrance. When discriminating between modified molecules or other similar species using the same nanopore, this factor is less dependent on the nanopore diameter. However, when comparing different mutant nanopores, the nanopore diameter significantly influences the ionic current response: smaller nanopores yield a larger response, resulting in better spatial resolution. The introduction of chemical groups by PTMs, such as phosphate moieties or glycan chains, directly alters the molecule's effective volume and contour, thereby modulating the extent of ionic current blockade. The steric contribution can be quantitatively expressed as:

$$\Psi_{Steric} = k_s \cdot \Delta V_{eff}$$

where k_s is a proportionality constant related to the geometry of the nanopore, ΔV_{eff} is the change in the effective excluded volume of the molecule at the narrowest constriction of the pore. For a PTM, this corresponds to the volumetric difference between the modified and unmodified states, i.e., $V_{modified} - V_{unmodified}$. In the case of citrullination, the associated mass change is extremely small (~0.984 Da). As a result, the contribution of the steric term, denoted by the weighting coefficient w_1, is likely minimal. This subtlety helps explain why citrullination is challenging to detect and motivates the exploration of interaction modes beyond steric-dominated sensing. In this formulation, ΔV_{eff} captures one geometric contribution to analyte-induced current

modulation, but does not explicitly account for analyte conformation, orientation, or local ion and water organization within the pore.[1,2]

To quantify this effect, we estimated the molecular volumes of arginine and citrulline based on the Van der Waals radii of their constituent elements: r_vdw(C)=1.70 Å, r_vdw(N)=1.55 Å, r_vdw(O)=1.52 Å and r_vdw(H)=1.20 Å. Treating atoms as hard spheres, the volume was calculated as $\frac{4}{3}\pi r^3$. This yields $V_{Arg} = 316.6$ Å^3 for arginine (C6H14N4O2) and $V_{Cit} = 308.5$ Å^3 for citrulline (C6H13N3O3), corresponding to a volume change of −2.57%, which is quite small so that the steric hindrance is ignorable. Normalized to a range of 0 to 500 Å^3 yields:

$$\Psi_{Steric\_Norm}(WT\_Arg) = 0.633$$

$$\Psi_{Steric\_Norm}(WT\_Cit) = 0.617$$

Such a minor shift confirms that steric hindrance plays a negligible role in this context. In comparison, volume calculations performed using the published software MoloVol[3] gave $V_{Arg} = 185.1$ Å^3 and $V_{Cit} = 175.2$ Å^3, corresponding to a –5.35% change, approximately twice that of the simple hard-sphere estimate. While the absolute values differ between methods, both consistently indicate a small volumetric perturbation. Although the MoloVol-derived decrease appears more pronounced, this estimation approach remains broadly applicable and informative for assessing steric contributions in other PTM systems.

Second term, binding-related descriptor. It is approximated using DFT-derived local residue–analyte interaction energies from simplified Gaussian calculations. Although quantum chemically derived binding energies encompass both Van der Waals

and electrostatic contributions, methodological variations exist in their computational treatment. The composite binding energy metric cannot delineate the specific contributions of distinct physical interactions, which may exhibit differential weighting coefficients, particularly relevant for quantifying the separate mechanistic roles of steric packing and electrostatics in citrullination and mutant nanopores. The binding energies for WT nanopore was calculated and subsequently normalized according to the following rule:

- Repulsive interactions ($\Psi_{Binding} > 0$) are mapped to 0 ($\Psi_{Binding\_Norm} = 0$)
- Attractive interactions ($\Psi_{Binding} < 0$) are scaled as $\Psi_{Binding\_Norm} = -\Psi_{Binding}/500$

$$\Psi_{Binding}(WT\_Arg) = 17.6\ kJ/mol$$

$$\Psi_{Binding}(WT\_Cit) = -157.4\ kJ/mol$$

Thus,

$$\Psi_{Binding\_Norm}(WT\_Arg) = 0$$

$$\Psi_{Binding\_Norm}(WT\_Cit) = 0.3148$$

The normalized DFT-derived quantity is used here only as a relative descriptor of local residue-specific interaction preference. Within the present reduced framework, a larger difference in this local descriptor may contribute to stronger signal contrast.

Third term, VDW interaction. Its magnitude depends strongly on the distance between the interacting parties and the atomic polarizability. This short-range interaction between the molecule and the pore wall is commonly described by the Lennard-Jones potential:

$$\Psi_{VDW} = \sum_{i,j} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

Where ε_{ij} and σ_{ij} are Lennard-Jones potential parameters describing the interaction strength and equilibrium distance for atom pair i-j (analyte atom i and pore atom j). r_{ij} is instantaneous distance between atoms i and j. When the engineered nanopore (e.g., K238W) positions the indole ring of tryptophan in close proximity the analyte, differences in contact distance and polarizability between the arginine and citrulline side chains will contribute to the recognition signal through the VDW term.

The VDW interactions were estimated based on the Lennard-Jones potential with parameters derived from the CHARMM36 force field,[4] as summarized in Table S1.

Table S1. CHARMM36 Lennard-Jones parameters.

Atom	σ(Å)	ε(kj/mol)
C	3.564	0.4602
N	3.296	0.8368
O	3.029	0.502
S	3.564	1.88
H	0.400	0.192

The calculation considered only the key residue at position 238 of the nanopore and the side chains of the target arginine and citrulline. The Lorentz-Berthelot combination rules were applied to determine interaction parameters between different atom types:

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2}$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \times \varepsilon_j}$$

The resulting atomic pair parameters are summarized in Table S2.

Table S2. CHARMM36 Lennard-Jones pair parameters.

Atoms	σ(Å)	ε(kj/mol)
N-N	3.296	0.8368
N-O	3.1625	0.648
O-O	3.029	0.502
O-C	3.2965	0.480
C-C	3.564	0.4602
H-O	1.7145	0.310
H-N	1.848	0.401
C-N	3.430	0.6204
H-C	1.982	0.2972
S-N	3.43	1.254
S-O	3.2965	0.9715
S-C	3.564	0.9302
S-H	1.982	0.601

The lysine residue at the WT 238 site contains 4 carbon atoms and 1 nitrogen atom. The arginine side chain includes 3 nitrogen atoms in the guanidinium group, 4 carbon atoms in the alkyl chain, and several hydrogen atoms. Assuming an interaction distance of $r = 0.95\sigma$, the VDW interaction energy was calculated as:

$$\Psi_{VDW} = 4\varepsilon\left[\left(\frac{1}{0.95}\right)^{12} - \left(\frac{1}{0.95}\right)^{6}\right] = 1.96\varepsilon$$

For the interaction between WT and arginine, which involves 3 N–N pairs, 2 C–C pairs, and 2 H–N pairs, the total energy was computed as:

$$\Psi_{VDW}(WT\_Arg) = 3 \times 1.96 \times 0.8368 + 2 \times 1.96 \times 0.4602 + 2 \times 1.96 \times 0.401$$

$$= 8.295 \text{ kJ/mol}$$

For the interaction between WT and citrulline, which includes 2 N–N pairs, 1 N–O pair, 2 C–C pairs, and 1 H–O pair, the total energy was:

$$\Psi_{VDW}(WT\_Cit)$$

$$= 2 \times 1.96 \times 0.8368 + 1 \times 1.96 \times 0.648 + 2 \times 1.96 \times 1.804$$

$$+ 1 \times 1.96 \times 0.310 = 6.961\ kJ/mol$$

These values were normalized to a range of [0, 15] kJ/mol, yielding:

$$\Psi_{VDW\_Norm}(WT\_Arg) = 0.553,\ \Psi_{VDW\_Norm}(WT\_Cit) = 0.464$$

Fourth term, electrostatic interaction. The interaction energy is given by Coulomb's law:

$$\Psi_{Electrostatic} = \frac{1}{4\pi\varepsilon_0\varepsilon_r}\sum_{i,j}\frac{q_i Q_j}{r_{ij}}$$

Where ε_0 and ε_r are vacuum permittivity and relative dielectric constant. q_i and Q_j are partial charges on analyte atom i and the pore residue atom j.

Using specific physical parameters (ε_0 = 8.854×10^{-12} $C^2/J\cdot m$, e = 1.602×10^{-19} C, interaction distance r = 5 Å, based on aerolysin channel structure, and ε_r = 30 typical for protein interiors), we calculated the electrostatic interactions for WT nanopore. The resulting $\Psi_{Electrostatic}$ values were normalized to [0, 1] scale using a sigmoid function,

$$\Psi_{norm} = 0.5 + 0.5 \times \tanh\left(-\frac{\Psi_{Electrostatic}}{40}\right)$$

which maps $\Psi_{Electrostatic} = 0$ to a neutral value of 0.5. Strong repulsive interactions ($\Psi_{Electrostatic} \to +\infty$) approach 0, while strong attractive interactions ($\Psi_{Electrostatic} \to -\infty$) approach 1.0. Applying this method:

$$\Psi_{Electrostatic}(WT\_Arg) = \frac{1}{4\pi \times 8.854 \times 10^{-12} \times 30} \times \frac{(+1e) \times (+1e)}{5 \times 10^{-1}} = +46.2\ kJ/mol$$

$$\Psi_{Electrostatic}(WT\_Cit) = 0\ kJ/mol$$

normalized to [0, 1] yields,

$$\Psi_{Electrostatic\_Norm}(WT\_Arg) = 0.09$$

$$\Psi_{Electrostatic\_Norm}(WT\_Cit) = 0.5$$

The WT pore demonstrates pronounced electrostatic repulsion against arginine (Ψ = 0.09), while maintaining electrostatic neutrality toward citrulline (Ψ = 0.5), yielding a substantial contrast ($\Delta\Psi_{WT}$ = 0.41) in electrostatic behavior. Despite this marked difference, electrostatic interactions contribute minimally to citrullination detection in the WT nanopore: the strong repulsion toward arginine leads to excessively short dwell times that compromise detection sensitivity, whereas the neutral interaction with citrulline fails to generate a discernible signal. This inherent inadequacy underscores the need for sensing mechanisms that exploit non-electrostatic interactions or involve strategic pore redesign to effectively leverage electrostatic contributions.

Fifth term, hydrophobic interaction. This interaction influences both the positioning and conformational dynamics of molecules within the nanopore. While challenging to express with a simple analytical equation, this interaction is generally proportional to the change in solvent-accessible surface area ($SASA$):

$$\Psi_{Hydrophobic} = \gamma \cdot SASA$$

Here, γ represents the free energy coefficient per unit area, typically ranging between 0.1–0.2 kJ/mol/Å^2, while $SASA$ corresponds to the solvent-accessible surface area of hydrophobic regions upon formation of the molecule-pore complex. PTMs can

modify molecular hydrophobicity, consequently affecting this term. The guanidino headgroup of arginine is highly hydrophilic, while citrullination modestly increases the hydrophobicity of the side chain. This subtle alteration in hydrophobicity could affect the peptide's dwell time and spatial configuration when interacting with hydrophobic regions of the pore (such as those formed by phenylalanine, leucine, or other hydrophobic residues), thereby contributing to the sensing signal through the hydrophobic interaction term.

To quantify this effect, we assumed hydrophobic surface areas of $SASA_{hydrophobic}(Arg) =$ 85 Å^2 for arginine and $SASA_{hydrophobic}(Cit) =$ 88 Å^2 for citrulline, reflecting a $\Delta SASA$ of +3 Å^2 (+3.5%). This increase aligns with reported trends of enhanced hydrophobicity following citrullination, attributed to reduced hydrogen bonding with water upon loss of the charged guanidinium group.

Using a moderate γ value of 0.15 kJ/mol/Å^2, the hydrophobic interaction energies were calculated as:

$$\Psi_{Hydrophobic}(WT\_Arg) = 0.15 \times 85 = 12.75\ kJ/mol$$

$$\Psi_{Hydrophobic}(WT\_Cit) = 0.15 \times 88 = 13.20\ kJ/mol$$

Normalizing with a 0-15 kJ/mol reference range yielded:

$$\Psi_{Hydrophobic\_Norm}(WT\_Arg) = 0.638$$

$$\Psi_{Hydrophobic\_Norm}(WT\_Cit) = 0.660$$

These estimates are qualitatively consistent with the idea that citrullination may modestly alter hydrophobic character. Although the ureido group retains some polarity, its neutral nature, compared to the positively charged guanidinium group, reduces

hydrogen bonding with water, resulting in a net increase in hydrophobic character. While this hydrophobic enhancement is modest relative to electrostatic contributions, it provides valuable supplementary discrimination capability within the multi-parameter sensing framework.

Sixth term, electrophoretic contribution. The electrophoretic force is defined as:

$$F_{electrophoretic} = q \cdot E$$

Where q is the net charge of the analyte. And E is the electric field strength within the nanopore. $E = {V_{bias}}/{L}$, where V_{bias} is the applied bias voltage and L is the effective pore length. While not directly contributing to the sensing signal (S), the electrophoretic force governs the analyte's translational velocity and dwell time within the pore via Newton's second law ($F = m \cdot a$). These kinetic parameters (velocity and dwell time) serve as crucial characteristic features of the ultimate sensing signal S. In nanopore study, this force functions as a driver rather than a direct sensing interaction. The charge difference between arginine and citrulline results in distinct electrophoretic forces, which directly determine their translocation rates through the pore, affecting both capture efficiency and dwell time.

To quantify the electrophoretic energy contribution, we express the interaction term as:

$$\Psi_{Electrophor} = q_{eff} \times V_{bias}$$

Using an applied voltage of 0.1 V, typical for aerolysin nanopore experiments:

$$\Psi_{Electrophoretic}(WT\_Arg)$$

$$= (+1) \times (1.602 \times 10^{-19} C) \times 0.1\, V \times \frac{(6.022 \times 10^{23} mol^{-1})}{1000}$$

$$= 9.65\, kJ/mol$$

$$\Psi_{Electrophoretic}(WT\_Cit) = 0\, kJ/mol$$

Normalization within a 0—15 kJ/mol reference range, based on typical electrophoretic energy scales in nanopore systems, yields:

$$\Psi_{Electrophoreti\ \_Norm}(WT\_Arg) = 0.643$$

$$\Psi_{Electrophoreti\ \_Norm}(WT\_Cit) = 0$$

$$\Delta\Psi_{Electrophor} \quad (WT) = -0.643$$

This pronounced disparity in electrophoretic contribution, corresponding to a normalized $\Delta\Psi_{Electrophore}$ of −0.643. In the present framework, electrophoretic differences are treated as one potentially informative transport-related descriptor that may contribute to different translocation kinetics between Arg- and Cit-containing peptides. The positively charged arginine residue experiences substantial electrophoretic propulsion toward the cathode, which markedly increases its translocation probability while simultaneously reducing its dwell time within the pore. In stark contrast, the neutral citrulline moiety lacks this directed electrophoretic impetus, depending principally on electroosmotic flow and diffusive processes for its translocation through the nanoconstriction.

Consequently, the electrophoretic term functions as a powerful discriminant between modified and unmodified peptides, establishing distinct translocation kinetic profiles that effectively complement the electrostatic interaction characteristics. This integrated

electrical discrimination strategy, which combines both static electrostatic and dynamic electrophoretic charge-based phenomena, significantly augments the overall detection sensitivity for citrullination. This multi-parameter approach enables robust identification of this subtle post-translational modification through comprehensive signal analysis.

The last term, electroosmotic contribution. This electroosmotic force can be expressed as:

$$\mathrm{F}_{Electroosmotic} \propto -\frac{\varepsilon_r \zeta_p E}{\eta}$$

Where ζ_p is zeta potential at the pore surface, and η is the viscosity of the solution. By treating viscosity η and dielectric constant ε_r as system constants, the corresponding energy term can be simplified to:

$$\Psi_{Electroosmotic} = \kappa \times |\zeta_p| \times V_{bias}$$

Here, κ is a proportionality constant relating zeta potential to effective electroosmotic charge, empirically estimated at 0.25 e/mV.

For WT, ζ_p is +10 mV due to slightly positively charged lysine residue, $\Psi_{Electroosmotic}\,(WT) = 0.25 \times 10 \times 0.1 \times (96.5\ kJ/mol\ per\ e \cdot V) = 24.1\ kJ/mol$

Normalized within a 0–100 kJ/mol reference range yields:

$$\Psi_{Electroosmotic\_Norm}\,(WT) = 0.241$$

The magnitude of electroosmotic energy is solely dependent on the surface charge characteristics of the nanopore channel and is independent of the chemical properties of the analyte. A stronger electroosmotic flow provides an effective translocation

driving force for all dissolved molecules, including both charged and neutral species, with particular benefit to neutral molecules such as citrulline. Similar to electrophoretic effects, electroosmotic force directly influences the temporal characteristics of the sensing signal S, particularly dwell time distributions, by modulating the net translocation velocity of molecules. This regulation provides an additional dimension for molecular discrimination, enabling precise differentiation through multiparameter analysis.

To achieve specific recognition of citrullination, we designed nanopore mutants tailored for this post-translational modification detection and selected the following variants based on their distinct side-chain functional groups: the amide group of asparagine (Asn, N), the mercapto group of cysteine (Cys, C), the aromatic indole group of tryptophan (Trp, W), and the carboxyl group of glutamic acid (Glu, E). Below is the detailed calculation process.

1.1.3 Mutation-specific implementation in the present aerolysin system

For K238N,

✧ Steric

$$\Psi_{Steric}(Arg) = 0.633$$

$$\Psi_{Steric}(Cit) = 0.617$$

✧ Binding energy

$\Psi_{Binding}(K238N\_Arg) = -127.6\ kJ/mol,\ \Psi_{Binding\_Norm}(K238N\_Arg) = 0.2552$

$\Psi_{Binding}(K238N\_Cit) = -84.9\ kJ/mol,\ \Psi_{Binding\_Norm}(K238N\_Cit) = 0.1698$

✧ VDW interactions

Arg: 3×N-N + 2×N-O + 2×C-C + 1×C-O + 2×H-N, $\Psi_{VDW}(K238N\_Arg) =$ 11.777 kJ/mol, $\Psi_{VDW\_Norm}(K238N\_Arg) = 0.785$

Cit: 2×N-N + 1×N-O + 1×O-N + 2×C-C + 1×C-O + 1×H-O, $\Psi_{VDW}(K238N\_Cit) =$ 9.173 kJ/mol, $\Psi_{VDW\_Norm}(K238N\_Cit) = 0.612$

✧ Electrostatic interactions

$$\Psi_{Electrostatic}(K238N\_Arg) = \Psi_{Electrostatic}(K238N\_Cit) = 0$$

$$\Psi_{Electrostatic\_Norm}(K238N\_Arg) = \Psi_{Electrostatic\_Norm}(K238N\_Cit) = 0.5$$

✧ Hydrophobic

$$\Psi_{Hydrophobic\_Norm}(K238N\_Arg) = 0.638$$

$$\Psi_{Hydrophobi\ \_Norm}(K238N\_Cit) = 0.660$$

✧ Electrophoretic

$$\Psi_{Electrophoretic\_Norm}(K238N\_Arg) = 0.643$$

$$\Psi_{Electrophor\quad\_Norm}(K238N\_Cit) = 0$$

✧ Electroosmotic

$$\zeta_p = +5\ mV$$

$$\Psi_{Electroosmotic}(K238N\_Arg) = \Psi_{Electroosmotic}(K238N\_Cit)$$

$$= 0.25 \times 5 \times 0.1 \times 96.5 = 12.06\ kJ/mol$$

$\Psi_{Electroosmotic\_Norm}(K238N\_Arg) = \Psi_{Electroosmotic\_Norm}(K238N\_Cit)$=0.121

For K238C,

✧ Steric

$$\Psi_{Steric}(Arg) = 0.633$$

$$\Psi_{Steric}(Cit) = 0.617$$

✧ Binding energy

$\Psi_{Binding}(K238C\_Arg) = -54.7\ kJ/mol,\ \Psi_{Binding\_Norm}(K238C\_Arg)$ =0.1094

$\Psi_{Binding}(K238C\_Cit) = -31.8\ kJ/mol,\ \Psi_{Binding\_Norm}(K238C\_Cit) = 0.0636$

✧ VDW interactions

Arg: 3×S-N + 2×S-C + 2×C-C + 2×H-S, $\Psi_{VDW}(K238C\_Arg) = 15.184$ kJ/mol, $\Psi_{VDW\_Norm}(K238C\_Arg) = 1.012 \rightarrow 1.0$ (cutoff)

Cit: 1×S-N + 1×S-O + 1×S-C+ 1×C-C + 1×S-H, $\Psi_{VDW}(K238C\_Cit) = 10.088$ kJ/mol $\Psi_{VDW\_Norm}(K238C\_Cit) = 0.673$

✧ Electrostatic interactions

$$\Psi_{Electrostatic}(K238C\_Arg) = \Psi_{Electrostatic}(K238C\_Cit) = 0$$

$$\Psi_{Electrostatic\_Norm}(K238C\_Cit) = \Psi_{Electrostatic\_Norm}(K238C\_Cit) = 0.5$$

✧ Hydrophobic

$$\Psi_{Hydrophobic\_Norm}(K238C\_Arg) = 0.638$$

$$\Psi_{Hydrophobic\_Norm}(K238C\_Cit) = 0.660$$

✧ Electrophoretic

$$\Psi_{Electrophoreti\ \_Norm}(K238C\_Arg) = 0.643$$

$$\Psi_{Electrophoreti\ \_Norm}(K238C\_Cit) = 0$$

✧ Electroosmotic

$$\zeta_p = 0\ mV$$

$$\Psi_{Electroosmotic\_Norm}(K238C\_Arg) = \Psi_{Electroosmotic\_Norm}(K238C\_Cit) = 0$$

For K238W,

✧ Steric

$$\Psi_{Steric}(Arg) = 0.633$$

$$\Psi_{Steric}(Cit) = 0.617$$

✧ Binding energy

$$\Psi_{Binding}(K238W\_Arg) = -69.4\ kJ/mol,\ \Psi_{Binding\_Norm}(K238W\_Arg) = 0.1388$$

$$\Psi_{Binding}(K238W\_Cit) = -27.2\ kJ/mol,\ \Psi_{Binding\_Norm}(K238W\_Cit) = 0.0544$$

✧ VDW interactions

The indole ring exhibits enhanced van der Waals interactions attributable to aromatic ring polarizability, corresponding to an enhancement factor of 1.3,

Arg: (3×N-N + 4×C-C) ×1.3,

$$\Psi_{VDW}(K238W\_Arg) = 11.086\ \text{kJ/mol},$$

$$\Psi_{VD\ \ Norm}(K238W\_Arg) = 0.739$$

Cit: (2×N-N + 1×N-O + 3×C-C) × 1.3,

$$\Psi_{VDW}(K238W\_Cit) = 9.433\ \text{kJ/mol}$$

$$\Psi_{VDW\_Norm}(K238C\_Cit) = 0.0.629$$

✧ Electrostatic interactions

$$\Psi_{Electrostatic}(K238W\_Arg) = \Psi_{Electrostatic}(K238W\_Cit) = 0$$

$$\Psi_{Electrostati\ \_Norm}(K238W\_Arg) = \Psi_{Electrosta\ \ \_Norm}(K238W\_Cit) = 0.5$$

✧ Hydrophobic

$$\Psi_{Hydrophobic\_Norm}(K238W\_Arg) = 0.638$$

$$\Psi_{Hydrophobic\_Norm}(K238W\_Cit) = 0.660$$

✧ Electrophoretic

$$\Psi_{Electrophoretic\_Norm}(K238W\_Arg) = 0.643$$

$$\Psi_{Electrophoretic\_Norm}(K238W\_Cit) = 0$$

✧ Electroosmotic

$$\zeta_p = 0\ mV$$

$$\Psi_{Electroosmotic\_Norm}(K238W\_Arg) = \Psi_{Electroosmotic\_Norm}(K238W\_Cit) = 0$$

For K238E,

✧ Steric

$$\Psi_{Steric}(Arg) = 0.633$$

$$\Psi_{Steric}(Cit) = 0.617$$

✧ Binding energy

$\Psi_{Binding}(K238E\_Arg) = -467.1 kJ/mol,\ \Psi_{Binding\_Norm}(K238E\_Arg) = 0.9342$

$\Psi_{Binding}(K238E\_Cit) = -136.9\ kJ/mol,\ \Psi_{Binding\_Norm}(K238E\_Cit) = 0.2738$

✧ VDW interactions

Arg: 3×N-O + 2×O-C + 2×C-C + 2×H-O,

$$\Psi_{VDW}(K238E\_Arg) = 8.711\ \text{kJ/mol},$$

$$\Psi_{VDW_{Norm}}(K238E\_Arg) = 0.581$$

Cit: 2×N-O + 1×O-O + 2×O-C + 2×C-C + 1×H-O,

$$\Psi_{VDW}(K238E\_Cit) = 7.818\ \text{kJ/mol}$$

$$\Psi_{VDW\_Norm}(K238E\_Cit) = 0.521$$

✧ Electrostatic interactions

$$\Psi_{Electrostatic}(K238E\_Arg) = \frac{1}{4\pi\times 8.854\times 10^{-12}\times 30} \times \frac{(+1e)\times(-1e)}{5\times 10^{-10}} = -46.2\ kJ/mol$$

$$\Psi_{Electrostatic}(K238E\_Cit) = 0\ kJ/mol$$

$$\Psi_{Electrostatic\_Norm}(K238E\_Arg) = 0.91$$

$$\Psi_{Electrostatic\_Norm}(K238E\_Cit) = 0.5$$

✧ Hydrophobic

$$\Psi_{Hydrophobic\_Norm}(K238E\_Arg) = 0.638$$

$$\Psi_{Hydrophobic\_Norm}(K238E\_Cit) = 0.660$$

✧ Electrophoretic

$$\Psi_{Electrophoretic\_Norm}(K238E\_Arg) = 0.643$$

$$\Psi_{Electrophore\quad\_Norm}(K238E\_Cit) = 0$$

✧ Electroosmotic

$\zeta_p = -40\ mV$ due to strongly negatively charged glutamate residue,

$$\Psi_{Electroosmotic}(K238E\_Arg) = \Psi_{Electroosmotic\_}(K238E\_Cit)$$

$$= 0.25 \times 40 \times 0.1 \times (96.5\ kJ/mol\ per\ e \cdot V) \ = \ 96.4\ kJ/mol$$

$$\Psi_{Electroosmotic\_Norm}(K238E\_Arg) = \Psi_{Electroosmotic\_Norm}(K238E\_Cit) = 0.965$$

Table S3. Summary of normalized values for WT, K238N, K238C, K238W, and K238E corresponding to Fig. 1.

Interactions	**WT**		**K238N**		**K238C**		**K238W**		**K238E**	
	Arg	Cit	Arg	Cit	Arg	Cit	Arg	Cit	Arg	Cit
$\boldsymbol{\Psi_{Steric}}$	0.633	0.617	0.633	0.617	0.633	0.617	0.633	0.617	0.633	0.617
$\boldsymbol{\Psi_{Binding}}$	0	0.315	0.255	0.17	0.109	0.064	0.139	0.054	0.934	0.274
$\boldsymbol{\Psi_{VDW}}$	0.553	0.464	0.785	0.612	1	0.673	0.739	0.629	0.581	0.521
$\boldsymbol{\Psi_{Electrostatic}}$	0.09	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.91	0.5
$\boldsymbol{\Psi_{Hydrophobic}}$	0.638	0.66	0.638	0.66	0.638	0.66	0.638	0.66	0.638	0.66

$\Psi_{Electrophoretic}$	0.643	0	0.643	0	0.643	0	0.643	0	0.643	0
$\Psi_{Electroosmotic}$	0.241	0.241	0.121	0.121	0	0	0	0	0.965	0.965

Assume that the weighting factors are equal,

$$w_1 = w_2 = w_3 = w_4 = w_5 = w_6 = w_7 = 1/7$$

Table S4. Calculated S and ΔS values used in this work.

pore	analyte	S	ΔS
WT	Arg	0.4	0
	Cit	0.4	
K238N	Arg	0.511	0.128
	Cit	0.383	
K238C	Arg	0.503	0.144
	Cit	0.359	
K238W	Arg	0.47	0.119
	Cit	0.351	
K238E	Arg	0.758	0.253
	Cit	0.505	

1.1.4 Scope, assumptions, and limitations of the present implementation

The present PIToN implementation is intentionally simplified and should be interpreted as a reduced descriptor model rather than as a full mechanistic model of nanopore transport. For tractability, K238 is treated as an engineering-relevant position within a local constriction-centered sensing region, rather than as an exclusive single-

site reader. Local residue–analyte interactions are represented using simplified pairwise configurations and a single-protomer approximation, rather than the full many-body environment of the heptameric pore. The DFT-derived binding-related term is used as a heuristic local interaction descriptor under gas-phase conditions and is not intended to represent solution-phase binding free energies under the 4 M KCl experimental conditions. Electrophoretic and electroosmotic contributions are retained as phenomenological transport-related descriptors that may influence dwell time, rather than as a rigorous decomposition of the in-pore force balance. Equal weighting is adopted as a transparent first-order approximation without claiming that the physically optimal weighting scheme has been established. All seven terms are retained for conceptual completeness, although not all of them are equally informative in the present citrullination system. Accordingly, PIToN is used here to support comparative mutation prioritization within the present aerolysin-based citrullination system, while broader applicability to other nanopores, analytes, and PTM classes remains to be established.

Rationale for the additive linear combination model in PIToN

The nanopore model described here assumes that the sensing response is a weighted linear combination of various physical factors. The additive form used in PIToN is a pragmatic first-order hypothesis motivated by physico-chemical considerations and model tractability.

The following points provide the key justifications for this approach:

1. Physico-chemical basis: approximate independence of molecular forces

- Independence of forces: at the molecular scale, different physical interactions,

such as van der Waals forces, electrostatic interactions, hydrophobic effects, and hydrogen bonding, can be considered independent to a first approximation. The total interaction energy (Ψ) between an analyte and the pore wall can be approximated as the sum of these individual interaction energies:

$$\Psi_{total} \approx \Psi_{elec} + \Psi_{vdw} + \Psi_{hbond} + \Psi_{hydrophobic} + \cdots$$

- From energy to signal: the nanopore signal (e.g., blockade current I_b or dwell time Δt) is closely related to the effective potential energy landscape of the analyte within the pore, which is determined by this total interaction energy, Ψ_{total}. Therefore, if the total energy is a sum of individual contributions, the observed signal S, as a function of the total energy, can be expressed as a linear combination of the individual interaction contributions within a small perturbation range.

$$S = f(\Psi_{total}) \approx a * \Psi_{elec} + b * \Psi_{vdw} + c * \Psi_{hbond} + d * \Psi_{hydrophobi} + \cdots + constant$$

A simple analogy: consider an object subjected to multiple forces (gravity, push, friction). The object's total acceleration (the observed signal) is directly related to the vector sum (linear addition) of these forces, not their product.

2. Mathematical justification: first-order Taylor expansion

From a mathematical perspective, any complex, unknown functional relationship can be approximated by a linear (first-order) model within a sufficiently small operating range. This is the fundamental concept of a Taylor expansion.

- Even if the true physical relationship $S = f(\Psi_{elec}, \Psi_{vdw}, \cdots)$ is a complex

nonlinear function, it can be expanded around a point of interest (e.g., for a specific class of analytes), and higher-order terms can be neglected:

$$S \approx S_0 + (\partial f/\partial \Psi_{elec}) * \Delta G\_elec + (\partial f/\partial \Psi_{vdw}) * \Psi_{vdw} + \cdots$$

- The partial derivatives $(\partial f/\partial \Psi_{xxx})$ correspond to the "weights" mentioned in the main text, and S_0 is the baseline signal. This directly constitutes a weighted linear combination model.

3. Why not a multiplicative relationship?

A multiplicative relationship $S = k * (\Psi_{elec}) * (\Psi_{vdw}) * \cdots$ typically implies strong synergistic or coupling effects between the factors.

- Synergistic effects: for instance, a dramatic change in signal would only occur when electrostatic attraction and hydrophobic interaction are simultaneously present; the absence of one would render the other nearly ineffective. While this can exist in highly complex biological systems (e.g., "induced fit" in enzyme-substrate binding), it is not the default scenario for a relatively simpler sensing system like a nanopore, especially in an initial model.
- Model complexity and tractability:
  - The additive form is used here as a pragmatic first-order approximation rather than as a claim that all interaction terms are truly independent under experimental nanopore conditions. Its main advantage in the present study is interpretability and tractability within a reduced descriptor framework.
  - The multiplicative model is a nonlinear model, making parameter fitting and estimation much more difficult, prone to overfitting, and susceptible to local

minima, for an initial framework intended to organize comparative mutation screening in a transparent and interpretable way.

These observations support the additive model hypothesis that factors contribute independently. Naturally, this does not mean the additive model is an absolute truth. We acknowledge this by explicitly labeling PIToN a “semi-rational theoretical framework.” It is therefore best interpreted as a working approximation whose limitations can help identify where additional nonlinear or coupled effects may need to be incorporated in future refinements.

1.2 Quantum-chemical evaluation of residue–analyte pairwise interaction energies

Density functional theory calculations were performed using the Gaussian software to estimate residue–analyte pairwise interaction energies. All calculations employed the B3LYP exchange-correlation functional with the 6-31G basis set, and the SMD implicit solvent model for water was used to mimic the aqueous environment. The interaction energy was defined as

$$\Delta E_{int} = E_{A\cdot B} - E_A - E_B$$

where A denotes a residue side-chain analog and B denotes the analyte analog (Arg or Cit). In this reduced model, ΔE_{int} was used to compare the relative interaction tendencies of Arg and Cit with different residue types.

The calculations in this section were intended to provide simplified local interaction descriptors for comparative ranking within the PIToN model. Specifically, the calculations were carried out on isolated residue–analyte dimers representing local interaction motifs under gas-phase conditions. Accordingly, the resulting values do not account for solvent screening, high ionic strength, thermal sampling, full pore geometry, or the many-body environment of the nanopore. These quantities should therefore be interpreted as heuristic relative interaction descriptors, reflecting residue-specific interaction preference, rather than as quantitative pore-level binding free energies under the experimental conditions.

Table S5. Side-chain analogs used in the pairwise interaction calculations.

Residue	Side-chain analog	Charge	Multiplicity
Alanine	methane	0	1

Arginine	propylguanidinium	+1	1
Asparagine	acetamide	0	1
Aspartate	acetate	−1	1
Citrulline	propylureido analog	0	1
Cysteine	methanethiol	0	1
Glutamate	propionate	−1	1
Glutamine	propionamide	0	1
Glycine	hydrogen	0	1
Histidine	neutral methyl-imidazole analog	0	1
Isoleucine	n-butane	0	1
Leucine	isobutane	0	1
Lysine	n-butylammonium	+1	1
Methionine	ethyl methyl sulfide	0	1
Phenylalanine	toluene	0	1
Proline	pyrrolidine	0	1
Serine	methanol	0	1
Threonine	ethanol	0	1
Tyrosine	p-cresol	0	1
Tryptophan	3-methylindole	0	1
Valine	propane	0	1

For each Arg–residue and Cit–residue pair, 3–5 distinct initial dimer conformations were manually constructed based on plausible local interaction motifs, including

hydrogen bonding, salt-bridge-like contacts, cation–π interactions, polar–π contacts, and hydrophobic side-chain contacts where appropriate. All dimers were optimized at the same level of theory as used for the monomers. For the main-text comparison, the lowest-energy optimized conformer was taken as the representative value for each pair. Representative conformational screening results, including 3–5 initial poses and their corresponding interaction energies, are provided here using Arg–Asn as an example in Table S6.

Table S6. An example of DFT calculations with different initial posess. The energy unit is kJ/mol.

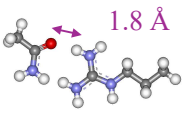
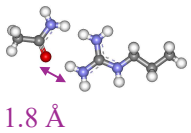
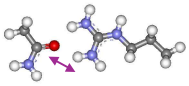
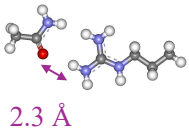
Pose ID	1	2	3	4
Initial pose	NH···O hydrogen bond	shifted NH···O hydrogen bond	rotated NH···O pose	alternative offset pose
Scheme				
ΔE_{int}	−127.615	−127.568	−127.565	−127.568

Table S7. Calculated interaction energies for Arg and Cit with the 20 residue types.

Residue	**One-letter code**	**ΔE with Arg (kJ/mol)**	**ΔE with Cit (kJ/mol)**
Lys	**K**	17.6	−157.4
Asp	**N**	−127.6	-84.9
Cys	**C**	−54.7	-31.8
Try	**W**	-69.4	-27.2
Glu	**E**	-467.1	-136.9

Gln	**Q**	-126.0	-84.2
Tyr	**Y**	-88.5	-61.4
Ala	**A**	-6.7	-8.5
Asp	**D**	-475.9	-262.7
Phe	**F**	-0.02	-22.0
Gly	**G**	-4.0	-7.4
His	**H**	-187.1	-114.5
Ile	**I**	-15.8	-12.7
Leu	**L**	-20.0	-11.5
Met	**M**	-65.9	-26.8
Pro	**P**	-97.7	-57.4
Arg	**R**	29.5	-386.7
Ser	**S**	-102.3	-67.0
Thr	**T**	-105.3	-46.4
Val	**V**	-332.6	-471.0

1.3 Calculations for other mutants by our PIToN model

For K238Q,

✧ Binding energy

$\Psi_{Binding}(K238Q\_Arg) = -126.0\ kJ/mol,\ \Psi_{Binding\_Norm}(K238Q\_Arg) = 0.252$

$\Psi_{Binding}(K238Q\_Cit) = -84.2\ kJ/mol,\ \Psi_{Binding\_Norm}(K238Q\_Cit) = 0.1684$

✧ VDW interactions

Arg: 3×N-N + 2×N-O + 2×C-C + 1×C-O + 2×H-N, $\Psi_{VDW}(K238Q\_Arg) =$ $11.777\ \mathrm{kJ/mol},\ \Psi_{VDW\_Norm}(K238Q\_Arg) = 0.785$

Cit: 2×N-N + 1×N-O + 1×O-N + 2×C-C + 1×C-O + 1×H-O, $\Psi_{VDW}(K238Q\_Cit) =$ $9.173\ \mathrm{kJ/mol}\ \Psi_{VDW\_Norm}(K238Q\_Cit) = 0.612$

✧ Electroosmotic

$$\zeta_p = +5\ mV$$

$$\Psi_{Electroosmotic}(K238Q\_Arg) = \Psi_{Electroosmotic}(K238Q\_Cit)$$

$$= 0.25 \times 5 \times 0.1 \times 96.5 = 12.06\ kJ/mol$$

$$\Psi_{Electroosmotic\_Norm}(K238Q\_Arg) = \Psi_{Electroosmotic\_Norm}(K238Q\_Cit) = 0.121$$

For K238Y,

✧ Binding energy

$\Psi_{Binding}(K238Y\_Arg) = -88.5\ kJ/mol,\ \Psi_{Binding\_Norm}(K238Y\_Arg) = 0.177$

$\Psi_{Binding}(K238Y\_Cit) = -61.4\ kJ/mol,\ \Psi_{Binding\_Norm}(K238Y\_Cit) = 0.1228$

✧ VDW interactions (plus enhancement factor of 1.3)

Arg: (3×N-O + 4×C-C + 2×H-O) ×1.3, $\Psi_{VDW}(K238Y\_Arg) = 11.2236\ \mathrm{kJ/mol}$, $\Psi_{VDW\_Norm}(K238Y\_Arg) = 0.748$

Cit: （1×N-O + 1×O-O + 4×C-C + 1×H-O) ×1.3, $\Psi_{VDW}(K238Y\_Cit) = 8{,}4105\text{kJ/mol}$ $\Psi_{VDW\_Norm}(K238Y\_Cit) = 0.561$

For K238A

✧ Binding energy

$\Psi_{Binding}(K238A\_Arg) = -6.7\ kJ/mol,\ \Psi_{Binding\_Norm}(K238A\_Arg) = 0.0134$

$\Psi_{Binding}(K238A\_Cit) = -8.5\ kJ/mol,\ \Psi_{Binding\_Norm}(K238A\_Cit) = 0.017$

✧ VDW interactions

Arg: 2×C-C + 2×H-C, $\Psi_{VDW}(K238A\_Arg) = 2.996\ \text{kJ/mol}$, $\Psi_{VDW\_Norm}(K238A\_Arg) = 0.200$

Cit: 1×C-O + 1×H-O, $\Psi_{VDW}(K238A\_Cit) = 1.548\ \text{kJ/mol}$ $\Psi_{VDW\_Norm}(K238A\_Cit) = 0.103$

For K238F

✧ Binding energy

$\Psi_{Binding}(K238F\_Arg) = -0.02\ kJ/mol,\ \Psi_{Binding\_Norm}(K238F\_Arg) = 0.00004$

$\Psi_{Binding}(K238F\_Cit) = -22.0\ kJ/mol,\ \Psi_{Binding\_Norm}(K238F\_Cit) = 0.044$

✧ VDW interactions

The benzene ring exhibits enhanced van der Waals interactions attributable to aromatic ring polarizability, corresponding to an enhancement factor of 1.3,

Arg: (3×C-C + 2×H-C) ×1.3, $\Psi_{VDW}(K238F\_Arg) = 10.906\ \text{kJ/mol}$, $\Psi_{VDW\_Norm}(K238F\_Arg) = 0.727$

Cit: （1×C-O + 1×H-O) ×1.3, $\Psi_{VDW}(K238F\_Cit) = 4.516\ \text{kJ/mol}$ $\Psi_{VDW\_Norm}(K238F\_Cit) = 0.301$

For K238D

✧ Binding energy

$\Psi_{Binding}(K238D\_Arg) = -475.9\ kJ/mol,\ \Psi_{Binding\_Norm}(K238D\_Arg) = 0.9518$

$\Psi_{Binding}(K238D\_Cit) = -262.7\ kJ/mol,\ \Psi_{Binding\_Norm}(K238D\_Cit) = 0.5254$

✧ VDW interactions

Arg: 2×C-N + 2×C-C, $\Psi_{VDW}(K238D\_Arg) = 5.664$ kJ/mol , $\Psi_{VDW\_Norm}(K238D\_Arg) = 0.378$

Cit: 1×C-O + 1×C-N, $\Psi_{VDW}(K238D\_Cit) = 2.156$ kJ/mol $\Psi_{VDW\_Norm}(K238D\_Cit) = 0.144$

✧ Electroosmotic

$$\zeta_p = -40\ mV$$

$$\Psi_{Electroosmotic}(K238D\_Arg) = \Psi_{Electroosmotic}(K238D\_Cit) = 96.4\ kJ/mol$$
$$\Psi_{Electroosmotic\_Norm}(K238D\_Arg) = \Psi_{Electroosmotic\_Norm}(K238D\_Cit) = 0.965$$

For K238G

✧ Binding energy

$\Psi_{Binding}(K238G\_Arg) = -4.0\ kJ/mol,\ \Psi_{Binding\_Norm}(K238G\_Arg) = 0.008$

$\Psi_{Binding}(K238G\_Cit) = -7.4\ kJ/mol,\ \Psi_{Binding\_Norm}(K238G\_Cit) = 0.0148$

✧ VDW interactions

Arg: 2×H-N + 2×H-C, $\Psi_{VDW}(K238G\_Arg) = 2.96$ kJ/mol , $\Psi_{VDW\_Norm}(K238G\_Arg) = 0.197$

Cit: 1×H-O + 1×H-N, $\Psi_{VDW}(K238G\_Cit) = 1.208$ kJ/mol $\Psi_{VDW\_Norm}(K238G\_Cit) = 0.081$

For K238H

✧ Binding energy

$\Psi_{Binding}(K238H\_Arg) = 187.1\ kJ/mol,\ \Psi_{Binding\_Norm}(K238H\_Arg) = 0$

$\Psi_{Binding}(K238H\_Cit) = -114.5\ kJ/mol,\ \Psi_{Binding\_Norm}(K238H\_Cit) = 0.229$

✧ VDW interactions

The imidazole ring exhibits enhanced van der Waals interactions attributable to aromatic ring polarizability, corresponding to an enhancement factor of 1.3,

Arg: (2×N-N + 1×C-N + 2×H-N) × 1.3, $\Psi_{VDW}(K238H\_Arg) = 9.333$ kJ/mol, $\Psi_{VDW\_Norm}(K238H\_Arg) = 0.622$

Cit: (1×N-O + 1×C-O + 1×H-O) × 1.3, $\Psi_{VDW}(K238H\_Cit) = 3.859$ kJ/mol $\Psi_{VDW\_Norm}(K238H\_Cit) = 0.257$

For K238L

✧ Binding energy

$\Psi_{Binding}(K238L\_Arg) = -20.0\ kJ/mol,\ \Psi_{Binding\_Norm}(K238L\_Arg) = 0.04$

$\Psi_{Binding}(K238L\_Cit) = -11.5\ kJ/mol,\ \Psi_{Binding\_Norm}(K238L\_Cit) = 0.023$

✧ VDW interactions

Arg: 4×C-C + 3×H-C, $\Psi_{VDW}(K238L\_Arg) = 8.235$ kJ/mol, $\Psi_{VDW\_Norm}(K238L\_Arg) = 0.549$

Cit: 1×C-O + 2×H-O, $\Psi_{VDW}(K238L\_Cit) = 3.362$ kJ/mol $\Psi_{VDW\_Norm}(K238L\_Cit) = 0.224$

For K238I

✧ Binding energy

$\Psi_{Binding}(K238I\_Arg) = -15.8\ kJ/mol,\ \Psi_{Binding\_Norm}(K238I\_Arg) = 0.0316$

$\Psi_{Binding}(K238I\_Cit) = -12.7\ kJ/mol,\ \Psi_{Binding\_Norm}(K238I\_Cit) = 0.0254$

✧ VDW interactions

Arg: 4×C-C + 3×H-C, $\Psi_{VDW}(K238I\_Arg) = 8.854$ kJ/mol , $\Psi_{VDW\_Norm}(K238I\_Arg) = 0.590$

Cit: 1×C-O + 2×H-O, $\Psi_{VDW}(K238I\_Cit) = 3.614$ kJ/mol $\Psi_{VDW\_Norm}(K238I\_Cit) = 0.241$

For K238M

✧ Binding energy

$\Psi_{Binding}(K238M\_Arg) = -65.9\ kJ/mol,\ \Psi_{Binding\_Norm}(K238M\_Arg) = 0.1258$

$\Psi_{Binding}(K238M\_Cit) = -26.8\ kJ/mol,\ \Psi_{Binding\_Norm}(K238M\_Cit) = 0.0536$

✧ VDW interactions

Arg: 2×S-N + 2×C-C + 2×H-N, $\Psi_{VDW}(K238M\_Arg) = 9.015$ kJ/mol , $\Psi_{VDW\_Norm}(K238M\_Arg) = 0.601$

Cit: 1×S-O + 1×C-O + 1×H-O, $\Psi_{VDW}(K238M\_Cit) = 3.780$ kJ/mol $\Psi_{VDW\_Norm}(K238M\_Cit) = 0.252$

For K238P

✧ Binding energy

$\Psi_{Binding}(K238P\_Arg) = -97.9\ kJ/mol,\ \Psi_{Binding\_Norm}(K238P\_Arg) = 0.1954$

$\Psi_{Binding}(K238P\_Cit) = -57.4\ kJ/mol,\ \Psi_{Binding\_Norm}(K238P\_Cit) = 0.1148$

✧ VDW interactions

Arg: 2×C-N + 2×C-C, $\Psi_{VDW}(K238P\_Arg) = 5.664$ kJ/mol , $\Psi_{VDW\_Norm}(K238P\_Arg) = 0.378$

Cit: 1×C-O + 1×C-N, $\Psi_{VDW}(K238P\_Cit) = 2.156$ kJ/mol

$\Psi_{VDW\_Norm}(K238P\_Cit) = 0.144$

For K238R

✧ Binding energy

$\Psi_{Binding}(K238R\_Arg) = 29.5\ kJ/mol,\ \Psi_{Binding\_Norm}(K238R\_Arg) = 0$

$\Psi_{Binding}(K238R\_Cit) = -387.7\ kJ/mol,\ \Psi_{Binding\_Norm}(K238R\_Cit) = 0.7754$

✧ VDW interactions

Arg: 3×N-N + 2×C-N + 2×H-N, $\Psi_{VDW}(K238R\_Arg) = 11.777$ kJ/mol ,

$\Psi_{VDW\_Norm}(K238R\_Arg) = 0.785$

Cit: 1×N-O + 1×C-O + 1×H-O, $\Psi_{VDW}(K238R\_Cit) = 4.800$ kJ/mol

$\Psi_{VDW\_Norm}(K238R\_Cit) = 0.320$

✧ Electroosmotic

$$\zeta_p = +10\ mV$$

$$\Psi_{Electroosmotic}(K238R\_Arg) = \Psi_{Electroosmotic}(K238R\_Cit) = 24.1\ kJ/mol$$

$$\Psi_{Electroosmotic\_Norm}(K238R\_Arg) = \Psi_{Electroosmotic\_Norm}(K238R\_Cit) = 0.241$$

For K238S

✧ Binding energy

$\Psi_{Binding}(K238S\_Arg) = -102.3\ kJ/mol,\ \Psi_{Binding\_Norm}(K238S\_Arg) = 0.2046$

$\Psi_{Binding}(K238S\_Cit) = -67.0\ kJ/mol,\ \Psi_{Binding\_Norm}(K238S\_Cit) = 0.134$

✧ VDW interactions

Arg: 1×O-N + 2×C-N + 2×H-N, $\Psi_{VDW}(K238S\_Arg) = 6.724$ kJ/mol ,

$\Psi_{VDW\_Norm}(K238S\_Arg) = 0.448$

Cit: 1×O-O + 1×C-O + 1×H-O, $\Psi_{VDW}(K238S\_Cit) = 2.744$ kJ/mol

$\Psi_{VDW\_Norm}(K238S\_Cit) = 0.183$

For K238T

✧ Binding energy

$\Psi_{Binding}(K238T\_Arg) = -105.3\ kJ/mol,\ \Psi_{Binding\_Norm}(K238T\_Arg) = 0.2106$

$\Psi_{Binding}(K238T\_Cit) = -46.4\ kJ/mol,\ \Psi_{Binding\_Norm}(K238T\_Cit) = 0.0928$

✧ VDW interactions

Arg: 1×O-N + 2×C-N + 2×H-N, $\Psi_{VDW}(K238T\_Arg) = 7.233$ kJ/mol , $\Psi_{VDW\_Norm}(K238T\_Arg) = 0.482$

Cit: 1×O-O + 1×C-O + 1×H-O, $\Psi_{VDW}(K238T\_Cit) = 2.950$ kJ/mol $\Psi_{VDW\_Norm}(K238T\_Cit) = 0.197$

For K238V

✧ Binding energy

$\Psi_{Binding}(K238V\_Arg) = -332.6\ kJ/mol,\ \Psi_{Binding\_Norm}(K238V\_Arg) = 0.6652$

$\Psi_{Binding}(K238V\_Cit) = -471.0\ kJ/mol,\ \Psi_{Binding\_Norm}(K238V\_Cit) = 0.942$

✧ VDW interactions

Arg: 4×C-C + 1×H-C, $\Psi_{VDW}(K238V\_Arg) = 6.756$ kJ/mol , $\Psi_{VDW\_Norm}(K238V\_Arg) = 0.450$

Cit: 1×C-O + 1×H-O, $\Psi_{VDW}(K238V\_Cit) = 2.156$ kJ/mol $\Psi_{VDW\_Norm}(K238V\_Cit) = 0.144$

✧ Electrostatic interactions

$$\Psi_{Electrostatic}(K238V\_Arg) = \Psi_{Electrostatic}(K238V\_Cit) = 0$$

$$\Psi_{Electrostatic\_Norm}(K238V\_Arg) = \Psi_{Electrostatic\_Norm}(K238V\_Cit) = 0.5$$

Table S8. Summary table for mutant nanopores.

Interactions		Ψ_{Steric}	$\Psi_{Binding}$	Ψ_{VDW}	$\Psi_{Electrostatic}$	$\Psi_{Hydrophobic}$	$\Psi_{Electrophoretic}$	$\Psi_{Electroosmotic}$
K238Q	Arg	0.633	0.252	0.785	0.5	0.638	0.643	0.121
	Cit	0.617	0.1684	0.612	0.5	0.66	0	0.121
K238Y	Arg	0.633	0.177	0.748	0.5	0.638	0.643	0
	Cit	0.617	0.1228	0.561	0.5	0.66	0	0
K238A	Arg	0.633	0.0134	0.200	0.5	0.638	0.643	0
	Cit	0.617	0.0107	0.103	0.5	0.66	0	0
K238F	Arg	0.633	0.00004	0.727	0.5	0.638	0.643	0
	Cit	0.617	0.044	0.301	0.5	0.66	0	0
K238D	Arg	0.633	0.9518	0.378	0.91	0.638	0.643	0.965
	Cit	0.617	0.5254	0.144	0.5	0.660	0	0.965
K238G	Arg	0.633	0.008	0.197	0.5	0.638	0.643	0
	Cit	0.617	0.0148	0.081	0.5	0.660	0	0
K238H	Arg	0.633	0	0.622	0.5	0.638	0.643	0
	Cit	0.617	0.229	0.257	0.5	0.660	0	0
K238L	Arg	0.633	0.04	0.549	0.5	0.638	0.643	0
	Cit	0.617	0.023	0.224	0.5	0.660	0	0
K238I	Arg	0.633	0.0316	0.590	0.5	0.638	0.643	0
	Cit	0.617	0.0254	0.241	0.5	0.660	0	0
K238M	Arg	0.633	0.1258	0.601	0.5	0.638	0.643	0
	Cit	0.617	0.0536	0.252	0.5	0.660	0	0
K238P	Arg	0.633	0.1954	0.378	0.5	0.638	0.643	0
	Cit	0.617	0.1148	0.144	0.5	0.660	0	0
K238R	Arg	0.633	0	0.785	0.09	0.638	0.643	0.241
	Cit	0.617	0.7754	0.320	0.5	0.660	0	0.241
K238S	Arg	0.633	0.2046	0.448	0.5	0.638	0.643	0
	Cit	0.617	0.134	0.183	0.5	0.660	0	0
K238T	Arg	0.633	0.2106	0.482	0.5	0.638	0.643	0
	Cit	0.617	0.0928	0.197	0.5	0.660	0	0
K238V	Arg	0.633	0.6652	0.450	0.5	0.638	0.643	0
	Cit	0.617	0.942	0.144	0.5	0.660	0	0

Table S9. Summary of all ΔS values.

	$S(Arg)$	$S(Cit)$	ΔS
WT	0.4	0.4	0
K238N	0.511	0.383	0.128
K238C	0.503	0.359	0.144
K238W	0.47	0.351	0.119
K238E	0.758	0.505	0.253
K238Q	0.510	0.383	0.127
K238Y	0.477	0.352	0.125
K238A	0.376	0.271	0.105
K238F	0.449	0.304	0.145
K238D	0.731	0.487	0.244
K238G	0.374	0.268	0.106
K238H	0.434	0.323	0.111
K238L	0.430	0.305	0.125
K238I	0.433	0.292	0.141
K238M	0.448	0.305	0.143
K238P	0.427	0.314	0.113
K238R	0.433	0.445	0.012
K238S	0.438	0.299	0.139
K238T	0.444	0.310	0.134
K238V	0.504	0.409	0.095

Table S10. Comparison of PIToN-predicted signal contrast (ΔS) and experimentally measured discrimination resolution (R_s) for selected aerolysin nanopore mutants in citrullination discrimination.

Experimentally tested mutants							
Nanopore	**WT**	**K238N**	**K238Q**	**K238C**	**K238Y**	**K238W**	**K238E**
ΔS	0	0.128	0.127	0.144	0.125	0.119	0.253
R_s	0.01	0.55	0.49	0.84	0.62	0.82	0.90
Computationally screened mutants only							
Nanopore	**K238A**	**K238F**	**K238D**	**K238G**	**K238H**	**K238L**	**K238I**
ΔS	0.105	0.145	0.244	0.106	0.111	0.125	0.141
Nanopore	**K238M**	**K238P**	**K238R**	**K238S**	**K238T**	**K238V**	
ΔS	0.143	0.113	0.012	0.139	0.134	0.095	

1.4 A minimal empirical surrogate linking PIToN-derived ΔS to experimental R_s

To better connect the PIToN-derived signal-contrast descriptor ΔS with the experimentally measured discrimination resolution R_s, we tested a minimal empirical correction in which the contrast term was penalized by the mean binding-related descriptor. This extension was not intended as a universal model of nanopore resolution, but as a first-order surrogate to account for the empirical observation that pores with stronger average local interaction propensity may also exhibit broader blockade distributions.

The baseline model used a simple linear scaling of ΔS:

$$R_{s,pred} = a \cdot \Delta S$$

The empirical bridge model was defined as:

$$R_{s,pred} = a \cdot \frac{\Delta S}{1 + b \cdot \overline{\Psi_{Binding}}}$$

where $\overline{\Psi_{Binding}}$ is the mean normalized binding-related descriptor for the two analyte states in a given pore. This term was introduced as a minimal penalty reflecting the possibility that stronger average local interaction propensity may contribute to broader event distributions and therefore reduce effective experimental separability.

Across the seven experimentally tested mutant pores, the binding-penalty surrogate improved the approximation of experimental R_s relative to the baseline linear scaling of ΔS. In contrast, an analogous VDW-based penalty collapsed to the baseline solution, indicating that mean VDW was not an informative standalone proxy for broadening in the present pore set.

Table S11. Comparison of baseline and binding-penalty surrogate models for

approximating experimental R_s.

Model	Expression	SSE	RMSE
Baseline linear scaling	$R_{s,pred} = a \cdot \Delta S$	0.185	0.163
Binding-penalty surrogate	$R_{s,pred} = a \cdot \frac{\Delta S}{1 + b \cdot \overline{\Psi_{Binding}}}$	0.050	0.084

The binding-penalty surrogate reduced SSE (sum of squares due to error) by approximately 73.0% relative to the baseline model and reduced RMSE (root mean square error) by approximately 48.1%, indicating a substantially improved correspondence between PIToN-derived contrast and experimental separability.

Leave-one-out validation

To evaluate whether this empirical bridge retained predictive value beyond full-data fitting, a leave-one-out analysis was performed across the seven experimentally tested pores. In each round, one pore was excluded, the parameters a and b were refitted using the remaining six pores, and the excluded pore was then predicted.

Table S12. Leave-one-out cross-validation for predicted R_s using the binding-penalty surrogate method.

Left-out pore	Experimental R_s	Predicted R_s	Absolute error
WT	0.01	0.000	0.010
K238N	0.55	0.625	0.075
K238C	0.84	0.828	0.012
K238W	0.82	0.675	0.145

K238E	0.90	0.843	0.057
K238Y	0.62	0.660	0.040
K238Q	0.49	0.641	0.151

The leave-one-out analysis gave:

$$MAE\ (mean\ absolute\ error)\ =\ 0.0702$$

$$RMSE\ =\ 0.0885$$

These results suggest that the binding-penalty surrogate retained qualitative predictive value for held-out pores, although systematic deviations remained for specific mutants, particularly K238W and K238Q. In particular, K238W tended to be underpredicted, which may indicate that favorable aromatic-contact effects contributing to separability were not fully captured by the present mean-binding penalty term. K238Q, in contrast, remained overpredicted, suggesting that the current surrogate does not yet fully represent the distribution-broadening behavior of amide-containing pores.

Overall, these observations support the interpretation that the original PIToN ΔS term primarily captures signal-contrast tendency, whereas experimentally observed R_s additionally reflects broadening effects that are not explicitly represented in the base formulation. The present binding-penalty surrogate should therefore be viewed as a first-order empirical bridge rather than as a new universal predictive model. Because experimental R_s depends not only on signal separation but also on peak broadening, we further explored whether a minimal empirical penalty term could improve the

correspondence between PIToN-derived ΔS and experimental separability.

1.5 Relationship of frequency against concentration for $H4_{1\text{-}10}$ and $H4_{1\text{-}10}$R3(Cit)

To establish quantification curves for $H4_{1\text{-}10}$ and $H4_{1\text{-}10}$R3(Cit) using the R282A/K238E nanopore, both peptides were assayed at a gradient of concentrations. The final concentrations of the two peptides were set to 100, 300, 500, 800 and 1000 nM, respectively. Analyte solutions of known concentrations were introduced into the cis chamber, and current recordings were acquired under standard nanopore assay conditions. Within this concentration range, the event frequencies of $H4_{1\text{-}10}$ and $H4_{1\text{-}10}$R3(Cit) were linearly proportional to their respective concentrations.

2. Materials and Methods

2.1 Reagents and materials

1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC, powder, 99%) was purchased from Avanti Polar Lipids, Inc. (Alabaster, AL, USA). n-Decane (99%) was purchased from Thermo Fisher Scientific Inc. (Waltham, MA, USA). Trypsin agarose (T1763, buffered aqueous suspension, from bovine pancreas trypsin) and potassium chloride (for molecular biology, 99%) were purchased from Sigma-Aldrich (Shanghai) Trading Co., Ltd. 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES, reagent grade, 99.5%) was purchased from Sigma-Aldrich. All solutions were prepared using ultrapure water (18.2 M $\Omega \cdot cm^{-1}$ at 25 °C) from a Milli-Q system. The Ag/AgCl electrode was prepared by electroplating from a silver wire (1 mm in diameter) as the anode (a platinum wire served as the cathode) in 0.1 M HCl under a constant 1 mA current for 15 minutes. The ethanol extract of *Terminalia chebula* was prepared following the reported protocol,[5] and 40 mg of extract powder was weighed and dissolved in 1 ml of 50% dimethyl sulfoxide (DMSO) in water. The peptides were purchased from Anhui Guoping Pharmaceutical Co., LTD (Anhui, China) and were purified individually by HPLC (> 95%). N-Benzoyl-L-arginine Ethyl Ester (BAEE), antipyrine, 2, 3-butanedione monoxime, isopropyl β-D-thiogalactoside (IPTG), and glutathione were purchased from Shanghai Yuanye Biochemical Technology Co., Ltd (Shanghai, China). Dithiothreitol (DTT), ammonium iron sulfate dodecahydrate, and phosphoric acid were acquired from Shanghai McLean Biochemical Technology Co., Ltd (Shanghai, China).

2.2 Quality-control characterization of peptide samples

Vendor-supplied quality-control data, including high performance liquid chromatography (HPLC) chromatograms and electrospray ionization mass spectrometry (ESI-MS) spectra, supported the expected molecular identity and high analytical purity of the peptide samples used in this study (**Fig. S3-S12**). These data support that the samples predominantly consisted of the intended peptide species, although trace low-abundance impurities cannot be completely excluded.

2.3 Nanopore experiments

The genes of mutants (K238C, K238N, K238Q, K238Y, K238W, R282A, R282A/K238E, R282A/K238D, and R282A/K238E/K258A) were codon optimized and synthesized by GenScript, Inc. The details on the preparation of recombined wild-type proaerolysin and various mutants were described in previous studies.[6,7] Nanopore recordings were performed using a planar lipid bilayer setup with separate *cis* and *trans* chambers of 1 mL each. The two chambers were separated by a Delrin perfusion cup containing a single aperture of approximately 150 μm in diameter. A DPhPC lipid bilayer was formed across the aperture. Ag/AgCl electrodes were used to apply the transmembrane potential and record ionic currents. Unless otherwise specified, peptides were added to the cis compartment. The recording buffer consisted of 4 M KCl and 25 mM HEPES (pH 7.5). Unless otherwise specified, nanopore recordings were performed at an applied voltage of –50 mV, sampled at 250 kHz, and low-pass filtered at 5 kHz. Currents were recorded using an Axopatch 200B amplifier (Molecular

Devices, USA) and digitized with a 1550B digitizer (Molecular Devices, USA). Ionic current traces were analyzed using the ADEPT 2-state algorithm implemented in MOSAIC software[8] to extract event dwell time ($\Delta \boldsymbol{t}$) and relative current blockade depth ($\boldsymbol{I_b/I_o}$). Because blockade-depth estimation becomes less reliable for very short events under the finite temporal response of the recording system, only events with dwell times of at least 0.2 ms were retained for subsequent quantitative analysis. For all datasets, scatter plots, histograms, and Gaussian fitting were generated from the same filtered event set after exclusion of events with dwell times shorter than 0.2 ms.

2.4 Biolayer interferometry (BLI)

The affinity between $H4_{1\text{-}10}$ peptide and individual amino acids was measured through a Biolayer Interferometry (BLI) experiment using the Octet K2 System with amine reactive 2nd generation (AR2G) biosensors. Before the test, $H4_{1\text{-}10}$ peptide (1 $mg \cdot mL^{-1}$) was immobilized onto the carboxyl-terminated biosensors surface over 24 h via EDC/NHS-mediated coupling between carboxyl groups on the sensor and amine residues of the $H4_{1\text{-}10}$ peptide, performed in ultrapure water. After that, the sensor surface was thoroughly washed with ultrapure water to remove the unreacted sample. Subsequently, the peptide-modified biosensors were exposed to separate test solutions each containing a single amino acid (E, Q, Y, W, C, K) at a uniform concentration of 1 μM, followed by a dissociation phase in ultrapure water. All sensor preparation and interaction analyses were conducted at 25 °C. Data were processed using ForteBio Data Analysis 11.0 software.

The affinity between EE and peptides ($H4_{1\text{-}10}$ and $H4_{1\text{-}10}$R3(Cit)) was measured through a BLI experiment using the Octet K2 System with amine reactive 2nd generation (AR2G) biosensors. Before the test, EE (1 $mg \cdot mL^{-1}$) were immobilized onto the carboxyl-terminated biosensors surface for 24 h using a coupling reaction between the carboxyl group and the amine residue of EE, facilitated by EDC/NHS in ultrapure water. After that, the sensor surface was thoroughly washed with ultrapure water to remove the unreacted sample. Subsequently, EE-modified biosensors were exposed to peptide solutions with different concentrations (0.005, 0.03, 0.125, 0.25, and 1 μM), The interaction analysis was followed by the dissociation in ultrapure water. All sensor preparation and interaction analyses were conducted at 25 °C. Data were analyzed using ForteBio Data Analysis 11.0 software.

2.5 Molecular dynamics simulations

Model construction. The model constructions were similar to our previous studies. Briefly, the initial all-atom model of wild-type aerolysin nanopore embedded in a DPhPC membrane solvated in 1 M KCl with a 200 Å×200 Å×160 Å water box of TIP3P water was taken from previous studies.[9] The present simulations retained the inherited structural setup, including the default protonation-state assignment of titratable residues from the original model, and no independent reassignment of residue protonation states was performed. This fixed setup was used consistently across all simulated systems to preserve internal comparability between the WT and mutant pores. Mutant nanopores (K238E, R282A, R282A/K238E) were constructed by VMD mutator plugin.[10] The $H4_{1\text{-}}$

$_{10}$ peptide was built by Avogadro 1.2.0,[11] while H4$_{1\text{-}10}$R3(Cit) was built based on H4$_{1\text{-}10}$ model by changing the atoms. Only the citrulline was described by CHARMM22 force field from SwissSidechain,[12] while others were described by CHARMM36 force field. The peptides were placed with their centers of mass at the entrance of the nanopore and vertically oriented toward the channel. The initial peptide placement used in the simulations corresponded to the configuration illustrated in Figure 3a, with the peptide positioned at the *cis* entrance of the pore and aligned approximately along the pore axis.

MD simulations. All MD simulations were performed using NAMD2,[13,14] while visualization and analysis were performed by VMD.[10] All simulations used 2 fs time step. The particle-mesh Ewald method was used to evaluate long range electrostatic interactions with a 1 Å-spaced grid. Van der Waals forces were calculated using a cutoff of 12 Å and a switching distance of 10 Å. Periodic boundary conditions were applied in all directions. For the nanopore system, an initial energy minimization was performed with all alpha-carbon atoms of the protein restrained to their initial coordinates for 2000 steps. The system was then heated to 295 K and equilibrated in the NPT ensemble with all alpha-carbon atoms of the protein restrained to their initial coordinates. The combined protein and glycan system took 5000 minimization steps, followed by heating to 295 K and then equilibration. A transmembrane bias of –100 mV was applied along the z-axis and implemented in the simulation as a uniform electric field across the periodic simulation box. Constant-velocity steered molecular dynamics simulations (SMD) were applied by attaching a dummy atom to center of mass of

peptides using a virtual spring with a spring constant of 69.5 pN/nm. The dummy atom was driven transversely through the nanopore from *cis* to *trans* at a constant velocity of 0.5 $nm \cdot ns^{-1}$, corresponding to an SMD trajectory length of 20 ns. The NAMDEnergy plugin was used for calculating the non-bonded pair interactions between the nanopore and the peptide molecule based on the trajectories obtained from the SMD simulation. Peptide translocation trajectories were generated by SMD, and the corresponding open-pore and peptide-occupied ionic currents were estimated from the resulting pore-peptide conformations using the steric exclusion model (SEM), following the procedures described by Wilson et al.[15] and as implemented in Ouldali et al.[9] In this approach, SMD simulations provide the structural trajectories, whereas ionic currents are derived from SEM-based analysis of the resulting pore-peptide conformations. Accordingly, current blockade was obtained by comparing the SEM-predicted peptide-occupied current with the corresponding open-pore current under the same bias conditions.

Equilibrium MD simulations. To sample different configurations of the $H4_{1-10}$ peptide, two snapshots were extracted from the SMD trajectory when the peptide was located near the K238 region. Each snapshot was subjected to a 5 ns equilibrium MD simulation. The occupancies of the two peptide configurations were averaged over the total simulation time using the VMD volmap plugin.

2.6 PAD4 expression and purification

The recombinant human PAD4 protein fused with a glutathione S-transferase (GST) tag was expressed and purified as previously described.[5,16] Briefly, the cDNA encoding

human PAD4 was subcloned into a pGEX vector to generate a GST-PAD4 fusion construct. This plasmid was transformed into chemically competent Escherichia coli (*E.coli*) BL21(DE3) cells. The transformed cells were plated on the lysogeny broth (LB) agar containing ampicillin (100 μg/mL) and incubated overnight at 37 °C. A single colony was picked in 20 mL of LB liquid medium under 280 rpm shaking overnight at 37 ℃. Next, 10 mL of the initial culture solution was added to 200 mL of the LB liquid medium with ampicillin (100 mg/mL) and shaken overnight at 37 ℃. The protein expression was induced with a final concentration of 0.5 mM IPTG overnight at 16 ℃ when the OD_{600} value reached 0.6. The bacteria were then collected and cell pellets were resuspended in a lysis buffer containing 20 mM Tris-HCl (pH = 7.6), 200 mM NaCl, 10% glycerin, and sonicated at 200 W for 10 cycles (3 s pulse on, 5 s pulse off) in an ice-water bath. After that, the lysate was cleared by centrifugation at 20,000 *g* for 60 min at 4 ℃, and the resulting supernatant was purified by the Protino's glutathione agarose 4 B according to the manufacturer's protocol. Finally, the purified GST-PAD4 protein were stored in aliquots at -80 ℃.

2.7 Nanopore based PAD4-catalyzed citrullination reaction

Table S13. Composition of the PAD4 reaction buffer.

	1 M Tris-HCl (pH =7.6)	**1 M DTT**	**1 M NaCl**	**200 mM $CaCl_2$**	**dd H_2O**
Volume/μl	200	10	500	250	40
Final concentration/mM	200	10	500	50	-

Table S14. The PAD4-catalyzed citrullination reaction in Fig. 3k-m and Fig. 5b, e of the manuscript.

	$H4_{1-10}$ PAD4 reaction	Final concentration
10 mM Sample	3 μl	1.36 mM
0.2 mg/mL PAD4	15 μl	1.36 μM
Water	1 μl	-
Buffer	3 μl	-

Table S15. The PAD4-catalyzed citrullination reaction in addition with ethanol extract of *Terminalia chebula* in Fig. 3m the manuscript.

	10 mM $H4_{1-10}$	0.1 mg/mL PAD4	Water	Inhibitor	Buffer
Volume	2 μl	4 μl	2 μl	2 μl	2 μl
Final concentration	1.67 mM	0.33 μM	-	3.33 mg/mL	-

For the time-dependent reactions depicted in Fig. S12, these reactions were conducted in separate tubes and terminated by incubating in boiling water for 15 minutes. The reaction solution was then cooled to room temperature, and 4 μl of it was sampled for subsequent nanopore detection.

2.8 Machine learning (ML) analysis

Machine learning (ML)-based classification analyses were performed using Orange data-mining software (version 3.23).[17] Seven ML models were employed, namely: 1) AdaBoost; 2) k-Nearest Neighbors (kNN); 3) Logistic Regression; 4) Naïve Bayes; 5) Neural Network; 6) Random Forest; and 7) Support Vector Machine (SVM). After evaluating the area under the curve (AUC) and classification accuracy of each model, the Neural Network exhibited the highest classification performance. The trained Neural Network model was subsequently applied to predict the percentage of citrullination products generated in the PAD4-catalyzed reaction.

2.9 Herbal medical plant extracts preparation

Terminalia chebula herbal plant was purchased from Beijing Tongrentang Chinese Medicine Pharmacy Co., Ltd (Beijing, China). In this study, 75% ethanol-water was used to extract bioactive components of herbal plant extracts as previous described.[16] Briefly, 25 g (± 0.1 g) of freeze-drying plant powder were weighted and placed in a 250 mL round bottomed flask by addition of 150 mL of 75% ethanol. The solution was then sealed and soaked overnight. After that, the flask solutions were extracted three times by heating reflux at 85℃ for 2 h. The filtrates were merged and concentrated to about 50 mL at 60℃ with a rotary evaporator. The concentrated liquid was volatilized in a water bath at 70℃ until no fluid can be seen, and then freeze-dried to obtain dry plant ethanol extract. Finally, the plant extracts were ground into powder and stored at -80℃.

A certain amount (20 mg) herbal plant extracts powder was weighed into a 1.5 mL centrifuge tube, followed by addition of 0.5 mL of 50% dimethyl sulfoxide (DMSO). The tube was then dissolved by vortex and ultrasound for 20 min. After that the extracts solution was cleared by centrifugation at 25,000 g for 10 min at 10 ℃, the supernatant was used as the treated stock extract solution of plant with concentration of 20 mg/mL. All the plant extracts reconstitution solutions were stored at -80 ℃ for further use.

2.10 PAD4 activity assay using a colorimetric method

PAD4 activity was determined using a previously described colorimetric assay method with slight modification.[16] Briefly, the reaction buffer containing 50 mM Tris-HCl, 2 mM BAEE, 5 mM Ca^{2+} and 100 nM PAD4 with a total volume of 100 μL was incubated

at 37 ℃ for 1 h. Then, the reaction was quenched by adding 200 μL of freshly prepared color reagent containing 5 mM antipyrine, 20 mM diacetylmonoxime, 1.2 mM ammonium iron (III) sulfate dodecahydrate, 25% sulfuric acid, and 15% phosphoric acid. Then, the solution was incubated at 95 ℃ for 30 min for color development. The absorbance intensity was detected at 490 nm by a SuPerMax 3000FA multifunctional microplate reader (Shanghai Flash Spectrum Biotechnology Co., Ltd).

Assessment of peptides catalytic activity on PAD4 based the colorimetric method

The assessment method was performed following the colorimetric method with slight modification. Briefly, the synthetic peptides (either $H3_{1\text{-}10}$ or $H3_{1\text{-}10}$K9(Me3)) were used as substrates to determine PAD4 activity. To minimize the peptide consumption, the reaction volume was reduced to 25 μL. The reaction mixture comprised 50 mM Tris-HCl, 400 nM PAD4, 5 mM Ca^{2+}, either 6 mM or 12 mM synthetic peptide. After incubation at 37 °C for 2 h, 100 μL of the color reagent was added to quench the reaction. The samples were then heated at 95 °C for 30 min for color development prior to absorbance measurement.

3. Supplementary Figures

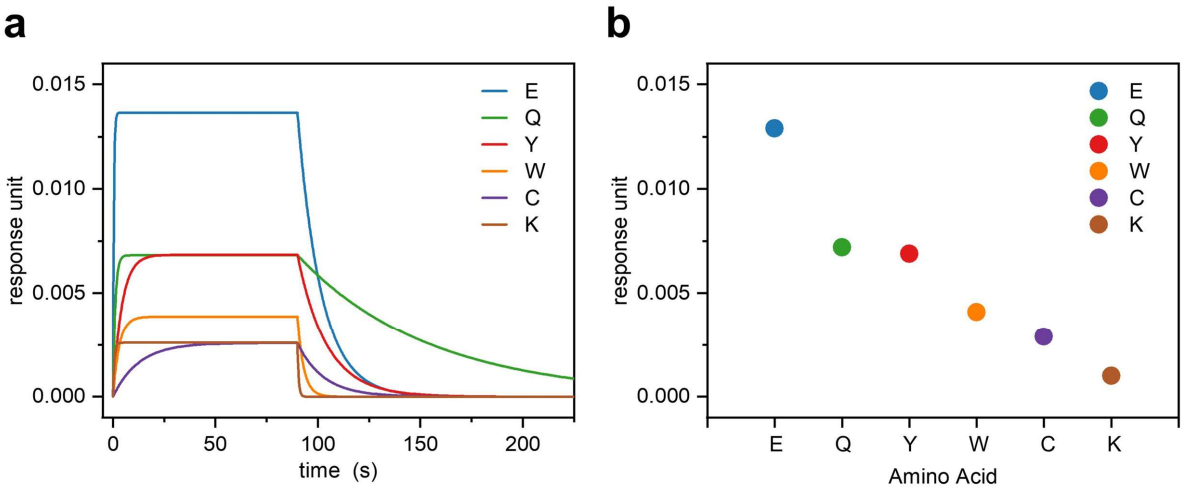


Figure S1. Binding kinetics of individual amino acids on a BLI sensor surface immobilized with H41-10 peptides. The binding affinity trend determined from BLI measurements is consistent with that computed for the binding-related term of the PIToN model via DFT calculations.

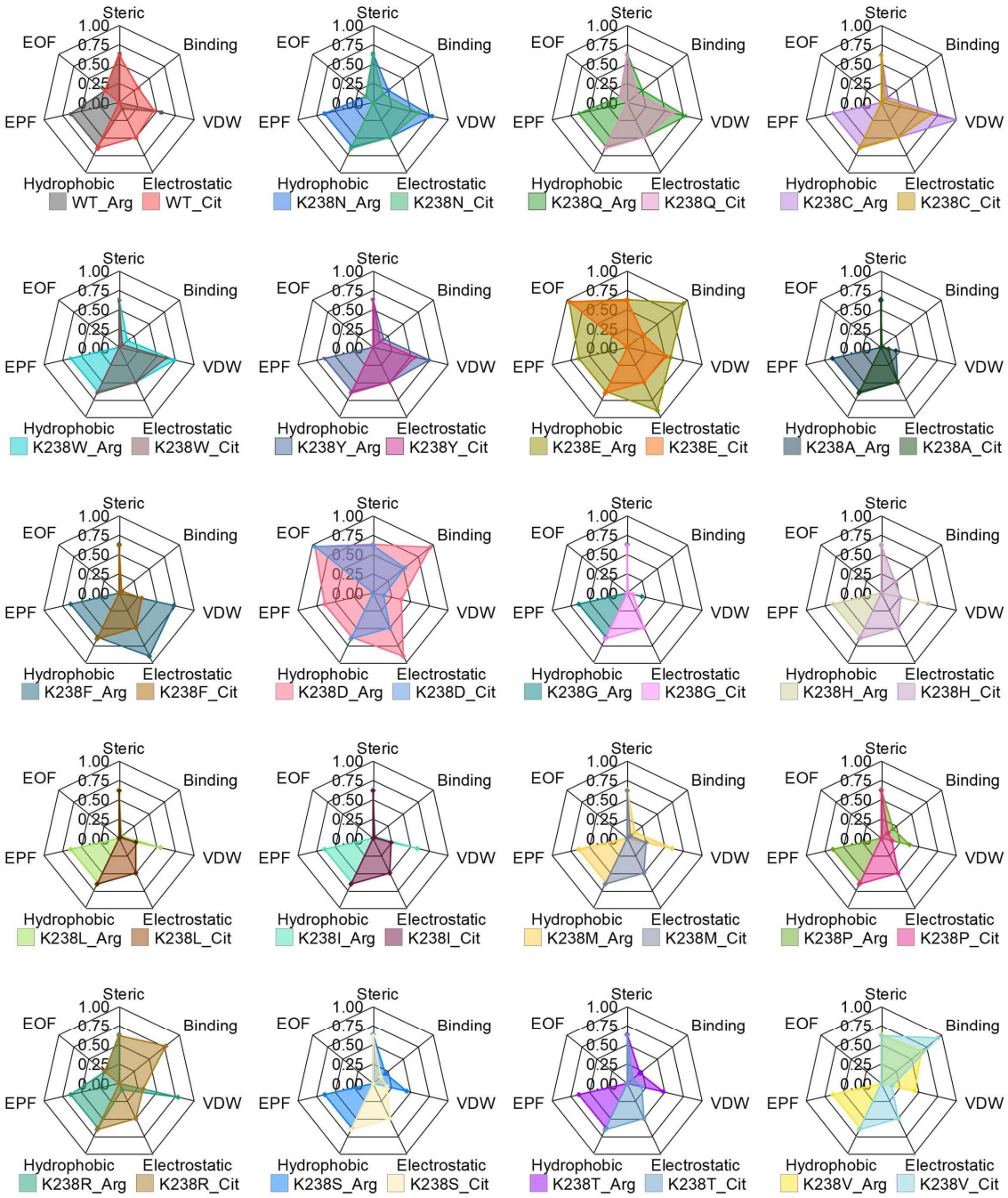


Figure S2. Calculated radar plots with various mutants for citrullination discrimination by our PIToN model.

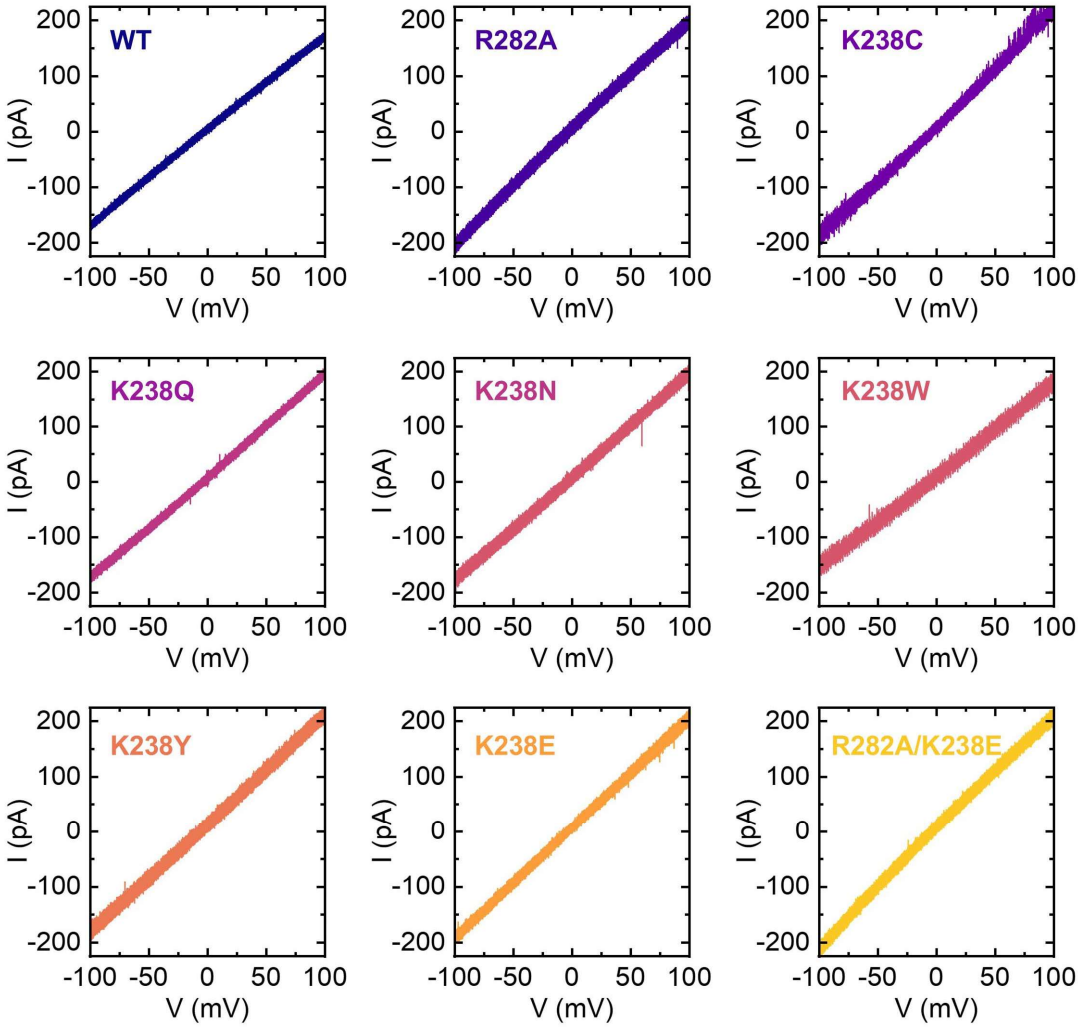
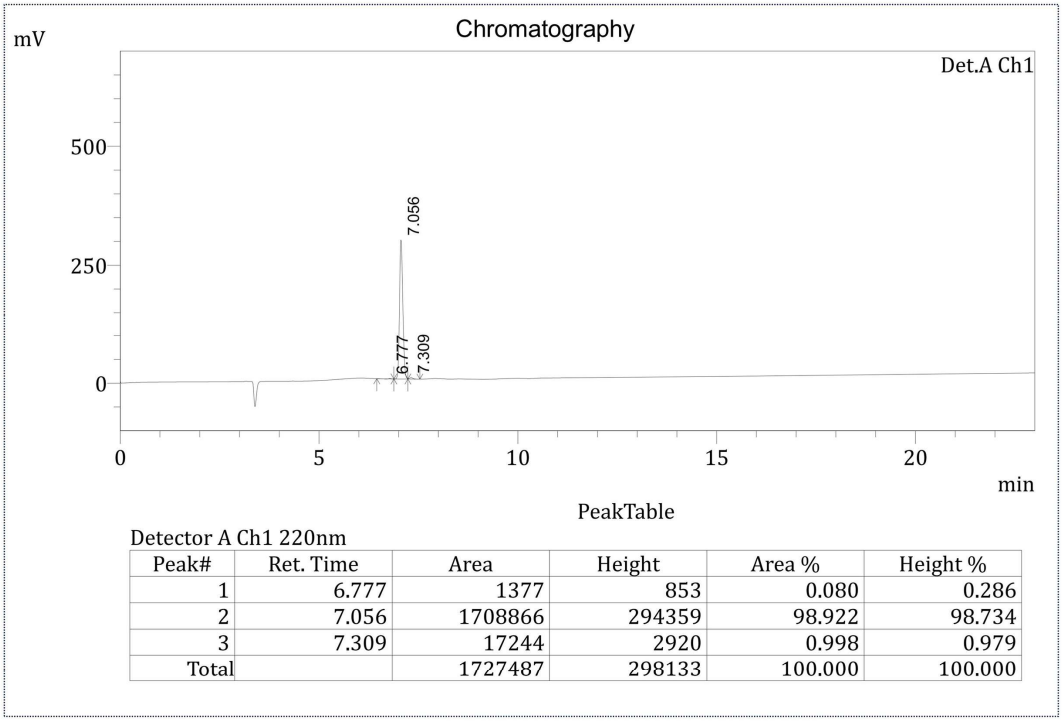


Figure S3. I-V plots for various mutant nanopores. All experiments were conducted in 4 M KCl, 25 mM HEPES, pH 7.5, at 25 ºC, with an applied voltage of –50 mV.

Sample: $H4_{1-10}$ SGRGKGGKGL



PeakTable

Detector A Ch1 220nm

Peak#	Ret. Time	Area	Height	Area %	Height %
1	6.777	1377	853	0.080	0.286
2	7.056	1708866	294359	98.922	98.734
3	7.309	17244	2920	0.998	0.979
Total		1727487	298133	100.000	100.000

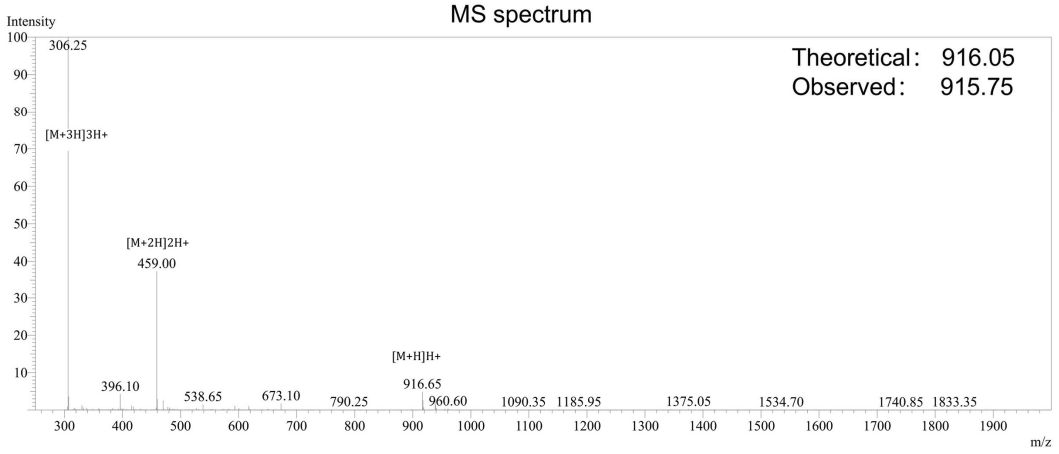
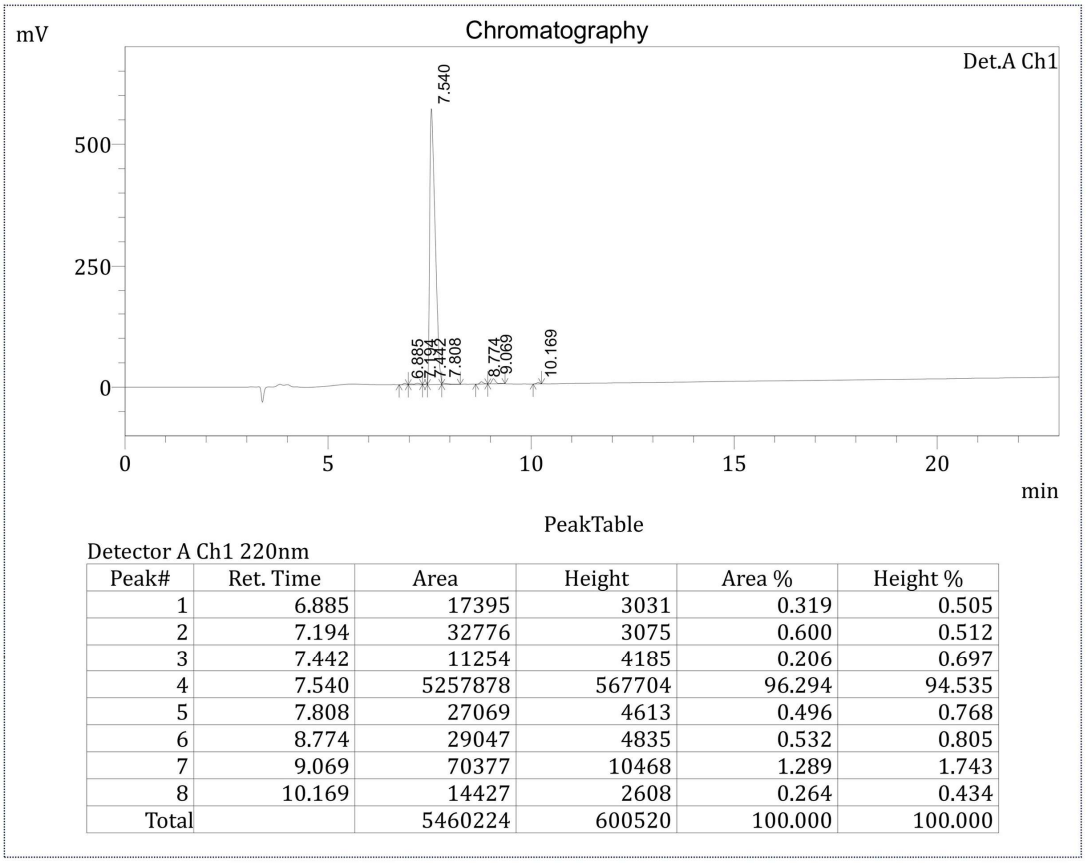


Figure S4. Vendor-supplied HPLC chromatogram of $H4_{1-10}$, showing a single dominant peak with an analytical purity of 98.922%. Vendor-supplied ESI-MS spectrum of $H4_{1-10}$, consistent with the expected molecular mass of the target peptide.

Sample: $H4_{1-10}R3(Cit)$ SG-R(Cit)-GKGGKGL



PeakTable

Detector A Ch1 220nm

Peak#	Ret. Time	Area	Height	Area %	Height %
1	6.885	17395	3031	0.319	0.505
2	7.194	32776	3075	0.600	0.512
3	7.442	11254	4185	0.206	0.697
4	7.540	5257878	567704	96.294	94.535
5	7.808	27069	4613	0.496	0.768
6	8.774	29047	4835	0.532	0.805
7	9.069	70377	10468	1.289	1.743
8	10.169	14427	2608	0.264	0.434
Total		5460224	600520	100.000	100.000

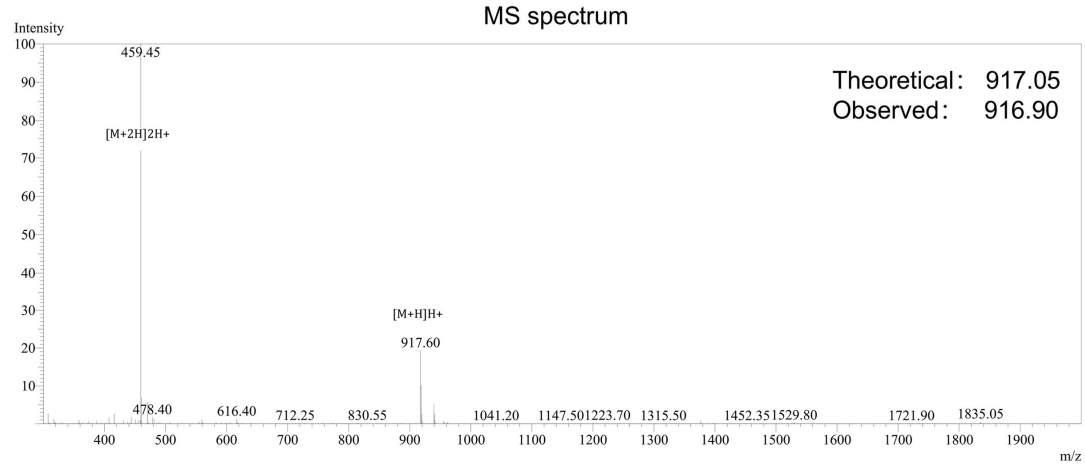
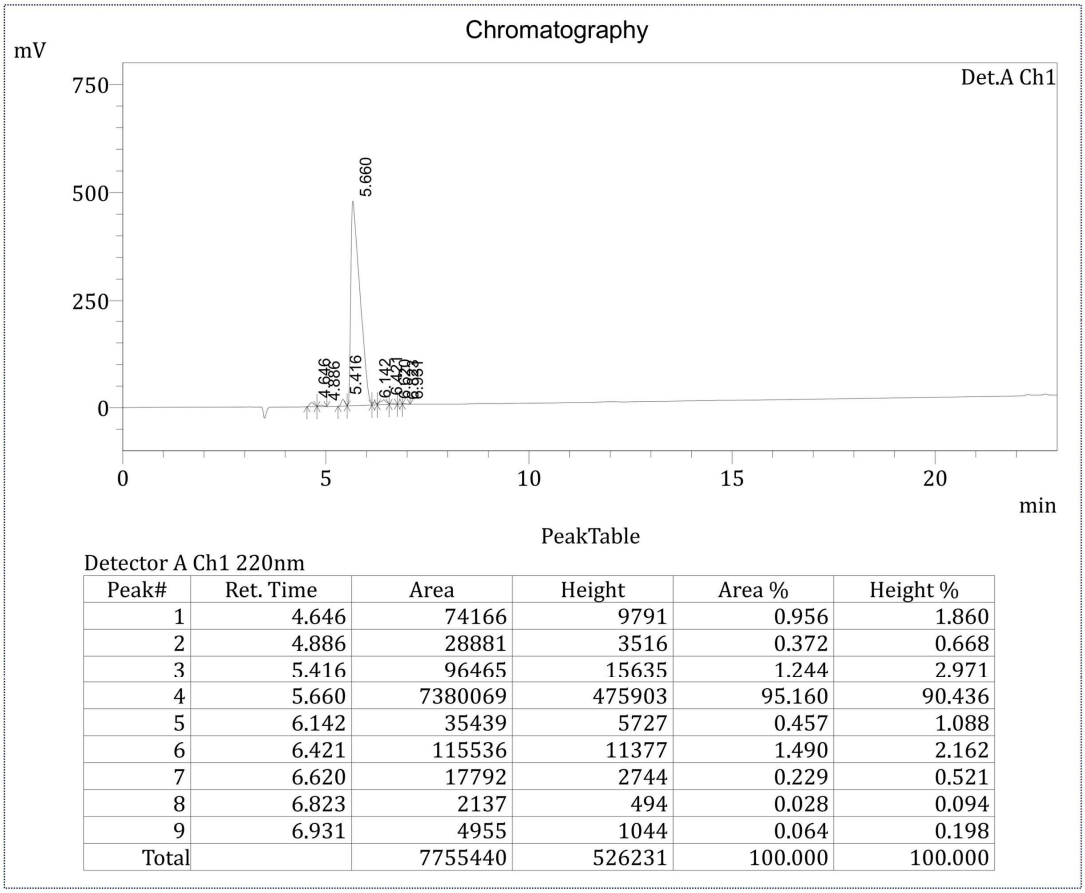


Figure S5. Vendor-supplied HPLC chromatogram of $H4_{1-10}R3(Cit)$, showing a single dominant peak with an analytical purity of 96.294%. Vendor-supplied ESI-MS spectrum of $H4_{1-10}R3(Cit)$, consistent with the expected molecular mass of the target peptide.

Sample: $H3_{1-10}$ ARTKQTARKS



PeakTable

Detector A Ch1 220nm

Peak#	Ret. Time	Area	Height	Area %	Height %
1	4.646	74166	9791	0.956	1.860
2	4.886	28881	3516	0.372	0.668
3	5.416	96465	15635	1.244	2.971
4	5.660	7380069	475903	95.160	90.436
5	6.142	35439	5727	0.457	1.088
6	6.421	115536	11377	1.490	2.162
7	6.620	17792	2744	0.229	0.521
8	6.823	2137	494	0.028	0.094
9	6.931	4955	1044	0.064	0.198
Total		7755440	526231	100.000	100.000

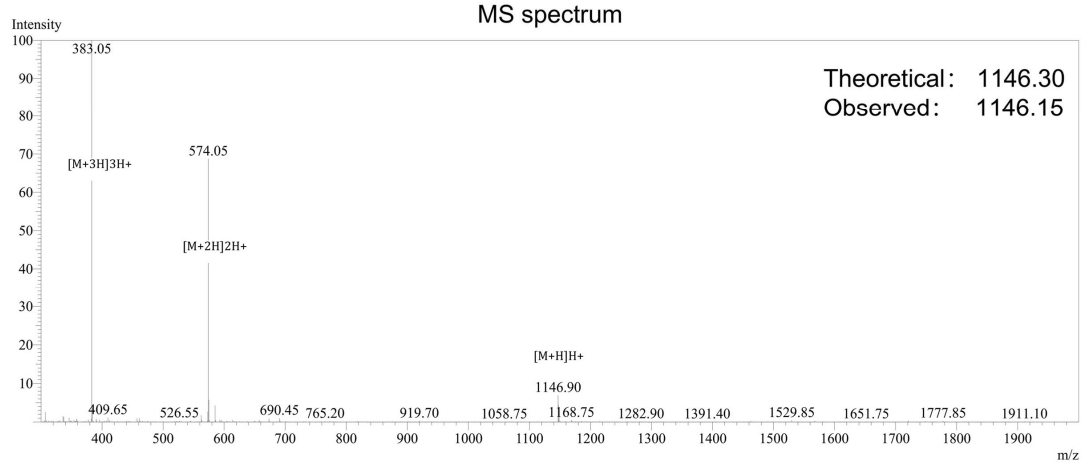
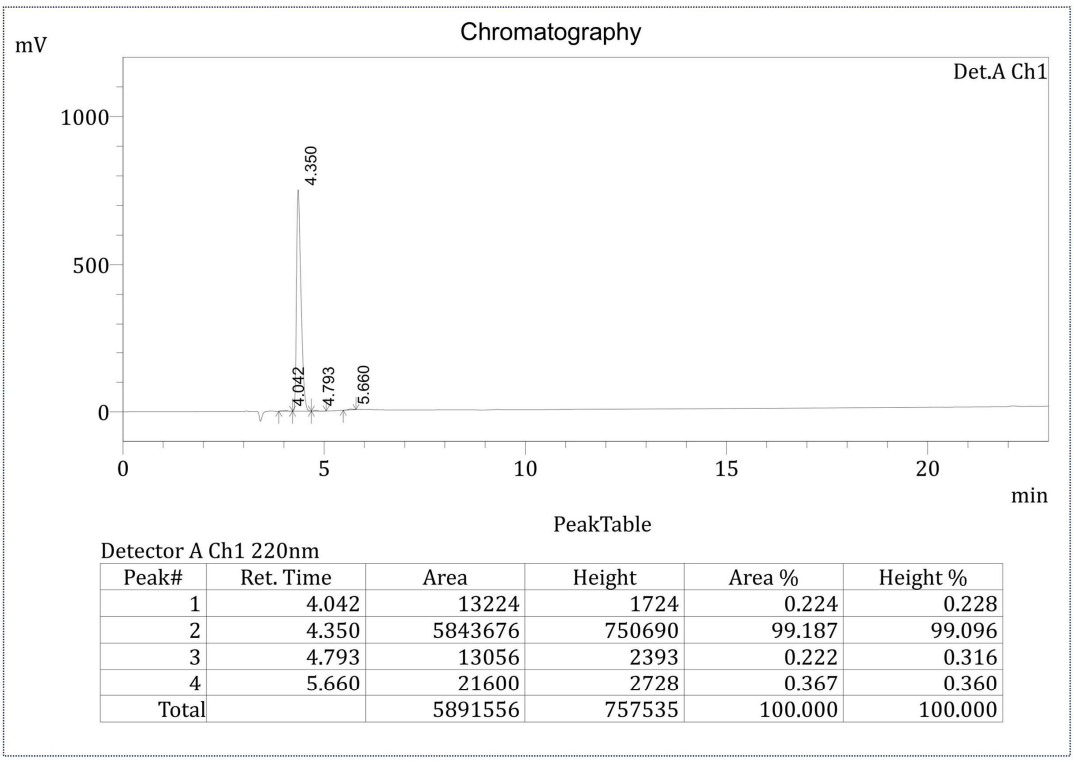


Figure S6. Vendor-supplied HPLC chromatogram of $H3_{1-10}$, showing a single dominant peak with an analytical purity of 95.160%. Vendor-supplied ESI-MS spectrum of $H3_{1-10}$, consistent with the expected molecular mass of the target peptide.

Sample: $H3_{1-10}R2(Cit)$ A-R(Cit)-TKQTARKS



PeakTable

Detector A Ch1 220nm

Peak#	Ret. Time	Area	Height	Area %	Height %
1	4.042	13224	1724	0.224	0.228
2	4.350	5843676	750690	99.187	99.096
3	4.793	13056	2393	0.222	0.316
4	5.660	21600	2728	0.367	0.360
Total		5891556	757535	100.000	100.000

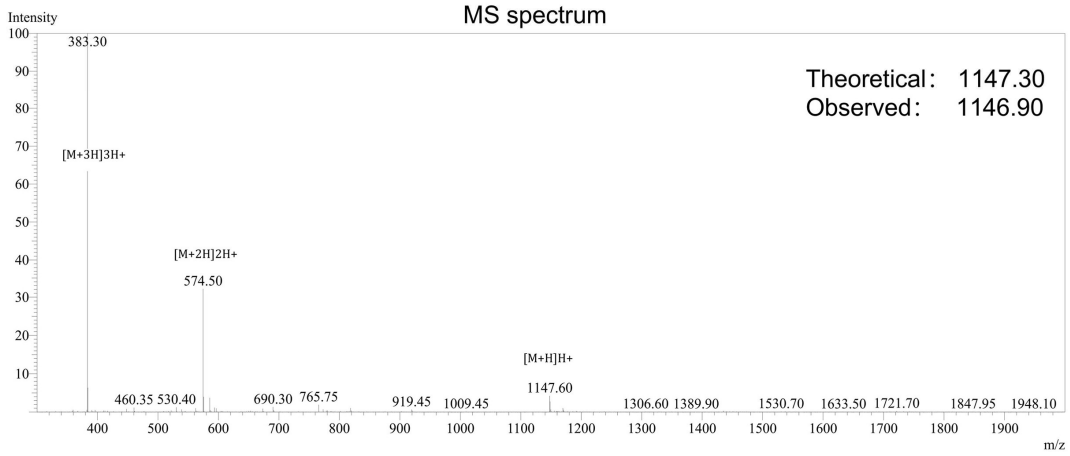
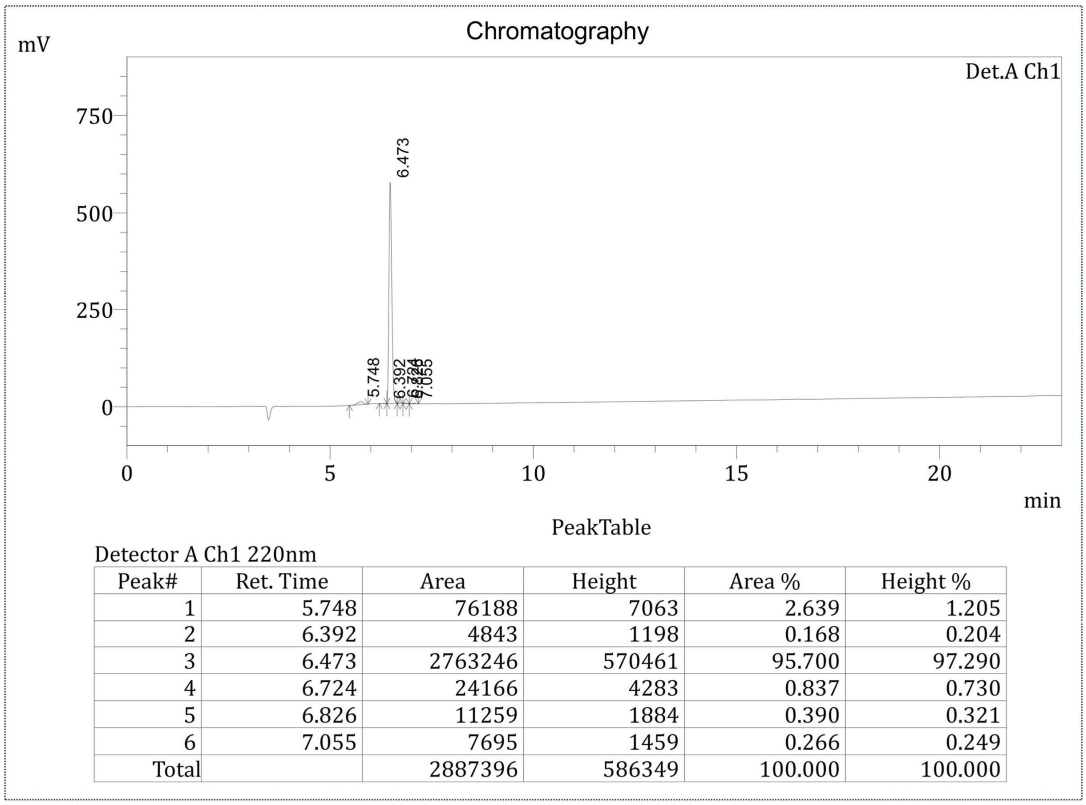


Figure S7. Vendor-supplied HPLC chromatogram of $H3_{1-10}R2(Cit)$, showing a single dominant peak with an analytical purity of 99.187%. Vendor-supplied ESI-MS spectrum of $H3_{1-10}R2(Cit)$, consistent with the expected molecular mass of the target peptide.

Sample: $H3_{1\text{-}10}R2(Me)$ A-R(Me)-TKQTARKS



PeakTable

Detector A Ch1 220nm

Peak#	Ret. Time	Area	Height	Area %	Height %
1	5.748	76188	7063	2.639	1.205
2	6.392	4843	1198	0.168	0.204
3	6.473	2763246	570461	95.700	97.290
4	6.724	24166	4283	0.837	0.730
5	6.826	11259	1884	0.390	0.321
6	7.055	7695	1459	0.266	0.249
Total		2887396	586349	100.000	100.000

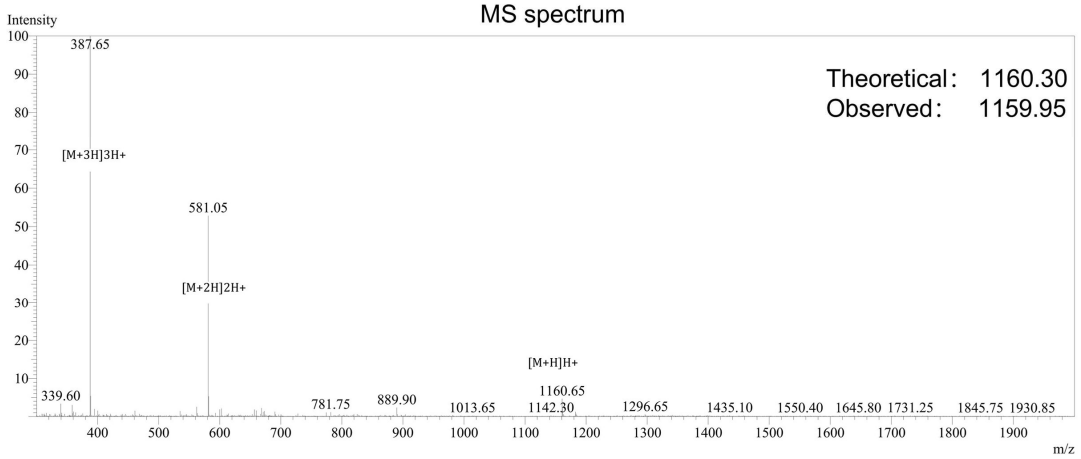


Figure S8. Vendor-supplied HPLC chromatogram of $H3_{1\text{-}10}R2(Me)$, showing a single dominant peak with an analytical purity of 95.700%. Vendor-supplied ESI-MS spectrum of $H3_{1\text{-}10}R2(Me)$, consistent with the expected molecular mass of the target peptide.

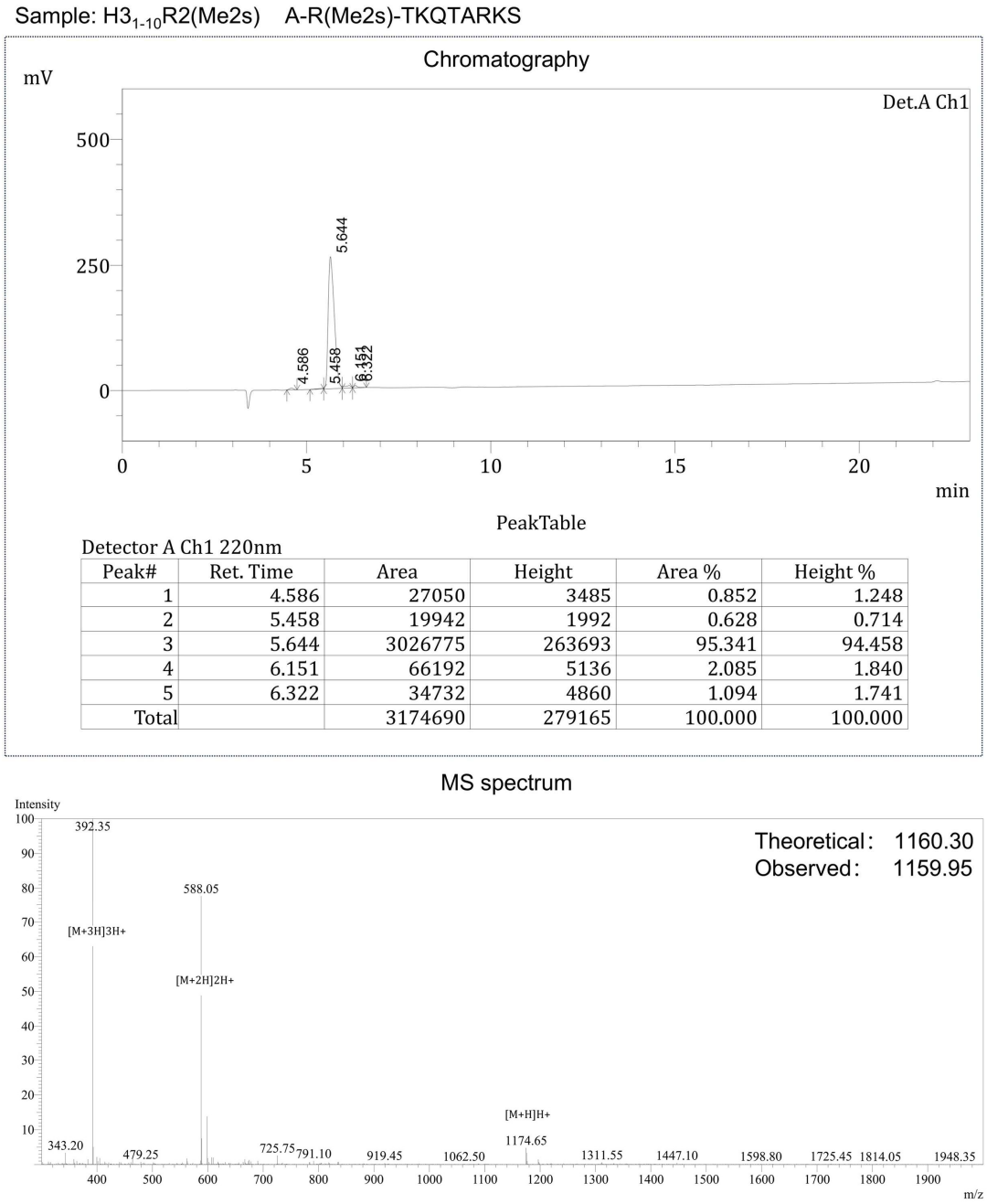
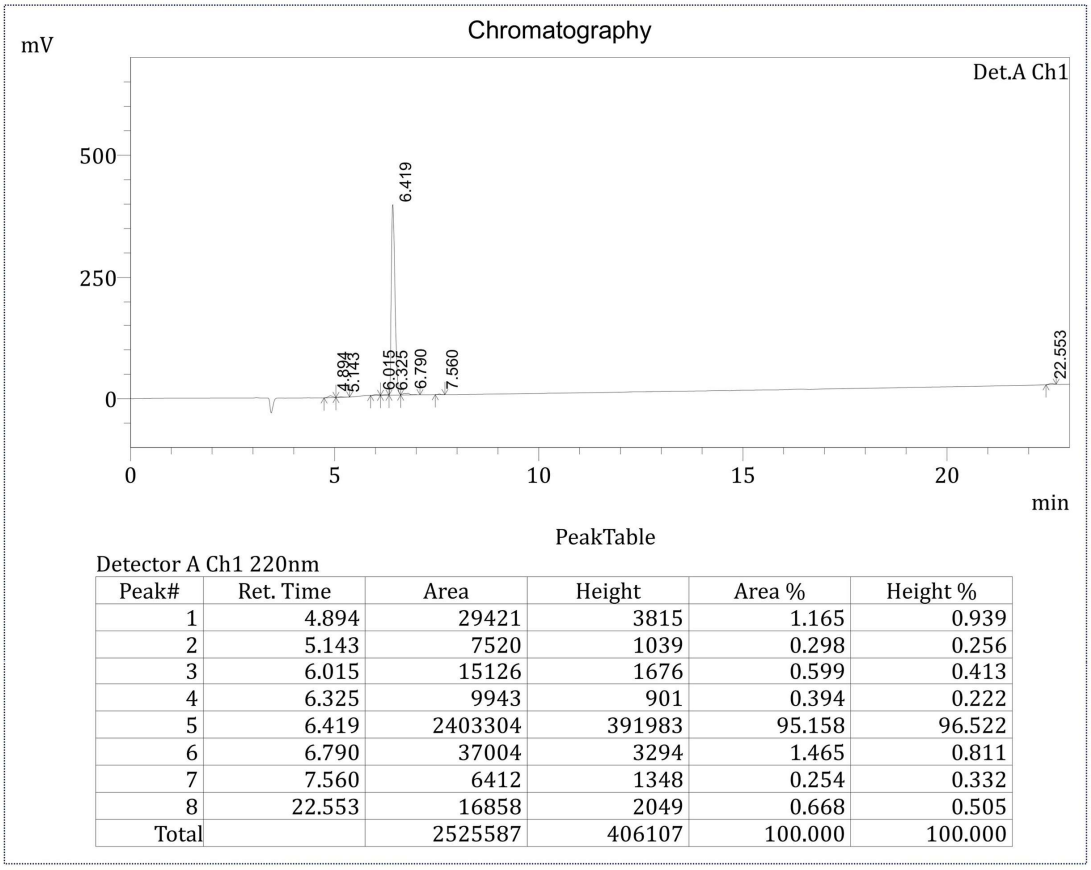


Figure S9. Vendor-supplied HPLC chromatogram of H3$_{1-10}$R2(Me2s), showing a single dominant peak with an analytical purity of 95.341%. Vendor-supplied ESI-MS spectrum of H3$_{1-10}$R2(Me2s), consistent with the expected molecular mass of the target peptide.

Sample: $H3_{1-10}R2(Cit)K9(Me3)$ A-R(Cit)-TKQTAR-K(Me3)-S



PeakTable

Detector A Ch1 220nm

Peak#	Ret. Time	Area	Height	Area %	Height %
1	4.894	29421	3815	1.165	0.939
2	5.143	7520	1039	0.298	0.256
3	6.015	15126	1676	0.599	0.413
4	6.325	9943	901	0.394	0.222
5	6.419	2403304	391983	95.158	96.522
6	6.790	37004	3294	1.465	0.811
7	7.560	6412	1348	0.254	0.332
8	22.553	16858	2049	0.668	0.505
Total		2525587	406107	100.000	100.000

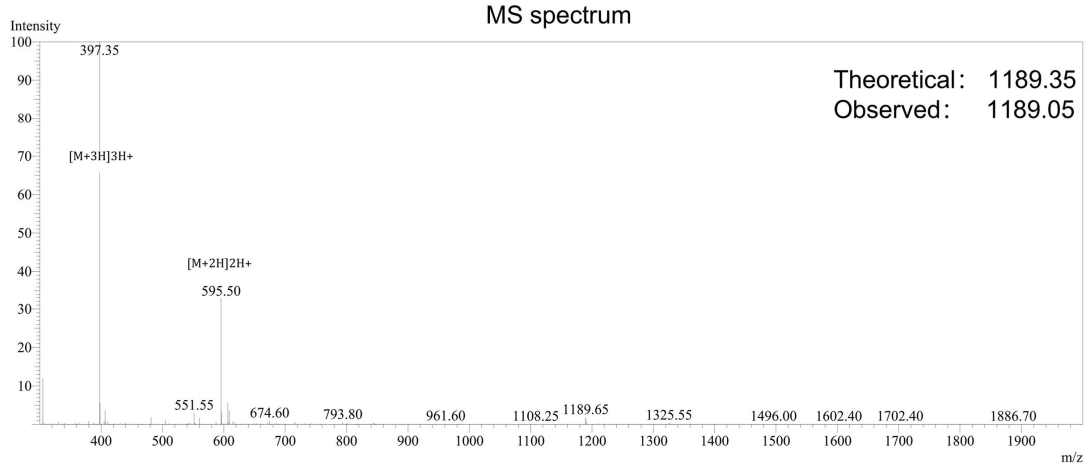
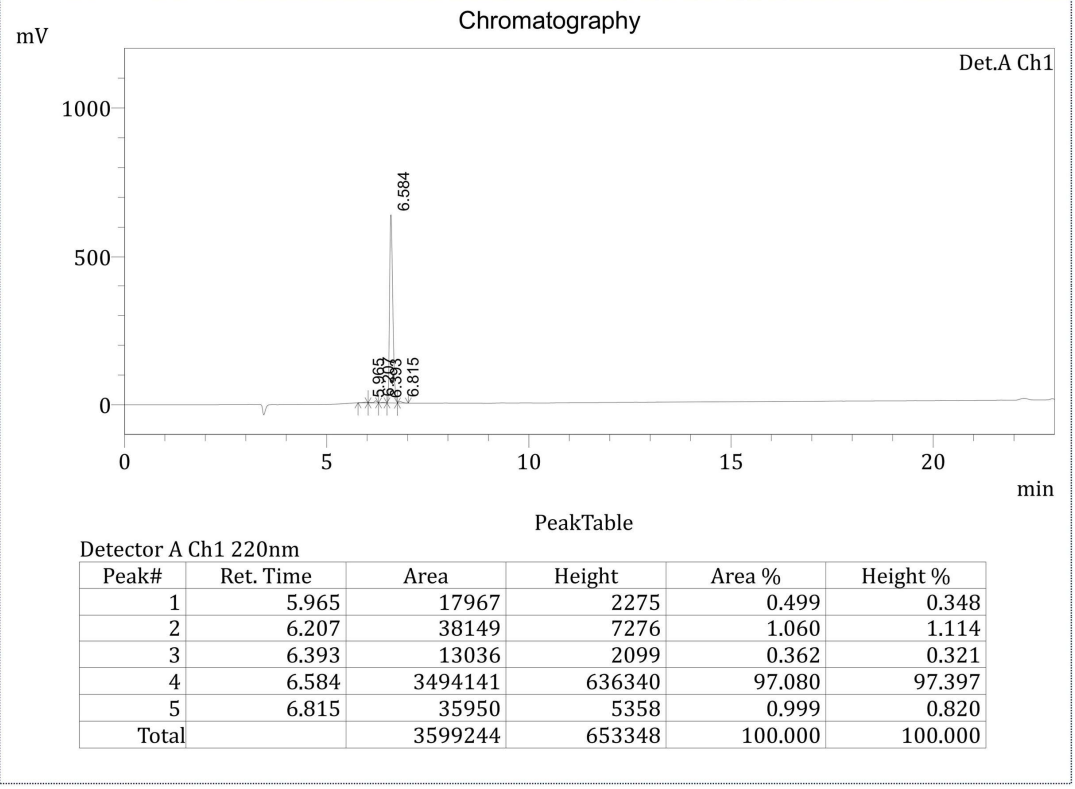


Figure S10. Vendor-supplied HPLC chromatogram of $H3_{1-10}R2(Cit)K9(Me3)$, showing a single dominant peak with an analytical purity of 95.158%. Vendor-supplied ESI-MS spectrum of $H3_{1-10}R2(Cit)K9(Me3)$, consistent with the expected molecular mass of the target peptide.

Sample: $H3_{1-10}K9(Me3)$ ARTKQTAR-K(Me3)-S



PeakTable

Detector A Ch1 220nm

Peak#	Ret. Time	Area	Height	Area %	Height %
1	5.965	17967	2275	0.499	0.348
2	6.207	38149	7276	1.060	1.114
3	6.393	13036	2099	0.362	0.321
4	6.584	3494141	636340	97.080	97.397
5	6.815	35950	5358	0.999	0.820
Total		3599244	653348	100.000	100.000

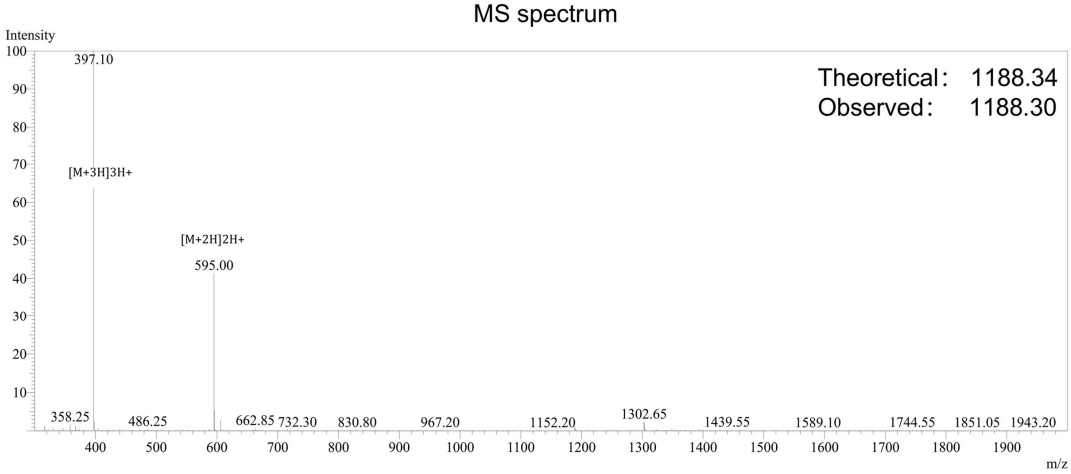
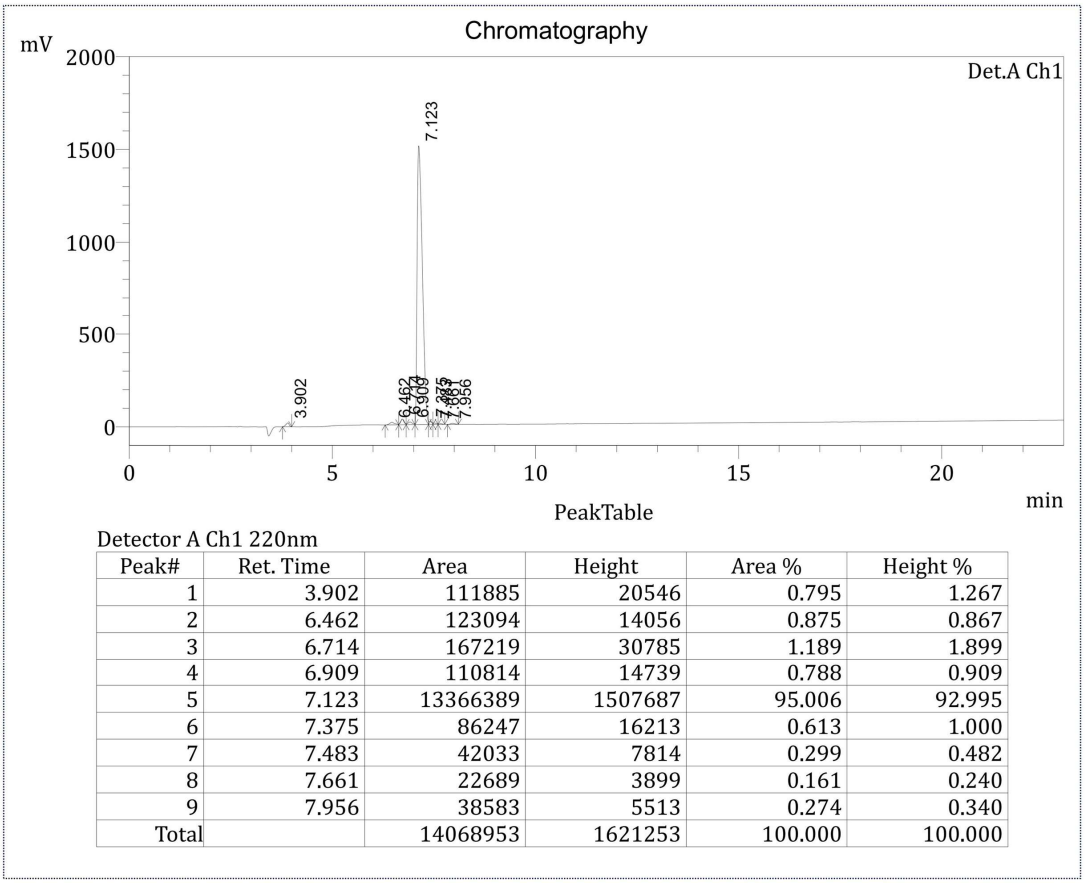


Figure S11. Vendor-supplied HPLC chromatogram of $H3_{1-10}K9(Me3)$, showing a single dominant peak with an analytical purity of 97.080%. Vendor-supplied ESI-MS spectrum of $H3_{1-10}K9(Me3)$, consistent with the expected molecular mass of the target peptide.

Sample: $H3_{1-10}$R2(Me)K9(Me3) A-R(Me)-TKQTAR-K(Me3)-S



Detector A Ch1 220nm

Peak#	Ret. Time	Area	Height	Area %	Height %
1	3.902	111885	20546	0.795	1.267
2	6.462	123094	14056	0.875	0.867
3	6.714	167219	30785	1.189	1.899
4	6.909	110814	14739	0.788	0.909
5	7.123	13366389	1507687	95.006	92.995
6	7.375	86247	16213	0.613	1.000
7	7.483	42033	7814	0.299	0.482
8	7.661	22689	3899	0.161	0.240
9	7.956	38583	5513	0.274	0.340
Total		14068953	1621253	100.000	100.000

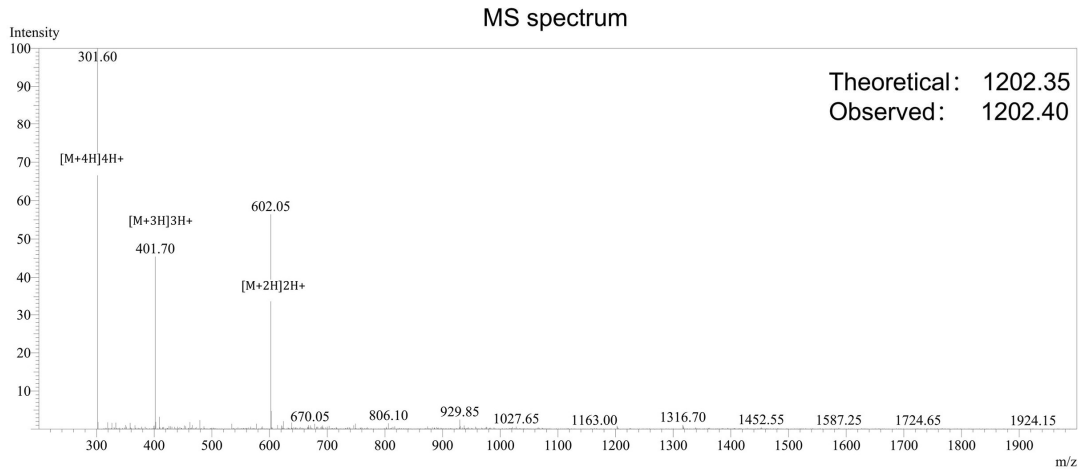
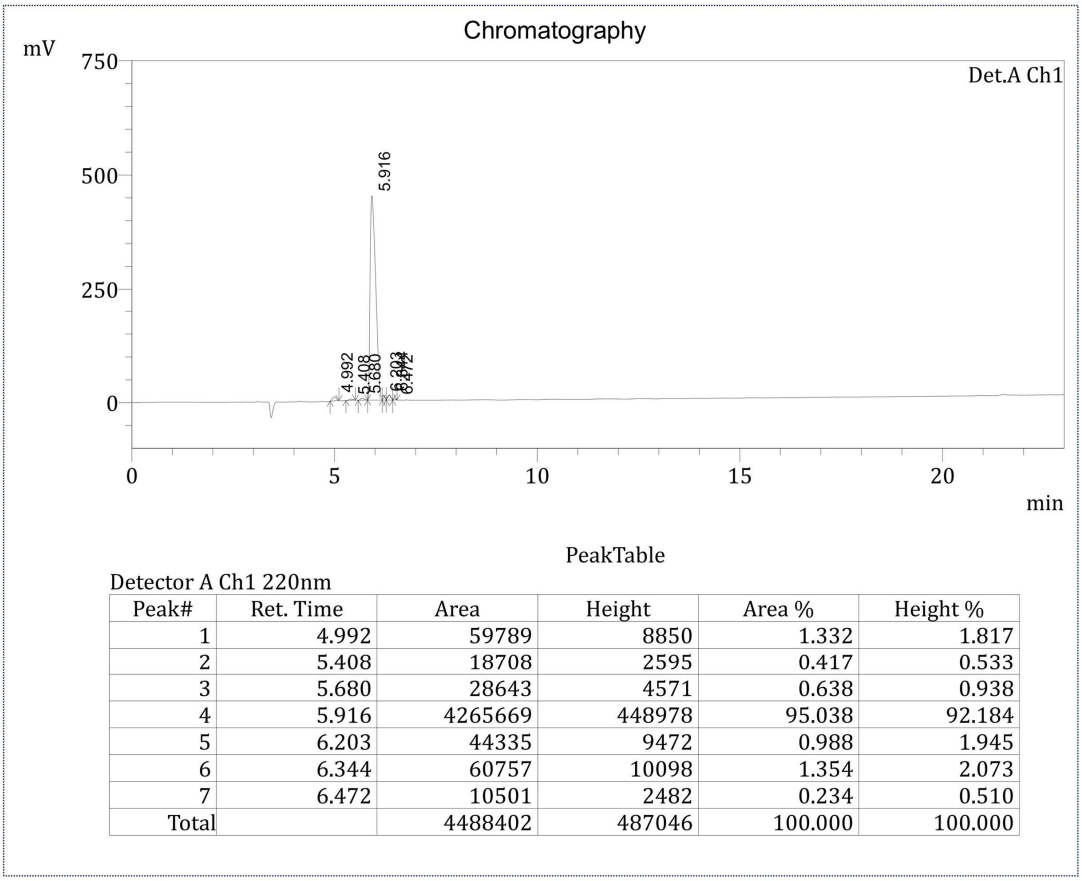


Figure S12. Vendor-supplied HPLC chromatogram of $H3_{1-10}$R2(Me)K9(Me3), showing a single dominant peak with an analytical purity of 95.006%. Vendor-supplied ESI-MS spectrum of $H3_{1-10}$R2(Me)K9(Me3), consistent with the expected molecular mass of the target peptide.

Sample: $H3_{1-10}R2(Me2s)K9(Me3)$ A-R(Me2s)-TKQTAR-K(Me3)-S



PeakTable

Detector A Ch1 220nm

Peak#	Ret. Time	Area	Height	Area %	Height %
1	4.992	59789	8850	1.332	1.817
2	5.408	18708	2595	0.417	0.533
3	5.680	28643	4571	0.638	0.938
4	5.916	4265669	448978	95.038	92.184
5	6.203	44335	9472	0.988	1.945
6	6.344	60757	10098	1.354	2.073
7	6.472	10501	2482	0.234	0.510
Total		4488402	487046	100.000	100.000

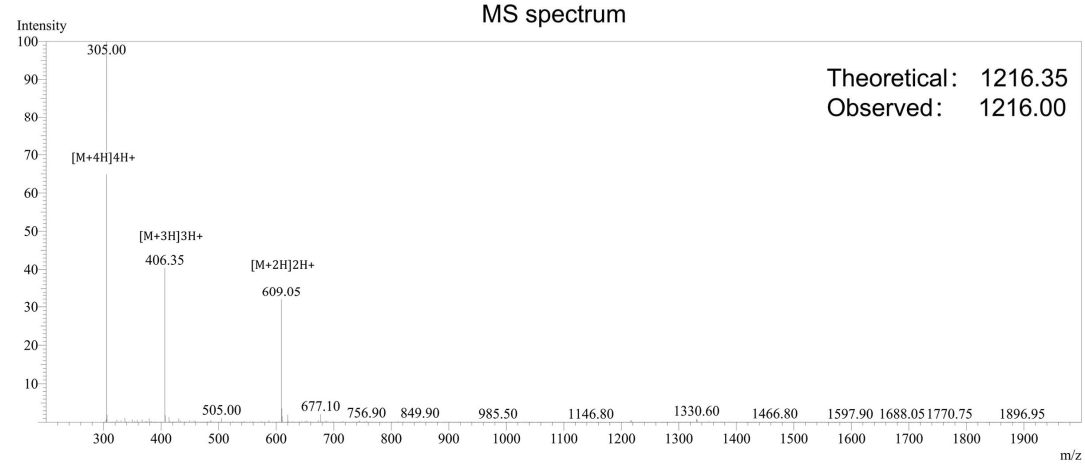


Figure S13. Vendor-supplied HPLC chromatogram of $H3_{1-10}R2(Me2s)K9(Me3)$, showing a single dominant peak with an analytical purity of 95.038%. Vendor-supplied ESI-MS spectrum of $H3_{1-10}R2(Me2s)K9(Me3)$, consistent with the expected molecular mass of the target peptide.

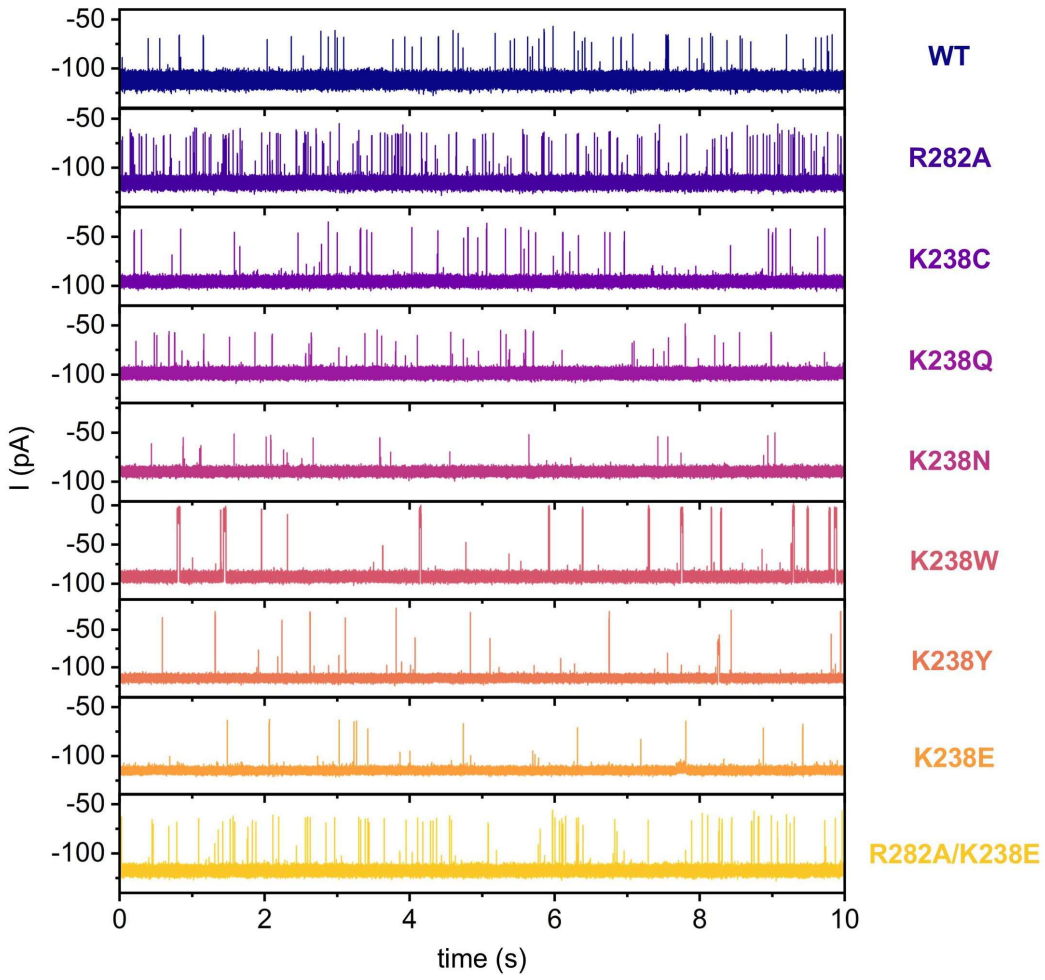


Figure S14. Current traces for $H4_{1\text{-}10}$ with various mutant nanopores. All experiments were conducted in 4 M KCl, 25 mM HEPES, pH 7.5, at 25 °C, with an applied voltage of –50 mV. The concentration of $H4_{1\text{-}10}$ was 20 μM.

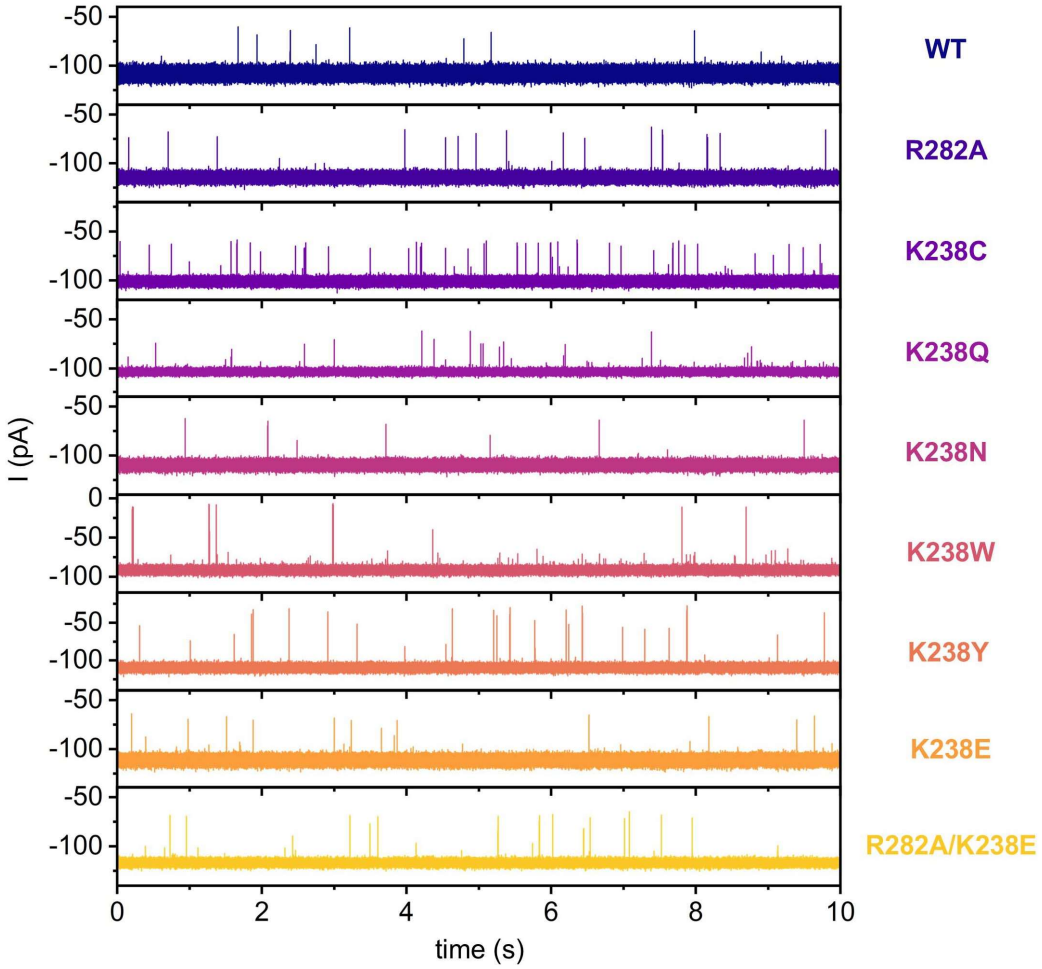


Figure S15. Current traces for $H4_{1\text{-}10}R3(Cit)$ with various mutant nanopores. All experiments were conducted in 4 M KCl, 25 mM HEPES, pH 7.5, at 25 ºC, with an applied voltage of –50 mV. The concentration of $H4_{1\text{-}10}R3(Cit)$ was 20 µM.

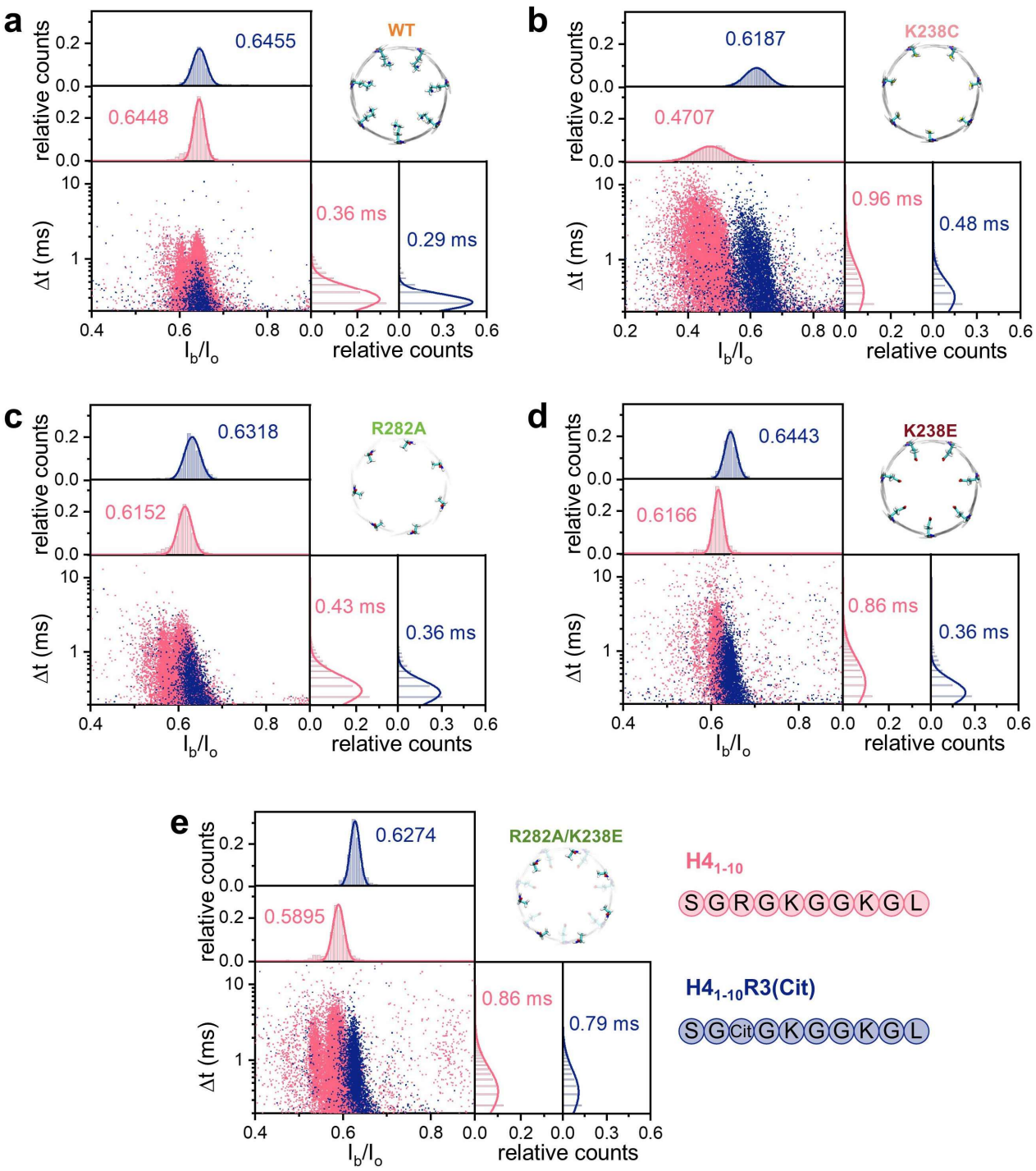


Figure S16. Statistical data for $H4_{1-10}$ and $H4_{1-10}$R3(Cit) through (a) WT, (b) K238C, (c) R282A, (d) K238E and (e) R282A/K238E. All experiments were conducted in 4 M KCl, 25 mM HEPES, pH 7.5, at 25 °C, with an applied voltage of –50 mV. The concentrations for both $H4_{1-10}$ and $H4_{1-10}$R3(Cit) were 20 μM. Scatter plots and the corresponding histograms/Gaussian fittings were generated from the same filtered

event set; events with dwell times shorter than 0.2 ms were excluded from both the scatter-plot display and the histogram analysis.

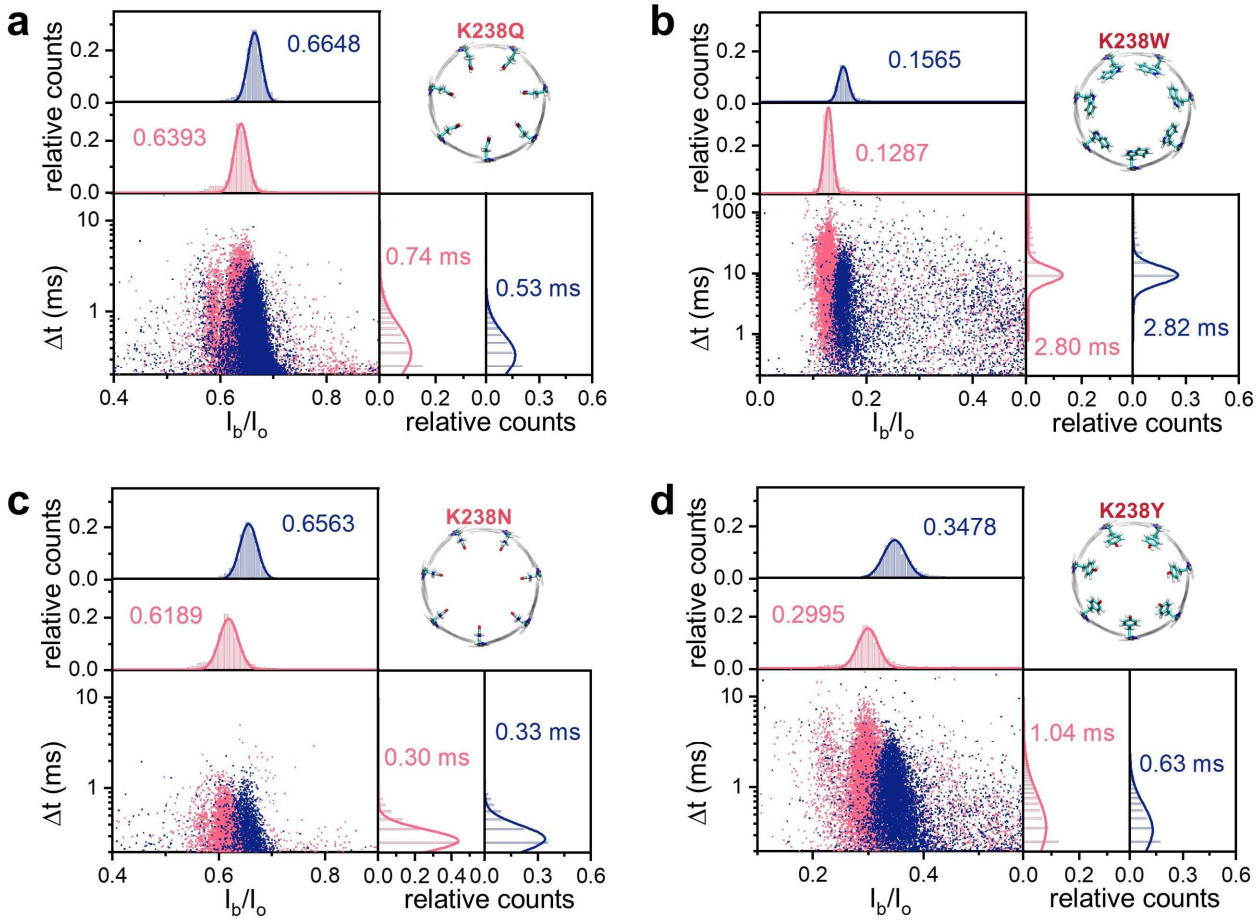


Figure S17. Statistical data for $H4_{1\text{-}10}$ and $H4_{1\text{-}10}R3(Cit)$ through (a) K238Q, (b) K238w, (c) K238N and (d) K238Y. All experiments were conducted in 4 M KCl, 25 mM HEPES, pH 7.5, at 25 ºC, with an applied voltage of –50 mV. The concentrations for both $H4_{1\text{-}10}$ and $H4_{1\text{-}10}R3(Cit)$ were 20 μM. Scatter plots and the corresponding histograms/Gaussian fittings were generated from the same filtered event set; events with dwell times shorter than 0.2 ms were excluded from both the scatter-plot display and the histogram analysis.

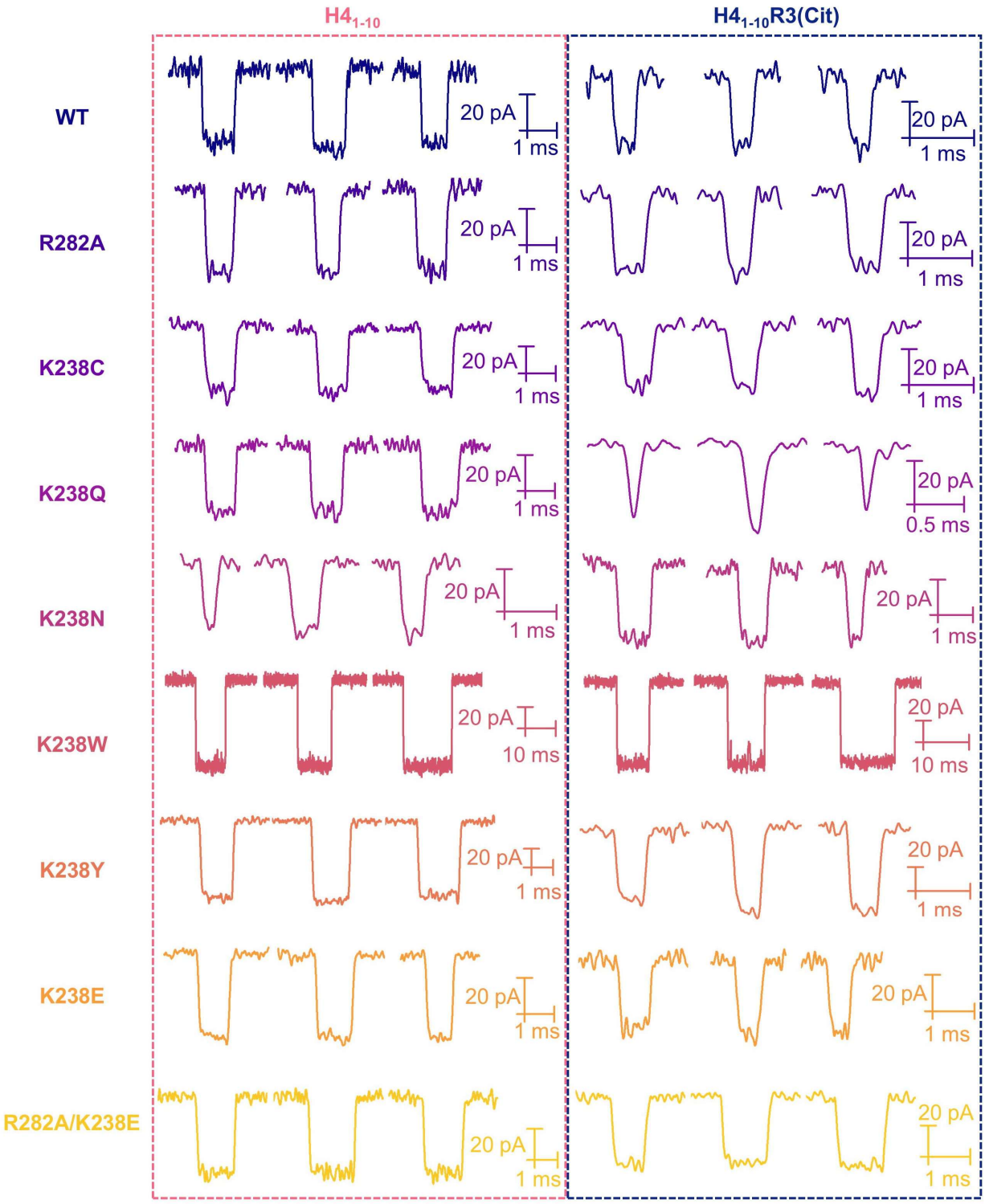


Figure S18. Typical events for for H4$_{1-10}$ and H4$_{1-10}$R3(Cit) through various mutant nanopores. All experiments were conducted in 4 M KCl, 25 mM HEPES, pH 7.5, at 25 ^{o}C, with an applied voltage of –50 mV. The concentrations for both H4$_{1-10}$ and H4$_{1-10}$R3(Cit) were 20 µM.

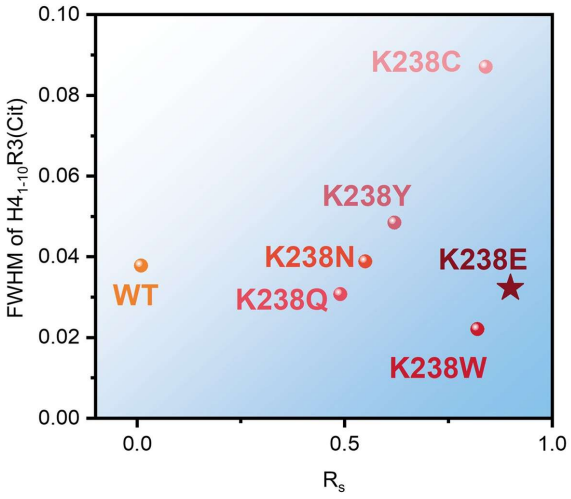


Figure S19. A summary plot of FWHM and R_S for all seven nanopores

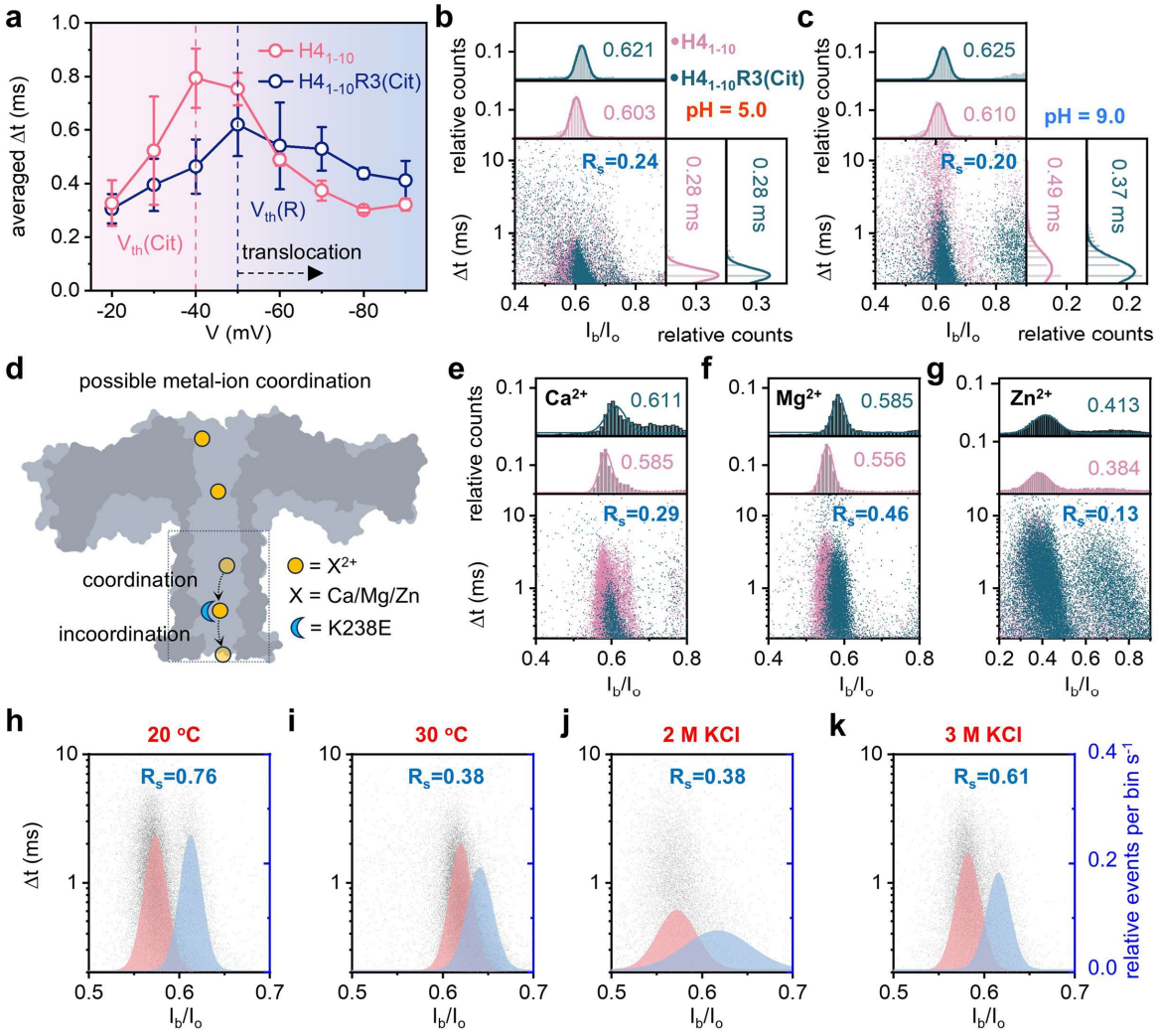


Figure S20. Systematic perturbations to driving force and environment compromise citrullination recognition sensitivity. (a) Applied voltage-dependence of averaged dwell times of $H4_{1\text{-}10}$ (pink) and $H4_{1\text{-}10}R3(Cit)$ peptides (blue) translocating through the EA nanopore. Dashed lines represent the threshold voltages that the peptide could transport through the nanopore. (b,c) Scatter plots and Gaussian fittings showing the influence of pH on the discrimination capacity for $H4_{1\text{-}10}$ (pink) and $H4_{1\text{-}10}R3(Cit)$ peptides (green), (b): pH 5.0, (c): pH 9.0. (d) Schematic illustrating the proposed coordination mechanism between divalent metal ions and engineered pore residues. (e-g) Scatter plots of translocation events in the presence of 10 mM Ca^{2+} (e), Mg^{2+} (f) and Zn^{2+} (g). (h-i) Scatter plots and Gaussian fittings showing the influence of temperatures on the discrimination capacity for $H4_{1\text{-}10}$ (pink) and $H4_{1\text{-}10}R3(Cit)$ peptides (blue), (h): 20 °C,

(i): 30 °C. (j-k) Scatter plots and Gaussian fittings showing the influence of electrolyte concentration on the discrimination capacity for $H4_{1\text{-}10}$ (pink) and $H4_{1\text{-}10}$R3(Cit) peptides (blue), (j): 2 M KCl, (k): 3 M KCl. All nanopore experiments were conducted in 4 M KCl, 25 mM HEPES, pH 7.5, at 25 °C, with an applied voltage of –50 mV, the peptide concentrations were 20 μM, unless otherwise specified. Scatter plots and the corresponding histograms/Gaussian fittings were generated from the same filtered event set; events with dwell times shorter than 0.2 ms were excluded from both the scatter-plot display and the histogram analysis.

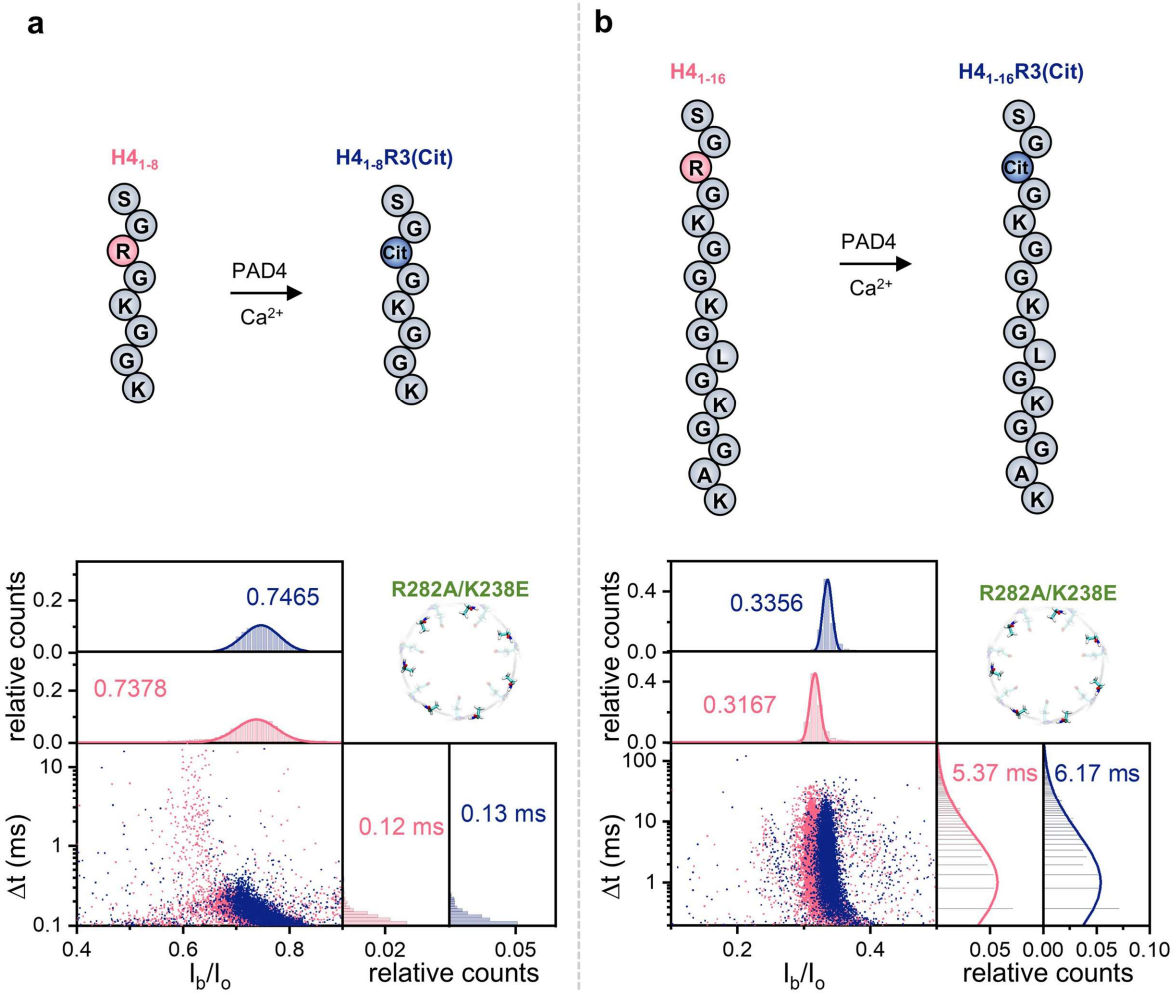


Figure S21. Statistical data for $H4_{1-8}$ and $H4_{1-8}$R3(Cit) (a), $H4_{1-16}$ and $H4_{1-16}$R3(Cit) (b) through R282A/K238E nanopore. All experiments were conducted in 4 M KCl, 25 mM HEPES, pH 7.5, at 25 ºC, with an applied voltage of –50 mV. The concentrations for all the peptides were 20 μM. Scatter plots and the corresponding histograms/Gaussian fittings were generated from the same filtered event set; events with dwell times shorter than 0.2 ms were excluded from both the scatter-plot display and the histogram analysis.

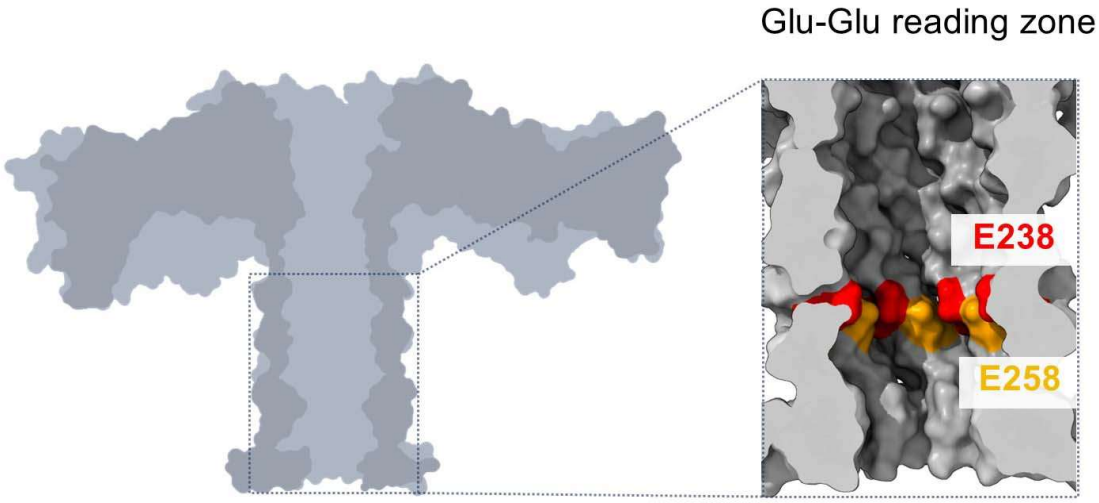


Figure S22. Illustration of dual-glutamate "reading zone" (E238 and E258) in the EA nanopore, where the two residues act synergistically to enhance citrullination recognition.

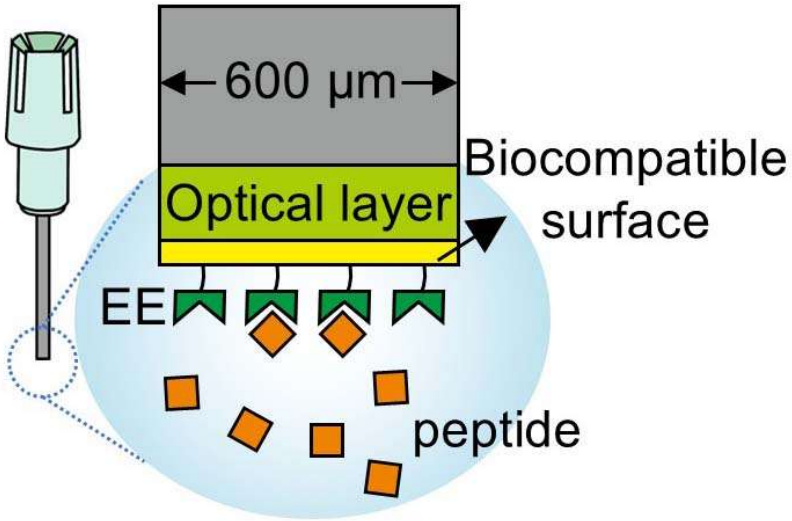


Figure S23. Schematic illustration of bio-layer interferometry (BLI). The Glu-Glu (EE) peptides were immobilized on the probe surface, and $H4_{1\text{-}10}$ as well as $H4_{1\text{-}10}$R3(Cit) at concentrations of 0.005, 0.03, 0.125, 0.25, and 1 µM were employed to determine binding kinetics.

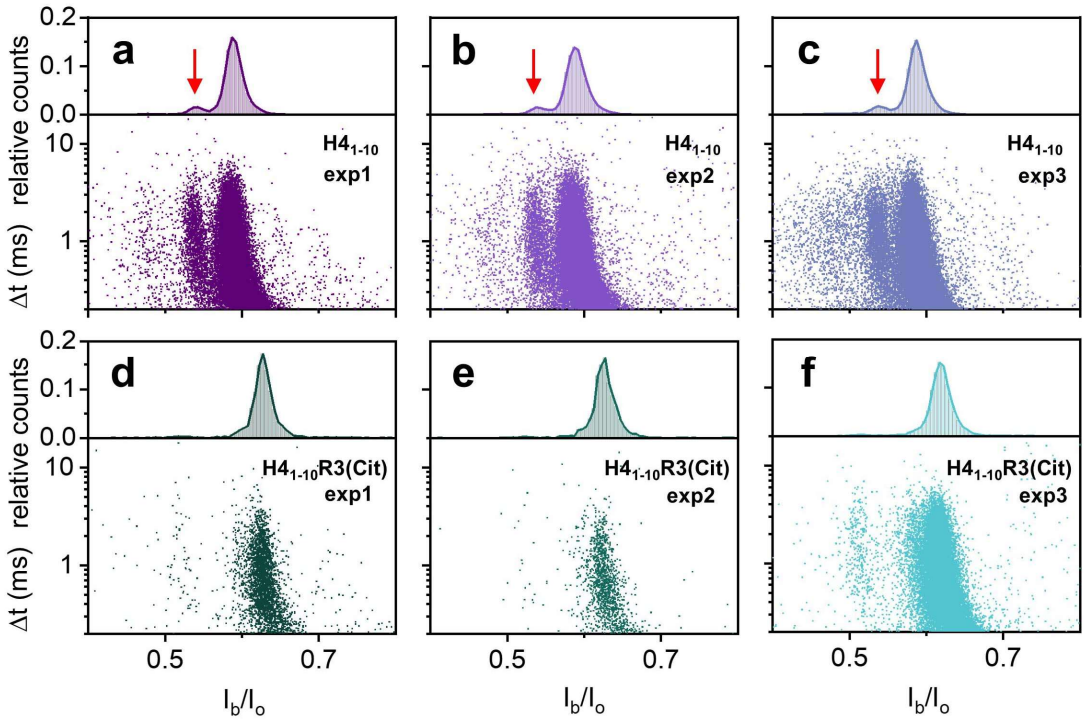


Figure S24. Three individual experiments for $H4_{1-10}$ and $H4_{1-10}$R3(Cit) through R282A/K238E nanopore, respectively. The red arrow indicates extra peak for $H4_{1-10}$. All experiments were conducted in 4 M KCl, 25 mM HEPES, pH 7.5, at 25 °C, with an applied voltage of –50 mV. The concentrations for both $H4_{1-10}$ and $H4_{1-10}$R3(Cit) were 20 μM. Scatter plots and the corresponding histograms/Gaussian fittings were generated from the same filtered event set; events with dwell times shorter than 0.2 ms were excluded from both the scatter-plot display and the histogram analysis.

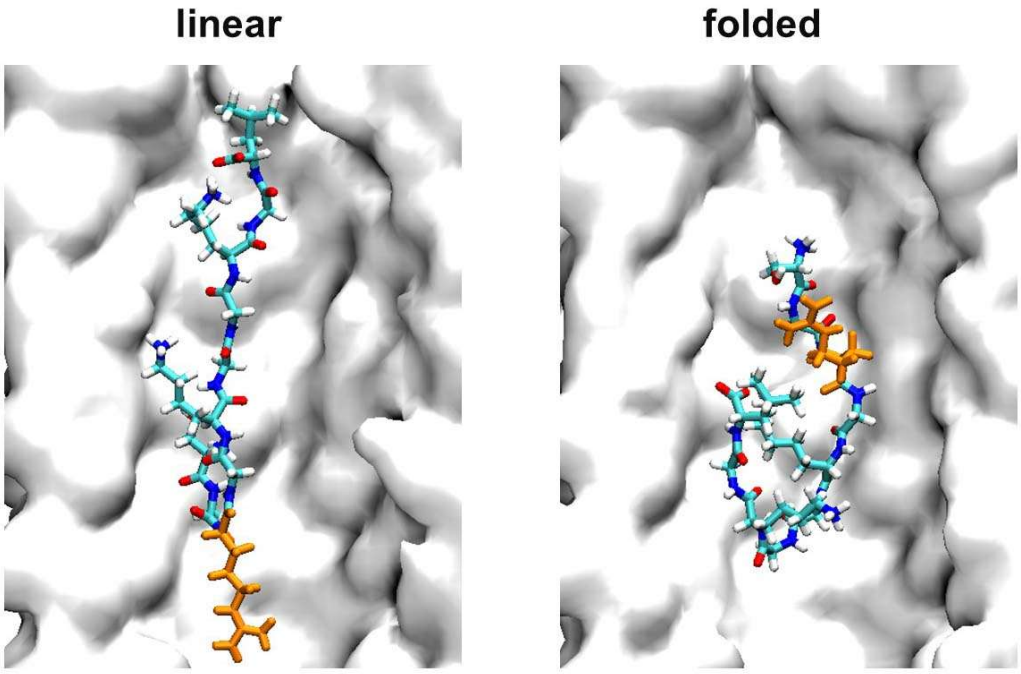
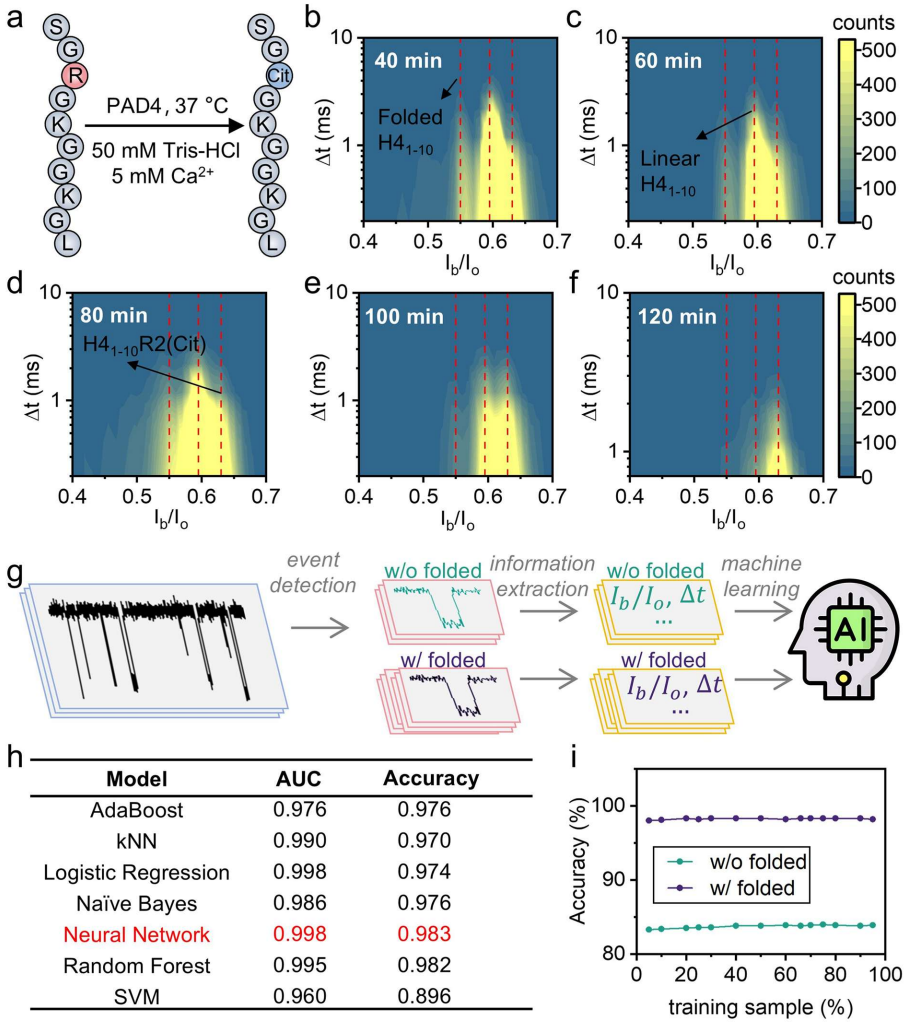


Figure S25. Schematic model proposing two distinct translocation conformations to explain the deep and shallow blockade events of $H4_{1\text{-}10}$.



Model	AUC	Accuracy
AdaBoost	0.976	0.976
kNN	0.990	0.970
Logistic Regression	0.998	0.974
Naïve Bayes	0.986	0.976
Neural Network	0.998	0.983
Random Forest	0.995	0.982
SVM	0.960	0.896

Figure S26. Monitoring and quantification of the PAD4-catalyzed citrullination reaction. (a) PAD4-catalyzed citrullination of $H4_{1\text{-}10}$ (1.36 mM), was carried out with the addition of 1.36 μM PAD4, and the reaction buffer containing 200 mM Tris-HCl, 50 mM $CaCl_2$, incubated at 37 °C. (b-h) Scatter plots (I_b/I_o vs. Δt) of single-molecule events from the production of (a) at indicated reaction time points: 40 (b), 60 (c), 80 (d), 100 (e), and 120 (f) minutes. (g) Schematic workflow illustrating the machine learning-based classification of single-molecule events. (h) Performance comparison of seven different machine learning models, highlighting the superior accuracy of the Neural Network. (i) Robustness of the Neural Network's performance against varying training size of datasets both with (w/) and without (w/o) folded $H4_{1\text{-}10}$ events, indicating better accuracy w/ folded $H4_{1\text{-}10}$ events. Events with dwell times shorter than 0.2 ms were excluded from the scatter-plot display.

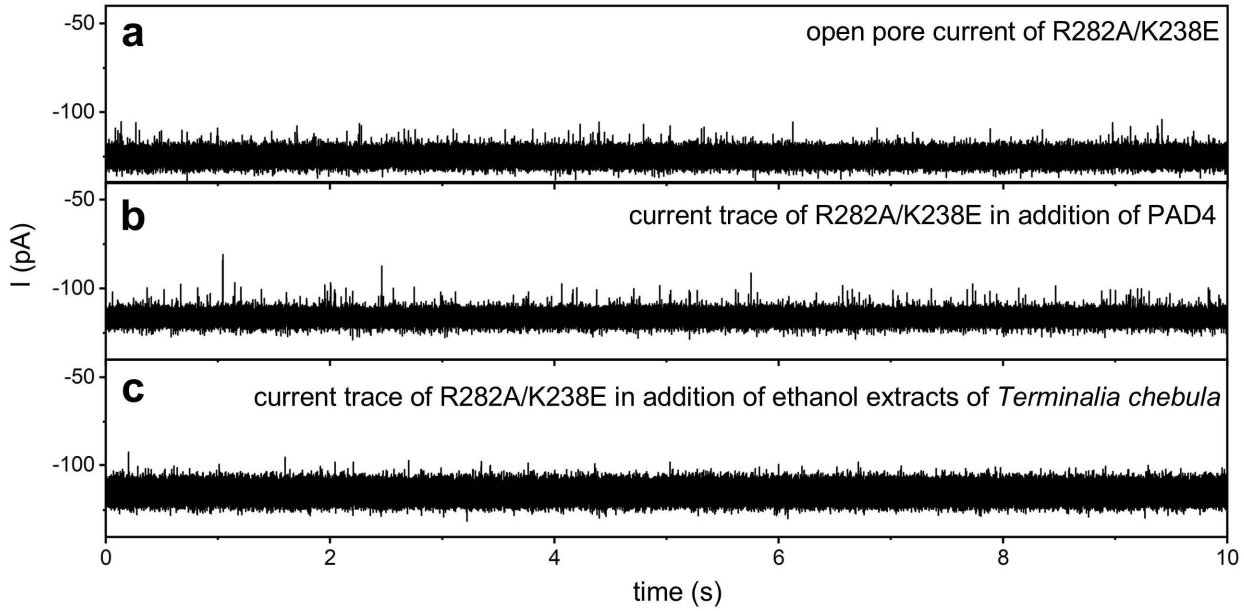


Figure S27. Current traces of the EA nanopore, and following addition of PAD4 enzyme and inhibitor. (a) Raw current trace in the absence of any analyte. (b) Current trace after addition of 4 μL of 0.2 mg/mL PAD4 enzyme solution. (c) Current trace after addition of 4 μL of 20 mg/mL ethanolic extracts of *Terminalia chebula*. All nanopore experiments were performed in a buffer containing 4 M KCl and 25 mM HEPES (pH 7.5) at 25 °C, with an applied voltage of –50 mV.

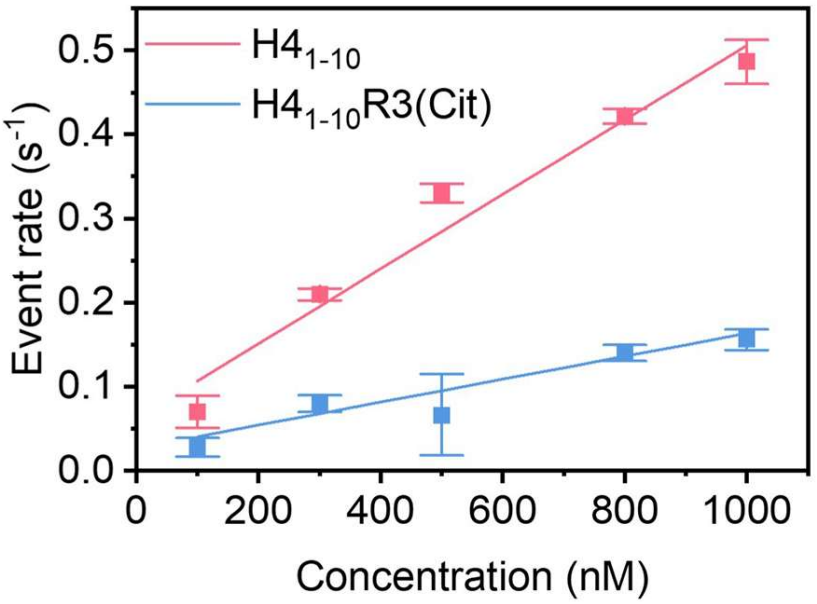


Figure S28. Capture rates for H4$_{1-10}$ (pink) and H4$_{1-10}$R3(Cit) (blue) across a range of concentrations were measured via the EA nanopore. All experiments were performed in 4 M KCl, 25 mM HEPES, pH 7.5, at 25 ºC, with an applied voltage of –50 mV. Data points represent the mean values from triplicate independent experiments.

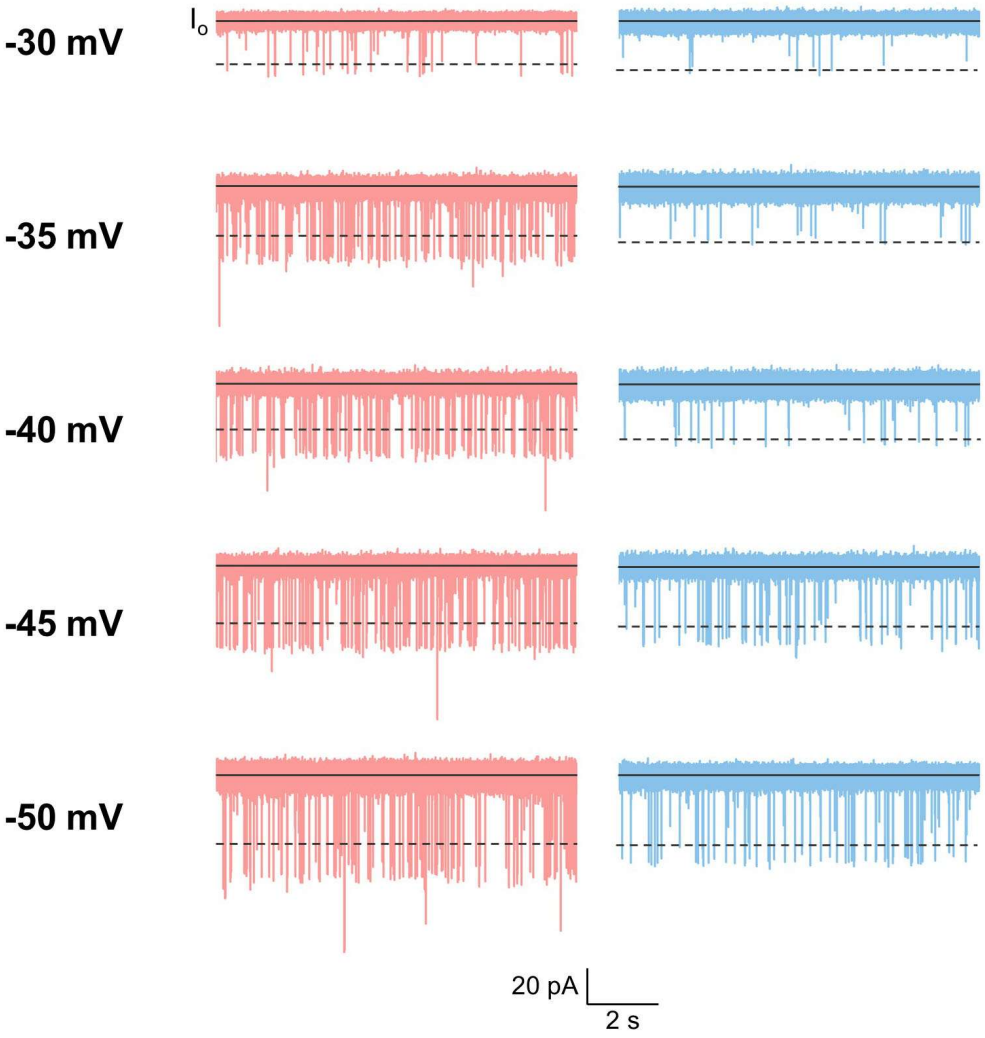


Figure S29. Current traces of $H4_{1\text{-}10}$ (pink) and $H4_{1\text{-}10}R3(Cit)$ (blue) at various applied voltages through EA nanopore. The dashed line indicates the event threshold, defined as $I_o - 12\sigma$, where σ is the baseline noise standard deviation. All experiments were conducted in 4 M KCl, 25 mM HEPES, pH 7.5, at 25 °C. The concentrations for all the peptides were 20 μM.

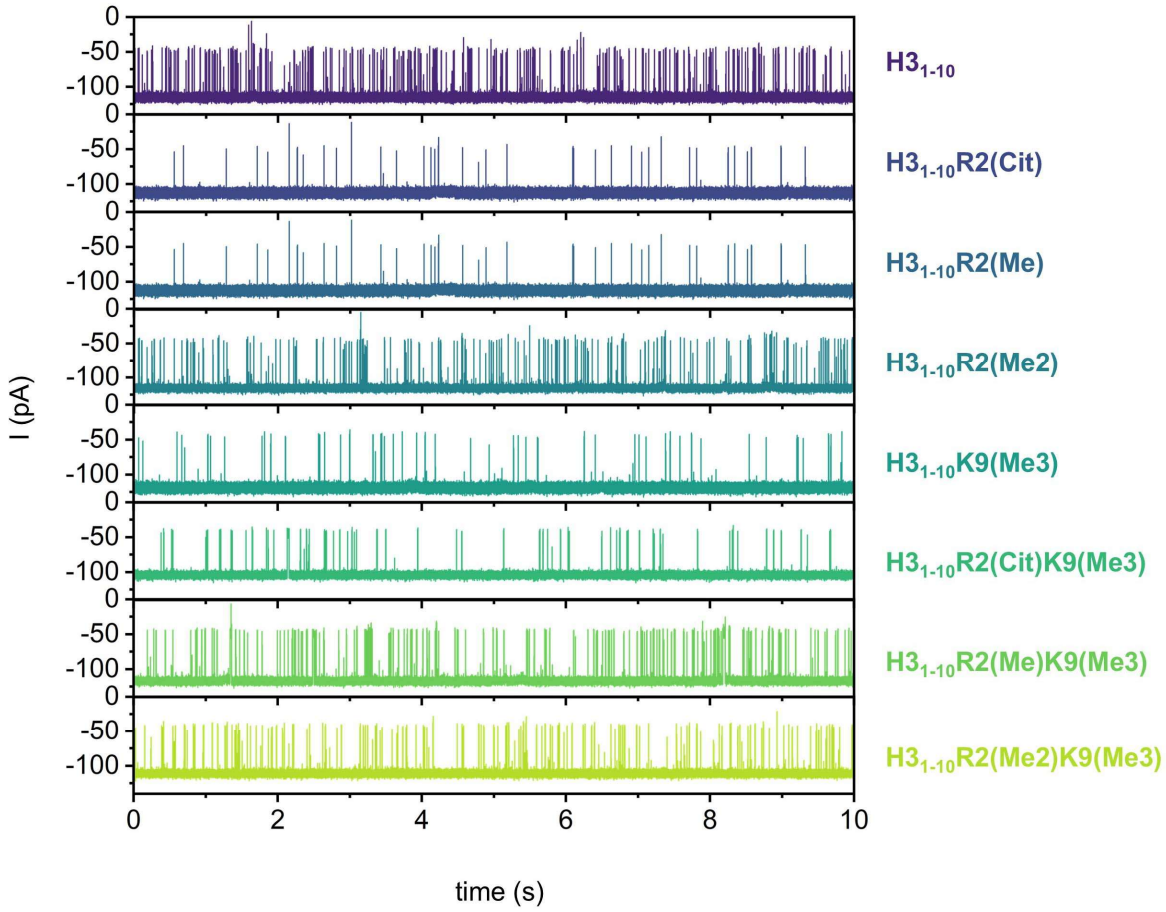


Figure S30. Current traces for $H3_{1-10}$ and its modifications with various PTMs through R282A/K238E nanopore. All experiments were conducted in 4 M KCl, 25 mM HEPES, pH 7.5, at 25 ºC, with an applied voltage of –50 mV. The concentrations for all the peptides were 20 μM.

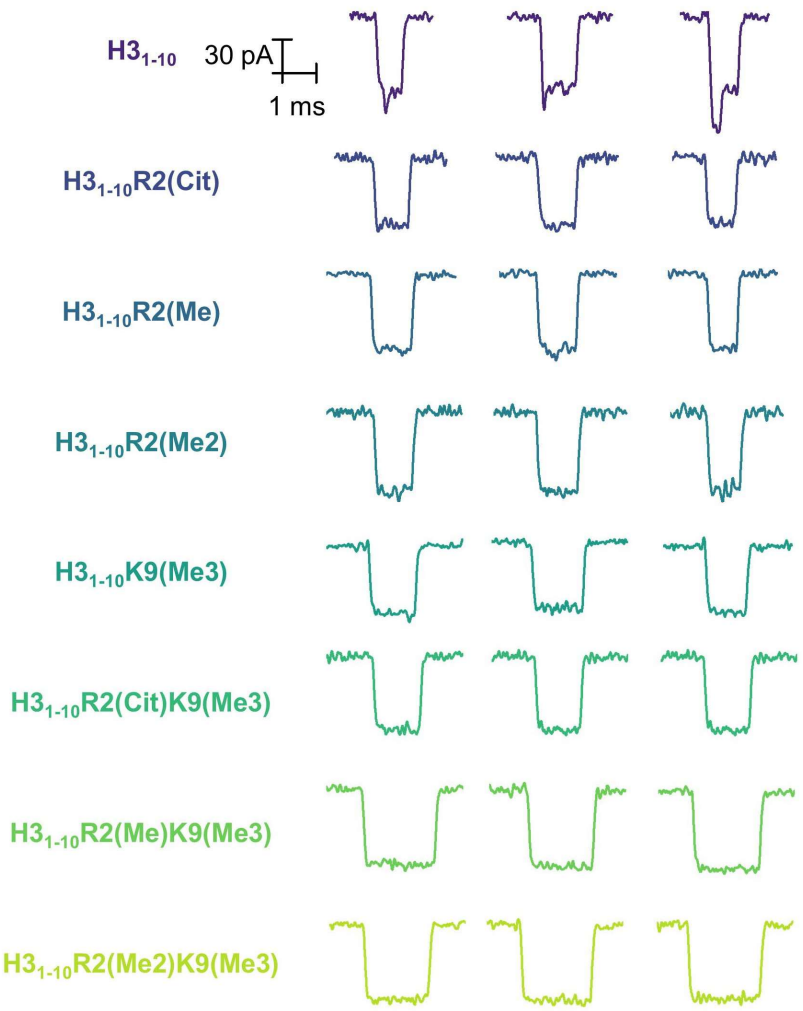


Figure S31. Typical events for $H3_{1-10}$ and its modifications with various PTMs through R282A/K238E nanopore. All experiments were conducted in 4 M KCl, 25 mM HEPES, pH 7.5, at 25 ºC, with an applied voltage of –50 mV. The concentrations for all the peptides were 20 μM.

4. References

1. Di Muccio, G., Rossini, A. E., Di Marino, D., Zollo, G. & Chinappi, M. Insights into protein sequencing with an α-Hemolysin nanopore by atomistic simulations. *Scientific Reports* **9**, 6440, (2019).
2. Liu, J. & Aksimentiev, A. Molecular determinants of current blockade produced by peptide transport through a nanopore. *ACS Nanoscience Au* **4**, 21-29, (2024).
3. Maglic, J. B. & Lavendomme, R. MoloVol: an easy-to-use program for analyzing cavities, volumes and surface areas of chemical structures. *J. Appl. Crystallogr.* **55**, 1033-1044, (2022).
4. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671-690, (2010).
5. Zhao, Z. *et al.* Immobilized PAD4 enzyme on magnetic nanoparticles for screening natural inhibitors from traditional Chinese medicines. *Talanta* **278**, 126492, (2024).
6. Li, M. *et al.* Single-molecule identification and quantification of steviol glycosides with a deep learning-powered nanopore sensor. *ACS Nano* **18**, 25155-25169, (2024).
7. Lu, W. *et al.* Precise structural analysis of neutral glycans using aerolysin mutant T240R nanopore. *ACS Nano* **18**, 12412-12426, (2024).
8. Forstater, J. H. *et al.* MOSAIC: A modular single-molecule analysis interface for decoding multistate nanopore data. *Anal. Chem.* **88**, 11900-11907, (2016).
9. Ouldali, H. *et al.* Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nat. Biotechnol.* **38**, 176-181, (2020).
10. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **14**, 33-38, (1996).
11. Hanwell, M. D. *et al.* Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminf.* **4**, 17, (2012).
12. Gfeller, D., Michielin, O. & Zoete, V. SwissSidechain: a molecular and structural database of non-natural sidechains. *Nucleic Acids Res.* **41**, D327-D332, (2013).
13. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781-1802, (2005).
14. Phillips, J. C. *et al.* Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **153**, 044130, (2020).
15. Wilson, J., Sarthak, K., Si, W., Gao, L. & Aksimentiev, A. Rapid and accurate determination of nanopore ionic current using a steric exclusion model. *ACS Sens.* **4**, 634-644, (2019).
16. Zhao, J. *et al.* Screening of natural inhibitors against peptidyl arginine deiminase 4 from herbal extracts by a high-performance liquid chromatography ultraviolet-visible based method. *J. Chromatogr. A* **1716**, 464643, (2024).
17. Demšar, J. *et al.* Orange: data mining toolbox in Python. *J. Mach. Learn. Res.* **14**, 2349-2353, (2013).