Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore

Hadjer Ouldali[1], Kumar Sarthak[2], Tobias Ensslen[3], Fabien Piguet[1,11], Philippe Manivet[4,5], Juan Pelta[6], Jan C. Behrends [3,7,8], Aleksei Aksimentiev [2,9,10]* and Abdelghani Oukhaled [1]*

[1]LAMBE UMR 8587, Université de Cergy-Pontoise, CNRS, CEA, Université Paris-Seine, Cergy-Pontoise, France. [2]Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [3]Laboratory for Membrane Physiology and Technology, Department of Physiology, Faculty of Medicine, University of Freiburg, Freiburg, Germany. [4]APHP, GHU APHP.Nord, DMU BioGem, Hôpital Lariboisière, BIOBANK Lariboisière Department BB-0033-00064, Plateforme de BioPathologie et de Technologies Innovantes en Santé, Paris, France. [5]INSERM UMR 1141 "NeuroDiderot", Université de Paris, Paris, France. [6]LAMBE UMR 8587, Université d'Evry-Val-d'Essonne, CNRS, CEA, Université Paris-Saclay, Evry, France. [7]Freiburg Materials Research Centre (FMF), University of Freiburg, Freiburg, Germany. [8]Freiburg Centre for Interactive Materials and Bioinspired Technologies (FIT), University of Freiburg, Freiburg, Germany. [9]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [10]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [11]Present address: DreamPore S.A.S., 33 Boulevard du Port 95000, Cergy, France. *e-mail: aksiment@illinois.edu; abdelghani.oukhaled@u-cergy.fr
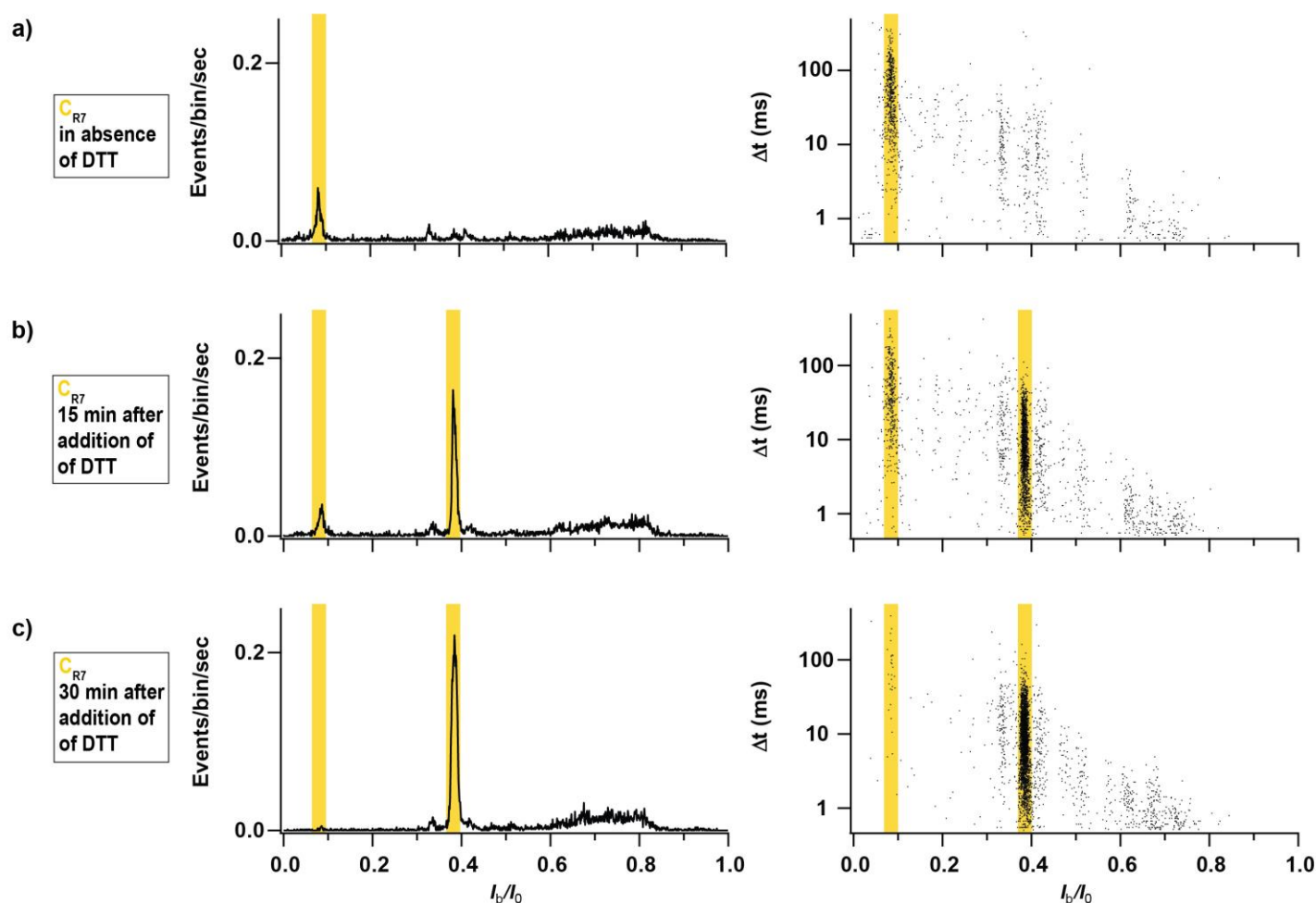
Supplementary Figure 1

Electrical signatures of real-time reduction of C_{R7} peptides by dithiothreitol.

Histograms of relative residual current I_b/I_0 (left) and scatter plots of the blockade duration *versus* I_b/I_0 (right) measured for C_{R7} peptides in the absence of the disulfide-bond reducing agent dithiothreitol (DTT) (a, n = 3311 events), 15 (b, n = 5211 events) and 30 (c, n = 5926 events) minutes after the addition of 25 mM DTT. Colored rectangles indicate the mean (centerline) and the standard deviation (widths) of the I_b/I_0 values recorded individually for C_{R7} dimers (data from Fig.1f-j) and recorded for C_{R7} monomers after disulfide bonds reduction. In absence of DTT, the main population of blockade current has a mean I_b/I_0 value of 0.084 ± 0.005, which we attribute to C_{R7} dimers. Fifteen minutes after the DTT addition, frequency of blockades produced by C_{R7} dimers is reduced as another population of blockade currents of $I_b/I_0 = 0.383 \pm 0.005$ emerges, indicating reduction of C_{R7} dimers to C_{R7} monomers by DTT. After 30 minutes of the DTT activity, the 0.084 ± 0.005 population has almost entirely disappeared. In the absence of DTT, the minority clusters of events at relative residual current of around 0.4 correspond to impurities present in our sample except for the cluster of events characterized by the relative residual current of 0.383 ± 0.005, which corresponds to events induced by a minority population of C_{R7} monomers. The mean value (respectively uncertainty) of relative residual current of each peptide was obtained as the mean value (respectively standard deviation) of a gaussian fit of the corresponding I_b/I_0 distribution; from single independent experiments. The data were acquired in 4 M KCl, 25 mM HEPES buffer, at 7.5 pH and $20.0 \pm 0.5^{\circ}$C, and under a –50 mV bias applied to the *trans* compartment.
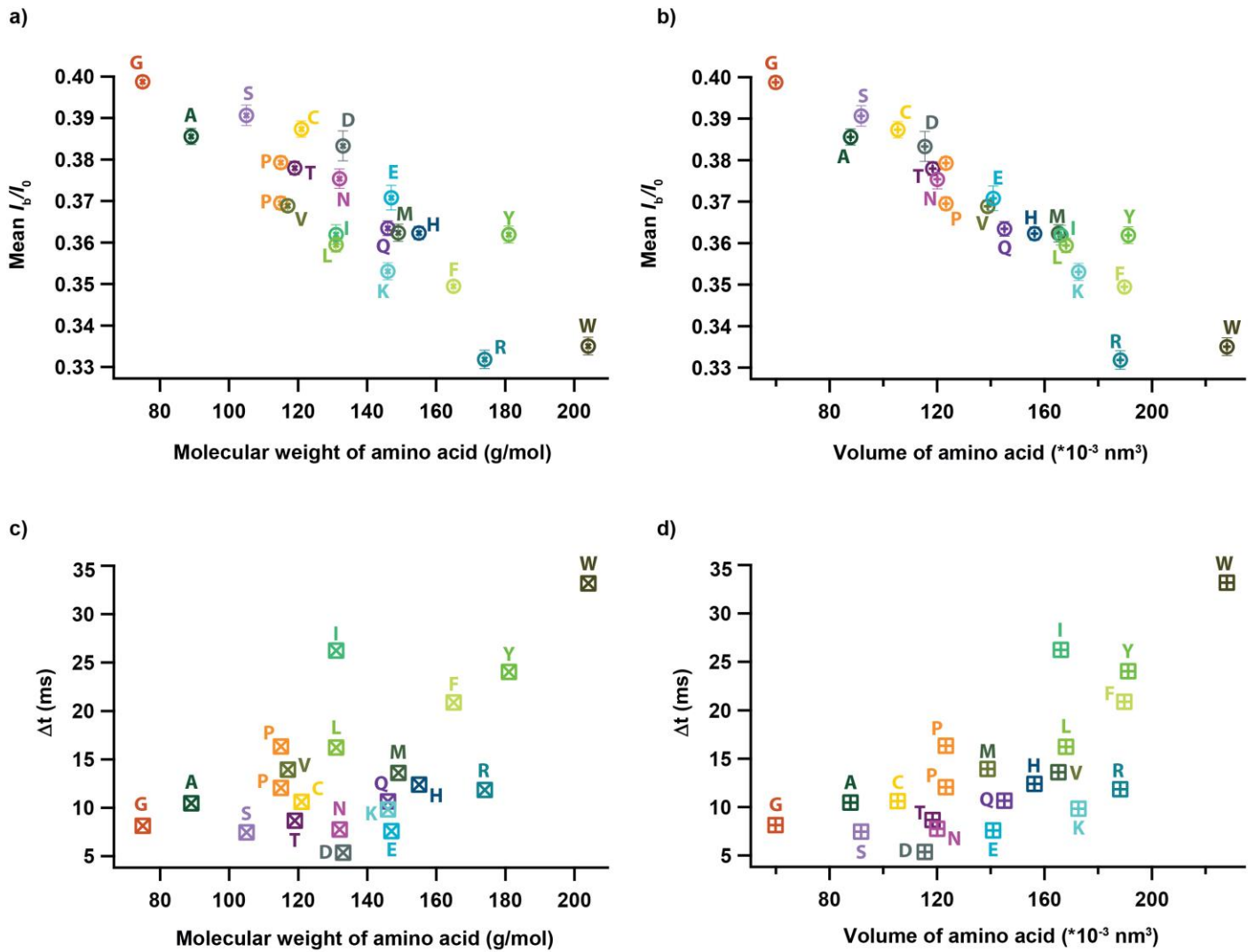
Supplementary Figure 2

Dependence of the mean relative residual current and the blockade duration on the molecular weight or the volume of amino acid X.
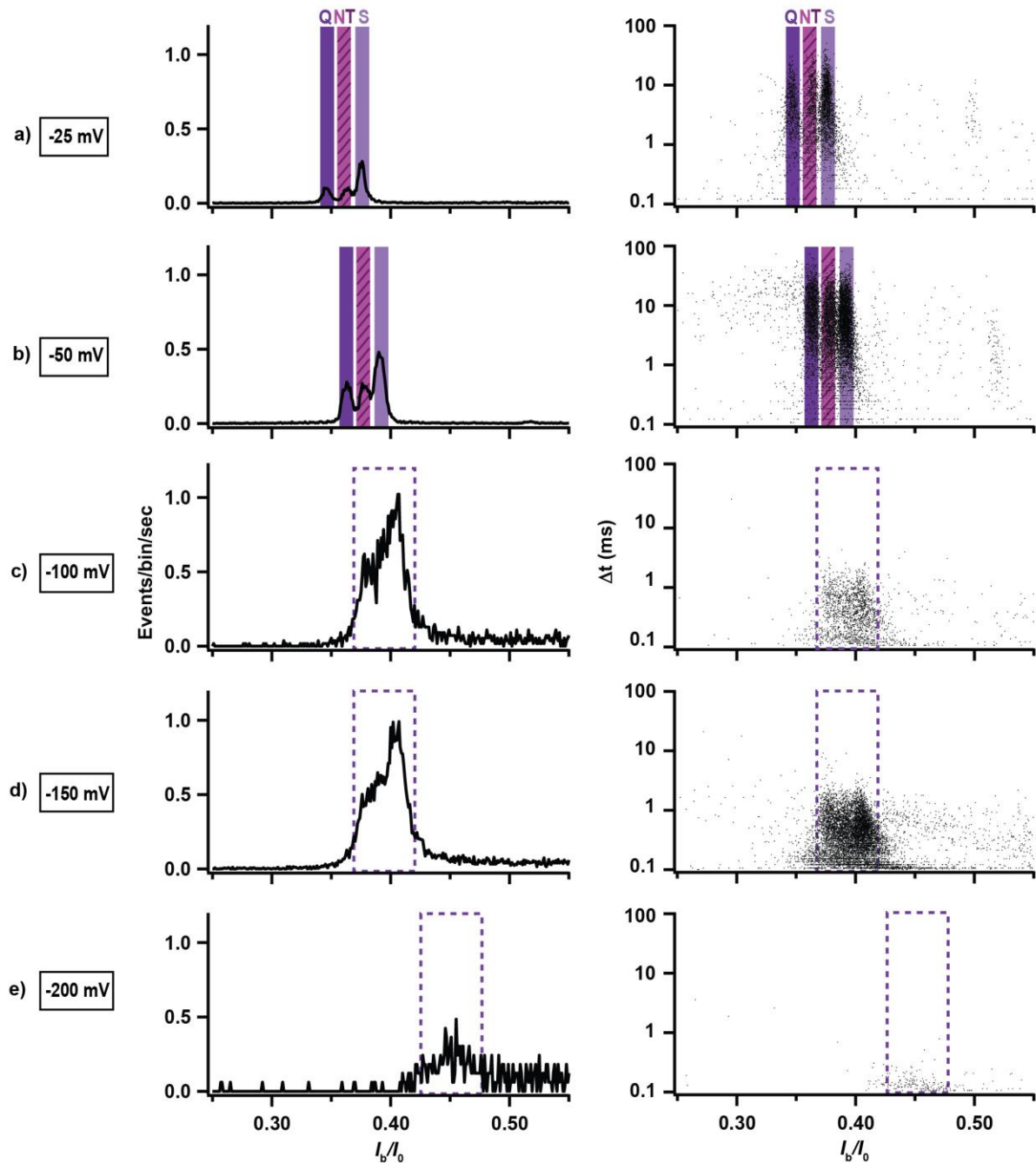
Mean relative residual current $\langle I_b/I_0 \rangle$ and its standard deviation (a,b) and mean blockade duration (c,d) for the X_{R7} peptide *versus* molecular weight (a,c) or volume (b,d) of amino acid X (n = 34025 (R_{R7}), 1328 (W_{R7}), 6228 (F_{R7}), 1969 (K_{R7}), 2319 (L_{R7}), 3411 (Y_{R7}), 5185 (I_{R7}), 1449 (H_{R7}), 2317 (M_{R7}), 3822 (Q_{R7}), 5369 (V_{R7}), 1681 (P_{R7}), 1655 (E_{R7}), 3165 (N_{R7}), 2978 (T_{R7}), 1866 (D_{R7}), 1717 (A_{R7}), 3311 (C_{R7}), 8169 (S_{R7}), 3266 (G_{R7}), 2626 ((NO_2)-Y_{R7}), 2374 ((sulfoxide)-M $_{R7}$) events). Figure 1k from the main text is reproduced here as panel b to facilitate comparison of the two dependences. The mean value (respectively uncertainty) of relative residual current of each peptide was obtained as the mean value (respectively standard deviation) of a gaussian fit of the corresponding I_b/I_0 distribution; from single independent experiments. The data were acquired in 4 M KCl, 25 mM HEPES buffer, at 7.5 pH and 20.0 ± 0.5°C, and under a −50 mV bias applied to the *trans* compartment.
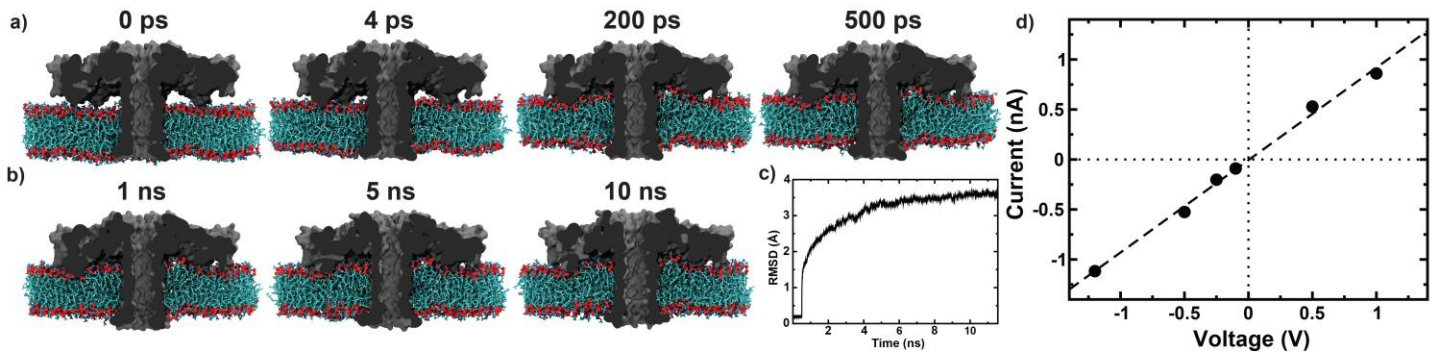
Supplementary Figure 3

MD simulation of open-pore aerolysin system

(a) Equilibration of the lipid bilayer. During this 0.5 ns simulation, all non-hydrogen atoms of the aerolysin channel were restrained to their initial coordinates while preventing water molecules from entering the lipid-protein interface. The aerolysin channel is shown as a grey cut-away molecular surface, the head groups and tails of the DPhPC lipid bilayer are shown in red and cyan; 1 M KCl solution is not shown. (b) Free equilibration of the aerolysin system. The last frame of the equilibration trajectory was used as the starting structure for subsequent electric field simulations. (c) Root-mean-square deviation (RMSD) of the protein's alpha-carbon atoms from their initial coordinates during the equilibration simulations. Each data point was averaged over the protein's 2976 alpha-carbon atoms. (d) The simulated current-voltage curve of the aerolysin nanopore. Each data point derives from an independent MD simulation of the aerolysin system, each lasting from 13 to 37 ns, at the specified transmembrane voltage and 1 M KCl concentration. The average current was obtained by averaging instantaneous ionic current over the course of the MD trajectory. The number of data points used for calculating the average ionic current of each system is as follows: 4521 (-1.2 V), 6193 (-0.5 V), 15084 (-0.25 V), 12733 (-0.1 V), 5914 (0.5 V), 5925 (1 V).

Supplementary Figure 4

Effect of voltage on the discrimination of X_{R7} peptides.

Histogram of relative residual current I_b/I_0 (left) and scatter plot of the blockade duration *versus* I_b/I_0 (right) obtained from nanopore experiment performed using an equimolar mixture of Q_{R7}, N_{R7}, T_{R7}, and S_{R7} peptides under a –25 mV (a), –50 mV (b), –100 mV (c), –150 mV (d) and –200 mV (e) bias applied to the *trans* compartment. The increase of the voltage magnitude reduces discrimination of the blockade current populations, shifts the I_b/I_0 values and decreases the blockade duration. In panels a and b, solid rectangles indicate the location of distributions produced by the Q_{R7} and S_{R7} peptides; the striped rectangles indicate overlapping distributions of the N_{R7} and T_{R7} peptides. In panels c-e, the dashed rectangle indicates the expected location of the Q_{R7}, N_{R7}, T_{R7}, and S_{R7} distributions. The data were acquired in 4 M KCl, 25 mM HEPES buffer, at 7.5 pH and 20.0 ± 0.5°C, and under a –50 mV bias applied to the *trans* compartment. For each histogram, at least 1000 events were analyzed.
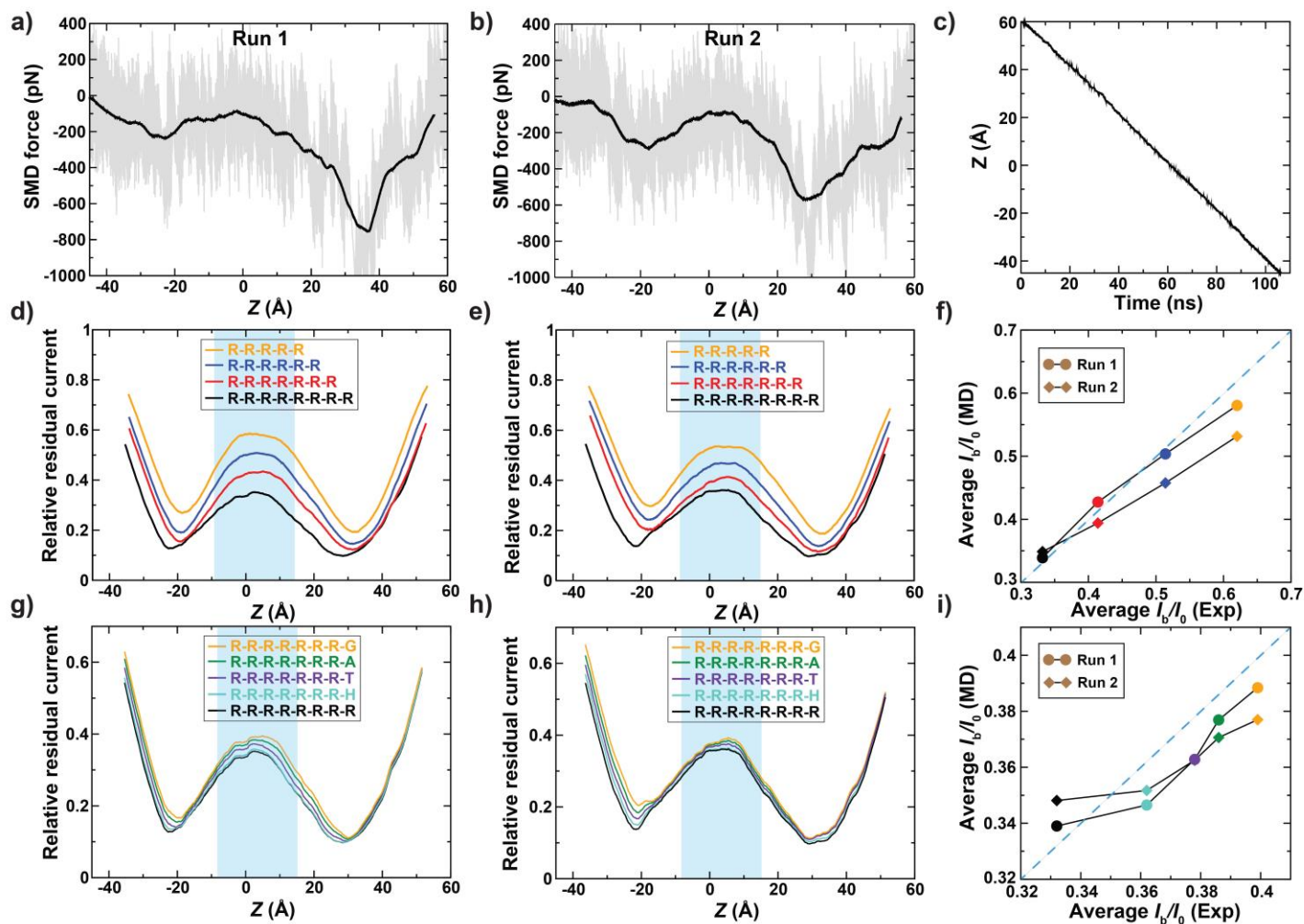
Supplementary Figure 5

Replicate simulation of peptide translocation through aerolysin.

(a,b) Force exerted on the peptide during the two SMD simulations *versus* the *z* coordinate of the peptide's CoM. Hereafter Run 1 refers to the simulation reported in the main text whereas Run 2 refers to the replicate simulation. Instantaneous SMD forces are shown in gray; and their running average (5 Å window) is shown in black. (c) The *z* coordinate of the peptide's CoM *versus* simulation time for Run 1 and Run 2; the two dependences follow closely one another. (d,e) Relative residual current *versus* the CoM coordinate of R_{R7}, R_{R6}, R_{R5} and R_{R4} peptides for Run 1 (d) and Run 2 (e). The peptides' coordinates were derived from the conformations of R_{R7} sampled during the SMD simulation. The currents were computed using the steric exclusion model and averaged using a 10 Å running average. (f) Simulated *versus* experimental average relative currents produced by the arginine peptides within the sensing region of aerolysin. (g-i) Same as d-f but for G_{R7}, A_{R7}, T_{R7}, H_{R7}, and R_{R7} peptides; the experimental values are reproduced from Fig. 1. For panels f & i, the average simulated current was calculated from 130 relative residual current values for runs 1 and 2, corresponding to the presence of arginine peptides in the sensing region. For panels g & h, the currents were averaged using a 10 Å (50 values) running average.
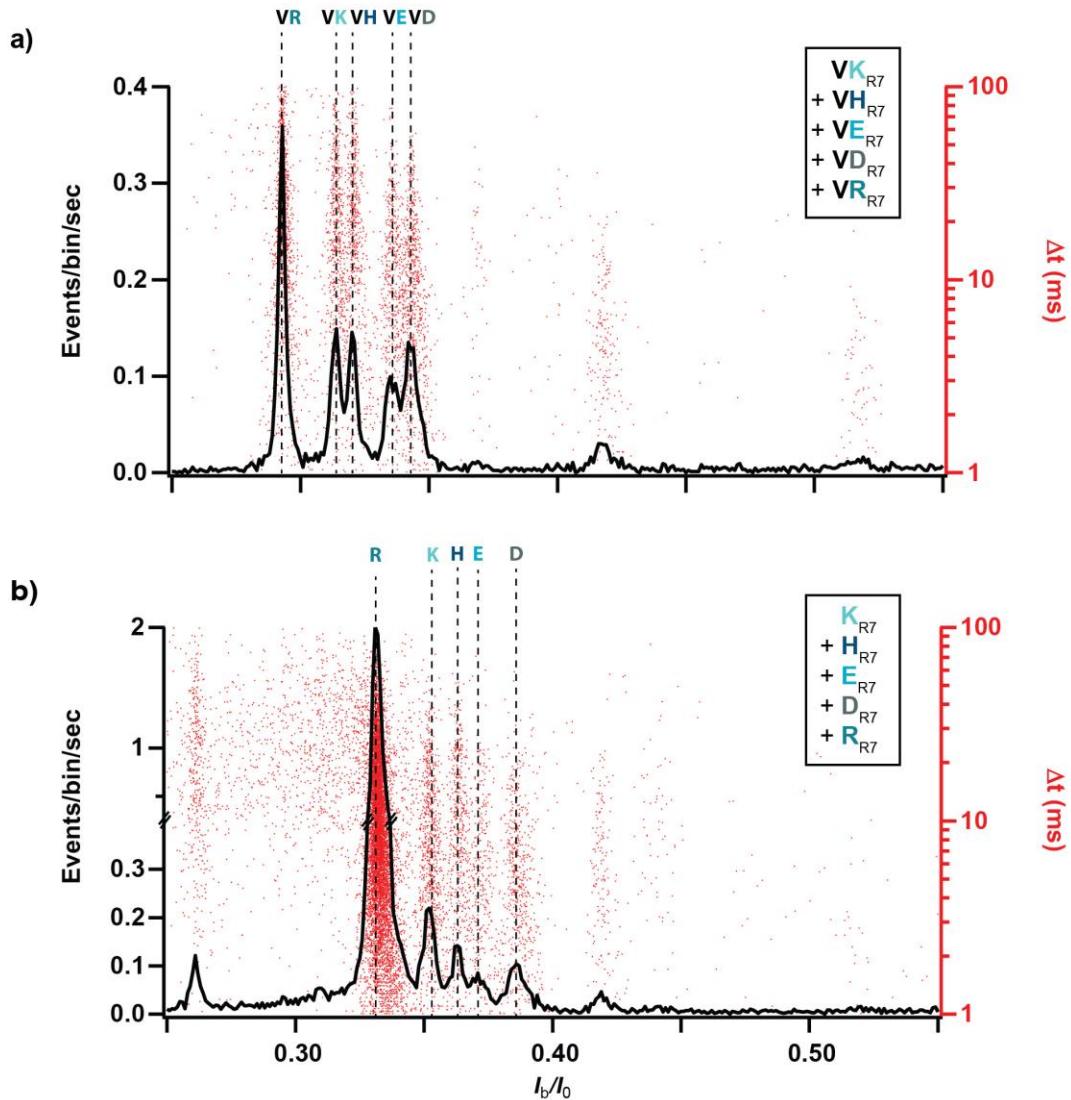
Supplementary Figure 6

Effect of amino acid addition on blockade current.

Histograms of I_b/I_0 (left axis) and scatter plots of blockade duration *versus* I_b/I_0 (right axis) obtained from nanopore experiments performed using an equimolar mixture (1 μM) of VK_{R7}, VH_{R7}, VE_{R7}, VD_{R7}, and VR_{R7} peptides (a, n = 18066 events) or an equimolar mixture (1 μM) of K_{R7}, H_{R7}, E_{R7}, D_{R7}, or R_{R7} peptide species (b, n = 19641 events). In both cases, the addition of the mixture produces five distinct peaks matching the location of the peaks observed in the individual peptide species experiments. The addition of a V amino acid to each X_{R7} peptide leads to a shift of the population to lower I_b/I_0 values. The mean I_b/I_0 values are: 0.353 ± 0.003 (K_{R7}) *versus* 0.314 ± 0.002 (VK_{R7}); 0.362 ± 0.003 (H_{R7}) *versus* 0.320 ± 0.002 (VH_{R7}); 0.371 ± 0.004 (D_{R7}) *versus* 0.336 ± 0.004 (VD_{R7}); 0.332 ± 0.002 (R_{R7}) *versus* 0.293 ± 0.001 (VR_{R7}). The mean value (respectively uncertainty) of relative residual current of each peptide was obtained as the mean value (respectively standard deviation) of a gaussian fit of the corresponding I_b/I_0 distribution; from single independent experiments. The data were acquired in 4 M KCl, 25 mM HEPES buffer, at 7.5 pH and 20.0 ± 0.5°C, and under a –50 mV bias applied to the *trans* compartment.
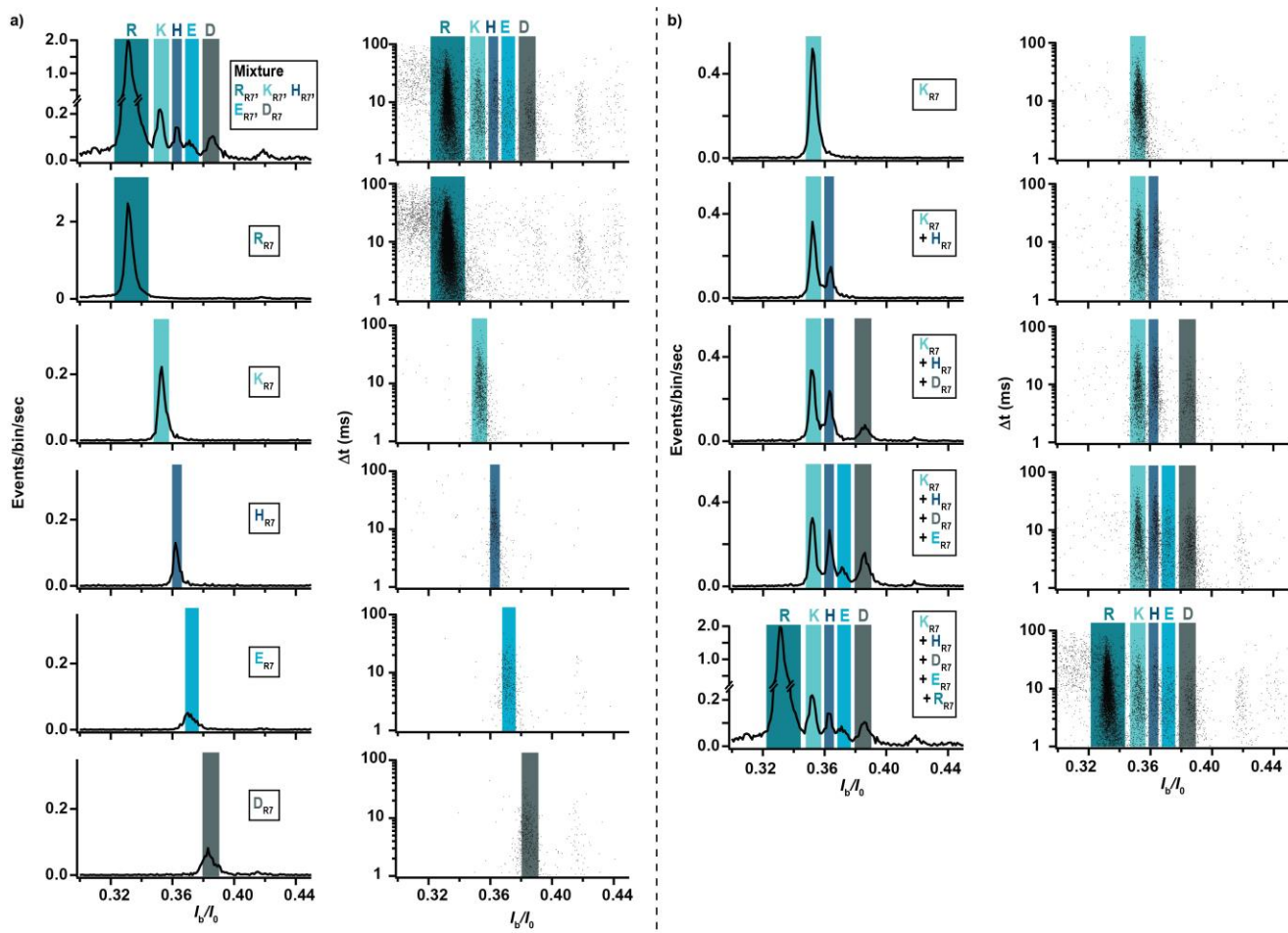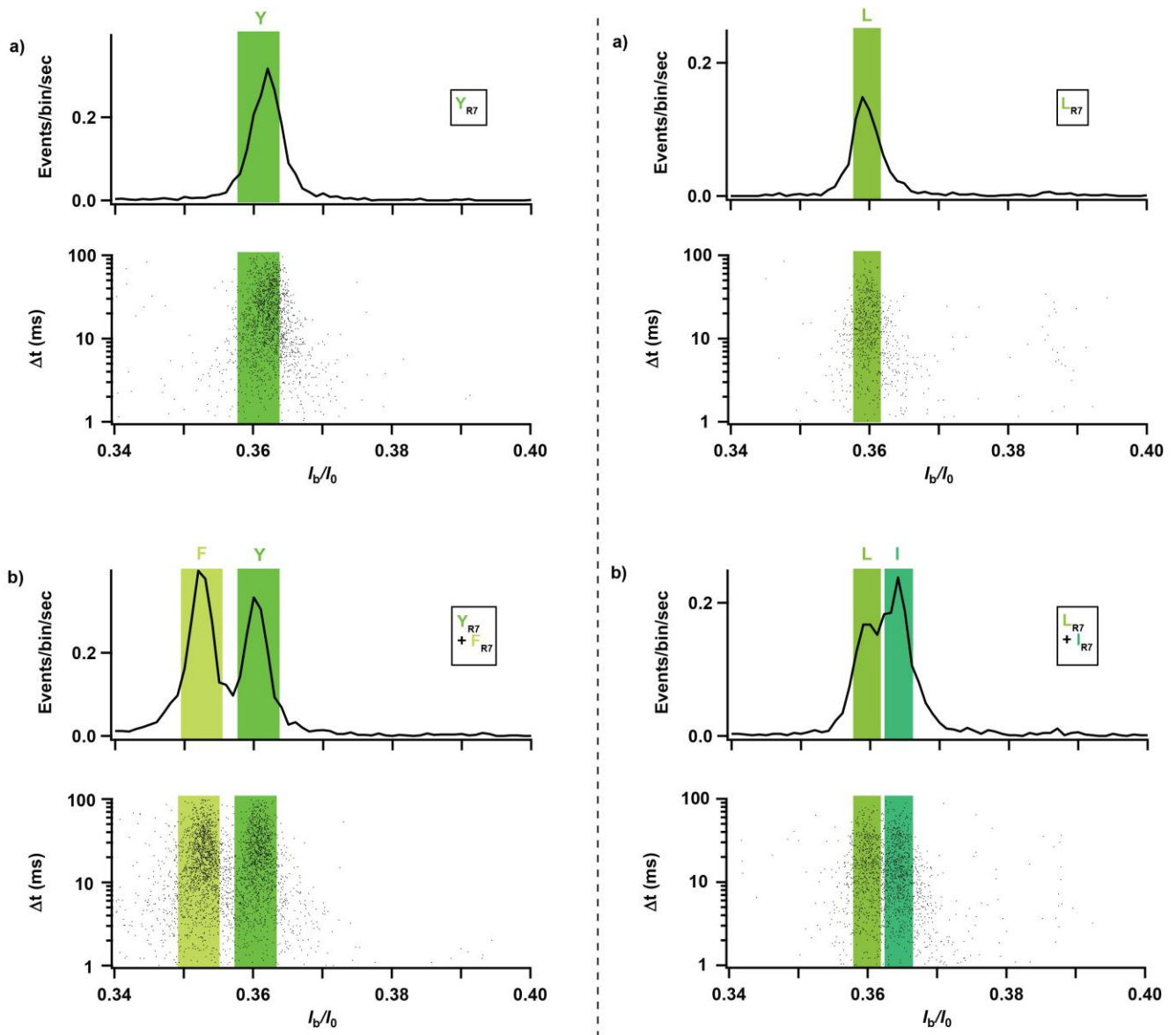
Supplementary Figure 7

Identification of individual peptide species from a mixture.

(a) Histograms of relative residual current I_b/I_0 (left) and scatter plots of blockade duration *versus* of I_b/I_0 (right) obtained from nanopore experiment using an equimolar mixture of K_{R7}, H_{R7}, E_{R7}, D_{R7}, and R_{R7} peptides (top row) or individual K_{R7}, H_{R7}, E_{R7}, D_{R7}, or R_{R7} peptide species (rows 2-6). The location of each peak observed in the mixture experiment matches the location of the peaks observed in the individual peptide species experiments, allowing us to identify populations present in the mixture. (b) Successive addition experiments. Histograms of relative residual current I_b/I_0 (left) and scatter plots of blockade duration *versus* I_b/I_0 obtained from nanopore experiment during successive addition of K_{R7}, H_{R7}, D_{R7}, E_{R7}, and R_{R7} peptides. Each successive addition produces a distinct peak in the I_b/I_0 histogram, allowing us to associate individual peaks with the peptide species. The data were acquired in 4 M KCl, 25 mM HEPES buffer, at 7.5 pH and $20.0 \pm 0.5^{\circ}$C, and under a -50 mV bias applied to the *trans* compartment. For each histogram, at least 1000 events were analyzed.
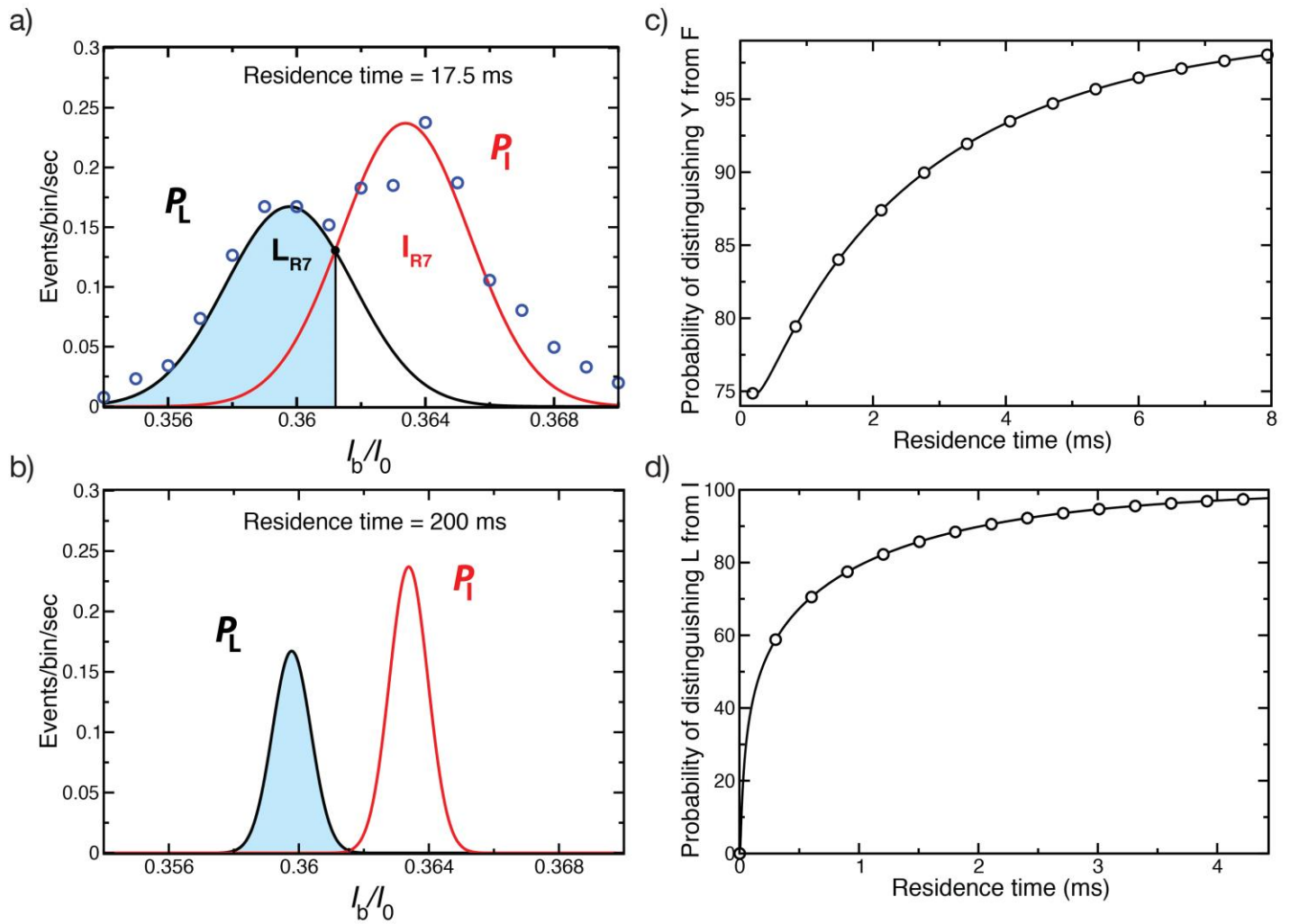
Supplementary Figure 8

Identification of peptide species different by a single hydroxylation or a structural isoform.

Histogram of I_b/I_0 (top plot of each panel) and scatter plot of blockade duration *versus* I_b/I_0 (bottom plot of each panel) obtained from nanopore experiments during successive addition of Y_{R7} and F_{R7} peptides (a,b), where Y and F differ by an hydroxyl group or L_{R7} and I_{R7} peptides (c,d), where L and I are structural isomers. In both cases, well-separated distributions can be associated to the presence of individual peptide species. For each histogram, at least 1000 events were analyzed. The data were acquired in 4 M KCl, 25 mM HEPES buffer, at 7.5 pH and $20.0 \pm 0.5^{\circ}$C, and under a -50 mV bias applied to the *trans* compartment.

Supplementary Figure 9

Theoretical assessment of peptide distinguishability from a single nanopore passage.

(a) Number of events per bin per second for L_{R7} and I_{R7} (open blue circles) derived from our experimental measurements (Fig. 3d). Gaussian fits, $P_L(I_b/I_0)$ and $P_I(I_b/I_0)$, to the experimental data for L_{R7} and I_{R7} are shown in black and red, respectively. The shaded area under the P_L curve indicates the region of single passage distinguishability of L_{R7} peptides (see Supplementary Note 1). (b) $P_L(I_b/I_0)$ and $P_I(I_b/I_0)$ distributions when the peptide residence time is 200 ms. The shaded area under the P_L curve indicates the region of single passage distinguishability. (c) Probability of distinguishing tyrosine (Y) and phenylalanine (F) from a single nanopore passage as a function of the passage duration. (d) Probability of distinguishing leucine (L) from isoleucine (I) when ion counting error is the only noise source.

Supplementary Figure 10

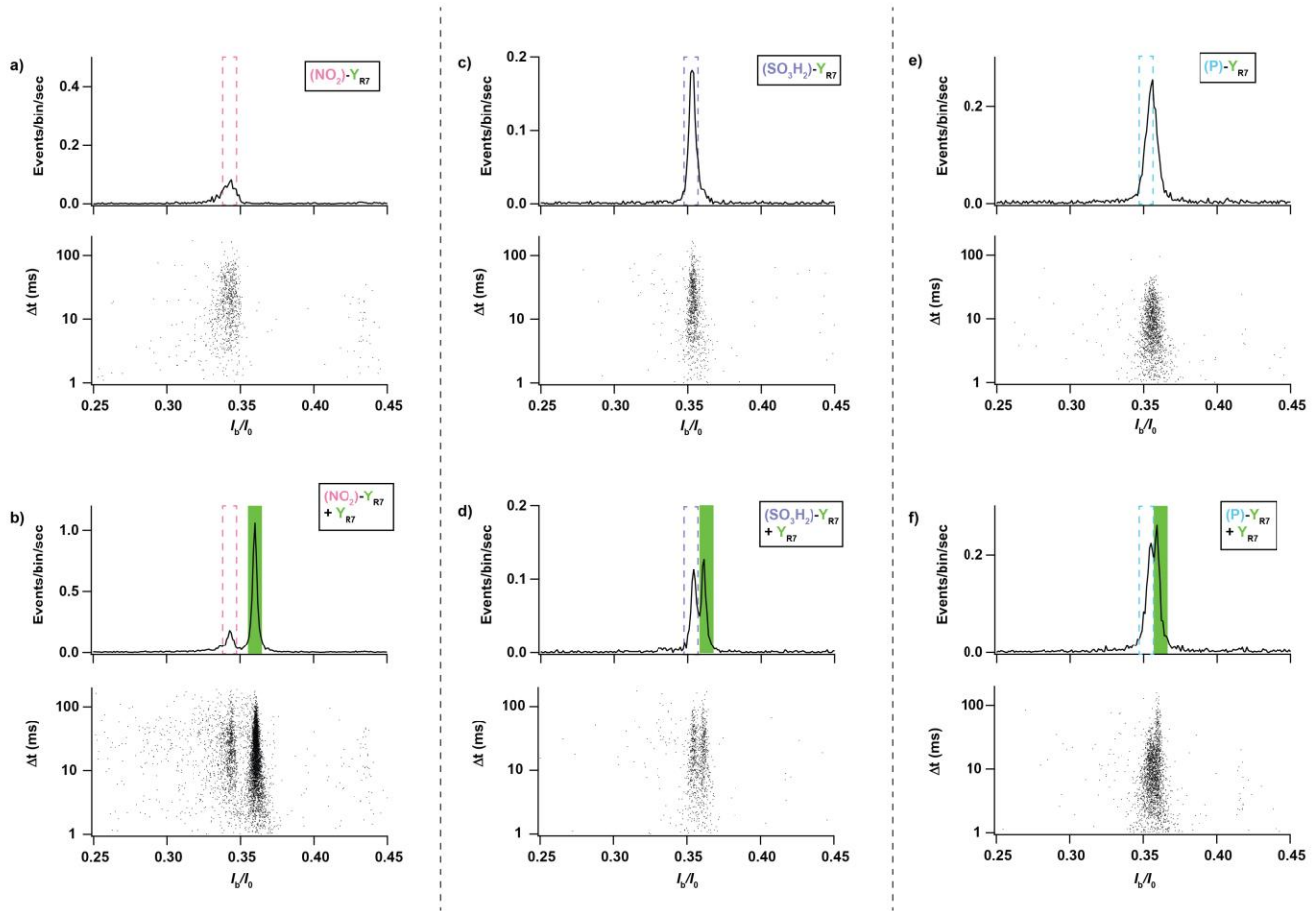Blockade currents of chemically modified Y_{R7} peptides.

Histogram of I_b/I_0 (top plot of each panel) and scatter plot of blockade duration *versus* I_b/I_0 (bottom plot of each panel) obtained from nanopore experiments during analysis of: (NO_2)-Y_{R7} peptide, where the Y amino acid is modified by an additional -(NO_2) group (a, n = 5201 events); an equimolar mixture of Y_{R7} and (NO_2)-Y_{R7} peptides (b, n = 10615), (SO_3H_2)-Y_{R7} peptide, where the Y amino acid is modified by an additional -(SO_3H_2) group (c, n = 3183 events); an equimolar mixture of Y_{R7} and (SO_3H_2)-Y_{R7} peptides (d, n = 4227 events); (P)-Y_{R7} peptide, where the Y amino acid is modified by an additional -(P) group (e, n = 2626); and a mixture of Y_{R7} (1 μM final concentration) and (P)-Y_{R7} (10 μM final concentration) peptides (f, n = 10281 events). The chemical modification of the peptide leads to a shift of the mean I_b/I_0 value from 0.360 $\pm$ 0.002 (Y_{R7}) to 0.343 $\pm$ 0.002 ((NO_2)-Y_{R7}), 0.354 $\pm$ 0.002 ((SO_3H_2)-Y_{R7}) and 0.355 $\pm$ 0.002 ((P)-Y_{R7}). The mean value (respectively uncertainty) of relative residual current of each peptide was obtained as the mean value (respectively standard deviation) of a gaussian fit of the corresponding I_b/I_0 distribution; from single independent experiments. The data were acquired in 4 M KCl, 25 mM HEPES buffer, at 7.5 pH and 20.0 $\pm$ 0.5^{o}C, and under a –50 mV bias applied to the *trans* compartment.
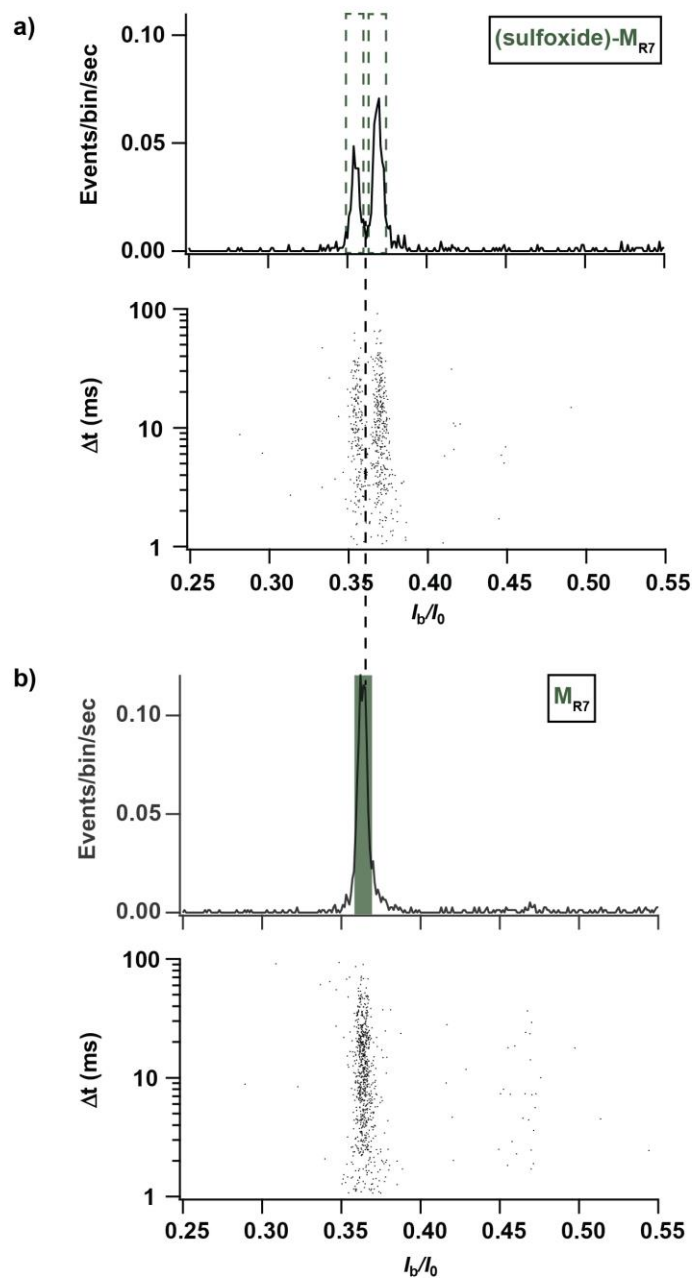
Supplementary Figure 11

Blockade currents of chemically modified M_{R7} peptide.

Histogram of I_b/I_0 (top plot of each panel) and scatter plot of blockade duration *versus* I_b/I_0 (bottom plot of each panel) obtained from nanopore experiments during analysis of: (sulfoxide)-M_{R7} peptide (a, n = 2374 events), where the M amino acid is modified by an additional -(sulfoxide) group; and M_{R7} peptide (b, n = 2317 events). The chemical modification of the peptide leads to the emergence of two main peaks of mean I_b/I_0 values of 0.355 ± 0.002 and 0.369 ± 0.003, different from the mean I_b/I_0 value of 0.363 ± 0.003 corresponding to M_{R7} peptide. The mean value (respectively uncertainty) of relative residual current of each peptide was obtained as the mean value (respectively standard deviation) of a gaussian fit of the corresponding I_b/I_0 distribution; from single independent experiments. The data were acquired in 4 M KCl, 25 mM HEPES buffer, at 7.5 pH and 20.0 ± 0.5°C, and under a –50 mV bias applied to the *trans* compartment.
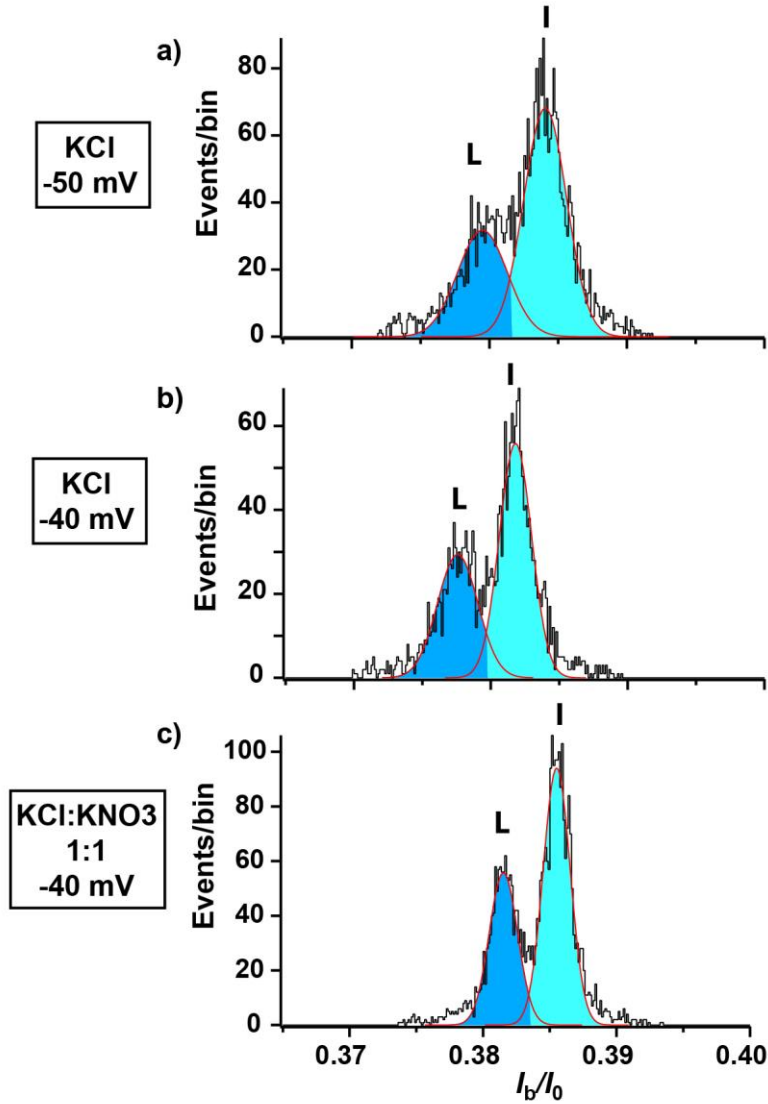
Supplementary Figure 12

Discrimination of L_{R7} and I_{R7} under optimized recording conditions.

Histograms of relative residual current I_b/I_0 measured using a low-noise setup (see Supplementary Note 3). The nanopore experiments were performed using an equimolar mixture of L_{R7} and I_{R7} peptides and under the following experimental conditions: –50 mV in 4M KCl (a, n = 3474 events); –40 mV in 4M KCl (b, n = 2373 events); –40 mV in 2M KCl/2M KNO$_3$ mixture (c, n = 3437 events). In 4M KCl, the diminution of the voltage magnitude from –50 mV to –40 mV leads to a reduction of the standard deviations of the two main populations, and a shift of the I_b/I_0 values from 0.380 ± 0.0018 to 0.378 ± 0.0013 (L_{R7}) and from 0.384 ± 0.0016 to 0.382 ± 0.0012 (I_{R7}). At –40 mV, the substitution of 50% of the Cl$^-$ ions by 50% of NO$_3^-$ ions further reduces the standard deviations of the two main populations now located at I_b/I_0 values of 0.382 ± 0.0010 (L_{R7}) and 0.386 ± 0.0010 (I_{R7}). Continuous lines show the underlying Gaussian distributions. The mean value (respectively uncertainty) of relative residual current of each peptide was obtained as the mean value (respectively standard deviation) of a gaussian fit of the corresponding I_b/I_0 distribution; from single independent experiments. The data were acquired at room temperature and under a voltage bias applied to the *trans* compartment.
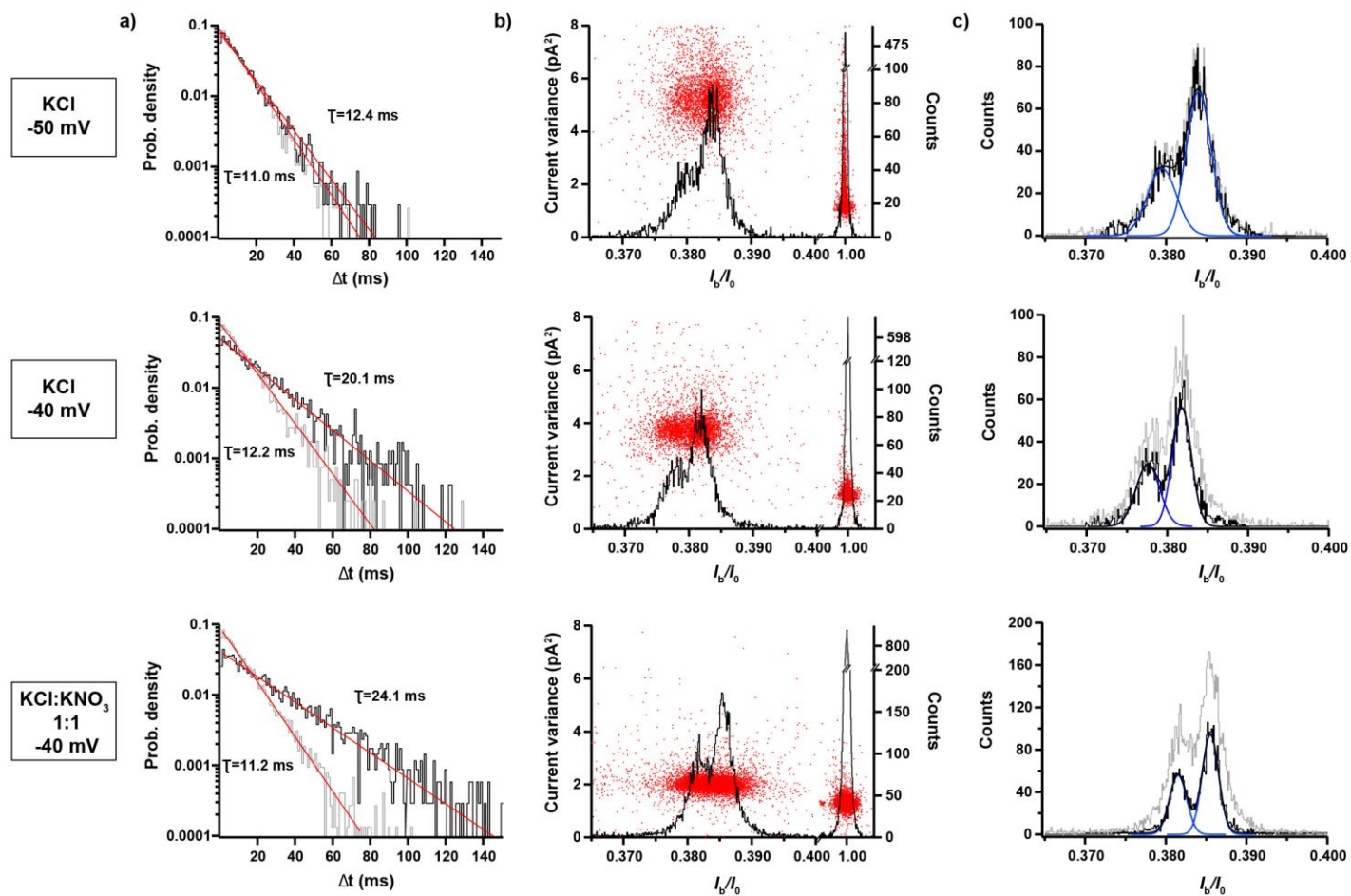
Supplementary Figure 13

Detailed analysis of blockade currents recorded using a low-noise setup.

All measurements were conducted using an equimolar mixture of L_{R7} and I_{R7} peptides under the following three conditions: 4 M KCl, –50 mV (top row); 4 M KCl, –40 mV (middle row); 2 M KCl/2 M KNO_3, –40 mV (bottom row). (a) Blockade duration histograms for individual levels (grey) and resistive pulses (black) with monoexponential fits (red lines). (b) Scatterplots of current variance *vs* I_b/I_0 for individual levels. Excess variance is determined as the difference between the center of mass of the point clouds for the blocked states and the open state ($I_b/I_0 \approx 1$). (c) Histograms values in terms of I_b/I_0 for individual levels (grey cityscape) and resistive pulses (black cityscape). Continuous black and blue lines are the sum of two Gaussians and the underlying Gaussian distributions, respectively. For each histogram, at least 1000 events were analyzed.
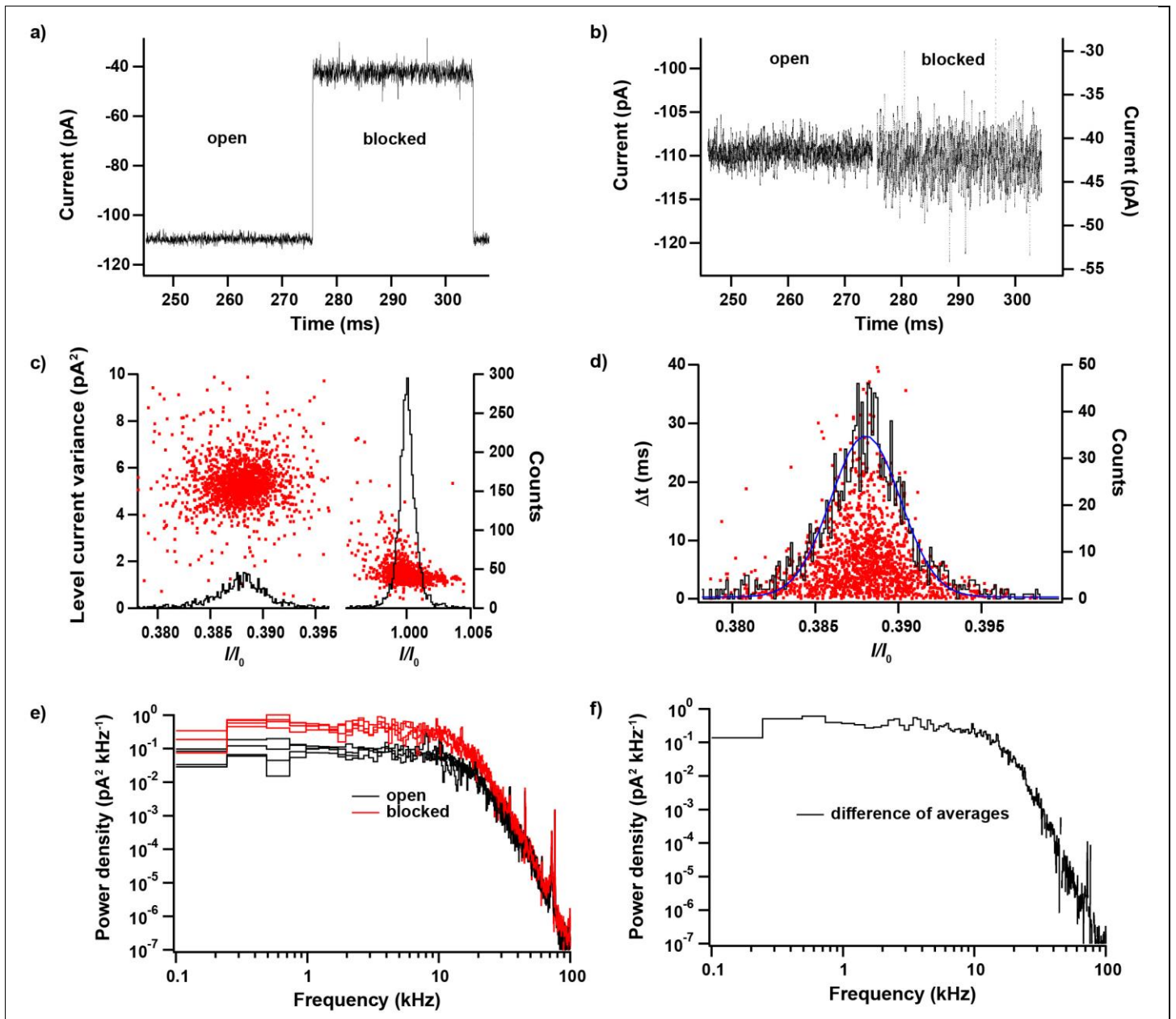
Supplementary Figure 14

Detailed analysis of the L_{R7} blockades measured using a low-noise setup.

(a) Representative fragment of ionic current recording showing an example of a transition from an open pore current state to a blockade current state. (b) Close-up of both levels from panel a to illustrate the increased noise of the blocked current state in comparison to the open pore current state. (c) Scatterplot (red dots) of the variance of detected current levels *versus* I/I_0. Note the approximately threefold increase in variance in the blocked state ($I/I_0 \approx 0.385$) with respect to the open state ($I/I_0 = 1$). (d) Histogram of the relative residual current I/I_0 (black trace), scatterplot of blockade duration *versus* I_b/I_0 (red dots) and a Gaussian fit of the histogram (red trace). Note that the estimated mean values get progressively closer to the population mean with longer dwell times, indicating that current noise and not inter-event-variance is responsible for the width of the Gaussian peak. For panels c & d, histograms and scatter plots are based on the analysis of 7309 events. (e) Power spectral density plots for four blocked levels and four preceding baseline levels (as in panels a and b). (f) Difference of the averages of the four blocked state and four open state spectra, respectively, from panel e. Note the flat curve between 100 Hz and 10 kHz (the cut-off frequency of the 4-pole Bessel filter used). The mean value (respectively uncertainty) of relative residual current of each peptide was obtained as the mean value (respectively standard deviation) of a gaussian fit of the corresponding I_b/I_0 distribution; from single independent experiments. The data were acquired in 4 M KCl, 25 mM HEPES buffer, at 7.5 pH and at room temperature, and under a –50 mV bias applied to the *trans* compartment, and with a 10 kHz filter and a 1 MHz sampling rate.
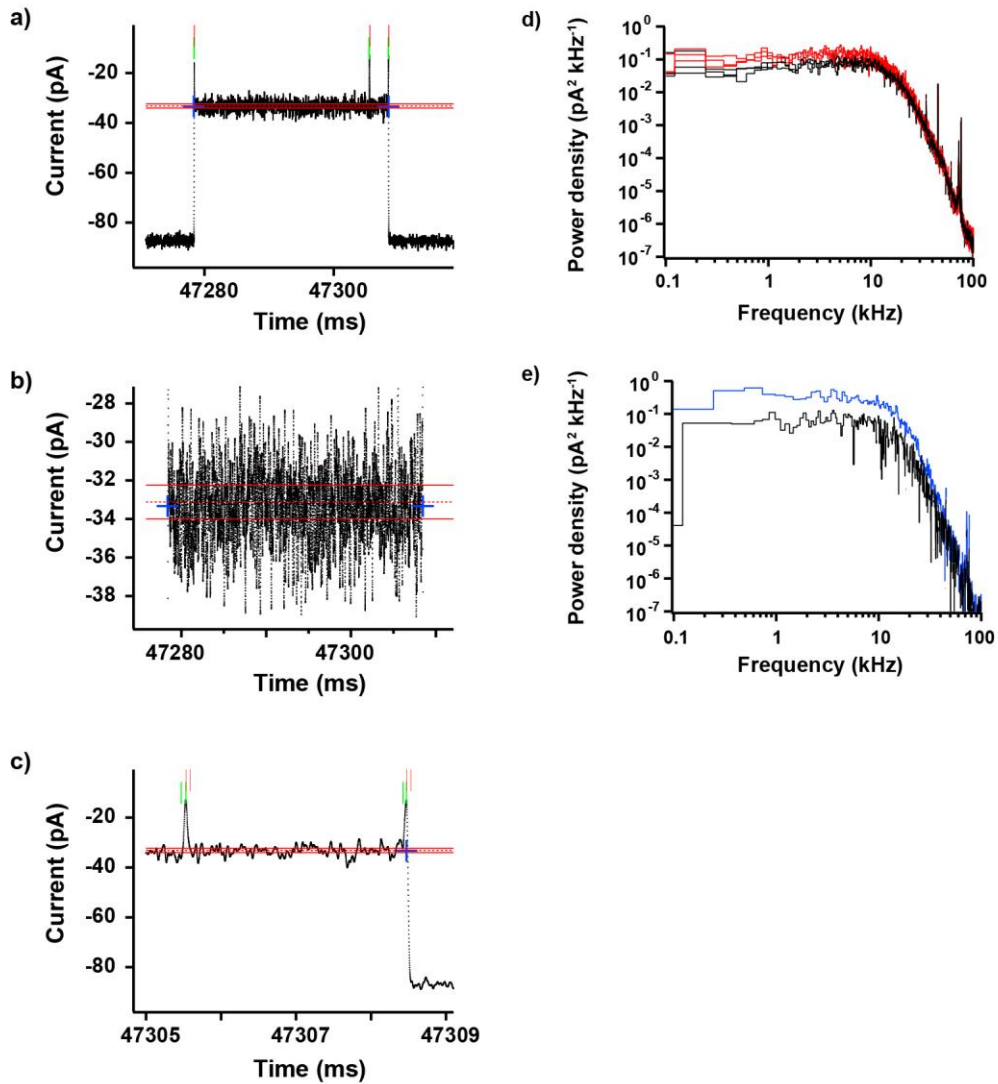
Supplementary Figure 15

Analysis of individual resistive pulses from low-noise ionic current measurements and the effect of solvent.

(a) Typical resistive pulse measured using a low-noise setup and an equimolar mixture of L_{R7} and I_{R7} peptides at –40 mV in 4 M KCl. This resistive pulse consists of the following five levels: two visits to the long-lasting level and three short visits to a deeper blocked level. The dashed red line corresponds to the current level at the midpoint between the two maxima for L_{R7} and I_{R7}: the red continuous lines show the extremes of the range of mean current levels that are included in determining the resistive pulse mean $\langle I_b/I_0 \rangle$ value. Blue markers show start and end of resistive pulse (vertical arm) and mean current value (horizontal arm). Green and red vertical lines show onset and end of transitions as detected by the DetectiVent algorithm. Vertical (b) and horizontal (c) expansions of (a) to show details. (d) Power spectral density of three blocked (red) and unblocked (black) current levels in the presence of 2 M KCl/2 M KNO_3. (e) Comparison of the difference of averages in 4 M KCl (blue, same as in Supplementary Fig. 14f) and 2 M KCl/2 M KNO_3 (black).

Supplementary Table

(a)

Peptides X_1R_7	$I_b/I_o(X_1R_7)$	$[X_1] = I_b/I_o(X_1R_7) - I_b/I_o(RR_7)$
VR_7	0.369	0.037

Peptides X_2R_7	$I_b/I_o(X_2R_7)$	$[X_2] = I_b/I_o(X_2R_7) - I_b/I_o(RR_7)$
KR_7	0.353	0.021
HR_7	0.362	0.03
ER_7	0.371	0.039
DR_7	0.383	0.051
RR_7	0.332	0

Peptides $X_1X_2R_7$	$I_b/I_o(X_1X_2R_7)$	$[X_1X_2] = I_b/I_o(X_1X_2R_7) - I_b/I_o(RRR_7)$
VKR_7	0.314	0.058
VHR_7	0.32	0.064
VER_7	0.336	0.08
VDR_7	0.343	0.087
VRR_7	0.293	0.037

(b)

X_1	X_2	$[X_1] + [X_2]$	$[X_1X_2]$
V	K	0.058	0.058
V	H	0.067	0.064
V	E	0.076	0.08
V	D	0.088	0.087
V	R	0.037	0.037

Supplementary Table 1

Summary of $X_1X_2R_7$ experiments.

(a) Mean relative residual current $\langle I_b/I_0 \rangle$ measured for the X_1R_7, X_2R_7, and $X_1X_2R_7$ peptide sequences and the contribution of the amino acid $[X_1]$, $[X_2]$ or $[X_1X_2]$ to the blockade current. The data were acquired in 4 M KCl, 25 mM HEPES buffer, at 7.5 pH and $20.0 \pm 0.5^{o}C$, and under a -50 mV bias applied to the *trans* compartment. (b) Comparison of the contributions of X_1X_2 to the $X_1X_2R_7$ blockade to the sum of individual contributions of X_1 and X_2 to the X_1R_7 and X_2R_7 blockades.

Supplementary Note 1: Experimental test of the volume exclusion mechanism

Our theoretical analysis suggests that the blockade current produced by the peptides in aerolysin is determined by the total excluded volume of the peptide. Therefore, one can expect the blockade current of an $X_1X_2R_7$ peptide (formed by the addition of an X_1 amino acid to an X_2R_7 peptide) to be a linear combination of the X_1R_7 and X_2R_7 blockade currents. We experimentally tested this prediction by measuring current blockades produced by an equimolar mixture of five $X_1X_2R_7$ peptides, where X_1 = V and X_2 = K, H, D, E, R, Supplementary Fig. 6a. For reference, Supplementary Fig. 6b shows blockade currents produced by an equimolar mixture of X_2R_7 peptides, where X_2 = K, H, D, E, R. The blockade current of the X_1R_7 peptide (with X_1 = V) is reported in the main text Fig. 1i. The results of these measurements are summarized in Supplementary Table 1a.

To determine the contribution of an X_1 amino acid (defined as $[X_1]$ henceforth) to the blockade current of X_1R_7, we subtract the baseline contribution from the arginine homopeptide consisting of eight amino acids i.e. $[X_1] = I_b/I_0(X_1R_7) - I_b/I_0(R_{R7})$. Similarly, we calculate $[X_2]$ as $I_b/I_0(X_2R_7) - I_b/I_0(R_{R7})$. For the $X_1X_2R_7$ experiments, the contribution of X_1X_2 is calculated by subtracting the blockade current of arginine homopeptide consisting of nine amino acids i.e. $[X_1X_2]$ = $I_b/I_0(X_1X_2R_7) - I_b/I_0(RR_{R7})$. In other words, $[X_1]$ is the shift in the mean I_b/I_0 value induced by mutating the C-terminal amino acid in an arginine octapeptide to V, while $[X_1]$ is the shift induced by mutating it to K, H, D, E or R. $[X_1X_2]$ is the shift in the mean I_b/I_0 value induced by mutating the two last C-terminal amino acids in an arginine nonapeptide so that the last amino acid is always a V and the penultimate amino acid can be K, H, D, E or R. What is shown here is that the sum of $[X_1]$ and $[X_2]$ agrees well with $[X_1X_2]$, so that we can conclude that the size-exclusion effect of two amino acids combines linearly. Note that, in the above expressions, we use the baseline blockade current of the arginine homopeptides of the same length as the probed sequence. We do that because our peptide identification strategy is sensitive to the total length of the peptide confined within the sensing region of aerolysin and not just to the individual amino acid, which is a fundamental difference from strand sequencing. From our homopeptide experiments, we found the average baseline blockade currents to be 0.332 and 0.256 for R_{R7} and RR_{R7}, respectively.

Supplementary Table 1b compares the contributions of X_1X_2 to the $X_1X_2R_7$ blockade to the sum of individual contributions of X_1 and X_2 to the X_1R_7 and X_2R_7 blockades. The excellent agreement of the $[X_1X_2]$ and $[X_1] + [X_2]$ values proves that the average blockade values are additive, reflecting the volume exclusion nature of the current blockade.

Supplementary Note 2: Single passage identification of X_{R7} peptides

In our work, the mean amino acid-specific blockade values, $\langle I_b/I_0 \rangle_X$, and their standard deviations, σ_X^{exp}, (Fig. 1e-j) were experimentally determined for each peptide species by first determining $\langle I_b/I_0 \rangle_{single}$, the mean blockade value for each of the $N_0 \sim 1000$ translocation events, each event lasting approximately $\tau^{exp} \sim 20$ ms. For each species, the resulting distribution of the $\langle I_b/I_0 \rangle_{single}$ values was fit by a Gaussian to determine the mean of the distribution, $\langle I_b/I_0 \rangle_X$, and its standard deviation, σ_X^{exp}. Supplementary Fig. 9a displays the gaussian fits, $P_L(I_b/I_0)$ and $P_I(I_b/I_0)$, to the experimental $\langle I_b/I_0 \rangle_{single}$ data recorded for L_{R7} and I_{R7} peptides, respectively, among which L_{R7} has the shortest average residence time of 17.5 ms, Supplementary Fig. 2c,d. For a binary mixture of L_{R7} and I_{R7}, a single translocation event can be identified as being produced by an L_{R7} peptide if its $\langle I_b/I_0 \rangle_{single}$ value is closer to $\langle I_b/I_0 \rangle_L$ than to $\langle I_b/I_0 \rangle_I$. The percentage of population identified correctly is 100 times the normalized area under the leucine gaussian for which $P_L/(P_L + P_I) > 0.5$. This area is shown as a shaded region (from $-\infty$ to the intersection of the two gaussians) in Supplementary Fig. 9a. For the wild type aerolysin nanopore, the probability of identifying L_{R7} from the L_{R7}/I_{R7} mixture is ~69.8%.

Next, we estimate the theoretical resolving power of our amino acid detection method using, as a free parameter, the residence time, τ, of individual peptides within the sensing volume of the aerolysin nanopore. We assume that increasing the residence time does not affect the mean, amino acid-specific blockade values, $\langle I_b/I_0 \rangle_X$ (which we experimentally found to be correct, see Supplementary Fig. 13c), the ionic current noise level and the peptide capture rate. We also assume that all twenty peptide-specific $\langle I_b/I_0 \rangle_X$ values have been determined to very high accuracy through independent calibration measurements. Increasing the peptide residence time τ is expected to make the P_L and P_I distributions narrower, increasing the likelihood of single passage peptide identification. According to our assumptions, these gaussians will have the same $\langle I_b/I_0 \rangle_L$ and $\langle I_b/I_0 \rangle_I$ values, but smaller standard deviations $\sigma_X(\tau) = \sigma_X^{exp}\sqrt{\tau^{exp}/\tau}$. Here we assumed that averaging over N translocation events of duration τ is equivalent to averaging over one translocation event of duration $N\tau$. Supplementary Fig. 9b shows P_L and P_I distributions for the peptide residence time of 200 ms. According to these distributions, the probability of identifying L_{R7} from the L_{R7}/I_{R7} mixture is 99.9%. Repeating such calculations for various values of τ produced data shown in the main text Fig. 3e. Similar calculations for distinguishing Y_{R7} and F_{R7} peptides are shown in Supplementary Fig. 9c.

We have applied the same theoretical model to estimate single passage distinguishability of individual peptides in the presence of all twenty peptide species. The mean values of the twenty gaussians arranged in an increasing order are shown in Fig. 3f. The distinguishability of an X_{R7} peptide is conditioned by the proximity of its $\langle I_b/I_0 \rangle_X$ value to the nearest $\langle I_b/I_0 \rangle$ values of other peptide species. To estimate the number of correctly identified single passage translocations of peptide X_{R7}, we find the intersection points of the X_{R7} gaussian with the two nearest gaussians, P1 and P2. The area under the X_{R7} gaussian between point P1 and P2, normalized with its total area, gives the percentage of correctly identified peptides for a given peptide residence time. Doing the same calculation for all twenty amino acid species, except for the end points R_{R7} (P1 = $-\infty$) and G_{R7} (P2 = ∞), results in the bar graph shown in Fig. 3g. Note that the P_{R7} peptide was found to produce two peaks in experiment recordings, see Fig. 1j. To simplify our analysis, we used the higher, more probable value of $\langle I_b/I_0 \rangle_P$, 0.3793.

Our calculations thus far were based on experimentally derived distributions of $\langle I_b/I_0 \rangle_{single}$, which is influenced by many noise sources. For illustrative comparison, we estimate single passage identification of L_{R7} peptides from the L_{R7} / I_{R7} mixture under ideal conditions of ion counting being the only noise source. In our experiments, the typical blockade current is of the order of 10 pA, which corresponds to the passage of 62,500,000*τ ions for a current blockade of duration τ (expressed in the units of seconds). The standard deviation of the relative residual current is then $5.05 \times 10^{-5}/\sqrt{\tau}$. Using the same process as before, we determined the likelihood of correctly identifying a single passage of L_{R7} peptide from an equimolar L_{R7} / I_{R7} mixture, Supplementary Fig. 9d. Thus, in the absence of other noise sources, a 5 ms residence time would be sufficient to correctly identify 99% of individual L_{R7} translocations from an equimolar L_{R7} / I_{R7} mixture.

Supplementary Note 3: Improving single passage differentiation by optimizing the apparatus and recording conditions

While both nanopore engineering (Supplementary Note 2) and chemical modification of amino acids (Supplementary Figs. 10 and 11) are promising avenues toward improving the amino acid identification, here we show that accuracy of amino acid identification can also be improved by increasing fidelity of the ionic current measurement. As a test case, we focus here on the discrimination of I and L peptides, for which we had found a bimodal distribution with considerable overlap, Fig. 3d, and a < 70% probability of correct peptide identification (over a 30% error).

Using chip based lipid membranes spanning 50 μm-diameter apertures (MECA16, Ionera Technologies GmbH, Freiburg, Germany), we were able to reduce open pore current noise fivefold to ≤ 1 pA r.m.s. at 0-10 kHz bandwidth, thus achieving > 85% correct identification for both peptides (see Supplementary Fig. 12a) with a characteristic resistive pulse dwell time of 12.4 ms at −50 mV (Supplementary Fig. 13a). This improvement was accompanied by a substantial decrease of the standard deviation of the fitted Gaussian distributions from σ > 0.005 to σ = 0.0018 (L) and 0.0016 (I). At the same time, these low-noise measurements revealed an important component: intrinsic excess low-frequency noise during the blocked state of the pore (Supplementary Fig. 14) which limits the utility of further reductions of noise originating in the recording apparatus.

As mentioned before, the simplest means of increasing discrimination in the presence of low-frequency noise is prolongation of dwell times. Indeed, longer resistive pulses have means that lie increasingly closer to the population mean (Supplementary Fig. 14d), showing that uncertainty in determining the true mean current level rather than inter-event variance underlies the width of distributions.

We therefore decided to repeat the low-noise measurement at the voltage giving the longest dwell times for arginine homopeptides with wild-type aerolysin in Piguet et al.[1] (−40 mV). With careful analysis procedures to remove the influence of short-lived substates occurring at increased frequency at this voltage (Supplementary Fig. 15), this yielded single passage identification at > 92% (L) and > 99% (I) with a characteristic resistive pulse dwell time of 20.1 ms (Supplementary Fig. 12b). At −40 mV, Gaussian fits had standard deviations of I_b/I_0 = 0.0013 (L) and 0.0012 (I).

Besides voltage, an important degree of freedom of nanopore sensing is the ionic composition of the electrolyte solution. Several ion-specific effects have been observed ranging

from prolonging dwell-times of synthetic polymers with fluoride anions[2] to slowing down translocation of DNA with lithium or tetraethylammonium cations[3,4]. These effects may be due to specific ions more or less strongly acting as counterions with charges on the analyte or on the pore wall, thereby modifying analyte-pore interactions. We hypothesized that the excess low-frequency noise produced by L_{R7} and I_{R7} peptides might be reduced by modifying the charge density of the anions, and, therefore, chose to replace 50% of Cl^- by the less densely charged NO_3^- anion. As shown in Supplementary Fig. 15d,e and Supplementary Fig. 13, this resulted in a clear reduction in excess current variance (from approximately 4 to 2 pA^2). The effect on discrimination is illustrated in Supplementary Fig. 12c: single passage identification was >95% for both peptides with a characteristic dwell time of 24.1 ms. Standard deviations of Gaussian fits were further reduced to σ = 0.0010 for both L and I peptides.

In summary, we have shown in this section that discrimination of I and L as trailing amino acids can be improved considerably even using the wild-type aerolysin pore by optimizing both apparatus and recording conditions. Taken together with the potential of improvement of the sensing capabilities of the sensor itself using site-directed mutagenesis, in our judgement, this augurs well for future employment of this method in protein characterization and sequencing.

[1] Piguet, F. *et al.* Identification of single amino acid differences in uniformly charged homopolymeric peptides with aerolysin nanopore. *Nat. Commun.* **9**, 966 (2018).

[2] Rodrigues, C. G., Machado, D. C., da Silva, A. M., Júnior, J. J. & Krasilnikov, O. V. Hofmeister effect in confined spaces: halogen ions and single molecule detection. *Biophys. J.* **100**, 2929–2935 (2011).

[3] Kowalczyk, S. W., Wells, D. B., Aksimentiev, A. & Dekker, C. Slowing down DNA translocation through a nanopore in lithium chloride. *Nano Lett.* **12**, 1038–1044 (2012).

[4] Wang, Y., Yao, F. & Kang, X.-f. Tetramethylammonium-filled protein nanopore for single-molecule analysis. *Anal. Chem.* **87**, 9991–9997 (2015).
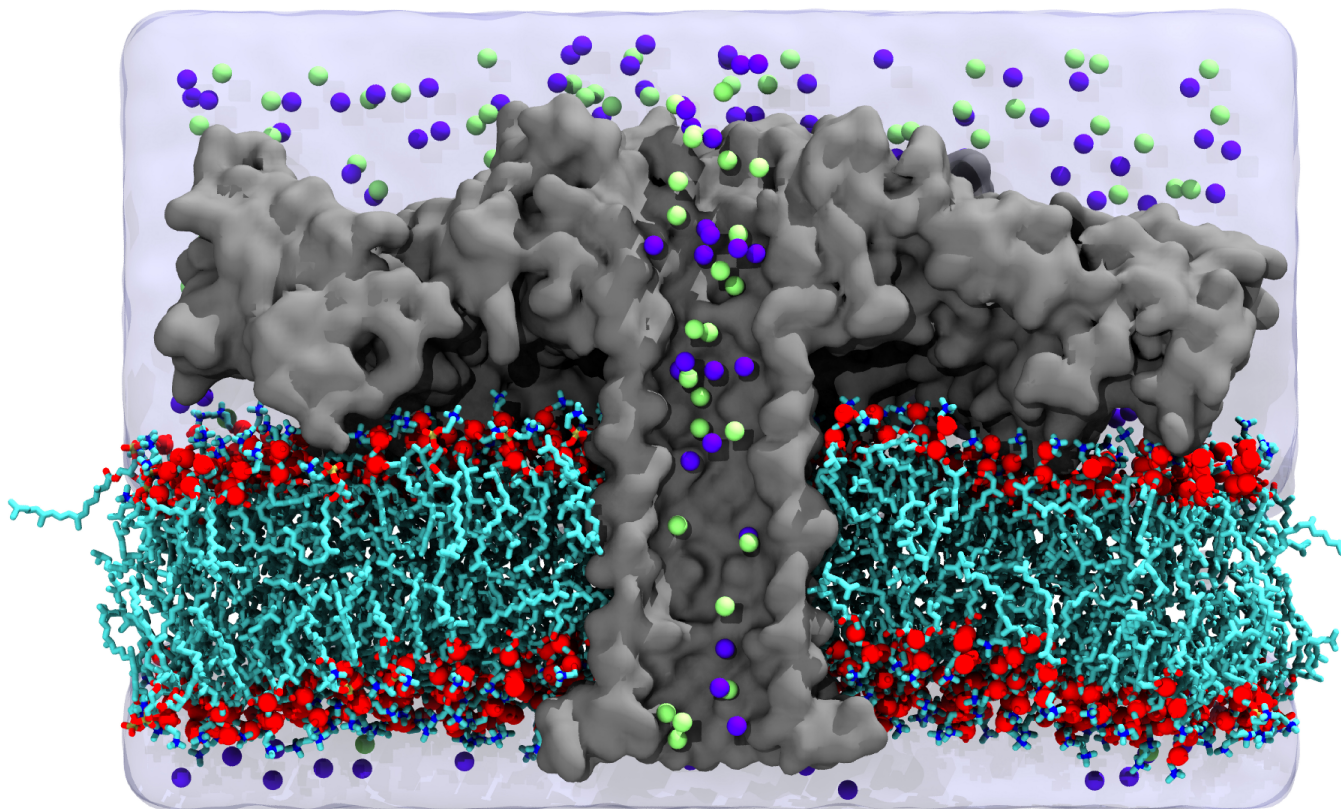
Captions to supplementary videos
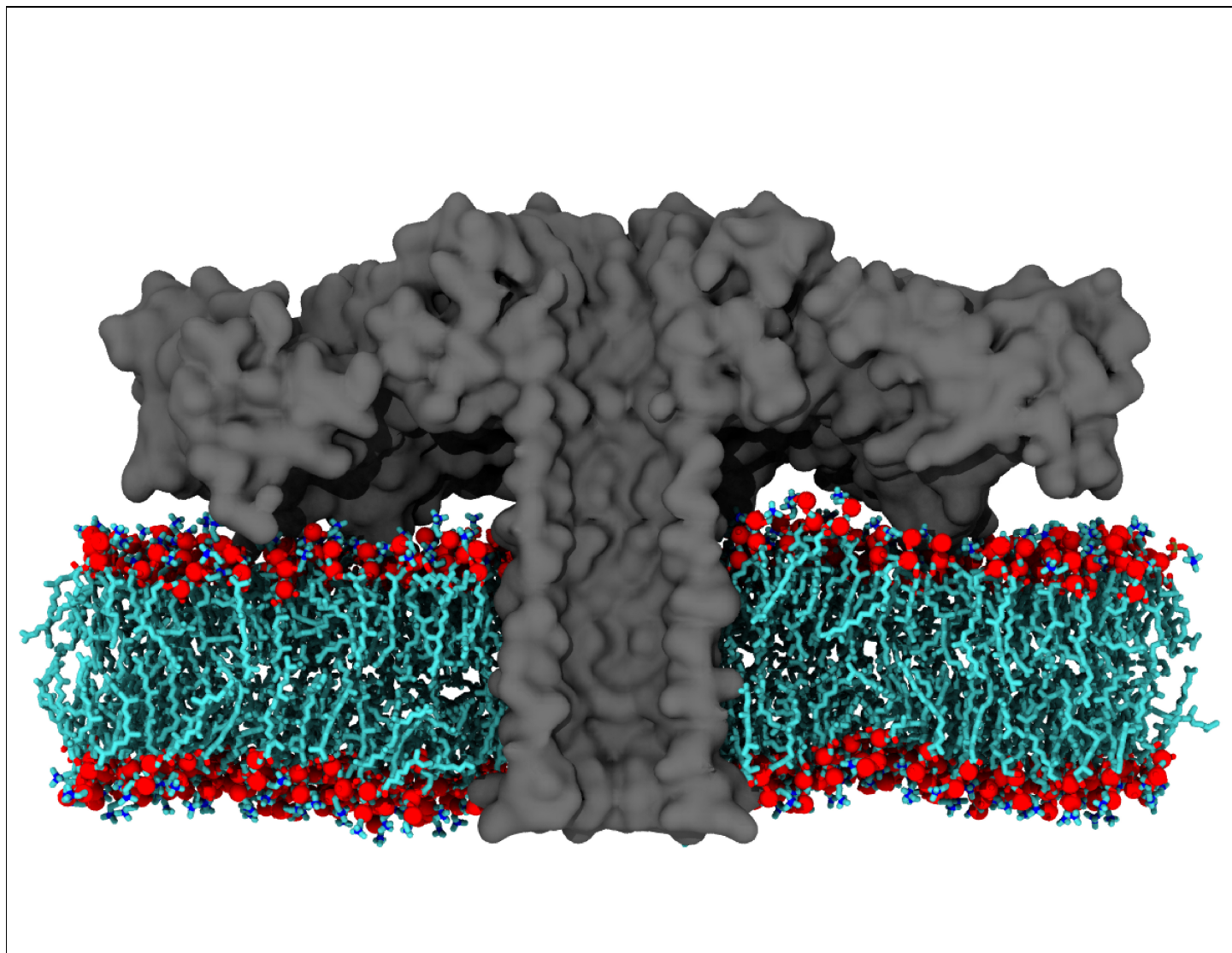
Supplementary Video 1

Equilibration of aerolysin nanopore in DPhPC membrane.

The video illustrates a 10 ns MD trajectory of the aerolysin system consisting of the aerolysin nanopore (grey, cut-away molecular surface) embedded in a DPhPC lipid bilayer (red for head groups and cyan for tails) and 1 M KCl solution (not shown for clarity).
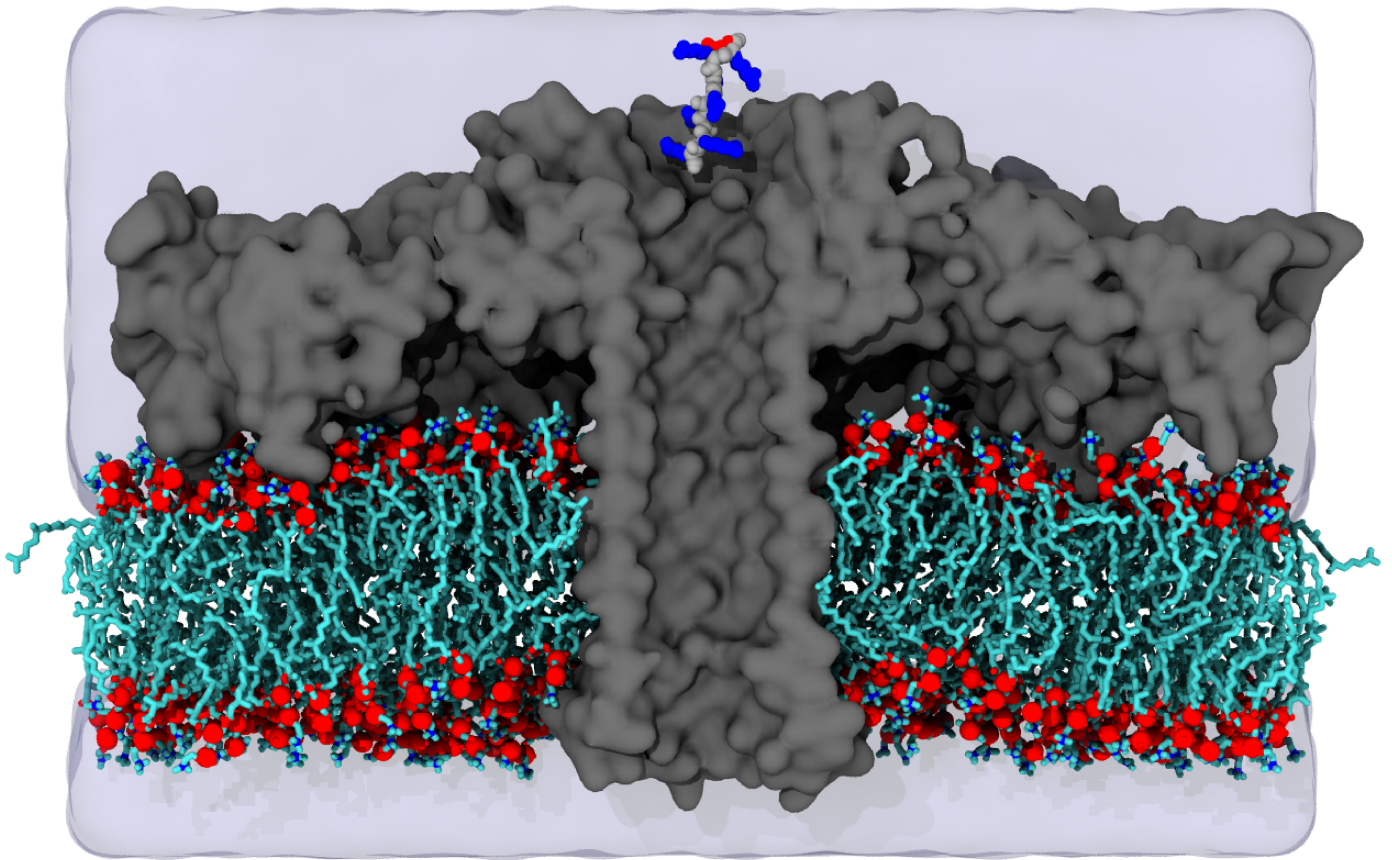
Supplementary Video 2

MD simulation of open pore ionic current.

The video illustrates a 2 ns fragment of an MD trajectory of the aerolysin system under a transmembrane bias of -100 mV. The aerolysin nanopore is shown using a grey, cut-away molecular surface; the head groups and tails of the DPhPC lipid bilayer are shown in red and cyan, respectively; the volume occupied by 1 M KCl electrolyte is represented by a blue translucent surface; potassium and chloride ions are shown as violet and green spheres, respectively.

Supplementary Video 3

SMD simulation of R_{R7} peptide translocation through aerolysin.

The video illustrates a 100 ns MD trajectory where an R_{R7} peptide was driven through the aerolysin nanopore using the SMD protocol. The aerolysin nanopore is shown as a grey, cutaway molecular surface; the DPhPC lipid bilayer is shown in red (head groups) and cyan (lipid tails); the backbone of the R_{R7} peptide is shown in grey while the side chains of all arginine residues are shown in blue except for the last one, which is shown in red.