Electronic Supplementary Information for:

Nanopore Sensing of Single-Biomolecule: a New Procedure to Identify Protein Sequence Motifs from Molecular Dynamics

Adrien Nicolaï, Aniket Rath, Patrice Delarue and Patrick Senet

Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 6303 CNRS-Université Bourgogne Franche-Comté, 9 Av. A. Savary, BP 47 870, F-21078 Dijon Cedex, France; E-mail: adrien.nicolai@u-bourgogne.fr

- MD simulations details
- Statistical analysis of sensing events
- Mechanisms of translocation: electrostatic *vs* steric effects
- Adding Noise to ionic current time series computed from MD
- Figure S1
- Figure S2
- Figure S3
- Figure S4
- Movie S1: Snapshots extracted from MD simulation of translocation of KTKEGV peptide through SL-MoS_2 nanopore. Applied voltage is $V = 1$ V. The color code is the same as in Fig. 1 of the main text. Snapshots every $\Delta t = 1$ ns are represented for a total length $T = 280$ ns. Ionic current is shown as inset.

MD simulation details

Initially, MoS_2 single-layer is constructed using 2D unit cell lattice vectors $\vec{a} = (3.1, 0, 0)$ Å and $\vec{b} = (0, 5.4, 0)$ Å. Each rectangular unit cell for MoS_2 is comprised of 6 atoms, 2 Mo and 4 S. The Mo-S bond length was taken as $d_{Mo-S} = 2.4$ Å and the S-S distance was taken as $d_{S-S} = 3.2$ Å. MoS_2 pores were constructed by removing atoms from the 2D membrane whose coordinates satisfy $x^2 + y^2 < R^2$, where $D = 2R$ is the diameter of the pore and considering the center of the pore at the origin of the box. The pore diameter considered in this work is $D = 1.5$ nm. In total, the simulation box is comprised of around 100,000 atoms. In the present simulations, we do not consider the membrane to be rigid. Consequently, a Stillinger-Weber potential[1] is used to simulate the dynamics of Mo-S bonded interactions. The peptide is modeled using the Amber ff99SB-ILDN force-field[2]. The water model used in the present work is the TIP3P model[3]. Non-bonded interactions between MoS_2, peptide atoms, water and ions are described using a Lennard-Jones (LJ) plus a Coulomb potential. LJ parameters for Mo and S atoms were taken from reference[4]. Partial charges of Mo and S atoms of the nanoporous membrane were computed in vacuum using ReaxFF[5,6] and charge equilibration algorithm[7,8], available in LAMMPS. The parameters for ReaxFF were taken from reference[9]. LJ parameters for K^+ and Cl^- ions were taken from reference[10] where specific parameters were developed for the water model employed here. Originally, the biomolecule is placed above the MoS_2 nanoporous membrane at a vertical distance of 20 Å. By doing that, we avoid a common biased threading when the peptide is originally placed into the pore. Before running MD simulations, an equilibration of the system in the NPT ensemble ($T = 300$ K and $P = 1$ bar) without any applied electric field was performed during 100 ps to relax the system at the target temperature and volume. Equilibration was followed by MD production runs of 500 ns duration, which were carried out in the NVT ensemble using the velocity-Verlet algorithm[11] with a time step of 2 fs. A Nosé-Hoover thermostat[12,13] was used to maintain the temperature at 300 K with a time constant of 0.1 ps. Particle-particle particle-mesh method[14] was used to describe long-range electrostatic interactions. A cutoff of 1.0 nm was applied to LJ and Coulomb potentials for non-bonded interactions. A SHAKE algorithm[15,16] was used to constrain bonds of the peptide containing H atoms and the bond length and angle of TIP3P water molecules.

Statistical analysis of sensing events

Fig. S2 shows characteristics of one specific event (index 41, $\tau_d = 19.5$ ns and $\Delta I^{MAX} = 0.34$ nA, see Fig. S1A). This event is associated to the presence of residues T2, K3, E4 and G5 inside the pore (see Fig.S2A). Particularly, T2 and K3 residues are in the pore at the beginning of the event and completely translocate through the pore after 5 ns. At the same time, E4 and G5 residues enter the pore and stay in for 3 ns. Then, T2 and K3 residues come back inside the pore, translocating in the opposite direction compared to the electric field direction and a second translocation occurs around 10 ns. From the sequence of sub-events described above, we computed the probability to find a single amino acid inside the pore within an event and its associated average current drop $< \Delta I >$ (Fig. S2B). On one hand, the residue that resides in the pore for the longest period of time is E4 ($\sim 70\%$), which is a negatively charged amino acid. In addition, surprisingly G5, which is the smallest amino acid, stays in the pore for $\sim 65\%$ of the time within an event.

On the other hand, T2 and K3 stays in the pore during ~ 36 and 30% of the event dwell time, respectively. Average current drops are around 0.1 nA for T2 and K3 and around 0.2 nA for E4 and G5. The fact that the dwell time per residue and average current drops are very similar for E4, G5 and for K2, T3 comes from the fact that not only one amino acid resides in the pore at a time, although SL-MoS_2 is only 3.1 Å thick. In that sense, it is better suited when analyzing nanopore translocation data to discuss about the presence of protein sequence motifs in the pore than single amino acids. For example, the fact that G5 behaves similarly to E4 comes from the fact that they are bonded to each other via the peptide bond. However, not only residues connected via peptide bond form sequence motif. The local folding/unfolding of the peptide plays a role in the protein sequence motif detected. Fig. S2C represents the same data as the one presented in Fig. S2B but along protein sequence motif instead of single amino acid. For the specific event presented in Fig.S2, a total of 6 motifs are associated to the event, the two most probable being E4-G5 ($\sim 60\%$) and T2-K3 ($\sim 30\%$). At the transition between the full translocation of T2-K3 through the pore, other motifs such as T2-E4 and T2-E4-G5 are observed but do not last for a long time inside the pore. These two motifs are examples of motifs made of 2 or 3 amino acids that are not following each other in the primary sequence.

Mechanisms of translocation: electrostatic *vs* steric effects

The volume and charge due to the presence of the peptide inside the pore presented in Fig. S3 was computed as follows: at each time step t of MD trajectories, atoms of the peptide in the pore, (with radial coordinate $\rho \leq R$ and normal coordinate $|z| \leq h^*$) were identified. The volume time series $V(t)$ was computed as the sum of the van der Waals volume of each atom. The charge time series $q(t)$ was computed from the partial charge of each atom given by the Amber ff99SB-ILDN force-field[2]. Then, for each of the 49 events presented in Fig. S1A, we identified the starting time of the event, t_s and the variation of volume and charge are given by:

$$\Delta V(t) = V(t) - V(t_S) \text{ and } \Delta q(t) = q(t) - q(t_S)$$

Fig. S3A shows the variation of volume and charge for the event already presented in Fig. S2. Overall, the temporal correlation between ΔI and ΔV or Δq is not obvious. For instance, at $\tau_d = 8$ ns, E4G5 motif that is inside the pore is replaced by T2K3, which involves a decrease of ionic current. In the meantime, the volume inside the pore increase since glycine residue G5 left the pore. Therefore, this particular drop of current cannot be associated with ion blockade due to the presence of the peptide only since the occupied volume is smaller than before the current decreases. It is even more drastic for $\tau_d > 10$ ns, with a large increase of current followed by a plateau at $\tau_d = 15$ ns, the volume of the peptide in the pore being constant (E4G5 residues inside). This bump observed in the ionic current is specifically due to the decrease of the charge inside the pore. However, by computing the total correlation for both ΔV and Δq properties with ΔI, *i.e.* $\rho(\Delta I, \Delta V)$ and $\rho(\Delta I, \Delta q)$ respectively, it appears that all different scenarios exist, *i.e.* drops are due to increase/decrease of volume or increase/decrease of charge. Statistically, 35% of the events are correlated to an increase of volume and a decrease of charge, 27% are correlated to a decrease of volume and charge, 24% are correlated to a decrease of volume and an increase of charge and 14% are correlated to a decrease of volume and charge simultaneously. However, a majority of events (60%) show that volume and charge variation are anti-correlated. In addition, some events are characterized by a very low correlation with volume and charge variation, which means that other mechanisms than electrostatic end steric effects exist, as imagined originally. Finally, as already stated in the previous section, protein sequence motifs associated with drops of currents are more relevant signatures of peptide sequencing. For this reason, we computed the temporal correlation between current drops and volume/charge variations in the pore for sub-events associated with protein sequence motifs. We focused our analysis only to the four motifs detected from the largest dwell time probabilities, i.e. EG, E, TK and EGV. Same trends are observed as for all the events, with all correlation scenarios existing. The only difference that we were able to highlight to separate motifs is the fact that for E and EG motifs, ΔV and Δq are correlated and for EGV and TK, it is the opposite (Fig. S3C).

Adding Noise to ionic current time series computed from MD

We tested the procedure based on PE on MD ionic current time series by adding Gaussian white noise to $I(t)$ signal. The standard deviation of the Gaussian noise was chosen to be 0.2 nA, the same standard deviation as the one in an open pore configuration. Results are presented in Fig.S4. After adding Gaussian noise to ionic current time series, the data were filtered in order to remove high-frequency fluctuations. The procedure is described in the main text. From the filtered signal, we applied PE procedure to identify regular patterns in sensing events. results are shown for event 41 of duration 19.5 ns, the same as the one chosen to illustrate the procedure in the main text. As shown in Fig. S4B, PE follows the same trend for data with additional noise as for original data. Moreover, PE values are larger, which can be explained by the fact that the probability of patterns π_j decreased due to the addition of random noise in the data. Then, we extracted from data with additional noise sub-events for which PE is lower than $< PE > -\sigma_{PE} = 0.43$. As shown in Fig. S4B, 4 sub-events were detected corresponding to regular linear drops of ionic current. Comparted to original data without additional noise, sub-events are identical, and the same regular patterns are detected with similar slopes. Finally, the correlation coefficient of linear fitting R^2 is larger than 0.97 for all the sub-events detected from data with additional noise.

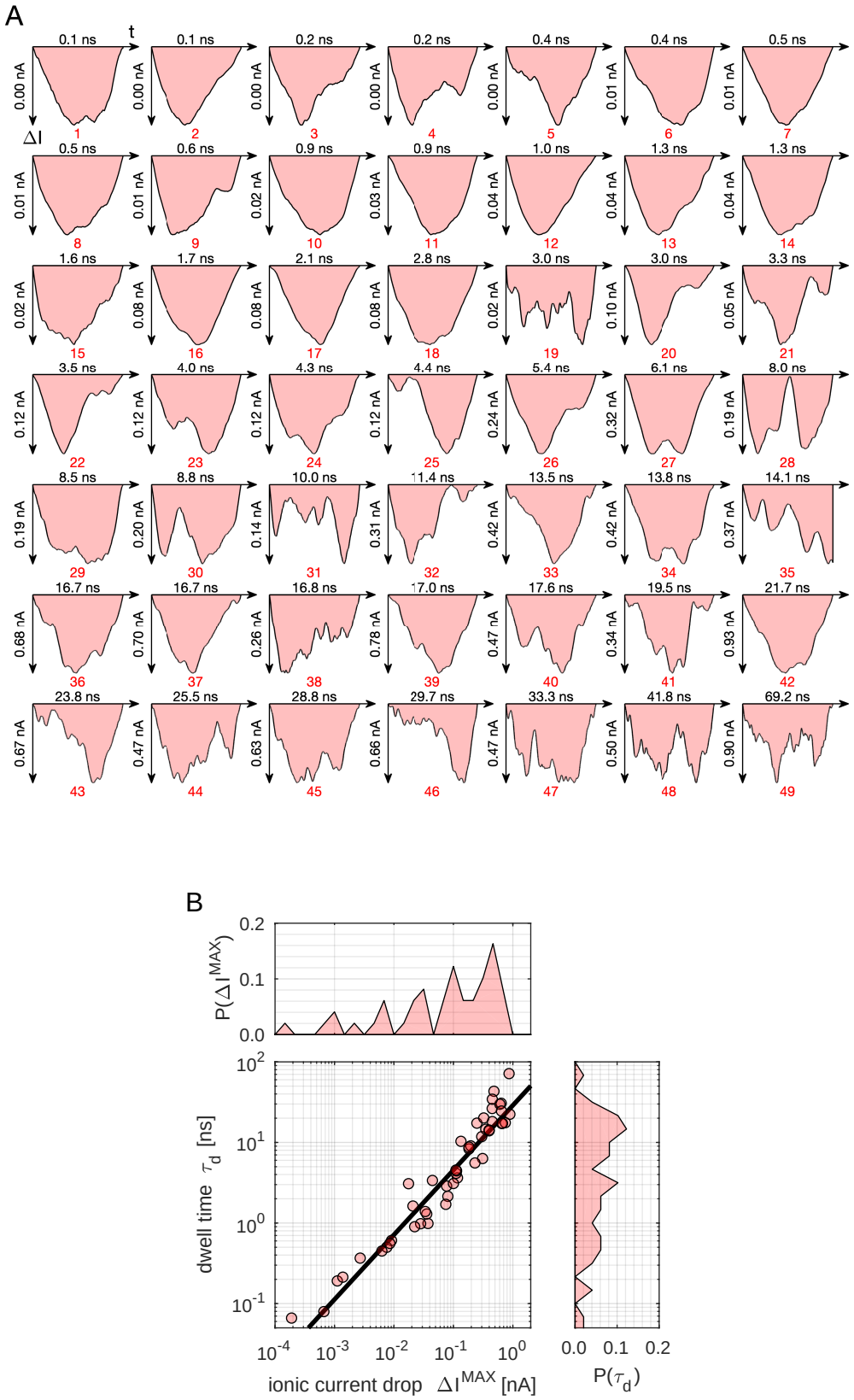

Fig. S1 A) Ionic current *vs* time signatures of the 49 sensing events detected in the present work for the MD concatenated trajectory of KTKEGV peptide (Fig. 2A of the main text). Event index is given in red. B) Dwell time τ_d *vs* maximum ionic current drops ΔI^{MAX} recorded using a $5\sigma_o$-threshold. Probability distributions are also shown in gray on the graphic sides. Black thick line represents the correlation observed between both characteristics of events.

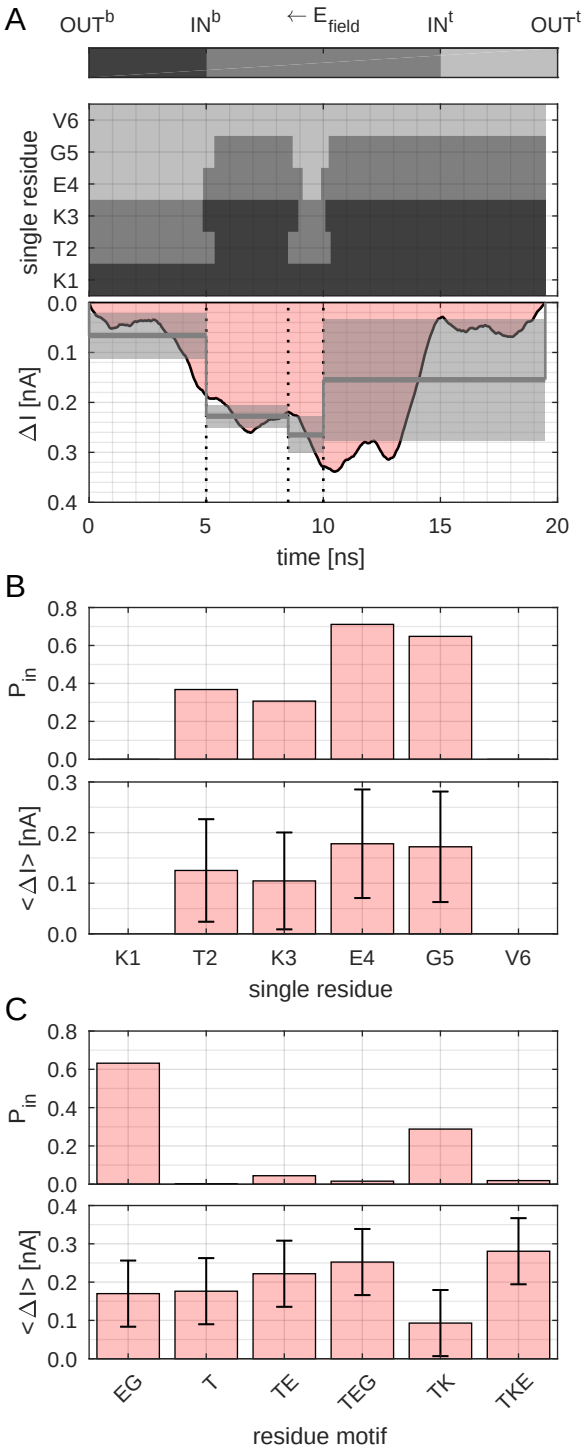

Fig. S2 A) Position (top panel) and current drop (bottom panel) *vs* time for a specific event (index 41, see Fig. S1A). In the top panel, the color code is the following: peptide outside the pore at the bottom surface (dark gray), peptide inside the pore (gray) and peptide outside the pore at the top surface (light gray). In the bottom panel, dashed lines represent sub-events described in the main text. Horizontal bronze lines represent the average current per sub-event. B) Probability to find a single amino acid inside the pore within an event (top panel) and average current drop [in nA] associated to it. Error bars correspond to the standard deviation. C) Same as panel B but for protein sequence motifs.

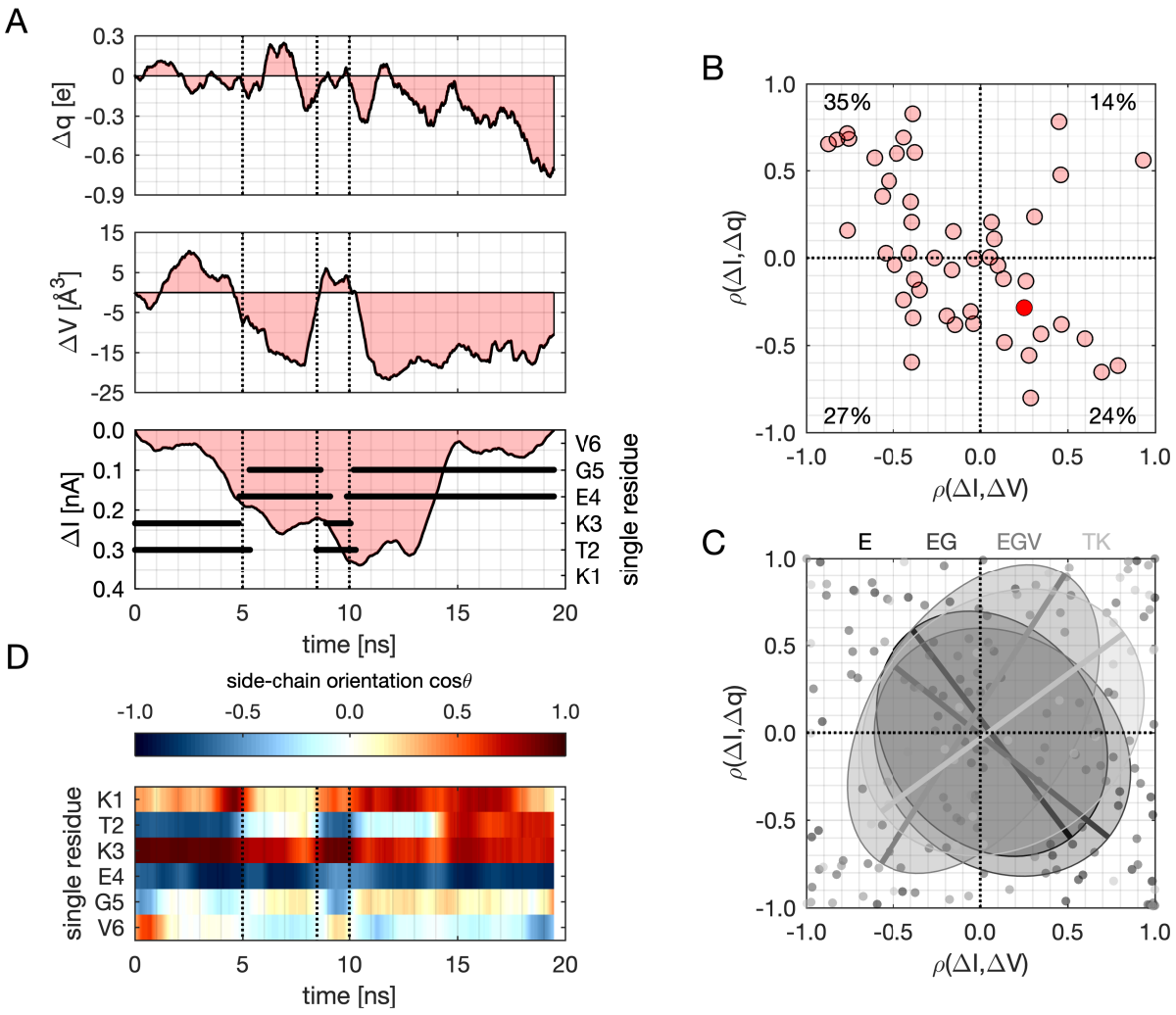

Fig. S3 A) Variation of charge (top panel), of volume (middle panel) and of current (bottom panel) *vs* time for event 41 also presented in Fig.S2. Black thick lines represent the presence of single amino acid inside the pore during the event. B) Correlation map $\rho(\Delta I, \Delta q)$ *vs* $\rho(\Delta I, \Delta V)$ for all events presented in Fig. S1A for KTKEGV peptide. The red circle represents event 41 shown in panel A. C) Correlation map $\rho(\Delta I, \Delta q)$ *vs* $\rho(\Delta I, \Delta V)$ associated with protein sequence motifs for the same data shown in panel B. Ellipses represent the confidence ellipses and thick lines represent the principal axis obtained from covariance analysis. D) Side-chain orientation of amino acids according to the electric field *vs* time for event 41 also presented in Fig.S2 (top panel).

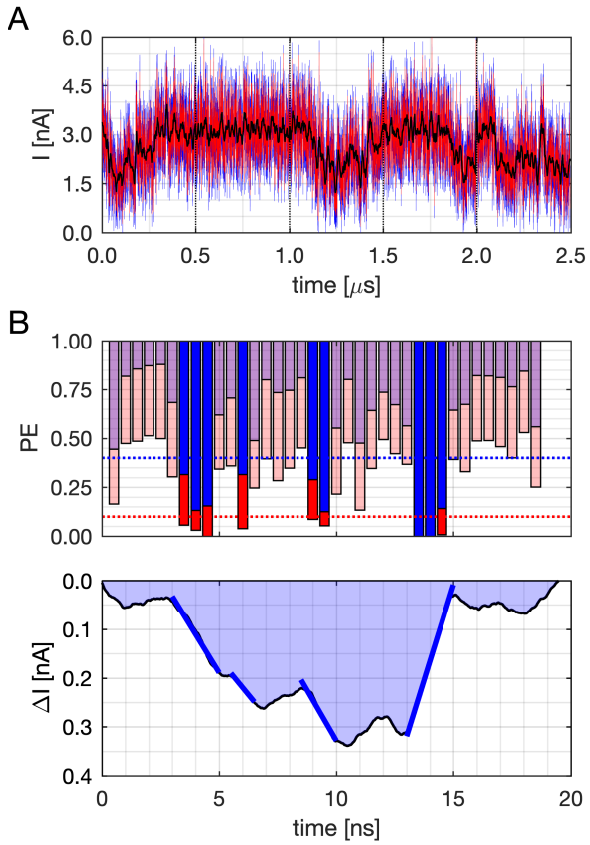

Fig. S4 A) Ionic current I [in nA] as a function of time [in μs] computed from five independent concatenated MD simulations (500 ns each) of the translocation of KTKEGV peptide through SL-MoS_2 nanopore. Red and blue thin lines represent raw data without and with addition of white noise, respectively. Black thick line corresponds to filtered data containing the addition of noise. B) PE *vs* time (top panel) for a specific sensing event (index 41, Fig. S1A) of 19.5 ns duration. Ionic current drop *vs* time (bottom panel) is also shown for comparison. Histogram with light red/red and light blue/blue bars correspond to the signal without (Fig. 4B of the main text) and with an additional white Gaussian noise, respectively. Red and blue dashed lines correspond to parts of the signal for which PE< 0.11 (without white noise) and PE< 0.43 (with white noise), respectively.

References

1 J.-W. Jiang, *Nanotechnology*, 2015, **26**, 315706.

2 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins*, 2010, **78**, 1950–1958.

3 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *The Journal of Chemical Physics*, 1983, **79**, 926–935.

4 Z. Gu, P. D. Luna, Z. Yang and R. Zhou, *Phys. Chem. Chem. Phys.*, 2017, **19**, 3039–3045.

5 K. Chenoweth, A. C. T. van Duin and W. A. Goddard, *J. Phys. Chem. A*, 2008, **112**, 1040–1053.

6 H. M. Aktulga, J. C. Fogarty, S. A. Pandit and A. Y. Grama, *Parallel Computing*, 2012, **38**, 245–259.

7 A. K. Rappe and W. A. Goddard, *J. Phys. Chem.*, 1991, **95**, 3358–3363.

8 A. Nakano, *Computer Physics Communications*, 1997, **104**, 59–69.

9 A. Ostadhossein, A. Rahnamoun, Y. Wang, P. Zhao, S. Zhang, V. H. Crespi and A. C. T. van Duin, *J Phys Chem Lett*, 2017, **8**, 631–640.

10 I. S. Joung and T. E. Cheatham, *J Phys Chem B*, 2008, **112**, 9020–9041.

11 W. C. Swope, H. C. Andersen, P. H. Berens and K. R. Wilson, *The Journal of Chemical Physics*, 1982, **76**, 637–649.

12 S. Nosé, *The Journal of Chemical Physics*, 1984, **81**, 511–519.

13 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.

14 R. E. Isele-Holder, W. Mitchell and A. E. Ismail, *The Journal of Chemical Physics*, 2012, **137**, 174107.

15 J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *Journal of Computational Physics*, 1977, **23**, 327–341.

16 H. C. Andersen, *Journal of Computational Physics*, 1983, **52**, 24–34.