Supplementary Information

Mechanism of polypeptide translocation through gold nanopores in view of sequencing applications

Mirko Vanzan,[1,2] *Agostino Migliore,*[1,*] *Maria Blanco-Formoso,*[3,*] *Francesco De Angelis,*[3] *Stefano Corni*[1,4,*]

[1] Department of Chemical Sciences, University of Padova, Via Marzolo 1, 35131 Padova, Italy

[2] Department of Physics, University of Milan, Via Celoria 16, 20133 Milano, Italy

[3] Istituto Italiano di Tecnologia, Via Morego 30, 16163, Genova, Italy

[4] CNR Institute of Nanoscience, 41125 Modena, Italy

* Emails: agostino.migliore@unipd.it, Maria.BlancoFormoso@iit.it, stefano.corni@unipd.it

Author ORCIDs – Mirko Vanzan: 0000-0003-3521-8045; Agostino Migliore: 0000-0001-7780-2296; Francesco De Angelis: 0000-0001-6053-2488; Stefano Corni: 0000-0001-6707-108X

S1. Summary of the computational analysis

Table S1. Summary of the simulations performed.

nanopore	counterion	I(M)	E_z(V/nm)	V_t(mV)	Δ_c(nm)	N_w	t_0(ns)
no	Na^+	NS	0.25	0		2093	50
		NS	0.5	0	3.997	2093	50
		NS	0.75	0		2093	50
		NS	1.0	0	3.981	2093	50
		NS [a]	0.5	0	3.859	1107	50
		0.5	0.5	0	3.997	2055	50
		NS [b]	0.5	0		6903	150
no	Li^+	NS	0.5	0	4.005	2093	200
		0.5	0.5	0	3.991	2055	100
no	K^+	NS	0.5	0	4.004	2093	150
		0.5	0.5	0	4.017	2055	50
yes	Na^+	NS	0.5	0		4016	150
		NS	0.5	160		4016	50
		NS	0.5	180		4016	150
		NS	0.5	200		4016	50
		NS	0.5	210		4016	50
		NS	0.5	215		4016	50
		NS	0.5	220		4016	150
		NS	0.5	230		4016	310
		NS	0.5	240		4016	150
		NS	0.5	250		4016	100
		NS	0.5	260		4016	100
		0.3	0.5	0		3978	300
yes	Li^+	NS	NS	0		4016	100
		0.3	0.3	0		3978	50

E_z = intensity of the driving electrostatic field, oriented along the direction of the nanopore where the latter is present; Δ_c= MD cell size in the final snapshot of the given simulation; N_w = number of water molecules in the simulation; t_0 = time after which the dynamics of 10GLU reaches a plateau according to the structural descriptors and the statistical analysis is thus performed; v = velocity of the center of mass of the polypeptide in the z direction. [a] The solution contains urea with a 6 M concentration, which amounts to adding 308 molecules of urea. [b] 20GLU is used.

S2. Correlation between structural descriptors

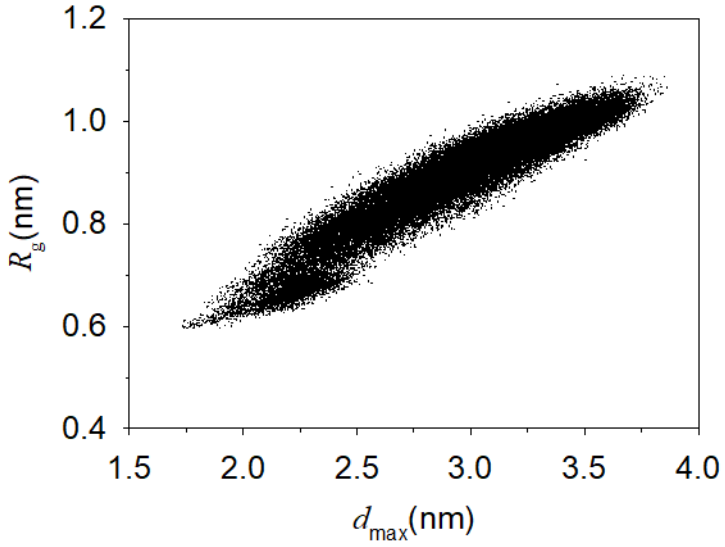

Fig. S1 Approximately linear correlation between the maximum atomic distance in the 10GLU peptide ($d_{\max}$) and its radius of gyration (R_{g}) under the same simulation conditions as in Fig. 2. This behavior supports that both quantities are approximately proportional to the number of amino acid units, and thereby proportional to each other, as is expected for a rod-like polymer.

S3. Length dependence of the peptide dynamics in solution

As shown by the comparison of Fig. 2 with Fig. S2 below, the dynamics of 20GLU is overall similar to that of 10GLU. The average value of R_{g} scales almost linearly for 10GLU and 20GLU (compare the results of the statistical analysis reported in Figs. 2a and S2a). The R_{g} value scales slightly less than linearly with the number of amino acid residues, which can be ascribed to a non-full elongation of the peptides.

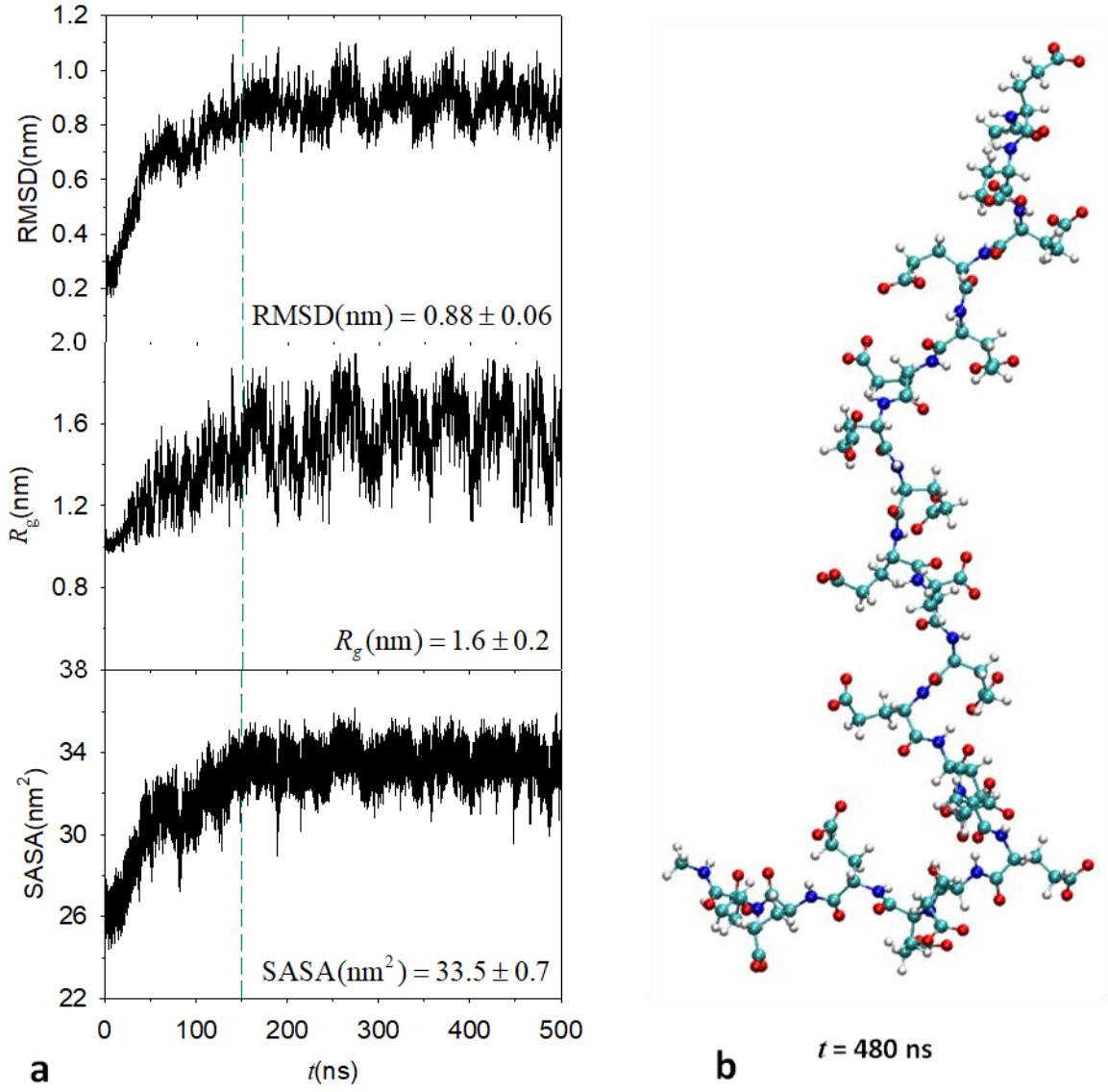

Fig. S2 MD evolution of the 20GLU structure in a solution neutralized by Na^+. (a) RMSD, R_g, and SASA. Average and standard deviation of each parameter after 150 ns are reported on the respective panels. (b) Peptide MD snapshots at the indicated time.

S4. Analysis of peptide-ion interactions

To investigate relevant aspects of polypeptide-ion interactions, we started by considering the radial distribution function, $g(r)$, of the counterions with respect to the carboxyl oxygens of the ten GLU

residues. The $g(r)$ itself does not provide information on the dynamical aspects of the ion-peptide interactions, as it only describes the average distribution of counterions around the GLU functional groups. However, combining its knowledge with information on the ion-dependent fluctuations of the peptide structure during the MD simulations also contributes to understanding dynamic aspects of the ion-peptide interaction (see main text).

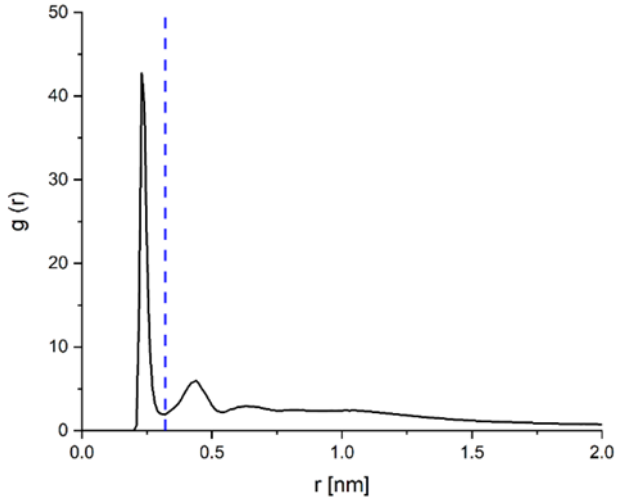

Fig. S3 Radial distribution function $g(r)$ of the Na^+ ions with respect to the oxygens of the carboxylates in 10GLU (averaged over the ten groups). The blue dashed line marks the first local minimum of $g(r)$, used in Equation 1 to delimit the first peak of the function.

Fig. S3 shows the $g(r)$ calculated for the Na^+ ions in the solution surrounding 10GLU, for the first simulation reported in Table 2. The first local minimum of the $g(r)$ approximately delimits its first peak (blue dashed line at 0.32 nm in Fig. S3; a very close value of 3.15 nm was obtained in ref. 1). Integrating the $g(r)$ over this first peak, and using the average density of Na^+ ions in solution in the MD simulation, $\rho = 0.16\,\mathrm{nm}^{-3}$, we obtain the average coordination number

$$n_{Na^+-COO^-} = 4\pi\rho\int_0^{0.32} g(r)r^2dr = 0.16 \qquad \text{(S1)}$$

That is, on average, each carboxylic group is coordinated by 0.16 Na^+ ions, and 1.6 (namely, 1 or 2) COO^- groups of 10GLU engage in significant interactions with these ions during the MD.

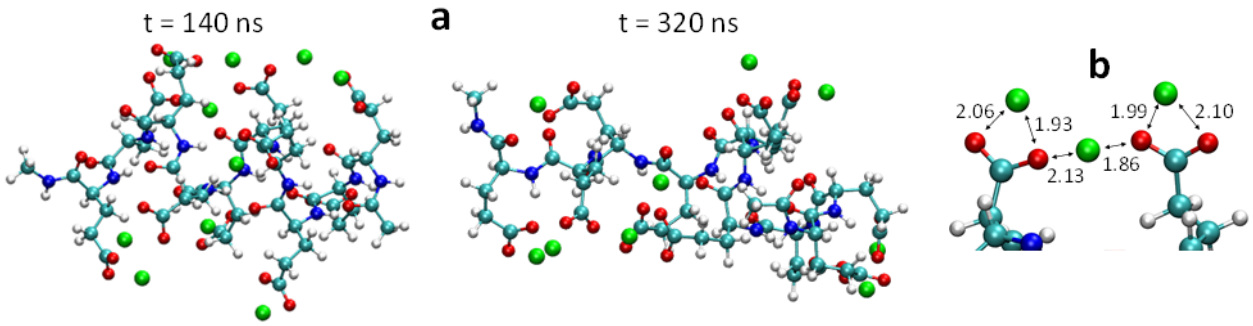

Fig. S4 (a) Two sample MD snapshots of the Li^+-containing system with zero ionic strength (for which $n_{Li^+-COO^-}=1$) at the indicated times. (b) Structural detail showing carboxylates coordinated by more Li ions (distances are in Å). This kind of situation is more frequent in the 0.5 M solution.

The average coordination number depends on the type of counterion and changes with the ion concentration (this concentration was varied adding the chloride salt corresponding to each given counterion), as is shown in Table 1. For example, without adding salt, we obtain $n_{Li^+-COO^-}=1$, that is, on average each COO^- is coordinated by one Li^+ (see Fig. S4a for example). The coordination number increases to $n_{Li^+-COO^-}=1.6$ as the Li^+ ion concentration increases to its value in the 0.5 M solution, which implies that on average each COO^- is coordinated to more than one Li^+ ion, as is shown in Fig. S4b. The stronger interaction of the peptide with the Li^+ ions than with the Na^+ and K^+ ions is consistently characterized by the occurrence of the first peak in the $g(r)$ at 0.28, 0.32, and 0.34 nm, respectively.

S5. Effect of a driving electrostatic field on free peptide in solution

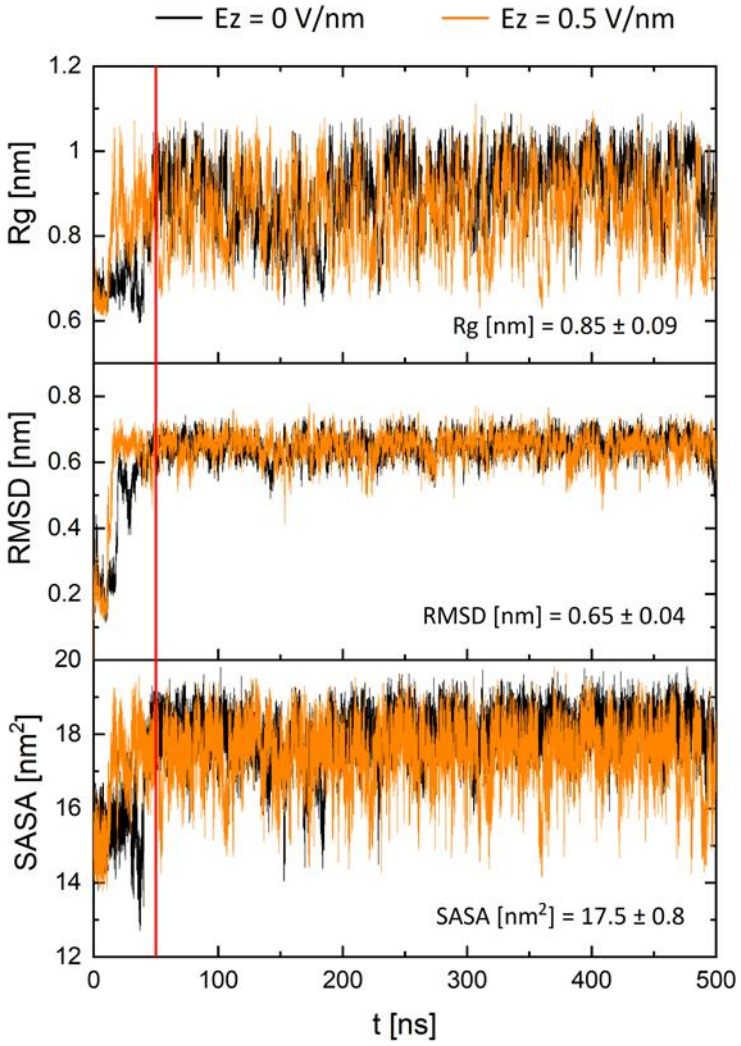

Fig. S5 Time evolution of RMSD, R_g, and SASA in the MD simulation of 10GLU in water with Na^+ counterions, for zero electrostatic field (black) and $E_z = 0.5\,\mathrm{V/nm}$ (orange). Mean values and standard deviations of the structural descriptors along the MD simulations with the external field are reported on the panels. The statistical analysis excluded the first 50 ns. The field causes a translation of the peptide but no significant change in its internal structure dynamics.

S6. Dependence of the peptide velocity on the intensity of the driving field

Fig. S6 shows the velocity of the polypeptide (as measured by the speed of its center of mass) as a function of the intensity of the driving electrostatic field. The numerical values of the data points are reported in Table S2. The small statistical errors of the order of $0.1\ \mathrm{m/s}$ are not used in the figure.

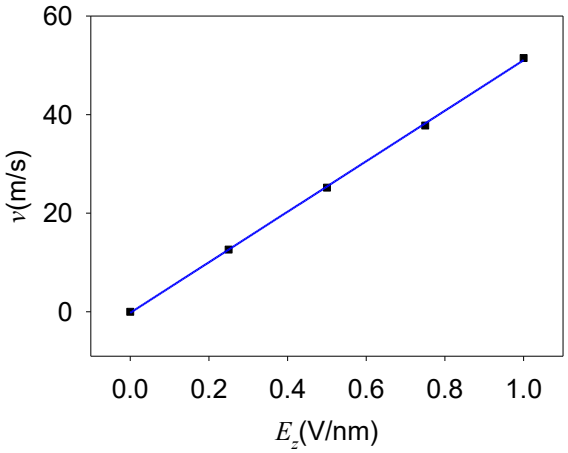

Fig. S6 Velocity of 10GLU in water solvent with Na^+ counterions vs intensity of the driving field (the velocity, obtained by averaging over MD snapshots as detailed in section 2.3, was clearly zero for $E_z = 0$). The regression line is $v = v_0 + bE_z$, where the intercept $v_0 = (-0.2 \pm 0.3)\ \mathrm{ms}^{-1}$ is consistent with the expected value of zero and the slope is $b = (51.3 \pm 0.5) \cdot 10^{-9}\ \mathrm{m^2V^{-1}s^{-1}}$. The linear correlation is significantly supported by a Pearson's correlation coefficient $R = 0.9999$ and a P-value smaller than 0.0001.

Table S2

$E_z(\mathrm{V/nm})$	$v(\mathrm{m/s})$
0.25	12.6 ± 0.1
0.50	25.2 ± 0.2
0.75	37.8 ± 0.1
1.00	51.5 ± 0.3

S7. Effect of a driving electrostatic field on solvated peptide in the nanopore

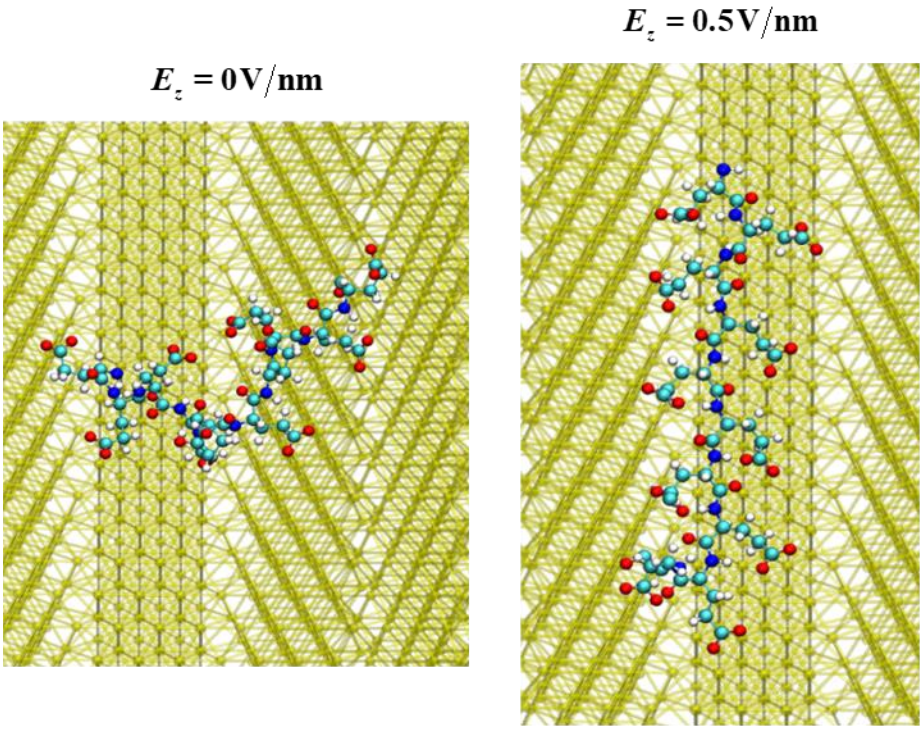

Fig. S7 MD snapshots of 10GLU inside the Au nanopore at 250 ns. Water and counterions are omitted for clarity. The inner surface of the nanopore is shown in pale yellow.

S8. Standard errors of translocation speeds obtained by block averaging

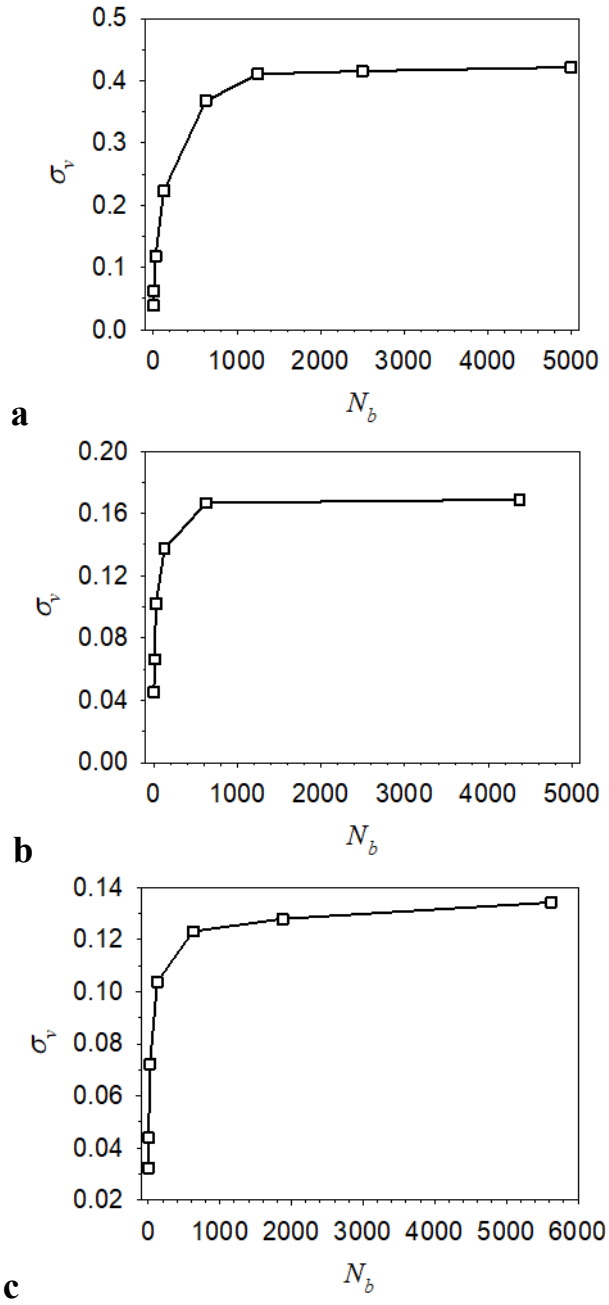

Fig. S8 Standard deviation associated with the translocation velocity, σ_v, as a function of the number of data points collected in each block, N_b, for the peptide in solutions (a) 0.5 M in Li^+, (b) 0.5 M in K^+, and (c) 0.5 M in Na^+, 6 M in urea. The statistical analysis was carried out starting from the times t_0 reported in Table 2. Interpolating lines are shown.

S9. Dynamics of the center of mass for a peptide stuck on the nanopore surface

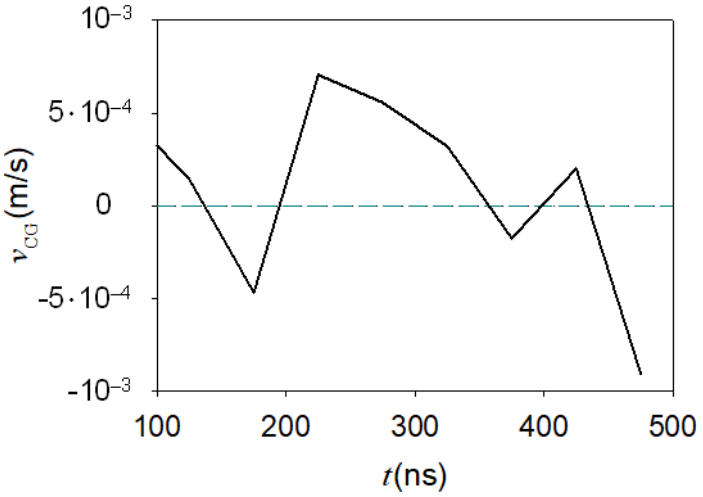

Fig. S9 v_{CG} vs MD time for a later potential difference of 240 mV, obtained by averaging $\boldsymbol{v}$ over 50 ns-wide time intervals. 10GLU remains stuck on the surface of the nanopore after approaching it during the first 100 ns, while its center of mass fluctuates as shown.

S10. Replica analysis

Fig. S10 reports some replicas of the MD simulation with $V_t = 220$ mV, whose analysis is deepened in the main text. The comparison is performed in the worst conditions, that is, before achieving a plateau of the structural descriptors adopted (as reported in Table 3, the time t_0 to reach this plateau can be approximately set to 150 ns). Nonetheless, by removing an increasing amount of initial MD simulation (compatible with the 100 ns length of MD replicas), the average velocity of the peptide approaches a similar range of values in the different simulations.

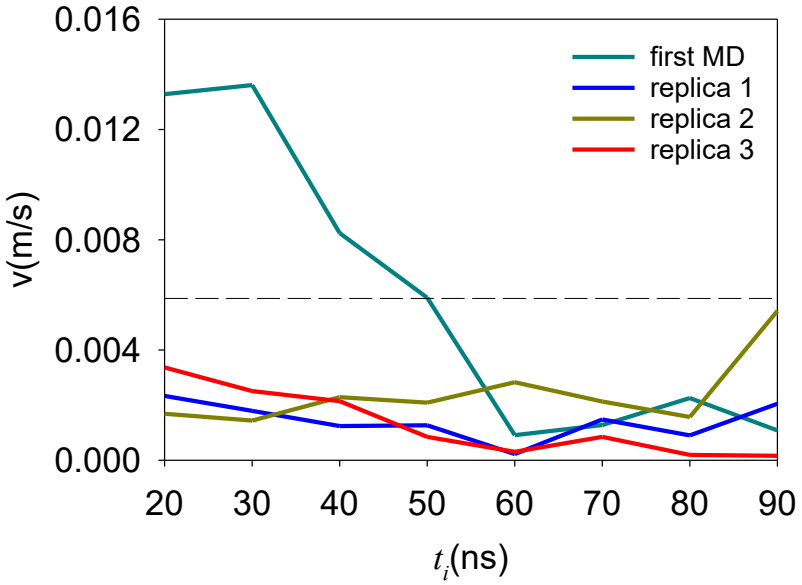

Fig. S10 Mean velocity v of 10GLU over the time intervals $[t_i, 100\ \text{ns}]$ in the main MD simulation and in three MD replicas obtained by changing the random seed for the initialization of the pseudo-random number generator used to produce the initial velocity distribution. The initial coordinates were the same as at the beginning of the main MD production run, and the MD replicas lasted 100 ns.

(1) Dzubiella, J., Molecular Insights into the Ion-Specific Kinetics of Anionic Peptides. *J. Phys. Chem. B* **2010,** *114*, 7098-7103.