THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Multistep Current Signal in Protein Translocation through Graphene Nanopores

Emma Letizia Bonome,[†] Rosalba Lepore,[‡] Domenico Raimondo,[‡] Fabio Cecconi,[§] Anna Tramontano,[‡,∥] and Mauro Chinappi*[,⊥]

[†]Dipartimento di Ingegneria Meccanica e Aerospaziale, Sapienza Università di Roma, Via Eudossiana 18, 00184 Roma, Italy

[‡]Dipartimento di Fisica, Sapienza Università di Roma, 00185 Rome, Italy

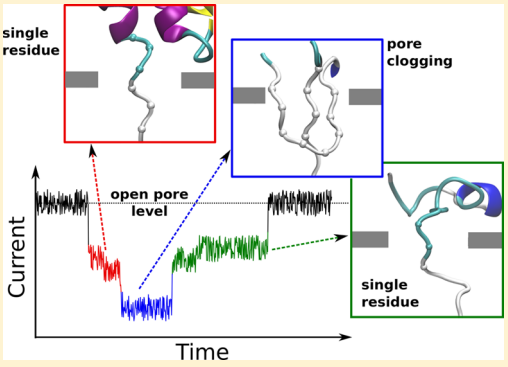
[§]CNR-Istituto dei Sistemi Complessi UoS Sapienza, Via dei Taurini 19, 00185 Roma, Italy

[∥]Istituto Pasteur - Fondazione Cenci Bolognetti, Sapienza Università di Roma, Viale Regina Elena, 291 Rome 00185, Italy

[⊥]Center for Life Nano Science@Sapienza, Istituto Italiano di Tecnologia, Via Regina Elena 291, 00161 Roma, Italy

Supporting Information

ABSTRACT: In nanopore sensing experiments, the properties of molecules are probed by the variation of ionic currents flowing through the nanopore. In this context, the electronic properties and the single-layer thickness of graphene constitute a major advantage for molecule characterization. Here we analyze the translocation pathway of the thioredoxin protein across a graphene nanopore, and the related ionic currents, by integrating two nonequilibrium molecular dynamics methods with a bioinformatic structural analysis. To obtain a qualitative picture of the translocation process and to identify salient features we performed unsupervised structural clustering on translocation conformations. This allowed us to identify some specific and robust translocation intermediates, characterized by significantly different ionic current flows. We found that the ion current strictly anticorrelates with the amount of pore occupancy by thioredoxin residues, providing a putative explanation of the multilevel current scenario observed in recently published translocation experiments.



INTRODUCTION

In the past few years, a number of innovative protocols for biomolecule analysis based on nanopores have been proposed and developed.[2−8] In essence, the working principle of a nanopore based sensing device is very simple. The nanopore connects two chambers containing an electrolyte solution. A voltage applied between the chambers induces the passage of the ions from one chamber to the other resulting in an electric current that depends on the electrolyte solution and on the nanopore features. When a macromolecule translocates through the pore, a change in the current is measured. The intensity and the duration of the current alteration are determined not only by the physicochemical properties of the portion of the molecule inside the nanopore but also by the conformations it assumes at the pore ends.[2−6] The crucial issue in this nanopore sensing approach, also known as resistive pulse sensing, is the capability to correlate the current signal with specific features of the molecule. In the last couple of decades a typical application of interest was DNA analysis[9−13] and, only recently, the technique was applied to proteins and polypeptides. Remarkable examples are the use of nanopores to discriminate between multiple conformations of a peptide induced by specific ligands[2,14] (an issue potentially relevant for addressing protein misfolding associated with neurodegenerative diseases[3]), the analysis of the effect of specific mutations on the unfolding transition[15,16] and, very recently, the detection of post-translational modifications.[4]

A crucial advance in the protein nanopore sensing occurred in 2013 when two independent studies[17,18] provided evidence for the possibility to translocate protein molecules, paving the way to strategies for controlling the translocation speed. In particular, in Nivala et al.[17] the translocation was carried out by the action of the AAA+ unfoldase ClpX protein demonstrating that molecular motors can reproducibly drive proteins through nanopores at controlled speed. Rodriguez-Larrea and Bayley[18] showed how tagging the protein substrates with oligonucleotides enables voltage-driven unidirectional movement through a nanopore. In both experiments, the translocation is accompanied by a specific sequence of different current levels, suggesting the presence of an underlying multistep dynamics. Multistep translocation was also previously observed in coarse grained simulations.[19−23] However, the employment of such reduced models did not provide access to the current levels.

Most of these studies were performed using the α-hemolysin as a biological nanopore, although, recently, a number of other biological pores have been proposed[24−29] Beside biological

Received: March 5, 2015
Revised: April 5, 2015
Published: April 13, 2015

DOI: 10.1021/acs.jpcb.5b02172
J. Phys. Chem. B 2015, 119, 5815−5823

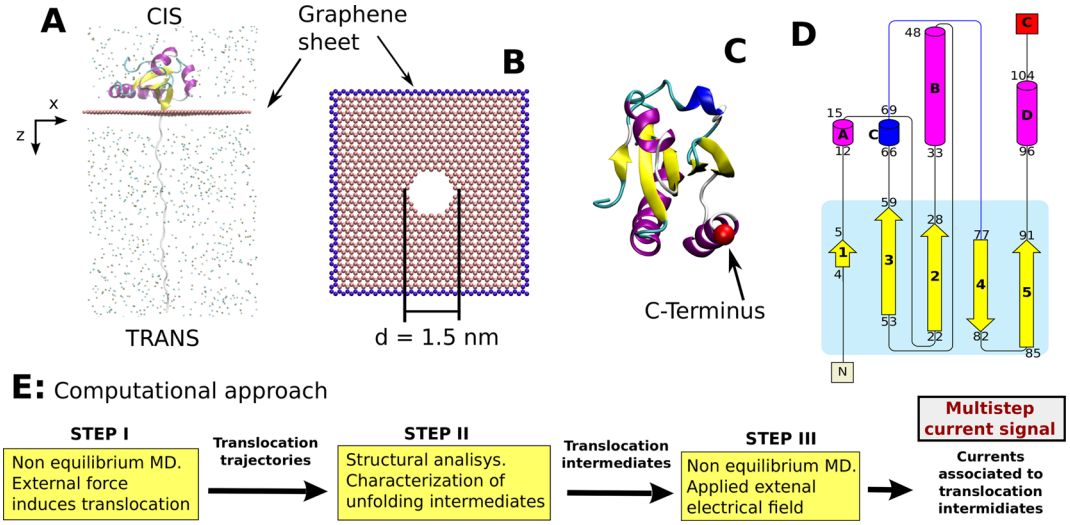


Figure 1. System overview. (A) Translocation of thioredoxin (Trx) from cis to trans compartment through the graphene nanopore. (B) The graphene sheet lies on the Oxy plane (d = diameter 1.5 nm). (C) Ribbon representation of Trx three-dimensional structure. The protein is colored according to its secondary structure (helices in magenta, β-strands in yellow and 3(10) helix in blue) and the C-terminus is highlighted by a red sphere. (D) Trx topological diagram[1] showing the label system used in the text. The color code is the same as in panel (C). For each secondary structure element, the first and the last residue are reported. (E) Flowchart of the computational approach.

nanopores, solid state nanopores are extremely promising as leading technologies for biosensing.[30] In particular graphene appears to be a promising material for nanopore fabrication.[31−33] Recently, it was shown that Si atoms can stabilize nanopore edges avoiding the pore self-healing process often observed for H-passivated pores.[34] The single atom thickness of graphene should result in a higher signal-to-noise ratio, as the current would mainly depend on the translocating residue and, to a lesser extent, on the neighboring ones. The ion conductance of graphene nanopore and their possible application for macromolecule characterization has been investigated through numerical simulation employing different approaches[35−41] However, these studies are limited to nucleic-acids.

In this paper, we employ a computational approach, mainly based on all-atom molecular dynamics (MD) simulations, to analyze the translocation pathway of thioredoxin (the same protein employed in Rodriguez-Larrea and Bayley[18]) through a graphene nanopore. Our computational study clearly shows that the thioredoxin traslocates in a multistep process where it gets stuck in the pore in specific and reproducible conformations, hereafter dubbed translocation intermediates or *stalls*. Different current levels are observed for the different translocation intermediates and strong correlation between electric current and pore occupancy indicates that the steric hindrance is the main responsible for the observed different current levels.

RESULTS AND DISCUSSION

In the present study, we illustrate the results from molecular dynamics simulations of the Thioredoxin (Trx) protein translocation through a graphene nanopore of diameter d = 1.5 nm. Trx is a 108 amino acid protein (PDB accession code: 2TRX[42]) consisting of a central five-strand β-sheet (β1−β5) and three long helices (A, B, D) (Figure 1). Two of these helices connect strands of the β-sheet: helix A, between β1 and β2, and helix B between β2 and β3. Strands β4 and β5 are connected by a short β-bulge loop. Strands β3 and β4 are connected by a long loop containing a 3(10) helical element (helix C). The β-sheet is the core of the molecule and it is surrounded on either side by hydrophobic residues. The graphene sheet lies on the Oxy plane and the protein, pulled from its C-terminus, translocates from the $z > 0$ emispace (cis side) to the $z < 0$ emispace (trans side); see Figure 1A. The overall strategy, summarized in Figure 1E, relies on the combination of three different MD simulation protocols integrated with a structural bioinformatic analysis and it is organized as follows. The first step employs constant velocity (cvSMD) and constant force (cfSMD) simulations to explore the translocation pathway. In the second step, representative conformations of the translocation intermediates are extracted from the MD trajectories using structural clustering. We finally calculate current levels associated with the intermediates via dedicated nonequilibrium MD runs (step III).

Constant Velocity Translocation (cvSMD). The protein is imported into the pore at a constant velocity through a spring that applies a force F to the C-terminus (see Methods). As shown in Figure 2, the force F is plotted against the coordinate z_C of the α-carbon of the terminal residue ($C\alpha_{108}$). The importing force remains quite constant up to z_C = 160 Å. During this stage ($z_C \in (0,160)$ Å), ending with the translocation of residue 57, the pore is occupied only by a single translocating residue and the process occurs as a smooth single-residue passage of consecutive amino-acids; see Figure 2b for a representative snapshot. After that, the increase of F ($z_C \sim 200$ Å) is accompanied by a qualitative change in the translocation dynamics. During this second stage, the translocation is hindered by a partial clogging of the pore not only by the translocating residue, but also by those natively belonging to the strand β2, see Figure 2c for a representative snapshot. The presence of different force peaks indicates a not uniform translocation where a series of structural rearrangements occurs. The largest peak corresponds to the conformation reported in Figure 2d, where three amino acids block the pore and the hairpin, temporarily formed by residues from Ala22 to Trp28 is going to cross the pore. Once the hairpin is translocated ($z_C \sim 300$ Å) no more peaks are present and the single-residue passage already observed in the early

DOI: 10.1021/acs.jpcb.5b02172
J. Phys. Chem. B 2015, 119, 5815−5823

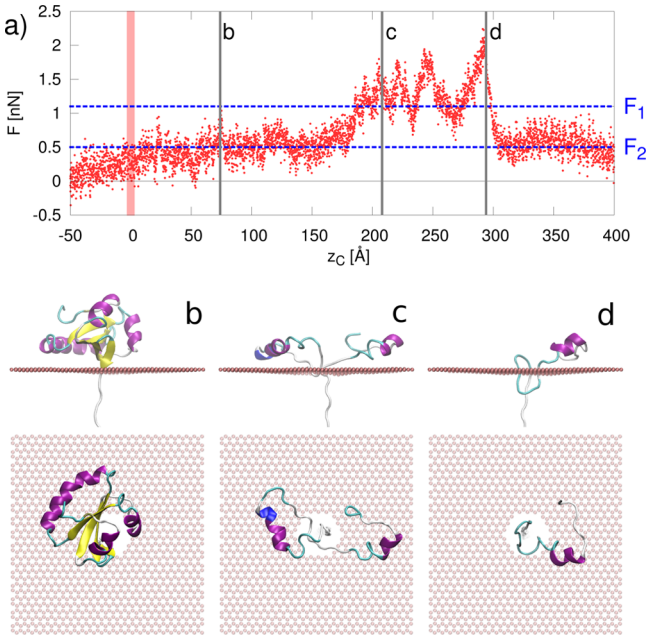


Figure 2. Constant velocity steered molecular dynamics (cvSMD) results. (a) The force F applied to the protein C-terminus is plotted versus the position of the pulled atom z_C. The vertical shaded red line marks the position of the graphene sheet at $z_C = 0$. The horizontal blue dashed lines correspond to the selected values of the force at which we run the cfSMD simulations (F_1 = 1100 pN and F_2 = 550 pN). Protein conformations explored under the force peaks (black vertical lines in (a) are shown in panels (b)−(d) (upper panels, side view; lower panels, top view).

translocation, sets-in again. Accordingly, the resulting force F recovers the same initial intensity.

Constant Force Translocation (cfSMD). To further characterize the translocation intermediates, we performed a set of constant force steered MD simulations (cfSMD) where a constant importing force F_z is applied to Cα_{108}. We selected two forces F_1 and F_2, where F_1 is the intermediate value in the force−extension curve reported in Figure 2 and $F_2 = F_1/2$ is the typical force associated with the single-residue transport regime. Hereafter, we report on the simulations performed at $F_z = F_1$, while the details of the lower-force experiment (F_2) can be found in the Supporting Information.

We performed six independent runs with a time window T_w = 40 ns. In five runs out of six, the protein successfully translocates (Figure 3a,b and Supporting Information Figure S1a, c, and d), while in the remaining case the translocation did not complete in the allotted time window (0, T_w) (Figure S1b). We define the progress variable as the index of the latest translocated amino acid N_p. The time course of N_p is shown in Figure 3a and b for two cases. The presence of multiple stall points is clearly visible. Notably, the stalls are conserved among all the simulations (see Figure S1) and involve values of N_p in the range 30−60, again indicating these events to be associated with the formation of specific and robust translocation intermediates. To support this scenario, we performed an unsupervised structural clustering of the Trx conformations over the six translocations. Protein conformations were clustered based on their global structure similarity using the GDT-TS score (see Methods). The silhouette criterion is adopted to select the number of clusters leading to the best separation. This analysis identified 11 clusters, labeled by roman numerals from I to XI, and sorted according to the N_p value of the representative element of the cluster (the medoid). Since the protein is pulled from the C-terminus, cluster I corresponds to the highest N_p = 105, while cluster XI to the lowest one, N_p = 3. A summary of the clustering results is reported in Table 1. As it can be appreciated in Figure 3a and b,

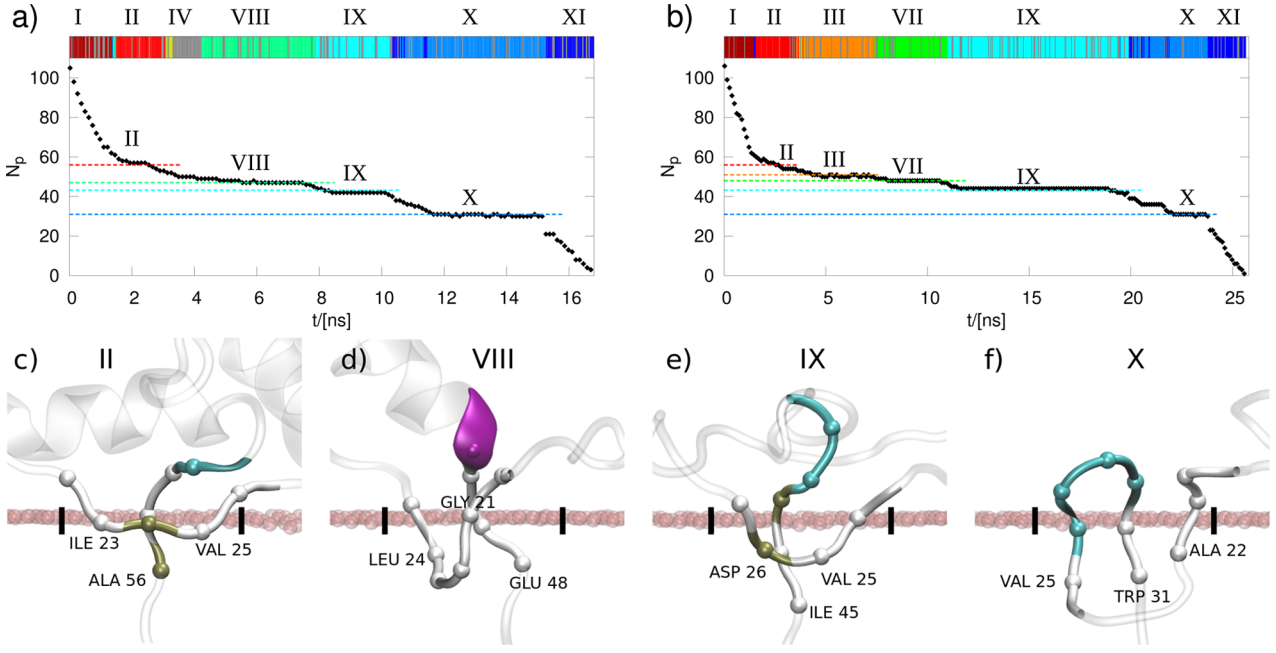


Figure 3. Constant force Steered Molecular Dynamics (cfSMD) results. (a,b) Time evolution of N_p for two cfSMD translocations at F_1. Different translocation stalls are visible and marked by horizontal dashed lines. Upper color bars, corresponding to different protein conformations, are colored according to their cluster membership. (c−f) Traslocation region of Trx for the cluster representatives II, VIII, IX, X. They correspond to major translocation intermediates. Black lines on the graphene sheet indicate the pore size.

DOI: 10.1021/acs.jpcb.5b02172
J. Phys. Chem. B 2015, 119, 5815−5823

Table 1. Cluster Features and Electric Current[a]

ID	N_p	i_p	I [nA]
I	105		8.78
II	56	23/25	7.28
III	51	25/26	5.75
IV	50	22/23	5.41
V	49	21/25	2.23
VI	49	21/26	1.17
VII	48	25/26	6.43
VIII	47	21/24	2.21
IX	44	25/26	8.51
X	31	21/25	11.85
XI	3		9.94
O			15.87

[a]For each of the 11 clusters, we report: the index N_p of the latest translocated amino-acid of the representative conformation of the cluster (the medoid), the indexes i_p of those residues, other than N_p, that occupy the pore, and the electric current I generated by a 5 V applied potential. The last row O refers to the open pore case. The current averages are all significantly different with the exception of few cases, for example, the pair V and VIII; see Figure S7.

the distribution of structural clusters along the translocation curve nicely reflects the sequence of the translocation steps. The only exception is provided by cluster IV, whose members are almost evenly distributed along the curve.

Analysis of the Translocation Pathway. Here we provide a full description of the major events observed during the translocation process. For the sake of clarity, we selected as a reference the simulation reported in Figure 3a.

The first part of the translocation curve shows as Trx rapidly moves away from the native state, forming an elbow which tightly separates the translocation starting point (cluster I, N_p = 105) from the second translocation phase (cluster II, N_p = 56), where a first stall occurs. During the first stage, $t \in (0,1)$ ns, the pore is only occupied by the translocating residue and the C-terminal portion of the protein smoothly translocates (N_p decreases almost linearly, from N_p = 108 to N_p = 56). From a structural point of view, this early stage of the translocation pathway is characterized by the unfolding of the C-terminal α-helix D and the disruption of the h-bond network between the two antiparallel β-strands, β5 and β4. The loss of interactions between the C-terminal β-strands contributes to the destabilization of the core five-stranded β-sheet, having a huge effect on the whole protein fold. Indeed, when the first stall occurs (N_p = 56, cluster II), the protein has already lost its native shape. The translocation of Ala56 causes its parent strand β3 to lose the hydrogen bonding pattern with β2 (residue range Ala22-Trp28), which becomes unstructured as well. The only intact secondary structure elements are the helices A and B (residue ranges Phe12-Asp15 and Gly33-Glu48 respectively). These are connected by a long unstructured region, from Val16 to Cys32, which originally constituted β2 and the coil region linking β2 with helix A. This region adopts a hairpin-like structure which lies on the entrance of the pore.

These early events constitute the major transition, in terms of secondary structure content and native contacts, during the whole translocation process, showing a conserved trend among the different simulations. The same behavior is also observed in the cvSMD (see Figure 2). The hairpin formation is followed by a slow down of the translocation process. Here, as it can be seen in Figure 3a, the translocation curve is characterized by a long plateau, where the untranslocated protein region undergoes local structural rearrangements leading to consecutive translocation bottlenecks. These are associated with distinct structural clusters (in the case at hand, cluster II N_p = 56, cluster VIII N_p = 47, cluster IX N_p = 44, and cluster X N_p = 31). In all these cases, the pore is occupied not only by the translocating residue (i.e., Ala56, Asp47, Asp44, Trp31) but also by residues in the range Gly21-Asp26, see Table 1 column i_p, and Figure 3 panels c, d, e and f. During the first stall (N_p = 56, Cluster *II*), these residues begin to form the hairpin structure and partially cross the pore. The next stall (cluster *VIII*, N_p = 47) is reached after a major structural rearrangement on the pore Cis side. In particular, a reorientation between the two α-helices A and B and the local unfolding of the C-terminal segment of helix B, hosting the translocation residue Asp47, are observed (Figure S2). The translocation of helix B drags the hairpin into the pore in a kinked conformation, occluding it (see Figure 3d and Figure S2). The hairpin conformation persists until helix B entirely translocates, however the kink is less pronounced while the region corresponding to helix B loses its secondary structure (cluster IX, N_p = 44, and Figure 3e). The last stall observed occurs when helix B is pulled away and the unfolding of the hairpin takes place (cluster X, N_p = 31). This is followed by a jump in the translocation curve, due to the joined translocation of residues from Trp31 to Val25.

The last step of the process is characterized by the single-residue translocation of the N-terminal region of the protein. Here, as observed in the first stage, the translocation curve rapidly drops. A similar scenario is observed for the other five simulations. In all cases, the translocation bottlenecks are associated with the rearrangement of the untranslocated protein region, originally constituting the strand β3, on the Cis side and close to the pore entrance, with a partial or full clogging of the pore itself.

The same analysis has been carried out for the three independent simulations at lower force, $F_z = F_2$ = 550 pN. All the details concerning these simulations, and the clusters found, are reported in the Supporting Information. Despite that none of the runs result in complete protein translocation in the defined time window $[0, T_w]$, also in these cases the translocation curves are characterized by a faster single-residue translocation phase followed by a long plateau where different structural intermediates are formed. In particular, we found 5 clusters, the medoids of which have N_p = 105, 59, 54, 49, and 45 (see Table S2). These values are very similar to the ones obtained at F_1. Moreover, in two of them (III_2 and IV_2), the pore is occupied not only by the translocating residue, but also by residues originally belonging to β2, an occurrence already discussed for F_1.

We finally computed the pairwise structure similarity among medoids from F_1 and F_2 simulations. The result is reported in Figure 4. The structure similarity between medoids III_2, IV_2, and V_2 (case F_2 with medoids in the range from II to IX (case F_1) is a further indication that our strategy is effectively able to identify translocation intermediates and that these are reproducible even through different simulation protocols.

Current Measurement. The structures of the 11 representative translocation intermediates (cluster medoids) are taken as initial configurations for non equilibrium simulations where an electric field perpendicular to the graphene layer is applied, corresponding to a voltage ΔV = 5 V. We selected ΔV = 5 *V*, representing a trade-off between the computational requirement of having a low noise and the need

DOI: 10.1021/acs.jpcb.5b02172
J. Phys. Chem. B 2015, 119, 5815−5823

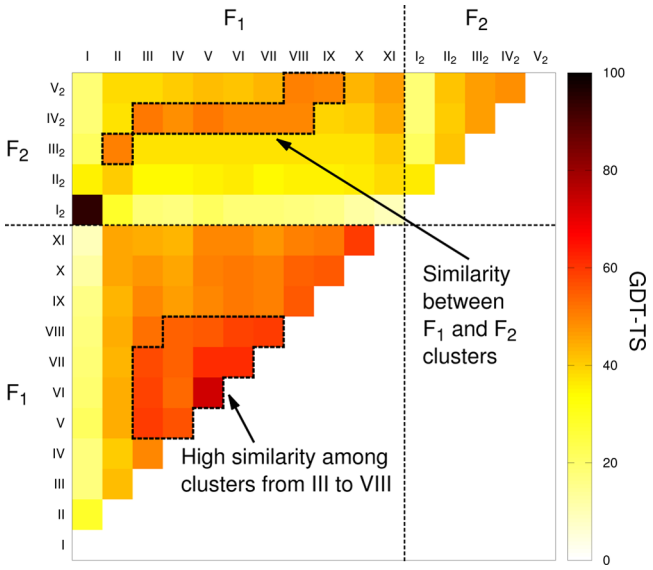


Figure 4. Similarity among cluster medoids of F_1 and F_2. The color scale represents the GDT-TS ranging from 0 to 100 (the higher the values, the higher the similarity among the structures).

of working in the ohmic regime (see Figure S6). The mean electrostatic potential in the system is reported in Figure S8. The potential maps show that most of the potential drop is localized across the membrane (in agreement with ref 37) and that the protein does substantially alter the potential across the membrane.

The induced ionic current I is measured for each independent run. The average current values are reported in Table 1. The data clearly show that (i) the protein translocation results in a strong decrease of I with respect to the open pore condition, and, more importantly, (ii) the different translocation intermediates are associated with distinct current levels. Figure 5 illustrates, using as reference case the already discussed traslocation of Figure 3a, the correspondence between the stalls and the average current levels associated

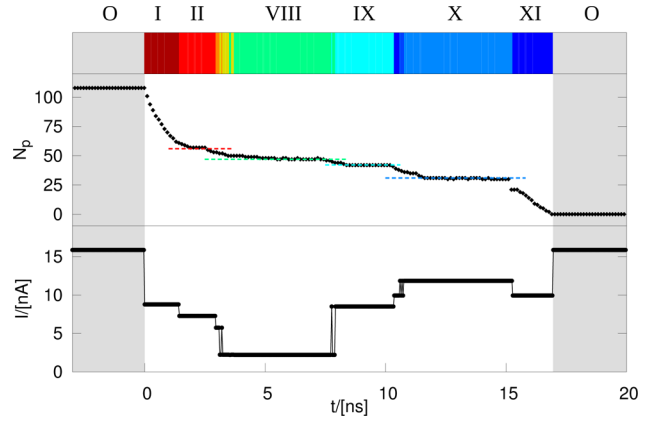


Figure 5. Translocation pathway and current levels. The top panel shows the same plot as in Figure 3a. The bottom panel reports the average current values I measured in the applied-voltage simulations performed on the representative translocation intermediates. The length of each current plateaux is arbitrarily taken equal to the duration of the corresponding stall. For the sake of clarity, we reassigned conformations not belonging to most represented clusters (I, II, VIII, IX, X, XI) to the neighbor ones.

with the translocation intermediates. In the first steps of the translocation, the pore is only occupied by the translocating residue and the currents are in the range 6−9 nA (clusters I and II). In a second stage, the hairpin Ala22-Trp28 clogs the pore and a strong decrease in the current is observed (∼2 nA, cluster VIII). Finally, as soon as the hairpin leaves the pore, the single-residue traslocation regimes is recovered. Moreover, the current (∼10 nA) is slightly larger that the values at the beginning of the translocation (cluster I and II) as there are no partially folded regions able to hinder the ion flux.

The phenomenological picture described above suggests that the different current levels have to be ascribed to the steric occlusion of the ion permeation pathway through the nanopore. To further investigate this hypothesis, we defined the pore occupancy as the average number of non-hydrogen protein atoms within 0.7 nm from the center of the pore. We measured the average pore occupancy for each independent run, and found that it strikingly anticorrelates with the average current values (Pearson's product-moment correlation $r = -0.98$, p-value $= 3.14 \times 10^{-13}$; see Figure 6). Notably, the same

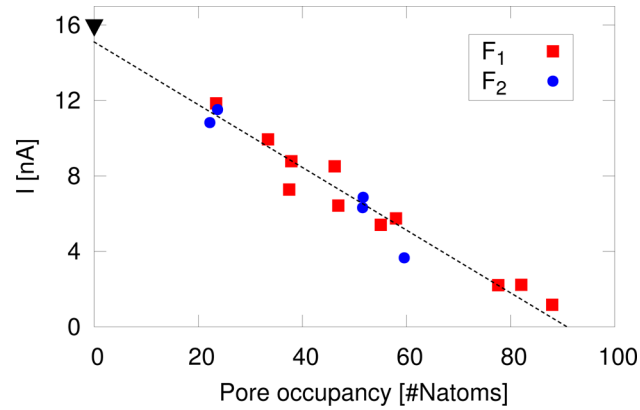


Figure 6. Correlation between current and pore occupancy. Average current I is reported as a function of the average pore occupancy n for F_1 (red squares) and F_2 (blue circles) medoids. The black triangle at zero pore occupancy marks the open-pore value (O in Table 1). The dashed line corresponds to a linear fit ($I = -0.16n + 15.01$).

relationship also holds for the F_2 case. These results indicate that the main cause of the observed current drops ΔI is the steric hindrance, and are consistent with recent experimental findings on translocation through α-hemolysin channels.[43]

It is worth to mention that the observed currents do not trivially anticorrelate with the number of amino acids clogging the pore in the medoid conformation (third column of Table 1). Indeed, cluster X, despite the massive occupation of the pore by the Ala22-Trp28 loop, shows a large current, similar to the one measured in single-residue passage medoids I and XI. A closer inspection of the actual MD trajectory shows that the Ala22-Trp28 loop soon unravels and then translocates recovering the single-residue pore occupation with the corresponding high current level.

Comparison with Experiments. The picture emerging by our simulations can be qualitatively compared with recent experimental results of the Bayley group.[4,18] These authors studied the translocation of thioredoxyn through an α-hemolysyn (αHL) pore. αHL is a mushroom-shaped pore of length ∼100 Å. Its interior is constituted of two main regions: the vestibule and the β-barrel. The vestibule has a maximum diameter of ∼45 Å while the β-barrel, embedded in the lipid

DOI: 10.1021/acs.jpcb.5b02172
J. Phys. Chem. B 2015, 119, 5815−5823

membrane, has a cylindrical shape with a diameter of ∼20 Å. The two regions are connected by a constriction that constitute the pore narrowest section[44] the radius of which is ≃15 Å, that is, similar to our graphene pore.[44] The shape, the ion permeability and the conductance of α-HL[45,46] where quite different from a single layer pore, such as graphene, nevertheless we expect that some of the configurations observed in the constriction, can be similar to the one observed in our graphene pore.

In refs 4 and 18, the transport was driven by an oligonucleotide covalently bonded to the C-terminus of Trx. They observed three distinct current levels. The first one (step 2 in refs 4, 18, and 47) is associated with the oligonucleotide passage and, consequently, no comparison is possible with our simulations. The second level (step 3 in refs 4, 18, and 47) was shown by the authors to correspond to a folded portion of the protein blocked at the pore mouth, while the C-terminal segment occupies the pore, presumably, in a single file conformation. Interestingly, the authors provided convincing evidence to assert that this first stall is associated with the presence at the pore mouth of the loop containing Ile60. Our results support this interpretation. Indeed, in the simulations, the first stalls occur always at $N_p \in (55-60)$, that is, few residues before the translocation of β3. In addition, our simulations show that, after the passage of β3, helices A and B are the only conserved secondary structure elements, and hence, the protein is mostly unfolded. The translocation is ruled, no more by native structure topology, but by the relative position of unstructured loops and, in particular, by the hindering due to the Ala22-Trp28 segment. This is in agreement with experimental findings showing how the last observed current level (step 4 in refs 4, 18, and 47), associated with the passage of the last 40−50 N-terminal residues, is not altered by addition of denaturating agents since it involves already unfolded elements. This step is characterized by a very low current.

Our results allow to speculate on the molecular origin of this strong current drop. The section of our graphene pore is of the same size as the narrowest section of αHL. We measured very low currents when the pore is occupied not only by the translocation residues but also by a portion of the hairpin structure formed by residues natively belonging to β2 (Ala22-Trp28). A putative explanation of the low current observed in refs 4, 18, and 47 is that, after the passage of the region Val55-Ile60, the Trx core completely unfolds and the molecule penetrates into the αHL vestibule. At this point, we argue that what happens in our simulations is similar to what happens in the αHL narrower section. More specifically, the region Ala22-Trp28 partially engages the narrowest section of αHL that, hence, is occupied by more than one residue (as for clusters V, VI, and VIII). This high pore occupancy is associated with a strong decrease in the current that lasts until the complete passage of Ala22-Trp28 region (transition from cluster X to XI in our case). Once this region has crossed the narrowest αHL section, the translocation rapidly proceeds until the channel becomes completely free.

CONCLUSION

This work reports a computational study on the translocation pathway of thioredoxin across a graphene nanopore and the associated ion currents in a voltage driven condition. Recent experiments[18,47] have revealed that a multilevel current signal associated with the thioredoxin transport in a biological nanopore can be interpreted as a defined sequence of steps in the translocation pathway. Our results clearly show that even the traslocation of thioredoxin across a graphene nanopore is a multistep process, where the protein gets stuck in specific conformations, a feature that is robust and reproducible.

We characterized the thierodoxin multistep dynamics in terms of translocation intermediates. The analysis can be sketched as a pipeline where in the first stage, nonequilibrium MD simulations are employed to explore the complete phenomenology of the translocation. The second stage amounts to a structural clustering applied to the sampled conformations in order to identify the common structural features to be associated with the stalling. This analysis produces 11 clusters for the high forcing case and 5 clusters for the low forcing case. Finally the representative intermediate of each cluster is taken as an initial condition for dedicated nonequilibrium simulation where an applied voltage induces a ionic current flowing through the pore.

We found that the amount of current blockage is strikingly proportional to the steric hindrance of the chain portion engaging the pore Figure 6. Moreover the stalls and the related current drops originate from the pore clogging by the residue belonging to native β2-strand (residue 22−28) the unfolding of which is triggered by the passage of strand β3 (residue 53−59). This finding remarkably matches the interpretation of the multistep current pattern of Rodriguez-Larrea and Bayley[47] for the thioredoxin translocation across an α-hemolysin channel. The authors indeed, suggest that one of the intermediates occurs soon after the passage of the first ∼50 C-terminal residues, corresponding to the beginning of the β3 translocation. In our system the major current drops are associated with conformations where the pore is occupied not only by the translocating residue but also by a small hairpin-like motif from the β2 belonging residues. Moderate current drops are instead measured when the thioredoxin migrates in the single-residue transport regime.

Our study shows the potentiality of graphene pores as sensitive tools to determine the volume of the molecule portion engaging the pore. This suggests possible applications to the analysis of protein sequences as well as to the single-residue detection of post translational modifications characterized by alteration of the residue volume.

METHODS

All the simulation (∼0.5 μs) were been performed using NAMD,[48] while VMD[49] was employed for system preparation and analysis.

Graphene Pore Simulation, System Preparation. A graphene sheet of 6 × 6 cells (1500 atoms) was created using the *nanotube builder* utilities of the VMD software.[49] The sheet lies in the Oxy plane and its dimensions are L_x = 61.3 Å and L_y = 63.8 Å. The nanopore is obtained by deleting all the carbon atoms in a circle of radius r_p = 7 Å centered in the middle point of the graphene sheet, corresponding to 54 removed atoms. The psf file for the graphene sheet topology is obtained using the topotools package while the parameters for carbon atoms of graphene were those of type CA in the CHARMM force field.[50] The thioredoxin structure was taken from PDB file 2TRX.[42] In particular, chain A was translated and rotated in order to have the Cα of the C-terminal residue (alanine 108) in the position $\mathbf{r}_{\alpha_{108}} = (x_c, y_c, 50)$ Å, where x_c and y_c are the center of the pore. Missing atoms were added with the VMD *psfgen* package. In

DOI: 10.1021/acs.jpcb.5b02172
J. Phys. Chem. B 2015, 119, 5815−5823

order to allow a single file translocation, the two cysteines, CYS32 and CYS35, are modeled as free cysteines, that is, no disulfide bond (note that in the[18] experiment the two cysteines have been mutated to serines). Water molecules and K^+ and Cl^- ions corresponding to a 2 M salt concentration were added with *solvate* and *ionize* VMD packages, respectively. The resulting system consists of 172 076 atoms. The system is equilibrated with an NPT protocols (T = 310 K, P = 1 atm, during which the Cα of the C-terminus (Cα_{108}) and the two external layers of graphene sheet (dark atoms in Figure 1B) were kept fixed. During the equilibration (5 ns is total), only the z box dimension L_z is allowed to vary, reaching a stationary value of L_z = 426.8 Å, enough to accommodate the elongated protein.

Graphene Pore Translocation. In constant velocity steered molecular dynamics simulation (cvSMD), the Cα_{108} is pulled through the pore by an external force the potential of which reads $U = 1/2k(\mathbf{r}_{C\alpha_{108}} - \mathbf{r}^0_{C\alpha_{108}} - v\,\hat{\mathbf{n}})^2$ with k = 7 kcal mol^{-1}/Å^2, $\hat{n}$ = (0, 0,−1), v = 0.025 Å/ps, and $\mathbf{r}^0_{C\alpha_{108}}$ being the position of Cα_{108} at the end of the equilibration. With the given v, an 18 ns simulation is sufficient for a complete translocation. During the cvSMD run, a configuration with the translocation protein at the pore mouth ($z_{C\alpha_{108}}$ = −0.4 Å) was selected to be used as initial configuration of an equilibrium NVT run where Cα_{108} was kept fixed. Random configurations were extracted from this simulation to be used as independent initial states for constant forcing runs (cfSMD), where a force $F = (0, 0, -F_z)$ is applied to Cα_{108}. The cfSMD simulations were stopped when the protein completely translocates or after 40 ns.

Identification of Representative Translocation Intermediates. Structural superpositions were performed using the LGA package with default parameters (−3 −sda −d:0.5).[51] As a measurement of the distance between two configurations, we employed the Global Distance Test.[52] The Global Distance Test (GDT) analysis was performed in order to identify the set of residues that could fit under different values of distance cutoff. The GDT Total Score (GDT-TS) was calculated as follows: GDT-TS$_{i,j}$ = $(P_1 + P_2 + P_4 + P_8)/4$ where S_i and S_j are two structures under comparison and P_d is the percentage of residues within a distance d, d = 1, 2, 4, 8 Å after structure superposition. The GDT-TS value ranges from 0 to 100. The higher the score, the higher the similarity between the structures. GDT-TS has been shown to provide a good estimate of global structure similarity[52] and the main benefits of using the GDT-TS over other metrics, such as the RMSD, is that it is less sensitive to local structural deviations. Pairwise GDT-TS scores have been calculated over the MD trajectory conformations in an all-versus-all comparison, and used to build a dissimilarity matrix $D = (d_{i,j}) = 100 - \text{GDT-TS}_{i,j}$. The computed matrix D has been used as input for an unsupervised partitional based clustering algorithm (PAM, Partitioning Around Medoids[53] implemented in the R package) with a number of clusters ranging from k = 2 to 20. The silhouette analysis was performed to select the k value leading to the best cluster separations.[54]

Current Measurement. A homogeneous and constant electric field $\mathbf{E} = (0, 0, E_z)$ acting along the z direction is applied. This is equivalent to the application of a constant voltage $\Delta V = E_z L_z$.[55] Although the typical voltage applied in translocation experiments is usually lower than 1 V, in order to have a well-defined signal we decided to operate at a higher voltage. We selected ΔV = 5 V, representing a trade-off between the computational requirement of having a low noise and the need of working in the ohmic region (see Figure S6). Currents traces were calculated as block averages over 20 ps over 4 ns simulation. Their averages were computed cutting out the first 1 ns transient. Statistical analysis of the current traces have been compared using the t test (null hypothesis of equal means) with a multiple testing correction for the p-values (Bonferroni correction[56]). Pore occupancy n is calculated as the average number of massive (non-hydrogen) protein atoms in a sphere of radius 7 Å centered in the pore center. Also for this observable, the first 1 ns is excluded from the averaging. Remarkably, despite the large instantaneous fluctuation on the picosecond scale, all the observed differences among the clusters result to be statistically significant with the exception of few pairs (see Figure S7).

■ ASSOCIATED CONTENT

Supporting Information

Additional data for constant force SMD for the case F_1 ($N_p(t)$ plot and medoids conformation), complete data for case F_2, and current−voltage curve for open pore case. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.5b02172.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: mauro.chinappi@iit.it.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research used the resources of the Supercomputing Laboratory at King Abdullah University of Science & Technology (KAUST) in Thuwal, Saudi Arabia and of the CINECA (GRAPUNA project). Funding: KAUST Award No. KUK-I1-012-43 made by King Abdullah University of Science and Technology (KAUST).

■ REFERENCES

(1) Stivala, A.; Wybrow, M.; Wirth, A.; Whisstock, J. C.; Stuckey, P. J. Automatic generation of protein structure cartoons with Pro-origami. *Bioinformatics* **2011**, *27*, 3315−3316.

(2) Cressiot, B.; Oukhaled, A.; Bacri, L.; Pelta, J. Focus on Protein Unfolding Through Nanopores. *BioNanoScience* **2014**, 1−8.

(3) Madampage, C.; Tavassoly, O.; Christensen, C.; Kumari, M.; Lee, J. Nanopore analysis: An emerging technique for studying the folding and misfolding of proteins. *Prion* **2012**, *6*, 110−118.

(4) Rosen, C. B.; Rodriguez-Larrea, D.; Bayley, H. Single-molecule site-specific detection of protein phosphorylation with a nanopore. *Nat. Biotechnol.* **2014**, *32*, 179−181.

(5) Venkatesan, B. M.; Bashir, R. Nanopore sensors for nucleic acid analysis. *Nat. Nanotechnol.* **2011**, *6*, 615−624.

(6) Makra, I.; Gyurcsányi, R. E. Electrochemical sensing with nanopores: A mini review. *Electrochem. Commun.* **2014**, *43*, 55−59.

(7) Zhang, X.; Wang, Y.; Fricke, B. L.; Gu, L.-Q. Programming Nanopore Ion Flow for Encoded Multiplex MicroRNA Detection. *ACS Nano* **2014**, *8*, 3444−3450.

(8) Schreiber, J.; Wescoe, Z. L.; Abu-Shumays, R.; Vivian, J. T.; Baatar, B.; Karplus, K.; Akeson, M. Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 18910−18915.

(9) Guy, A. T.; Piggot, T. J.; Khalid, S. Single-stranded DNA within nanopores: conformational dynamics and implications for sequencing;

DOI: 10.1021/acs.jpcb.5b02172
J. Phys. Chem. B 2015, 119, 5815−5823

a molecular dynamics simulation study. *Biophys. J.* **2012**, *103*, 1028−1036.
(10) Schneider, G. F.; Kowalczyk, S. W.; Calado, V. E.; Pandraud, G.; Zandbergen, H. W.; Vandersypen, L. M.; Dekker, C. DNA translocation through graphene nanopores. *Nano Lett.* **2010**, *10*, 3163−3167.
(11) Schneider, G.; Dekker, C. DNA sequencing with nanopores. *Nat. Biotechnol.* **2012**, *30*, 326−328.
(12) Chinappi, M.; Casciola, C. M.; Cecconi, F.; Marconi, U. M. B.; Melchionna, S. Modulation of current through a nanopore induced by a charged globule: Implications for DNA-docking. *EPL* **2014**, *108*, 46002.
(13) Fyta, M.; Melchionna, S.; Succi, S. Translocation of biomolecules through solid-state nanopores: Theory meets experiments. *J. Polym. Sci., Part B: Polym. Phys.* **2011**, *49*, 985−1011.
(14) Mereuta, L., Roy, M., Asandei, A., Lee, J. K., Park, Y., Andricioaei, I., Luchian, T. Slowing down single-molecule trafficking through a protein nanopore reveals intermediates for peptide translocation *Sci. Rep.* **2014***4*.
(15) Oukhaled, G.; Mathe, J.; Biance, A.; Bacri, L.; Betton, J.; Lairez, D.; Pelta, J.; Auvray, L. Unfolding of proteins and long transient conformations detected by single nanopore recording. *Phys. Rev. Lett.* **2007**, *98*, 158101.
(16) Merstorf, C.; Cressiot, B.; Pastoriza-Gallego, M.; Oukhaled, A.; Betton, J.-M.; Auvray, L.; Pelta, J. Wild type, mutant protein unfolding and phase transition detected by single-nanopore recording. *ACS Chem. Biol.* **2012**, *7*, 652−658.
(17) Nivala, J.; Marks, D. B.; Akeson, M. Unfoldase-mediated protein translocation through an [alpha]-hemolysin nanopore. *Nat. Biotechnol.* **2013**, *31*, 247−250.
(18) Rodriguez-Larrea, D.; Bayley, H. Multistep protein unfolding during nanopore translocation. *Nat. Nanotechnol.* **2013**, *8*, 288−295.
(19) Bacci, M.; Chinappi, M.; Casciola, C.; Cecconi, F. Role of Denaturation in Maltose Binding Protein Translocation Dynamics. *J. Phys. Chem. B* **2012**, *116*, 4255−4262.
(20) Bacci, M.; Chinappi, M.; Casciola, C. M.; Cecconi, F. Protein translocation in narrow pores: Inferring bottlenecks from native structure topology. *Phys. Rev. E* **2013**, *88*, 022712.
(21) Cecconi, F.; Bacci, M.; Chinappi, M. Protein Transport Across Nanopores: A Statistical Mechanical Perspective From Coarse-Grained Modeling and Approaches. *Protein Pept. Lett.* **2014**, *21*, 227−234.
(22) Huang, L.; Kirmizialtin, S.; Makarov, D. Computer simulations of the translocation and unfolding of a protein pulled mechanically through a pore. *J. Chem. Phys.* **2005**, *123*, 124903.
(23) Tian, P.; Andricioaei, I. Repetitive pulling catalyzes co-translocational unfolding of barnase during import through a mithochondrial pore. *J. Mol. Biol.* **2005**, *350*, 1017−1034.
(24) Derrington, I. M.; Butler, T. Z.; Collins, M. D.; Manrao, E.; Pavlenok, M.; Niederweis, M.; Gundlach, J. H. Nanopore DNA sequencing with MspA. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 16060−16065.
(25) Pastoriza-Gallego, M.; Rabah, L.; Gibrat, G.; Thiebot, B.; van der Goot, F. G.; Auvray, L.; Betton, J.-M.; Pelta, J. Dynamics of unfolded protein transport through an aerolysin pore. *J. Am. Chem. Soc.* **2011**, *133*, 2923−2931.
(26) Housden, N. G.; Hopper, J. T.; Lukoyanova, N.; Rodriguez-Larrea, D.; Wojdyla, J. A.; Klein, A.; Kaminska, R.; Bayley, H.; Saibil, H. R.; Robinson, C. V.; et al. Intrinsically disordered protein threads through the bacterial outer-membrane porin OmpF. *Science* **2013**, *340*, 1570−1574.
(27) Mohammad, M.; Iyer, R.; Howard, K.; McPike, M.; Borer, P.; Movileanu, L. Engineering a rigid protein tunnel for biomolecular detection. *J. Am. Chem. Soc.* **2012**, *134*, 9521−9531.
(28) Trick, J. L.; Wallace, E. J.; Bayley, H.; Sansom, M. S. Designing a Hydrophobic Barrier within Biomimetic Nanopores. *ACS Nano* **2014**, *8*, 11268−11279.
(29) Barati Farimani, A.; Min, K.; Aluru, N. R. DNA Base Detection Using a Single-Layer MoS2. *ACS Nano* **2014**, *8*, 7914−7922.
(30) Miles, B. N.; Ivanov, A. P.; Wilson, K. A.; Doğan, F.; Japrung, D.; Edel, J. B. Single molecule sensing with solid-state nanopores: novel materials, methods, and applications. *Chem. Soc. Rev.* **2013**, *42*, 15−28.
(31) Russo, C. J.; Golovchenko, J. Atom-by-atom nucleation and growth of graphene nanopores. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 5953−5957.
(32) OHern, S. C.; Boutilier, M. S.; Idrobo, J.-C.; Song, Y.; Kong, J.; Laoui, T.; Atieh, M.; Karnik, R. Selective Ionic Transport through Tunable Subnanometer Pores in Single-Layer Graphene Membranes. *Nano Lett.* **2014**, *14*, 1234−1241.
(33) Garaj, S.; Liu, S.; Golovchenko, J. A.; Branton, D. Molecule-hugging graphene nanopores. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 12192−12196.
(34) Lee, J.; Yang, Z.; Zhou, W.; Pennycook, S. J.; Pantelides, S. T.; Chisholm, M. F. Stabilization of graphene nanopore. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 7522−7526.
(35) Suk, M. E.; Aluru, N. Ion transport in sub-5-nm graphene nanopores. *J. Chem. Phys.* **2014**, *140*, 084707.
(36) Hu, G.; Mao, M.; Ghosal, S. Ion transport through a graphene nanopore. *Nanotechnology* **2012**, *23*, 395501.
(37) Sathe, C.; Zou, X.; Leburton, J.-P.; Schulten, K. Computational investigation of DNA detection using graphene nanopores. *ACS Nano* **2011**, *5*, 8842−8851.
(38) Wells, D. B.; Belkin, M.; Comer, J.; Aksimentiev, A. Assessing graphene nanopores for sequencing DNA. *Nano Lett.* **2012**, *12*, 4117−4123.
(39) Avdoshenko, S. M.; Nozaki, D.; Gomes da Rocha, C.; Gonzalez, J. W.; Lee, M. H.; Gutierrez, R.; Cuniberti, G. Dynamic and electronic transport properties of DNA translocation through graphene nanopores. *Nano Lett.* **2013**, *13*, 1969−1976.
(40) Lv, W.; Liu, S.; Li, X.; Wu, R. Spatial blockage of ionic current for electrophoretic translocation of DNA through a graphene nanopore. *Electrophoresis* **2014**, *35*, 1144−1151.
(41) Shankla, M., Aksimentiev, A. Conformational transitions and stop-and-go nanopore transport of single-stranded DNA on charged graphene *Nat. Commun.* **2014***5*.
(42) Katti, S. K.; LeMaster, D. M.; Eklund, H. Crystal structure of thioredoxin from *Escherichia coli* at 1.68 Å resolution. *J. Mol. Biol.* **1990**, *212*, 167−184.
(43) Asandei, A.; Chinappi, M.; Lee, J.-k.; Seo, C. H.; Mereuta, L.; Park, Y.; Luchian, T. Placement of oppositely charged aminoacids at a peptide termini determines the voltage-controlled, dynamic braking of peptide passage across nanopores. *Sci. Rep* **2015**, accepted for publication.
(44) Song, L.; Hobaugh, M. R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. E. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **1996**, *274*, 1859−1866.
(45) Aksimentiev, A.; Schulten, K. Imaging α-hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophys. J.* **2005**, *88*, 3745−3761.
(46) Bhattacharya, S.; Muzard, J.; Payet, L.; Mathé, J.; Bockelmann, U.; Aksimentiev, A.; Viasnoff, V. Rectification of the current in α-Hemolysin pore depends on the cation type: the alkali series probed by molecular dynamics simulations and experiments. *J. Phys. Chem. C* **2011**, *115*, 4255−4264.
(47) Rodriguez-Larrea, D.; Bayley, H. Protein co-translocational unfolding depends on the direction of pulling. *Nat. Commun.* **2014**, *5*, 4841.
(48) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(49) Humphrey, W.; Dalke, A.; Schulten, K.; et al. Humphrey, William and Dalke, Andrew and Schulten, Klaus and others VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(50) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. Vanommeslaeghe, Kenno and Hatcher, Elizabeth and Acharya,

DOI: 10.1021/acs.jpcb.5b02172
J. Phys. Chem. B 2015, 119, 5815−5823

Chayan and Kundu, Sibsankar and Zhong, Shijun and Shim, Jihyun and Darian, Eva and Guvench, Olgun and Lopes, P and Vorobyov, Igor and others CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671−690.

(51) Zemla, A. LGA: a method for finding 3D similarities in protein structures. *Nucleic Acids Res.* **2003**, *31*, 3370−3374.

(52) Zemla, A.; Venclovas, Č; Moult, J.; Fidelis, K. Processing and analysis of CASP3 protein structure predictions. *Proteins: Struct., Funct., Bioinf.* **1999**, *37*, 22−29.

(53) Kaufman, L.; Rousseeuw, P. J. *Finding groups in data: an introduction to cluster analysis*; John Wiley & Sons: New York, 2009; Vol. *344*.

(54) Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53−65.

(55) Gumbart, J.; Khalili-Araghi, F.; Sotomayor, M.; Roux, B. Constant electric field simulations of the membrane potential illustrated with simple systems. *Biochim. Biophys. Acta, Biomembr.* **2012**, *1818*, 294−302.

(56) Bland, J. M.; Altman, D. G. Multiple significance tests: The Bonferroni method. *BMJ* **1995**, *310*, 170.

DOI: 10.1021/acs.jpcb.5b02172
J. Phys. Chem. B 2015, 119, 5815−5823