Computational analysis of maltose binding protein translocation

Mauro Chinappi[a]; Fabio Cecconi[b]; Carlo Massimo Casciola[a]

[a] Dipartimento di Ingegneria Meccanica e Aerospaziale, Sapienza Università di Roma, 00184 Roma, Italia [b] CNR, Istituto Sistemi Complessi, 00185 Roma, Italia

First published on: 11 March 2011

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis
Taylor & Francis Group

Computational analysis of maltose binding protein translocation

Mauro Chinappi[a], Fabio Cecconi[b]* and Carlo Massimo Casciola[a]

[a]*Dipartimento di Ingegneria Meccanica e Aerospaziale, Sapienza Università di Roma, Via Eudossiana 18, 00184 Roma, Italia;* [b]*CNR, Istituto Sistemi Complessi, Via dei Taurini 19, 00185 Roma, Italia*

We propose a computational model for the study of maltose binding protein translocation across α-hemolysin nanopores. The phenomenological approach simplifies both the pore and the polypeptide chain; however it retains the basic structural protein-like properties of the maltose binding protein by promoting the correct formation of its native key interactions. By considering different observables characterising the channel blockade and molecule transport, we verified that MD simulations reproduce qualitatively the behaviour observed in a recent experiment. Simulations reveal that blockade events consist of a capture stage, to some extent related to the unfolding kinetics, and a single file translocation process in the channel. A threshold mechanics underlies the process activation with a critical force depending on the protein denaturation state. Finally, our results support the simple interpretation of translocation via first-passage statistics of a driven diffusion process of a single reaction coordinate.

Keywords: protein translocation; mechanical pulling; first-passage time; maltose binding protein; molecular dynamics; Gō-model

1. Introduction

Transport across cell membranes is the basis of many processes occurring in living organisms, such as metabolic reactions, signal and chemical transmission, and information transfer [1–3]. Nowadays, emerging nanotechnologies offer the opportunity to explore transport at the cellular level (translocation) quantitatively. For example, in the last decade, experiments exploited the stability of the α-hemolysin (αHL) pore (an ion channel from *staphylococcus aureus*) and its spontaneous propensity to be integrated into a lipid bilayer [4], to realise voltage-driven translocation of polynucleotides and polypeptides across membranes in highly controlled conditions. Moreover, several research groups have pointed out that the phenomenology of biopolymer transport through a confining environment can even be reproduced by replacing αHL with solid state nanoscale channels, e.g. carbon nanotubes [5,6]. In a typical voltage-driven translocation, a single nanopore is integrated into a phospholipid bilayer separating two chambers containing ion solutions; a small applied voltage ($V \simeq 100\,\mathrm{mV}$) induces ionic currents through the

nanopore which can be measured by standard electrophysiological techniques. The mixing of biopolymers with the solution produces a variation of the ionic current strongly dependent on the chemical and physical properties of the passing biomolecule which temporarily occupies the channel. For this reason single nanopore systems are considered to be efficient devices for the detection and discrimination of molecules in solution, with powerful applications to the sequencing of biological macromolecules [7].

In this paper, we propose a simplified computational model for the translocation of maltose binding protein (MBP) with the primary purpose of relating our model predictions to some experimental results by Oukhaled et al. [8] on voltage-driven MBP translocation. MBP is a single chain protein of 370 residues without disulfide bonds. Its structure was resolved by Spurlino et al. [9] and successive experimental studies indicated a two-state thermal and chemical unfolding, with a denaturation temperature $T \simeq 336\,\mathrm{K}$ [10]. Such a protein is known to undergo an unfolded translocation through the interaction with the transport protein SecB involving the competition between folding of MBP and the binding of SecB to unfolded forms of MBP to keep them in a translocation competent state. Oukhaled et al. [8] studied the blockade events in a voltage-driven translocation of MBP into a αHL nanopore as a function of the concentration of the denaturing agent. The results revealed short and long blockade times of the αHL channel. Short blockades are due to the passage of completely unfolded proteins; their frequency increases as the concentration of the denaturing agent increases, following a sigmoid denaturation curve. Long blockades reveal partially folded conformations and their duration grows as the proteins are more compact. The interesting physical issue in the study of protein translocation is their resistance to unfolding which tends to inhibit the translocation in narrow channels. The structural protein-like properties are expected to influence the translocation mechanism, so that unfolding kinetics couples to transport to yield the overall biological translocation process. In order to address this issue, we propose a coarse grained model of the MBP and αHL pore as well. Such an approximate computational setup constitutes an 'almost obliged alternative' to simulations at atomic resolution when dealing with translocation of relatively large proteins. Current atomistic methods, indeed, are unable to explore time-scales compatible with real translocation events due to the large number of degrees of freedom involved in describing the biopolymer, the nanopore with its portion of lipid membrane, and the solvent as well. In this paper, we model the relevant structural protein-like properties of the MBP by a well known Gō-model scheme on the C_α backbone proposed by Clementi et al. [11]. We perform MD simulations to extract the behaviour of the main observables characterising the transport. In particular we shall see that the distribution of blockade times is amenable to an interpretation in terms of a model involving a preliminary capture stage followed by an elementary first-passage process [12], where random walkers under a constant bias are injected from the *cis* side of the channel and absorbed to the *trans* side. Simulations allow a complete definition of the model by accessing the full set of parameters involved. Although the model provides a reasonable description of the time distributions of translocation stages, it needs an essential refinement to account for the significant effects induced by the protein-like structure of the MPB in interaction with the nanopore. In principle, such effects are taken into account by addressing the effective

free-energy profiles of the translocation pathways. However, since the evaluation of the free-energy landscape in a meaningful reaction coordinate remains a difficult and open issue, we resort to a model employing a predefined shape of the free-energy combined with a driven diffusion Smoluchowski equation discussed for a related problem in [13] and here reported in Appendix A.2. The simplifications of the model do not allow a perfect overlap with experimental conditions, therefore the significance of our simulation results relies more in suggesting the trends of the observables rather than matching experimental data. For example, in our simulations, we found that even folded proteins can always undergo a translocation provided a force threshold is exceeded. This is in contrast to experiment [8], where no folded proteins were found to translocate because the necessary voltage will probably destroy the lipid bilayer. However, our model provides the possibility of a qualitative comparison between translocation of folded and unfolded conformations and, at the same time, it emphasises the influence of some external parameters on the translocation mechanism.

The paper is organised as follows. Section 2 presents the model for the MBP and the channel, and in Section 3, the simulation results are shown and discussed. Conclusions are drawn in the last section.

2. Model and methods

The maltose binding protein (PDB_id: 4MBP) is modelled via the C_α backbone Gō-model (Clementi et al. [11]) defined by the potential energy

$$V = \sum_{i=1}^{N-1} \frac{k_h}{2}(r_{i,i+1} - R_{i,i+1})^2 + \sum_{i=2}^{N-1} \frac{k_\theta}{2}(\theta_i - \Theta_i)^2 + \sum_{i=3}^{N-1} k_\phi^{(1)}[1 - \cos(\phi_i - \Phi_i)]$$
$$+ k_\phi^{(3)}[1 - \cos 3(\phi_i - \Phi_i)] + \sum_{i,j > i+3} V_{nb}(r_{ij}), \tag{1}$$

where r_{ij} is the distance between the i and j C_α, and θ_i, ϕ_i denote the (pseudo) bond angles, and torsion angles, respectively, identified by the three consecutive C_α's, $i-1$, i, $i+1$, and by the two planes C_α's, $i-2$, $i-1$, i, $i+1$. The upper-case symbols are the corresponding quantities in the native conformation. The strengths of the various terms are chosen as in previous studies [11,14], i.e. $k_h = 1000\epsilon/d_0^2$ ($d_0 = 3.8$ Å), $k_\theta = 20\epsilon$, $k_\phi^{(1)} = \epsilon$, and $k_\phi^{(3)} = 0.5\epsilon$. The parameter ϵ sets the energy scale of the model. The last potential term $V_{nb}(r_{ij})$ is the sum of pairs of native and non-native interactions (contacts). Specifically, amino acids i and j are considered to be in contact if $|i - j| > 3$ and their distance in the PDB structure is within a cutoff R_c; an attractive 12-10 LJ interaction is assigned between them. Amino acids not in native contact repel each other with a soft excluded volume force with core $\sigma = 4.5$ Å. Here, we consider different cutoff values R_c to simulate the translocation of MBP under different denaturation conditions as in the work of Oukhaled et al. [8]. In our model, the increase of denaturant would correspond to a reduction of R_c. Indeed, as R_c controls the number of attractive interactions, its decrease means fewer contacts, hence a destabilisation of the native structure. We chose as a reference cutoff $R_c = 7.0$ Å which selects 824 native contacts. Preliminary folding simulations for

Table 1. List of abbreviations used for different protein denaturation states.

Symbol	Description	R_c/Å	T
RS	Reference state	7.0	0.75
LD	Low denaturation	6.5	0.75
HD	High denaturation	0	0.75
HT	High temperature	7.0	1.5

$R_c = 7.0$ Å identify the folding temperature at $T_f = 0.89$, which corresponds to the experimental value $T \simeq 336$ K [10]. This allows the unique energy scale of the model to be set to $\epsilon = 0.75$ kcal mol^{-1} and accordingly the time and force units are $t_u \simeq 9.4$ ps, $f_u \simeq 13$ pN. We chose as a reference temperature $T = 0.75$ (283 K) which for the reference cutoff $R_c = 7.0$ Å, yields MBP conformations with a high degree of nativeness. Therefore, the state defined by the couple of values $R_c = 7.0$ Å and $T = 0.75$ will be indicated throughout this paper as the reference state (RS).

We will consider a low denatured (LD) MBP at $R_c = 6.5$ Å (676 native contacts) and a highly denatured MBP state (HD), obtained by removing from the force field (1) all the native attractions (setting $R_c = 0$) and the angular terms, to study the translocation of fully denatured and unstructured polypeptide chains (random coil conformations). For a further comparison we also simulated the case of high temperature (HT) $T = 1.5$ for $R_c = 7$ Å, where full denaturation is thermally achieved. For the sake of clarity and readability, it is convenient to refer to Table 1, summarising the different cases considered in this paper.

The temperature in the simulations is controlled by an integrator implementing a Langevin thermostat, time step $= 0.005 t_u$, friction $\gamma = 0.25 t_u^{-1}$.

The αHL pore, which the MBP is imported into, is modelled by a potential with cylindrical symmetry around the x-axis (the translocation direction) [13,15]:

$$V_p(x, y, z) = V_0 \psi(y, z)[1 - \tanh(\alpha x(x - L))], \qquad (2)$$

where $V_0 = \epsilon$, $\psi(y, z) = [(y^2 + z^2)/R_p^2]^q$, and L and R_p denote the pore length and radius, respectively. The values $L = 100$ Å and $R_p = 10$ Å are comparable to the size of the αHL pore as reported in experimental studies [4,8,16]. The parameter q tunes the potential (soft wall) stiffness and α modulates the soft step-like profile in the x-direction; the larger is α, the steeper the step. In this work, we considered $q = 1$ and $\alpha = 3$ Å^{-2}. A constant force F inside the channel and parallel to the x-axis mimics the transport system. Thus, the pore–protein interaction, apart from the homogeneous pulling force, is exemplified to a confinement effect within a cylindrical region of volume $\pi L R_p^2$. Since $R_p = 10$ Å is smaller than the native MBP gyration radius, translocation occurs only in unfolded conformations.

The native MBP has been properly rotated and translated such that last residue 370 is placed on the left side (*cis*) of the pore (Figure 1) in the $(-1, 0, 0)$Å position. The system is then pre-equilibrated at the proper temperature. During equilibration residue 370 is constrained with a harmonic spring, and different uncorrelated configurations selected from the equilibrium run are used as initial conditions for
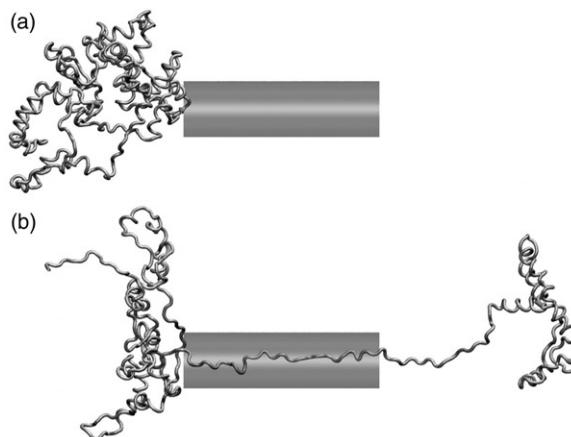
Figure 1. Sketch of the simulation setup. The origin of the reference system coincides with the centre of the left pore mouth and the x-axis is along the channel axis, pointing from left (*cis*) to right (*trans*). MBP is equilibrated on the *cis* side of the pore with the C-terminal constrained to the pore entrance (upper panel). After equilibration, a constant force is applied to the foremost amino acid occupying the pore.

pulling simulations. The MPB is imported into the channel by a constant force applied to the foremost residue which is still in the channel: $x_* \in (-2, L)$. Pulling simulations were run on a finite time window $[0, t_w]$ with $t_w = 10^5 t_u$.

3. Results

Voltage-driven translocation experiments have easy access to *blockade* events, detected through the breakdown of the ion current, indicating that the pore is obstructed. In the specific case of MBP, this could be ascribed both to the passage of the protein through the pore and to the capturing process where the biomolecule is on the *cis* side (left). We performed MD simulations of MBP translocation by mechanical pulling, starting from an ensemble of $M = 400$ partially folded MBP configurations, as in Figure 1. In our implementation, the blockade time corresponds to the first time of a pulling run at which all the residues of the MBP are outside the pore on the *trans* side (i.e. when the x-coordinates of all the residues are such that $x_i > L + 3\,\text{Å}$).

The total blockade time, t_{bl}, which in our numerical implementation is the total duration of a run, could be interpreted as the sum of two different contributions $t_c + t_{tr}$: (a) the capture time, t_c, meant as the time interval the protein spends on the first entrance of the channel without starting a translocation; and (b) the translocation time, t_{tr}, being the time of the net transport. However, it should be remarked that such a distinction is not so sharp and hardly detectable on high force regimes, where the capture stage becomes virtually negligible. During its capture, the MBP inserts some terminal residues into the pore; however, due to thermal fluctuations, they do not constitute a stable nucleus capable of triggering the transport. Figure 2 shows an example of a successful translocation run that can be discussed in terms of the number of residues, $n_{cis}, n_{in}, n_{trans}$, on the *cis*-side, inside, and
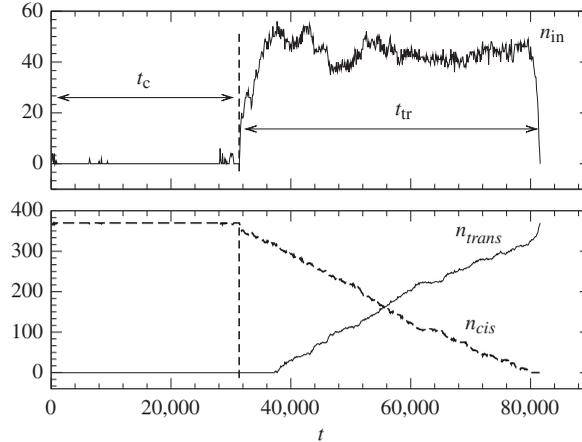
Figure 2. Phenomenology of the channel blockade in terms of MBP residues: $n_{cis}, n_{in}, n_{trans}$ indicate the number of residues, on the left (*cis*) side, inside the pore, and on the right (*trans*) side, respectively. The plot shows the splitting of the blockade time t_{bl} (total time of the run) into capture time t_c and transport (or translocation) time t_{tr}.
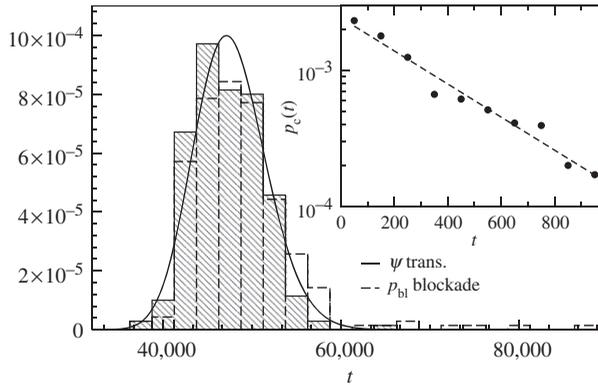
Figure 3. Probability distributions ψ for translocation time (dashed) and blockade p_{bl} time (solid) for $R_c = 6.5$ Å (LD), $F = 0.5 f_u$. Histograms are collected over 280 runs. The continuous line is the fit of $\psi(t)$ via formula (3). The inset shows the distribution of capture time p_c in a semi-log plot, where the exponential decay is apparent.

trans-side, respectively; obviously $n_{cis} + n_{in} + n_{trans} = N = 370$, the total number of MBP residues. In the capture process, the whole molecule lies on the *cis* side, $n_{cis} = 370$, with the exception of a few events where some terminal residues enter and suddenly exit ($n_{in} \neq 0$). At a certain instant, the translocation starts and n_{cis} vanishes almost linearly in time, while n_{trans} increases accordingly. During this stage, at most $n_{in} = 60$ residues constantly occupy the pore. Figure 3 shows the distribution $p_{bl}(t)$ and $\psi(t)$ of blockade and translocation time for $R_c = 6.5$ Å (low denaturation, LD) and a pulling force $F = 0.5$, while the capture time distribution $p_c(t)$ is plotted in the inset. It is apparent that a contribution to the skewness of the blockade time
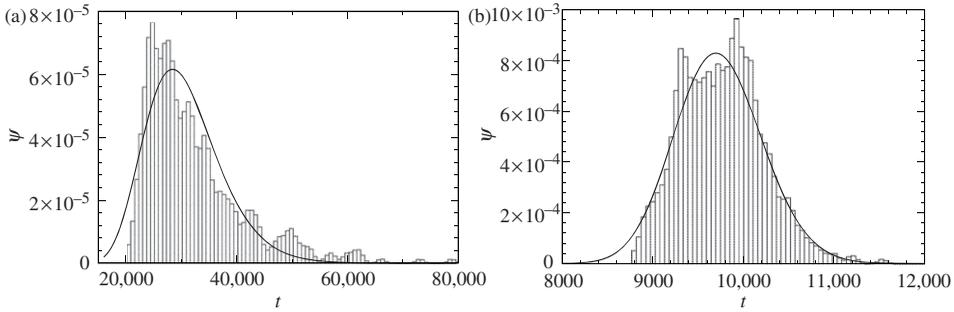
Figure 4. Probability distributions ψ of translocation times for $R_c = 7.0$ Å (RS); $F = 1.0 f_u$ (left) and $F = 2.0 f_u$ (right), respectively. In these cases, capture, when it occurs, is so sudden that translocation events coincide with total channel blockades. Solid lines are fits via formula (3).

distribution is due to the capture process. The distributions display a shape in qualitative agreement with those measured by Oukhaled et al. [8]. A slightly different behaviour is observed for $R_c = 7.0$ Å (reference state, RS), Figure 4, where the capture time is found to be negligible with respect to the transport time. In this case, the skewness of p_{bl} is probably a consequence of the first-passage process character [12] of sharp translocation events. Indeed, previous studies [13,17–20] supported the theoretical interpretation of blockade time distributions as distributions of first arrival time of an appropriate driven diffusion process [13,17–20]. Moreover, as already observed [13], the degree of asymmetry (skewness) decreases for higher force regimes.

A first analysis of the simulation data suggests that the capture time PdF, $p_c(t)$, reduces to an exponential $\lambda_c \exp(-\lambda_c t)$ (see the inset of Figure 3) where λ_c is the capture rate whose dependence on F looks like a simple Arrhenius-like behaviour, as we will discuss later; see also [21] for a theoretical treatment of the phenomenological aspects of the capture process. As usual in this kind of problem, a first crude approximation assumes that the transport time PdF, ψ, could be described by the inverse Gaussian [22]

$$\psi(t) = \frac{L}{\sqrt{4\pi D t^3}} \exp\left\{-\frac{(L - \mu F t)^2}{4Dt}\right\}, \tag{3}$$

which is the first arrival time distribution at $x = L$ of the biased random walk in the domain $[-\infty, L]$, where walkers injected at the origin are absorbed in L. In formula (3), D and μ denote the effective diffusion coefficients and mobility of the protein centre of mass, which, in general, both depend on the pulling field F. In Figures 3 and 4, we used formula (3) to fit the histograms. The only free parameter is D provided that μ is estimated from $\mu = v/F$, v being the translocation velocity

$$v = \frac{L}{M} \sum_{i=1}^{M} \frac{1}{t_{tr}(i)}, \tag{4}$$

where M is the number of pulling runs where the protein successfully translocates in the time window $[0, t_w]$. Another reasonable approximation can be based on the assumption of statistical independence between capture and translocation events implying that p_{bl} would be the convolution

$$p_{bl}(t) = \int_0^t dt' p_c(t') \, \psi(t - t')$$

of the transport ψ and capture p_c distributions. The above argument prescribes the average blockade time

$$\tau_{bl} = \frac{1}{\lambda_c} + \frac{L}{\mu F}, \tag{5}$$

with the first term pertaining to capture and the second to transport, which, for distribution (3), coincides with a trivial ballistic estimate. A direct estimation of the average blockade from simulation is

$$\tau_{bl} = \frac{1}{M} \sum_{i=1}^{M} t_{bl}(i) \tag{6}$$

(M is the same as in Equation (4)). Similar formulas are used to determine the average capture (τ_c) and translocation (τ_{tr}) times. The estimations (4) and (6) are both affected by a systematic error due to the finite time windows considered, and this problem becomes critical at low forcing, where τ_{tr} approaches t_w. The behaviour of τ_{bl} is shown in the left panel of Figure 5 for all the cases of Table 1 as a function of the importing force, and the inset shows the behaviour of τ_c for RS and LD.
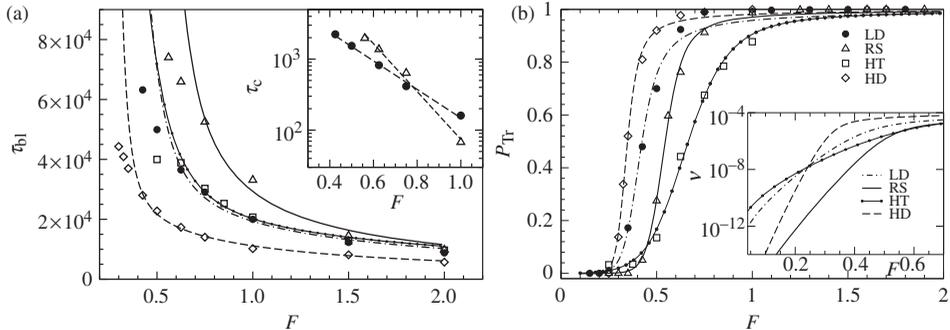
Figure 5. Left: average blockade time τ_{bl} as a function of the importing force F, referred to the four cases considered in Table 1: LD full circles, RS (open triangles), HT (open squares) and HD (open diamonds). Lines are the fits obtained via expression (7). Inset: Arrhenius-like behaviour of average capture time τ_c as a function of F for LD and RS MBP; the fits are $1.54 \cdot 10^5 e^{-4.62F}$ and $1.71 \cdot 10^5 e^{-7.74F}$, respectively. In the other two cases, the capture process is negligible. Right: translocation probability of MBP across the channel as a function of the driving force. The symbols for the four considered cases are the same as for left panel. Simulation data are fit through Equation (22) derived from the driven diffusion theory, see Appendix A.2. The inset reports the range of exponential dependence $\nu \propto \exp(F/F_0)$ of the blockade event frequency.

Capture time is negligible in the two fully denatured states HD and HT, as the open protein conformation highly facilitates the sudden introduction of a critical number of residues (stable nucleus) in the pore which triggers the translocation. Our data suggest that the average capture-time scales exponentially with the forcing, i.e. $\lambda_c \propto \exp(-\beta F l_c)$, where l_c is a constant with the dimension of a length. The phenomenological expression (5) does not reproduce the average blockade times unless we assume an *ad hoc* dependence of both μ and D on F. A more refined approach based on the driven-diffusion description of the translocation process as discussed in Appendix A.2 provides the following formula:

$$\tau_{bl}(F) = \frac{e^{-\beta FL}}{R_L} M_-(F) + \frac{1}{D_0} M_0(F), \tag{7}$$

where R_L is a rate taking into account that the *cis*-boundary is not perfectly absorbing (radiation boundary condition [23]), D_0 is the effective diffusion coefficient and $M_-(F)$ and $M_0(F)$ are specific functions depending on the free energy profile in the reaction coordinate associated with the translocation process. The fit by formula (7) well characterises the behaviour of $\tau_{bl}(F)$ data with the exception of the low force region where the values of τ_{bl} are systematically underestimated due to the finiteness of the simulation time window of $[0, t_w]$. In fact, only a fraction of the runs corresponds to successful translocation events; the remaining ones not contributing to the averages of transport observables are killed as soon as $t > t_w$. This calls for the definition of the *translocation probability*, P_{Tr}, estimated as the number of translocation successes, within time $t_w \simeq 10^5 t_u$, over the total number of runs. Given the finite time window, P_{Tr} depends on t_w and represents the probability for the run not to be yet absorbed in a time $t < t_w$. Figure 5 shows P_{Tr} versus F for all the cases listed in Table 1. The simulation data of P_{Tr} are fit by the formula

$$P_{Tr}(F) = \frac{1}{1 + R_0/R_L e^{-\beta FL} + R_0/D_0 M_+(F)} \tag{8}$$

derived by the driven diffusion approach summarised in Appendix A.2, where R_0 is the counterpart of R_L above defined for the *cis* boundary. We can define the force threshold F_c as the value of the field where $P_{Tr} = 1/2$; below this threshold the probability becomes small. It is clear from the right panel of Figure 5 that for the cases at the same temperature $T = 0.75$ (RS, LD, HD), F_c decreases when R_c runs from 7 Å to 0 Å, the latter corresponding to a random coil. The result is clear when considering that the greater the interaction cutoff R_c, the more stable the structures, hence a greater importing force is required to trigger the denaturation and successive transport. The transition region (the F-range around F_c where P_{Tr} shifts from 0 to 1) is narrow, meaning that small variations of the force near F_c produce strong changes in the translocation activation. On the other hand, comparing the results of RS and HT (both at $R_c = 7$ Å), we observe that high temperature implies a higher F_c, as expected because the force needs to overwhelm a stronger thermal agitation of the chain which hinders the passage of the denatured MBP. Moreover, thermal fluctuations significantly widen the transition region resulting in a less steep P_{Tr} curve. The results have a natural interpretation in terms of free energy barriers.

At constant temperature the free-energy barrier is lower at small R_c, thus a less intense pulling field F is needed for the MBP to translocate, in agreement with the experimental findings as the parameter R_c works as a denaturant concentration. In order to relate our results to the laboratory observations by Oukhaled et al. [8], we can assume the threshold force F_c corresponds to an applied voltage $V_0 = F_c L / Q_{eff}$ which, using an effective charge of $Q_{eff} = 0.6$ electronic charges as suggested in [8], provides the value $V_0 \simeq 1.4$ V. It is important to stress that this very high voltage threshold, more than one order of magnitude greater than the typical experimental value of 100 mV [8] enables the translocation of even initially folded MBP, a regime presumably not accessible to laboratory experiments with biological nanopores. However, a qualitative comparison with experiment could be attempted by considering other results reported by Oukhaled et al. [8]. In that paper, an exponential dependence of the blockade frequency ν on the applied voltage V, $\nu \propto \exp(qV/k_B T)$ is found in the range [50,150] mV. As such a frequency depends both on the possibility for the MBP to translocate and on the translocation rate per molecule, ν could be estimated, in our model, as the ratio

$$\nu(F) \propto \frac{P_{Tr}(F)}{\tau_{bl}(F)}, \tag{9}$$

which is plotted on a semilog scale in the inset of the right panel of Figure 5. Our prediction (9) for the event frequency exhibits a similar exponential regime below F_c in the RS and HD cases although a quantitative comparison is not possible. The behaviour is not so clean in the other two cases (LD, HT) presumably due to poor statistics in the region of very small probabilities which strongly affect the accuracy of the fits. Moreover in the nearly ballistic regime (large F), the simple analysis of Appendix A.2 shows that ν grows linearly with F, as $P_{Tr}(F) \to 1$ and $\tau_{bl}(F) \propto 1/F$. A similar dependence is thus expected in experiments on voltages well above 150 mV, which, though not supported by biomembrane embedded pores, should perhaps be possible for solid state nanopores.

4. Conclusions

We considered and used a coarse grained model for the maltose binding protein to perform molecular dynamics (MD) simulations of its transport across a single α-hemolysin (αHL) pore by mechanical pulling. The protein is described by a Gō-like model [11] frequently used in the literature on protein folding. Although, the model neglects many atomic details, it retains the essential structural protein-like properties of the polypeptide chains which are basic to investigate the coupling of the unfolding with the translocation process. In the same philosophy, also the αHL is described as a simple cylindrical channel where the macromolecule is imported by a constant force. The work has been motivated by a recent experiment on the voltage-driven translocation of MBP through the αHL pore [8], where the blockade events of the channel have been studied as a function of the protein chemical denaturation. Although not fully equivalent, in our model, the role of denaturation is played by a reduced number of native interactions conferring less stability on the MBP native crystallographic structure. Our study focuses on four different conditions (Table 1).

In two of them, we consider the translocation of MBP from its folded or partially folded conformation (RS, LD); in the other two (HD, HT) the translocation involves random coil structures achieved by thermal unfolding of the native state (HT) or by removing attractive and angular interaction from the Gō-like force-field (HD). A whole range of pulling forces, even outside the experimental capabilities, has been explored, to obtain translocation events even of initially folded MPB structures not observed in experiment [8].

We show that a computational approach combined with a driven-diffusion model is able, through simple physical assumptions, to interpret some results from the experiments. Our work supports the theory of a first-passage process as the natural conceptual framework to classify the protein translocation phenomenology. Indeed, the shape of the distributions of the first-passage time in our study agrees with the blockade time distributions in the experiment by Oukhaled et al. [8]. Comparison of the translocation of MBP in different denatured conditions indicates that a threshold mechanism underlies the process activation. The critical forcing F_c, keeping the temperature constant, depends on the MBP denaturation state, such that F_c decreases with the degree of denaturation. The passage of thermally denatured MBP (HT) requires larger thresholds with respect to 'chemically denatured' MBP (HD) due to the higher entropy contribution to the free energy. The exponential dependence of the blockade frequency ν on the voltage V just above the threshold is qualitatively verified by our model; at the same time our analysis predicts a linear scaling for high applied voltages.

In the present study, the effects of the chain on translocation rates are taken into account by assuming a simple free energy translocation profile that looks like a trapezoid function of the reaction coordinate. In this respect, a better comprehension of the process could be achieved by investigating how the details of the free energy profile are related to the non-trivial kinetics of MBP under mechanical pulling, as discussed by Bertz and Rief [24].

Acknowledgements

References

[1] H. Lodish, D. Baltimore, A. Berk, S.L. Zipursky, P. Matsudaira and J. Durnell, *Molecular Cell Biology*, Freeman, New York, 1996.
[2] W. Wickner and R. Schekman, Science 310 (2005) p.247.
[3] G. Schatz and B. Dobberstein, Science 271 (1996) p.1519.
[4] J. Kasianowicz, E. Brandin, D. Branton and D. Deamer, Proc. Natl. Acad. Sci. USA 93 (1996) p.13770.
[5] L. Jiali, M. Gershow, D. Stein, E. Brandin and J. Golovchenko, Nature Mater. 2 (2003) p.611.
[6] C. Dekker, Nat. Nanotech. 2 (2007) p.209.
[7] A. Meller, L. Nivon, E. Brandin, J. Golovchenko and D. Branton, Proc. Natl. Acad. Sci. 97 (2000) p.1079.

[8] G. Oukhaled, J. Mathé, A. Biance, L. Bacri, J. Betton, D. Lairez, J. Pelta and L. Auvray, Phys. Rev. Lett. 98 (2007) p.158101.

[9] J. Spurlino, G. Lu and F. Quiocho, J. Biol. Chem. 266 (1991) p.5202.

[10] C. Ganesh, A. Shah, C. Swaminathan, A. Surolia and R. Varadarajan, Biochem. 36 (1997) p.5020.

[11] C. Clementi, H. Nymeyer and J. Onuchic, J. Mol. Biol. 298 (2000) p.937.

[12] S. Redner, *A Guide to First-Passage Processes*, Cambridge University Press, Cambridge, 2001.

[13] A. Ammenti, F. Cecconi, U. Marini-Bettolo-Marconi and A. Vulpiani, J. Phys. Chem. B 113 (2009) p.10348.

[14] F. Cecconi, C. Guardiani and R. Livi, Biophys. J. 91 (2006) p.694.

[15] L. Huang and D. Makarov, J. Chem. Phys. 129 (2008) p.121107.

[16] A. Meller, L. Nivon and D. Branton, Phys. Rev. Lett. 86 (2001) p.3435.

[17] D. Lubensky and D. Nelson, Biophys. J. 77 (1999) p.1824.

[18] M. Muthukumar, Phys. Rev. Lett. 86 (2001) p.3188.

[19] A. Berezhkovskii, M. Pustovoit and S. Bezrukov, J. Chem. Phys. 116 (2002) p.9952.

[20] A. Berezhkovskii and I. Gopich, Biophys. J. 84 (2003) p.787.

[21] M. Muthukumar, J. Chem. Phys. 132 (2010) p.195101.

[22] V. Seshadri, *The Inverse Gaussian Distribution: A Case Study in Exponential Families*, Oxford University Press, New York, 1993.

[23] A. Szabo, K. Schulten and Z. Schulten, J. Chem. Phys. 72 (1980) p.4350.

[24] M. Bertz and M. Rief, J. Mol. Biol. 378 (2008) p.447.

[25] G. Arfken and H. Weber, *Mathematical Methods for Physicists*, Academic Press, New York, 2001, p.618.

[26] D. Makarov, Acc. Chem. Res. 42 (2009).

[27] S. Luccioli, A. Imparato, S. Mitternacht, A. Irbäck and A. Torcini, Phys. Rev. E 81 (2010) p.10902.

Appendix

This appendix briefly summarises the two theories used to interpret the simulation data on maltose binding protein translocation. The first theory, the simpler one, provides an expression for the blockade time distribution in terms of the capture and translocation times; however, since it limits us to successful translocations only, it does not contain any information about the translocation probability. The second theory is more general, taking into account the role of the pore–protein interaction and the effects of pore boundaries. It provides an analytical expression for the translocation probability and a more appropriate expression of the blockade time.

A.1. *Two stage theory: capture and transport*

The channel blockade can be split into two stages, capture and translocation. Capture involves the approach of the protein to the pore and its transition to a state which is the precursor of translocation. Simulation data suggest modelling the capture as if the molecule were in two complementary states: 'captured' and 'uncaptured', where uncaptured means that the *C*-terminal region of the MBP explores the *cis* side of the channel without, however, triggering successful transport. Indicating by P_c the probability of being captured, its rate equation reads

$$\frac{dP_c}{dt} = -\lambda_0 P_c + \lambda_c(1 - P_c),\tag{10}$$

where λ_c is the transition rate from the *uncaptured* to the *captured* state and λ_0 is the rate of the reverse process. We assume that once the protein is captured it certainly translocates, hence $\lambda_0 \simeq 0$. Thus the solution of Equation (10) with initial condition $P_c(0) = 0$ is $P_c(t) = 1 - \exp(-\lambda_c t)$ whose time derivative yields the distribution of the capture time,

$$p_c = \frac{dP_c}{dt} = \lambda_c e^{-\lambda_c t}, \tag{11}$$

which is employed to fit the data (see the inset of Figure 3). Simulation data (Figure 5 inset, left) seem to indicate that the rate of the *uncaptured* $\rightarrow$ *captured* transition exhibits an Arrhenius-like behaviour, $\lambda_c \propto \exp(-\beta F l_c)$, where l_c is a constant with the dimension of a length.

Once the MBP is captured, its translocation is interpreted in terms of a first-passage process [12] from the *cis* to the *trans* side of the channel. The standard theoretical approach that captures the essential phenomenology of biomolecular translocation under driving fields is based on a driven diffusion Smoluchowski equation [17,19,20]

$$\frac{\partial P}{\partial t} + \frac{\partial J}{\partial x} = 0 \tag{12}$$

for the probability $P = P(x, t)$ to find the MBP with a reaction coordinate x at time t, $J = J(x, t)$ being the current associated with the probability. In our context, a reasonable reaction coordinate is the centre of mass of the molecule [13].

The simplest model amounts to considering the first arrival process at the adsorbing boundary for a random walk with bias $\mu_0 F$, where walkers are injected at a distance L from the boundary. The associated probability flux is

$$J = -D_0 \frac{\partial P}{\partial x} + \mu_0 F P, \tag{13}$$

where D_0, μ_0 are the effective diffusion and mobility constants, respectively. Finally, the initial $P(x, 0) = \delta(x)$ and boundary conditions $P(-\infty, t) = P(L, t) = 0$, select the solution of the problem. The solution of the Smoluchowski problem given by Equations (12) and (13) is a linear combination

$$P(x, t) = g(x - \mu_0 F t) - \exp\left(\frac{L \mu_0 F}{D_0}\right) g(x - \mu_0 F t - 2L) \tag{14}$$

of the Gaussian originating from $x = 0$ and travelling with velocity $v = \mu_0 F$: $g(x - vt)$ with a second Gaussian $g(x - vt - 2L)$ originating from $x = 2L$ where $g(u) = (4\pi D_0 t)^{-1/2} \times \exp(-u^2/(4D_0 t))$. It is easy to verify that the solution (14) automatically satisfies the boundary conditions.

The key quantity to derive the first passage time statistics is the survival probability of a particle in the domain $[-\infty, L]$, which is defined as

$$S(t) = \int_{-\infty}^{L} dx P(x, t). \tag{15}$$

Hence the probability of the particle exiting in the time interval $[0, t]$ is $1 - S(t)$ whose time derivative is the PdF of the residence times in the channel (blockade times)

$$\psi(t) = -\frac{dS}{dt} = -\int_{-\infty}^{L} dx \frac{\partial P(x, t)}{\partial t} = \int_{-\infty}^{L} dx \frac{\partial J(x, t)}{\partial x}, \tag{16}$$

which, recalling the boundary conditions for $x \rightarrow -\infty$, gives

$$\phi(t) = J(L, t) = -D_0 \frac{\partial P}{\partial x}\bigg|_L = \frac{L}{\sqrt{4\pi D t^3}} \exp\left\{-\frac{(L - \mu F t)^2}{4D t}\right\}. \tag{17}$$

Assuming statistical independence between capture and translocation events the blockade time distribution p_{bl} is the convolution

$$p_{bl}(t) = \int_0^t dt' p_c(t') \, \psi(t - t')$$

of the transport ψ and capture p_c distributions. The above argument predicts the average blockade time to be $\tau_{bl} = 1/\lambda_c + L/(\mu F)$.

This two stage model provides expressions for the distribution of capture and translocation times in terms of the forcing F and a series of phenomenological parameters such as l_c, D_0 and μ_0 that could, in principle, be fit from the data. The model, however, refers only to the translocated protein ($P_c \to 1$ for $t \to \infty$ in the first stage) hence it is not able to predict the translocation probability (Figure 5).

A.2. *Smoluchowski drift–diffusion equation with radiation boundary*

A more realistic approach cannot neglect the role of the channel and the possibility that the protein spontaneously escapes from the *cis* boundary coming back to the ion solution. The above description does not incorporate the enormous conformational entropy reduction that polypeptide chains undergo when constrained to perform an almost single-file translocation in a channel. This process is expected to occur only through overcoming large and probably steep free energy barriers. These effects could be included by considering the Smoluchowski Equation (12) in an interval $[0, L]$ with a probability current

$$J(x, t) = -D_0 \frac{\partial P}{\partial x} + u(x)P, \tag{18}$$

where

$$u(x) = \mu_0 \left\{ F - \frac{dG}{dx} \right\} \tag{19}$$

is the sum of the constant drift and the free energy contribution resulting from intrachain and pore–chain interactions. The initial condition, $P(x, 0) = \delta(x - x_0)$, with $x_0 \in (0, L)$, accounts for the situation where the molecules are released in position x_0, with $x_0 \to 0$ to match the simulation initial state. Following [19], we assign the radiation boundary conditions (RBC) at the channel ends $x = 0$ and $x = L$: $J(0, t) = -R_L P(0, t)$ and $J(L, t) = R_L P(L, t)$ which take into account partial absorption. A radiation boundary, with coefficient R, reduces to an absorbing, reflecting one, in the limits $R \to \infty$ and $R \to 0$, respectively. The survival probability in $[0, L]$ is

$$S(t) = \int_0^L dx P(x, t), \tag{20}$$

and using the analogue of Equation (16), the blockade time distribution is $\psi(t) = J(L, t) - J(0, t)$, which, due to RBC, takes on the final form

$$\psi(t) = R_0 P(0, t) + R_L P(L, t). \tag{21}$$

Again, the nature of the exit process is determined by the flux properties at the boundaries. As we are interested in successful translocation events, it is convenient to split the blockades into *translocation* (a molecule from $x = 0$ is absorbed at $x = L$) and its complementary process *no translocation* (the molecule from $x = 0$ is never absorbed at $x = L$). The successful translocation probability is simply given by integrating $J(L, t)$ over all time,

$$P_{Tr} = R_L \int_0^\infty dt P(L, t) \tag{22}$$

and the average translocation time of our simulations coincides with the average first arrival time to the *trans* side $x = L$. Therefore, by introducing the conditional time distribution $\psi_L(t) = R_L P(L, t)/P_{\mathrm{Tr}}$ which is conditioned to successful translocation, we obtain

$$\tau = \frac{R_L}{P_{\mathrm{Tr}}} \int_0^\infty \mathrm{d}t \, t P(L, t).$$

Explicit expressions for P_{Tr} and τ could be found by solving the Smoluchowski problem (12), (18), (19) through a time Laplace transform. More specifically, it is possible to show [12,19] that the quantities of interest are obtained by solving the boundary value problem

$$\frac{\partial}{\partial x} \left\{ D_0 \frac{\partial \widehat{P}}{\partial x} + \widehat{P} \left[\mu_0 F - \frac{\mathrm{d}G(x)}{\mathrm{d}x} \right] \right\} = s\widehat{P} - \delta(x - x_0) \tag{23}$$

$$\widehat{J}(0, t) = -R_0 \widehat{P}(0, t), \quad \widehat{J}(L, t) = R_L \widehat{P}(L, t) \tag{24}$$

via the standard technique of Green's functions [13,20,25] for $s = 0$ and in the limit $x_0 \to 0$ corresponding to molecules injected close to the *cis* side of the channel. The final expressions are

$$P_{\mathrm{Tr}}(F) = \frac{1}{1 + R_0/R_L e^{-\beta FL} + R_0/D_0 M_+(F)} \tag{25}$$

and

$$\tau_{\mathrm{bl}}(F) = \frac{D_0 M_-(F) + R_L e^{\beta FL} M_0(F)}{D_0 R_L e^{\beta FL}}, \tag{26}$$

where $\beta = \mu_0/D_0$ and the explicit dependence on the free energy profile $G(x)$ is hidden in the functions

$$M_\pm(F) = \int_0^L \mathrm{d}x \, e^{\pm\beta[G(x)-Fx]}, \quad M_0(F) = \int_0^L \mathrm{d}x \, e^{-\beta[G(x)-Fx]} \int_x^L \mathrm{d}y \, e^{\beta[G(y)-Fy]}$$

(we assume that $G(L) = G(0) = 0$).

When FL is sufficiently larger than the free energy barrier, we are allowed to neglect $G(x)$ and expand for large F, thus $M_+ \simeq 1/(\beta F)$, $M_- \simeq (\beta F)^{-1} \exp(\beta FL)$ and $M_0 \simeq 1/(\beta F)$, we have $P_{\mathrm{Tr}}(F) \to 1$ and $\tau_{\mathrm{bl}}(F) \simeq F^{-1}$.

The computation of $G(x)$ for the MBP directly from umbrella sampling simulations presents some technical problems; however preliminary results and previous indications [13,26] from the context of protein mechanical pulling [27] too, suggest considering a symmetric trapezoid shape in the reaction coordinate

$$G(x) = G_0 \begin{cases} x/h & 0 \leq x < h \\ 1 & h \leq x < L - h \\ (L - x)/h & L - h \leq x \leq L \end{cases} \tag{27}$$

defined by the parameters G_0 and h. In conclusion, the expressions (25) and (26) depend on the parameters D_0, R_0, R_L, G_0 and h which are considered to be free and tunable to fit either the simulation or experimental data.