JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

# A High-Throughput Steered Molecular Dynamics Study on the Free Energy Profile of Ion Permeation through Gramicidin A

Toni Giorgino* and Gianni De Fabritiis*

Computational Biochemistry and Biophysics Laboratory (GRIB-IMIM), Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB), C/Doctor Aiguader 88, 08003 Barcelona, Spain

**ABSTRACT:** Steered molecular dynamics (SMD) simulations for the calculation of free energies are well suited for high-throughput molecular simulations on a distributed infrastructure due to the simplicity of the setup and parallel granularity of the runs. However, so far, the computational cost limited the estimation of the free energy typically over just a few pullings, thus impeding the evaluation of statistical uncertainties involved. In this work, we performed two thousand pulls for the permeation of a potassium ion in the gramicidin A pore by all-atom molecular dynamics in order to assess the bidirectional SMD protocol with a proper amount of sampling. The estimated free energy profile still shows a statistical error of several kcal/mol, while the work distributions are estimated to be non-Gaussian at pulling speeds of 10 Å/ns. We discuss the methodology and the confidence intervals in relation to increasing amounts of computed trajectories and how different permeation pathways for the potassium ion, knock-on and sideways, affect the sampling and the free energy estimation.

## 1. INTRODUCTION

Biologically relevant events often take place at time scales far beyond those accessible by fully atomistic simulations, for example, conduction of ions through narrow channels.[1−3] A successful approach for describing molecular phenomena at longer time scales is to average out all but a few degrees of freedom of the system by selecting a reaction coordinate. The forces affecting the process are then described as an effective potential of mean force (PMF), i.e., the free energy profile along the reaction coordinate.[4−6] The PMF can, in principle, be computed by sampling the equilibrium statistical distribution of the system. However, the time required for the system to cross high free energy barriers may be long enough to make the computation infeasible.

This problem has been successfully addressed with biasing protocols,[7,8] such as umbrella sampling (US),[9] which overcomes this limitation by sampling several biased equilibrium distributions, which are later merged by histogram-based techniques.[10−13] Jarzynski equation (JE) and Crooks fluctuation theorem (CFT) equalities showed that the PMF can also be recovered from nonequilibrium steered molecular dynamics.[14,15] Steered molecular dynamics (SMD) is a well-known computational protocol to exploit nonequilibrium sampling, in which the application of time-dependent biasing forces guides the system according to a predefined protocol.[16,17] In SMD experiments, several pulls are simulated in one (forward)[15,18] or two (forward and reverse) directions.[14,19−21]

A number of previous studies have used SMD simulations to compute free energy profiles on realistic biomolecules. The JE applied to one-directional (forward only) SMD experiments has been used by several authors to compute the free energy profile in large biomolecular systems. Among the most recent works, Cuendet et al.[22] used two groups of $n \sim 150$ single directional trajectories (total sampled time of approximately 2 $\mu$s) to compute the PMF of the T cell receptor with a major histocompatibility complex peptide (TCR-pMHC) complex and a mutant. Martin et al.[23] used single directional SMD on a large system to compute the energetics of translocation of a polynucleotide through a nanopore. The pullings were conducted with various parameters, and the PMF curves computed on 2−6 samples. Liu et al.[24] studied the permeation of $Na^+$ through gramicidin A (gA) with $n \simeq 8$ single directional trajectories. Zhang et al.[25] used $n \sim 35$ trajectories and four different computational methods based on the JE to compute the unbinding of acetylcholine from the α-7 nicotinic receptor along four different paths. Jensen et al. computed the energetics of sugar permeation through lactose permease[26] and glycerol through aquaglyceroporin,[27] respectively, via four SMD runs with cumulant expansion. Comparatively, fewer works have discussed the use of bidirectional pulling experiments in large proteins. De Fabritiis et al.[28] computed the PMF in gA with 25 pulls in each direction. Forney et al.[19] computed the PMF of gA, as well, with 10 pulls in each direction comparing various ionic strengths and backbone restrain types. Due to limitations in computational resources, most studies could only sample a low number of trajectories and were therefore unable to assess the impact of increased sampling on the precision of the PMF profiles.

Here, we expand the amount of sampling with respect to previous bidirectional SMD studies by almost two orders of magnitude in a realistic system, the gA dimer embedded in a membrane and explicit solvent, to test the methodology in a properly sampled system. Gramicidin A (see Figure 1) is a helical antibacterial dimer (15 amino acids each) which increases permeability of biological membranes to inorganic ions.[29] The backbone of the gA dimer forms a narrow pore, allowing a single file of water molecules (or potassium ions) to fill it. Due to the diameter of the pore, the transport of a single ion drags with it a

**Received:** December 9, 2010
**Published:** April 27, 2011

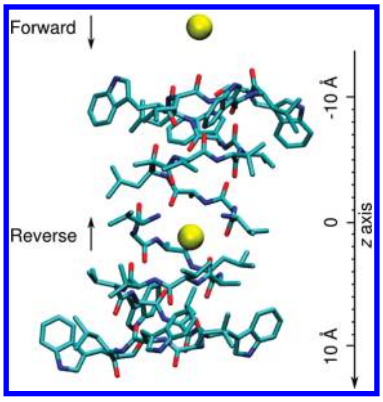


**Figure 1.** Structure of the gramicidin A dimer, with the starting $K^+$ ion positions for the forward and reverse steered molecular dynamics runs (yellow). In the former, the cation is pushed from the outside toward the inside of the channel ($z$ increasing from −15 toward 0 Å). In the reverse simulations, the ion is pulled along the time-reversed path. The $z$ axis runs through the center of the pore. Lipid bilayer (117 DMPC molecules), water (8668 molecules), and other solvated ions have been omitted for clarity.

column of six to nine water molecules in a single file, reducing drastically the possibility that water molecules slip past each other.[30] The small size and the early availability of its structure made gA a frequently used model for a membrane channel.[31,32] Despite its simplicity, permeation is not so well reproduced computationally; several studies reported with a barrier to permeation of 10−20 kcal/mol, several kcal/mol higher than the experimental one.[3,33,34] The barrier height was recently shown to be much improved with the use of a polarizable force field.[35]

We performed an extensive set of all-atom molecular dynamics (MD) experiments on the gA channel[36] and computed the PMF from bidirectional pulling experiments[21,37] using 1000 pulls per direction. Confidence bands for an increasing number of pulls, computed with a variable-size bootstrap procedure, are also presented. The importance of sampling effectively the degrees of freedom orthogonal to the reaction coordinate is well illustrated by two permeation pathways shown by the potassium ion in the interface between the two monomers of gA with different free energy profiles.

## 2. MATERIALS AND METHODS

**2.1. The Potential of Mean force.** The PMF is a convenient description of the energetics of the system obtained integrating out all of the degrees of freedom with the exception of one reaction coordinate, $z = z(\mathbf{R})$, which should capture the interesting features of the system. The PMF $G(z)$ would then be

$$e^{-\beta\Delta G(z')} = \frac{\int d\mathbf{R}d\mathbf{P}\delta(z(\mathbf{R}) - z') \exp(-\beta H)}{\int d\mathbf{R}d\mathbf{P} \exp(-\beta H)}$$

where $H = H(\mathbf{R},\mathbf{P})$ is the Hamiltonian of the system, $\mathbf{R} = (\mathbf{r}_1, ..., \mathbf{r}_N)$, $\mathbf{P} = (\mathbf{p}_1, ...,\mathbf{p}_N)$ are the positions and momenta of the $N$ atoms, $\delta(\cdot)$ is the Dirac delta function, and $\beta = 1/(k_BT)$, where $k_B$ is the Boltzmann constant and $T$ the temperature of the system.

The Crooks fluctuation relation[18,38,39] allows one to compute the equilibrium free energy difference $\Delta G$ between two states "0" and "1" described by two Hamiltonians $H_0$ and $H_1$ as

$$\frac{P_F(+\beta W)}{P_R(-\beta W)} = \exp(\beta(W - \Delta G)) \quad (1)$$

where $W$ is the external work done on the system by forcing it to change from state 0 to 1, and $P_F$ and $P_R$ are the probability distributions of releasing the work $W$ into the system during a transformation in the forward (F) $0 \rightarrow 1$ and reverse (R) $1 \rightarrow 0$ direction, respectively, in a finite time. The Crooks fluctuation relation is a generalization of the Jarzynski equality (JE):[15]

$$\langle \exp(-\beta W)\rangle_F = \exp(-\beta\Delta G)$$

recovered from Crooks fluctuation relation by integrating both sides of eq 1. The Crooks fluctuation relation can be estimated using the optimal Bennett acceptance ratio method.[40] Interestingly, these two fundamental relations have their equilibrium counterparts obtained for an infinite pulling speed, where CFT resembles an equilibrium relation previously derived by Shing and Gubbins[41] and JE corresponds to Widom's formula used to compute the chemical potential by test particle insertion[42] (with well-known poor convergence properties).[43] The exponential average of the JE causes few rare low-energy trajectories dominating the estimate of $\Delta F$. When only few trajectories are available, the estimate can be improved using a first-order cumulant expansion, which is exact in the limit when the distribution of $W$ values is Gaussian.[22,44] The CFT has much better convergence properties than a direct application of JE and was therefore used in a previous study[28] and in this study. Both the JE and the CFT have been confirmed experimentally in atomic force and single-molecule pulling experiments.[18,39]

In this work, we use the estimator for the PMF $G(z)$ proposed by Minh and Adib:[21]

$$e^{-\beta G(z)} = \sum_t \left[ \left\langle \frac{n_F \delta(z - z_t) e^{-\beta W_0^t}}{n_F + n_R e^{-\beta(W - \Delta F)}} \right\rangle_F + \left\langle \frac{n_R \delta(z - z_{\tau - t}) e^{\beta W_\tau^{\tau - t}}}{n_F + n_R\, e^{\beta(W + \Delta F)}} \right\rangle_R \right] \times \frac{e^{\beta \Delta F_t}}{\sum_t e^{-\beta[V(z;t) - \Delta F_t]}} \quad (2)$$

where $n_F$ and $n_R$ are the number of forward and reverse trajectories, respectively, $W_a^b$ is the partial work performed in the interval between time $a$ and $b$, $\tau$ is the final simulation time, and $\Delta F_t$ is the free energy difference between the initial state and the one at time $t$. In this work we used the implementation of eq 2 provided in the FERBE package.[45]

**2.2. Steered MD Protocol.** To describe the permeation experiments of one potassium ion through the gA channel, the collective reaction coordinate was assumed to be the $z$ coordinate of one of the cations, which will be called $K^+_{SMD}$ in the following, i.e., $z(\mathbf{R}) = z_K$. The chosen cation was driven through the channel applying a simple harmonic biasing potential parallel to the $z$ axis producing the force:

$$F_z(z_K, t) = -k(z_K - b(t)) \quad (3)$$

where $k$ is the spring constant, $z_K$ is the instantaneous $z$ coordinate of the ion, and $b(t)$ is the time-dependent equilibrium point for the biasing force. The spring constant $k = 10$ kcal/mol of the biased system was set in order to fulfill the strong spring approximation, $k > \max(2\alpha/\delta z^2, 2U_{max}/\delta z^2)$, where $\delta z$ is the spatial resolution that we are seeking for the PMF, $U_{max}$ is the maximum energetic barrier that we expect in $\delta z$, and $\alpha \gg 1$. The

biasing position was displaced linearly with time:

$$b(t) = \begin{cases} z_F + vt & \text{(forward)} \\ z_R - vt & \text{(reverse)} \end{cases}$$

The cumulative work profile was obtained by integrating the instantaneous forces over the corresponding interval:

$$W(t) = \int_{t'=0}^{T} F(t')v\mathrm{d}t' \quad (4)$$

with $F(t)$ given by eq 3. For the numerical computation of eq 4, the integral was approximated as a discrete summation over time intervals of length $\Delta t$:

$$W_a^b = \sum_{t_j=a}^{b} -k(z(t_j) - z_0 \mp vt_j)v\Delta t \quad (5)$$

where $t_j = j\Delta t$ is the time corresponding to the $j$-th interval, $z_0$ is the starting position of the pull, and the sign is taken according to the pull direction. The $z$ axis was divided in 100 bins, equally spaced over the interval $z = -10, ..., 0$ Å.

**2.3. Preparation of the System.** The gA dimer was prepared based on the Protein Data Bank entry PDB:1JNO,[46] extending the protocol already presented in De Fabritiis et al.[28] The structure used in the previous study, comprising the gA dimer and dimyristoylphosphatidylcholine (DMPC) lipid bilayer, was solvated with 8668 TIP3 water molecules and ionized at a ionic strength of 150 mM with 24 pairs of $K^+$ and $Cl^-$ ions. The final system, comprising 40 410 atoms, was then equilibrated at 1 atm and 305 K with the CHARMM27[47] force field in the NPT ensemble for approximately 13 ns. The lipid bilayer is oriented in the $xy$ plane, and the $z$ axis goes through the gA pore (see Figure 1). The simulation box resulting from the NPT equilibration was 66.1 × 65.8 × 88.9 Å$^3$. The preparation runs were performed with the NAMD program,[48] with particle-mesh Ewald (PME) electrostatics,[49] rigid bonds, cubic periodic boundary conditions, and a time step of 2 fs.

In order to generate the initial configurations for the forward run, the position of one potassium ion was exchanged with that of a water molecule located close to the entrance of the pore, i.e., approximately at (0,0,−15) Å. For the reverse runs, the ion was exchanged with the water molecule closest to the middle of the channel. The two systems were subject to a further 20 ns of equilibration in the NVT ensemble, while restraining $K^+_{SMD}$ to its initial position with a spring constant of 10 kcal/mol/Å$^2$. After the initial 20 ns of equilibration, the runs were extended further for 20 ns in the same conditions, taking snapshots at 200 ps intervals, thus yielding 100 snapshots for each of the two systems. Each snapshot was used as an initial configuration for 10 SMD runs. Figure 1 shows the ion at the initial positions $z_F$ and $z_R$ for one of the forward and reverse pulls, respectively.

Further analysis on the configurations, e.g., the water occupancy of the pore, and statistics on ion−water−protein relative positions were performed using the scripting facilities of the VMD program.[50]

**2.4. Production Runs.** We performed 1000 forward and 1000 reverse SMD runs, starting from the 200 distinct initial configurations prepared according to the protocol explained above. The pulling speed and the SMD spring constants were set at $v = \pm 10$ Å/ns and $k = 10$ kcal/mol/Å$^2$, respectively. Each SMD run lasted 2 ns, long enough for the pulled ion to reach the starting position for the opposite direction, i.e., until $z < z_R$ for the forward runs and $z > z_F$ for the reverse ones. When the ion was outside of the channel, a flat bottom potential was applied to keep the ion in line with the pore during the approach (the region $z <$ 10 was excluded from the PMF calculation). No constraints were imposed on the ion on the plane orthogonal to $z$. Only the center of mass of the Cα atoms of the pore was restrained to its initial position with a harmonic potential of 100 kcal/mol/Å$^2$, to avoid the pore being displaced out of the membrane by the net force applied throughout the simulations. The restrain was just applied to the center of mass, in order not to artificially constrain the local helix radius, interdimer distance, or orientations of the side-chain. The pore was therefore free to expand under the influence of the permeating cation; this flexibility has been shown to play an important role in the permeation energetics.[19,51,52]

All of the production runs were performed with ACEMD,[53] which leverages off-the-shelf accelerated graphic processing units (GPUs) allowing one to achieve approximately 100 ns per day of simulated time on a single GPU for system size of the order of 23 000 atoms, performance decreasing linearly with system size. Production runs have been performed in the constant volume and temperature (NVT) ensemble, Langevin thermostat at 305 K with a relaxation of 0.1/ps, computing the electrostatic interactions with the PME algorithm.[54] The integration time step was set to 4 fs, enabled by the hydrogen mass repartitioning scheme[53,55] available in ACEMD. This scheme allows for longer timesteps by using the property that the equilibrium distribution is not affected by individual atom masses provided that the total mass of the system stays the same. Transport properties change by less than 10%, a small amount compared of the errors intrinsic in the TIP3P water model compared to real water.[55]

The runs were performed on a distributed computing grid called GPUGRID.net.[56] We set up a server based on the Berkeley Open Infrastructure for Network Computing (BOINC) to automatically distribute the runs through the Internet.[57] In order to be executed remotely, each forward and reverse run was arranged as a separate work unit. As soon as each participating computer finished computing the assigned simulation time, it returned a log file with the trajectory $z(t)$ and the force $F(t)$ exerted on the SMD ion, recorded at intervals of 200 fs. The computational effort used for computing the PMF curves amounted to the generation of 4 $\mu$s of total simulation data.

**2.5. Bootstrapping Procedure.** The convergence properties of the PMF were estimated with respect to increased configuration sampling by recomputing the potential curves with a varying sample size bootstrapping technique,[58] similar to the one employed by Cuendet et al.[22] In this procedure we constructed resampled sets of the available bidirectional pulls of increasing cardinality. Each of the available pulling trajectories was randomly taken zero or more times, in order to build a resampled set containing $R$ bidirectional trajectories. A PMF profile was computed considering only the resampled set, and the PMF depth was obtained. The process was iterated until $B$ bootstrapped replicas were obtained, finally yielding the standard deviation of the PMF depth, $\sigma(R)$. The procedure was repeated for resampled sets of sizes $R = 10, 50, 100, 250, 500, 750, n$, with $n = 1000$ being the count of all available pulls. The case $R = 1000$ corresponds to the plain bootstrap procedure, which creates resampled sets as large as the number of trajectories originally available.

## 3. RESULTS

We computed cumulative work profiles using eq 5 for all of the 2000 pulling experiments. The upper and lower part of Figure 2

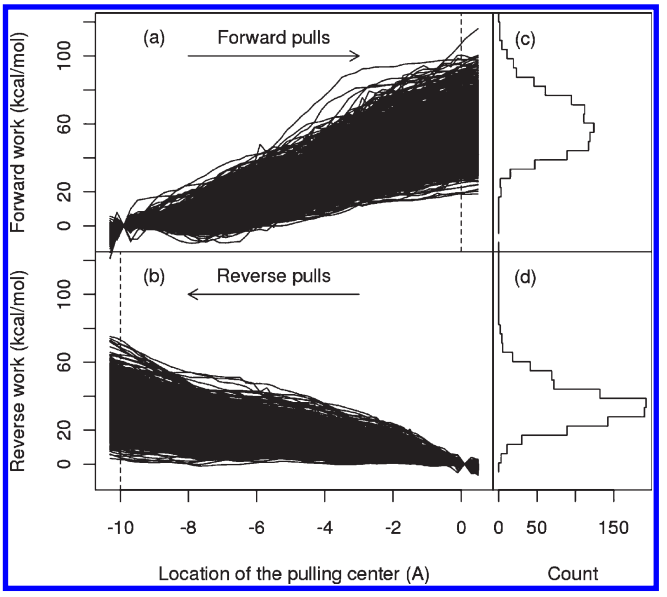


**Figure 2.** Profiles of the accumulated work $W(z)$ spent to pull the ion inside (a, forward) or outside (b, reverse) the gA channel at 10 Å/ns (1000 pulls per direction). The panels on the right-hand side show the distribution of final work values for the forward (c) and reverse (d) directions, respectively. The mean work performed in the forward direction at the end of the pulls was 55.5 kcal/mol (SD 14.9 kcal/mol); for the reverse direction it was 31.9 kcal/mol (SD 11.2 kcal/mol).

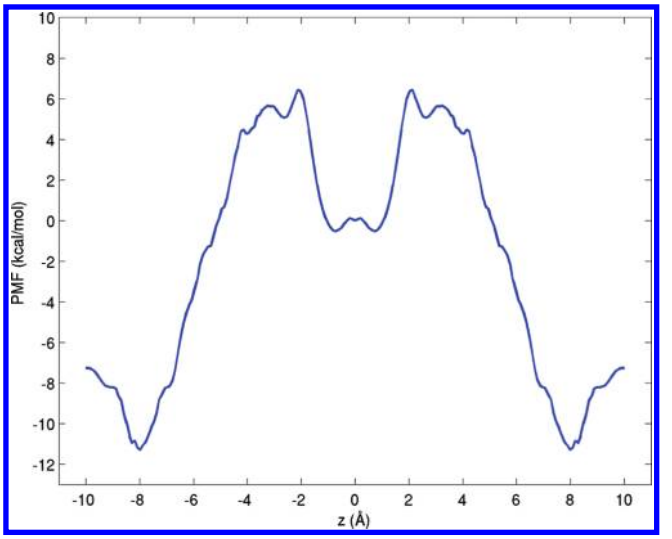


**Figure 3.** PMF curves for a $K^+$ atom to cross the gA channel, computed from 1000 steered MD experiments at pulling speed of 10 Å/ns in each direction.

shows the work profiles for the forward (a) and reverse work (b), respectively. For the forward pulls, the work was taken as the one required to push the $K^+_{SMD}$ ion from $z = -10$ to 0; for the reverse pulls, the end points are reversed. The distributions of final work values for the forward and reverse pulls are shown on the right-hand side of Figure 2, respectively in panels (c) and (d). The mean work performed in the forward direction at the end of the pulls (dashed line) was 55.5 kcal/mol [standard deviation (SD) 14.9 kcal/mol]; for the reverse direction it was 31.9 kcal/mol (SD 11.2 kcal/mol).

**3.1. Final Work Distributions.** The CFT implies that if the forward work values follow a normal distribution with a variance $\sigma$, the reverse work values should also be a Gaussian with the same variance.[59] To check whether the final work values obtained in the simulations follow a normal distribution, we applied the well-known Shapiro−Wilk normality test.[60] The test quantifies the probability $p$ that a given set of values could have been taken from a Gaussian distribution (null hypothesis). If the $p$ value computed by the test is smaller than a fixed threshold, usually taken as 0.05, then one concludes that there is strong evidence against normality. The Shapiro−Wilk test rejected the null hypothesis that the final work values follow normal distributions in either direction ($p < 4 \times 10^{-7}$ for both the forward and reverse runs). Given that the pulling speed used here was 10 Å/ns, it is unlikely that the work distribution is Gaussian, at least for a system with enough dissipation like pulling an ion through a pore.

**3.2. PMF Profiles.** The PMF curves recovered from the bidirectional pulls with the analysis protocol cited above[21] are shown in Figure 3. For clarity, the potential profile has been symmetrized around the $z = 0$ axis and offset so that the PMF is 0 at $z = -10$ Å. The PMF exhibits a binding site at $z \simeq 8.5$ Å and a total barrier height with respect to the bulk of ∼14 kcal/mol. The location of the binding site is approximately consistent with the ones reported in the literature.[19,34,35,61]

The barrier height obtained here is lower with respect to the one computed from the ref 28 data set (19 kcal/mol), obtained from 25 bidirectional pulls. As discussed in the same paper, the induced dipole of the water molecules surrounding the channel provides an important component to the barrier to permeation. In particular, when the $K^+$ ion is not in the middle of the channel, a large fraction of its electrostatic interaction energy is due to atoms between 6 and 16 Å of distance, i.e., in the second coordination shell. This fact underlines the role played by the water molecules' polarization and the finite time required for their reorientation. Part of the relatively higher barrier to permeation found in this study with respect to others performed with US[34,61] or SMD[19] may be therefore ascribed to the biological (150 mM) KCl ionic strength employed outside the pore, compared to higher concentrations used in other studies.

Furthermore, the previous study ref 28 sampled only 25 bidirectional pulls and therefore could not provide a measure of the statistical uncertainty. We shall show later that that this amount of sampling still incurs a statistical uncertainty of tens of kcal/mol.

**3.3. Permeation Pathways.** Ion and water permeate through the narrow gA channel as a single file;[30,61] the sequence of ion and water molecules should therefore be preserved during a pull. However, during some of the reverse (outgoing) pulls, $K^+_{SMD}$ was observed to exchange places with the preceding water molecule. These trajectories could be distinguished according to the value of the work $W$ cumulated at 300 ps, as shown in Figure 4a. We inspected the structural features of the two groups analyzing a subset of the reverse runs and labeled the trajectories for which $W(300\ \text{ps}) > 10$ kcal/mol as belonging to the "H" group (37% of the trajectories) and "L" otherwise (63%). To analyze the atomistic basis for this difference, we computed 52 additional trajectories recording the state of the system every 10 ps and labeled the pulls in groups H (14 pulls) and L (38 pulls) as above. Inspection of the trajectories in group H revealed that the order in the water file was partially lost around 300 ps after the beginning of the pull (Figure 4b and c). When this event occurred, the $K^+_{SMD}$ ion overtook the preceding water molecule

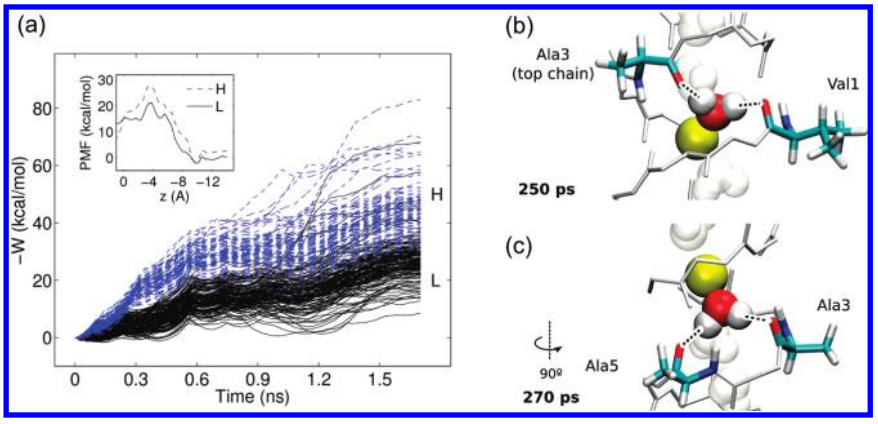


**Figure 4.** (a) Work profiles for the pulls in groups H (37% of the pulls at $v$ = 10 Å/ns) and L (63%), with the distinction drawn at $t$ = 0.3 ns. Inset: PMF profiles reconstructed considering the pulls of the two groups separately. (b and c) The putative structural explanation of the water file disruption in group H. In these runs the water molecule preceding the pulled ion, W2, formed hydrogen bonds with the carbonyl atoms of the backbone close to the dimer interface. The bonds were most frequently observed with residues Val1 of chain A (about 50% of the runs in group H) and Ala3 of chain B (70%).

(indicated as W2 in Figure 5, left). Conversely, in the pulls in group L, the order of the water file is preserved (Figure 5, right).

The interruption of the water file may be traced back to the formation of hydrogen bonds between W2 and carbonyl atoms in the gA backbone. Figure 4b and c shows the residues whose carbonyl atoms were most often acceptors of a hydrogen atom of W2 at the time when it was overtaken by $K^+_{SMD}$. A hydrogen bond was observed with residue Val1 of chain A in about 50% of the runs in group H and with residue Ala3 of chain B in about 70% (chains A and B being the monomer placed at positive and negative $z$, respectively). Radial and angular cutoffs for hydrogen bonds were taken as 4 Å and 30°, respectively.

Finally, we performed a control simulation to check the stability of the dimer's embedding in the hydrophobic membrane environment. Specifically, we checked the equilibrium configuration of the dimer when the permeating ion was held close to the dimer interface ($z$ = 0 Å) by a constant biasing potential of $k$ = 10 kcal/mol/Å$^2$, analogous to a setup that would be used for a US window in the middle of the channel. A simulation of 50 ns was sufficient to disrupt the pore structure, with water fingering from bulk on the side of the channel in order to balance the ion charge in the middle of the membrane. Thus, the biased equilibrium with the ion forced to stay within the pore is substantially different from the unbiased permeation event, which could imply that SMD is better than US for this system.

**3.4. Sampling and Convergence.** We characterized the convergence of PMF estimates with respect to increased sampling with two methods. First, the available trajectories were split in nonoverlapping blocks of different sizes, computing the PMF profiles using the data contained in each, assuming $G(0)$ = 0. Figure 6a shows the PMF curves obtained using 10 bidirectional trajectories each, i.e., 1−10 (first block), 11−20 (second block), and so on, for a total of 100 profiles. Analogously, Figure 6b shows PMF profiles computed with blocks of 50 bidirectional pulls each (20 profiles), and Figure 6c shows the same using 250 bidirectional pulls at a time (4 profiles). For reference, in panels (a−c) the thick blue line shows the PMF computed with all available data. Increased sampling clearly improves the reproducibility of PMF curves. In particular, it appears that curves obtained with only 10 bidirectional pulls each are affected by a statistical error of the order of 10 kcal/mol, comparable to the PMF depth.

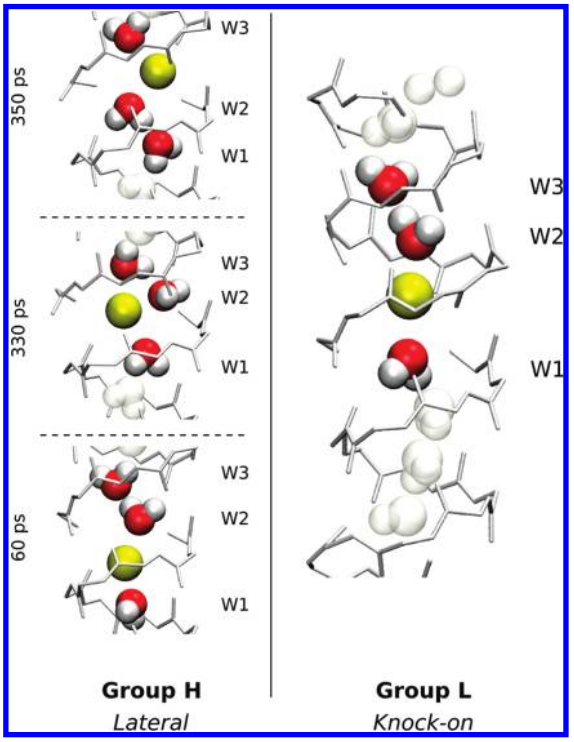


**Figure 5.** Left: the position of water molecules and $K^+_{SMD}$ at three snapshots (60, 330, and 350 ps after the beginning of the run) during a representative trajectory of group H. The ion steered upward (yellow) laterally "slips past" the preceding water molecule in the file (W2). The water molecule is held in place by hydrogen bonds with carbonyl oxygens of the backbone and exchanges sides with the ion. Right: during runs in group L, the water filling the channel is displaced together with the ion, preserving the sequence of the file (knock-on).

We used the bootstrapping procedure outlined in Section 2.5 to quantify the effect of increased sampling on the variability of the PMF depth, $G(10)-G(0)$. Figure 6d shows the SD of the PMF depth over the bootstrapped profiles, obtained including different number of trajectories. The results confirm that the SD for $R$ = 10 trajectories is of the order of 5 kcal/mol, which decreases to 3 for $R$ = 50 and to 2 for $R$ = 250. When using all of

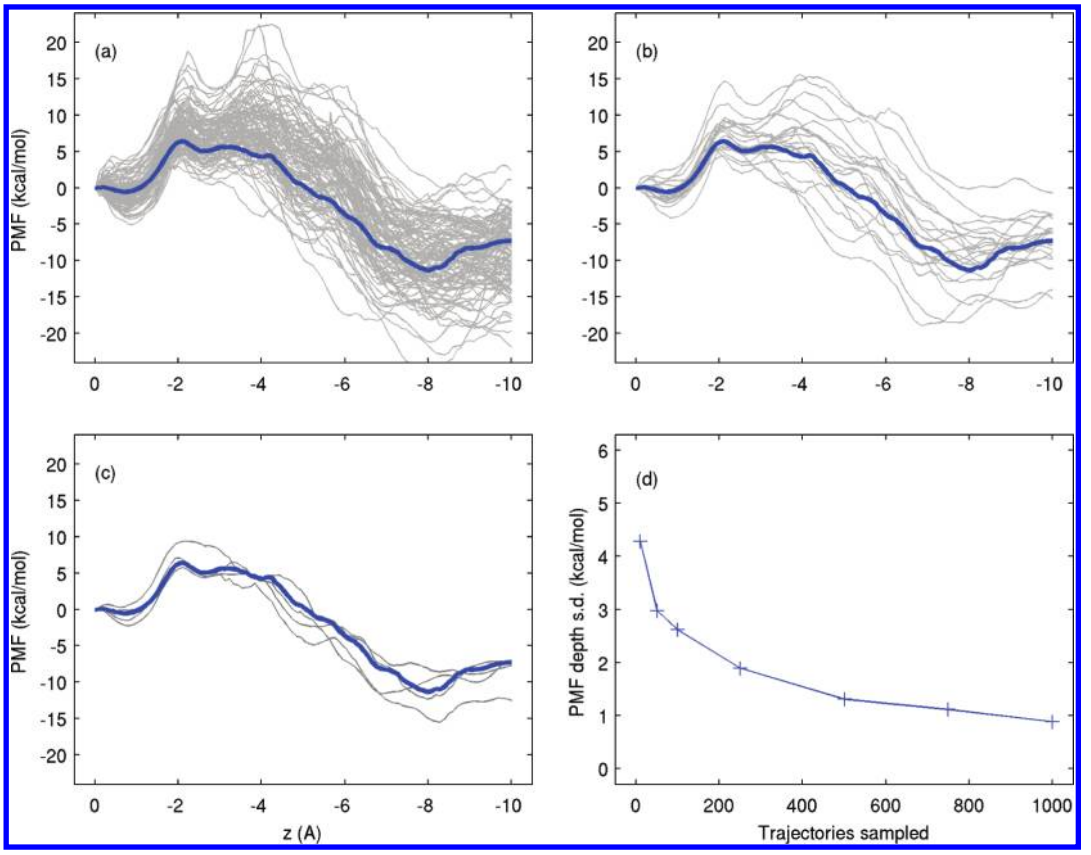


**Figure 6.** PMF curves computed in nonoverlapping blocks of (a) 10 bidirectional, (b) 50, and (c) 250 pulls. SD of the PMF depth with various amounts of sampled trajectories, obtained by 200-fold bootstrapping (d).

the 1000 available trajectories, the bootstrap analysis estimates a statistical error of 1 kcal/mol.

Finally, we performed two additional simulation sets as controls to check the influence of pulling speed on the PMF depths and profiles. The structures for these runs were taken from the previous study[28] and were slightly smaller than the production simulations (∼ 29 000 atoms in total). In the two control sets the ion was pulled at $v$ = 10 (221 bidirectional pulls) and 2.5 Å/ns (171 bipulls), respectively. We used the aforementioned bootstrapping technique to compute the convergence of the PMF depth in the two data sets (Table 1). The final PMF profiles are qualitatively similar with each other and with one obtained from the production simulations. Consistently to what is observed in the production runs, decreasing the statistical error below 2 kcal/mol requires significant computational effort for both pulling speeds. Given that slow pulls require four times as much simulation time as the fast ones, performing SMD at $v$ = 10 Å/ns appears to be more computationally efficient for this system.

**Table 1. Statistical Uncertainty of the PMF Depth Estimate (in kcal/mol), at Ion Pulling Speeds of $v$ = 10 and 2.5 Å/ns, Computed through Bootstrapping for Increasing Number of Pulls**[a]

| | PMF depth CI (kcal/mol) | |
|---|---|---|
| pulls | 10 Å/ns | 2.5 Å/ns |
| 10 | 8.3 | 8.7 |
| 25 | 5.4 | 6.9 |
| 50 | 3.9 | 5.8 |
| 100 | 3.2 | 4.3 |
| 150 | 2.8 | 3.9 |
| 171 | – | 3.6 |
| 221 | 2.5 | – |

[a] The figures show the width of the 68% confidence interval (CI), matching a 1-$\sigma$ interval around the mean. Computation of each of the slower pulls requires four times the computational effort of a fast one.

## 4. DISCUSSION AND CONCLUSIONS

In this paper we performed an extensive set of SMD forward−reverse experiments, sampling a simple but realistic biomolecular test system, gA, well beyond the state of the art, to analyze the effect of increased sampling on the precision of the estimate of the PMF of permeation of a potassium ion steered through the channel.

The advent of accelerator processors and codes able to exploit them, like ACEMD for graphical processing units,[53] should play an important role in reducing sampling limitations to obtain free energy estimates up to the accuracy of the force fields. The SMD protocol is well suited for high-throughput molecular simulations in distributed computing infrastructures, like GPUGRID. Still, considering the particle being steered, the case of ion permeation through gA is a relatively simplistic test case. In particular, a single ion does not have orientational nor conformational degrees of freedom; this is not the case when dealing with generic ligands, whose internal degrees of freedom have to be sampled as well.

The scientist has to prepare two initial systems (forward and reverse) from which many simulations are spawned. Umbrella sampling offers a similar degree of parallelism, as long as several uncorrelated configurations are generated to start each US window or a sufficient equilibration time (100 ns or more) is allowed for each window.[56] US simulations with stratification of windows showed much less variability than the one reported here for SMD, but the question remains of how much is actually due to poorer sampling. As each US window is starting from a single initial configuration, the US protocol is averaging the effective potential on a small area of the configurational space close to it. If multiple configurations are used, as in ref 56, similar levels of fluctuations in the free energy are obtained. With US, the method used to generate the initial configurations of the biasing windows has a crucial importance, but it is subject to the specific choice of the scientist; the SMD protocol mitigates the problem of generating initial conditions because better sampling is achieved simply by increasing number of trajectories. However, for gA, the US biased equilibrium with the ion forced to stay within the pore was structurally different from the unbiased configuration and thus probably less representative of a permeation event than a SMD pull.

Finally, the choice of the pulling speed influences the amount of pulls for a fixed computational cost: pulling too fast would produce higher energy pathways but allows for more pulls, while slower pulls would be closer to equilibrium but more computationally demanding. Given that the system is in nonequilibrium, however, there may be regions of the phase space that only become accessible after some transient time, like for the two permeation pathways showed for the potassium ion in gA, and faster pulling speeds may prevent some conformational transitions from happening. Therefore, the preference of SMD versus US is probably system dependent, and the biasing methodology that produces the lesser perturbation compared to the unbiased case should be chosen.

In the case studied here, thousands of pulls were required to reach a statistical precision within 1 kcal/mol. Even though the results were obtained on the basis of extensive experiments on a specific, admittedly simple, system, similar considerations may apply to more complex cases.

## AUTHOR INFORMATION

### Corresponding Author

*E-mail: toni.giorgino@upf.edu, gianni.defabritiis@upf.edu.

## ACKNOWLEDGMENT

The authors would like to thank Dr. S. K. Sadiq, I. Buch, and M. J. Harvey for reading the manuscript. We acknowledge Sony Computer Entertainment, Spain. G.D.F. acknowledges support from the Ramon y Cajal scheme and by the Spanish Ministry of Science and Innovation (ref. FIS2008-01040). T.G. acknowledges support from the Programa Beatriu de Pinós from the Generalitat de Catalunya. This work was partially supported by the virtual physiological human EU network of excellence (VPH-NoE). The authors would like to express a special acknowledgment to the volunteers who contribute to GPUGRID (formerly PS3GRID) distributed computing networks.

## REFERENCES

(1) Allen, T. W.; Baştuğ, T.; Kuyucak, S.; Chung, S.-H. *Biophys. J.* **2003**, *84*, 2159.
(2) Jordan, P. C. *IEEE Trans. Nanobiosci.* **2005**, *4*, 3.
(3) Allen, T. W.; Andersen, O. S.; Roux, B. *Biophys. J.* **2006**, *90*, 3447.
(4) Åqvist, J.; Luzhkov, V. *Nature* **2000**, *404*, 881.
(5) Treptow, W.; Tarek, M. *Biophys. J.* **2006**, *91*, L81–L83.
(6) Åqvist, J.; Warshel, A. *Biophys. J.* **1989**, *56*, 171.
(7) Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 2600–2607.
(8) Zwier, M. C.; Chong, L. T. *Curr. Opin. Pharmacol.* **2010**, *10*, 745–752.
(9) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.
(10) Bernèche, S.; Roux, B. *Nature* **2001**, *414*, 73.
(11) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275.
(12) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011.
(13) Kato, M.; Warshel, A. *J. Phys. Chem. B* **2005**, *109*, 19516.
(14) Crooks, G. E. *Phys. Rev. E* **2000**, *61*, 2361.
(15) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690.
(16) Izrailev, S.; Stepaniants, S.; Isralewitz, B.; Kosztin, D.; Lu, H.; Molnar, F.; Wriggers, W.; Schulten, K. Steered Molecular Dynamics. In *Lecture Notes in Computational Science and Engineering 4*; Deuflhard, P., Hermans, J., Leimkuhler, B., Mark, A. E., Reich, S., Skeel, R. D., Eds.; Springer-Verlag: Berlin, Germany, 1998.
(17) Isralewitz, B.; Baudry, J.; Gullingsrud, J.; Kosztin, D.; Schulten, K. *J. Mol. Graphics Modell.* **2001**, *19*, 13.
(18) Hummer, G.; Szabo, A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3658.
(19) Forney, M. W.; Janosi, L.; Kosztin, I. *Phys. Rev. E* **2008**, *78*, 051913.
(20) Kosztin, I.; Barz, B.; Janosi, L. *J. Chem. Phys.* **2006**, *124*, 064106.
(21) Minh, D. D. L.; Adib, A. B. *Phys. Rev. Lett.* **2008**, *100*, 180602.
(22) Cuendet, M. A.; Michielin, O. *Biophys. J.* **2008**, *95*, 3575.
(23) Martin, H. S. C.; Jha, S.; Howorka, S.; Coveney, P. V. *J. Chem. Theory Comput.* **2009**, *5*, 2135.
(24) Liu, Z.; Xu, Y.; Tang, P. *J. Phys. Chem. B* **2006**, *110*, 12789.
(25) Zhang, D.; Gullingsrud, J.; McCammon, J. A. *J. Am. Chem. Soc.* **2006**, *128*, 3019.
(26) Jensen, M. O.; Yin, Y.; Tajkhorshid, E.; Schulten, K. *Biophys. J.* **2007**, *93*, 92.
(27) Jensen, M. O.; Park, S.; Tajkhorshid, E.; Schulten, K. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6731.
(28) De Fabritiis, G.; Coveney, P. V.; Villà-Freixa, J. *Proteins* **2008**, *73*, 185.
(29) Miloshevsky, G. *Biophys. J.* **2004**, *86*, 92.
(30) Hille, B. *Ion Channels of Excitable Membranes*, 2nd ed.; Sinauer Associates, Inc: Sunderland, Massachusetts, 2001; Chapter 11, p 299.
(31) Chung, S. H.; Kuyucak, S. *Biochim. Biophys. Acta* **2002**, *1565*, 267.
(32) Mamonov, A. B.; Coalson, R. D.; Nitzan, A.; Kurnikova, M. G. *Biophys. J.* **2003**, *84*, 3646.
(33) de Groot, B. L.; Tieleman, D. P.; Pohl, P.; Grubmüller, H. *Biophys. J.* **2002**, *82*, 2934.
(34) Baştuğ, T.; Kuyucak, S. *J. Chem. Phys.* **2007**, *126*, 105103.
(35) Patel, S.; Davis, J. E.; Bauer, B. A. *J. Am. Chem. Soc.* **2009**, *131*, 13890–13891.
(36) Andersen, O. S.; Koeppe, R. E., II; Roux, B. *IEEE Trans. Nanobiosci.* **2005**, *4*, 10.
(37) Ytreberg, F. M.; Swendsen, R. H.; Zuckerman, D. M. *J. Chem. Phys.* **2006**, *125*, 184114.
(38) Crooks, G. E. *Phys. Rev. E* **1999**, *60*, 2721.
(39) Collin, D.; Ritort, F.; Jarzynski, C.; Smith, S. B.; Tinoco, I.; Bustamante, C. *Nature* **2005**, *437*, 231.
(40) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245.
(41) Shing, K. S.; Gubbins, K. E. *Mol. Phys.* **1982**, *46*, 1109.
(42) Delgado-Buscalioni, R.; De Fabritiis, G.; Coveney, P. V. *J. Chem. Phys.* **2005**, *123*, 054105.
(43) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*, 2nd ed.; Academic Press: San Diego, California, 2001.
(44) Gore, J.; Ritort, F.; Bustamante, C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12564.

(45) Minh, D. D. L. *FERBE: Free Energy Reconstruction from Biased Experiments*, version 0.2; https://simtk.org/home/ferbe/. Accessed April 18, 2011.
(46) Townsley, L. E.; Tucker, W. A.; Sham, S.; Hinton, J. F. *Biochemistry* **2001**, *40*, 11676–11686.
(47) MacKerell, A. D.; Banavali, N.; Foloppe, N. *Biopolymers* **2000**, *56*, 257.
(48) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781.
(49) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
(50) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.
(51) Bastug, T.; Grayweale, A.; Patra, S.; Kuyucak, S. *Biophys. J.* **2006**, *90*, 2285.
(52) Corry, B.; Chung, S. *Eur. Biophys. J.* **2005**, *34*, 208.
(53) Harvey, M. J.; Giupponi, G.; De Fabritiis, G. *J. Chem. Theory Comput.* **2009**, *5*, 1632.
(54) Harvey, M. J.; De Fabritiis, G. *J. Chem. Theory Comput.* **2009**, *5*, 2371.
(55) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. *J. Comput. Chem.* **1999**, *20*, 786.
(56) Buch, I.; Harvey, M. J.; Giorgino, T.; Anderson, D. P.; Fabritiis, G. D. *J. Chem. Inf. Model.* **2010**, *50*, 397.
(57) Giorgino, T.; Harvey, M.; De Fabritiis, G. *Comput. Phys. Commun.* **2010**, *181*, 1402.
(58) Efron, B.; Tibshirani, R. *Stat. Sci.* **1986**, *1*, 54.
(59) Procacci, P.; Marsili, S.; Barducci, A.; Signorini, G. F.; Chelli, R. *J. Chem. Phys.* **2006**, *125*, 164101.
(60) Shapiro, S. S.; Wilk, M. B. *Biometrika* **1965**, *52*, 591.
(61) Allen, T. W.; Andersen, O. S.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 117.