JCTC Journal of Chemical Theory and Computation

Biased Molecular Simulations for Free-Energy Mapping: A Comparison on the KcsA Channel as a Test Case

Enrico Piccinini,*[,†,‡] Matteo Ceccarelli,[§] Fabio Affinito,[†,||,⊥] Rossella Brunetti,[†,||] and Carlo Jacoboni[†,||]

CNR-INFM National Research Center on nanoStructures and bioSystems at Surfaces (S³), Via Campi 213/A, I-41100 Modena, Italy, Dipartimento di Ingegneria Elettronica, Informatica e Sistemistica DEIS, Alma Mater Studiorum Università di Bologna, Viale Risorgimento 2, I-40136 Bologna, Italy, Dipartimento di Fisica and Sardinian Laboratory for Computational Materials Science - SLACS, Università di Cagliari, Cittadella Monserrato, I-09042 Monserrato (CA), Italy, and Dipartimento di Fisica, Università di Modena e Reggio Emilia, Via Campi 213/A, I-41100 Modena, Italy

Received July 31, 2007

Abstract: The calculation of free-energy landscapes in proteins is a challenge for modern numerical simulations. As to the case of potassium ion channels is concerned, it is particularly interesting because of the nanometric dimensions of the selectivity filter, where the complex electrostatics is highly relevant. The present study aims at comparing three different techniques used to bias molecular dynamics simulations, namely Umbrella Sampling, Steered Molecular Dynamics, and Metadynamics, never applied all together in the past to the same channel protein. Our test case is represented by potassium ions permeating the selectivity filter of the KcsA channel.

1. Introduction

Molecular Dynamics (MD) simulation is considered today the most powerful computational method to explore or interpret specific protein functions, provided that a high-resolution structure is known, and it has been widely applied to study specific features of single-ion translocations through nanometric membrane channels that underlie many important physiological features.[1]

The power of the method relies on the possibility of linking specific features of the permeation path with the peculiar interactions existing among the permeating ions and between each of them and the protein residues facing the permeation pathway.

The main limitations of the method are due to (a) the parametrization of the force field and its accuracy to take into account the strong electrostatic interaction between ions and proteins, (b) the computationally expensive very large number of atoms forming the simulated system, and (c) the way the complex electrostatics of the nanometric environment is tackled.

With reference to point (a) above, Allen et al. recently compared the most widely used biomolecular force fields using gramicidin A as a prototypical narrow ion channel showing that a polarizable reliable force field would introduce a significant enhancement of state-of-the-art results.[2]

The second problem listed above prevents the possibility to directly compare results from MD simulations with experimental results. It is in fact known that single-ion translocations require typical times ranging from 10 to 100 ns. With the present hardware and software computation tools only a few events can be observed with MD: they are useful

* Corresponding author phone: +39 059 205 5292; fax: +39 059 205 5616; e-mail: enrico.piccinini@unimore.it. Corresponding author address: CNR-INFM National Research Center on nanoStructures and bioSystems at Surfaces (S³), Via Campi 213/A, 41100 Modena, Italy.

† CNR-INFM National Research Center on nanoStructures and bioSystems at Surfaces (S³).

‡ Università di Bologna.

§ Università di Cagliari.

|| Università di Modena e Reggio Emilia.

⊥ Current address: International School of Advanced Studied (SISSA/ISAS), Via Beirut 2/4, 34014 Trieste, Italy.

10.1021/ct7001896 CCC: $40.75
Published on Web 12/11/2007

to study the full permeation pathway but still not enough for the straightforward simulation of a macroscopic ion flux. To fill this gap computational approaches able to calculate ion fluxes and including as much as possible the molecular information of the protein in the input parameters and in the model have been recently presented in the literature.[3−5] They all rely on more or less detailed information about the potential of mean force (PMF) of the system formed by the ions and the protein to identify the relevant occupation configurations involved during the permeation process and the probabilities associated with the transitions between them. The use of these "mesoscale" simulation procedures allows the linking of the atomistic description to the real functional properties of the proteins and promises to become in the future one of the investigation tools to be used for engineering protein functionality and fixing failures.

Mapping all the relevant structures of a complex PMF from MD simulations is a very difficult task, especially in view of the fact that an uncertainty of few k_BT in the evaluation of a free-energy barrier can be very relevant when the barrier height is used to estimate the transition probabilities in the simulation of the conduction process.[5] An accurate MD estimate can be obtained with the use of many reaction coordinates and long computer runs, which in some cases makes the calculation in practice impossible. Thus very relevant care is devoted to study when and how the simulation problem can be suitably simplified to find a reasonable trade-off between accuracy and resource demand.

For this purpose many computational techniques have been proposed and applied in the literature to artificially "bias" a simulation and force the time evolution of the system toward a given transition of interest, depending on the case at hand. Among them, we have chosen three methodologies relying on very different computational strategies, namely Steered MD[6] (SMD), the most widely used Umbrella Sampling[7] (US), and Metadynamics[8] (MetaD).

The main aim of this investigation is to establish the degree of reliability and the vulnerable aspects of these computational techniques on the basis of a common test on a nanometric channel. To this purpose we report both the evaluations of the free-energy profiles and the corresponding technique-dependent error. This critical analysis seldom accompanies this kind of calculations. Our test case is potassium ions permeating the selectivity filter of the bacterial potassium channel KcsA from *Streptomices lividans*.[9]

The KcsA structure is known from X-ray investigations, further confirmed by MD simulations. The potassium permeation of this channel takes place through a short and narrow region of the protein, called the "selectivity filter''. Seven stable binding sites have been identified, usually referred to as S_{ext}, S_0, ..., S_4, S_{cav} in which, alternatively, potassium ions and water molecules are found. The two outermost sites S_{ext} and S_0 were both first predicted in MD simulations[10] and subsequently observed in the higher resolution X-ray structure.[9] S_{ext} can fictitiously be made collapsing onto S_0 when the conduction properties of the channel are investigated because its position is diffuse and quite close to the bulk water phase. Two ions must always reside in the region S_1, ..., S_4 in a stable conductive situation, otherwise the protein changes its conformation and switches to a nonconductive state. The conduction process involves the simultaneous and concerted movement of ions in a single file, giving origin to a cycle of different occupancy configurations, as indicated in the sketch reported in Figure 1 (site S_{ext} is not represented). A proper free-energy barrier identifies each transition between different configurations. In this study we are interested in mapping the free-energy profile associated with internal transitions (i.e., transitions not involving new ion entries or exits), labeled as (t) or (d) in Figure 1.

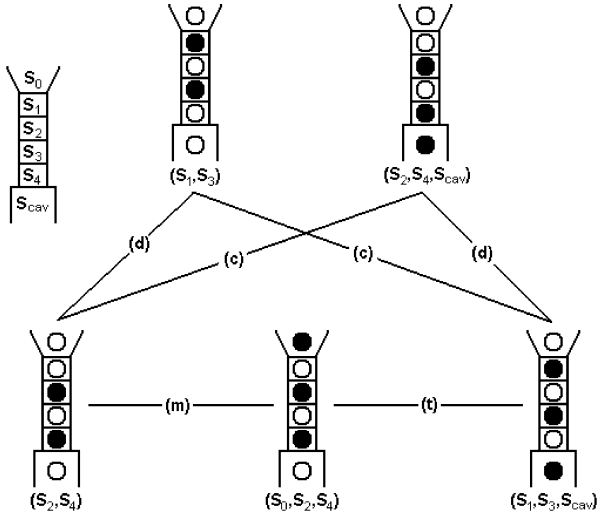

Figure 1. Configurations considered in the model and transitions among them. A sketch of the selectivity filter is reported on the left; site S_{ext} is located on top of site S_0, and it is not represented. Open circles stand for water molecules, and solid circles stand for potassium ions. Labels (m), (c), (d), and (t) mean an ion entry/exit into/from site S_0, for an exit/entry from/into the cavity site S_{cav}, for a two-ion concerted motion, and for a three-ion concerted motion, respectively.

The choice of the KcsA channel as the test case for our analysis is justified by the fact that in this highly selective nanometric channel ions move in single file along a pore which is roughly their size and strongly interact with each other and with the protein environment, thus producing physical challenging conditions for any molecular simulation.

Moreover, this system was deeply investigated in the past by means of the US technique, and many results to compare it with can be found in the literature.[3,10−12] To our knowledge MetaD was exploited in the past to study chlorine channels,[13] and it is here applied to the KcsA case for the first time.

2. Methods

2.1. Modeled System. Our starting point is the most recent X-ray structure of the KcsA solved at 2.0 Å resolution (PDB code 1K4C), inserted in a slab composed by 500 octane molecules mimicking the cell membrane. We solvated with 8802 water molecules, and 24 chlorine ions were used to keep the system electrically neutral. The final system, analogous to the one reported in the literature,[14] is composed of 34 434 atoms.

Calculations have been carried out with the GROMACS 3.3 package[15,16] for SMD and US simulations and with the ORAC code[17] for MetaD, using in all cases the standard GROMACS force field (also known as GROMOS 87) for the protein and SPC model for water. This combination of force fields showed its capability to keep the positions of the binding sites in the selectivity filter stable after equilibration time.

The system has been initially fully equilibrated for 1.2 ns using GROMACS, in which the first 200 ps were useful to heat the structure from 100 to 300 K. After equilibration the simulation box is approximately $6.9 \times 6.9 \times 9.7$ nm; then a further short equilibration (a few hundred picoseconds) was performed after the introduction of the biasing potential to obtain a starting configuration for each of the selected techniques and respective MD codes.

Electrostatic interactions have been computed using the smooth particle-mesh Ewald (PME) algorithm[18] implemented in the two codes with a fourth-order interpolation function, a grid of $72 \times 72 \times 100$ points (corresponding to a mesh of less than 0.1 nm wide), and a cutoff in the direct and reciprocal space of 1.2 nm and 0.042 nm^{-1}, respectively. Considering the nanometric width of the selectivity filter, we also tested higher-order functions. However, a finer treatment of the PME did not add any further contribution or improvement to the present results and does not justify the increased computational burden.

All simulations have been carried out in the NVT ensemble, using the Nose-Hoover temperature coupling (reference temperature 300 K, time constant between 3 and 5 ps). A time step of 1 fs was used in GROMACS runs; the r-RESPA algorithm[19] was used in ORAC in conjunction with the rattle-shake algorithm to fix covalent bonds involving hydrogen atoms. The time steps used in the RESPA algorithm were 0.5-1-2-4 and 12 fs.

2.2. Computational Strategies To Bias MD Simulations. In the following a short overview of the three techniques applied in our calculations to bias the system toward the transitions of interest is presented, with the purpose of pointing out the theoretical assumptions and the computational strategies, which can possibly produce qualitative and quantitative effects on the results. Details are also provided about the values of the parameters used in our simulations for the KcsA case. These parameters have been fixed in order to optimize the convergence of the methods used to study the physical system at hand, but the sensitivity of each technique on its parameter set is beyond the scope of this work and it was not tackled with detail.

In the following we use PMF as a synonym for free-energy profile as a function of a set of chosen coordinates, which are also called reaction coordinates.

2.2.1. Umbrella Sampling. The problem of calculating the PMF is tackled with US technique[20] by adding a fictitious term to the Hamiltonian $H(\mathbf{x},\mathbf{p})$ of the system under investigation

$$\tilde{H}(\mathbf{x},\mathbf{p}) = H(\mathbf{x},\mathbf{p}) + h_i(r) \tag{1}$$

where $\mathbf{x}$ and $\mathbf{p}$ are the positions and the momenta of all the atoms of the system of interest.

This term is a static harmonic biasing potential, function of a chosen reaction coordinate $r = r(\mathbf{x})$, and center around at a given position r_i:

$$h_i(r) = \frac{1}{2} k(r - r_i)^2 \tag{2}$$

$h_i(r)$ is used to restrain the reaction coordinate r in the neighborhood of r_i, thus enhancing sampling of that region of the configuration space.

The center position r_i of the biasing potential is varied step by step along a defined path to obtain a set of N partially overlapping windows, each of them providing an ion probability distribution function $\rho_i(r)$. These distributions are then combined together to give the unbiased PMF by means of the weighted histogram analysis method (WHAM).[21]

Following the scheme suggested by Souaille and Roux,[22] the unbiased total distribution probability $\rho^u(r)$ is defined as

$$\rho^u(r) = C \sum_{i=1}^{N} \frac{n_i \exp[-\beta(h_i(r) - f_i)]}{\sum_{j=1}^{N} n_j \exp[-\beta(h_j(r) - f_j)]} \rho_i(r) \tag{3}$$

where $\beta = 1/k_B T$, C can play the role of a normalization constant, n_i is the weight of the simulation of the ith window, i.e., the number of configuration samples used to compute $\rho_i(r)$, and coefficients f_i, coming from the adding of the biasing potential, are calculated by an iterative solution of the formula

$$\exp(-\beta f_i) = C \int \mathrm{d}r \sum_{k=1}^{N} \frac{n_k \exp(-\beta h_i(r))}{\sum_{j=1}^{N} n_j \exp(-\beta(h_j(r) - f_k))} \rho_i(r) \tag{4}$$

The PMF is finally calculated by means of a generalization of the reversible work theorem[23]

$$G(r) = G(r_0) - k_B T \ln \frac{\rho^u(r)}{\rho_0^u(r_0)} \tag{5}$$

where r_0 is a reference point.

If more than one reaction coordinate is used in the simulation, the overall biasing potential is given by a sum of terms of the type reported in eq 2; in this case the center positions of the biasing potentials are varied on a grid on a multidimensional surface.

2.2.2. Steered Molecular Dynamics. SMD as well makes use of an added biasing potential to force the system to visit high free-energy regions. Contrary to the US technique, where the system is driven to the region identified by r_i (eq 2) and there statically remains until the following window is considered, in this case the added potential is continuously varied in time until the system reaches its ending configuration.

Therefore eq 1 is still valid, but in this case the new Hamiltonian is also a function of time that appears explicitly in the expression of the new biasing potential

$$h(\mathbf{x}, t) = \frac{1}{2} k[r(\mathbf{x}) - (\mu_0 + \upsilon t)]^2 \tag{6}$$

where μ_0 is the initial center position of the restraining potential, and v is the pulling velocity. This framework resembles atomic-force microscope experiments where a molecule is pulled between two positions, being subject to a time-varying external force.

The evaluation of the PMF relies on the Jarzynski's identity.[24] This equality links the equilibrium free-energy differences ΔG between the states A and B to the work W done on the system through all the nonequilibrium processes leading it from A to B. According to the second law of thermodynamics, ΔG represents the lower limit of $\langle W \rangle$: $\langle W \rangle \geq \Delta G$, the equality being valid in the limit of quasi-static (or equilibrium) processes. $\langle\rangle$ denotes an ensemble average. Jarzynski, however, proved that the following equality holds true regardless of the speed of the process:

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta G} \tag{7}$$

The general validity of eq 7 depends on a small number of trajectories where $W_i \leq \Delta G$. The probability of these events decreases exponentially as the speed of the process increases, thus a large number of simulations is needed to handle a reliable statistical ensemble with even a relatively high pulling velocity. In practice, despite its theoretical speed-free validity, the applicability of this equation is limited to slow processes, whose energy fluctuations are comparable to $k_B T$, and a number of trajectories have to be combined together to obtain significant results.

By means of eq 7, one gets the free-energy of the system described by $\tilde{H}$ that must be corrected by subtracting the term due to the perturbing potential. Following the procedure described by Hummer and Szabo,[25] one finally gets the expression for the free-energy $G(r)$, as a function of the chosen reaction coordinate

$$G(r) = -\frac{1}{\beta} \ln \frac{\sum_t \frac{\langle \delta(r - r_t) \exp(-\beta w_t) \rangle}{\langle \exp(-\beta w_t) \rangle}}{\sum_t \frac{\exp[-\beta h(r,t)]}{\langle \exp(-\beta w_t) \rangle}} \tag{8}$$

where the two summations are over time steps t, $h(r,t)$ is the perturbing potential defined in eq 6, and w_t is the work done on the system until time t. Note that for this kind of simulation a single reaction coordinate is used, so that a one-dimensional analysis is performed.

An improvement of SMD results can in principle be obtained by the Crooks equation[26] that makes use of both the forward and the backward average work and extends the Jarzynski identity. Whenever the hypotheses of the transient fluctuation theorem are satisfied, the works of the forward and backward transitions can be mixed together to give a better estimate of the PMF profile.

2.2.3. Metadynamics. MetaD[8,27−29] is a recently introduced technique based on the idea of the complexity reduction, being able to speed up the evolution of some defined reaction coordinates $r_k(\mathbf{x})$ with the introduction of a "history-dependent" biasing potential $V(r_k,t)$. The latter is the sum of repulsive functions that are added at given time steps during the simulation in order to constitute a "penalty" term for configurations already visited in the space of the reaction coordinates. These repulsive functions fill the minima of the PMF and, after a long simulation, tend to compensate exactly the underlying PMF that, in turn, can be approximated by their sum. If Gaussians are used as repulsive functions for the potential, then eq 1 rewrites

$$\tilde{H}(\mathbf{x}, \mathbf{p}, t) = H(\mathbf{x}, \mathbf{p}) + \sum_i w_i \exp\left[-\sum_{k=1}^{N} \frac{(r_k(\mathbf{x}, t) - r_k(\mathbf{x}_i, t_i))^2}{\Delta r_k} \right] \tag{9}$$

where w_i and Δr_k are the height of the repulsive potential and the scale factor for the kth coordinate, respectively. The outer summation (index i) is over time steps. The scale factor defines the range of action of the repulsive potential and represents a sort of resolution of the reconstructed PMF.

The main advantage of MetaD with respect to US is that it is not required to define a priori the range of variation of the reaction coordinates, letting the system evolve toward the lowest transition state, thus obtaining the minimum free-energy landscape along the path connecting the two minima. This prevents the sampling of uninteresting regions, and, in principle, it allows the introduction of a high number of reaction coordinates. Similar approaches have previously been exploited to explore the configuration space, such as the taboo search,[30] and the local elevation method.[31] Moreover, MetaD can also be considered as an extension of the Wang-Landau algorithm,[32] and it is closely related to the recent adaptive force-bias algorithm,[33,34] where the derivative of the free-energy along a reaction coordinate is reconstructed by means of an adaptive time-dependent bias.

However, the use of a time-dependent biasing potential is in some way a nonequilibrium procedure with respect to the other degrees of freedom, especially to the so-called slow modes. When the latter are not included in the chosen set of reaction coordinates, the choice of the parameters controlling the repulsive potentials (i.e., deposition time step, height and scale factor) is crucial to let the system equilibrate each time a new term is added. The efficient sampling of nonexplicit slow modes within MetaD can be tackled in different ways, either improving the sampling with the replica exchange method,[35] or by means of the bias-exchange metadynamics,[36] or correcting the reconstructed PMF with a subsequent refining US.[37]

2.3. Choice of the Reaction Coordinates. A computational mapping of the free-energy profile for potassium ions in the KcsA protein was already done in the past[3,10] by means of multidimensional US. In that case, the authors used a number of occupancy configurations corresponding to the case of two ions within the selectivity filter, independently varying the position of these two ions together with that of a third ion in the cavity. More than 300 simulations were needed (leading to an aggregate total simulation time of 36 ns) which, in turn, implied a significant computation time. The obtained free-energy maps confirmed that the conduction process takes place as several consecutive steps in which, if the two ions within the filter move, they always move concertedly. This picture of the permeation process suggests the possibility of using a single curvilinear coordinate to describe ion motion within the US framework, thus avoiding

to explore paths along which the system will never evolve due to excessively high energy. A further confirmation of what above stated can be deduced by analyzing the PMF plots reported in the reference work by Bernèche and Roux.[10] It can be observed that the preferred pathways can be split into segments where only one of the selected coordinates moves significantly, the others being confined in a narrow well centered on their initial value. The identification of a curvilinear coordinate to reduce the amount of CPU time is possible, though not trivial at all because it must not force unphysical movements of the ions.

The SMD technique seems to be suitable for a similar procedure. Under physiological conditions the ion flux through the channel is driven by a transmembrane potential, resulting from a charge imbalance between intra- and extracellular environments. As a consequence a potential that changes along the channel axis should exist, and the reaction coordinate should follow this observation. A natural choice of reaction coordinate is thus the z value of an ion's coordinate in a Cartesian orthogonal reference system, also taking into account the strong ion confinement in the xy plane.

Three simulations have been performed, using the position of the top outermost ion (initially labeled as K_2), the position of the middle ion (K_4), or the position of their center of mass (K_{com}) as reaction coordinates, respectively. We would like to point out again that this technique implicitly accounts for one-dimensional analysis, bounding the position of one ion (or group of ions) to the pulling spring and, in practice, forcing in this way the ions' motion. An analogous choice of reaction coordinates has also been performed for the US simulations, for comparison purposes.

The case of MetaD requires a different approach. The chance to sample *at the same time* different reaction coordinates together with the use of a history-dependent potential term automatically drives the dynamics of the system along the minimum energy path, thus avoiding the exploration of undesired regions, where one can suppose a priori that the system will hardly pass through. The reader easily understands that the choice of a suitable minimal set of coordinates is far from being a trivial point. In order not to waste time, coordinates must be independent from each other and represent a minimum set able to describe the evolution of the system under investigation, including the slow-modes. Preliminary investigations focused on the definition of the appropriate coordinates are often needed. In this case we found that the most effective set of coordinates is represented by the positions of the ion in the cavity K_{cav} and of the middle ion in selectivity filter K_4. It should also be noticed that a fair comparison of the US and/or SMD sampling with MetaD under equivalent conditions implies to project the n-D free-energy profile from MetaD along the minimum-energy path, i.e., as a function of a single curvilinear coordinate making use, for instance, of the nudged elastic band (NEB) method.[38] A second benefit of MetaD is represented by the possibility to introduce coordinates not linked to the physical position of the ions, e.g., the water coordination number of the ion in the cavity. It is likely to suppose that the hydration/dehydration process affecting the ion in the cavity may be significant in the permeation process as much as the position of the ions in the selectivity filter. This concept has been underlined in different channels by Gervasio et al.[39] and Braun-Sand et al.[40]

2.4. Parameters and Computational Details. One of the major points concerning the comparison among different techniques aiming at the same result is the proper choice of simulation parameters for each technique to avoid that one technique outperforms the others only because of an unfair set of parameters.

The state of the art of US simulations for the KcsA channel is reported by Bernèche and Roux,[10] where the system was investigated with great detail, and it is used as a guide in the following. We used a force constant $k = 8368$ kJ/mol nm^2 (20 kcal/mol Å^{-2}) and a step between two consecutive centers of the biasing potential (r_i) $\Delta r = 0.05$ nm to respect the overlapping constraint. Up to 18 steps, 515 ps-long each, have been combined together, and the biased distributions have been reconstructed by 250 bins ranging from $3 \cdot 10^{-3}$ to $3.8 \cdot 10^{-3}$ nm in width, depending on the chosen reaction coordinate, as described in the next subsection.

The first 15 ps of each simulation were used to adjust the biasing potential to the new position, starting from the previous configuration. Then the following 250 ps were discarded as equilibration time in the presence of the biasing potential. The remaining 250 ps were then split into 5 blocks of 50 ps, and, finally, the results were averaged. The reaction coordinate was recorded at every time step, for a total productive sampling roughly about 2 million configurations per reaction coordinate. By this protocol it was possible to calculate the statistical error of the estimated PMF, which results to be about $\pm 1\ k_B T$ at room temperature (corresponding to 2.5 kJ/mol).

In the absence of previously published studies on the KcsA channel with SMD, we performed several preliminary tests to determine both a suitable force constant and a pulling velocity. In a work of Jensen et al. on the permeation of glycerol through aquaglyceroporin GlpF[41] the SMD technique was intensively used to compute the energetics, and the influence of its parameters on the final result was also revised with details.

Following that suggested scheme, a harmonic constraint with a spring constant $k = 1673.6$ kJ/mol nm^2 (4 kcal/mol Å^{-2}) was attached to the selected reaction coordinate. This latter constant ensures a thermal fluctuation of the constrained coordinate of about $\sqrt{k_B T/k} = 0.04$ nm and a corresponding force fluctuation of approximately 100 pN. We determined that a pulling velocity $v = 1 \cdot 10^{-3}$ nm/ps is slow enough to guarantee that the system always evolves through intermediate quasi-equilibrium states. This was further confirmed by performing reverse transitions at a double steering velocity and by observing that they converge to the same value of the free-energy barrier (see also Figure 3).

A preliminary run keeping the perturbing potential fixed was performed. The first 150 ps were discarded, then system configurations were saved every 50 ps, as input starting configuration for subsequent productive runs. The 50 ps interval ensures that saved configurations are uncorrelated. Each productive run lasted 500 ps, and the reaction coordi-

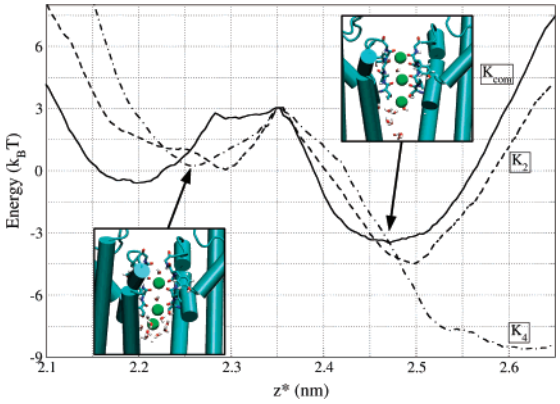

Figure 2. PMFs from US for the transition (S_2, S_4, S_{cav}) ↔ (S_1, S_3, S_{cav}). Solid, dashed, and dash-dotted lines refer to different reaction coordinates (see framed labels); abscissas have been shifted in order to allow direct comparison of the results. The maximum has been used as the pivotal point; the zero-level of the free energy is at arbitrary position. The two insets represent the configurations corresponding to the two minima.

nate was saved at each time step, for a total simulation time of 4.5 ns and 4 million sampled points. The PMF profile was reconstructed by means of eq 8 adapting the weighted histogram method[21] with 250 bins combining eight uncorrelated trajectories together. That proved to be enough for convergence. The statistical uncertainty on the PMF was also investigated by averaging 7 blocks of 8 trajectories, thus obtaining an error of $\pm 2\ k_BT$ (5 kJ/mol) on the energy scale and ± 0.05 nm for the position of the minimum.

MetaD simulations were performed by means of two reaction coordinates, namely the positions of two ions in the filter-cavity region along the channel axis. We adopted the following protocol for the simulations: 2 kJ/mol-high hills were added every 4 ps, and the scale factor of reaction coordinates was set in order not to exceed the thermal fluctuations of the two coordinates in absence of any bias. Values of 0.02 and 0.015 nm looked appropriate for the ion in the cavity and the ions inside the filter, respectively. The latter value is less than the former due to their reduced mobility, as also stated in ref 14. A previous metadynamics study on the motion of ions inside chloride channels[13] guided this choice and actually ensures that the error on the reconstructed PMF is of the order of 2 k_BT (5 kJ/mol). The overall MetaD simulation lasted approximately 9 ns.

3. Results and Discussion

MD free-energy results usually depend both on the adopted biasing methodology and on the particular force field used in the simulations. An exhaustive comparison of different force-field models applied to gramicidin A as a test case can be found in the recent literature.[2]

The focus of our calculations is on the comparison among the three different biasing techniques illustrated in section 2 with the use of the GROMOS87 force field. It was recently proved that GROMOS*nn* force fields introduce a systematic overestimation of the energy barriers,[2] due to the peculiar set of electric charges included in their parametrization of the electrostatic interaction. This evidence, however, does not affect the validity of our conclusions because it equally influences all of the results under comparison. The obtained energy profiles fairly agree with those reported in the literature.[10]

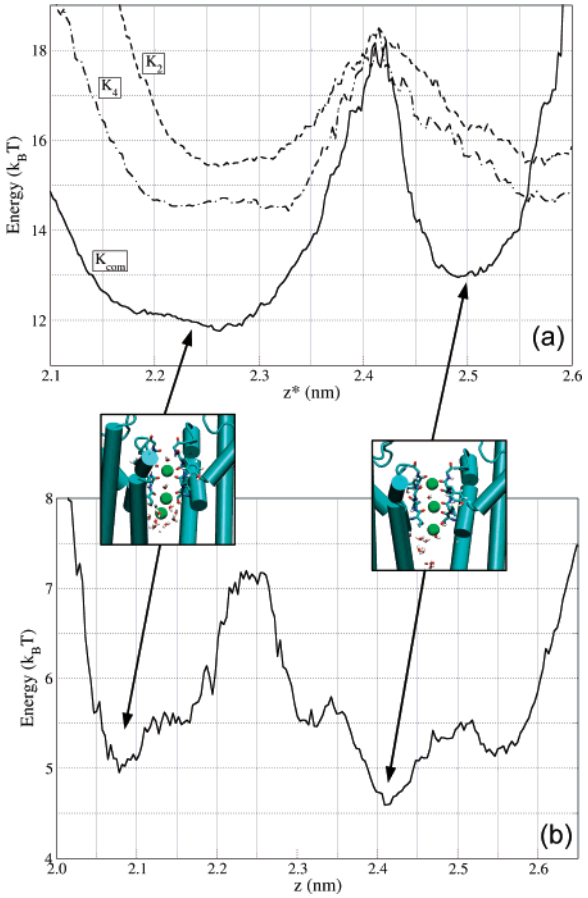

Figure 3. (a) (top) PMFs from SMD for the transition (S_2, S_4, S_{cav}) → (S_1, S_3, S_{cav}). Solid, dashed, and dash-dotted lines refer to different reaction coordinates (see framed labels); abscissas have been shifted in order to allow direct comparison of the results. The maximum has been used as the pivotal point; the zero-level of the free energy is at arbitrary position. The two insets represent the filter configurations corresponding to the two minima. (b) (bottom) PMF for the reverse transition (see text).

Furthermore, we point out that energy values are given using k_BT units. This choice is convenient since the results are used within the framework of the reaction-rate theory, with the purpose of studying conduction properties.[5]

3.1. US vs SMD. PMFs from US and SMD for transition (S_2, S_4, S_{cav}) → (S_1, S_3, S_{cav}) are reported in Figures 2 and 3a, respectively.

For comparison purposes, the curves within each figure corresponding to different reaction coordinates have been shifted using the free-energy relative maximum as the pivotal point for both *x*- and *y*-axes. Spatial differences in the position of minima and maxima obtained with US and SMD are limited within 0.1 nm or less, and they are attributed both to rigid shifts of the filter structure with respect to the internal reference during the simulation, and to thermal fluctuations. They do not affect either the PMF determination or the comparison between the two methods. Data are shown

in the range 2.1−2.65 nm for US and 2.1−2.6 nm for SMD to include the abscissas corresponding to the initial and final configurations. The free-energy zero-level is at an arbitrary position in the two cases presented, which again does not influence the evolution of the energy difference separating the configurations.

Two minima are found, as expected, since the two corresponding configurations (S_2, S_4, S_{cav}) and (S_1, S_3, S_{cav}) of Figure 1 have been already identified as stable configurations of the selectivity filter.[10,11,42] All of the curves obtained are qualitatively similar, even though it should be remarked that the minimum of the final state in the US calculation is deeper than the one found with SMD. In this latter case, the second minimum is approximately as deep as the starting point.

In both plots two of the three curves (K_{com} and K_2 in US, K_2 and K_4 in SMD) are quite similar in shape and in numerical values, with differences less than 1 k_BT that can be easily associated with the statistical variance; nonetheless, the third curves (K_4 in US and K_{com} in SMD), though confirming a similar overall shape, present significantly different numerical values.

These differences are not related to computational uncertainties but instead to spatial fluctuations attributed to variations of the ion position within the cavity and to torsions of the residues facing the selectivity filter, as already observed in the past.[12] These changes, which take place on the time scale of our MD simulations, give origin to variations of the energy barriers larger than the computational uncertainty of MD but still small with respect to barriers separating conducting from nonconducting states. For this reason we have considered the previous results as a more likely estimate of the barrier height provided by the two techniques considered in this paragraph.

The interpretation of the results coming from the application of the SMD technique requires some care. With reference to Figure 3a, the method provides a reliable estimate of the energy barrier associated with the transition out of the minimum of the PMF, which is the starting point of the pulled atoms. After barrier crossing the initial spring length is not fully recovered, i.e., some elongation is still present. This fact produces an artificial overestimate of the energy level associated with the final state of the pulled transition. To get a correct estimate, the reverse transition, namely (S_1, S_3, S_{cav}) → (S_2, S_4, S_{cav}), has been simulated starting from an initially fully relaxed spring and pulling the ion K_3 toward site S_4. The results of this simulation, reported in Figure 3b, compare with the curve labeled K_4 in Figure 3a. Here the overestimate due to the pulling action of the spring does not allow a correct sampling of the minimum located at 2.07 nm. By analyzing the two results together we can conclude that, within the numerical uncertainty associated with the SMD method, the energy barriers associated with forward and reverse transitions result in being equal, in contrast to what is observed with US. From the US runs we obtain an estimate of the barrier of about 3 k_BT for the forward transition and of about 6 k_BT for the reverse transition (except curve K_4, where a 9 k_BT barrier is found), which means a deeper second minimum.

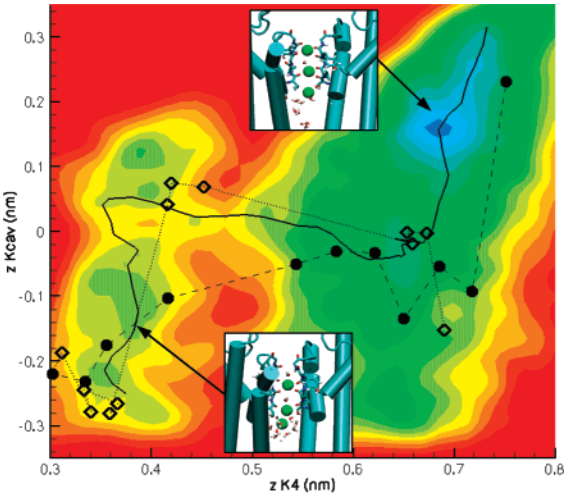

Figure 4. PMF from MtD for the transition (S_2, S_4, S_{cav}) ↔ (S_1, S_3, S_{cav}); each color level corresponds to an energy of 1 k_BT. The full line represents the minimum-energy path within MetaD; the dots and the diamonds are snapshots taken from trajectories followed by the curvilinear coordinate in US and SMD runs, respectively. The dotted and dashed lines are drawn to guide the eye and not as real paths. The two insets illustrate the position of the ions in the selectivity filter corresponding to the two main minima.

The comparison of the above results with those obtained with MetaD helps to justify this discrepancy.

Some enhancements may occur, and a more precise PMF profile can be achieved by means of the Crooks equation. However for the present case the error affecting SMD calculations is fair enough to let us infer that the strong qualitative difference existing among these profiles and those obtained by US and MetaD can only be attributed to the different number of sampled coordinates. The main disadvantage of SMD over US and MetaD is thus represented by the intrinsic one-dimensional behavior of eq 7 that leads to a unique coordinate analysis. When a multiplicity of coordinates is required, as the present case looks like, the SMD picture is clearly too poor.

3.2. MetaD vs US. MetaD results for the energy landscape involved in the transition (S_2, S_4, S_{cav}) → (S_1, S_3, S_{cav}) are reported in Figure 4 as functions of the position of ion K_4 and of ion K_{cav}. A 3 k_BT barrier is estimated by all of the selected techniques, and the general trend of a deeper minimum for the (S_1, S_3, S_{cav}) configuration, as obtained from US, is also confirmed. It is worth noticing that the two minima do not correspond to the same position of the ion in the cavity: when the latter moves closer to the near vacant site inside the selectivity filter the free-energy surface shows a deeper minimum.

To better link the results obtained with the three different methods we can analyze the trajectories followed by the moving ions in US and SMD runs and the time needed to observe the transition. Values reported in Figure 4 (and in Figure 6) must be interpreted as the time-averaged position of the K_4 and K_{cav} ions within US windows (K_4 trajectory) and as snapshots of a representative run in the SMD case. The standard deviation of ion positions range from 0.013 to 0.021 nm for K_4 and from 0.04 to 0.13 nm for K_{cav} as US is

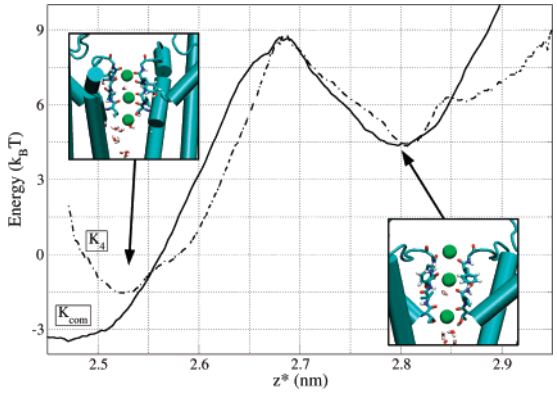

Figure 5. PMFs from US for the transition (S_1, S_3, S_{cav}) ↔ (S_0, S_2, S_4). Solid and dash-dotted lines refer to different reaction coordinates (see framed labels); abscissas have been shifted in order to allow direct comparison of the results. The maximum has been used as the pivotal point; the zero-level of the free energy is at arbitrary position. The two insets represent the configurations corresponding to the two minima.

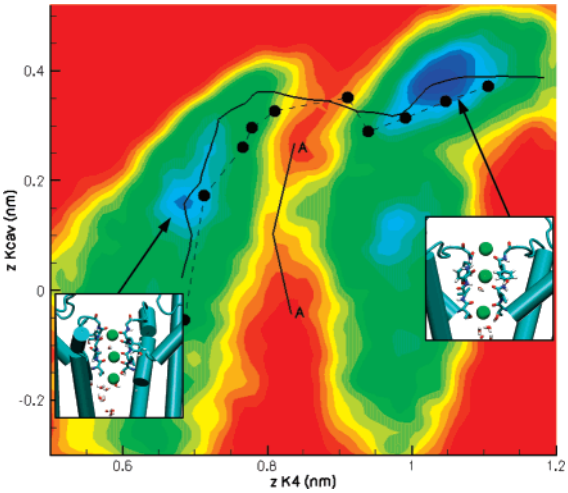

Figure 6. PMF from MtD for the transition (S_1, S_3, S_{cav}) ↔ (S_0, S_2, S_4); each color level corresponds to an energy of 1 k_BT. The full line represents the minimum-energy path within MetaD; the dots are snapshots taken from trajectories followed by the curvilinear coordinate in US (the dashed lines are drawn only to guide the eyes and do not correspond to a physical path). Line A−A represents a wall separating the two regions, the saddle point in that position is due to an unavoidable artifact of the interpolation surface. The two insets illustrate the configurations corresponding to the two minima.

concerned and approximately 0.05 nm (K_4) and 0.1 nm (K_{cav}) in the case of SMD. At the beginning, the ions reside in their original positions, which correspond to the first minimum in the energy plot; then, as we start pulling one of the ions in the selectivity filter with SMD, their energy is increased, and finally the barrier is crossed. Meanwhile, the ion in the cavity cannot rise up to the neighboring site in the selectivity filter until the transition takes place, and it is observed to strongly oscillate within the cavity. The configurations with the ion in the cavity and the ion close to site S_4 are in fact almost isoenergetic (Figure 4). After the transition, some time is needed for the ion in the cavity to stabilize and enter the selectivity filter; this, in turn, lets the system evolve toward its new free-energy minimum. The full translocation sequence can, in principle, be observed with US, but not with SMD, since the continuous pulling action in practice does not allow the system to reach the final equilibrium. A similar explanation can also be given for the deeper minimum of curve K_4 of US analysis (Figure 2): both curves labeled K_{com} and K_2 correspond to simulations where the ion in the cavity is not substantially moving. According to this interpretation, the role of ion K_4 is crucial for the conduction process, because it influences directly the position of the other ions in the cavity and in the selectivity filter.

To further confirm the hypothesis above, we have also investigated the transition (S_1, S_3, S_{cav}) → (S_0, S_2, S_4) with US and MetaD. SMD was not used any longer because it proved not to be able to correctly identify the position of the minimum corresponding to the (S_1, S_3, S_{cav}) configuration in the previous analyzed transition. For the situation at hand, it must be remarked that, if the selectivity filter is not populated by two ions, the protein undergoes a significant conformational change leading to a nonconductive state.[9] For this reason we always have to consider reaction coordinates directly or indirectly involving the intermediate ion in the filter. If we used the outermost ion position as the only reaction coordinate, we would have run the risk of emptying the selectivity filter and, thus, driving the system to a nonconductive state. On the other hand, the ion in the cavity can only fill site S_4: in the final configuration of the previous analyzed transition this ion resided quite close to the filter's mouth. Figures 5 and 6 show the results for US and MetaD, respectively. A general agreement from both a qualitative and a quantitative point of view is found between the two techniques: the path followed with 1-D US coordinate is the minimum energy path also identified within the MetaD framework, and the exit barriers do not differ very much from each other (10−12 k_BT for US, approximately 13 k_BT for MetaD). The MetaD analysis reveals two minima corresponding to transition (S_3 → S_2), depending on the final position of the ion originally in the cavity, which can either reside close to or fully enter site S_4. The most stable one is represented by the latter case, since it corresponds to a deeper minimum. The minimum-energy path calculated with the NEB method and reported in Figure 6 shows that the transition happens when the ion in the cavity enters site S_4 and not when it is adjacent to it, even if this corresponds to a slightly higher barrier. This interpretation is further confirmed by the fact that in the transition (S_1, S_3, S_{cav}) → (S_0, S_2, S_4) potassium ions can occupy two adjacent sites for short periods of time, breaking the rule of concerted motion. In particular this sequence of events, reported in Figure 7, takes place: the ion originally located in S_{cav} enters the site S_4, being S_3 filled; then, the water molecule in S_2 exchanges its position with the ion in S_3, leading to (S_1, S_2, S_4) configuration. The latter configuration is quite unstable (top of the barrier) and evolves to (S_0, S_2, S_4), which represents the final state of this transition. These intermediate three-ion states, in which potassium ions occupy two adjacent binding sites, have already been reported in the literature,[3] and they have recently been observed also in the analysis of the permeation paths of homologous channel Kv1.2.[43]

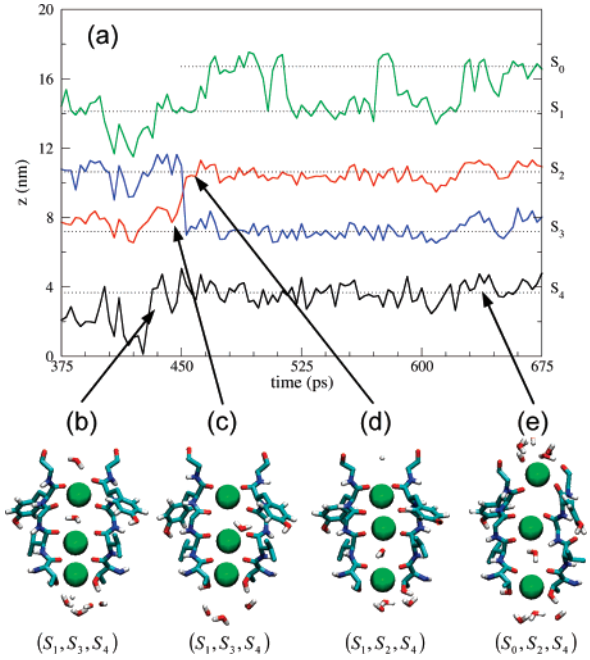

Figure 7. Trajectories of the potassium ions (green, red, and black lines) and of the water molecule (blue line) in the selectivity filter, projected onto the symmetry axis of the channel, during transition (S_1, S_3, S_{cav}) → (S_0, S_2, S_4) (a) and significant MD snapshots of this transition (b)−(e). For clarity purposes, only two subunits are represented. The positions of the binding sites are represented by dotted lines. Data are recorded every 3 ps. At around 450 ps the water molecule and the ion initially in S_3 exchange their position within the imposed time step: snapshots (c) and (d) suggest that the water molecule first moves off-axis close to the backbone, facilitating the ion to rise up into site S_2, and then slips down into its final position. A fast flip of a carbonyl oxygen of Val76, separating sites S_2 and S_3, is also possible on a shorter time scale. A similar reorientation is reported also in snapshot (e), when a hydrogen bond between the carbonyl oxygen of Val76 and the protein backbone behind it (not represented in the figure) is established.

Furthermore, we did not observe a substantial rearrangement of the selectivity filter in connection with the water−potassium exchange between two subsequent simulation snapshots (10 ps); however, we cannot exclude that this may take place on a shorter time scale. The overall PMF calculated via MetaD and projected onto the minimum-energy path is reported in Figure 8.

When the (S_0, S_2, S_4) configuration is reached, we have observed a symmetry breaking, also reported in a previous work by Bernèche and Roux.[44] As it is shown in Figure 7e, the amide plane Val76-Gly77 undergoes a 180° reorientation, while the backbone carbonyl oxygen of Val76 points away from the conduction path. As a consequence the water molecule in site S_3 forms a hydrogen bond with the hydrogen of Gly77. At the same time the plane of the aromatic ring of Tyr78 bends down, getting closer to the reoriented carbonyl oxygen of Val76, but it keeps its own specific mobility preserved.

In conclusion, both MetaD and US allow the easy use of many reaction coordinates, even though with different

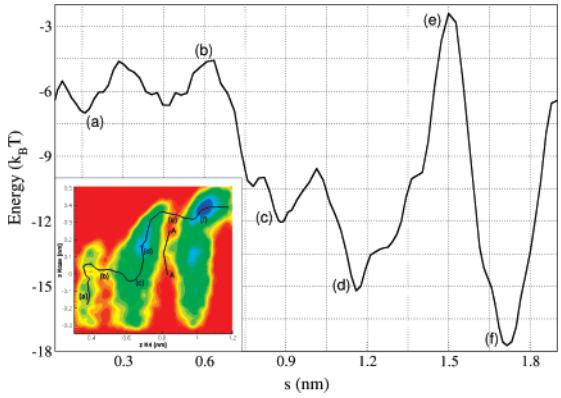

Figure 8. MetaD PMF projected onto the minimum-energy path by means of the NEB method. Data are presented as a function of a curvilinear coordinate, and labels are used to identify corresponding points in the free-energy landscape. In the inset the whole PMF landscape and trajectory are also reported.

sampling strategies: MetaD performs a simultaneous sampling of the whole set of reaction coordinates, while within the US framework only one coordinate is varied at a time, the others being temporarily kept fixed as parameters. As a consequence, MetaD is less resource-demanding than US and provides results faster. In our case the 2D plot obtained with MetaD in Figures 4 and 6 would have required up to ten times the simulation time used, if calculated using US.

US, however, provides a more accurate sampling than MetaD when the same set of coordinates is considered, in particular when the PMF exhibits many competing pathways in the explored area. In our case, for example, Figure 6 can be compared with Figure 2 on the left in the reference work by Bernèche and Roux.[10] A general qualitative (and, to a less extent, even quantitative) agreement between the two PMFs is found, but some differences exist. The free-energy path explored by MetaD corresponds to the major path found by US, but the secondary path identified by US has not been mapped by MetaD. This secondary path can be associated with either a slightly higher energy or slow modes not sampled by the chosen set of reaction coordinates. It should be noticed, however, that the region corresponding to the (S_0, S_2, S_{cav}) configuration is visited after that (S_0, S_2, S_4) is reached, thus warning about the possible existence of a competing path. The saddle point linking (S_1, S_3, S_{cav}) to (S_0, S_2, S_{cav}) is an artifact of the interpolating surface, as the NEB analysis testifies. Last, it must also be pointed out that the (S_1, S_3, S_4) configuration corresponds to an intermediate state along the transition path close to the energy maximum. The associated secondary minimum reported in the reference work by Bernèche and Roux is not revealed by the MetaD simulation.

Both techniques seem to be successful and accurate enough (within their own constraints and limitations) to describe the permeation process of an ion in a narrow pore. A statement about which of the two outperforms the other strictly depends also on the system under investigation, and, for this reason, the choice of MetaD instead of US (or vice versa) must be done after a careful evaluation of the trade-off level between computational efficiency and sampling accuracy.

4. Conclusion

A compared analysis of three different numerical techniques (Umbrella Sampling, Steered Molecular Dynamics, and Metadynamics) aimed at the reconstruction of the free-energy landscape via molecular dynamics has been presented. As a case study, two key transitions of the permeation process in the KcsA channel have been chosen.

The obtained results suggest few conclusive statements, which can be considered general and applicable also to other permeation cases in nanometric pores.

All of the three techniques represent computational tools able to reconstruct the PMF profile composed by a number of valleys and hills of different height. From a qualitative point of view the identification of the minimum energy path connecting two consecutive valleys, i.e., finding the lower barrier existing between two ion occupancy configurations, is a straightforward activity that can be performed with an average error of few k_BT. Barriers giving origin to an effective permeation path must, in fact, be smaller than the ones existing among configurations not involved in the conduction. However, the statistical errors allow only approximate quantitative estimate of transport properties, such as, for instance, the ionic current. One more critical point is represented by the choice of the modeling force field that can introduce a further systematic source of uncertainty.

With regards to the three selected techniques, SMD proved to be the less suitable one. In fact, this technique was originally developed for studying the stretching and/or unfolding of proteins and, then, adapted to investigate free-energy landscapes. This technique provided poor results for the present case: when the PMF cannot be described by means of a unique physical reaction coordinate, SMD may be inaccurate because the continuous pulling of the reaction coordinate prevents the other coordinates from reaching values leading to the global energy minimum.

The US technique is the straightforward, most used, and most accurate technique for this kind of analyses. However it demands intensive computational efforts when many coordinates are used. MetaD is faster in achieving results, once an appropriate set of reaction coordinates is chosen. Moreover, it also allows the use of nonintuitive reaction coordinates not directly related to the spatial coordinates of the permeating ions (e.g., ion hydration). For this reason it can be preferred as computationally advantageous. With respect to US, MetaD tends to drive the simulation through the main paths of the free-energy profile, whereas US provides a description of the whole landscapes, investigating also secondary pathways that could be accessed via MetaD only by increasing the number of variables and, consequently, the computational burden. Furthermore, MetaD provides an upper limit to barriers, because it may happen that some slow modes are not explicitly taken into account. On the other hand, US provides a lower limit, due to the fact that the transition state itself is always considered an equilibrium distribution by the computational procedure. The combined use of MetaD and US, the former being able to scope out the dominant reaction coordinates and the latter to refine results, represents a good suggestion to achieve accurate results with an affordable computational cost.

Acknowledgment. We gratefully acknowledge the CINECA National Supercomputing Center for providing CPU time and the CASPUR Computer Center (Rome) for additional CPU time allocation. The Italian Ministry for University and Research (MIUR) supported funding through project PON-Cybersar.

References

(1) Roux, B.; Allen, T. W.; Bernèche, S.; Im, W. *Quart. Rev. Biophys.* **2004**, *37*, 15.

(2) Allen, T. W.; Andersen, O. S.; Roux, B. *Biophys. J.* **2006**, *90*, 3447.

(3) Bernèche, S.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 8644.

(4) Mapes, E. J.; Schumaker, M. F. *Bull. Math. Biol.* **2006**, *68*, 1429.

(5) Piccinini, E.; Affinito, F.; Brunetti, R.; Jacoboni, C.; Rudan, M. *J. Chem. Theory Comput.* **2007**, *3*, 248.

(6) Gullingsrud, J. R.; Braun, R.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 190.

(7) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275.

(8) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562.

(9) Zhou, Y.; Morais-Cabral, J.; Kaufman, A.; MacKinnon, R. *Nature* **2001**, *414*, 43.

(10) Bernèche, S.; Roux, B. *Nature* **2001**, *414*, 73.

(11) Aqvist, J.; Luzhkov, V. *Nature* **2002**, *404*, 881.

(12) Bernèche, S.; Roux, B. *Biophys. J.* **2002**, *78*, 2900.

(13) Gervasio, F. L.; Parrinello, M.; Ceccarelli, M.; Klein, M. L. *J. Mol. Biol.* **2006**, *361*, 390.

(14) Compoint, M.; Carloni, P.; Ramseyer, C.; Girardet, C. *Biochim. Biophys. Acta* **2004**, *1661*, 26.

(15) Lindhal, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306.

(16) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43.

(17) Procacci, P.; Paci, E.; Darden, T. A.; Marchi, M. *J. Comput. Chem.* **1997**, *18*, 1848.

(18) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T. A.; Lee, H.; Pedersen, L. *J. Chem. Phys.* **1995**, *103*, 8577.

(19) Tuckerman, M. E.; Berne, B. J.; Matryna, G. J. *J. Chem. Phys.* **1992**, *97*, 1990.

(20) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.

(21) Kumar, S.; Bouzida, D.; Swensen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.

(22) Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001**, *135*, 40.

(23) Chandler D. Statistical Fluids. In *Introduction to modern statistical mechanics*; Oxford University Press: Oxford, New York, 1987; pp 188−233.

(24) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690.

(25) Hummer, G.; Szabo, A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3658.

(26) Crooks, G. E. *J. Stat. Phys.* **1998**, *90*, 1481.

(27) Iannuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 238302-1.

(28) Ceccarelli, M.; Danelon, C.; Laio, A.; Parrinello, M. *Biophys. J.* **2004**, *87*, 58.

(29) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. *J. Phys. Chem. B* **2005**, *109*, 6714.

(30) Cvijovic, D.; Klinowski, J. *Science* **1995**, *267*, 664.

(31) Huber, T.; Horda, A.E.; van Gunsteren, W. F. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695.

(32) Wang, F.; Landau, D. P. *Phys. Rev. Lett.* **2001**, *86*, 2050.

(33) Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *108*, 1964.

(34) Henin, J.; Chipot, C. *J. Chem. Phys.* **2004**, *121*, 2904.

(35) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435.

(36) Piana, S.; Laio, A. *J. Phys. Chem. B* **2007**, *111*, 4553.

(37) Babin, V.; Roland, C.; Darden, T. A.; Sagui, C. *J. Chem. Phys.* **2006**, *125*, 204909.

(38) Jónsson, H.; Mills, G.; Jacobsen, K. W. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; Berne, B. J., Ciccotti G., Coker, D. F., Eds.; World Scientific: Singapore, 1998; pp 385−403.

(39) Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 2600.

(40) Braun-Sand, S.; Burykin, A.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **2005**, *109*, 583.

(41) Jensen, M. Ø.; Park, S.; Tajkhorshid, E.; Schulten, K. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6731.

(42) Morais-Cabral, J. H.; Zhou, Y.; MacKinnon, R. *Nature* **2001**, *414*, 37.

(43) Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. *Biophys. J.* **2006**, *91*, L72.

(44) Bernèche, S.; Roux, B. *Structure* **2005**, *13*, 591.

CT7001896