International Journal of Modern Physics C

Multiscale modelling of permeation through membrane channels using pregenerated molecular dynamics trajectories

G. De Fabritiis

Computational Biochemistry and Biophysics Lab (GRIB-IMIM/UPF), Barcelona Biomedical Research Park (PRBB), C/ Dr Aiguader 88, 08003 Barcelona, Spain
gdefabritiis@imim.es

J. Villà-Freixa

Computational Biochemistry and Biophysics Lab (GRIB-IMIM/UPF), Barcelona Biomedical Research Park (PRBB), C/ Dr Aiguader 88, 08003 Barcelona, Spain
jvilla@imim.es

P. V. Coveney

Centre for Computational Science, Department of Chemistry, University College London 20 Gordon street, London WC1H 0AJ, United Kingdom
p.v.coveney@ucl.ac.uk

Permeation of small molecules across membrane channels can be measured by a multiscale computational protocol based on Brownian dynamics and the potential of mean force formalism. In this article we look at ways to compute the potential of mean force by reusing pre-existing molecular dynamics trajectories via a protocol centered on instantaneous forward/reverse transformations. We apply the method to the energetics of water across the narrow channel formed by Gramicidin A and reproduce several features of the energy barrier across the channel albeit at a coarse level of detail due to limits imposed by the exponential averages intrinsic to the method and the small size of the channel. The implications for larger channels are briefly discussed.

Keywords: Multiscale modelling; forward-reverse transformations; potential of mean force

pacs nos.:

1. Introduction

The permeation of small molecules through trans-membrane proteins represents a major challenge for current molecular simulations. For ions, this quantity can be directly observed experimentally by measuring the conductivity at given electrostatic potential differences between the inner and outer parts of the membrane and is important for the normal functioning of the cell [1,2]. Computationally, permeability cannot be obtained directly by brute force calculations because the time scales involved are too long (of the order of milliseconds). Rather, typically a multiscale approach is used by splitting the problem into two fundamental components: Brow-

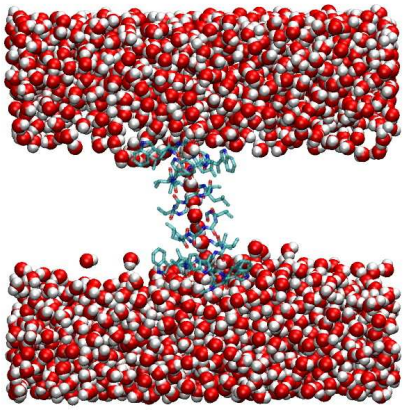

Fig. 1. The Gramicidin A channel solvated in a DMPC membrane with TIP3P water and 150mM KCl. A single file of water molecules is present inside the channel. Lipids are not shown for clarity.

nian permeation along the z direction (orthogonal to the membrane plane) and an effective potential of the averaged degrees of freedom of the atoms of the channel itself, the lipid membrane and the solvation water [3,4]. The choice of the effective potential can be made in several ways including the evaluation of the solvation free energy of the moving molecule through the channel by semimacroscopic techniques [4] or by mean of the potential of mean force (PMF) through the channel[5]. This effective potential is then used in combination with Brownian dynamics simulations or escape time calculations in order to recover the permeation time [6]. In this methodology, the evaluation of the effective potential is computationally the most critical and expensive part of the calculation.

The aim of the present paper is to describe a computational protocol to compute the PMF for water [7] using pre-generated molecular dynamics (MD) trajectories based on instantaneous insertions and removals of molecules. The calculations are performed on a simple membrane channel, Gramicidin A, a helical antibacterial dimer used to increase the permeability of biological membranes to inorganic ions. The channel formed by Gramicidin A is one of the best characterised both biochemically and structurally [7] and for this reason it has been a test case in a number of computer simulations [8,9,10,7,11,12]. One characteristic of this molecule is that it forms a narrow channel, only allowing a single file of water molecules (or potassium ions) to fill it (see figure 1). We restrict the analysis to the study of the permeability of a single water molecule in order to demonstrate the method, leaving the case of potassium ion permeability for future work [13].

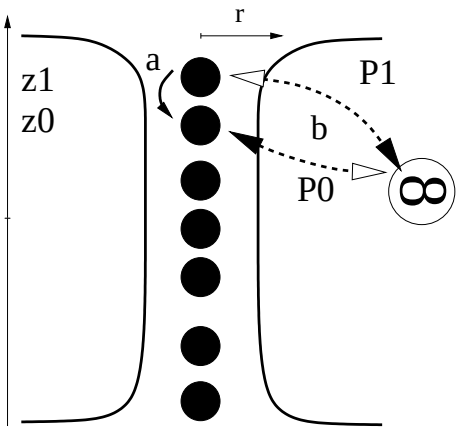

Fig. 2. The potential of mean force (PMF) of water buried inside the channel along the z direction is computed via a thermodynamic cycle. The PMF $W(z_1) - W(z_0)$ to move a water molecule from position z_0 to z_1 (a) is computed using path b: A water molecule is removed from the system (equivalent to moving it to an ideal gas condition, no longer interacting) or inserted with the inverse transformation.

2. Methods

We define a reaction coordinate $\xi = \xi(\mathbf{R}^N)$ and the probability density of encountering ξ given by the ensemble average

$$\rho(\xi) = \frac{\int d\mathbf{R}^N d\mathbf{P}_R^N \delta(\xi(\mathbf{R}) - \xi) \exp(-\beta H)}{\int d\mathbf{R}^N d\mathbf{P}_R^N \exp(-\beta H)}, \tag{1}$$

which also allow us to compute the free energy difference $\Delta A_{\xi_b - \xi_a} = A(N, V, T, \xi_b) - A(N, V, T, \xi_a)$ between two states $\xi = \xi_b$ and $\xi = \xi_a$ in terms of $\rho(\xi)$ as

$$\Delta A_{\xi_b - \xi_a} = -kT \log \frac{\rho(\xi_b)}{\rho(\xi_a)}. \tag{2}$$

The quantity in Eq.(2) is also called the potential of mean force (PMF), that is the free energy along a given generalised coordinate (reaction coordinate) ξ and written, up to an arbitrary constant, as $W(\xi) = -k_B T \log \rho(\xi)$.

A direct way to compute the PMF is to measure the density ρ in Eq. (2) over a long MD trajectory in order to accumulate enough statistics. Water molecules populate the channel for long enough to get a good estimate of the PMF which we shall use as a reference to compare with our calculations. In the general case, e.g. ions, this procedure is bound to produce high errors when ξ is scarcely populated because $\rho(\xi)$ is close to zero where the logarithm function is singular. Because of this, the PMF is usually computed with the help of a bias potential (umbrella sampling)[14] in order to improve the sampling along the coordinate. This procedure requires running simulations explicitly designed for the task [12].

Instead, here, we aim to use pre-existing molecular dynamics trajectories [15,16,17] in order to determine the PMF by simple analysis of such data. We restrict our-

selves to the case where the reaction coordinate ξ is just a projection over a simple coordinate such as $\xi = \xi(\mathbf{R}) = \mathbf{r}_{n+1}$, where $\mathbf{R} = \mathbf{r}^{n+1}\mathbf{X}$. For the Gramicidin A channel (Fig. 1) $\mathbf{r}_i$ are the coordinates of the n water molecules inside the channel and $\mathbf{X}$ represents the other degrees of freedom including lipids, protein ions and other water molecules outside the protein. We also define two systems characterised by the potentials $U_1(\mathbf{r}^{n+1}, \mathbf{X}^m)$ and $U_0(\mathbf{r}^n, \mathbf{X}^m)$ where the reaction coordinate is not interacting with the rest of the system any longer. Our protocol is focused on the thermodynamic cycle described in Fig. (2). The PMF along the coordinate z between the coordinate z_0 and z_1 indicated by (a) in Fig. (2) can be computed following the free energy difference in path (b): the water PMF is computed by free energy perturbation methods corresponding to the transformations from system 1 to 0 and *vice-versa*, i.e. moving a molecule at the coordinate z_0 to an ideal gas condition (no interaction and infinite distance) and from the ideal gas to z_1.

2.1. *Free energy perturbation methods*

Following Powles *et al.* [18], we can rewrite the ensemble average of Eq. 1 in terms of a one-dimensional probability distribution of the energy difference $u = U_0 - U_1$

$$P_i(u) =< \delta(u - [U_0(\mathbf{R}^N) - U_1(\mathbf{R}^N)]) >_i, \tag{3}$$

where U_0 and U_1 represent the potential energy of the two reference states. This gives us a very general formula for the free energy difference, first derived by Shing and Gubbins [19]

$$P_0(u)\exp(-\beta(A_0 - A_1)) = P_1(u)\exp(-\beta u), \tag{4}$$

which implies that the free energy difference $A_1 - A_0$ between two equilibrium thermodynamic states "0" and "1" can be written in terms of the probability distributions of the energy difference $u = U_1 - U_0$ between the two states upon forward ($0 \to 1$) and reverse ($1 \to 0$) transformations. The simplicity of this formula is striking: the free energy difference is the value of the energy at which the probabilities of the system to move from one state to the other upon forward and reverse transformations are equal.

In practice, the estimation of the free energy difference is not performed by using directly Eq. 4 but rather in combination with the acceptance ratio method [20] which is directly derivable from Eq. 4, and is designed to minimise the variance of the estimator. The Bennett formula reads

$$\exp\left(-\beta(A_0 - A_1)\right) = \frac{< F(-\beta[u - C]) >_1}{< F(\beta[u - C]) >_0} \exp(\beta C), \tag{5}$$

where C is an arbitrary constant and $F(x) = 1/(1 + \exp(x))$ is the Fermi function. The value of C providing the minimum variance is $C = A_0 - A_1$, which is what we want to calculate. Therefore, the Bennett formula is often used in an iterative manner where C is determined self-consistently by starting from an initial value (for instance 0) and iterating it until Eq. 5 is satisfied up to a given tolerance.

2.2. *Sampling forward and reverse distributions*

The formula in Eq. 5 requires us to sample the forward energy distribution $P_0(u)$ for passing from system '0' with energy U_0 to system '1' with energy U_1 and the reverse energy distribution for the reverse transformation $P_1(u)$. In the case of Gramicidin, the forward probability distribution consists of sampling, by test particle insertion, the energy landscape of the system, i.e. computing the energy gained by the system by random insertions of a water molecule within the channel. Indeed, the configurational space of the channel is so small that the potential energy landscape underlying P_0 can be probed by brute force with many random test insertions.

A good estimation of the reverse distribution P_1 requires a good sampling of the $\{\mathbf{r^n}\}$ configurations according to the canonical distribution $\exp(-\beta U(\mathbf{R}^N))$, where $\mathbf{R}^N = \mathbf{r}^n\mathbf{X}^m$. Ideally, this task is simply performed by simulating an MD trajectory of the molecular system for long enough and saving the energy of each individual water molecule inside the channel in order to estimate $P_1(u)$.

For almost the same computational cost of running the MD trajectory, it is possible to obtain better statistics of P_1 by mixing MD and Monte Carlo moves. The number of degrees of freedom $\mathbf{X}^m$ that we are averaging out into the potential of mean force is usually very large, hence a lot of computational time is spent simulating the thermal bath far away from the channel rather than collecting statistics for the coordinates $\{\mathbf{r}^n\}$ which will make a larger individual contribution to the energy u.

In more detail, for any configuration $\{\mathbf{X}^m\}$ of the bath, we generate several configurations $\{\mathbf{r}^n\}$ which suit equally well the canonical distribution and improve the estimation of the histogram. Given an MD trajectory $\{\mathbf{R}^N\}$ sampling the canonical distribution, each point of this trajectory can be taken as a starting point of a new Markov chain following a different realisation. To implement this, the potential energy of the entire molecular system can be divided into several parts according to the decomposition in orthogonal coordinates $\mathbf{R}^N = \mathbf{r}^n\mathbf{X}^m$,

$$U(\mathbf{r}^n, \mathbf{X}^m) = U(\mathbf{r}^n) + U(\mathbf{r}^n; \mathbf{X}^m) + U(\mathbf{X}^m), \tag{6}$$

and from any single configuration $(\mathbf{r}; \mathbf{X}^m)$ we start a Metropolis Monte Carlo [21] scheme with potential $U' = U(\mathbf{r}; \mathbf{r}) + U(\mathbf{r}; \mathbf{X})$ while leaving the bath coordinate completely frozen. The second term $U(\mathbf{r}; \mathbf{X})$ can be seen as an external force field due the bath degrees of freedom. For each reference position of the bath, we are able to generate several configurations of molecules inside the channel, improving the estimation of P_1.

The boot-strapped estimation of $P_1(u)$ is performed using a standard Metropolis MC algorithm [21] for a rigid water molecule with maximum trial displacement of the centre of mass $\delta r = 0.3$ Å and maximum random rotation of 30° producing an acceptance ratio for the MC moves of approximately 0.3. For every configuration of the trajectory file, each water molecule inside the channel is subjected to 300 MC moves sampling new positions and energies every 5 moves. During each MC move

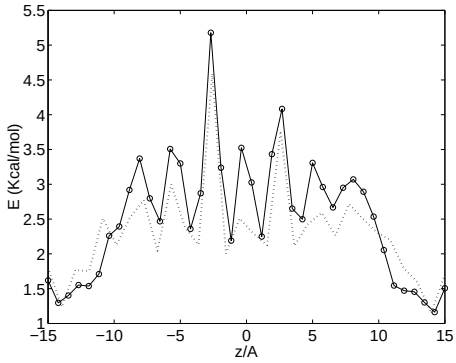

Fig. 3. Potential of mean force computed using the bootstrap method to generate configurations and the density to estimate the free energy $A = -k_b T \log(\rho)$. The PMF calculated from the trajectory data alone is also shown (dotted line).

the rest of the system is kept frozen. With these settings, the data is augmented by around two orders of magnitude for P_1.

3. Results and discussion

From the starting structure (PDB:1JNO) the system was progressively equilibrated [13] up to generate a 10 ns production run. From the analysis of this long run, we estimate the channel size in cylindrical coordinates to be 22 Å x r = 2.5 Å. We also define n_t as the number of water molecules within 22 Å of the channel. The system naturally fluctuates between a water molecule occupation number of 7 or 8 with very short excursions to 6 or 9 molecules in the channel. Instead of simulating two different systems with potential energies defined by $U_0 = U(\mathbf{r}_1, ..., \mathbf{r}_7, \mathbf{X})$ and $U_1 = U(\mathbf{r}_1, ..., \mathbf{r}_8, \mathbf{X})$, we simply split the trajectory configurations into two different trajectories according to the number of water molecules in the channel. For each frame of the MD trajectory, the occupancy number n_t is computed and is used to split the trajectory frames into two subsets: configurations where n_t is less or equals 7 water molecules and configurations with more than 8 water molecules. This procedure separates system 0 and 1 from the equilibration run without having to relay on forcing a fixed number of molecules into the channel with a flat-bottom potential.

The PMF computed directly with Eq. 2 is shown in Fig. 3 and also compared with bootstrapped estimations. The position of the seven stable water molecules within the channel is clearly visible in both cases. Overall, we find a mean energy barrier of around 2 Kcal/mol with a peak of 5 Kcal/mol.

We now move to the calculation of the PMF using the Bennett method by sampling P_0. The forward distribution is sampled by randomly inserting a water molecule onto a regular lattice with 0.5 Å spacing spanning the entire cylinder of the channel. This generates 450 energy values u for each frame. We then apply the

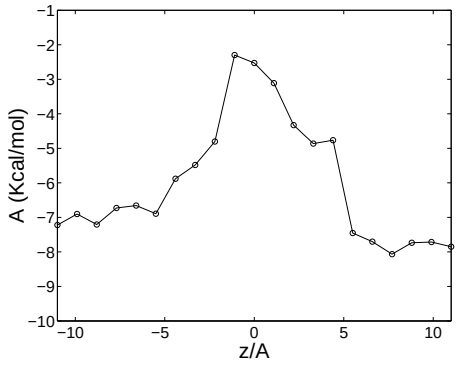

Fig. 4. The potential of mean force for water inside the channel and along the z coordinate given by the free energy perturbation method. Note that the values of the free energy at both extremes of the channel tends to approach the excess chemical potential for bulk water.

Bennett iterative formula (5) to the forward and reverse distributions and obtain the free energy profile shown in Fig. 4.

A direct comparison of results from the Bennett method in Fig. 4 and the control profile of Fig. 3 shows that the first cannot reproduce the same level of detail. The problem lies in the fact that the average over P_0 in Eq. 5 has to be computed on larger volumes along the channel axis in order to converge. This is due to the fact that the forward and reverse distributions do not overlap owing to the narrow size and packing of the channel which produces a profile with variations within one Angstrom (see Fig. 3). Overlapping is well known to be one the most important aspects determining the accuracy of free energy perturbation approaches due to the exponential average intrinsic to formula (5). Nevertheless, an approximate barrier of around $5 - 6$ Kcal/mol can be recovered.

4. Conclusions

We performed a set of calculations for the potential of mean force of water along the Gramicidin A channel. We designed a protocol that can be performed on an MD trajectory of a molecular system allowing use of previously generated repositories of such trajectories[15,16,17]. However, the fact that the trajectory data is augmented by the test particle insertion protocol does not seem to overcome the fact that in a narrow channel like Gramicidin A the energy distributions P_i have very little overlap. This is a common problem of free energy perturbation methods [22] due to the exponential averages required to merge information from forward and reverse paths with the Shing and Gubbins formula. Although this problem limits the resolution of the free energy barrier in figure 4, we find the total energy barrier of around 5Kcal/mol is still correctly captured. Other transmembrane proteins with larger channels should improve the amount of overlap between forward and reverse distributions and produce more accurate free energy profiles.

Acknowledgments

We thank Barcelona Supercomputing Center (BSC) for access to MareNostrum, HPC-Europa for a visiting grant given to GDF and EPSRC for funding the Integrative Biology project (GR/S72023). This work was supported in part by BQU2003-04448 grant from the Spanish MCyT. We thank M. Johnston for carefully reading the manuscript.

References

1. J. Aqvist and V. Luzhkov, Nature **404**, 881 (2000).
2. S. Berneche and B. Roux, Nature **414**, 73 (2001).
3. B. Roux, Biophys. J. **77**, 139 (1999).
4. A. Burykin, C. N. Schutz, J. Villa, and A. Warshel, Proteins **47**, 265 (2002).
5. B. Roux, Comp. Phys. Commun. **91**, 275 (1995).
6. D. Lua, P. Grayson, and K. Schulten, **85**, 2977 (2003).
7. B. L. De Groot, D. P. Tieleman, P. Pohl, and H. Grubmuller, Biophys. J. **82**, 2934 (2002).
8. J. Aqvist and A. Warshel, Biophys. J. **56**, 171 (1989).
9. B. Roux and M. Karplus, Annu. Rev. Biophys. Biomol. Struct. **23**, 731 (1994).
10. S.-W. Chiu, S. Subramaniam, and E. Jakobsson, Biophys. J. **76**, 1929 (1999).
11. S. Braun-Sand, A. Burykin, Z. T. Chu, and A. Warshel, J. Phys. Chem. B **109**, 583 (2005).
12. T. W. Allen, O. S. Andersen, and B. Roux, Biophys. J. **90**, 3447 (2006).
13. G. De Fabritiis, P. V. Coveney, and J. Villa-Freixa, preprint (2006).
14. D. Frenkel and B. Smit, *Understanding molecular simulation* (Accademic Press, 1996).
15. `http://www.dynome.org`.
16. `http://www.biosimgrid.org`.
17. M. Johnston and J. Villa-Freixa, in *Distributed, High-Performance and Grid Computing in Computational Biology* (Springer, 2006), in press Lecture Notes in Bioinformatics.
18. J. G. Powles and W. A. B. Evans, Mol. Phys. **46**, 1347 (1982).
19. K. S. Shing and K. E. Gubbins, Mol. Phys. **46**, 1109 (1982).
20. C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).
21. M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids* (Oxford University Press, Oxford, 1987).
22. N. Lu, J. K. Singh, and D. A. Kofke, J. Chem. Phys. **118**, 2977 (2003).