Received: 14 September 2018 | Revised: 18 December 2018 | Accepted: 22 December 2018

DOI: 10.1111/cbdd.13485

SPECIAL ISSUE ARTICLE

WILEY

Lessons learned about steered molecular dynamics simulations and free energy calculations

Fernando Martín Boubeta[1,2] | **Rocío María Contestín García**[1]* | **Ezequiel Norberto Lorenzo**[1]* | **Leonardo Boechi**[3] | **Dario Estrin**[1,2] | **Mariela Sued**[3] | **Mehrnoosh Arrar**[1,2]

[1]CONICET-Facultad de Ciencias Exactas y Naturales, Instituto de Química-Física de los Materiales, Medio Ambiente y Energía, Universidad de Buenos Aires, Buenos Aires, Argentina

[2]Departamento de Química Inorgánica, Analítica y Química Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

[3]CONICET-Facultad de Ciencias Exactas y Naturales, Instituto de Cálculo, Universidad de Buenos Aires, Buenos Aires, Argentina

Correspondence
Mehrnoosh Arrar, CONICET-Facultad de Ciencias Exactas y Naturales, Instituto de Química-Física de los Materiales, Medio Ambiente y Energía, Universidad de Buenos Aires, Buenos Aires, Argentina.
Email: mehrnoosh.arrar@qi.fcen.uba.ar

Funding information
Consejo Nacional de Investigaciones Científicas y Técnicas; Fundación Bunge y Born

Abstract

The calculation of free energy profiles is central in understanding differential enzymatic activity, for instance, involving chemical reactions that require QM-MM tools, ligand migration, and conformational rearrangements that can be modeled using classical potentials. The use of steered molecular dynamics (sMD) together with the Jarzynski equality is a popular approach in calculating free energy profiles. Here, we first briefly review the application of the Jarzynski equality to sMD simulations, then revisit the so-called stiff-spring approximation and the consequent expectation of Gaussian work distributions and, finally, reiterate the practical utility of the second-order cumulant expansion, as it coincides with the parametric maximum-likelihood estimator in this scenario. We illustrate this procedure using simulations of CO, both in aqueous solution and in a carbon nanotube as a model system for biologically relevant nanoheterogeneous environments. We conclude the use of the second-order cumulant expansion permits the use of faster pulling velocities in sMD simulations, without introducing bias due to large dispersion in the non-equilibrium work distribution.

KEYWORDS

free energy, Jarzynski, maximum likelihood, steered molecular dynamics

1 | INTRODUCTION

The change in free energy governs all chemical processes. The calculation of free energy profiles, that is, the change in free energy along a transformation or reaction coordinate, has thus been critical in understanding many complex processes such as protein–protein recognition (Gohlke, Kiel, & Case, 2003; Israelewitz, Baudry, Gullingsrud, Kosztin, & Schulten, 2001), drug design (Israelewitz et al., 2001; Shirts, Mobley, & Brown, 2010), ligand binding kinetics (Boechi et al., 2008; Boubeta, Bari, Estrin, & Boechi, 2016; Hu, Xu, & Wang, 2015; Selvam, Wereszczynski, & Tikhonova, 2012; Xiong, Crespo, Marti, Estrin, & Roitberg, 2006), large-scale conformational reorganization in proteins (Bringas, Petruk, Estrin, Capece, & Martí, 2017; Israelewitz, Gao, & Schulten, 2001; Marsico et al., 2018), and in the elucidation of chemical reaction mechanisms in enzyme active sites (Bieza et al., 2014; Crespo, Martí, Estrin, & Roitberg, 2005).

Here, we will specifically address the use of steered molecular dynamics (sMD) simulations and the Jarzynski equality to calculate a free energy profile along a particular transformation coordinate λ. We note that for all intents and purposes

*Contributed equally to this work.

here the terms reaction and transformation coordinate will be used interchangeably. Though we will not provide an in-depth review of the theory, we will need to introduce a few key equations in order to set the notation.

Namely, the free energy change between two states, in which $\lambda = \lambda_0$ and $\lambda = \lambda_1$, can be defined as:

$$\Delta F = \Delta F_{\lambda_1} - \Delta F_{\lambda_0} = -\frac{1}{\beta} \ln\left(\frac{Z_{\lambda_0}}{Z_{\lambda_1}}\right), \tag{1}$$

where Z is the partition function for the system with the indexed value of the reaction coordinate, that is, $Z_{\lambda_1} = \int \partial \boldsymbol{q} \exp[-\beta H(\boldsymbol{q}, \lambda = \lambda_1)]$, and the Hamiltonian is a function of the positions of all atoms of the system $\boldsymbol{q}$ and the reaction coordinate λ.

In 1997, Jarzynski showed that this free energy change between two states can be exactly related to the ensemble average of the Boltzmann-weighted work performed in many non-equilibrium transformations from the initial to final states (Jarzynski, 1997):

$$\Delta F = -\frac{1}{\beta} \ln\langle e^{-\beta W}\rangle, \tag{2}$$

where $\beta = 1/k_B T$, with k_B and T being the Boltzmann constant and temperature, respectively, and the work W represents the external work performed in changing the reaction coordinate λ from λ_0 to λ_1 in some time-dependent manner. The Jarzynski equality only requires that the initial configurations be taken from the equilibrium ensemble with $\lambda = \lambda_0$, and we note that the final states with $\lambda = \lambda_1$ do not belong to an equilibrium ensemble.

In sMD simulations (e.g., as implemented in Amber or NAMD simulation suites), the reaction coordinate is not usually constrained to a particular value, rather it is *restrained* to a particular center (generally according to a harmonic potential of force constant k) that changes according to a constant velocity v during a simulation of the irreversible transformation from λ_0 to λ_1:

$$V(\boldsymbol{q}, t) = \frac{1}{2} k[\lambda(t) - \lambda'(\boldsymbol{q})]^2, \tag{3}$$

where we use the prime notation to indicate the evaluation of the reaction coordinate for a given microscopic state $\boldsymbol{q}$ of the system, as opposed to the time-dependent value of the reaction coordinate center, which is $\lambda(t) = \lambda_0 + vt$.

Even though this implementation introduces an additional biasing potential to the system, the underlying potential of mean force (PMF) $G(\lambda)$ can still be recovered using Jarzynski's equality and re-weighting estimators analogous to the WHAM approach (Bartels & Karplus, 1997; Ferrenberg & Swendsen, 1989; Kumar, Rosenberg, Bouzida, Swendsen, & Kollman, 1992; Shirts & Chodera, 2008). Alternatively, if the force constant k is sufficiently large (the so-called "stiff-spring approximation"), such that $\lambda'(\boldsymbol{q}) \approx \lambda(t)$, $G(\lambda)$ of the system without the presence of a biasing potential can be approximated through the following relation (Park & Schulten, 2004):

$$G[\lambda = \lambda(t)] \approx \Delta F(t) - \frac{1}{2kv^2}\left(\frac{\Delta \ddot{F}(t)}{\beta} - \ddot{F}(t)^2\right) \tag{4}$$

We note that in general, if k is large, the higher order terms in Equation 4 become small enough to disregard.

Conventionally, when invoking the Jarzynski equality, ΔF is estimated with the so-called Jarzynski estimator:

$$\Delta \hat{F}_J = -\frac{1}{\beta} \ln\left(\frac{1}{N}\sum_{i=1}^{N} e^{-\beta W_i}\right) \tag{5}$$

Nevertheless, this estimator can be grossly inaccurate as it is dominated by poorly sampled low-work values, particularly if the variance of the work distribution is greater than $(k_B T)^2$ (Arrar et al., 2018; Hummer, 2001, 2007; Jarzynski, 2006; Pohorille, Jarzynski, & Chipot, 2010; Yunger Halpern & Jarzynski, 2016). Many have characterized this bias in terms of parameters of the underlying work distribution; Kofke and co-workers have proposed a heuristic to evaluate whether the free energy estimate is biased or not (Wu & Kofke, 2005, 2005); Zuckermann and colleagues have proposed extrapolation methods to correct for this bias (Bucher, Walker, & McCammon, 2014; Echeverria & Amzel, 2012; Ytreberg & Zuckerman, 2004; Zuckerman & Woolf, 2001, 2002). Additionally, methods have been proposed to specifically limit the variance of the work distribution itself (Chelli, 2012; Nicolini, Frezzato, & Chelli, 2011; Ozer, Valeev, Quirk, & Hernandez, 2010; Ramírez, Zeida, Jara, Roitberg, & Martí, 2014; Schmiedl & Seifert, 2007; Vaikuntanathan & Jarzynski, 2008; Zerbetto, Piserchia, & Frezzato, 2014).

Interestingly, several have shown that by satisfying the stiff-spring approximation, and with sufficiently fast relaxations of the system, the work distribution obtained from an sMD simulation is anticipated to be Gaussian (Gore, Ritort, & Bustamante, 2003; Jarzynski, 2011; Park & Schulten, 2004). In this scenario, the second-order cumulant expansion of the right-hand side of Equation 2 is an exact relation:

$$\Delta F = \mu - \frac{\beta}{2}\sigma^2. \tag{6}$$

Furthermore, we have highlighted in previous work (Arrar et al., 2018) that by estimating the parameters μ and σ^2 from a particular sample of N work values using the maximum-likelihood estimators $\bar{W}$ and $\hat{\sigma}_W^2$:

$$\bar{W}_N = \frac{1}{N}\sum_{i=1}^{N} W_i \quad \text{and} \quad \hat{\sigma}_W^2 = \frac{1}{N}\sum_{i=1}^{N}(W_i - \bar{W})^2 \qquad (7)$$

we obtain the parametric maximum-likelihood estimator $\Delta\hat{F}_{\mathrm{ML}}$, which, unlike $\Delta\hat{F}_{\mathrm{J}}$, is an unbiased estimator that performs well, even with large variance in the Gaussian work distribution. Nevertheless, there are many examples in practice in which the non-equilibrium work distributions are not Gaussian (Chipot & Pohorille, 2007), and the second-order cumulant expansion in general must be used with caution, to avoid possibly large truncation errors. Here, we evaluate the impact of different spring force constants and velocities on the accuracy of free energies estimated with $\Delta\hat{F}_{\mathrm{ML}}$, under the assumption of a Gaussian work distribution, and $\Delta\hat{F}_{\mathrm{J}}$. We consider systems for which the overall change in free energy is known (and equal to zero) and illustrate that the regime in which such an assumption is valid is limited to minimal values of k that satisfy the stiff-spring approximation, whereas artifacts are introduced for higher values of k, an effect that is particularly pronounced when combined with low pulling velocities.

2 | METHODS

2.1 | Test systems

The two systems considered here are (a) a single molecule of carbon monoxide (CO), (b) a single molecule of CO and a carbon nanotube of radius 7 Å. Both systems were solvated in a box of TIP3P water molecules (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983). The Carbon/Boron Nitride Nanostructure Builder Plugin in VMD (Humphrey, Dalke, & Schulten, 1996) was used to generate the carbon nanotube structure, to which no charges were assigned. CO parameters were taken from previous work. All topology files (AMBER (Case et al., 2010) format) are available as Supplementary Information.

2.2 | Initial configurations

To generate the initial configurations of each system for sMD simulations, an initial structure was first thermalized to 300 K and 1 atm, which were maintained using the Langevin thermostat was used with a damping coefficient $\gamma = 1\ \mathrm{ps}^{-1}$; pressure was maintained using the Langevin piston Nose–Hoover method (Feller, Zhang, Pastor, & Brooks, 1995; Martyna, Tobias, & Klein, 1994). A cutoff of 11 Å was used for long-range electrostatic calculations, with a switching distance of 10 Å. The SHAKE algorithm (Ryckaert & Ciccotti, 1977) was used to constrain bonds to non-polar hydrogens, and a 2-fs timestep was used for all simulations. A total of 5 ns were simulated for each system, and snapshots of coordinates and velocities were saved at 10-ps intervals, for a total of 500 initial configurations for each system. The Cartesian coordinates of the carbon atom of the CO molecule as well as three carbons of the carbon nanotube were constrained in these simulations. For both systems, periodic boundary conditions were employed in all three dimensions. For the nanotube-containing system, the CO molecule was initially 2 Å away from the central axis of the nanotube.

2.3 | Steered molecular dynamics

For each system, a pulling vector was defined from the initial coordinates of the carbon atom of the CO molecule. For the case of CO in water, the vector was simply along the x-axis (1,0,0), and in the case of migration of CO through the carbon nanotube, the vector was the central axis of the carbon nanotube. The reaction coordinate was defined as the projection of the distance vector between the Cartesian coordinates of the carbon atom from its initial coordinates onto the pulling vector. The carbon atoms that were constrained in the generation of the initial configurations were also constrained during the sMD simulations, to avoid reorientation of the nanotube during simulations. Two different pulling velocities $v = 0.001$ and $v = 0.0001$ Å/timestep were considered, as well as a range of force constants (k = 20, 200, 400, 800 kcal/mol/Å^2) for the simplest system of CO in water; only the largest three force constants were considered for the system containing the carbon nanotube. For each combination of k and v, 500 sMD simulations were carried out. All simulations were carried out using the NAMD simulation package (Phillips et al., 2005).

3 | RESULTS AND DISCUSSION

A necessary component for running sMD simulations is the initial configurations of the system. The Jarzynski equality requires that the initial configurations be sampled from an equilibrium distribution with $\lambda = \lambda_0$ (Jarzynski, 1997). These configurations can be obtained through a Monte Carlo simulation, or by periodically saving snapshots from an MD simulation of the system with the reaction coordinate restrained to the desired initial value; we used the latter approach here. We note that it is important to ensure that the snapshots be taken at sufficiently long time intervals so that they are uncorrelated; a 10-ps interval was used for both systems considered here.

The two key parameters to set before performing constant velocity sMD simulations are the force constant k and the pulling velocity v. In what follows we will highlight results from a simple test system in which we pull a molecule of CO in a box of water molecules, and we later illustrate that the same trends can be observed for the other test system in which CO is pulled through the carbon nanotube. For both systems, we considered a range of values for k and different pulling velocities. For each set of conditions k and v, 500 sMD trajectories were generated. The same initial snapshots were used for all conditions.

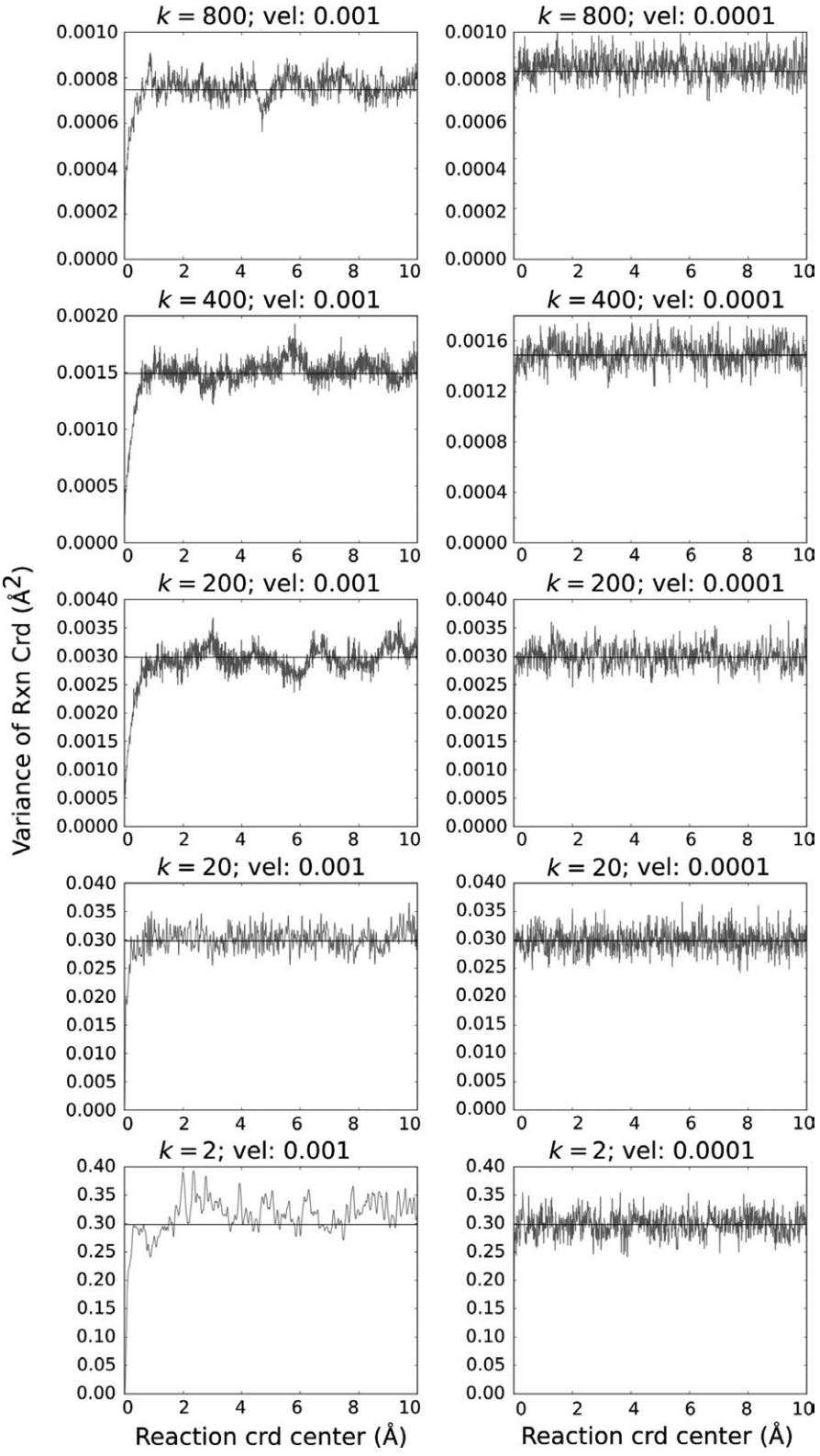

FIGURE 1 Time evolution of variance of the reaction coordinate λ' in sMD simulations of CO in water. The values of k and pulling velocity are specified in subplot titles. The grid is organized such that the value of k decreases from top to bottom, and v decreases from left to right. $N = 500$ trajectories were simulated for each set of conditions

3.1 | CO in water

As mentioned in the Section 1, the spring constant should be stiff enough such that $\lambda'(\boldsymbol{q})$ closely follows $\lambda(t)$. To simplify notation, we will refer to $\lambda'(\boldsymbol{q})$ and $\lambda(t)$ as λ' and λ, respectively. At any given time, the value of λ' for the group of sMD trajectories will be more or less clustered around λ, depending on the force constant k. In fact, the variance of λ' at a given time t is inversely proportional to k:

$$\frac{1}{\beta k} \approx \mathrm{var}\lambda' \tag{8}$$

In Figure 1, we verify this relation for values of k ranging from 20 to 800 kcal/mol/Å^2 and two different pulling velocities of 0.001 and 0.0001 Å/timestep. In each subplot, the time evolution of Var(λ') fluctuates around $1/\beta k$ (horizontal line in each subplot). This relation is key in determining the optimal value of k, which depends on the magnitude of an observable fluctuation in the reaction coordinate, $\delta(\lambda')$:

$$k^* = \frac{1}{\beta\delta(\lambda')} \tag{9}$$

here, since λ' represents the projection of a distance onto the pulling vector, a fluctuation of $\delta(\lambda') = 0.1$ Å, for instance, would be reasonable lower limit, resulting in an optimal value of approximately 60 kcal/mol/Å^2 at 300 K. Values of k larger than k^* unnecessarily increase the perturbation in the system and introduce large forces in oscillating directions. In Figure 2, we show the time evolution of the applied force for the same set of conditions. In practice, the work performed on the system is obtained by numerically integrating the force over time, for instance, by invoking the trapezoid rule; however, the large oscillations in the force values introduce a numerical drift in the integration, introducing error in the calculated work values, and subsequently, in the estimation of ΔF. In Figure 3, we show the estimated free energy profiles obtained by using either $\Delta\hat{F}_J$ (red series) or $\Delta\hat{F}_{ML}$ (blue series), and highlight that at slower pulling velocities the artifact resulting from large force constants is exacerbated, with both estimators resulting in large deviations from the true value of ΔF. We note that in this high-k regime, the first- and second-order terms in Equation 4 are small, meaning that the approximation of ΔF values for the underlying PMF is valid. Furthermore, even though the work distributions upon visual inspection appear to be Gaussian, the use of the second-order cumulant expansion results in large inaccuracy in the estimated free energy.

If we specifically consider the lowest value of k that satisfies the stiff-spring approximation (in this case, $k = 200$ kcal/mol/Å^2), we note that even at the fast pulling velocity of $v = 0.001$ Å/timestep, the use of $\Delta\hat{F}_{ML}$ permits a more accurate estimation of the free energy change than does $\Delta\hat{F}_J$, which performs worse at faster pulling velocities due to increased dispersion in the work distribution.

3.2 | Migration of CO through carbon nanotube

We performed analogous analysis for a system in which CO is pulled through a carbon nanotube. In this way, we approximate the problem of ligand migration in proteins while maintaining a known true value of $\Delta F = 0$ kcal/mol. Again, because the reaction coordinate is a distance, we can approximate an optimal value of $k^* \approx 60$ kcal/mol/Å^2. We summarize in Figure 4 the absolute error obtained in the free energy after crossing the nanotube under the different conditions evaluated. We note that the higher order terms in Equation 4 are small in all cases here (dashed series), yet the absolute error in estimated ΔF increases with k, particularly for the lower pulling velocity.

3.3 | Fast relaxations

Up until now, the trends for all the test systems would appear to suggest that the pulling velocity can do no harm, as long as the spring constant is close to the minimum value that satisfies the stiff-spring approximation, which is not true. In fact, the extrapolation of an sMD simulation to instantaneous switching results in a free energy perturbation calculation, for which the work distributions are generally Exponential or Gamma distributions, not Gaussian. Indeed, the expectation of Gaussian work distribution requires not only the stiff-spring approximation, but also sufficiently fast relaxations of the system.

A fairly straightforward metric to evaluate fast relaxations proposed by Park and Schulten (2004) is to calculate the time evolution of the diffusion coefficient of λ' in the sMD simulations:

$$D(t) = \frac{2v^2}{\beta^2}\left(\frac{\partial}{\partial t}\mathrm{var}\left(\lambda'\right)\right)^{-1} \tag{10}$$

As long as $v/\beta kD \ll l$, with l being a characteristic range of λ in which the change in D is large, we can consider that the fast relaxation criteria have been met. For the test systems evaluated here, the relaxations involve the reorganization of water molecules around the molecule of CO, and this criterion was satisfied under all conditions considered.

4 | CONCLUSION

Here, we have illustrated the practical effects of the choice of the spring force constant and pulling velocity in sMD

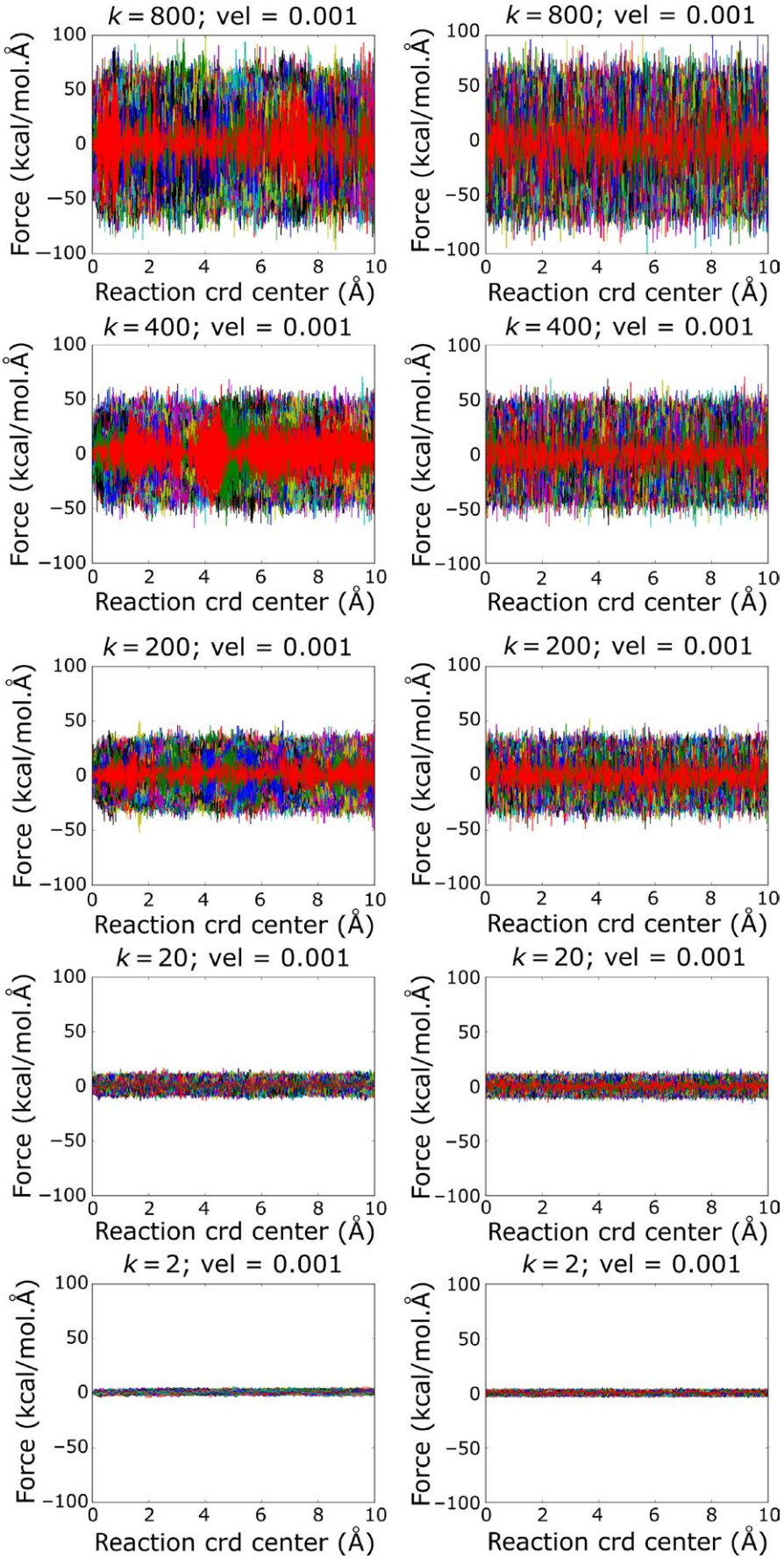

FIGURE 2 Time evolution of applied force in sMD simulations of CO in water. The values of k and pulling velocity are specified in subplot titles. The grid is organized such that the value of k decreases from top to bottom, and v decreases from left to right. 500 trajectories were simulated for each set of conditions

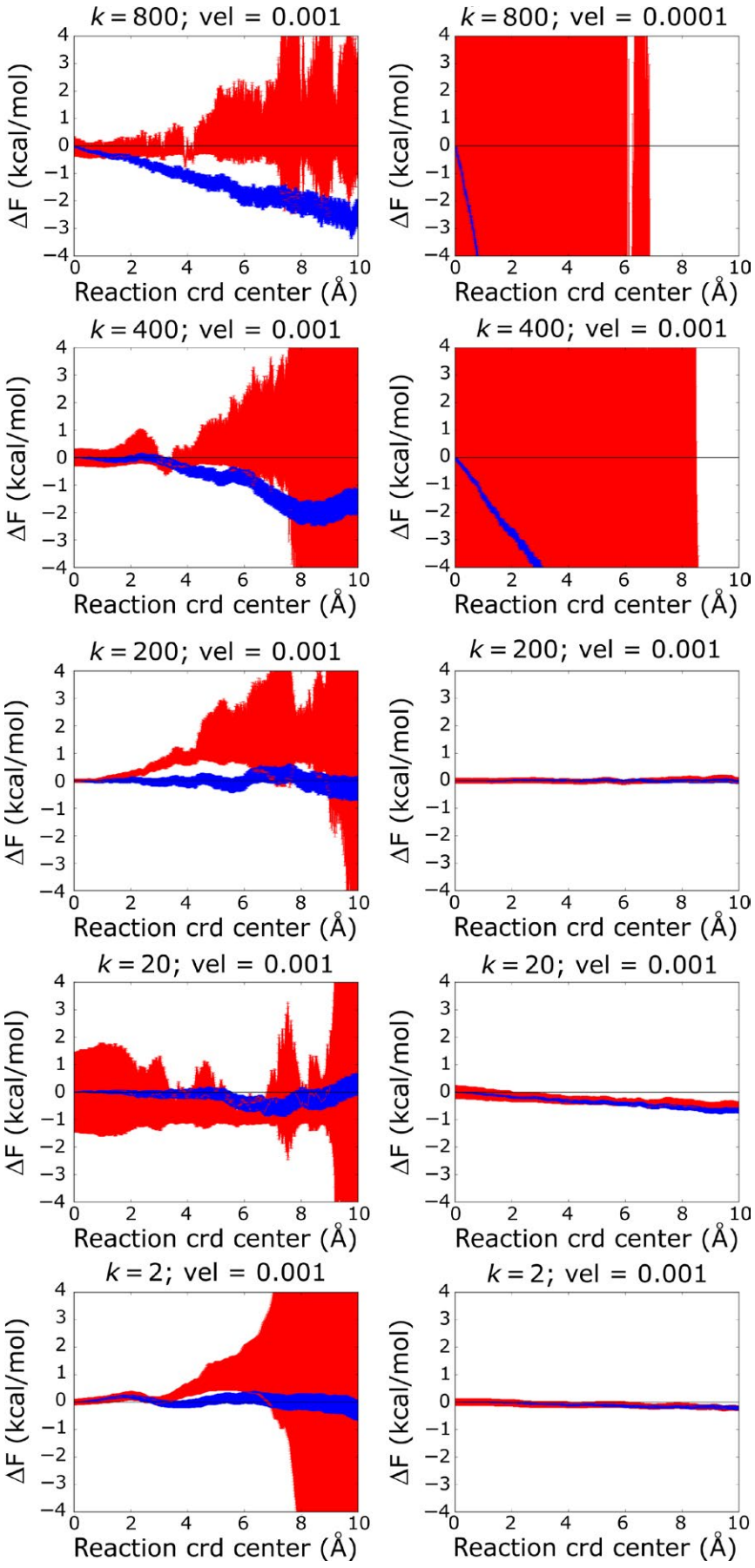

FIGURE 3 Free energy profiles for CO in water using $\Delta\hat{F}_{ML}$ (blue) and $\Delta\hat{F}_J$ (red) estimators. The values of k and pulling velocity are specified in subplot titles. The grid is organized such that the value of k decreases from top to bottom, and v decreases from left to right. 500 trajectories were simulated for each set of conditions. Error bars were calculated using error propagation as detailed elsewhere (Arrar et al., 2018); error bars for $\Delta\hat{F}_J$ exceed the scale of the figures in the high-k and low-v regime (top right)

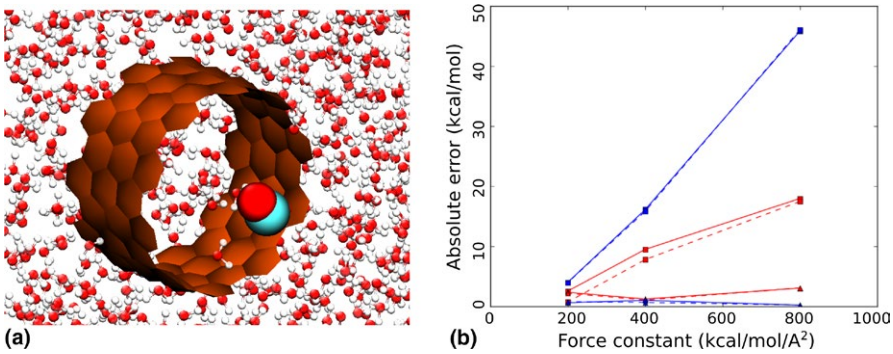

FIGURE 4 CO migration through carbon nanotube. Absolute error of ΔF estimate obtained with $\Delta\hat{F}_J$ (red) or $\Delta\hat{F}_{ML}$ (blue) for different force constants and pulling velocities. Square markers correspond to $v = 0.0001$ Å/timestep; triangle markers correspond to $v = 0.001$ Å/timestep. Dashed series show the effect of adding the higher order terms in Equation 4 to the free energy estimate

simulations. In particular, we have highlighted the importance of setting the spring force constant to the lowest possible value that satisfies the stiff-spring approximation. Only in this regime, and with the criteria of fast relaxations met, we conclude that the use of the second-order cumulant expansion is indeed advantageous for accurate estimation of ΔF. These methodological criteria should be useful in the determination of free energy profiles for investigating enzymatic chemical reactions, as well as in studies of ligand migration or conformational changes in proteins.

ACKNOWLEDGMENTS

We thank Dr. Daniela Rodriguez and Dr. Maria Eugenia Szretter for many helpful discussions. This work was supported in part by CONICET and Fundación Bunge and Born. Computer simulations were run in part on the CeCAR computing cluster.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

ORCID

Mehrnoosh Arrar https://orcid.org/0000-0001-6404-0552

REFERENCES

Arrar, M., Boubeta, F. M., Szretter, M. E., Sued, M., Boechi, L., & Rodriguez, D. (2018). On the accurate estimation of free energies using the Jarzynski equality. *The Journal of Computational Chemistry*, *40*, 688–696. https://doi.org/10.1002/jcc.25754

Bartels, C., & Karplus, M. (1997). Multidimensional adaptive umbrella sampling: Applications to main chain and side chain peptide conformations. *Journal of Computational Chemistry*, *18*, 1450–1462. https://doi.org/10.1002/(ISSN)1096-987X

Bieza, S. A., Boubeta, F., Feis, A., Smulevich, G., Estrin, D. A., Boechi, L., & Bari, S. E. (2014). Reactivity of inorganic sulfide species toward a heme protein model. *Inorganic Chemistry*, *54*, 527–533.

Boechi, L., Martí, M. A., Milani, M., Bolognesi, M., Luque, F. J., & Estrin, D. A. (2008). Structural determinants of ligand migration in *Mycobacterium tuberculosis* truncated hemoglobin O. *Proteins Structure, Function, and Bioinformatics*, *73*, 372–379. https://doi.org/10.1002/prot.22072

Boubeta, F. M., Bari, S. E., Estrin, D. A., & Boechi, L. (2016). Access and binding of H2S to hemeproteins: The case of HbI of *Lucina pectinata*. *The Journal of Physical Chemistry B*, *120*, 9642–9653. https://doi.org/10.1021/acs.jpcb.6b06686

Bringas, M., Petruk, A. A., Estrin, D. A., Capece, L., & Martí, M. A. (2017). Tertiary and quaternary structural basis of oxygen affinity in human hemoglobin as revealed by multiscale simulations. *Scientific Reports*, *7*, 10926. https://doi.org/10.1038/s41598-017-11259-0

Bucher, D., Walker, R. C., & McCammon, J. A. (2014). Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *Journal of Chemical Theory and Computation*, *10*, 2677–2689.

Case, D. A., Darden, T. A., Cheatham, T. E. I. I. I., Simmerling, C. L., Wang, J., Duke, R. E., … Kollman, P. (2010). *Amber 11*. San Francisco: University of California.

Chelli, R. (2012). Local sampling in steered Monte Carlo simulations decreases dissipation and enhances free energy estimates via nonequilibrium work theorems. *Journal of Chemical Theory and Computation*, *8*, 4040–4052. https://doi.org/10.1021/ct300348w

Chipot, C., & Pohorille, A. (2007). *Free energy calculations*. New York, NY: Springer. https://doi.org/10.1007/978-3-540-38448-9

Crespo, A., Martí, M. A., Estrin, D. A., & Roitberg, A. E. (2005). Multiple-steering QM-MM calculation of the free energy profile in chorismate mutase. *Journal of the American Chemical Society*, *127*, 6940–6941. https://doi.org/10.1021/ja0452830

Echeverria, I., & Amzel, L. M. (2012). Estimation of free-energy differences from computed work distributions: An application of jarzynski?s equality. *The Journal of Physical Chemistry B*, *116*, 10986–10995. https://doi.org/10.1021/jp300527q

Feller, S. E., Zhang, Y., Pastor, R. W., & Brooks, B. R. (1995). Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics*, *103*, 4613–4621. https://doi.org/10.1063/1.470648

Ferrenberg, A. M., & Swendsen, R. H. (1989). Optimized monte carlo data analysis. *Computers in Physics*, *3*, 101–104. https://doi.org/10.1063/1.4822862

Gohlke, H., Kiel, C., & Case, D. A. (2003). Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the Ras–Raf and Ras–RalGDS complexes. *Journal of Molecular Biology*, *330*, 891–913. https://doi.org/10.1016/S0022-2836(03)00610-7

Gore, J., Ritort, F., & Bustamante, C. (2003). Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proceedings of the National Academy of Sciences*, *100*, 12564–12569. https://doi.org/10.1073/pnas.1635159100

Hu, G., Xu, S., & Wang, J. (2015). Characterizing the free-energy landscape of MDM2 protein–ligand interactions by steered molecular dynamics simulations. *Chemical Biology & Drug Design*, *86*, 1351–1359. https://doi.org/10.1111/cbdd.12598

Hummer, G. (2001). Fast-growth thermodynamic integration: Error and efficiency analysis. *The Journal of Chemical Physics*, *114*, 7330–7337. https://doi.org/10.1063/1.1363668

Hummer, G. (2007). *Free energy calculations* (pp. 171–198). New York, NY: Springer.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD – Visual molecular dynamics. *Journal of Molecular Graphics*, *14*, 33–38. https://doi.org/10.1016/0263-7855(96)00018-5

Israelewitz, B., Baudry, J., Gullingsrud, J., Kosztin, D., & Schulten, K. (2001a). Steered molecular dynamics investigations of protein function. *Journal of Molecular Graphics and Modelling*, *19*, 13–25. https://doi.org/10.1016/S1093-3263(00)00133-9

Israelewitz, B., Gao, M., & Schulten, K. (2001b). Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology*, *11*, 224–230. https://doi.org/10.1016/S0959-440X(00)00194-9

Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Physical Review Letters*, *78*, 2690–2693. https://doi.org/10.1103/PhysRevLett.78.2690

Jarzynski, C. (2006). Rare events and the convergence of exponentially averaged work values. *Physical Review E*, *73*, 046105. https://doi.org/10.1103/PhysRevE.73.046105

Jarzynski, C. (2011). Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale. *Annual Review of Condensed Matter Physics*, *2*, 329–351. https://doi.org/10.1146/annurev-conmatphys-062910-140506

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, *79*, 926–935. https://doi.org/10.1063/1.445869

Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., & Kollman, P. A. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, *13*, 1011–1021. https://doi.org/10.1002/(ISSN)1096-987X

Marsico, F., Burastero, O., Defelipe, L. A., Lopez, E. D., Arrar, M., Turjanski, A. G., & Marti, M. A. (2018). Multiscale approach to the activation and phosphotransfer mechanism of CpxA histidine kinase reveals a tight coupling between conformational and chemical steps. *Biochemical and Biophysical Research Communications*, *498*, 305–312. https://doi.org/10.1016/j.bbrc.2017.09.039

Martyna, G. J., Tobias, D. J., & Klein, M. L. (1994). Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, *101*, 4177–4189. https://doi.org/10.1063/1.467468

Nicolini, P., Frezzato, D., & Chelli, R. (2011). Exploiting configurational freezing in nonequilibrium Monte Carlo simulations. *Journal of Chemical Theory and Computation*, *7*, 582–593. https://doi.org/10.1021/ct100568n

Ozer, G., Valeev, E. F., Quirk, S., & Hernandez, R. (2010). Adaptive steered molecular dynamics of the long-distance unfolding of neuropeptide Y. *Journal of Chemical Theory and Computation*, *6*, 3026–3038. https://doi.org/10.1021/ct100320g

Park, S., & Schulten, K. (2004). Calculating potentials of mean force from steered molecular dynamics simulations. *Journal of Chemical Physics*, *120*, 5946–5961. https://doi.org/10.1063/1.1651473

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., … Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*, 1781–1802. https://doi.org/10.1002/(ISSN)1096-987X

Pohorille, A., Jarzynski, C., & Chipot, C. (2010). Good practices in free-energy calculations. *Journal of Physical Chemistry B*, *114*, 10235–10253. https://doi.org/10.1021/jp102971x

Ramírez, C. L., Zeida, A., Jara, G. E., Roitberg, A. E., & Martí, M. A. (2014). Improving Efficiency in SMD simulations through a hybrid differential relaxation algorithm. *Journal of Chemical Theory and Computation*, *10*, 4609–4617. https://doi.org/10.1021/ct500672d

Ryckaert, J., & Ciccotti, G. (1977). ScienceDirect—choose organization. *Journal of Computational Physics*, *23*, 327–341. https://doi.org/10.1016/0021-9991(77)90098-5

Schmiedl, T., & Seifert, U. (2007). Optimal finite-time processes in stochastic thermodynamics. *Physical Review Letters*, *98*, 108301. https://doi.org/10.1103/PhysRevLett.98.108301

Selvam, B., Wereszczynski, J., & Tikhonova, I. G. (2012). Comparison of dynamics of extracellular accesses to the β1 and β2 adrenoceptors binding sites uncovers the potential of kinetic basis of antagonist selectivity. *Chemical Biology & Drug Design*, *80*, 215–226. https://doi.org/10.1111/j.1747-0285.2012.01390.x

Shirts, M. R., & Chodera, J. D. (2008). Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics*, *129*, 124105. https://doi.org/10.1063/1.2978177

Shirts, M. R., Mobley, D. L., & Brown, S. P. (2010). Free-energy calculations in structure-based drug design. *Drug Design: Structure- and Ligand-Based Approaches*, 61–86. https://doi.org/10.1017/CBO9780511730412

Vaikuntanathan, S., & Jarzynski, C. (2008). Escorted free energy simulations: Improving convergence by reducing dissipation. *Physical Review Letters*, *100*, 190601. https://doi.org/10.1103/PhysRevLett.100.190601

Wu, D., & Kofke, D. A. (2005a). Phase-space overlap measures: I: Fail-safe bias detection in free energies calculated by molecular simulation. *The Journal of Chemical Physics*, *123*, 054103. https://doi.org/10.1063/1.1992483

Wu, D., & Kofke, D. A. (2005b). Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations. *The Journal of Chemical Physics*, *123*, 084109. https://doi.org/10.1063/1.2011391

Xiong, H., Crespo, A., Marti, M., Estrin, D., & Roitberg, A. E. (2006). Free energy calculations with non-equilibrium methods: Applications

of the Jarzynski relationship. *Theoretical Chemistry Accounts*, *116*, 338–346. https://doi.org/10.1007/s00214-005-0072-2

Ytreberg, F. M., & Zuckerman, D. M. (2004). Efficient use of nonequilibrium measurement to estimate free energy differences for molecular systems. *Journal of Computational Chemistry*, *25*, 1749–1759. https://doi.org/10.1002/(ISSN)1096-987X

Yunger Halpern, N., & Jarzynski, C. (2016). Number of trials required to estimate a free-energy difference, using fluctuation relations. *Physical Review E*, *93*, 052144. https://doi.org/10.1103/PhysRevE.93.052144

Zerbetto, M., Piserchia, A., & Frezzato, D. (2014). Looking for some free energy? *Journal of Computational Chemistry*, *35*, 1865–1881. https://doi.org/10.1002/jcc.23701

Zuckerman, D. M., & Woolf, T. B. (2001). Extrapolative analysis of fast-switching free energy estimates in a molecular system. arXiv preprint physics/0107066

Zuckerman, D. M., & Woolf, T. B. (2002). Overcoming finite-sampling errors in fast-switching free-energy estimates: Extrapolative analysis of a molecular system. *Chemical Physics Letters*, *351*, 445–453.

How to cite this article: Boubeta FM, Contestín García RM, Lorenzo EN, et al. Lessons learned about steered molecular dynamics simulations and free energy calculations. *Chem Biol Drug Des*. 2019;00:1–10. https://doi.org/10.1111/cbdd.13485