# Comment on ``Statistical efficiency of methods for computing free energy of hydration''[ J. Chem. Phys. 149, 144111 (2018)]

# AUTHOR QUERY FORM

| | | |
|---|---|---|
| AIP Publishing | **Journal: J. Chem. Phys.**<br><br>**Article Number: 058912JCP** | **Please provide your responses and any corrections by annotating this PDF and uploading it to AIP's eProof website as detailed in the Welcome email.** |

Dear Author,

Below are the queries associated with your article. Please answer all of these queries before sending the proof back to AIP.

**Article checklist:** In order to ensure greater accuracy, please check the following and make all necessary corrections before returning your proof.
1. Is the title of your article accurate and spelled correctly?
2. Please check affiliations including spelling, completeness, and correct linking to authors.
3. Did you remember to include acknowledgment of funding, if required, and is it accurate?

| Location in article | Query/Remark: click on the Q link to navigate to the appropriate spot in the proof. There, insert your comments as a PDF annotation. |
|---|---|
| Q1 | Please check that the author names are in the proper order and spelled correctly. Also, please ensure that each author's given and surnames have been correctly identified (given names are highlighted in red and surnames appear in blue). |
| Q2 | Please define MBAR at first occurrence. |
| Q3 | Please define BAR at first occurrence. |
| Q4 | Please check definition of MSD. |
| | Please confirm ORCIDs are accurate. If you wish to add an ORCID for any author that does not have one, you may do so now. For more information on ORCID, see https://orcid.org/.<br>Piero Procacci – 0000-0003-2667-3847 |

Thank you for your assistance.

The Journal of Chemical Physics COMMENT scitation.org/journal/jcp

# Comment on "Statistical efficiency of methods for computing free energy of hydration" [J. Chem. Phys. 149, 144111 (2018)]

Cite as: J. Chem. Phys. 150, 000000 (2019); doi: 10.1063/1.5086743
Submitted: 22 December 2018 • Accepted: 10 March 2019 •
Published Online: XX XX XXXX



Piero Procacci[a)]

AFFILIATIONS

Department of Chemistry, University of Florence, Florence, Italy

[a)]Electronic mail: procacci@unifi.it

https://doi.org/10.1063/1.5086743



In Ref. 1, Yildirim *et al.* compared the statistical efficiency for the computation of the hydration free energy (HFE) of a set of small organic molecules using nonequilibrium work (NEW) methods and the standard equilibrium free energy perturbation (FEP) approach with stratification (MBAR). Based on the analysis of the 34 rigid molecules set, the authors conclude that "the nonequilibrium methods tested here for the prediction of HFE have lower computational efficiency than the MBAR method."

This conclusion is based on the comparison of a so-called "statistical efficiency" $\epsilon_i = 1/T_i\sigma_i^2$, where $T_i$ and $\sigma_i^2$ are the total simulation time and the mean variance of method $i$, respectively. The authors state that the simulation time, $T$, "is used as a proxy for the amount of information in the simulation." Their definition of $\epsilon$ is not, by any means, a statistical efficiency and depends on their particular choice of the simulation protocol. The statistical efficiency of an unbiased estimator for a given parameter, say, $\Delta G$, is rigorously defined as the ratio of the inverse of the Fischer information (or minimum variance) and the sample variance for $\Delta G$ (e.g., computed with bootstrap with resampling), i.e.,

$$\epsilon = \frac{1/I(\Delta G)}{\sigma^2}, \tag{1}$$

where

$$I(\Delta G) = -E\left[\frac{\partial^2 \log(f(X,\Delta G))}{\partial \Delta G^2} \middle| \Delta G\right] \tag{2}$$

is the Fischer information and $f(X, \Delta G)$ is the distribution of the estimator $\Delta G$. The statistical efficiency is hence a number such that $0 \le \epsilon \le 1$. The inverse of the Fischer information for a fixed number of simulations, in the case of BAR, is known and is given by Eq. (10) of Ref. 2. If the authors had used the correct Fischer information instead of their "proxy," they would have found a statistical efficiency close to one. This is so since the sample variance, $\sigma^2$, obtained by computing many time $\Delta G$ using resampling with replacement, must be close to and *larger* than the theoretical variance $1/I(\Delta G)$ for a fixed number of studies given in Ref. 2. Hence, the true $\epsilon$ says nothing about the computational efficiency, except for the known fact[3] that BAR is a statistically efficient estimator. Nonetheless, $\epsilon$ could be regarded as a legitimate measure of *computational efficiency*. However, in order to test the efficiency of the NEW method, it would have been interesting to assess, at fixed total simulation time $T$, the effect on $\Delta G$ and its variance of the duration of the nonequilibrium trajectories and of the switching protocol during the decoupling/recoupling of the solute. By the same token, the optimal protocol[4] in terms of $\epsilon$ could have been determined for the FEP/MBAR simulation (number and schedule of $\lambda$ windows). Unfortunately, this kind of analysis is missing in the Yildirim paper.

Concerning the FEP-based equilibrium simulations, MBAR[5] is known to "perform similarly to BAR when the spacing between intermediate states is moderate and therefore only neighboring states have significant phase space overlap." This sentence is taken *verbatim* from Ref. 7[6] of the Yildirim paper. Besides, in Ref. 7, I have shown that the free energies computed with BAR or MBAR are virtually indistinguishable in a challenging equilibrium alchemical application, provided that the overlap between neighboring distribution is significant, which makes MBAR somewhat redundant in well-designed equilibrium alchemical simulations. Consequently, the variance in MBAR must be similar to that of BAR. The latter is basically given by the sum over all inner strata of terms such as those given by Eq. (10) of Ref. 2. However, in FEP with stratification, BAR

or MBAR can give a reliable estimate if and only if the sampling is adequate *in each stratum* of the alchemical coordinate. In Refs. 6, 8, and 9, it has been authoritatively pointed out that "sampling (in alchemical calculations with stratification) remains a critical issue as the solute size and flexibility grow and as the solvent dynamics or environment becomes heterogeneous, for example, for solvation free energies in octanol."[6]

In Ref. 1, the sampling issue has been analyzed *only for the end states*, which allows us to compute reliably the variance only for the NE approach, where the inner states are crossed at fast speed and just the final (end states) NE work distributions matter for producing the estimate $\Delta G$. The authors appear to be somehow aware of this fact, since they did perform "three replicates of the same simulation" to get a "more realistic" uncertainty for the equilibrium MBAR alchemical computation. Even given for granted that three trials are enough to get a "realistic" value of the modified signed-digit (MSD) for each of the 34 solutes in the set, it remains unclear why the authors failed to include the cost of these replicates (100 ns each) in the total simulation time for MBAR, lowering their efficiency measure to 3.9, i.e., in line with the unidirectional NEW Jarzynski approaches and less than that of NEW/BAR.

Finally, I do believe that, for the sound reasons explained above, while the Yilderim *et al.* paper is certainly technically valid and useful, their conclusions are not sufficiently motivated and that a much more thorough and challenging analysis is needed in order to rigorously compare the efficiency of NEW and equilibrium stratification techniques, especially with regard to the issue of the adequacy of sampling. In this respect, it should be taken into account that in NEW, adequate sampling is required only at the end states and that uncertainty on $\Delta G$ is essentially a function of the variance (or dissipation) of the work distributions.[9] In FEP with stratification, adequate sampling must be checked on each $\lambda$ stratum and the convergence rate is not guaranteed, by any means, to be the same on each of the strata.[10]

## REFERENCES


[1] A. Yildirim, T. A. Wassenaar, and D. van der Spoel, "Statistical efficiency of methods for computing free energy of hydration," J. Chem. Phys. **149**(14), 144111 (2018).

[2] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, "Equilibrium free energies from nonequilibrium measurements using maximum likelihood methods," Phys. Rev. Lett. **91**, 140601 (2003).

[3] C. H. Bennett, "Efficient estimation of free energy differences from Monte Carlo data," J. Comput. Phys. **22**, 245–268 (1976).

[4] L. N. Naden and M. R. Shirts, "Linear basis function approach to efficient alchemical free energy calculations. 2. Inserting and deleting particles with Coulombic interactions," J. Chem. Theory Comput. **11**, 2536–2549 (2015).

[5] M. R. Shirts and J. D. Chodera, "Statistically optimal analysis of samples from multiple equilibrium states," J. Chem. Phys. **129**, 124105 (2008).

[6] G. D. R. Matos, D. Y. Kyu, H. H. Loeffler, J. D. Chodera, M. R. Shirts, and D. L. Mobley, "Approaches for calculating solvation free energies and enthalpies demonstrated with an update of the freesolv database," J. Chem. Eng. Data **62**(5), 1559–1569 (2017).

[7] P. Procacci, "Multiple Bennett acceptance ratio made easy for replica exchange simulations," J. Chem. Phys. **139**, 124105 (2013).

[8] D. L. Mobley, "Let's get honest about sampling," J. Comput.-Aided Mol. Des. **26**(1), 93–95 (2012).

[9] A. Pohorille, C. Jarzynski, and C. Chipot, "Good practices in free-energy calculations," J. Phys. Chem. B **114**(32), 10235–10253 (2010).

[10] X. Wang, X. Tu, Z. John, H. Zhang, and Z. Sun, "Bar-based optimum adaptive sampling regime for variance minimization in alchemical transformation: The nonequilibrium stratification," Phys. Chem. Chem. Phys. **20**, 2009–2021 (2018).