DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service
National Institutes of Health
National Heart, Lung, and
Blood Institute
Bethesda, Maryland 20892

5635 Fishers Lane, Room T922
brb@nih.gov, MSC 9314
Phone: 301-496-0148
May 12, 2017

Professor David Mobley
Department of Pharmaceutical Sciences
Department of Chemistry
3134B Natural Sciences I
University of California, Irvine
Irvine, CA 92697

Re: SAMPL challenge testimonial

Dear David,

Our group has participated in the past three editions of the SAMPL blind free energy prediction challenge. In SAMPL3 we participated in the host-guest binding challenge,[1] for SAMPL4 we participated in the hydration free energy challenge (HFE)[2] and for SAMPL5 we participated in both the host-guest[3,4] and distribution coefficient challenges.[5–7] Throughout these various challenges, SAMPL has consistently provided us with an excellent means to evaluate the progress of our methodological development efforts in relative terms, against our competitors in the field, and in absolute terms, with respect to high quality experimental data.

Throughout our participation in the various SAMPL challenges, we have taken the lessons we have learned from previous challenges and applied them both to our future participation in general, and to our efforts to develop better computational methodologies.

From SAMPL3 we observed firsthand the superiority of the Bennett's Acceptance Ratio (BAR) free energy estimator to thermodynamic integration (TI), the critical importance of reproducing the experimental ionic strength in our simulations, and the sensitivity of our predicted results to the protonation state of titratable chemical groups. While these results were independently known to the field, it was important for our group to understand their relative importance and for us to calibrate our approaches to accommodate their practical impact upon our free energy predictions.

In SAMPL4 we focused on applying our suite of hybrid quantum/classical methods (QM/MM) to the problem of HFE prediction. Despite the great expense of these methods, our results were disappointing. Our primary lesson here was the sensitive nature of our choice of QM method in concert with our explicit solvent model. Popular methods like B3LYP yielded inferior results to older, and often overlooked functionals such as BLYP, because of a lack of proper error cancellation. This work also informed us of the challenges associated with explicit solvent models over-polarizing the QM region, especially relative to highly accurate implicit solvation models.[8] The high cost of our SAMPL4 QM/MM calculations also spurred

research in our group on developing approximate QM/MM methods such as MESS, which treat the solute as a rigid body, and precompute the solute's polarization response to an arbitrary external perturbation.[9] Such approximations yield up to two orders of magnitude decrease in computational expense, with a minor decrease in prediction accuracy. Finally, the HFE data from SAMPL4 provided us with a robust set of benchmarking data, which we used to explore and improve our various schemes for reweighting from classical simulations to QM based potentials.[10,11]

Our group participated in both the host-guest and partition coefficient challenges of the most recent SAMPL5 challenge. These challenges tested our group's ability to incorporate knowledge gleaned from our participation in previous SAMPL challenges and allowed our group to critically assess the progress we have made with our method development over the past two years. Taking these together, we were very pleased with the relatively strong showing of our results as compared with our competitors in the field. Despite submitting many high ranking predictions in the challenge, our optimism was tempered by the poor accuracy of our predictions in absolute terms, as measured by comparison with experiment. Some of the issues we encountered were related to the correct parametrization of a flexible host, the identification of the protonation states of ligands and the treatment of halogenated ligands.

The most recent challenge also afforded us the opportunity to develop new computational tools to address difficulties associated with free energy and ligand docking simulations, such as modifying Galaxy-Dock to work with host-guest systems (in collaboration with Prof. Chaok Seok, Seoul National University). We identified that using it to create optimal initial configurations for ligand binding free energy simulations leads to better quantitative predictions. We also developed a QM based pKa prediction pipeline to automate the critical, yet tedious, calculations required to accurately predict the correct protonation state of ligands in both the host guest and partition coefficient challenges.

Our poor predictions, in absolute terms, have implicated several deficiencies in current generation models that we will seek to address via polarizable force fields, constant pH ensembles and better hybrid QM/MM reweighting schemes. Since these method development topics are already a focus of research in our group, we believe that future SAMPL challenges will be an ideal showcase for our lab's methodological progress in an unbiased way.

There are too many published papers in our field presenting free energy protocols and methods that would not stand up well to the rigors of blind challenges such as SAMPL. And it is too easy to introduce unwarranted bias, even when the intent is to avoid such. There are too many studies where the data sets have been cherry picked, or even worse, cherry picking simulations. And It is too easy to chase agreement with experiment by tweaking a few force field parameters. But when it comes to truly improving methods and developing better ones, if we are going to separate the proverbial wheat from the chaff, then it is essential that we have

challenges and competitions like SAMPL. All of the worst practices are nearly impossible under a blind challenge. There is no other way to make rapid progress, efficiently and reliably, as a broader community. Challenges of this kind, such as SAMPL, must be funded. We sincerely hope that the NIH will ensure the continuing future of the SAMPL challenge through this important initiative. And here is my personal challenge the study section panel… If not SAMPL, then what?

Sincerely,

Bernard R. Brooks, Ph.D.
Chief, Computational Biophysics Section
Chief, Laboratory of Computational Biology
National Heart, Lung and Blood Institute

SAMPL related Laboratory of Computational Biology publications:

1. König, G. & Brooks, B. R. Predicting binding affinities of host-guest systems in the SAMPL3 blind challenge: the performance of relative free energy calculations. J. Comput. Aided Mol. Des. 26, 543–550 (2012).
2. König, G., Pickard, F. C., Mei, Y. & Brooks, B. R. Predicting hydration free energies with a hybrid QM/MM approach: an evaluation of implicit and explicit solvation models in SAMPL4. J. Comput. Aided Mol. Des. 28, 245–257 (2014).
3. Lee, J. et al. Absolute binding free energy calculations of CBClip host–guest systems in the SAMPL5 blind challenge. J. Comput. Aided Mol. Des. 31, 71–85 (2017).
4. Tofoleanu, F. et al. Absolute binding free energies for octa-acids and guests in SAMPL5. J. Comput. Aided Mol. Des. 31, 107–118 (2017).
5. König, G. et al. Calculating distribution coefficients based on multi-scale free energy simulations: an evaluation of MM and QM/MM explicit solvent simulations of water-cyclohexane transfer in the SAMPL5 challenge. J. Comput. Aided Mol. Des. 30, 989–1006 (2016).
6. Pickard, F. C. et al. Blind prediction of distribution in the SAMPL5 challenge with QM based protomer and pKa corrections. J. Comput. Aided Mol. Des. 30, 1087–1100 (2016).
7. Jones, M. R., Brooks, B. R. & Wilson, A. K. Partition coefficients for the SAMPL5 challenge using transfer free energies. J. Comput. Aided Mol. Des. 30, 1129–1138 (2016).
8. Pickard, F. C., König, G., Simmonett, A. C., Shao, Y. & Brooks, B. R. An efficient protocol for obtaining accurate hydration free energies using quantum chemistry and reweighting from molecular dynamics simulations. Bioorg. Med. Chem. 24, 4988–4997 (2016).
9. König, G. et al. Computation of Hydration Free Energies Using the Multiple Environment Single System Quantum Mechanical/Molecular Mechanical Method. J. Chem. Theory Comput. 12, 332–344 (2015).
10. Jia, X. et al. Calculations of Solvation Free Energy through Energy Reweighting from Molecular Mechanics to Quantum Mechanics. J. Chem. Theory Comput. 12, 499–511
11. Dybeck, E. C., König, G., Brooks, B. R. & Shirts, M. R. Comparison of Methods To Reweight from Classical Molecular Simulations to QM/MM Potentials. J. Chem. Theory Comput. 12, 1466–1480 (2016).