

## SAMPL2 challenge: prediction of solvation energies and tautomer ratios

A. Geoffrey Skillman · Matthew T. Geballe · Anthony Nicholls

Received: 13 April 2010 / Accepted: 15 April 2010 / Published online: 28 April 2010  
© Springer Science+Business Media B.V. 2010

This special issue reports on the third annual Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) challenge and workshop. The SAMPL2 challenge was open to all and allowed participants to make blind predictions of tautomer ratios and small-molecule hydration energies. Participants were asked to make specific predictions of experimental values that were not readily available. In order to maximize participation in SAMPL2, the challenge was announced at several conferences and on CCL (ccl.net, the computational chemistry list), and practitioners who had published articles about tautomer prediction or transfer energies within the last few years were invited directly. Prior SAMPL challenges covered transfer energies, affinity predictions, virtual screening and protein-ligand co-crystal pose prediction and have been discussed elsewhere [JMC 2008, 51(4):769–779; JPC B 2009, 113(14):4501–4537]. This introduction will comment briefly on the results from SAMPL2, discuss the importance and design of prospective predictions, and introduce the next challenge, SAMPL3.

The SAMPL2 evaluation was carried out in the spring of 2009 and discussed at a workshop held at McGill University in Montreal in June of 2009. This issue contains ten papers covering the results of the challenge and an eleventh manuscript will appear soon in JCAMD's special issue on Tautomers. The first paper is an overview of the data sets, methods of analyzing performance, and a perspective about the ability of computational chemists to correctly estimate these relatively simple experiments. The SAMPL organizers report that transfer energies can reliably be predicted with between 1 and 2 kcal/mol accuracy but that similar accuracy

in prediction of tautomer ratios requires computationally intense quantum-mechanical calculations, even for the simplest of tautomeric molecules. The following nine manuscripts describe individual efforts and include the very wide variety of solvation models currently under active development in the computational community, including all-atom explicit water simulations and, single- and multi-conformer implicit solvent models either with or without integrated quantum-mechanical calculations. The methods also included static charge models fit by a variety of methods and levels of theory as well as multiple approaches to solute polarizability. Despite the panoply of methods, it is encouraging to see that in many cases, the theoretical approaches produce similar results, and in some cases may even suggest the need to reexamine the experimental data.

### On the importance and design of blind challenges

The SAMPL challenge was designed to address computational prediction of biologically-relevant processes. Each challenge addresses one very basic challenge (hydration energies) and one more “applied” challenge (tautomer ratios, binding affinity, crystallographic protein-ligand geometry prediction). All of these experiments share the trait that, like most biological processes, they occur in condensed aqueous phase. It is for this reason the prediction of vacuum-water transfer energies (hydration energies) of simple small molecules is a recurring theme. If, as computational chemists, we cannot accurately predict the solvation of mono- or bi-functional small molecules, we should not expect (nor believe) more accurate predictions of dramatically more complex processes can be made without fortuitous cancellation of large terms, addition of critical fitting parameters, or model selection bias.

A. G. Skillman (✉) · M. T. Geballe · A. Nicholls  
OpenEye Scientific Software, Santa Fe, New Mexico, USA  
e-mail: skillman@eyesopen.com

Two of the most important aspects of the SAMPL challenge are first, working with blinded data or prospective prediction experiments and second, allowing scientists to make predictions while remaining anonymous. Despite the many potential shortcomings, prospective predictions, preferably in collaboration between experimentalists and theorists, provide a practical, fundamental “look in the mirror” for a field. Prospective predictions, or at least blind predictions of external data sets, eliminate many operational parameters commonplace in retrospective analyses, reveal over-fitting, and prevent system selection bias. In short, they make it much more difficult for us to fall afoul of Feynman’s first principle “you must not fool yourself and you are the easiest person to fool.” [Surely You’re Joking, Mr. Feynman!, W.W. Norton & Company, Inc., 1985, p. 343]

The most controversial aspect of our SAMPL experimental design is the acceptance of predictions from anonymous scientists. The publication and presentation of scientific data is generally limited to successes and this creates the well known “publication bias”, which makes it appear as if all experiments and predictions of note are successful. In order to allow SAMPL to give a more realistic picture our field, we specifically avoid the publication bias by allowing anonymous submissions and by allowing participants to declare their anonymity only after receiving the analysis of their results. This has proven to be a successful strategy and has resulted in submission of some high-risk predictions, that otherwise might not have been attempted. It is difficult to conjecture to what degree publication bias holds back our scientific progress, but it certainly does not help us face our challenges. Avoiding this bias is an important aspect of coming to correct conclusions about the strength of our field and these results are included in all of the aggregate results reported in the overview. Nevertheless, the scientific publication conventions remain and none of the anonymous submissions, whether successful or not, are found among the reports here.

Unfortunately, it is neither newsworthy, nor even noteworthy when hydration-energy predictions are made and, even more regrettable, *there seems no place in our current scientific environment for measuring these critical, yet relatively mundane values*. Despite our scientific skepticism, we are all invested in computational chemistry and are inclined to accept the anecdotal predictions on complex systems rather than to consider the implications that most solvation models cannot correctly predict the transfer-energy trend line for a series of substituted small molecule analogs.

Too often in prospective challenges participants who perform well and onlookers as well seek to “declare victory” for one method or technique. Declarations of winners can cause successful practitioners to avoid looking too

closely at their results. Meanwhile, poor performers seek to justify their shortcomings, ignore outliers and disparage the evaluation. However, the true value of an approach can only be ascertained using careful statistical analysis and examination of the population from which the evaluation data is drawn. For instance, in this issue the analysis of transfer energies by one group generated a mean error across all blinded compounds statistically significantly better than other methods, a result that is interesting but not insightful. Detailed analysis of the data set shows that the twenty-three compound dataset primarily falls into three classes. Further, the statistical superiority was due to especially good predictions for one of those, a series of substituted uracils.

In fact, such analysis is enlightening and the identification and the resultant understanding of such trends provide far more insights into our understanding of solvation models than merely celebrating the “defeat” of yet another scientific problem. A better approach is to cultivate a vision of a progressive field where we learn the strengths and weaknesses of each method, understand the difficulties, improve our understanding, and ultimately our predictions.

SAMPL attempts to bring all the practitioners together to critically analyze the dataset, the aggregate successes and failures of all methods and, perhaps most importantly, the areas where one or more models has a distinct advantage or disadvantage. It is through this honest and constructive analysis of both failures and successes that we can generate the greatest progress.

### **SAMPL3 in 2010: fragment-based design and the DINGO dataset**

The fourth annual SAMPL evaluation, SAMPL3, is being planned to begin in the summer of 2010. We are excited to announce SAMPL3 will be using the DINGO data set compiled by Tom Peat, Janet Newman, Kim Branson and co-workers [J. Biomol Screen, 2009, doi:[10.1177/1087057109348220](https://doi.org/10.1177/1087057109348220)]. This intriguing data set includes SPR binding measurements and x-ray crystallographic structures for the 500 member Maybridge Ro3 fragment library against Trypsin [[www.maybridge.com](http://www.maybridge.com)]. Participants in the three-stage exercise can attempt to (1) select binders from the 500 fragments, (2) predict the bound fragment geometries found in the fragment-protein crystallographic structures, and (3) estimate binding energies for a small set of compounds. In addition to the DINGO dataset, SAMPL3 will include a challenge to predict hydration free energies compiled by Peter Guthrie (University of Western Ontario). The SAMPL3 workshop will be hosted at Stanford University in April of 2011. For more information visit the challenge web site at [sAMPL.eyesopen.com](http://sAMPL.eyesopen.com).