SAMPL3: blinded prediction of host–guest binding affinities, hydration free energies, and trypsin inhibitors

A. Geoffrey Skillman

This special issue of the Journal of Computer-Aided Molecular Design is the culmination of the 4th Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) challenge and workshop. SAMPL3 had three datasets: blinded small-molecule hydration energies, provided by Peter Guthrie [1]; two novel host–guest systems, including eleven unpublished binding energies, provided by Adam Urbach and Lyle Issacs [2], and a monumental dataset including structural and affinity data for 500 fragments against Trypsin, provided by Tom Peat [3]. The SAMPL3 workshop saw over 40 attendees while the SAMPL3 challenge received 103 submissions from 23 participating groups using a variety of methods including: discrete and dynamic conformational sampling; implicit, semi-implicit and explicit water models; and myriad of force-fields and charge models. Gilson [2] and Geballe [1] have provided summaries of the host–guest challenge and solvation-energy challenge respectively. As with prior SAMPL challenges, many different approaches generated high-quality predictions yet no single technique distinguished itself significantly. Nevertheless, many important insights into the strengths and limitations of computational and experimental methods were developed through SAMPL.

SAMPL3 was the first blinded challenge to include prediction of host–guest binding affinities. Host–guest binding affinities provided an outstanding blind challenge, as they are simple enough to encourage participants to recognize, explore and address assumptions and errors. Most participants had great difficulty modeling the aspartyl-protease-like formal charges found in the host

molecules. This is concerning, for while ionization sites occur commonly in protein–ligand systems, rarely are they addressed at the level of detail participants found necessary for this host–guest system. More host–guest examples should be included in future SAMPL challenges as their streamlined nature highlights assumptions that can be too easily overlooked.

SAMPL is one of several projects that provide blinded or prospective experimental challenges to the computational community [4–6]. These projects are intended to serve as both a guidepost for computational progress and a meeting ground for experimental and computational scientist. However, it has been a struggle to generate mutual interest between computational and experimental scientists. This year, SAMPL had a breakthrough in the form of Lyle Issacs (host–guest affinities) and Tom Peat (trypsin structures and affinities). Lyle and Tom are experimentalists who provided data to SAMPL3, attended the SAMPL workshop, and provided insights and challenges to the computational scientists. We hope they are the first of many experimentalists to join SAMPL and challenge theorists with their data.

Unfortunately, no experimental collaborator has emerged to provide prospective hydration free energies, which have been part of each of the four SAMPL evaluations. Hydration energies are the most basic measure of the solvation of molecules in water. Aqueous solvation plays a critical role in most biophysical and biochemical phenomenon, and our ability to accurately predict biophysical processes is limited by our ability to accurately calculate solvation interactions. Hydration energies represent one of the simplest experiments that allow us to evaluate these predictions. As a consequence, comparison to hydration energies is a fundamental tool for evaluating force fields and electrostatic models (for example see [7]).

A. G. Skillman (✉)
OpenEye Scientific Software, Santa Fe, NM, USA
e-mail: skillman@eyesopen.com

Nevertheless, it is quite difficult to obtain new measurements. Despite efforts to identify collaborators from industry, government and academia, SAMPL remains dependent on the valiant efforts of Peter Guthrie to extract obscure new hydration energies from historic literature. Unfortunately, this is a limited supply and does not always include the most relevant drug-like compounds. Unless new sources of experimental measurements are identified, computational scientists may soon be limited to using this incisive measure of electrostatic model quality in a retrospective manner.

Previous SAMPL assessments [8] have primarily focused on designed retrospective data sets whereas this year two prospective data sets were presented. For example, during SAMPL1, JNK3 kinase and Urokinase data sets were designed by Vertex and Abbott respectively, to included tens of inhibitors spread evenly over 4–5 orders of magnitude in binding affinity. The spread of these data points were carefully chosen to effectively evaluate computational methods. The evenly spaced affinities lent statistical relevance to the calculation of Kendall's tau and the slope of affinity predictions. Unfortunately, with a prospective data set of a fixed size, one cannot control these properties. Many of the binding affinities in the prospective SAMPL3 challenge were separated by energies lower than the predictive limit of computational methods (optimistically 1 kcal/mol for intermolecular interactions in aqueous environment). While prospective experiments remain the gold standard for proving the effectiveness of a method, for blinded predictions, where insight and understanding trump unbiased proof of effectiveness, carefully designed retrospective datasets are most helpful.

One of the primary purposes of blinded challenges is to reduce the biases that enter the computational literature via retrospective analysis. For instance, scientists publish [9] and reviewers accept [10] successes much more readily than failures (publication bias). Further, in retrospective studies, hidden operational parameters are often introduced by re-running results with varied parameter sets until they generate suitable "predictions". Blind challenges mitigate these two important sources of bias. Simply putting forth blind or prospective data, however, does not entirely remove bias. While the SAMPL overview papers avoid publication bias by reporting all submitted data, there are at least three remaining source of obvious bias. First, the most successful participants each year publish their work while the less fortunate demonstrate less consistency. Likewise, the single retrospective work in this special issue may not have been pursued for publication had the results been poor. More subtly, one group mentioned in their publication that an ancillary algorithm used for setup performed quite well in making final predictions on the SAMPL data.

While the result is honest and valid, it is unlikely the authors would have mentioned the ancillary algorithm in print had they noticed poor results. Despite SAMPL's specific design to avoid bias, it is likely that the SAMPL3 special issue, like most scientific literature, gives a more optimistic view of computational methods than one would get by using the tools prospectively.

References

1. Geballe MT, Guthrie, JP (2012) The SAMPL3 blind prediction challenge: transfer-energy overview. J Comput Aided Mod Des. doi:10.1007/s10822-012-9568-8
2. Muddana HS, Daniel Varnado C, Bielawski CW, Urbach AR, Issacs L, Geballe MT and Gilson MK (2012) Blind prediction of host-guest binding affinities: a new SAMPL3 challenge. J Comput Aided Mod Des. doi:10.1007/s10822-012-9554-1
3. Newman J, Dolezal O, Fazio V, Cradoc-Davies T and Peat TS (2012) The DINGO dataset: a comprehensive set of data for the SAMPL challenge. J Comput Aided Mod Des. doi:10.1007/s10822-011-9521-2
4. Trapane TL, Lattman EE (2007) Seventh meeting on the critical assessment of techniques for protein structure prediction. Proteins: Struct Funct Bioinform 69(S8):1–2
5. Nielsen JE, Gunner MR, Garcia-Moreno EB (2011) The pK(a) cooperative: a collaborative effort to advance structure-based calculations of pK(a) values and electrostatic effects in proteins. Proteins 79:3249–3259
6. Bardwell DA, Adjiman CS, Arnautova YA, Bartashevich E, Boerrigter SXM, Braun DE, Cruz-Cabeza AJ, Day GM, Della Valle RG, Desiraju GR, van Eijck BP, Facelli JC, Ferraro MB, Grillo D, Habgood M, Hofmann DWM, Hofmann F, Jose KVJ, Karamertzanis PG, Kazantsev AV, Kendrick J, Kuleshova LN, Leusen FJJ, Maleev AV, Misquitta AJ, Mohamed S, Needs RJ, Neumann MA, Nikylov D, Orendt AM, Pal R, Pantelides CC, Pickard CJ, Price LS, Price SL, Scheraga HA, van de Streek J, Thakur TS, Tiwari S, Venuti E, Zhitkov IK (2011) Towards crystal structure prediction of complex organic compounds—a report on the fifth blind test. Acta Crystallographica B 67:535–551
7. Shi Y, Wu C, Ponder JW, Ren P (2010) Moltipole electrostatics in hydration free energy calculations. J Comput Chem 32:967–977
8. Nicholls A, Mobley DL, Guthrie JP, Chodera JD, Bayly CI, Cooper MD, Pande VS (2008) Predicting small-molecule solvation free energies: a blind challenge test for computational chemistry. J Med Chem 51:769–779
9. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R (2008) Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med 358:252–260
10. Emerson GB, Warme WJ, Wolf FM, Heckman JD, Leopold SS, Brand RA (2011) Testing for the presence of positive-outcome bias in peer review: a randomized controlled trial. Arch Int Med 171:1213–1214