Dr. Julien Michel
School of Chemistry,
The University of Edinburgh,
Edinburgh, EH9 3FJ, UK
julien.michel@ed.ac.uk
+44 (0)131 650 4797

Edinburgh, April 19th 2017

Dear David,

I am writing to in strong support of your proposal to generate high quality datasets for future SAMPL blind challenges. I lead a multidisciplinary research group in molecular modelling and molecular biophysics at the University of Edinburgh, United Kingdom. Our group develops software and methodologies to quantity the structure, dynamics and thermodynamics of protein-ligand interactions.

I have followed with interest SAMPL challenges since their very beginning. My group participated for the first time in the 5th iteration in 2015. The experience has been invaluable and had a significant impact on the way my lab validates computational methodologies. The reason SAMPL has been so effective is that blinded predictions eliminate any unconscious biases that can affect validation of a prediction when the answer is already known to the developer. In addition SAMPL competitions challenge developers to test their new methodologies on difficult, yet tractable, systems that pave the way for robust molecular modelling of even more complex protein-ligand datasets. In that sense SAMPL is highly complementary and synergistic with NIH initiative such as D3R.

To illustrate, through our previous participation in SAMPL5 we have discovered shortcomings in current methodologies to model protonation state changes upon transfer of drug-like molecules between different phases. We have also discovered the importance of finite-size effects on electrostatic interactions in host-guest recognition. In the latter case this has challenged us to come up with a new simulation protocol that deals more effectively with this problem, and we will publish soon a manuscript reporting our findings. In another positive development, we have recently started applying our improved simulation protocol to a very challenging protein-ligand dataset. I doubt I would have pursued this system without our previous success on a SAMPL5 host-guest dataset. This illustrates why SAMPL datasets are incredibly useful to the community even after the blinded competitions are over.

I firmly believe that repeated rounds of SAMPL competitions will dramatically advance the predictive power of molecular modelling, enabling computation as a cornerstone of the future science of molecular design. I wish you the best of luck with your proposal.

Best wishes,