

INTRODUCTION

The most profound criticism of the original proposal was that we did not propose to fund new computational methodology development, instead focusing on fielding blind challenges to enable other groups to innovate, assess, and cross-evaluate new methodologies, concluding “as such, the overall impact . . . is expected to be limited and would rely largely on others to make progress.” We acknowledge that our goal in this proposal is indeed to drive methods development widely in the field, rather than to funnel the data collection effort solely into our own methods development efforts. Indeed, the SAMPL challenge builds on a *proven* model leveraging blind challenges to motivate rapid community progress: this model has not only been shown to drive progress in computational biology by the CASP [? ? ?] and DREAM [? ? ?] challenges, but also in machine learning by the Netflix Prize [?] and innovation in general by the XPrize approach [103? , 104]. We now include a discussion of how this model of “crowdsourcing for innovation” has achieved widespread success in diverse fields [? ?] and is known to foster much more innovation than having groups work independently on their own problems [? ? ?]. With 99 papers arising from previous unfunded iterations of SAMPL, we have also proven the ability of SAMPL to drive innovation [105–204]; with NIH funding, overall impact will be greatly amplified. To assess how the biomolecular simulation *community* views the historical and future impact of SAMPL, we conducted a survey between April and June 2017 [?], and have included testimonials from several past participants highlighting different ways in which SAMPL has had impact by helping drive methodology development in their groups. Collectively, SAMPL’s publication record [105–204], community survey feedback [?], and the attached testimonials provide overwhelming evidence to support our view that our proposal for SAMPL will have high impact.

Aside from impact, reviewers were concerned that the data sets generated will not be large enough to provide a dramatically larger amount of training data for machine learning methods. We fully agree: SAMPL is about *evaluating* rather than *training* methods. As with past crowdsourcing efforts, SAMPL focuses on providing a venue to find out what works and does not work in a blind predictive setting. Existing datasets may be adequate for training, but it is also easy to overfit or unintentionally apply bias when studying existing data, so blind challenges like this (as Reviewer 2 noted) are particularly important by allowing all methods (both highly empirical methods relying extensively on training to existing data and physical methods like those we tend to prefer) to compete on equal footing. We now make this more clear in the proposal in our section on crowdsourcing.

Reviewers were also concerned that the real impact on the field would be minimal, in part because of the lack of methodological innovation. To address this, a variety of key leaders in the field have provided support letters highlighting specific ways in which SAMPL has helped advance the science already. We also believe the roughly 100 (typically well-cited) publications about the SAMPL series of challenges also support this point.

To assist with this revision, we surveyed the SAMPL community regarding our plans [?], and found overwhelming support. Out of 44 respondents (each typically representing an individual group) 95% saw SAMPL as a valuable resource to the community (with the remaining percentage seeing it as “somewhat valuable”) and more than 90% seeing it as important in driving progress in the field (with no respondents seeing it as unimportant). 91% are happy with the proposed future directions, and the remaining 9% had modest suggestions for refinements which we have incorporated, such as modest increases in dataset sizes, additional experimental follow-up measurements to address lingering questions in discrepancies with theory, and increased use of third-party testing. **[JDC: What do you mean by “third-party testing”? Outsourced measurements, or evaluation of tools by third parties?]** Full survey results are available online [?], and we can make respondent identities and contact info available if needed.

The reviewers also critiqued the perceived lack of participation in SAMPL, noting that only 20 groups participated in the last cycle (SAMPL5). We now address this in the proposal itself, but it is worth noting that the protein-ligand binding component of prior SAMPL challenges was subsumed into D3R for SAMPL5, and despite this, overall participation *continued to increase* relative to previous SAMPL iterations. The 20 groups who participated include more than 100 co-authors all over the world, and this occurred even without any significant publicity effort for the SAMPL challenge and with ongoing uncertainty as to whether any future SAMPL challenge will ever occur. Without stable funding, it has always been uncertain whether we can cobble together new experimental data for another iteration of the challenge and when this might occur, making it difficult to advertise and plan many cycles ahead.

Finally, the reviewers also objected that we have no apparent connection with the SAMPL challenge and that SAMPL’s presence on the D3R website is not that extensive. We have updated the D3R website to clearly indicate the SAMPL organizers (including Mobley and Chodera), though having a substantial web presence for SAMPL is to some extent impaired by our lack of funding.

Overall, we hope the overwhelming community support for our plans, the specific examples we now give of how SAMPL drives innovation in the field, and the ample precedent for our “crowdsourcing” approach to science

innovation (as we now discuss in the proposal), coupled to our improvements to how we explain our ideas in the proposal itself, demonstrate that SAMPL will indeed drive progress in the field even more than it already has.