INTRODUCTION

The reviewers of our original proposal raised a number of concerns we have addressed in the attached revised proposal and in this letter, as we detail below.

Overall, reviewers had several main points of criticism, the most profound of which is that our proposal will not actually develop any new methods, but instead relies on others to do the methodological innovations, thus they concluded that "as such, the overall impact... is expected to be limited and would rely largely on others to make progress." We agree with the reviewers that this work's goal is indeed to drive methods development in the field, rather than to produce our own methods developments. Indeed, the SAMPL challenge builds on a *proven* model of blind challenges to motivate progress much more broadly than one single group – a model already proven in computational biology by the CASP [ref] and DREAM [ref] challenges, but also proven at a more popular level by the Netflix challenge [ref] and to some extent the XPrize approach [ref]. This model of "crowdsourcing for innovation" has achieved widespread success in diverse fields [refs], as we now discuss in the proposal, and is known to foster much more innovation than having groups work independently on their own problems [ref]. Thus we, and the SAMPL community, disagree with the reviewers and believe the impact of SAMPL has been high and, with funding, will grow to be even higher. We now attach (and reference in the proposal) the results of a survey of of SAMPL's impact on the field, as well as testimonials from a number of past participants highlighting ways in which SAMPL has helped method development work in their groups. We hope these testimonials, together with more context on the crowdsourcing model employed here, will alleviate the concerns of the reviewers.

Reviewers were also concerned that the data sets generated will not be large enough to provide a dramatically larger amount of training data. We fully agree: SAMPL is about *testing* not *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 it is overwhelmingly supportive. Out of 45 respondents (each typically representing an individual group) 95.6% 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 to our directions, such as increasing the size of the datasets modestly, doing additional experimental follow-up after SAMPL challenges, and increased use of 3rd party testing. We discuss some of these new innovations in our proposal; typically these will require expanding our planned work and make funding even more of an imperative.

The reviewers also critiqued the perceived lack of participation in SAMPL, noting that only 20 groups participated this last cycle. We now address this in the proposal itself, but it is worth noting that this performance *is increased* relative to previous SAMPL iterations since the protein-ligand binding component of prior SAMPL challenges has now been subsumed into D3R. Additionally, 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 continual uncertainty as to whether any future SAMPL challenge will ever occur. Without funding, it is always uncertain whether we can even have another iteration of the challenge, let alone when it will occur – so it is difficult for us to advertise, and for researchers to plan on participating.

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.