UC Irvine
Mobley Lab Datasets

Title
Advancing predictive modeling through focused development of model systems to drive new modeling innovations

Permalink
https://escholarship.org/uc/item/7cf8c6cr

Authors
Mobley, David L
Chodera, John D
Isaacs, Lyle
et al.

Publication Date
2016-10-05

License
CC BY-NC-ND 4.0
ABSTRACT

This work seeks to advance quantitative methods for biomolecular design, especially for predicting biomolecular interactions, via a focused series of community blind prediction challenges. Physical methods for predicting binding free energies, or “free energy methods”, are poised to dramatically reshape early stage drug discovery, and are already finding applications in pharmaceutical lead optimization. However, performance is unreliable, the domain of applicability is limited, and failures in pharmaceutical applications are often hard to understand and fix. On the other hand, these methods can now typically predict a variety of simple physical properties such as solvation free energies or relative solubilities, though there is still clear room for improvement in accuracy. In recent years, competitions and crowdsourcing have proven an effective model for driving innovations in diverse fields. In our field, blind prediction challenges have played a key role in driving innovations in prediction of physical properties and binding, especially in the form of the SAMPL series of challenges. Here, we will continue and extend SAMPL prediction challenges to include new physical properties, more complicated host-guest binding data, and application to biomolecular systems. Carefully selected systems and novel experimental data will provide challenges of gradually increasing complexity spanning between systems which are now tractable to those which are marginally out of reach of today’s methods but still slightly simpler than those covered by the Drug Design Data Resource (D3R) series of challenges on existing pharmaceutical data. We will work with D3R to run blind challenges on the data we generate and to ensure it is designed to maximally benefit the field.

In Aim 1, we will collect new measurements on partitioning, distribution, and protonation of drug-like compounds, in collaboration with partners in the pharmaceutical industry. In Aim 2, we leverage our expertise in host-guest binding to generate new data on host-guest binding in cucubiturils and deep cavity cavitands. And in Aim 3, we use high-throughput robotic experiments to generate new protein-ligand binding data of biological relevance. Aim 4 focuses on using this data in the SAMPL series of challenges, applying proven crowdsourcing-based techniques to drive the development of new methods and new understanding of the strengths and weaknesses of existing techniques. We will also run reference calculations with the latest techniques.

This work will ensure the continued success of SAMPL challenges which have already driven considerable innovation in the field and been the focus of 100 different publications (each typically cited 5-50 times) since their inception around 2007, and will play a key role in driving the next several generations of improvements in computational techniques for molecular design. The research proposed here will lead to significant improvements in the predictive power of physical models for drug discovery, molecular design and the prediction of physical properties.
SPECIFIC AIMS

While computational techniques are widely used in pharmaceutical drug discovery, current generation technologies (such as docking) are unsuitable for true molecular design. Specifically, these techniques fail to predict small molecule binding affinities to target and antitarget biomolecules with sufficient accuracy for many applications. While virtual screening techniques can improve enrichment, they lack the accuracy to guide molecular design. A new generation of physical techniques, alchemical free energy calculations, are poised to fill this void by providing quantitative, predictive tools useful in multiple stages of the drug discovery pipeline.

Recent success of alchemical methods in predicting accurate affinities has sparked considerable enthusiasm, but the domain of applicability of these techniques is currently highly limited; broad application and routine use will require further evaluation, refinement, and development. There is a vast gulf between targets within the domain of applicability and those which are outside it, frustrating progress. Bridging this gulf will require focused study of carefully selected systems of intermediate complexity. Without such a bridge, these techniques may encounter the same problems faced by docking and related techniques: routine failure without clear insights into why, and years to decades spent making small methodological modifications without dramatic improvements in predictive power.

We propose a carefully designed, crowdsourcing-based model to drive innovation to bridge this gap. Specifically, we will collect targeted experimental datasets that highlight and focus the community on specific modeling difficulties, fielding blind prediction Challenges to spur new, crowdsourced methodology innovation from the community. This model has been proven to drive progress in a given area, both inside the computational biology community [1–6], and for even more visible projects like Netflix [7] and the XPrizes [8–10]. Here, we design an extension of the Statistical Assessment of Modeling of Proteins and Ligand (SAMPL) series Challenges, which allow a wide variety of different approaches to compete on equal footing on the same data, fostering rapid innovation, quick recognition of the best new technologies, and dissemination of these to the community. The data we generate will allow several Challenges of differing difficulty, between systems tractable with current methodologies and the pharmacologically-relevant drug targets featured in the NIH-funded D3R effort [11]. Central to this proposal is collecting targeted data and running blind Challenges to drive progress, but the data itself will have an even longer-term impact on the community. Our Aims are:

Aim 1. Collect new physical property datasets to assess accuracy and spur improvements in force fields and modeling of protonation states and tautomers.

We will develop new solution-phase datasets for druglike small molecules. These data can test critical aspects of small molecule modeling (including accounting for interactions and treatment of protonation/tautomeric state) and improve our ability to predict physical properties relevant to drug discovery in new regions of chemical space. We will initially focus on aqueous/nonpolar distribution coefficients and pKa measurements, advancing to solubilities and membrane permeabilities, while using these data to drive improvements in the modeling of ligand interactions.

Aim 2. Measure affinities of drug-like compounds in supramolecular hosts to challenge quantitative models of binding in systems not plagued by major receptor sampling issues.

We will measure new host-guest binding free energies (using cucurbiturils and deep-cavity cavitands as hosts) to field binding Challenges with varying complexity between physical property prediction and protein-ligand binding. Host guest systems are some of the simplest cases of molecular recognition, and thus these binding data will drive improvements in modeling of simple binding systems with techniques of relevance to drug discovery.

Aim 3. Develop model protein-ligand systems that isolate specific modeling challenges of drug targets.

We will identify suitable biological protein-ligand model systems that isolate individual modeling challenges (selected to push the limits of physical techniques) and develop these for blind Challenges based on new protein-ligand affinity measurements. While the initial year will feature fragment binding to human serum albumin, subsequent Challenge systems will be selected using a novel informatics platform to focus on timely modeling issues.

Aim 4. Crowdsolve innovation via community blind Challenges to advance biomolecular design.

The data collected in Aims 1–3 will drive annual SAMPL Challenges, allowing the field to test the latest methods and force fields to assess progress, compare them against one another head-to-head, and perform sensitivity analysis to learn how much different factors (protonation state, tautomer selection, solvent model, force field, sampling method, etc.) affect predictive power. Results will then feed back into improved treatment of these factors for subsequent Challenges, driving cycles of application, learning, and advancement—an approach proven to work in a variety of previous crowdsourcing approaches to science, as well as in prior iterations of SAMPL itself.

Overall, these SAMPL Challenges provide a catalyst to drive the next several generations of computational methodology innovation, development, and assessment for pharmaceutical drug discovery.
We propose an iterated Challenge-based approach to drive methodological innovation via crowdsourcing. Widespread availability of inexpensive graphics processing units (GPUs) provides a 100-fold increase in price-to-performance ratio over CPUs, while advances in automation and sampling protocols have helped simulation-based techniques reach the point where they now begin to be genuinely useful in guiding drug discovery for a limited domain of applicability. Specifically, in some situations, free energy calculations appear to be capable of achieving RMS errors of 1–2 kcal/mol with current force fields, even in prospective applications, sufficient to drastically reduce the number of molecules that must be synthesized and assayed. As a consequence, pharmaceutical companies have rapidly adopted these methods in active discovery projects.

Despite progress, current modeling methods suffer from severe limitations hindering their widespread use in molecular design. For example, even “small” protein conformational changes not gracefully handled by current methodologies can yield errors up to 5 kcal/mol in calculated binding free energies, force field limitations still pose major challenges, and the inability to treat important chemical effects like protonation and tautomer equilibria drastically limits the domain of applicability. For many pharmaceutically relevant systems, the most important sources of error—and modeling challenges—are not yet clear.

These hurdles frustrate progress in the field. Method developers focus on addressing some specific challenges, but new methods are tested on disparate datasets making it difficult to gauge progress and identify valuable methods when they are developed. And applications projects tend not to take advantage of the latest methods—in part because their utility is not yet clear. As a result, valuable methods can languish relatively unused in the literature for many years without impacting applications in drug discovery.

We propose an iterated Challenge-based approach to drive methodological innovation via crowdsourcing. A series of community blind prediction Challenges will test and push the limits of predictive techniques, providing a bridge between challenging but tractable problems and pharmacologically relevant but currently intractable problems. Our Statistical Assessment of Modeling of Proteins and Ligands (SAMPL) Challenges take advantage of proven crowdsourcing-based approaches to science (discussed below), allowing head-to-head comparisons of existing methods and fostering development of new methods. Regular, progressive Challenges provide an environment for rapid innovation from “unconventional innovators” from diverse fields who form collaborative teams, applying and integrating distinct approaches.

Tractable Challenges inspire rapid progress, while Challenges that are too difficult accomplish little.
We require high quality Challenges carefully designed specifically to maximize community learning by pushing the limits of current technology. While the Drug Design Data Resource (D3R [112]) also fields blind Challenges, these rely on contributed experimental protein-ligand datasets from now-defunct pharmaceutical discovery programs [112]. D3R Challenges provide a necessary benchmark for real-world performance, but provide only modest new insight to drive improvement because a multitude of issues are be responsible for poor performance [137].

Methodology improvement is most rapid when complex goals are achieved via tractable intermediate milestones which can be independently assessed using high quality data [106]. Thus, we focus on using Challenges to drive and evaluate solutions to specific component problems that make it difficult to predict biomolecular interactions. The Challenge format allows the entire community to learn from both failure and success. Much as unit testing is indispensable for identifying software bugs when complex integration tests fail, Challenges like SAMPL are valuable in pinpointing modeling errors, especially when projects like D3R highlight overall performance problems. Hence, SAMPL Challenges focus on predicting properties that isolate likely sources of error, such as physical properties of small molecules (hydration free energies, aqueous tautomer ratios, partition or distribution coefficients between aqueous and nonpolar phases) or binding in systems of reduced complexity (such as host-guest complexation, or binding of fragments to trypsin and HIV integrase).

The SAMPL series of Challenges has already proven it can drive dramatic improvements in modeling methodologies. SAMPL, born out of frustration with the lack of venues for comparing predictive accuracy on a level playing field, was initiated by Anthony Nicholls of OpenEye software in 2007/2008 [139], and has run Challenges approximately every two years since then [138, 140–148]. Governance transitioned to an unfunded academic collaboration among the PI and co-Investigators during SAMPL3 in 2012; this collaboration ran SAMPL4 (2014), SAMPL5 (2016), currently SAMPL6 (2017–2018). The PI of this proposal (Mobley) was a primary organizer of SAMPL4–6 (2014–present). SAMPL has demonstrated its utility as a valuable community resource, resulting in ~100 publications typically cited 5–50 times or more each [1–100]. SAMPL datasets are so valuable to the community that they foster retrospective tests long after the Challenges end (e.g. [147–150]).

Until now, SAMPL has relied on donated data and time, making it impossible to design Challenges specifically to improve modeling as we propose here. Despite these limitations, our recent survey reveals that SAMPL has already had a significant effect on the field; past participants are overwhelmingly enthusiastic about SAMPL and the plans we propose here [111]. One of our most frequent requests is for modestly larger datasets, which we propose here. Our proposed Challenges bridge the gap between calculations of simple physical properties that isolate force field inaccuracies from slow conformational sampling inherent to protein-ligand systems—such as hydration free energies, which can already be calculated fairly accurately [138]—and the D3R Grand Challenges on protein-ligand binding, which are a major source of community consternation [112, 151–153]. Unless this gap is bridged, there is a very real possibility that modeling will continue to fall far short of expectations in pharmaceutical systems for reasons which remain unclear [116].

Our major goal is to rapidly advance predictive modeling to enable it to guide biomolecular design; extending the SAMPL Challenges will accomplish this. This work will play a vital role in enhancing the work being done on existing data by D3R, helping prepare methods for application to D3R’s pharmaceutical Challenges.

INNOVATION

This work proposes a five-year plan for fielding SAMPL Challenges as an innovation engine for molecular design. Specifically, collaborative scientific competitions or Challenges are a proven model for inducing innovation [105, 106, 109], applying crowdsourcing to finding and identifying robust new methodologies [106]. The proposed series of SAMPL Challenges brings crowdsourcing to bear on molecular interactions in a way which focuses on solving specific, incremental problems that impair our ability to accurately and robustly predict these interactions. Thus, while the proposed work is indeed innovative (as we argue below), the most important innovation comes from participants in SAMPL Challenges, as is the case in all crowdsourcing approaches to innovation [106, 109].

Reliance on others for innovation is a strength of this model. It is not always obvious where the most important
innovations may come from, but Challenges draw together diverse entrants, including unconventional innovators who may apply unorthodox, risky, or radical technologies [106, 109]; progress comes both from the usual suspects and unexpected sources. It is not necessary to decide in advance who is best positioned to solve key problems [106]—the Challenge itself drives innovation and allows it to be recognized, regardless of source. Crowdsourcing models for innovation have an ample track record, with the XPrize driving major headlines [108–110], and the Netflix Prize also familiar to many [107]. While there are other Challenges in the area of biomolecular modeling, such as D3R [112], the pK_a cooperative [154], CAPRI [155] and CASP [101], **no other blind Challenge focuses specifically on datasets collected specifically to drive quantitative protein-ligand modeling.**

SAMPL is designed in the spirit of the wildly successful biomedical DREAM Challenge [104–106] which had more than 800 recent participants. As recommended by the DREAM Challenge organizers [106], SAMPL focuses on high-quality data, decomposing complex problems into tractable but difficult component problems in order to benefit from crowdsourcing, and drawing together diverse teams and methods to allow various solutions to be evaluated.

**Our recent survey indicates SAMPL has already been important in advancing protein-ligand modeling** [111]; attached testimonials highlight some specific ways SAMPL has advanced science. Several historical examples further serve to highlight how SAMPL can foster innovation (see also Figure 3 and our SAMPL bibliography). While the first SAMPL Challenges focusing on hydration free energies saw heterogeneous performance, they highlighted pitfalls of existing methods and force fields which led to marked improvements in PB models [140, 156, 157], identified some limitations of fixed-charge force fields [158, 159] that spurred their repair [158–160], and helped motivate alternate implicit or hybrid solvent models [161–163]. Shifts in protonation state and tautomer proved particularly important in SAMPL5’s log D Challenge [145, 164], as they will in binding. Host-guest binding Chal-
challenges [165] have highlighted the importance of salt effects [146, 165, 166] and in some cases revealed severe force field limitations [167, 168], pointing the way forward for improving predictive models [165, 169].

This proposal focuses on the development of SAMPL Challenges tailored to progressively advance computational methods for binding (Aim 4), gathering the requisite experimental data to drive these Challenges (Aims 1-3). **While the primary innovation will result from our Challenges, we also innovate on several other fronts.** Aim 3 develops a new informatics platform to facilitate the rapid identification and study of useful protein-ligand systems that are both experimentally tractable for high-throughput biophysical measurement and focus on specific challenges of interest; an innovative, automated wetlab helps produce our high quality data. Additionally, students in our computational groups perform reference calculations to benchmark current state-of-the-art techniques. This provides student training, but also drives innovation by (1) Benchmarking advanced method entries against our reference “current best practices” calculations; (2) Facilitating learning by allowing others to compare against our reference results to determine how methodology differences impact results; and (3) Focusing the field on key issues by analyzing sensitivity to conditions such as ionic strength, protein conformational sampling and specific ligand-receptor interactions. Many methods performed poorly in predicting log D in SAMPL5, with even the best methods having accuracies less than would be expected based on hydration free energies [145], pointing the way forward for improving predictive models [146], revealing the impact that targeted follow-up experiments (such as those we will conduct here) can have on improving models. Thus, targeted Challenges on solution-phase properties address some of the complexities of binding.

**APPROACH**

A series of data-driven Challenges will spur systematic advances in modeling for biomolecular design. We collect targeted experimental datasets focused on specific problems spanning a range of complexity (Aims 1–3), and use these to field SAMPL Challenges driving innovation (Aim 4). Data collection brings together multiple laboratories and industrial collaborators, including both theorists and experimentalists: graduate students from the Mobley and Chodera laboratories are paired with well-equipped experimental groups in industry to collect physical property data (Aim 1); Gibb and Isaacs, leading experimentalists in supramolecular chemistry, work with theorist Mobley to perform host-guest affinity measurements (Aim 2); and the Chodera lab applies new automated approaches to identify suitable protein-ligand systems and measure binding (Aim 3).

**Aim 1: Collect new physical property datasets to assess accuracy and spur improvements in force fields and modeling of protonation states and tautomers.**

Simple physical properties such as solvation, phase partitioning, and protonation equilibria can be calculated precisely (but not necessarily accurately) with physical methods, allowing quantitative comparison between calculations and experiment and revealing and isolating deficiencies in our models. These properties allow us to directly probe force field accuracy and chemical effects like protonation and tautomer handling in the absence of slow conformational changes and other effects which complicate assessment in protein-ligand systems.

**Rationale:** We will perform new solution-phase physical property measurements for drug-like molecules to motivate improvements in force fields and handling of protonation states and tautomers. This builds on our work on water-cyclohexane distribution coefficients (log D) for SAMPL5 (with Genentech), which revealed major issues with handling of protonation states and tautomers [145] as well as serious force field limitations [160] (Figure 3). Distribution coefficients give the equilibrium ratio of solute concentration between aqueous and nonpolar phases at a given pH. Thus, because these mimic transfer free energies from aqueous to protein- or membrane-like environments, they capture many of the same characteristics as binding but without the associated challenges of receptor conformational sampling and specific ligand-receptor interactions. Many methods performed poorly in predicting log D in SAMPL5, with even the best methods having accuracies less than would be expected based on hydration free energies [145], yet failures were informative and the major sources of error were issues which also plague ligand-receptor binding. In many respects, distribution coefficients are an ideal SAMPL Challenge, as they are sufficiently difficult for clear failures to be frequent, with ample room for improvement, but not so difficult that reasons for failure were generally unclear. Still, many methods consistently disagreed with experiment for some compounds [145, 160, 164, 171], revealing the impact that targeted follow-up experiments (such as those we will conduct here) can have on improving models. Thus, targeted Challenges on solution-phase properties address some of the complexities of binding.
Previous physical property SAMPL Challenges have already driven significant progress, but a frequent request is for modestly larger datasets and follow-up experiments to validate specific measurements [111]. This was impossible with donated data, but here, we will expand the size and quality of Challenge datasets, allowing us to expand to provide greater coverage of the full dynamic range. **Follow-up experiments will become routine, improving our ability to learn from failures.** For example, SAMPL5’s lack of dynamic range meant that, when calculated values often spanned a larger dynamic range than experiment, it was unclear if this was due to the choice of compounds, experimental limitations, or force field deficiencies [145, 160, 164, 171].

**Physical property datasets will be tailored for synergy with binding datasets.** Specifically, ligands studied in host-guest and protein-ligand binding Challenges will be prioritized for inclusion in solution-phase physical property datasets; if physical property data is already public for these, analogs or fragments will be included. Thus, if a binding free energy proves difficult to predict, participants can determine if its solution-phase properties (such as protonation states or tautomers) were similarly difficult, thereby narrowing down the source of the error. Below, we summarize plans for data collection for the SAMPL7-11 Challenges of Aim 4 (see also Figure 2). These specific datasets were selected to help the field first resolve problems encountered in SAMPL5 and then progress to accurate estimation of additional properties.Datasets will be larger than prior SAMPL Challenges—at least 96 compounds for good statistics—though when possible, a much larger amount of data will be collected.

**SAMPL7: Cyclohexane/water and octanol/water distribution coefficients, given pKa.** Building on the success of distribution coefficients at motivating rapid modeling advances [145], we will measure cyclohexane-water distribution coefficients at pH 7.4 for a new batch of commercially-available drug- and fragment-like molecules overlapping with compounds studied in Aims 2 and 3. We will also measure octanol-water distribution coefficients for the same compounds, given indications that their prediction may be computationally tractable [172, 173] despite the heterogeneous structure of the wet octanol phase [174]. Because pK_a prediction was difficult but critical for SAMPL5, we will focus SAMPL7 on force fields and tautomers by measuring and providing pK_a values, revisiting pK_a prediction in SAMPL8 and 9. A similar approach is currently being applied for SAMPL6 (data collection is in process in collaboration with Merck), so plans here may be adjusted modestly depending on SAMPL6 outcomes.

**SAMPL8 and SAMPL9: Distribution coefficients and pKa’s for drug-like molecules.** Distribution coefficient measurements conflate the challenging issues of protonation state and tautomer prediction, as well as transfer into different environments. After allowing participants to predict distribution coefficients given pK_a values for SAMPL6-7, the SAMPL8-9 datasets will include distribution coefficients and pK_a values, with participants asked to predict both. The datasets will include pK_a values and distribution coefficients for an extensive set of drug-like molecules. SAMPL8 will focus on octanol-water and cyclohexane-water, and SAMPL9 will shift to alternate solvents to ensure that models are capable of handling varied non-water phases. As in SAMPL7, we plan to focus partly on molecules and fragments overlapping with compounds being studied in Aims 2–3.

**SAMPL10-11: Solubility prediction and membrane permeability.** With solubility predictions now becoming tractable [122–124] (with Schrödinger also working on amorphous solubility), solubility measurements will be a valuable test for SAMPL10, combining the solvation aspects of SAMPL1-8 with a new solid phase component. New computational techniques are targeting membrane permeability [126, 127], and this is experimentally accessible (see support letters from Pfizer and Merck), leading to our interest in permeability for SAMPL11. However, it is likely that both solubility and permeability will still be extremely challenging—so we plan to also measure distribution coefficients and pKa’s for the same compounds to help participants isolate potential points of failure.

Based on community feedback [111], we are also interested in subsequent SAMPL Challenges on hydration and conformational free energies, and will seek suitable sources of reliable experimental data.

**Experimental plan:** Experimental data will be collected in collaboration with our pharma partners, roughly following the model used for SAMPL5, where Chodera lab student Bas Rustenburg worked with Genentech and adapted an existing workflow to measure cyclohexane-water log D [171]. A similar approach is currently being employed with Merck for SAMPL6. Thus, Mobley and Chodera lab students will visit industry partners collecting targeted datasets. This collaborative approach (see Letters of Collaboration from Genentech, Pfizer, GSK, and Merck) gives us substantial access to equipment and high-throughput measurement workflows—such as the Sirius T3 (which can measure partition/distribution coefficients, pK_a’s, and solubilities for molecules with titratable groups)—automation equipment, and compound libraries—for the purposes of rapidly collecting targeted datasets. This approach works [171], and our partners see the value of this data and SAMPL to the community.

Overall, Aim 1 extends prior SAMPL Challenges via data focused on quantitative prediction of physical properties relevant to accurately predicting biomolecular interactions, paving the way to applications in more complex systems.
Aim 2: Measure affinities of drug-like compounds in supramolecular hosts to challenge quantitative models of binding in systems not plagued by major receptor sampling issues.

Moving beyond solution-phase physical properties, we want Challenge data which introduces some of the additional complexity of binding in a controlled manner. Aim 1 focuses on the behavior of small molecules in different environments, in the absence of receptors and the associated potential for slow sampling, strong specific interactions, and other challenges like salt effects. Binding in host-guest systems (Aim 2) retains many of these same challenges and introduces strong specific interactions and other issues like salt effects [165], while still avoiding many of the issues with slow sampling (of protein conformational changes, ions, and ligand binding modes) seen in biomolecular interactions. Thus, binding in host-guest systems introduces new challenges relevant to biomolecular interactions, but without the full complexity of protein-ligand interactions, as recently reviewed [165].

Already, host-guest SAMPL Challenges have provided key tests for modeling of binding interactions [165], resulting in new attention paid to how co-solvents and ions modulate binding (resulting in errors of up to 5 kcal/mol when these effects are neglected) and the importance of adequately sampling water rearrangements [144, 146, 165, 176]. This new attention has resulted in clear improvements (Figure 4), though host-guest binding remains difficult to model accurately [177], in part due to force field limitations (resulting in new force field work [169]).

Here, we design a series of SAMPL Challenges focused on two classes of host-guest systems—cucurbiturils and analogs (SA 2.1) and Gibb’s deep-cavity cavitands (GDCCs, SA 2.2)—both of which build on prior SAMPLs. These two sets of systems exhibit different complexities [165], with the hosts of 2.1 bringing modest co-solvent and ion effects and some receptor sampling problems for the acyclic hosts, and the GDCCs of 2.2 bringing profound ion and co-solvent effects as well as water sampling challenges. Methods which perform well on one class may not perform well on the other [165] because of these distinct challenges. This diversity and complexity is important for Challenges which seek to drive researchers to focus on all of the important features rather than just a subset [106]. As the SAMPL community would like even more diversity in hosts and guests [111], we will also seek to include additional host-guest systems via data donation (as in SAMPL1-5).

SAMPL6 is currently in progress and includes components on GDCCs and cucurbiturils, so our plans in this section may be adjusted somewhat depending on the outcomes of SAMPL6.

Subaim 2.1: Cucurbituril-based receptors: model binding systems

Cucurbituril derivatives for host-guest binding. Building on previous success with cucurbit[n]uril (CB[n]) experiments for SAMPL [178–180], we will conduct a series of new experiments on these receptors, with experimental work conducted by co-investigator Isaacs, an expert on these systems who provided data for previous SAMPLs. CB[n] receptors are particularly well suited to our goals because they exhibit: (1) strong binding affinities in water, comparable to protein-ligand affinities (routinely μM to nM; occasionally pM to fM) [181–187], (2) high selectivities between structurally related guests which translate into large ΔΔG values [188], (3) low molecular weights (1–2 kDa) permitting high levels of theory to be used, and (4) highly restricted conformational degrees of freedom, reducing conformational sampling challenges often seen in protein-ligand binding. For SAMPL7-11, we will resynthesize a series of CB[n]-type receptors of increasing complexity, measure \( K_a \) values, and determine host-guest stoichiometry and geometry toward pharmaceutically relevant guests (selected drugs) in order to stringently test methods for predicting binding. Figure 5 shows the chemical structures of three hosts—\( \text{Me}_2\text{CB}[8] \) [189], glycoluril hexamer [190], and acyclic CB[n]-type receptors [191–196] which span a range in terms of level of preorganization and formal charge.

### Table 1. Selected drugs whose binding to CB[n] hosts will be assayed for SAMPL7, 9, and 11 Challenges (SA 2.1).

<table>
<thead>
<tr>
<th>Drug</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>memantine</td>
<td>adamantane; 1:1</td>
</tr>
<tr>
<td>saxaglpltin</td>
<td>adamantane; 1:1</td>
</tr>
<tr>
<td>premarin</td>
<td>steroid</td>
</tr>
<tr>
<td>pancuronium</td>
<td>steroid</td>
</tr>
<tr>
<td>varenicline</td>
<td>1:1 vs 1:2</td>
</tr>
<tr>
<td>valsartan</td>
<td>( pK_a ) 4.37</td>
</tr>
<tr>
<td>omeprazole</td>
<td>( pK_a ) 4.77</td>
</tr>
<tr>
<td>ranolazine</td>
<td>( pK_a ) 7.17; epitopes</td>
</tr>
<tr>
<td>pradaa</td>
<td>( pK_a ) 3.87; epitopes</td>
</tr>
<tr>
<td>nilotinin</td>
<td>epitopes; ( pK_a ) 6.3</td>
</tr>
<tr>
<td>senspar</td>
<td>epitopes; folding</td>
</tr>
<tr>
<td>vyyance</td>
<td>diamine; epitopes; folding</td>
</tr>
<tr>
<td>minocycline</td>
<td>tetracyclin; amino anline</td>
</tr>
</tbody>
</table>

Figure 4. The best host-guest binding predictions of SAMPL3 [175] and SAMPL5 [167]. Binding free energy predictions have shown clear improvements from SAMPL3 to SAMPL5 as the major obstacles become understood and are treated better by models, though a systematic offset remains in the best SAMPL5 predictions (yellow). Dashed lines denote errors of ±1.5 kcal/mol.
we will expand on the range of hosts by including e.g., which binding epitope is complexed) will be addressed by ITC the rim of the binding pocket in octa-acid derivative. However, the four benzoate groups are relocated from the extreme exterior in the case of Novel deep cavity hosts probe the effects of binding site charge constellations. issues complicate protein-ligand interactions as well. protonation state complexities [issues with host conformational sampling, slow water rearrangements, salt/buffer condition-dependence, and binding. Relative to cucurbiturils, the GDCCs introduce new complexities because of their tight exit portal, modest a range of 9 carboxylate guests, and (2) the binding of 6 carboxylate and trimethylammonium guests to both hosts (1, R = Me). These studies used ITC to measure the thermodynamics of (1) host GDCC hosts: the octa-acid 1 (R = H) and another octa-acid variant with four methyl groups at the portal of the history of GDCC SAMPL challenges. Subaim 2.2. Gibb deep cavity cavitands for host-guest studies History of GDCC SAMPL challenges. During SAMPL4 [208] and SAMPL5 [209] we focused on two specific GDCC hosts: the octa-acid 1 (R = H) and another octa-acid variant with four methyl groups at the portal of the binding pocket (1, R = Me). These studies used ITC to measure the thermodynamics of (1) host 1 (R = H) binding a range of 9 carboxylate guests, and (2) the binding of 6 carboxylate and trimethylammonium guests to both hosts (1, R = H and Me; Figure 6). In both cases 1H-NMR titration was also used to confirm ITC-derived free energies of binding. Relative to cucurbiturils, the GDCCs introduce new complexities because of their tight exit portal, modest issues with host conformational sampling, slow water rearrangements, salt/buffer condition-dependence, and protonation state complexities [146, 165]. Thus GDCC-oriented Challenges are particularly important since these issues complicate protein-ligand interactions as well.

Novel deep cavity hosts probe the effects of binding site charge constellations. For future GDCC datasets, we will expand on the range of hosts by including 2 and 3 in our ITC studies (Figure 6). Like cavitand 1, host 2 is an octa-acid derivative. However, the four benzoate groups are relocated from the extreme exterior in the case of 1, to the rim of the binding pocket in 2. This is expected to have a direct effect on the binding of charged guests as well as

Figure 5. SAMPL7-11 host-guest Challenges will feature cucurbituril hosts and analogs, including Me₄CB[8], glycoluril hexamer, and acyclic CB[n]-type receptors (SA 2.1). These receptors bind a variety of drug-like molecules, some with high affinity. For SAMPL7, we will measure $K_a$ and $\Delta H$ values, stoichiometry, and geometry for the interaction of Me₄CB[8] (a soluble CB[8] derivative) with 15 guests (selected top drugs, Table 1) by either direct or competition isothermal titration calorimetry (ITC), UV/Vis or fluorescence indicator displacement assay, or NMR competition experiments, as previously [180–182, 197]. Our selection of Me₄CB[8] binding to top drugs allows us to modulate the computational complexity by: 1) Changing host flexibility (e.g. Me₄CB[8] can exhibit ellipsoidal deformation) [189], 2) allowing the possibility of binary or ternary (e.g. 1:1 and/or 1:2 host:guest) complexes [198–200], 3) using drugs with several potential binding epitopes or modes to induce sampling issues. Host:guest stoichiometry and geometry (e.g., which binding epitope is complexed) will be addressed by ITC $n$ values, Job plots monitored by UV/Vis or NMR [201], and by 1H NMR complexation induced changes in chemical shifts [202]. All studies will be conducted in phosphate buffered saline (pH 7.4 with physiological salt) which introduces its own complexities due to salt competition for binding [165, 203]. SAMPL8 will revisit the same host, but use 15 different guests be selected from commercial sources on the basis of reference calculations (on a larger set of guests) to ensure that they cover substantial dynamic range and/or exhibit affinities that depend substantially on the force field or water model, thus effectively testing our force fields and methods. For SAMPL9, we will focus on binding of the same 15 drugs (Table 1), but to glycoluril hexamer. This host introduces the complication of increased conformational dynamics, and influences the number and energy of solvating (and unusually coordinated) water molecules implicated in the high binding constants for CB[n]-guest complexes [187, 204]. The selected drugs include several with pKₐ values in the 3.8 to 7.4 range; given that CB[n]-type receptors (like biomolecular receptors) can induce pKₐ shifts in their hosts of up to 4 pKₐ units [205–207], this will test how well models can predict these effects. Additionally, it will couple nicely with the focus on pKₐ values in Aim 1 – especially so given that Aim 1 compounds will include some of the same chemical functionality. SAMPL10 will revisit glycouril hexamer with the same 15 guests from SAMPL7. SAMPL11 will shift to acyclic CB[n]-type receptors (e.g. M1C₉, M1C₁₀, and M1PC₀ that contain anionic solubilizing groups attached via different linker lengths. As in SAMPL3 [175], these acyclic CB[n]-type receptors introduce conformational complexity, and water interactions play a key role.
an indirect effect on guest complexation via changes to the solvation of the empty host. Octa-trimethylammonium cavitand ("positand" 3) has the same overall architecture as host 1, but inverts the charges on the water solubilizing exterior coat. While it is not yet clear if this switch in groups relatively remote from the pocket will directly affect guest complexation, results from related systems suggest it can (unpublished).

Figure 6. Gibb deep cavity cavitands for the SAMPL7-11 datasets (SA 2.2). These hosts bind a variety of carboxylate and trimethylammonium guests in a strongly salt-dependent manner, providing a stringent test of our ability to model salt-dependent binding.

**SAMPL7-11 deep cavity cavitand datasets.** Data for SAMPL7 will focus on how well the effect of host carboxylate substituent location can be predicted, and will involve hosts 1 and 2 with a set of at least five previously uninvestigated guests. Guests will be selected from commercial sources on the basis of reference calculations in a similar manner to SAMPL8 in Subaim 2.1, specifically picking guests which have broad dynamic range and, here, have marked differences in affinities between hosts. SAMPL8 will provide a second iteration of this experiment to test algorithmic improvements in predictive modeling following SAMPL7 by comparing hosts 1 and 3 with a different set of guests. We anticipate that because of the relative remoteness of the charged groups in these two hosts, the effects of switching charges will be subtler than the differences between 1 and 2. SAMPL9 will consider the effect of common biologically-relevant counterions/salts on guest binding, comparing the effects of NaCl and NaI on the complexation of at least five guests to 1. We have previously shown that iodide has a weak affinity for the binding pocket of 1, while sodium ions have an affinity for the outer carboxylates [210], requiring modeling to capture the differential affinities of these ions in addition to guest affinities to successfully model the observed affinities. SAMPL10 will follow up on this by examining the effects of these same two salts on the complexation of at least five guests to 3, again giving the modeling community time to incorporate algorithmic improvements following SAMPL9. While we have not yet quantified salt affinities to host 3, we expect the iodide to have affinity for both the pocket and the positively charged solubilizing groups. For SAMPL11 we will consider the effects of co-solvents on the binding of five guests to 1 and 2 to probe the effect of co-solvent competition for the binding site, as well as effects co-solvents may have on weakening the hydrophobic effect. Participants frequently request larger datasets, so every effort will be made to include additional guests beyond the minimal number proposed if time allows or if human time can be reduced (such as via automated calorimetry). Regardless, the total number of binding affinities measured for the family is substantial, so the data will be of considerable value as a benchmark [165].

**Aim 3: Develop model protein-ligand systems that isolate specific modeling challenges of drug targets.**

To drive real improvement in quantitative modeling of protein-ligand interactions, we need to be able to revisit the same systems again and again in order to gauge and drive progress—in much the same way as participants in the Netflix Challenge had to genuinely improve their ability to predict user feedback to succeed [107]. While D3R [112] evaluates the accuracy of current protein-ligand binding methods, performance in each Challenge iteration varies widely depending on the nature of the donated pharmaceutical data and difficulty of the target. The high complexities of D3R targets make it difficult to identify clear points of failure [112, 151–153]. For example, while kinases are targets of great interest to drug discovery, blind Challenges involving kinase targets confound issues of slow protein conformational dynamics [211], protonation state effects of protein [212] and ligand [213, 214], phosphorylation, charged ligands, and other challenges; failures on such targets are thus often unexplained.

Only by revisiting the same systems repeatedly will we benefit from the wisdom of the crowds in identifying specific problems and validating solutions to these problems. Thus, here, we identify and develop model protein-ligand systems to isolate specific accuracy-limiting effects in a series of SAMPL Challenges. By developing model binding systems—biological protein-ligand systems comprised of single-domain proteins binding to a simple ligand series free of complex phenomena—we can study systems of complexity intermediate between completely artificial systems (like the valuable T4 lysozyme L99A model binding site [135, 165, 215]) and complex pharmaceutical targets where multiple modeling challenges make it difficult to learn from failure (Figure 8A). This allows Challenges focused on identifying and evaluating multiple solutions to selected accuracy-limiting effects (such as how to deal with ligand and protein protonation-state issues [216], slow protein conformational dynamics, etc.), with the ability to revisit the same systems repeatedly as needed to drive innovation.
Crowdsourced innovation enhances the potential of model systems for driving progress. For example, SAMPL3 focused on charged fragments binding to trypsin [220], and rapidly focused the field on the deficiencies of current alchemical free energy methodologies in treating the binding of charged ligands. Within two years, multiple laboratories developed solutions to this problem [221–223].

**SAMPL7-11 model protein-ligand Challenges.** We will introduce a new model protein-ligand system each SAMPL (revisiting the prior SAMPL’s system if this becomes too difficult), with multiple Challenges on each system (Figure 2) to allow iterative improvement and assessment. Our SAMPL7 data will focus on the binding of small soluble drug fragments to one particular protein (below), with fragments also studied in Aim 1. However, maximizing gains in this area requires adapting subsequent Challenges based on deficiencies identified by prior Challenges, so our new informatics platform (below) will aid in system selection.

**SAMPL7: Assessing predictive modeling of binding to multiple weak sites via measuring fragment binding to human serum albumin (HSA).** HSA, the most abundant blood plasma protein, has a remarkable ability to bind a variety of small molecule drugs in multiple binding sites (Figure 7B) [224]. Using HSA as a model receptor isolates the challenge of multiple weak ligands binding to a stable rigid protein that is also pharmacologically relevant because of its dramatic modulation of drug pharmacokinetics [217]. HSA has at least eight known binding sites, with numerous crystal structures available for drugs binding to two predominant sites (Site I and II) [217]. Small soluble molecules resembling drug fragments are highly likely to bind to HSA (≥90% of such fragments, as detected by SPR [225]), providing an experimentally-tractable diverse set of ligands spanning several orders of magnitude in affinity [225]. As many current physical modeling methods assume a single well-defined binding site with high affinity [226], this Challenge will allow the isolation of the effect of weak multiple binding from the majority of other confounding factors in protein-ligand binding. HSA is relatively rigid, and computational methods already show some promise in computing binding affinities to HSA [217, 227, 228], making this an optimal model system. Recombinant HSA will be expressed in *E. coli* and purified via refolding from inclusion bodies [229], then defatted at low pH [230] to ensure the resulting protein is free of the glycosylation and bound fatty acids found in plasma-isolated HSA [230]. Recombinant expression will allow mutant forms of HSA (engineered via single-primer mutagenesis) to be fielded for SAMPL8 (Figure 7). We will obtain a diverse library of at least 96 soluble drug-fragment-like molecules in pre-plated format as dry compound, and use our automated ITC pipeline (Figure 7C,D) to characterize overall binding affinities to HSA. The same ligands pre-plated in DMSO format will be used to conduct a separate set of fluorescence titration assays (monitoring tryptophan fluorescence quenching, Figure 7D).
and competition assays with site-specific fluorescent probes (Figure 7B) to resolve site-specific affinities to Sites I and II. We will field several levels of Challenges, including Challenges focused on affinities to Sites I and II, as well as Challenges focused on predicting overall affinity and stoichiometry.

**SAMPL8-11: Rapid development of new tailored model systems using a novel informatics platform.** We are developing a novel informatics platform, TargetExplorer, to identify protein targets that can be rapidly developed into tractable model systems focusing on individual challenges (Figure 8). This tool filters all known protein targets with structural data available in the PDB, first selecting for experimental tractability, then annotating experimentally tractable targets to determine which targets possess (or are likely to be free of) specific challenges for physical modeling. This will allow us to select systems isolating specific modeling challenges.

The Chodera lab has developed an automated wetlab to facilitate the development of such systems using bacterial expression (see Equipment and Facilities). Potential targets matching desired challenge criteria will be screened for bacterial expression using high-throughput cloning, transformation, and expression testing, with purity and yield assessed by capillary electrophoresis on a Caliper GXII. Targets will be screened for stability in various buffers using Thermofluor thermal shift assays [233]. Ligands identified via TargetExplorer as spanning a wide dynamic range of binding affinities will be purchased as dry powder stocks and prepared for assay by highly accurate gravimetric solution preparation techniques using a Quantos automated balance. Our lab has access to a wide variety of biophysical techniques for measuring protein-ligand binding affinities, including fluorescence (if fluorescent probe ligands are available), absorption (e.g. Soret band shifts), automated isothermal titration calorimetry (provided ligands are sufficiently soluble), surface plasmon resonance, microscale thermophoresis (MST), luminescence, and alphascreen; all except MST are fully automated.

We take a twofold approach to developing Challenge datasets: First, we will purchase and assay small molecules similar to known ligands, presuming that these molecules are likely to have measurable affinities. Second, using single-primer quick-change mutagenesis, we will introduce site-directed mutants to modulate the binding affinities of known ligands. This can be performed and screened for expression in 96-well format. Thus, datasets will consist of a matrix of protein mutants and ligands, providing opportunity to deeply explore the effects of interest.

**Aim 4: Crowdsourcing innovation via community blind Challenges to advance biomolecular design.**

The heart of our work lies here – driving innovation via community blind SAMPL Challenges centered around the datasets of Aims 1—3 (Figure 2). This approach utilizes crowdsourcing (see Innovation, above) to advance science. Our Challenges test the state of the art, provoke new methodological and force field innovations, allow comparative evaluation of methods so that advances are quickly recognized and spread, and drive downstream improvements. Our SAMPL Challenges are progressive, building on one another so that for success in later Challenges, participants must build on lessons learned from prior Challenges. Each iteration will likely yield its own incremental benefits (e.g., as in Figure 3) for molecular design, in addition to contributing to progress towards accurate prediction of biomolecular interactions.

**SAMPL blind Challenges.** Yearly SAMPL Challenges following our timeline (Figure 2) serve to advance the state-of-the-art. As is critical for such Challenges [106], we will advertise aggressively via press releases, commentaries
in relevant journals, CCL, e-mail, and Twitter. As experimental data for each component becomes available and is curated, input files and Challenge details will be made available at least six months prior to the Challenge deadline; data not yet available at that time will be held for a subsequent Challenge (with the exception of three months for year 1 due to startup timescales). Advertising will feature our full timeline (which will also be available at https://drugdesigndata.org/about/sampl along with details of plans) allowing planning of participation. Challenge datasets will now also be larger than prior SAMPLs, as frequently requested [111].

As in prior SAMPLs, submissions will be handled by a web upload service on the SAMPL website which validates submissions to ensure that they meet format standards for automated processing and analysis. Analysis will be conducted by our automated Python framework, and results returned automatically online. All participant submissions and methodology descriptions will (as before) be made available publicly on the website, along with participant information (except for participants who specifically request to remain anonymous prior to submission). Aggregate statistics and historical performance will also be made available on our website, along with a record linking publications to historical submissions.

As the project progresses, we will push to transition towards submission of methods rather than of predictions. Specifically, as automated workflow engines such as OpenEye’s Orion, AutoDesk’s MDT, and others begin to be adopted on a larger scale, we will push participants to submit their methods in a standardized package, and automatically apply their submitted method to the Challenge data via cloud computing resources. This is already done in some Challenge platforms [106] but will be new to SAMPL, resolving one major hurdle for participation. Participants, while enthusiastic about SAMPL and how it benefits science, often note the burden of participating [111]; submission of methods rather than results will solve this problem, allowing participation with minimal investment of human time. This will also allow for easy archival of methods and a focus on testing workflows rather than human expertise. We will work with the NSF-funded Molecular Sciences Software Sustainability Institute (MoSS) to identify best practices in this respect. One hurdle will be obtaining computer time, but we will seek donations of time from cloud computing vendors, support from pharma, and adapt our workflows so we can run on NSF supercomputing resources where we will seek grants. In the worst case we would ask participants submitting methods for a modest budget to run their methods, but having this level of automation results in substantial savings in human time which can itself result in monetary savings. This migration will also facilitate post-Challenge follow-up and method comparison, by allowing us to archive reproducible methods. Technology already allows this sort of archival and reproducibility, such as via Docker [234] containers with the participants’ executable programs [106].

**We will leverage SAMPL Challenges to identify modeling failures and identify potential solutions.** The learning process begins prior to each Challenge and continues afterwards. In advance, we provide guidance to participants as to what known modeling issues we expect may be relevant. For a host-guest system, for example, we might highlight known buffer/salt effects, protonation state challenges, and point out previous work on sampling challenges, with pointers to the relevant experimental work and to modeling work from past SAMPL Challenges and elsewhere [165]. This helps participants design their approach. Additionally, we will run reference calculations using current best practices. This serves several purposes: It provides a test of the current methods and force fields; it helps facilitate learning—we announce what calculations we plan to perform, make input files available in a wide variety of formats [145, 146, 235], and others can repeat our calculations with with method or force field variations to test protocol differences; it provides training to our students who are involved (as these reference calculations provide them familiarity with the latest methods and make them instrumental in community learning); and it allows sensitivity analysis, as by varying the conditions (protonation state, tautomer, etc., [145]) we can see how much this impacts calculated values and thereby how important each factor is. Reference calculations have, for example, helped us highlight the importance of a small amount of water in cyclohexane for accurately calculating log D values, show how an incorrect tautomer could affect calculated values by many log units [145], and discover that small force field modifications could significantly improve results on hydration free energies [138].

**Dissemination of Challenge results.** The most critical phase of each Challenge occurs after results are released. Typically, some participants uncover specific problems or solutions which are missed by others, so dissemination of these insights becomes critical to drive improvements. SAMPL workshops will occur after Challenges every two years (years 1, 3, and 5), co-hosted with D3R Grand Challenge workshops (see support letter) During the off years, discussion and dissemination of results will be via asynchronous means (below) and a “virtual workshop” consisting of talks and interaction over Google Hangouts or YouTube Live. For all workshops, virtual or otherwise, talks will be selected from abstracts written to describe what each participant learned from participating—allowing us to prioritize presentations from those who have learned a great deal or done exceptionally well, rather than those who are simply testing an existing method with no new insight gained. If the D3R effort is not renewed, we will run SAMPL workshops independently, controlling costs via the model we use for our Workshops on Free Energy Methods in Drug Design—specifically, most participants will pay their own way to the workshop, and we
will seek pharmaceutical and software industry sponsorships to defray costs.

Rapid dissemination of results is critical so that new insights can be used in subsequent Challenges. We will continue to publish special issues of the *Journal of Computer Aided Molecular Design* (JCAMD) collecting publications related to each year’s SAMPL Challenges (see Letter of Support). As requested by participants [111], we will also begin providing summaries of the key results and insights from each SAMPL that are suitable for a broad, non-technical audience – either in JCAMD or as commentary in another suitable journal. To ensure immediate availability of reports, we will strongly encourage prepublication sharing of results and analysis, including both slides and posters from SAMPL meetings (via F1000 Research) and paper preprints (via bioRxiv).

Post-Challenge analysis also drives progress, and often occurs outside of formal workshops and meetings. While this has happened in the past—for example, when participants using similar methods work together after the SAMPL meeting to identify the origin of these discrepancies [146, 165, 176, 236, 237]—we hope to accelerate this kind of collaboration. To facilitate rapid exchange of ideas and open communication between the community, we will use collaboration software—such as Slack, which facilitates scientific communication for the NASA/JPL Mars Rover teams and NSF antarctic scientific research teams—to build a community discussion platform, facilitating a process of learning from one another more rapidly than normal publication channels. Given the overlap between molecules/fragments used in Aim 1 and those studied in Aims 2–3, insights from one Challenge component will have implications for predictions in the other components, making this rapid exchange key. The students supported in this project will also play a role, working to understand the methods employed and bringing together researchers using similar methods to try and understand performance differences, including by running follow-up calculations.

**Each Challenge will have lasting effects.** SAMPL datasets are collected primarily to drive Challenges, but both the datasets and the lessons learned have lasting value. The lessons learned will have the most immediate near-term impact and are critical for progress because each Challenge builds on prior iterations. However, the datasets themselves are also key. In the past, due to the unfunded nature of SAMPL, we primarily emphasized compilation of the datasets and conducting Challenges. However, SAMPL datasets achieved a life of their own outside of SAMPL, becoming standard test sets [147, 148, 165, 238, 239]. This is one way SAMPL Challenges achieve their lasting effect, so we will perform additional curation on datasets post-Challenge, with follow-up experiments (as often requested by participants [111]) when needed, then release the data publicly as standard benchmark sets [165]. We will make the data (including primary data, processed data, and our analysis of Challenge submissions) available freely and publicly with permanent, citeable DOIs; ensure relevant data is deposited in standard community repositories (e.g. BindingDB [240]); and guarantee data longevity via backup hosting on library archival facilities (such as the UC’s DASH [https://dash.cdlib.org/]).

**COLLABORATION MANAGEMENT PLAN**

We have a strong previous history of successful collaboration, with Mobley and Chodera having co-authored roughly a dozen publications and organized several workshops and other initiatives. Mobley, Isaacs, and Gibb have also worked together to coordinate past SAMPL Challenges, and Mobley and Gibb a previous NSF workshop. PI Mobley will oversee the project, with teams for the other aims (Aim 1: Mobley & Chodera; Aim 2.1: Isaacs; Aim 2.2: Gibb; Aim 3: Chodera; Aim 4: Mobley & Chodera) involving the other co-investigators as needed. Meetings will consist of a monthly Google Hangout and a yearly in-person planning meeting. Chodera and Mobley will communicate more frequently due to the interlinked nature of their work. Publications are expected to be largely dictated by the overall Timeline, with an experimental publication associated with each Challenge component being prepared for distribution to participants along with their results. Conflict resolution is expected to be straightforward, but if any serious difficulties arise, Michael Gilson (UCSD) will arbitrate given our close connections with D3R.

**OUTLOOK**

Physical methods have been slow to achieve their promise in binding prediction, in part because truly significant innovations are so hard to recognize due to a lack of standard tests and benchmarks, and in part because of an “applications first” approach which seems to plague our community where we rush to apply our methods to problems of pharmaceutical relevance without ensuring they can tackle simpler, better-understood problems first. Here, we extend the successful series of SAMPL blind Challenges to crowdsource innovation focused on a series of key component problems centered around our major goal, quantitative prediction of biomolecular interactions. Collection of high quality, focused datasets will spur method innovation, beginning with relatively simple physical property prediction and progressing to challenging problems in biomolecular recognition via a series of carefully designed intermediate steps. Building on SAMPL’s strong prior track record, the proposed series of tailored Challenges will focus community innovation on critical problems we can resolve in the near term, resulting in dramatic improvements in computational molecular design.
Full List of SAMPL References


Bibliography and References Cited


D3R | What We Have Learned. https://drugdesigndata.org/about/what-we-have-learned.


