

# Predicting binding free energies: Frontiers and benchmarks (a perpetual review)

David L. Mobley,<sup>1,\*</sup> Germano Heinzelmann,<sup>2,†</sup> Niel M. Henriksen,<sup>3,2,‡</sup> and Michael K. Gilson<sup>3,§</sup>

<sup>1</sup>*Departments of Pharmaceutical Sciences and Chemistry,  
University of California, Irvine, CA, USA, 92697*

<sup>2</sup>*Departamento de Física, Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brazil 88040-900*

<sup>3</sup>*Skaggs School of Pharmacy and Pharmaceutical Sciences,  
University of California, San Diego, CA, USA, 92092*

(Dated: August 23, 2017)

Binding free energy calculations based on molecular simulations provide predicted affinities for biomolecular complexes. These calculations begin with a detailed description of a system, including its chemical composition and the interactions between its components. Simulations of the system are then used to compute thermodynamic information, such as binding affinities. Because of their growing promise for guiding molecular design, these calculations have recently begun to see widespread applications in early stage drug discovery. However, many challenges remain to make them a robust and reliable tool. Here, we highlight key challenges facing these calculations, describe known examples of these challenges, and call for the designation of standard community benchmark test systems that will help the research community generate and evaluate progress. In our view, progress will require careful assessment and evaluation of new methods, force fields, and modeling innovations on well-characterized benchmark systems, and we lay out our vision for how this can be achieved.

**Manuscript version 1.3 pre-release.** See <https://github.com/mobleylab/benchmarksets> for all versions.

**Permissions.** Portions of this work are posted with permission from the *Annual Review of Biophysics*, Volume 46 (c) 2017 by Annual Reviews, <http://www.annualreviews.org/>. The right to create derivative works is also executed here with permission from the *Annual Review of Biophysics*.

## I. INTRODUCTION

Molecular simulations provide a powerful technique for predicting and understanding the structure, function, dynamics, and interactions of biomolecules. Often, these techniques are valued because they provide a way to actually watch what might be going on at the atomic level. However, simulations also can be used to make quantitative predictions of thermodynamic and kinetic properties, with applications in fields including drug discovery, chemical engineering, and nanoengineering. A thermodynamic property of particular interest is the binding affinity between biomolecules and ligands such as inhibitors, modulators, or activators. With accurate and rapid affinity predictions, we could use simulations in varied health-related applications, such as the prediction of biomolecular interaction networks in support of systems biology, or rapid design of new medications with reduced side-effects and drug resistance. In this work, we give a view of how these simulations could impact drug discovery, briefly discuss where they stand now, and then argue for benchmark systems chosen to drive and assess

the advancement of these methods, helping to make them practical for drug discovery.

### A. Imagining a tool for drug discovery

A major aim in the development of molecular simulations is to create quantitative, accurate tools which will guide early stage drug discovery. Consider a medicinal chemist in the not-too-distant future who has just finished synthesizing several new derivatives of an existing inhibitor as potential drug leads targeting a particular biomolecule, and has obtained binding affinity or potency data against the desired biomolecular target. Before leaving work, he or she generates ideas for perhaps 100 new compounds which could be synthesized next, then sets a computer to work overnight prioritizing them. By morning, the compounds have all been prioritized based on reliable predictions of their affinity for the desired target, selectivity against alternative targets which should be avoided, solubility, and membrane permeability. The chemist then looks through the predicted properties for the top few compounds and selects the next ones for synthesis. If synthesizing and testing each compound takes several days, this workflow compresses roughly a year's work into a few days.

While this workflow is not yet a reality, huge strides have been made in this direction, with calculated binding affinity predictions now showing real promise [30, 31, 41,

---

\* [dmobley@mobleylab.org](mailto:dmobley@mobleylab.org)

† [germano.heinzelmann@ufsc.br](mailto:germano.heinzelmann@ufsc.br)

‡ [nhenriksen@ucsd.edu](mailto:nhenriksen@ucsd.edu)

§ [mgilson@ucsd.edu](mailto:mgilson@ucsd.edu)

[46, 144, 185, 214, 223], solubility predictions beginning to come online [123, 163, 182], and predicted drug resistance/selectivity also apparently tractable [119], with some headway apparent on membrane permeability [37, 115]. A considerable amount of science and engineering still remains to be done to make this vision a reality, but, given recent progress, the question now seems more one of *when* rather than *whether*.

## B. Increasing accuracy will yield increasing payoffs

Recent progress in computational power, especially the widespread availability of graphics processing units (GPUs) and advances in automation [125] and sampling protocols, have helped simulation-based techniques reach the point where they now appear to have sufficient accuracy to be genuinely useful in guiding pharmaceutical drug discovery, at least for a certain subset of problems [30, 41, 93, 139, 185, 214, 223]. Specifically, in some situations, free energy calculations appear to be capable of achieving RMS errors in the 1-2 kcal/mol range with current force fields, even in prospective applications. As a consequence, pharmaceutical companies are beginning to use these methods in discovery projects. The most immediate application of these techniques is to guide synthesis for lead optimization, but applications to scaffold hopping and in other areas also appear possible.

At the same time, it is clear that not all situations are so favorable, so it is worth asking what level of accuracy is actually needed. It is often suggested that we need binding free energy predictions accurate to within  $\sim 1$  kcal/mol, but we are not aware of a clear basis for this figure beyond the fact it is a pleasingly round number that is close to the thermal kinetic energy,  $RT$ . Instead of setting a single threshold requirement for accuracy, it is more informative to consider how accurate calculations must be to reduce the number of compounds synthesized and tested by some factor, relative to the number required without computational prioritization. If one targets a three-fold reduction, the answer appears to be that calculations with a 2 kcal/mol RMS error will suffice [144, 191]. Thus, one can gain substantial benefit from simulations that are good yet still quite imperfect.

More broadly, though, this analysis does not address the net value of computational affinity predictions in drug discovery. Costs include those of the software, computer time, and personnel required to incorporate calculations into the workflow; while benefits include the savings, revenue gains, and externalities attributable to reducing the number of low-affinity compounds synthesized and arriving earlier at a potent drug candidate. In addition, with sufficiently reliable predictions, chemists may choose to tackle difficult synthesis efforts they otherwise might have avoided, resulting in more novel and valuable chemical matter.

## C. Overview of free energy calculations

The present review focuses on a class of methods in which free energy differences are computed with simulations that sample Boltzmann distributions of molecular configurations. These samples are usually generated by molecular dynamics (MD) simulations [104], with the system effectively coupled to a heat bath at constant temperature, but Monte Carlo methods may also be used [35, 135, 136]. In either case, the energy of a given configuration is provided by a potential function, or force field, which estimates the potential energy of a system of solute and solvent molecules as a function of the coordinates of all of its atoms. Such simulations may be used in several different ways to compute binding free energies or relative binding free energies, as detailed elsewhere [29, 32, 136, 190] and summarized below. In all cases, however, the calculations yield the free energy difference between two states of a molecular system, and they do so by computing the reversible work for changing the initial state to the final one. Two broad approaches deserve mention.

The first general approach directly computes the standard binding free energy of two molecules via evaluating the reversible work of transferring the ligand from the binding site into solution. (This is sometimes called an absolute binding free energy calculation.) The pathway of this change may be one that is physically realizable, or one that is only realizable *in silico*, in which case it is sometimes called an “alchemical” pathway. Physical pathway methods provide the standard binding free energy by computing the reversible work of, in effect, pulling the ligand out of the binding site. Although, by definition, the pathway used must be a physical one that could occur in nature, it need not be probable, and improbable pathways, governed by an order parameter specifying how far the ligand is from the binding site, are often used [14, 89, 94, 213, 228, 238]. In addition, artificial restraints may be useful to avoid sampling problems in the face of often complex barriers along the pathway [14, 89, 94, 213, 228]. By contrast, alchemical pathway methods artificially decouple the ligand from the binding site and then recouple it to solution from the protein [16, 81, 91, 103, 140]. Although alchemical decoupling methods may avoid clashes of the ligand with the protein that might be problematic when pathway methods are applied to a protein with a buried binding site, they still can pose some of the same sampling challenges. For example, sampling of the unbound receptor must be adequate after the ligand is removed, and water molecules must have time to equilibrate in the vacated binding site. Given that free energy is a state function, it is not surprising that alchemical and physical pathway approaches yield apparently comparable results when applied to the same systems [44, 86, 118, 237].

The second general approach computes the difference between the binding free energies of two different ligands for the same receptor, by computing the work of arti-

ficially converting one ligand into another, first in the bound state and then free in solution [29, 32, 136, 209]. Because these conversions are not physically realizable, such calculations are, again, called alchemical. These calculations can be quite efficient if the two ligands are very similar to each other, but they become more complicated and pose greater sampling problems if the two ligands are very different chemically or if there is a high barrier to interconversion between their most stable bound conformations [125]. In addition, there may be concerns about slow conformational relaxation of the protein in response to the change in ligand. Nonetheless, alchemical relative free energy calculations currently are the best automated and most widely used free energy methods [125, 144, 223].

The accuracy and precision of all of these methods are controlled by three major considerations. First, many conformations typically need to be generated, or sampled, in order to obtain an adequate representation of the Boltzmann distribution. In the limit of infinite sampling, a correctly implemented method would yield the single value of the free energy difference dictated by the specification of the molecular system and the chosen force field. In reality, however, only finite sampling is possible, so the reported free energy will differ from the nominal value associated with infinite sampling. In addition, because sampling methods are typically stochastic and the dynamics of molecular systems are sensitive to initial conditions [4], repeated calculations, using different random number seeds or initial states, will yield somewhat different results. The problem of finite sampling is most acute for systems where low-energy (hence highly occupied) conformational states are separated by high effective barriers, whether energetic or entropic. Second, even if adequate sampling is achievable, free energy differences may disagree substantially with experiment if the force field is not sufficiently accurate. Third, errors may also arise if the representation of the system in the simulation does not adequately represent the actual system, e.g. if protonation states are assigned incorrectly and held fixed.

#### D. Challenges and the domain of applicability

Thus, in order for a free energy calculation to be reliable, it must use an appropriate representation of the physical system and an accurate force field, and it must adequately sample the relevant molecular configurations. In the case of the more widely used alchemical relative free energy approach, this means that the best results are expected when:

- a high quality receptor structure is available, without missing loops or other major uncertainties
- the protonation state of the ligand and binding-site residues (as well as any other relevant residues) can reliably be inferred
- the ligand binding mode is defined by crystallographic studies and is not expected to change much

on modification

- the receptor does not undergo substantial, slow conformational changes
- key interactions are expected to be well-described by underlying force fields

Beyond this domain of applicability—whose dimensions are, in fact, still somewhat vague [1]—substantial challenges may be encountered. For example, binding free energy calculations for a cytochrome C peroxidase mutant suggest limitations of fixed-charge force fields. In this case, the strength of electrostatic interactions in a buried, relatively nonpolar binding site appears to be overestimated by a conventional fixed-charge force field, likely due to underestimation of polarization effects [177]. Sampling problems are also common, with slow sidechain rearrangements and ligand binding mode rearrangements in model binding sites in T4 lysozyme posing timescale problems unless enhanced or biased sampling methods are carefully applied [19, 71, 98, 141, 143, 222]; and larger-scale protein motions induced by some ligands can also pose challenges [19, 121].

Although such problems need not prevent free energy calculations from being used, they can require specific adjustment of procedures and parameters based on experience and knowledge of the system at hand. Thus, a key challenge for the field is how to use insights from well-studied cases to enable automation and reduce the detailed knowledge of each system required to carry out high quality simulations.

Troubleshooting is also a major challenge. In most cases where calculations diverge substantially from experiment, the reason for the discrepancy is not apparent. Is the force field inaccurate? Would the results improve with more sampling? Were protonation states misassigned—or do they perhaps even change on binding? There might even be a software bug [52] or a human error in the use of the software. As a consequence, it is not clear what steps are most urgently needed to advance the field as a whole. In this work, we argue that many of these problems can be alleviated, and that the field will advance more rapidly, if we select a set of well-chosen benchmark systems on which free energy methods are regularly tested.

#### E. Improving modeling by cycles of testing, prediction, and improvement

Modeling can in some cases improve rapidly, but, in our experience, rapid advances are most common when computational models undergo regular cycles of improvement, predictive testing, learning, and then further improvement. This can be particularly difficult for academic groups which may not have the resources for predictive tests; however, these are essential, since it is only in predictive tests that we can be sure we are assessing the performance of a method as it works in real life, rather than relying on knowledge of the expected outcome to

inform setup of the calculations. With this in mind, the Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) blind challenges, as well as the Community Structure Activity Resource (CSAR) challenge, later replaced by the Drug Design Data Resource (D3R) grand challenges, have arisen to meet part of this need. Currently, D3R focuses on running blind challenges on protein-ligand binding with datasets from the pharmaceutical industry, allowing testing and evaluation of computational methods on systems of direct pharmaceutical relevance. SAMPL, in contrast, focuses on predictions in simpler physical settings [142], such as small molecules in aqueous and organic phases, and small molecules binding to supramolecular hosts. Together, the SAMPL and D3R challenges roughly span the spectrum from properties we can predict now (though they may be challenging in some cases [9, 142, 237]) to the drug binding we want to be able to reliably predict. These challenges are vital as they provide our only opportunity, at present, to routinely see how different methods compare when attempting to compute the same properties, and they provide the beginnings of a model for how we can best advance free energy techniques: routinely testing our methods on the same, well-understood systems to learn what does and doesn't work to improve performance. Thus, we need not just blind tests, but retrospective testing on well-understood, "benchmark" systems.

## II. THE NEED FOR WELL-CHOSEN BENCHMARK SYSTEMS

Although tests of individual free energy methods are not uncommon today [30, 41, 139, 214, 223], the use of nonoverlapping molecular systems and computational protocols makes it difficult to compare methods on a rigorous basis. In addition, few studies are designed to identify key sources of error and thereby focus future research and development. A few molecular systems have now emerged as *de facto* standards for general study (Section III). These selections result in part from two series of blinded prediction challenges (SAMPL [153], and CSAR [51] followed by D3R [74]), which have helped focus the computational chemistry community on a succession of test cases and highlighted the need for methodological improvements. However, broader adoption of a larger and more persistent set of test cases is needed. By coalescing around a compact set of benchmarks, well chosen to challenge and probe free energy calculations, practitioners and developers will be able to better assess and drive progress in binding free energy calculations. Our primary goals in this work are to explain how benchmark systems can be used to advance the field, to encourage adoption of a standard set of benchmark systems, and to propose some candidates for this set.

### A. Benchmark types and applications

We envision two classes of benchmark cases: "hard" benchmarks, which are simple enough that well-converged results can readily be computed; and "soft" benchmarks, for which convincingly converged results cannot readily be generated, but which are still simple enough that concerted study by the community can delineate key issues that might not arise in the simpler "hard" cases. The following subsections provide examples of how hard and soft benchmark systems may be used to address important issues in free energy simulations.

#### 1. Hard benchmarks

*a. Systems to test software implementations and usage* It is crucial yet nontrivial to validate that a simulation package correctly implements and applies the desired methods [189], and benchmark cases can help with this. First, all software packages could be tested for their ability to generate correct potential energies for a single configuration of a specified molecular system and force field. These results should be correct to within rounding error and the precision of the physical constants used in the calculations [189]. Second, different methods and software packages should give consistent binding free energies when identical force fields are applied with identical simulation setups and compositions. The benchmark systems for such testing can be simple and easy to converge, and high precision free energies (e.g., uncertainty  $\approx 0.1$  kcal/mol) should serve as a reference. Test calculations should typically agree with reference results to within 95% confidence intervals, from established methods [64, 188]. For this purpose, the correctly computed values need not agree with experiment; indeed, experimental results are unnecessary.

*b. Systems to check sampling completeness and efficiency* As discussed above, free energy calculations require thorough sampling of molecular configurations from the Boltzmann distribution dictated by the force field that is employed. This sampling is typically done by running molecular dynamics simulations, and for systems as large and complex as proteins, it is difficult to carry out long enough simulations. Calculations with inadequate sampling yield results that are imprecise, in the sense that multiple independent calculations with slightly different initial conditions will yield significantly different results, and these ill-converged results will in general be poor estimates of the ideal result obtained in the limit of infinite sampling. Advanced simulation methods have been developed to speed convergence [190, 205], but it is not always clear how various methods compare to one another. To effectively compare such enhanced sampling methods, we need benchmark molecular systems, parameterized with a force field that many software packages can use, that embody various sampling challenges, such as high dimensionality, and energetic and entropic bar-

riers between highly occupied states, but which are just tractable enough that reliable results are available via suitable reference calculations. Again, experimental data are not required, as the main purpose is to test and compare the power of various enhanced sampling technologies.

*c. Systems to assess force field accuracy* Some molecular systems are small and simple enough that current technology allows thorough conformational sampling, and hence well converged calculations of experimental observables. This has long been feasible for liquids [102]; for example, it is easy to precisely compute the heat of vaporization of liquid acetone with one of the standard force fields. More recently, advances in hardware and software have made it possible to compute binding thermodynamics to high precision for simple molecular recognition systems [89], as further discussed below. In such cases, absent complications like uncertain protonation states, the level of agreement with experiment reports directly on the accuracy of the force field. Thus, simple molecular recognition systems with reliable experimental binding data represent another valuable class of benchmarks. Here, of course, experimental data are needed. Ideally, the physical materials will be fairly easy to obtain so that measurements can be replicated, and new experimental conditions, such as different temperatures and solvent compositions, can be explored.

## 2. Soft benchmarks

*a. Systems to challenge conformational sampling techniques* Enhanced sampling techniques (Section II A 1 b), designed to speed convergence of free energy simulations, may not be adequately tested by any hard benchmark, because such systems are necessarily rather simple. Thus, despite the fact that reliable reference results are not available for soft benchmarks, they are still important for method comparisons. For example, it may become clear that some methods are better at sampling in systems with high energy barriers, and others in high-dimensional systems with rugged energy surfaces. Developers should test methods on a standard set of benchmark systems for informative comparisons.

*b. Direct tests of protein-ligand binding calculations* Although it is still very difficult to convincingly verify convergence of many protein-ligand binding calculations, it is still important to compare the performance of various methods for these real-world challenges. Appropriate soft benchmarks are likely to be cases which are still not overly challenging, involving small proteins and simple binding sites. We propose defining a series of benchmark protein-ligand systems that systematically introduce specific challenges. In particular, they should exemplify none, one, two, or  $N$  of the following challenges, in various combinations:

### 1. Sampling challenges

- (a) Sidechains in the binding site rearrange on

binding different ligands

- (b) Modest backbone conformational changes, such as loop motion
- (c) Large scale conformational changes, such as domain motions and allostery
- (d) Ligand binding modes change unpredictably with small chemical modifications
- (e) High occupancy water sites rearrange depending on bound ligand

### 2. System challenges

- (a) Protonation state of ligand and/or protein changes on binding
- (b) Multiple protonation states of the ligand and/or receptor are relevant, due to  $pK_a$ s near the experimental pH, or the presence of multiple relevant tautomers
- (c) Results are sensitive to buffer, salts or other environmental factors

### 3. Force field challenges

- (a) Strong electric fields suggest that omission of explicit electronic polarizability will limit accuracy
- (b) Ligands interact directly with metal ions
- (c) Ligands or co-factors challenge existing force fields

*c. Progression of soft benchmarks* We envision these more complex benchmark systems proceeding through stages, initially serving effectively as a playground where major challenges and issues are explored, documented, and become well-known. Eventually, some will become sufficiently well characterized and sampled that they become hard benchmarks.

## B. Applications and limitations of benchmark systems

Benchmark systems along the lines sketched above will allow potential advances in computational methods to be tested in a straightforward, reproducible manner. For example, force fields may be assessed by swapping new parameters, or even a new functional form, into an existing workflow to see the impact on accuracy for a hard benchmark test. Sampling methods may be assessed by using various enhanced sampling methods for either hard or soft sampling benchmarks, here without focusing on accuracy relative to experiment. And system preparation tools could be varied to see how different approaches to assigning protonation states, modeling missing loops, or setting initial ligand poses, affect agreement of receptor-ligand binding calculations with experiment—with the understanding that force field and sampling also play a role. Finally, comparisons across methods will be greatly facilitated by community acceptance of a set of standard cases: well-characterized and studied benchmarks utilized by the majority of developers and research groups, ideally on a routine basis.

At the same time, there is a possibility that that some

methods will inadvertently end up tuned specifically to generate good results for the set of accepted benchmarks. In such cases, the results for systems outside the benchmark set might still be disappointing. This means the field will need to work together to develop a truly representative set of benchmarks. This potential problem can also be mitigated by sharing of methods to enable broader testing by non-developers, and by participation in blinded prediction challenges, such as SAMPL and D3R, which confront methods with entirely new challenge cases.

### III. CURRENT BENCHMARK SYSTEMS FOR BINDING PREDICTIONS

No molecular systems have yet been designated by the field as benchmarks for free energy calculations, but certain test cases have proven particularly helpful and been widely studied, including specific host molecules, certain designed binding sites in the enzyme T4 lysozyme, and some bromodomain binding sites. Here, we describe the artificial and biological receptors considered and propose specific host-guest and ligand-receptor combinations as initial benchmark systems for free energy calculations.

We also point to several additional hosts and small proteins that also have potential to generate useful benchmarks in the future (Section IV). We focus on cases where experimental data are available and add value, rather than cases chosen to test conformational sampling methods, for which experimental data are not required (Section II A).

#### A. Host-guest benchmarks

Chemical hosts are small molecules, often comprising fewer than 100 non-hydrogen atoms, with a cavity or cleft that allows them to bind other compounds, called guests, with significant affinity. Hosts bind their guests via the same basic forces that proteins used to bind their ligands, so they can serve as simple test systems for computational models of noncovalent binding. Moreover, their small size, and, in many cases, their rigidity, can make it feasible to sample all relevant conformations, making for “hard” benchmarks as defined above (Section II A). Furthermore, experiments can often be run under conditions that make the protonation states of the host and guest unambiguous. Under these conditions, the level of agreement of correctly executed calculations with experiment effectively reports on the validity of the force field (Section II A 1c). For a number of host-guest systems, the use of isothermal titration calorimetry (ITC) to characterize binding provides both binding free energies and binding enthalpies. Binding enthalpies can often also be computed to good numerical precision [89], so they provide an additional check of the validity of simulations. A variety of cu-

rated host-guest binding data is available on BindingDB at <http://bindingdb.org/bind/HostGuest.jsp>.

Hosts fall into chemical families, such that all members of each family share a major chemical motif, but individuals vary in terms of localized chemical substitutions and, in some families, the number of characteristic monomers they comprise. For example, all members of the cyclodextrin family are chiral rings of glucose monomers; family members then differ in the number of monomers and in the presence or absence of various chemical substituents. For tests of computational methods ultimately aimed at predicting protein-ligand binding affinities in aqueous solution, water soluble hosts are, arguably, most relevant. On the other hand, host-guest systems in organic solvents may usefully test how well force fields work in the non-aqueous environment within a lipid membrane. Here, we focus on three host families, the cucurbiturils [65, 147], the octa-acids (more generally, Gibb deep cavity cavitands (GDCCs)) [77, 92], and the cyclodextrins [174]. These have already been the subject of concerted attention from the simulation community; in the case of cucurbiturils GDCCs due in part to their use in the SAMPL blinded prediction challenges [153, 155, 237], and in the case of cyclodextrins, due in part to the wealth of high quality experimental data available [90, 174, 227].

#### 1. Cucurbiturils

The cucurbiturils (**Figure 1**) are achiral rings of glycoluril monomers [65]. The first characterized family member, cucurbit[6]uril, has six glycoluril units, and subsequent synthetic efforts led to the five-, seven-, eight- and ten-monomer versions, cucurbit[n]uril ( $n=5,6,7,8,10$ ) [124], which have been characterized to different extents. The  $n=6,7,8$  variants accommodate guests of progressively larger size, but are consistent in preferring to bind guests with a hydrophobic core sized to fit snugly into the relatively nonpolar binding cavity, along with at least one cationic moiety (though neutral compounds do bind [116, 230]) that forms stabilizing interactions with the oxygens of the carbonyl groups fringing both portals of the host [124]. Although derivatives of these host molecules have been made [5, 38, 117, 217], most of the binding data published for this class of hosts pertain to the non-derivatized forms. A fairly extensive set of data is available in BindingDB at <http://bindingdb.org/bind/HostGuest.jsp>.

We propose cucurbit[7]uril (CB7) as the basis of one series of host-guest benchmark systems (**Figure 1**). This host is convenient experimentally, because it is reasonably soluble in water; and computationally, because it is quite rigid and lacks acidic or basic groups. In addition, it has attracted particular interest because of the high binding affinities of some guests, exceeding even the tightest-binding protein-ligand systems [24, 124, 148, 176]. Finally, CB7 is already familiar to a number of computational chemistry groups, as it figured in two of the

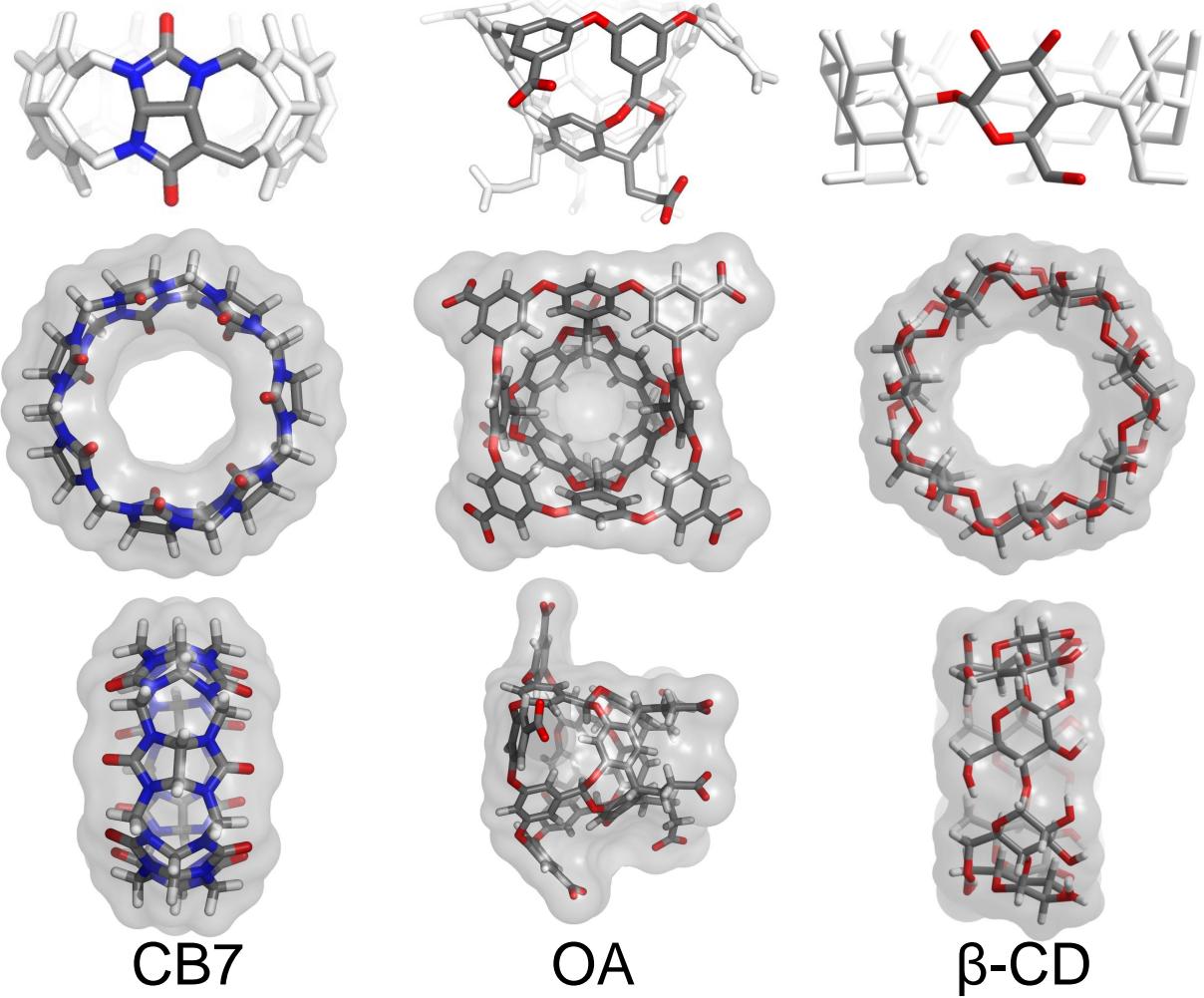


FIG. 1: Three host molecules, CB7, OA, and TEMOA, used in the host-guest benchmark sets. The first two are variants of the octa-acid GDCC, and the third is a cucurbituril. Guest structures are available in the supplemental material. Top row highlights the constituent monomers; middle and bottom rows show views into the binding sites and from the sides, respectively.

three SAMPL challenges that included host-guest components [153, 155], and it is currently the focus of the “hydrophobe challenge” [183].

*a. CB7 presents several challenges* Despite the simplicity of CB7, calculations of its binding thermodynamics are still challenging, with several known complexities:

1. **Tight exit portal:** Guest molecules with bulky hydrophobic cores, such as adamantyl or [2.2.2]bicyclooctyl [148, 149] groups, do not fit easily through the constrictive portals [212]. As a consequence, free energy methods which compute the work of binding along a physical dissociation pathway may encounter a high barrier as the bulky core exits the cavity, and this can lead to subtle convergence problems [89, 213]. One way to solve this problem is to reversibly add restraints that open the portal, then remove the guest, and finally reversibly remove the restraints [89], including all of

these contributions in the overall work of dissociation.

2. **Water binding and unbinding:** If one computes the work of removing the guest from the host by a nonphysical pathway, in which the bound guest is gradually decoupled from the host and surrounding water [81], large fluctuations in the number of water molecules within the host’s cavity can occur when the guest is partly decoupled, and these fluctuations can slow convergence [180].
3. **Salt concentration and buffer conditions:** Binding thermodynamics are sensitive to the composition of dissolved salts, both experimentally [148, 149, 153] and computationally [94, 156]. As a consequence, to be valid, a comparison of calculation with experiment must adequately model the experimental salt conditions.

#### 4. Finite-size artifacts due to charge modification:

**Because many guest molecules carry net charge, it should be ascertained that calculations in which guests are decoupled from the system do not generate artifacts related to the treatment of long-ranged Coulombic interactions [122, 172, 178, 193].**

*b. The proposed CB7 benchmark sets comprise two compound series* For CB7, we have selected two sets of guests that were studied experimentally under uniform conditions (50 mM sodium acetate buffer, pH 4.74, 298K) by one research group [24, 124]. Each series is based on a common chemical scaffold, making it amenable to not only absolute but also alchemical relative free energy calculations (Section IC). One set is based on an adamantane core (**Table I**), and the other on an aromatic ring (**Table II**). These systems can be run to convergence to allow detailed comparisons among methods and with experiment. Their measured binding free energies range from -5.99 to -17.19 kcal/mol, with the adamantane series spanning a particularly large range.

*c. Prior studies provide additional insight into CB7's challenges* Sampling of the host appears relatively straightforward in CB7 as it is quite rigid and its symmetry provides for clever convergence checks [89, 150]. Due to its top-bottom symmetry, flips of guests from “head-in” to “head-out” configurations are not necessary to obtain convergence [56]. However, sampling of the guest geometry can be a challenge, with transitions between binding modes as slow as 0.07 flips/ns [150], and flexible guests also presenting challenges [150]. As noted above, water sampling can also be an issue, with wetting/dewetting transitions occurring on the 50 ns timescale [180] when the guest is partly decoupled from the aqueous host in alchemical calculations.

Salt and buffer conditions are also key. In addition to the strong salt-dependence of binding [149], acetic acid (such as in a sodium acetate buffer) can compete with guests for the binding site [148]. This may partially explain systematic errors in some computational studies [94, 156]. Indeed, the difference between 50 mM sodium acetate buffer and 100 mM sodium phosphate buffer impacts measured binding free energies by 2.5-2.8 kcal/mol [153, 156]. Cationic guests could also have substantial and differing interactions with the counterions in solution, potentially lowering affinity relative to zero-salt conditions [153]. Additionally, CB7 can also bind cations fairly strongly [23, 95, 131]. Thus, one group found a 6.4-6.8 kcal/mol dependence on salt concentration [94] (possibly due to cation competition for the binding site), possibly impacting other studies as well [150].

Despite these issues, CB7 appears to be at the point where careful studies can probe the true accuracy of our force fields [73, 89, 94, 154, 235], and the results can be sobering, with root mean square (RMS) errors in the binding free energies as high as 8 kcal/mol [89, 150]. More encouragingly, the values of  $R^2$  values can be as high as 0.92 [89]. Calculated values are in many cases quite sensitive to details of force field parameters [149, 150, 154].

For example, modest modification of some Lennard-Jones parameters yielded dramatic improvements in calculated values [235], and host-guest binding data has, accordingly, been suggested as an input for force field development [73, 89, 235]. Water structure around CB7 and calculated binding enthalpies also appear particularly sensitive to the choice of water model [56, 73, 180], and water is clearly important for modulating binding [160]. The water model also impacts the number of sodium ions which must be displaced (in sodium-based buffer) on binding [73, 89].

In summary, CB7 is still a challenging benchmark that can put important issues into high relief. For example, in SAMPL4, free energy methods yielded  $R^2$  values from 0.1 to 0.8 and RMS errors of about 1.9 to 4.9 kcal/mol for the same set of CB7 cases [153]. This spread of results across rather similar methods highlights the need for shared benchmarks. Potential explanations include convergence difficulties, subtle methodological differences, and details of how the methods were applied. Until the origin of such discrepancies is clear, it is difficult to know how accurate our methods truly are.

To aid the adoption of these systems as benchmarks, input files for the CB7 systems proposed here are available in our GitHub repository.

## 2. Gibb Deep Cavity Cavitands (GDCC)

The octa-acids (OA) (**Figure 1**) are synthetic hosts with deep, basket-shaped, hydrophobic binding sites [77]. The eight carboxylic acidic groups for which they were originally named make these hosts water-soluble, but do not interact directly with bound hosts; instead, they project outward into solvent. Binding data have been reported for the original form of this host (OA) [77] and for a derivative with four added methyl groups at equivalent locations in the entryway, where they can contact a bound guest (TEMOA) [72, 202]. (Note that OA and TEMOA have also been called OAH and OAMe, respectively [237].) Additional family members with other substituents around the portal have been reported, as has a new series in which the eponymous carboxylic groups are replaced by various other groups, including a number of basic amines [92]. However, we are not aware of binding data for these derivatives. Because these closely related hosts are clearly in the same family but do not have eight acidic groups, and in recognition of the family's developer, we propose the more general name Gibb deep cavity cavitands (GDCCs) for this family of hosts. The binding cavities of the GDCCs are fairly rigid, though less so than the cucurbiturils. Some simulators report “breathing” motions that vary the diameter of the entry by up to 8 Å[138]; and, in some studies, the benzoic acid “flaps” around the entry occasionally flip upward and into contact with the guest [210, 236], though this motion has not been verified experimentally. Additionally, the four priopionate groups protruding into solution

TABLE I: Proposed CB7 Set 1 benchmark data

ID <sup>a</sup>	name	PC CID <sup>b</sup>	2D	SMILES	$\Delta G^c$ (kcal/mol)
1	memantine	4054		CC12CC3CC(C1)(CC(C3)(C2)N)C	$-5.99 \pm 0.05$ <sup>d</sup>
3	1,3-bis(trimethylaminio)adamantane	101379195		C[N+] (C) (C)C12CC3CC(C1)CC(C3)(C2)[N+] (C) (C)C	$-6.55 \pm 0.05$ <sup>d</sup>
5	N-(1-adamantyl)ethylenediamine	303798		C1C2CC3CC1CC(C2)(C3)NCCN	$-18.22 \pm 0.09$ <sup>e</sup>
17	adamantane-1,3-diamine	213512		C1C2CC3(CC1CC(C2)(C3)N)N	$-11.33 \pm 0.05$ <sup>d</sup>
18	1-adamantanecarboxylic acid	13235		C1C2CC3CC1CC(C2)(C3)C(=O)O	$-11.59 \pm 0.06$ <sup>d</sup>
22	1-Adamantyltrimethylaminium	3010127		C[N+] (C) (C)C12CC3CC(C1)CC(C3)C2	$-16.66 \pm 0.08$ <sup>d</sup>
23	amantadine	2130		C1C2CC3CC1CC(C2)(C3)N	$-17.19 \pm 0.08$ <sup>d</sup>
24	N-(1-adamantyl)pyridinium	3848257		C1C2CC3CC1CC(C2)(C3)[N+]4=CC=CC=C4	$-16.75 \pm 0.07$ <sup>d</sup>

<sup>a</sup> Compound ID from original paper; <sup>b</sup> PubChem Compound ID; <sup>c</sup> Standard binding free energy, where all measurements were done via NMR in 50mM sodium acetate buffer in  $D_2O$  at pH 4.74 and 298 K. Uncertainties are obtained by taking the reported standard deviations across triplicate measurements [96] and dividing by  $\sqrt{3}$ ; <sup>d</sup> drawn from [124]; <sup>e</sup> drawn from [24].

from the exterior base of the cavity are all flexible.

The octa-acids tend to bind guest molecules possessing a hydrophobic moiety that fits into the host's cavity and a hydrophilic moiety that projects into the aqueous solvent. Within these specifications, they bind a diversity of ligands, including both organic cations and anions, as well as neutral compounds with varying degrees of polarity [78, 80]. Compounds with adamantane or noradamantane groups display perhaps the highest affinities observed so far, with binding free energies ranging to about -8 kcal/mol [203]. Many of the experimental binding data comes from ITC, so binding enthalpies are often available.

Two experimental aspects of binding are particularly intriguing and noteworthy. First, the binding thermodynamics of OA is sensitive to the type and concentration of anions in solution. Although NaCl produces relatively modest effects, 100 mM sodium perchlorate, chloride and isothiocyanate can shift binding enthalpies by up to about 10 kcal/mol and free energies by around 2 kcal/mol [79]. These effects are due in part to binding of anions by the host; indeed, trichloroacetate is reported

to bind OA with a free energy of -5.2 kcal/mol [197], and competition of other guests with bound anions leads to entropy-enthalpy tradeoffs. Second, elongated guests can generate ternary complexes, in which two OA hosts encapsulate one guest, especially if both ends of the guest are not very polar [78].

*a. The proposed GDCC benchmark sets are drawn from SAMPL* We propose the establishment of two GDCC benchmark sets, based on data which formed part of the SAMPL4 and SAMPL5 challenges. One set is based on experiments carried out in phosphate buffer at pH 11.5, and the other on experiments in tetraborate buffer at pH 9.2. The guests in the first set (Table III) are adamantane derivatives and cyclic (aromatic and saturated) carboxylic acids (Table III), which bind to hosts OA and TEMOA with free energies of -3.7 to -7.6 kcal/mol. The second set of guests (Table IV) comprises carboxylic acids based on phenyl and cyclohexane cores. Both sets offer aqueous binding data with free energies spanning about 4 kcal/mol, frequently along with binding enthalpies. The hosts and many or all of their guests are small and rigid enough to allow convincing convergence

TABLE II: Proposed CB7 Set 2 benchmark data

ID <sup>a</sup>	name	PC CID <sup>b</sup>	2D	SMILES	$\Delta G^{c,d}$ (kcal/mol)
2	dopamine	681		c1cc(c(cc1CCN)O)O	-6.31 ± 0.05
4	o-phenylenediamine	7243		c1ccc(c(c1)N)N	-6.68 ± 0.05
5	m-phenylenediamine	7935		c1cc(cc(c1)N)N	-6.69 ± 0.02
7	4-(aminomethyl)pyridine	77317		c1cnccc1CN	-7.56 ± 0.06
8	p-phenylenediamine	7814		c1cc(ccc1N)N	-8.60 ± 0.06
9	p-toluidine	7813		Cc1ccc(cc1)N	-9.43 ± 0.05
20	p-xylylenediamine	68315		c1cc(ccc1CN)CN	-12.62 ± 0.06

<sup>a</sup> Compound ID from original paper; <sup>b</sup> PubChem Compound ID; <sup>c</sup> Standard binding free energy, where all measurements were done via NMR in 50mM sodium acetate buffer in  $D_2O$  at pH 4.74 and 298 K. Uncertainties are obtained by taking the reported standard deviations across triplicate measurements [96] and dividing by  $\sqrt{3}$ ; <sup>d</sup> drawn from [124].

of binding thermodynamics with readily feasible simulations; and, like the cucurbiturils, they are already emerging as *de facto* computational benchmarks, due to their use in the SAMPL4 and SAMPL5 challenges [153, 237].

*b.* *The GDCC hosts introduce new challenges beyond CB7* Issues deserving attention when interpreting the experimental data and calculating the binding thermodynamics of these systems include the following:

**1. Tight exit portal:** The methyl groups of the TEMOA variant narrow the entryway and can generate a barrier to the entry or exit of guest molecules with bulky hydrophobic cores, though the degree of constriction is not as marked as for CB7 (above). The TEMOA methyl groups can additionally hinder sampling of guest poses in the bound state, leading to convergence problems [237]

**2. Host conformational sampling:** Although the flexible propionate groups are not proximal to the binding cavity, they are charged and so can have long-ranged interactions. As a consequence, it may be important to ensure their conformations are well sampled, though motions may be slow [138]. Similarly, benzoic acid flips [210, 236] could potentially be an important challenge in some force fields.

**3. Water binding and unbinding:** Water appears to undergo slow motions into and out of the OA host, on timescales upwards of 5 ns [54]. This poses significant challenges for some approaches, such as metadynamics, where deliberately restraining water to stay out of the cavity when the host is not bound (and computing the free energy of doing so) can help convergence [14], and perhaps for other methods as well.

**4. Salt concentration and buffer conditions:** As in the case of CB7, binding to GDCCs is modulated by the composition of dissolved salts, both experimentally [79, 197] and computationally [162, 210]. As a consequence, to be valid, a comparison of calculation with experiment must adequately model the experimental salt conditions.

**5. Finite-size artifacts due to charge modification:** As for CB7, it should be ascertained that calculations in which charged guests are decoupled from the system do not generate artifacts related to long range Coulomb interactions. [122, 172, 178, 193].

**6. Protonation state effects:** Although experiments are typically run at pH values that lead to well-defined protonation states of the host and its guests, this may not always hold [54, 153, 210], particularly given experimental evidence for extreme binding-driven  $pK_a$  shifts of 3-4 log units for some carboxylate compounds [197, 221]. Thus, attention should be given to ionization states and their modulation by binding.

*c. Prior studies provide additional insight into the challenges of OA* As noted, two different host conformational sampling issues have been observed, with dihedral transitions for the propionate groups occurring on 1-2 ns timescales [138]); motions of the benzoic acid flaps were also relatively slow [210, 236] though perhaps thermodynamically unimportant. Guest sampling can also be an issue, at least in TEMOA [237], and this host's tight cavity may also have implications for binding entropy [236].

Salt concentration strongly modulates binding affin-

TABLE III: Proposed GDCC Set 1 benchmark data

ID <sup>a</sup>	SAMPL <sup>b</sup>	name	PC CID <sup>c</sup>	2D	SMILES	$\Delta G^c$ (kcal/mol)	$\Delta H^d$ (kcal/mol)
OA host							
3	OA-G1	5-hexynoic acid	143036		C#CCCCCC(=O)O	-5.400 ± 0.003	-7.71 ± 0.05
4	OA-G6	3-nitrobenzoic acid	8497		c1cc(cc(c1)[N+](=O)[O-])C(=O)O	-5.340 ± 0.005	-5.67 ± 0.01
5	OA-G2	4-cyanobenzoic acid	12087		c1cc(ccc1C#N)C(=O)O	-4.73 ± 0.01	-4.45 ± 0.08
6	OA-G4	4-bromo adamantane-1-carboxylic acid	12598766		C1C2CC3CC(C2)(CC1C3Br)C(=O)O	-9.37 ± 0.01	-14.78 ± 0.02
7	OA-G3	N,N,N-trimethylhexan-1-aminium	84774		CCCCC[N+](C)(C)C	-4.49 ± 0.01	-5.91 ± 0.10
8	OA-G5	trimethylphenethylaminium	14108		Cc1cccc1CC(C)(C)[NH3+]	-3.72 ± 0.01	-9.96 ± 0.11
TEMOA/OAME host							
3	OA-G1	5-hexynoic acid	143036		C#CCCCCC(=O)O	-5.476 ± 0.006	-9.961 ± 0.006
4	OA-G6	3-nitrobenzoic acid	8497		c1cc(cc(c1)[N+](=O)[O-])C(=O)O	-4.52 ± 0.02	-9.1 ± 0.1
5	OA-G2	4-cyanobenzoic acid	12087		c1cc(ccc1C#N)C(=O)O	-5.26 ± 0.01	-7.6 ± 0.1
7	OA-G3	N,N,N-trimethylhexan-1-aminium	84774		CCCCC[N+](C)(C)C	-5.73 ± 0.06	-6.62 ± 0.20

<sup>a</sup> Compound ID from [202]; <sup>b</sup> SAMPL5 ID from [237]; <sup>c</sup> PubChem Compound ID; <sup>d</sup> Standard binding free energy from [202], where all measurements were done via ITC in 50 mM sodium phosphate buffer at pH 11.5 and 298 K. Uncertainties, drawn from the experimental paper, were computed from triplicate measurements taken with freshly made solutions of host and guest. However, based on personal communication with the authors, it may be advisable to regard the accuracy more conservatively, at ~2% for  $\Delta G$  and ~6% for  $\Delta H$ ; <sup>e</sup> measured binding enthalpy [202], subject to the same conditions/caveats as <sup>c</sup>. <sup>e</sup> not done.

TABLE IV: Proposed GDCC Set 2 benchmark data for binding to the OA host.

ID <sup>a</sup>	name	PC CID <sup>b</sup>	2D	SMILES	Method	$\Delta G^c$ (kcal/mol)
1	benzoic acid	243		c1ccc(cc1)C(=O)O	NMR	-3.72 ± 0.03
2	4-methylbenzoic acid	7470		Cc1ccc(cc1)C(=O)O	NMR	-5.85 ± 0.06
3	4-ethylbenzoic acid	12086		CCc1ccc(cc1)C(=O)O	ITC	-6.27 ± 0.01
4	4-chlorobenzoic acid	6318		c1cc(ccc1C(=O)O)Cl	ITC	-6.72 ± 0.01
5	3-chlorobenzoic acid	447		c1cc(cc(c1)Cl)C(=O)O	NMR	-5.24 ± 0.02
6	cyclohexanecarboxylic acid	7413		C1CCC(CC1)C(=O)O	NMR	-5.62 ± 0.04
7	trans-4-methylcyclohexanecarboxylic acid	20330		[C@H]1(CC[C@H]1(CC1)C(=O)O[H])C	ITC	-7.61 ± 0.04

<sup>a</sup> Compound ID from original paper [80]; <sup>b</sup> PubChem Compound ID; <sup>c</sup> Standard binding free energy from [80], where all measurements were done in 10 mM sodium tetraborate buffer at pH 9.2 and 298 K. A quirk is that for the NMR measurements, the guest was titrated in from 50 mM sodium tetraborate buffer, so the buffer concentration changed during the titration. Uncertainty is the standard error of the mean in free energy, computed from the reported standard deviations in  $K_a$ . Again, based on personal communication with the authors, uncertainties of perhaps 10% may be more appropriate.

ity, at least for anions, and the nature of the salt also plays an important role [25]. Co-solvent anions can also increase *or* decrease binding depending on their identity [79]. Some salts even bind to OA themselves, with perchlorate [79] and trichloroacetate [197] being particularly potent, and thus will compete with guests for binding. Computationally, including additional salt beyond that needed for system neutralization changed binding free energies by up to 4 kcal/mol [210].

Although the protonation states of the guests might seem clear and unambiguous, experimental evidence also indicates major  $pK_a$  shifts on binding so that species such as acetate, formate and others could bind in neutral form at neutral pH [197, 221]. One study used absolute binding free energy calculations for different guest charge states, coupled with  $pK_a$  calculations, and found that inclusion of  $pK_a$  corrections and the possibility of alternate charge states of the guests affected calculated binding free energies by up to 2 kcal/mol [210]. Even the host protonation state may be unclear. Although OA might be assumed to have all eight carboxylic acids deprotonated at the basic pH of typical experiments, the four at the bottom are in close proximity, and these might make hydrogen bonds allowing retention of two protons [54]. Thus, there are uncertainties as to the host protonation state [54, 153], which perhaps also could be modulated by guest binding.

The challenges of system definition and conformational sampling appear to be greater for OA and TEMOA than for CB7, as discussed above, so it is harder to draw definite conclusions regarding force field accuracy and differences among force fields. Thus, although there have been some comparisons of charge [138, 150, 153] and water models [236], the resulting differences in computed binding thermodynamics so far seem mostly inconclusive or at least not that large. Similarly, OPLS and GAFF results did not appear dramatically different in accuracy [14]. It is worth noting that several groups using different computational approaches but the same force field and water model in SAMPL5 did not obtain identical binding free energies [14, 17, 237]. Some of these issues were resolved in follow-up work [14], bringing the methods into fairly good agreement for the majority of cases [17, 236].

To aid the adoption of these systems as benchmarks, input files for the GDCC systems proposed here are available in our GitHub repository.

### 3. Cyclodextrins

Cyclodextrins (CDs; **Figure 1**) are cyclic glucose polymers which feature a tapered, hydrophobic cavity ringed by secondary hydroxyls on the wider end and primary hydroxyls on the narrow end. They are produced from starch by an enzymatic process which generates several versions, but the most common are  $\alpha$ -CD and  $\beta$ -CD which consist of six and seven glucose units, respectively. CDs are known to bind a diverse set of ligands which

likely accounts for their widespread industrial use, including food applications, cosmetics, pharmaceuticals, and odor suppression. Indeed, a thorough review 1998 review tabulates hundreds of binding data [174], and many additional measurements have been published since then; e.g. references [26, 39, 40, 134, 179, 192, 229]. A number of these studies were done calorimetrically, and thus provide not only binding free energies but also binding enthalpies. Fewer data are available for the larger 8-membered  $\gamma$ -CD, and the greater diameter of this host makes for a less well-defined and floppier binding cavity.

We propose a benchmark series which includes both  $\alpha$ - and  $\beta$ -CD binding to a set of neutral and charged guests. The CD host has many experimental upsides, including adequate solubility, at least for these common variants, and they are available in gram quantities at low cost from multiple suppliers. Additionally, CDs are familiar host-guest systems to many computational researchers and are known to be tractable for simulation [11, 21, 28, 75, 89, 109, 128, 130, 184, 186, 226, 227, 240]. However, CDs offer a contrast to the previous two benchmark series in a number of ways. First, the hosts are more flexible than CB7 or the GDCCs, as their glucose units are joined by only a single bond, which allows greater rotation of each monomer relative to its neighbors. Second, the binding cavity has two, non-equivalent openings or “rims”, allowing asymmetric guests to bind in either a “head-in” or “head-out” orientation. It is also worth noting that, because the CDs are glucose polymers, they may be best modeled with dedicated carbohydrate force fields [27, 87, 111, 231], rather than generalized small molecule force fields. These features combine to add complexity in predicting the dominant binding poses for guests, although the default assumption is that polar functional groups will orient towards solvent near the wider rim (also called the secondary face) and hydrophobic groups will nestle into the hydrophobic cavity of the narrow rim (also called the primary face).

a. *The proposed CD benchmark series includes an  $\alpha$ -CD set and a  $\beta$ -CD set* For each CD set (Tables V and VI), we have selected guests which include n-alkyl ammonium cations, neutral cyclic alcohols, and n-alkyl or aromatic carboxylate anions. The experimental data all originates from the same ITC study [175] and experiments were performed under uniform conditions (50 mM sodium phosphate buffer, pH 6.90, 298K). The buffer constituents are not expected to compete for binding in the CD cavity [173]. Both binding free energy and binding enthalpy measurements are available for all guests in the series. The binding free energies range from -1.27 to -4.62 kcal/mol and the binding enthalpies range from 1.89 to -5.46 kcal/mol. In the case of  $\alpha$ -CD and its guests, the binding enthalpies are uniformly negative and therefore the binding entropies are uniformly positive. In contrast, the  $\beta$ -CD set includes both positive and negative binding enthalpies and binding entropies. In addition to the formal benchmark sets listed in Tables V and VI, which only include a representative sample of

eight and twelve guests, respectively, we have also made available (in our GitHub repository) an expanded set of guests which differ only slightly from the molecule classes mentioned above, by minor variations in the carbon substituents. Despite the flexibility of CD hosts, the small size of the guests combined with long timescale simulations enable by GPUs should allow adequate convergence of these thermodynamic values.

*b. Challenges specific to CDs* Researchers should anticipate the following complexities due to the dynamic nature of the CD host and its asymmetric binding site:

1. **Multiple binding orientations:** Unlike CB7 or OA, CDs have a binding cavity with two unique openings to solvent. Although the default expectation is that a guest's polar functional group will orient out of the wider (secondary) rim of the cavity [174, 175], and hydrophobic groups will favor the narrower (primary) side, it raises the possibility of the reverse type of binding or a mixture of both. Since the cavity is both narrow and hydrophobic, interchange between the two orientations is usually slow and requires release into the solvent and subsequent rebinding. Therefore unless frequent interchange between the two orientations is observed during simulation, it may be advisable to perform free energy calculations for both orientations and then combine the results according to their Boltzmann weights [89].
2. **Monomer flexibility:** Cyclodextrin molecules are known to be rather flexible [12, 39]. In particular, each glucose monomer can flip conformation along the glycosidic bond, which tends to partially occlude the binding cavity. In simulation, these flipping events occur on the multi-nanosecond timescale and thus can pose a challenge for convergence of binding thermodynamics. For affinity calculations which rely on physical or alchemical methods, the portion of the free energy pathway in which the guest is only partially interacting with the host is expected to be the most problematic. Conformational restraints, if properly accounted for, can help mitigate these convergence challenges [89].
3. **Buffer conditions:** CDs are not expected to bind phosphate ions from the buffer with detectable affinity [173, 174]; however, due to force field limitations, it is possible that the guest and buffer could compete for binding in simulations. The solution behavior of phosphate ions, particularly the  $\text{HPO}_4^{2-}$  species, can be sensitive to the particular choice of water model and may be worth investigating prior to extensive binding calculations. Alternatively, one could substitute an equivalent strength of monovalent salt for the phosphate buffer [90].
4. **Finite-size artifacts due to charge modification:** Once again, the presence of charged guests in the CD benchmarks necessitates care with re-

gard to charge decoupling procedures which could generate artifacts in the long range Coulomb calculations [122, 172, 178, 193].

*c. Prior studies provide additional insight into the challenges of CD* Although experimental literature for CD binding is widely available, computational studies including more than just a few guests are rare and, since CDs have not been included in any SAMPL challenges, the unique properties of this host might not be as familiar to many as those of CB7 and OA. As with these other hosts, the choice of buffer influences the binding measurements since any organic buffer components will likely competitively bind. Phosphate buffer, which is used in both of the proposed CD benchmarks, did not show any evidence of binding to the host (as measured by ITC), whereas acetate buffer demonstrated competitive activity [173, 174]. Ideally, simulations of these systems would match the experimental conditions, however, phosphate buffer can be difficult to model accurately, particularly the  $\text{HPO}_4^{2-}$  species [89, 90]. One potential remedy to this problem is to match the ionic strength of the phosphate buffer with monovalent salt [90], although the impact of this approximation is not fully understood.

Expulsion of structured water from the unoccupied host cavity is one of the often-cited driving forces for guests binding to CDs. Although the size of this effect is debated [39, 208], it is generally accepted that approximately two to ten water molecules occupy the unbound cavity, depending on how the cavity is defined and whether the host is  $\alpha$ - or  $\beta$ -CD [15, 39]. In simulations, the choice of host force field can play a large role in the appearance of structured water. If the glycosidic torsional barriers are low, the CD host can contort itself in such a way as to expel these presumably high energy waters [90].

There are some recent large-scale computational studies of CD binding free energies which deserve mention. One study, from the Levy group [227], used the BEDAM method [69] to compute binding free energies for 57 guests to  $\beta$ -CD. The method uses an implicit representation of solvent, although guest occupancy of water sites in the CD cavity are incorporated into the model. This study found an RMSE value of 1.44 kcal/mol and  $R^2$  value of 0.43. Levy and co-workers noted that some guests, particularly those with ammoniums, oriented their polar group out the narrow, primary rim of the CD cavity. Another study, by the van der Spoel group [241], computed umbrella sampling PMFs for removal of 75 neutral guests from  $\beta$ -CD using several Generalized Born implicit solvent models. These solvent models did not include consideration of structured water and the results were sobering: across all solvent models, RMSEs ranged between 1.9 - 4.8 kcal/mol and  $R^2$  values were less than 0.3. The authors noted that 70% of the guests preferred to orient the guest polar group through the wider, secondary opening of the host cavity, suggesting that although this orientation is preferred, the primary orientation must be considered as

TABLE V: Proposed Cyclodextrin Set 1 benchmark for the  $\alpha$ -CD host

ID <sup>a</sup>	name	PC CID <sup>b</sup>	2D	SMILES	Method	$\Delta G^c$ (kcal/mol)	$\Delta H^d$ (kcal/mol)
1	1-butylamine	8007		CCCCN	ITC	$-1.575 \pm 0.019$	$-2.17 \pm 0.05$
2	1-hexylamine	8102		CCCCCCN	ITC	$-3.533 \pm 0.004$	$-4.19 \pm 0.02$
3	1-octylamine	8143		CCCCCCCN	ITC	$-4.606 \pm 0.007$	$-5.46 \pm 0.03$
4	cyclopentanol	7298		C1CCC(C1)O	ITC	$-2.130 \pm 0.016$	$-2.74 \pm 0.02$
5	cycloheptanol	10399		C1CCCC(CC1)O	ITC	$-2.510 \pm 0.060$	$-2.99 \pm 0.23$
6	butanoic acid	264		CCCC(=O)O	ITC	$-1.506 \pm 0.036$	$-2.53 \pm 0.12$
7	hexanoic acid	8892		CCCCC(=O)O	ITC	$-3.380 \pm 0.005$	$-3.40 \pm 0.02$
8	octanoic acid	379		CCCCCC(=O)O	ITC	$-4.622 \pm 0.017$	$-4.89 \pm 0.03$

<sup>a</sup> Compound ID assigned in the present work; <sup>b</sup> PubChem Compound ID; <sup>c</sup> Standard binding free energy from [175], where all measurements were done in  $0.050 \text{ mol kg}^{-1}$  sodium phosphate buffer at pH 6.90 and 298.15 K. At these buffer conditions, the amine guests are expected to be protonated with a formal charge of +1, and the carboxylic acids are expected to be deprotonated with a net charge of -1. Uncertainty is the standard error of the mean in free energy or enthalpy, computed from the reported standard deviations in  $K_a$ .

well. Finally, a recent study by the Gilson lab [90] used the attach-pull-release method [89] to compute explicitly solvated binding free energies and enthalpies for 43 host-guest pairs, including both  $\alpha$ - and  $\beta$ -CD hosts, for which experimental calorimetry measurements are available [175]. The guests comprised three classes, including alkyl-ammoniums, cyclic alcohols, and alkyl- and aromatic-carboxylates. The two CD benchmark sets in this work were derived this study. Several water models and host force fields were investigated, but none of the force field combinations proved superior at both binding free energy and enthalpy. Across force fields, the RMSE values for binding free energy ranged from 0.9 to 1.8 kcal/mol and  $R^2$  from 0.44 to 0.60. For binding enthalpy, RMSE values ranged from 0.9 to 4.0 kcal/mol and  $R^2$  from 0.30 to 0.77. Once again, a preference of ammonium guests to orient out of the narrow, primary opening was observed. It remains to be determined whether experimental evidence will support this preference.

To aid the adoption of these systems as benchmarks, input files for the CD systems proposed here are available in our GitHub repository.

## B. Protein-ligand benchmarks: the T4 lysozyme model binding sites

Although we seek ultimately to predict binding in systems of direct pharmaceutical relevance, simpler protein-ligand systems can represent important stepping stones in this direction. Two model binding sites in T4 lysozyme have been particularly useful in this regard (**Figure 2**).

These two binding sites, called L99A [151, 152] and L99A/M102Q [84, 224] for point mutations which create the cavities of interest, have been studied extensively experimentally and via modeling. As protein-ligand systems, they introduce additional complexities beyond those observed in host-guest systems, yet they share some of the same simplicity. The ligands are generally small, neutral, and relatively rigid, with clear protonation states. For most ligands, substantial protein motions do not occur on binding at room temperature and ambient pressure, helping calculated binding free energies to reach apparent convergence relatively easily. However, like host-guest systems, these binding sites are still surprisingly challenging [19, 71, 98, 121, 140, 141, 143]. Thus, precise convergence is sometimes difficult to achieve, and it is in all cases essentially impossible to fully verify. As a consequence, these are “soft benchmarks” as defined above (Section II A). The utility of the lysozyme model sites is also driven by the large body of available experimental data. It has been relatively easy to identify new ligands and obtain high quality crystal structures and affinity measurements, and this has allowed two different rounds of blinded free energy prediction exercises [19, 143].

These binding sites do exhibit some surprising experimental complexities which make them interesting ongoing topics of study, such as the fact that the L99A site is empty of water when ligands are not bound [36, 120, 161] yet the protein can undergo pressure-induced filling [36, 120] or denaturation [161] which can be inhibited by binding of ligand [120, 161]. Pressure may also cause the protein to populate an excited state [112, 129]

TABLE VI: Proposed Cyclodextrin Set 2 benchmark for the  $\beta$ -CD host

ID <sup>a</sup>	name	PC CID <sup>b</sup>	2D	SMILES	Method	$\Delta G^c$ (kcal/mol)	$\Delta H^d$ (kcal/mol)
1	1-hexylamine	8012		CCCCCN	ITC	$-2.486 \pm 0.084$	$0.60 \pm 0.05$
2	1-octylamine	8143		CCCCCCN	ITC	$-3.585 \pm 0.120$	$-0.48 \pm 0.03$
3	cyclopentanol	7298		C1CCC(C1)O	ITC	$-3.050 \pm 0.010$	$-1.09 \pm 0.01$
4	trans-4-methylcyclohexanol	11524		C[C@H]1CC[C@@H](CC1)O	ITC	$-4.541 \pm 0.008$	$-2.17 \pm 0.02$
5	cycloheptanol	10399		C1CCCC(CC1)O	ITC	$-4.560 \pm 0.011$	$-2.96 \pm 0.01$
6	pentanoic acid	7991		CCCCC(=O)O	ITC	$-1.267 \pm 0.323$	$1.89 \pm 0.53$
7	heptanoic acid	8094		CCCCCC(=O)O	ITC	$-3.394 \pm 0.179$	$0.42 \pm 0.04$
8	benzoic acid	243		C1=CC=C(C=C1)C(=O)O	ITC	$-1.640 \pm 0.024$	$-2.51 \pm 0.08$
9	phenylacetic acid	999		C1=CC=C(C=C1)CC(=O)O	ITC	$-1.697 \pm 0.048$	$-1.79 \pm 0.11$
10	4-methylphenylacetic acid	248474		CC1=CC=C(C=C1)CC(=O)O	ITC	$-2.192 \pm 0.013$	$-2.89 \pm 0.05$
11	4-methoxyphenylacetic acid	7690		COC1=CC=C(C=C1)CC(=O)O	ITC	$-2.512 \pm 0.007$	$-1.96 \pm 0.01$
12	4-phenylbutanoic acid	4775		C1=CC=C(C=C1)CCCC(=O)O	ITC	$-3.599 \pm 0.018$	$-2.82 \pm 0.01$

<sup>a</sup> Compound ID assigned in the present work; <sup>b</sup> PubChem Compound ID; <sup>c</sup> Standard binding free energy from [175], where all measurements were done in  $0.050 \text{ mol kg}^{-1}$  sodium phosphate buffer at pH 6.90 and 298.15 K. At these buffer conditions, the amine guests are expected to be protonated with a formal charge of +1, and the carboxylic acids are expected to be deprotonated with a net charge of -1. Uncertainty is the standard error of the mean in free energy or enthalpy, computed from the reported standard deviations in  $K_a$ .

(but see [218]) which is already present to a very limited extent at equilibrium [18]. Still, as noted below, these issues do not seem to dramatically impact our ability to calculate binding free energies at standard temperature and pressure, probably in large part because these are effects which come into play only at high pressures [120, 129, 161], though as we discuss below, some ligands do induce a protein conformational change which affects the same helix as the proposed excited state [133]. It seems likely that the conformational heterogeneity observed experimentally will make lysozyme even more of a valuable benchmark system as test cases here can range from simple to challenging depending on the ligand and pressure being considered.

### 1. The apolar and polar cavities and their ligands

The L99A site is also called the “apolar” cavity. It is relatively flat and elongated, and binds mostly non-polar molecules such as benzene, toluene, p-xylene, and n-butylbenzene: basically, a fairly broad range of nonpolar planar five- and six-membered rings and ring systems (such as indole). The polar version, L99A/M102Q, introduces an additional point mutation along one edge of the binding site, providing a glutamine that introduces polarity and the potential for hydrogen bonding. It still binds a variety of nonpolar ligands such as toluene (though not benzene). One small downside of these binding sites is that the range of affinities is relatively narrow: about -4.5 to -6.7 kcal/mol in the apolar site [143, 151], and about -4 to -5.5 kcal/mol in the polar site [19]. Thus, even the strongest binders are not particularly strong, and the weakest binders tend to run up against their solubility limits. Still, these sites offer immensely useful

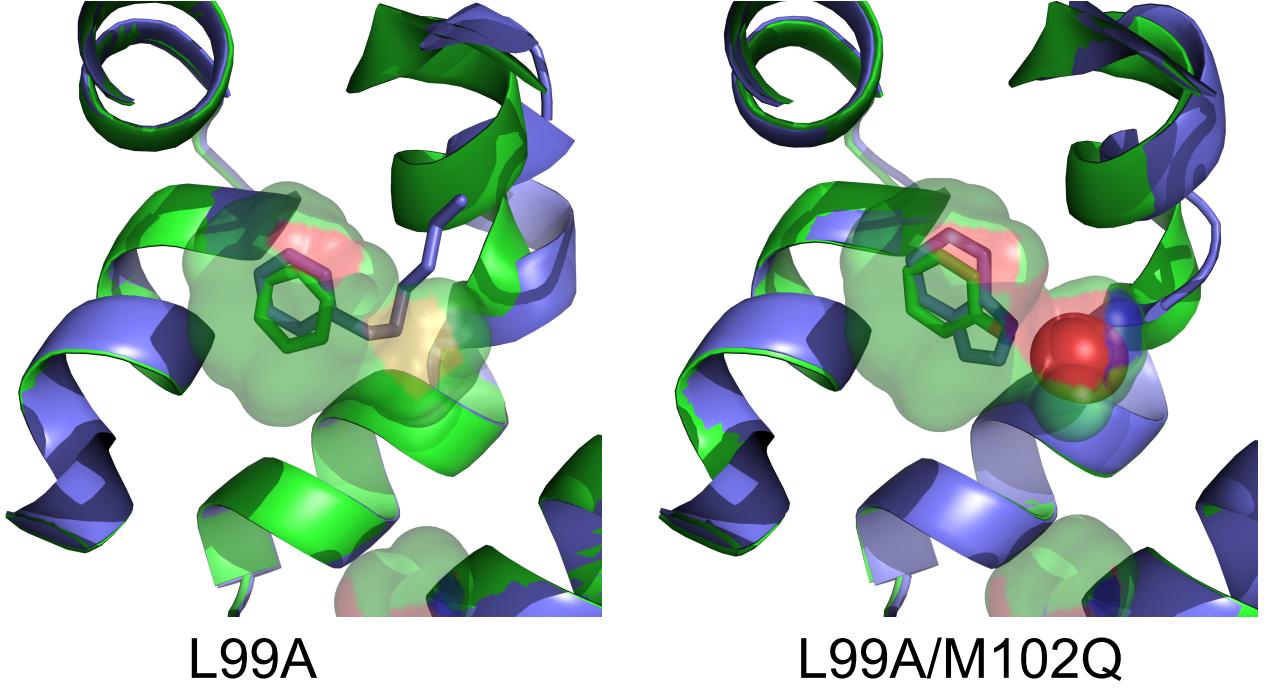


FIG. 2: Benzene and hexylbenzene in the lysozyme L99A site, and phenol and 4,5,6,7-tetrahydroindole in the L99A/M102Q site (PDBs 4W52, 4W59, 1LI2, and 3HUA, respectively). The binding site shape is shown as a semi-transparent surface, and the protein with cartoons. In both cases, the structure with the smaller ligand is shown in green and that with the larger ligand is shown in blue. The larger ligand induces a motion of helix F bordering the binding site. Phenol and 4,5,6,7-tetrahydroindole both also bind with an ordered water (red sphere), though this does not occur for all ligands in the polar L99A/M102Q site.

tests for free energy calculations.

For both sites, fixed charge force fields seem to yield reasonably accurate free energies, with RMS errors of 1-2 kcal/mol, and some level of correlation with experiment, despite limited dynamic range [19, 47, 71, 143, 220]. System composition/preparation issues also do not seem to be a huge factor. Instead, sampling issues predominate:

1. Ligand orientation: These oblong binding sites are buried, and their ligands are similar in shape. Ligands with axial symmetry typically have at least two reasonably likely binding modes, but broken symmetry can drive up the number of likely binding modes. For example, phenol has two plausible binding modes in the polar cavity [19, 85] but 3-chlorophenol has at least four, three of which appear to have some population in simulations [71], because the chlorine could point in either direction within the site. Timescales for binding mode interconversion are relatively slow, with in-plane transitions on the 1-10 nanosecond timescale, and out-of-plane transitions (e.g. between toluene's two symmetry-equivalent binding modes) taking hundreds of nanoseconds (Mobley group, unpublished data).

2. Sidechain rearrangement: Some sidechains reorganize when certain ligands, particularly larger ones, bind. This is common in particular for the side

chain of Val 111 in the L99A site [98, 141, 152] and Leu 118, Val 111, and Val 103 in L99A/M102Q [19, 85, 224, 225]. These sidechain motions can be slow, due to the tight packing of the binding site, and therefore can present sampling problems for standard MD simulations [19, 98, 141, 143, 222].

3. Backbone rearrangement: Larger ligands induce shifts of the F helix (residues 107 or 108 to 115), which is adjacent to the binding site, allowing the site to enlarge. This occurs in both binding sites [19, 133, 225], but is best characterized for L99A [133]. There, addition of a series of methyl groups from benzene up to n-hexylbenzene causes a conformational transition in the protein from closed to intermediate to open conformations; this affects the same region of helix F that undergoes a conformational change in the proposed excited state which is partially populated at equilibrium [18].

**Tables VII and VIII** introduce two initial benchmark sets based on the apolar and polar T4 lysozyme binding sites. These sets include ligands amenable to both absolute and relative free energy calculations, and have affinities that cover the currently available experimental range. Co-crystal structures are available in most cases; the corresponding PDB IDs are provided in the tables. The selected ligands span a range of challenges and levels of difficulty, ranging from fairly simple to includ-

ing most of the challenges noted above. Essentially all of them have been included in at least one prior computational study, and some have appeared in a variety of prior studies. Additional known ligands and nonbinders are available, with binding affinities available for 19 compounds in the L99A site [53, 143, 151] and 16 in L99A/M102Q [19, 84, 224]. Because of the extent of the sampling challenges in lysozyme, binding of most ligands will currently constitute a soft benchmark, though long-timescale simulations to turn these into hard benchmarks may already be feasible.

## 2. Computational challenges posed by the T4 lysozyme benchmarks

Early work on the lysozyme sites focused on the difficulty of predicting binding modes [19, 140, 143] because of the slow interconversions noted above. Docking methods often can generate reasonable poses spanning most of the important possibilities [19, 85, 140, 143] but do not accurately predict the binding modes of individual compounds [19, 85, 143]. Thus, binding calculations must explore multiple potential binding modes, especially as some ligands actually populate multiple poses at equilibrium [19]. In a number of studies, candidate binding modes from docking are relaxed with MD simulations, then clustered to select binding modes for further study. It turns out an effective binding free energy for each distinct candidate binding mode can be computed separately [140] and combined to find the population of each binding mode and determine the overall binding free energy. However, this is costly, since each binding mode requires a full binding free energy calculation.

Relative binding free energy calculations do not dramatically simplify the situation, as introduction of a ligand and modification can leave the binding mode uncertain. For example, adding a chlorine to phenol generates at least two binding mode variants, even if the binding mode of phenol is known) [19]. A naïve solution is to consider multiple possible binding modes in relative free energy calculations [19], but this generates multiple results; determining the true relative binding free energy requires additional information [144]. Enhanced sampling approaches provide one possible solution to the binding mode problem. For example,  $\lambda$  or Hamiltonian exchange techniques can incorporate sampling on an artificial energy surface where the ligand does not interact with the protein, and thus can readily reorient. Thus, approaches employing this strategy can naturally sample multiple binding modes [71, 220].

While sidechain sampling has been a significant challenge, it is possible to use biased sampling techniques such as umbrella sampling to deliberately compute and include free energies of sampling slow sidechain rearrangements [141]. However, this is not a general solution, since it requires knowing what sidechains might rearrange on binding and then expending substantial

computational power to map the free energy landscapes for these rearrangements. An apparently better general strategy is including sidechains in enhanced sampling regions selected for Hamiltonian exchange [98, 108] or REST2 [222], allowing sidechains to be alchemically softened or torsion barriers lowered (or both), to enhance sampling at alchemical intermediate states. Reduction of energy barriers at intermediate  $\lambda$  states in this scheme allows enhanced conformational sampling at these  $\lambda$  values. Then, with swaps between  $\lambda$  values, enhanced sidechain sampling at intermediate states can propagate to all states, improving convergence [98, 222].

Larger protein conformational changes in lysozyme have received less attention, partly because until very recently they seemed to be a peculiar oddity only rarely observed; i.e., for ligands 4,5,6,7-tetrahydroindole and benzyl acetate in the polar site [19]. However, recent work noted above highlighted how a helix in the apolar cavity can open to accommodate larger ligands [133]. Timescales for this motion appear to be on the order of 50 ns, so it can pose sampling challenges, even for relative free energy calculations [121]. Including part of the protein in the enhanced sampling region via REST2 (described above) provides some benefits [121], but sampling these motions will likely prove a valuable test for enhanced sampling methods.

Accounting for water exchange into and out of these buried binding sites could also be a challenge in the polar binding site, as phenol binds with one ordered water [224]. However, to our knowledge, the simulation timescales for water sampling have not yet been examined, other than noting that water does not enter the polar binding site as ligands are removed [19]. In contrast, the apolar cavity is dry, both with and without a bound ligand, so water sampling is unlikely to be a challenge in this case.

## C. Protein-ligand benchmarks: the first bromodomain of the BRD4 protein

We believe applications of free energy calculations to small protein domains called bromodomains can provide a suitable first benchmark on systems that have pharmaceutical relevance. Bromodomain binding, especially to the first bromodomain of the BRD4 protein (BRD4(1)), has been the subject of several experimental [6, 34, 58, 61, 83, 171, 233, 239] and computational [45, 48, 169, 170, 232, 234, 242] studies in the search for high affinity ligands. As in the case of T4 lysozyme, many BRD4(1) inhibitors are relatively rigid and neutral, with clear protonation states [57]. When it comes to the application and comparison of binding free energy methods, BRD4(1) is more representative of typical drug discovery targets than lysozyme, since its ligands are more chemically diverse, and the binding site has clear access to the solvent [57, 58, 233]. There

TABLE VII: Proposed Lysozyme L99A Set benchmark data

name	PC CID <sup>h</sup>	2D	SMILES	$\Delta G^a$ (kcal/mol)	PDB code	reference
benzene <sup>b</sup>	241		c1ccccc1	-5.19 ± 0.16	181L [152], 4W52 [133]	[151]
toluene <sup>b</sup>	1140		Cc1ccccc1	-5.52 ± 0.04	4W53 [133]	[151]
ethylbenzene <sup>b</sup>	7500		CCc1ccccc1	-5.76 ± 0.07	1NHB [152], 4W54 [133]	[151]
propylbenzene <sup>b</sup>	7668		CCCc1ccccc1	-6.55 ± 0.02	4W55 [133]	[151]
butylbenzene <sup>b</sup>	7705		CCCCc1ccccc1	-6.70 ± 0.02	186L [152], 4W57 [133]	[151]
hexylbenzene <sup>b</sup>	14109		CCCCCCc1ccccc1	UNK <sup>c</sup>	4W59 [133]	[151]
p-xylene <sup>d</sup>	7809		Cc1ccc(cc1)C	-4.67 ± 0.06	187L [152]	[151]
benzofuran	9223		c1ccc2c(c1)cco2	-5.46 ± 0.03	182L [152]	[151]
thieno[2,3-c]pyridine	9224		c1cncc2c1ccs2	NB <sup>e</sup>	ND <sup>f</sup>	[143]
phenol <sup>g</sup>	996		c1ccc(cc1)O	NB <sup>e</sup>	ND <sup>f</sup>	[143, 151]

<sup>a</sup>T=302K, with compounds from [151] measured in 50mM sodium acetate at pH 5.5 and thieno[2,3-c]pyridine measured at pH 6.8 in 50 mM potassium chloride and 38% (v/v) ethylene glycol; <sup>b</sup> part of the series of [133], so larger ligands in the series induce conformational change; <sup>c</sup> unknown due to solubility limitations, but likely binds strongly; <sup>d</sup> L99A sidechain undergoes rotation; <sup>e</sup> nonbinder; <sup>f</sup> not done; <sup>g</sup> included since it is a binder in the polar cavity; <sup>h</sup> PubChem compound ID.

are a large number of high-quality cocrystal structures and experimental affinities available for this system, as well as several computational studies showing encouraging agreement with experiments [2, 3, 88]. Even with the apparent convergence of binding free energy calculations on BRD4(1), these calculations still present a few challenges, most them related to the conformational sampling of the protein in the bound and unbound states. We believe that a careful analysis of these issues, combined with rigorous free energy calculations, have the potential of making the BRD4(1) bromodomain the first “hard” protein benchmark system.

Although the ligand complexity can be reduced by choosing small and rigid ligands for a benchmark set, protein conformational sampling still can be challenging for this system. In particular, the *apo* state exhibits substantial flexibility in the ZA-loop (Fig. 3(A)) [200, 201]. Together with the BC-loop, these two regions are part of the substrate binding site, and therefore are likely to be sensitive to the presence/absence of the substrate [58]. Indeed, a conformational change in the ZA-loop in the absence of the ligand was observed by independent studies [88, 113, 200], and this can affect the calculated binding free energies [88]. In addition, long simulations of the *apo* CREBBP bromodomain suggest that slow side-chain rearrangements are possible [198, 200], and these can be another complicating factor. As we show below, each of these challenges can be properly addressed, making them an integral part of the binding free energy calculations.

### 1. The first bromodomain of BRD4 and its ligands

In living cells the BRD4(1) bromodomain binds to acetylated lysines present in histone tails [33], but it is also capable of binding a broad spectrum of small molecules, with affinities as strong as nanomolar [57]. Although the binding site of BRD4(1) is largely hydrophobic, most of its strong inhibitors engage in a hydrogen bond with the Asn140 residue, mimicking the binding mode of the acetylated lysine. Small molecule inhibitors of BRD4(1) have exceptional variety, including distinct groups such as triazoles, diazepines, acetyl-pyrroles, and isoxazoles [58, 76, 127, 233]. They also have a broad range of binding free energies, from non-binders to ligands with nanomolar affinities [76, 215], and with crystal structures available for many. Charged ligands are possible [2], but most are neutral. Many inhibitors have double or triple ring systems. This reduces the conformational freedom of the inhibitors in the bound and unbound states, providing quicker convergence for free energy calculations, and partially avoiding symmetry issues that might appear [140].

All BRD4(1) ligands bind to the protein surface [57, 58, 233], in contrast to T4 lysozyme where they bind in a buried cavity, making this target more suitable for free energy calculations that rely on a physical pathway [88, 113, 158]. This has allowed the application of methods that involve both alchemical transformations [2, 3] and the physical pulling of the ligand from the binding site, such as the Attach-Pull-Release (APR)

TABLE VIII: Proposed Lysozyme L99A/M102Q Set benchmark data

ligand	PC CID <sup>k</sup>	2D	SMILES	$\Delta G^a$ (kcal/mol)	PDB code	reference
toluene <sup>b</sup>	1140		Cc1ccccc1	-4.93	ND <sup>c</sup>	[224]
phenol	996		c1ccc(cc1)O	-5.24	1LI2 [224]	[224]
catechol <sup>h</sup>	289		c1ccc(c(c1)O)O	-4.16 ± 0.03	1XEP [84]	[224]
2-ethoxyphenol <sup>d</sup>	66755		CCOc1ccccc1O	-4.02 ± 0.03	3HU8 [19]	[19]
benzyl acetate <sup>e,f</sup>	8785		CC(=O)OCCc1ccccc1	-4.48 ± 0.16	3HUK [19]	[19]
4,5,6,7-tetrahydroindole <sup>f</sup>	57452536		c1c[nH]c2c1CCCC2	-4.61 ± 0.09	3HUA [19]	[19]
n-phenylglycinonitrile <sup>g</sup>	76372		c1ccccc1NCC#N	-5.52 ± 0.18	2RBN [19]	[19]
3-chlorophenol	7933		c1cc(cc(c1)Cl)O	-5.51	1LI3 [224]	[224]
2-methoxyphenol	460		COCc1ccccc1O	NB <sup>i</sup>	ND <sup>c</sup>	[19]
4-vinylpyridine	7502		C=Cc1ccncc1	NB <sup>i</sup>	ND <sup>c</sup>	[224]

<sup>a</sup>T=283K, with measurements done at pH 6.8 in 50 mM potassium phosphate, 200 mM potassium chloride buffer in the case of [19]; <sup>b</sup> included for symmetry with the L99A site since this (unlike phenol and benzene) binds in both; <sup>c</sup> not determined; <sup>d</sup> fails to make crystallographic hydrogen bond [19]; <sup>e</sup> multiple binding modes; <sup>f</sup> induces helix F motion; <sup>g</sup> induces flip of Val111 sidechain; <sup>h</sup> induces flip of Leu118 sidechain; <sup>j</sup> nonbinder; <sup>k</sup> PubChem compound ID.

method [88, 89, 213]. In past work, alchemical calculations of BRD4(1) binding free energies yielded RMS errors as low as  $0.8 \pm 0.2$  kcal/mol and a Pearson's correlation coefficient ( $R$ ) over 0.80 for ligands spanning a  $\sim 4$  kcal/mol range. A similar correlation is achieved using APR, but with slightly larger RMSEs. The difference apparently traces, at least in part, to the fact that the longer APR calculations detected and accounted for a conformational relaxation of the protein after removal of the ligand, which was not observed in the shorter alchemical calculations [2](P. Biggin, personal communication). The two methods are also able to determine the correct binding pose of a ligand starting from multiple possibilities generated by docking, which is another application of free energy calculations [105].

When results from APR and alchemical calculations are compared relative to the same *apo* state of BRD(1) for three different ligands [88], excellent agreement is found, with the discrepancies observed all within the estimated uncertainties of the two methods. Although slightly different protein force fields were used (AMBER99SB-ILDN on Ref. [2] versus AMBER ff14SB on Ref. [88]), this agreement — given the diverse tech-

niques employed — appears to suggest that it is possible to obtain well-converged binding free energies in this system. The fact that the APR study used much longer simulation times suggests that convergence for this system can be obtained relatively quickly, aside from the BRD4(1) transition in the *apo* state which will be discussed in the next section.

**Table IX** shows the proposed initial benchmark set for BRD4(1), including one verified non-binder, with experimental binding free energies spanning a  $\sim 4.5$  kcal/mol range. Some ligands are similar and can be treated with relative binding free energy calculations, but others present significant structural diversity, and are likely more suitable for absolute calculations. There are crystal structures available for all ligands except the non-binder Compound 8a, for which there is an available structure for a very similar ligand (Compound 4). For this initial benchmark set, all ligands are neutral, but they present different levels of complexity, ranging from fairly rigid molecules to larger molecules with several rotating groups. Except for the first two, all the other, larger ligands have already been treated with alchemical and/or APR methods, including the three ligands for

which both types of calculations were compared. Even though our set initially is considered a soft benchmark, we are optimistic that further application of free energy methods to these cases, especially with a focus on enhanced conformational sampling in the bound and unbound states, can turn this diverse set of ligands into the first protein hard benchmark.

## 2. *BRD4 computational challenge: ZA-loop flexibility*

The *apo* crystal structures of BRD4(1) have the same conformation as the complexes [59, 127], but the ZA-loop near the binding site can display backbone RMSD values around  $\sim 3.0$  Å in simulations over 100 ns [113, 200]. These simulations point to a sharp irreversible transition in this region generally after 20 ns, suggesting an initial metastable state followed by stabilization of the protein backbone in the absence of the substrate. This behavior is not observed for BRD4(1) bound to various ligands in simulations up to 350 ns long, with backbone RMSD values close to  $\sim 1.0$  Å relative to the cocrystal structures [201] (Heinzelmann et. al, unpublished results). Further work has reproduced this *apo* state transition, obtaining similar RMSD values, finding that the ZA-loop conformation seen in the *apo* crystal structures (closed state) is metastable and transitions to an *apo* open state (Fig. 3(B)) after around 50 ns [88]. It is possible that the timescale depends on the force field and/or method; regardless, the observation of this conformational transition is consistent across studies. The open state appears to be  $\sim 2.5$  kcal/mol more favorable than the closed state, according to free energy calculations performed using the AMBER ff14SB protein force field and the TIP3P water model [88]. This discrepancy between the crystal data and the simulations might arise from artifacts due to the simulation force field, or crystal contacts involving loop residues and neighboring molecules. This last possibility has been proposed before in a study which noted the inconsistency between the relatively uniform crystallographic B-factors and the high flexibility of the ZA-loop observed in simulations of the *apo* CREBBP bromodomain [198].

The free energy of the *apo* BRD4(1) closed-to-open conformational change is independent of the substrate [88], so it will not impact the ligands' *relative* binding free energies, as long as this transition is treated consistently across the calculations. This would be the case when performing small alchemical transformations on the substrate inside the binding pocket [223], since the BRD4(1) bound (closed) state will likely be preserved due to the presence of the ligand. Even though this seems to be the case for most ligands, we cannot exclude the possibility that the open conformation of BRD4(1) might be favored even with a ligand bound for *some* potential ligand. If this were the case, relative calculations inducing the conformational change would then require an additional conformational contribution to be included in the

relative binding free energies.

To obtain the ligands' *absolute* binding free energies, the conformational change of the *apo* BRD4(1) has to be taken into account. The affinities can be computed either relative to the metastable *apo* closed state, consistent with the *apo* crystal structures of BRD4(1), or the *apo* open state, consistent with the AMBER ff14SB force-field. The results for a set of ligands will differ only by an free energy offset depending on this choice. The values relative to the metastable *apo* closed state of BRD4(1) can be calculated by including a simple restraint-wall scheme on the Asp88  $\psi$  dihedral, as described in Ref. [88]. This approach avoids sampling challenges that may occur if the conformational change occurs during the removal of the ligands. The binding free energies relative to the *apo* open state can then be determined by a separate calculation of the free energy change associated with the closed to open transition. This transition appears to work through a double hinge mechanism centered on the Asp88 and Asp96 residues [88], and the free energy for the conformational change can be observed using a single reaction coordinate (the Asp88  $\psi$  dihedral), making the conformational change problem suitable for various enhanced sampling methods.

## 3. *BRD4(1) computational challenge: sidechain sampling*

Protein side-chain rearrangements could pose another significant sampling challenge for BRD4(1), as suggested by several 0.5  $\mu$ s simulations of the *apo* CREBBP bromodomain [198, 200]. In these simulations, the conserved Tyr1167 residue in the BC-loop (equivalent to Tyr139 on BRD4(1)) is shown to adopt an outward conformation where it points towards bulk solvent, followed by the occupation of the binding site by the side-chain of the conserved Tyr1125 located in the ZA-loop (Tyr97 on BRD4(1)). The hydroxyl group from this last residue mimics the acetyl lysine, making a hydrogen bond with the conserved Asn1168 residue (Asn140 on BRD4(1)). However, even with the ZA-loop transition described above, 350ns simulations of the *apo* BRD4(1) did not exhibit these slow side-chain displacements and the occupation of the binding site by the Tyr97 side-chain (Fig. 3(C)). On the other hand, as in the *apo* CREBBP bromodomain [198, 200], rotations of the Asn140  $\chi_2$  dihedral frequently take place in simulations of the *apo* BRD4(1), but they are reversible and occur on a nanosecond time scale (Fig. 3(D)). Still, it is possible that slower side-chain transitions might still take place on BRD4(1) in the absence of the ligand, which would have to be accounted for if one is looking for gold standard results for this system.

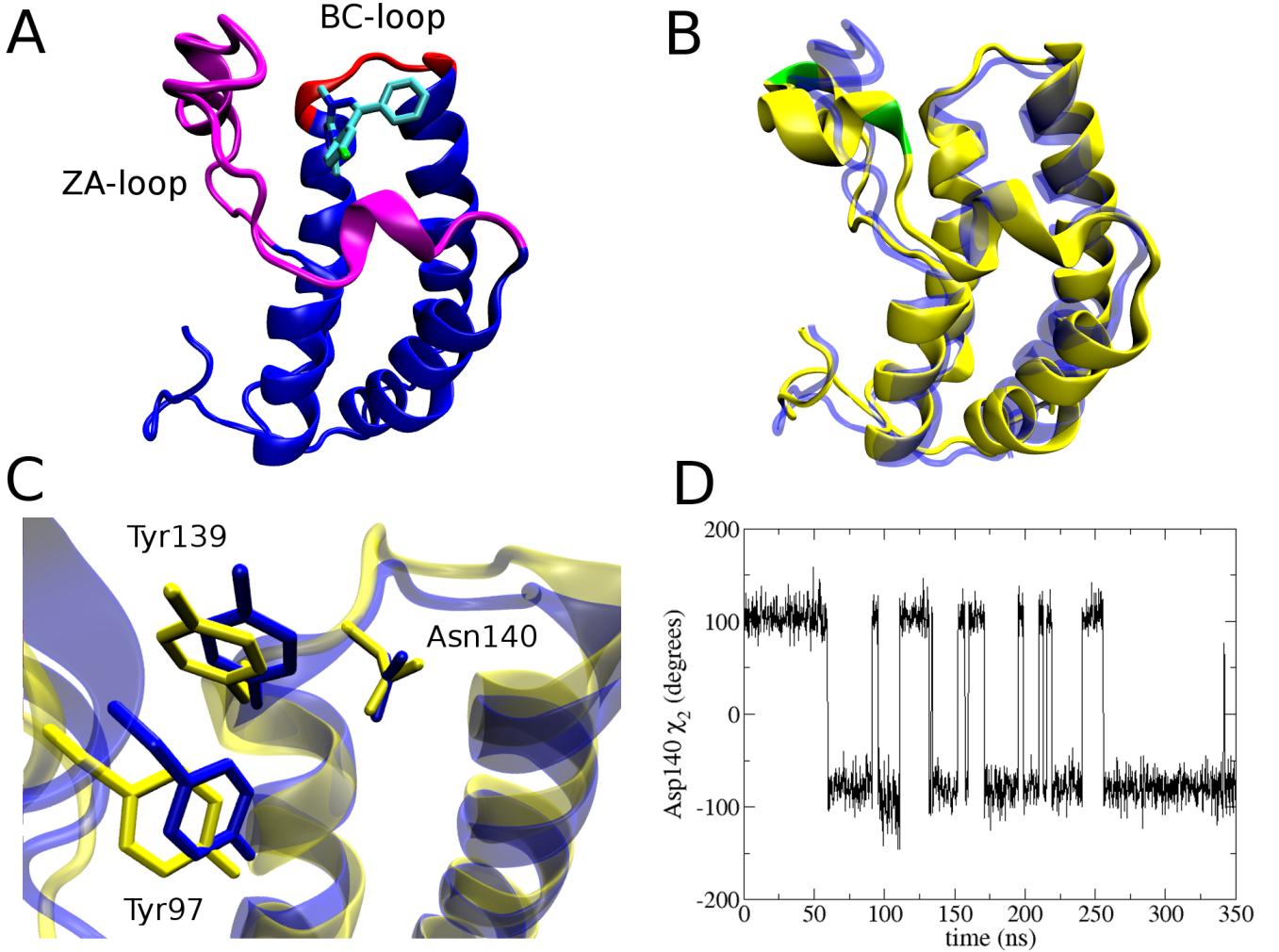


FIG. 3: (A) Crystal structure of BRD4(1) bound to the Bzt-7 ligand (PDB ID: 3U5L), showing the four helices (blue), the ZA-loop (magenta) and the BC-loop (red). (B) Comparison of the 4LYI crystal structure in the closed state (blue), and the open state obtained after 350 ns of simulations of the same structure (yellow). The two hinges, Asp88 and Asp96, are shown in green. (C) The same structures as in B, but showing the side chains of the Tyr97, Tyr139 and Asn140 residues, which do not show the rearrangements observed in the CREBBP bromodomain [200]. (D) Time series of the Asn140  $\chi_2$  dihedral during 350 ns of simulations of the *apo* BRD4(1) performed in Ref. [88] (data not shown in the paper), showing frequent and reversible transitions.

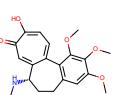
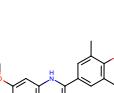
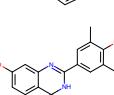
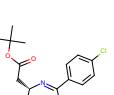
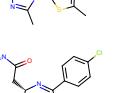
#### IV. FUTURE BENCHMARK SYSTEMS

Although the benchmark systems detailed above are useful, more systems are needed to expand the dataset, broaden the range of challenges, and bridge to biomedically relevant protein-ligand binding. The following subsections discuss additional systems that have already been used to test computational methods and that may be suited for development as new benchmark systems in the near future.

##### A. Host-Guest Systems

Depending on whether the existing host-guest systems end up being sufficient for the community's needs, there may be further opportunities for host-guest benchmark systems. The cyclodextrins (CDs; Section III A) are a particularly promising source of additional host-guest benchmark sets; one interesting possibility may be new CD derivatives, as the CDs appear easier to derivatize than the cucurbiturils. In particular, varied substituents may be appended by reactions involving secondary hydroxyls at the wide entry and primary hydroxyls at the narrow entry [50, 66, 101, 168, 204], with the caveat that generating pure products can be difficult, because there

TABLE IX: Proposed BRD4(1) Set benchmark data

ligand	PC	CID <sup>e</sup>	2D	SMILES	$\Delta G$ (kcal/mol)	PDB code	reference
Compound 8a	2295889			Cc1cc(n2c(n1)ncn2)c3cccc3	NB <sup>a,d</sup>	ND <sup>c</sup>	[215]
Compound 4	975713			Cc1cc(n2c(n1)nc(n2)N)c3cccc3	NS <sup>a,f</sup>	4MEQ [215]	[215]
XD1	234105			CC(=O)N[C@H]1CCc2cc(c(c(c2-c3c1cc(=O)c(cc3)O)OC)OC)OC	-5.95	<sup>b</sup> 4LYS [127]	[127]
Quinazolin	2927305			CN1Cc2cc(ccc2NC1=O)Br	-6.36 ± 0.17 <sup>a</sup>	4HBV [61]	[61]
Alprazolam	2118			Cc1nncc2n1-c3ccc(cc3C(=NC2)c4cccc4)Cl	-7.40 ± 0.05 <sup>b</sup>	3U5J [58]	[58]
RVX-208	24871506			Cc1cc(cc(c10CCO)C)c2[nH]c3cc(cc(c3c(=O)n2)OC)OC	-7.84 ± 0.06 <sup>b</sup>	4MR4 [165]	[165]
Bzt-7	20350618			Cc1nncc2n1-c3ccc(cc3C(=NN2C)c4cccc4)Cl	-8.16 ± 0.03 <sup>b</sup>	3U5L [58]	[58]
RVX-OH	24962985			Cc1cc(cc(c10C)C)c2[nH]c(=O)c3c(n2)cc(cc3O)OC	-8.99 ± 0.07 <sup>b</sup>	4MR3 [165]	[165]
JQ1(+)	46907787			Cc1c(sc-2c1C(=N[C@H](c3n2c(nn3)C)CC(=O)OC(C)(C)C)c4ccc(cc4)Cl)C	-9.64 ± 0.02 <sup>b</sup>	3MXF [60]	[60]
Compound 3	N/A			Cc1c(sc-2c1C(=N[C@H](c3c2c(no3)C)CC(=O)N)c4ccc(cc4)Cl)C	-10.41 ± 0.23 <sup>a</sup>	4LRG [76]	[76]

<sup>a</sup> Obtained using Alphascreen; <sup>b</sup> Obtained using Isothermal Titration Calorimetry; <sup>c</sup> not determined; <sup>d</sup> Non-binder (inactive at 250µM); <sup>e</sup> PubChem compound ID.; <sup>f</sup> No saturation (32% inhibition at 250µM). Experimental values for cases <sup>d</sup> and <sup>f</sup> may not be as reliable as required for detailed evaluation of computational accuracy, but they are nonetheless of interest as at least qualitatively informative cases. In addition, they may be useful in studies aimed at methodological validation and evaluating conformational sampling.

are so many hydroxyls that may be modified. Binding data on CD derivatives could be quite useful as a means of adding chemical diversity to host-guest benchmark sets, and such data are already available for some derivatives (see, e.g., [55, 174]).

A number of additional CD (and other host-guest) binding systems are available for download in electronic format at the BindingDB [82, 126] website ([www.bindingdb.org/bind/HostGuest.jsp](http://www.bindingdb.org/bind/HostGuest.jsp))

## B. Protein-Ligand Systems

On the order of a million experimental protein-ligand binding measurements are currently accessible through open-access databases, notably BindingDB [126], ChEMBL [13] and PubChem [110]. These databases can be valuable sources of, or at least starting points for, new protein-ligand benchmark datasets. In fact, automated procedures have already been used to extract about 700 downloadable validation sets ([www.bindingdb.org/validation](http://www.bindingdb.org/validation)

[sets/index.jsp](#)), each comprising a congeneric series of ~10-50 ligands with binding data against a defined protein, and a cocrystal structure for at least one of the ligands in the series. However, there is still a need for a smaller collection of highly optimized benchmark sets as research foci for the computational chemistry community. Such sets should exemplify specific challenges not well covered by existing benchmark systems; be based upon high quality binding measurements for at least 20 ligands, measured by consistent procedures across all ligands; and include crystal structures for the apo-protein and cocrystal structures for multiple ligands. Analysis and extraction of such sets from the big databases is a promising future direction. Here, however, we take the less systematic but more expedient approach of considering several systems that have already proven themselves to be computationally tractable and informative.

### 1. Constructed Binding Sites in Cytochrome C Peroxidase

Two artificial binding sites have been designed into the enzyme cytochrome C peroxidase (CCP): the “closed” site, created by the mutation W191G [62, 177]; and the “open” site mutant, created by supplementing mutation W191G with partial deletion of a loop, and thus opening the site to the outside of the protein [157, 177, 181]. As for the artificial binding cavities in T4 lysozyme (Section III B), the two CCP sites bind simple, fragment-like ligands with modest binding free energies (e.g., -3 to -7 kcal/mol for the open site [177]); discovery of new ligands is relatively straightforward [20, 177]; and new crystal structures can be obtained fairly readily. The protein has relatively modest size (around 280 residues) as in, for example, the 1KXM structure.

What is new is that the CCP binding sites contain an ASP which appears to be charged, at least in the presence of some ligands, as evidenced by observed interactions in structures, and by experiments where replacing the ASP with ASN abolishes binding of imidazoles, despite minimal changes to the binding site structure [63]. This side-chain interacts at short range with the ligands, which may be cationic or polar neutral compounds. As a consequence, the CCP sites challenge the ability of computational methods to accurately account for strong electrostatic interactions in the low-dielectric interior of a protein [177]. Indeed, free energy calculations with two different force fields and distinct computational methods were found to overestimate the range of affinities, across a series of ligands, about three-fold, for the closed site [7, 8], and a similar pattern appears to hold for the open site [177]. Further analysis suggests that these overestimates stem from overestimated electrostatic interactions in these buried sites [177], perhaps because the force fields used did not account explicitly for electronic polarization. Additionally, the fact that different CCP ligands have different net charges may pose methodological challenges for some free energy techniques [178]. One

modeling challenge is that CCP contains a heme, providing additional setup challenges (though since this is such an important cofactor, these may be worthwhile to face). In summary, these model CCP binding sites appear challenging yet tractable, and have already yielded insight regarding possible directions for force field improvements. Thus, they are good candidates to provide new benchmark sets in the near future.

### 2. Thrombin

Human thrombin, an enzyme of about 300 residues, is interesting both as a drug target related to blood coagulation and as a representative serine protease. Binding and structural data are available for a wide variety of ligands ([10, 167, 199, 211] and others), and have already been used as the basis for free energy simulations [22, 222, 223]. Some of the experimental binding data, obtained calorimetrically, highlight interesting trends, such as non-additivity (positive coupling) between substitutions at different locations on the chemical scaffolds of two compound series [10].

One computational study obtained relative free energies for two compound series that largely captured experimental trends and achieved an overall mean unsigned error of 0.74 kcal/mol over a range of roughly 5 kcal/mol. However, the accuracy of the results differed significantly between the two series [22]. Another relative binding free energy study for thrombin found that the results changed, depending on the initial conformation of the ligands, and showed that enhanced sampling techniques could reduce the dependence on starting conformation [222]. These prior studies suggest thrombin may be at a “sweet spot” for benchmark systems, where the system is relatively tractable, and encouraging results have been obtained, but where there are still clear challenges.

### 3. Other protein-ligand systems

Several other systems may be of possible interest because of the wealth and quality of experimental data, the extent of prior computational work, or the combination of pharmaceutical relevance with important challenges. None of these seem to be well-characterized or well-studied enough yet to be benchmark systems, but they may be interesting choices for the future. The first system we would put in this category is trypsin, and especially binding of benzamidine and its derivatives. While trypsin has been the focus of several free energy studies [43, 99, 100, 206, 216] and the SAMPL3 challenge [159, 194], it also appears to be subject to extremely slow protein motions (on the tens of microsecond timescale) which are only beginning to be characterized [166]. As a consequence, short free energy calculations may appear to converge [99, 100, 206], but

longer calculations can reveal a profound lack of convergence [166]. Therefore, trypsin may become a good benchmark for studying binding that is coupled with slow conformational dynamics; enhanced sampling methods and Markov state models [166] may be particularly helpful here.

HIV integrase is also of potential interest as a benchmark system. Indeed, blind predictions for a set of fragments binding to several sites formed part of the basis of the SAMPL4 challenge [145, 164]. Although the range of affinities in this set is modest (200 to 1450  $\mu M$  [164]), many crystal structures are available, as is a wealth of verified nonbinders [164]. One group actually did remarkably well using free energy calculations to recognize binding modes and predict nonbinders [70, 145]. Coupled with other HIV integrase data available in the literature and binding databases for larger ligands or other series, this may make this system attractive for future benchmarks.

The protein FKBP has also been the focus of several different free energy studies on the same series of ligands over the years [67, 68, 97, 118, 187, 219, 238]. However, in many cases, differences in calculated values between different studies – even with what appears to be the same force field and system preparation – are larger than differences relative to experiment [67, 97]. Some challenges are clear, such as conformational sampling for some of the larger ligands [187], and there have been some suggestions of force field issues and long equilibration times for the protein [68]. Perhaps this may be suitable as a benchmark system in the near-term as well.

A number of other proteins also have strong potential to generate useful benchmark sets. For example, free energy calculations have been carried out for influenza neuraminidase inhibitors [196] with some success [137], but other series include more complicated structure-activity relationships and are associated with protein loop motions that may be difficult to model [107]. Periplasmic oligopeptide binding protein A (OppA) binds a series of two to five-residue peptides, for which there exists a large amount of calorimetric and crystallographic binding data [42, 195, 207], and the system appears challenging but potentially tractable for free energy calculations [132]. The JNK kinase may pose an interesting conformational sampling challenge, due to its slow interconversion between binding modes for some ligands [106]. And the ongoing series of Drug Design Data Resource (D3R) [74] blinded challenges may also be source of informative protein-ligand systems ([drugdesign-data.org/about/datasets](http://drugdesign-data.org/about/datasets)) that are familiar to the computational chemistry community. As noted above, however, many other protein-ligand binding systems have been characterized experimentally, and a systematic filtering would undoubtedly yield more benchmark candidates.

## V. HOW TO USE BENCHMARK SYSTEMS

Benchmark systems will have multiple uses, as discussed above, but not all benchmark systems can cover all uses. Some will be particularly valuable for testing accuracy relative to experiment. For this purpose, relatively large numbers of ligands are needed to afford meaningful statistics. Other benchmark systems will be more useful for testing sampling techniques, and still others will, at least initially, serve as test beds to determine the sensitivity of computational results to various factors.

In our view, benchmark systems will serve also to help design careful computational experiments. For example, researchers can test whether a particular method is sampling the motions which others have already shown to be important, or how the choice of starting conformation impacts the rate of convergence to a known, gold standard value for a particular force field and system composition. The availability of benchmark systems will also facilitate comparisons where only a single piece of a workflow is modified. For example, one may ask how results change if a different protonation state assignment tool is used to prepare a protein. Of course, such comparisons can already be done, but the results will be far more useful in the context of generally accepted and widely used test cases.

We hope that, ultimately, results from reliable “gold standard” binding free energy computations will be available for a set of benchmark systems. These would be from fully converged binding free energy calculations, and give correct results for a particular force field and system preparation, allowing quantitative comparison of the force field results with experiment. Such results will also facilitate a great deal of science on method efficiency, as new methods which purport to be more efficient could easily and *automatically* be run on a standard set of systems to see how much more efficient they are than the (perhaps brute-force) method which yielded the gold standard results. Thus, various enhanced sampling methods could easily be observed to have strengths and weaknesses on known problem classes. These systems will allow automated testing of the efficiency of new methods on real-world problems.

## VI. WE NEED WORKFLOW SCIENCE

While the benchmark systems discussed here will already be useful, to fully realize their benefits a great deal of engineering needs to be done to facilitate workflow science. Currently, a wide variety of computational tools are available for different stages of the free energy calculation process, from system preparation (protonation state assignment, building in missing residues and loops, adding counterions, etc.) to force field assignment, to planning and conducting the calculations themselves (choice of method, simulation package, and so on). Often, each set of tools lives in its own ecosystem and is not

designed to be easily interchangeable with tools from another ecosystem. This makes it very difficult to systematically compare methods that differ only in one respect; instead, one must adopt an entirely different toolset to change one aspect of a procedure. For example, swapping different tools for assigning protein protonation states could yield valuable insights into the relative merits of these tools and the importance of protonation at specific residues, but currently, this is, at best, an arduous task.

### A. Workflow automation is needed

At the most basic level, we need to allow calculations to be easily repeated on all of the benchmark systems via automated workflows. One should not have to become an expert in the systems being studied in order to be able to successfully apply calculations to them; inputs should be easily available and repeating calculations should become fully automated so that a new method can be tested by simply specifying the set of benchmarks to run on.

To achieve this, at least two major innovations are needed. First, we need automated workflows that can proceed from the specification of a system to target to yielding the desired results without human intervention. Second, we need a standard data structure for input to and output from these workflows so that people can easily obtain inputs for benchmark systems and only change the component they want to change (such as the force field or system preparation) and leave the other components unchanged so that, in an automated manner, they can focus their testing on only the components they want to test.

### B. Analysis automation will also be needed

At the most basic level, we can simply check whether we are getting the expected answer for each calculation performed, at least for systems where a gold standard result is available. However, this does not provide nearly enough insight, especially in cases of failure, where we would like insight into *why* we failed. Are the relevant motions being sampled? Do we have the right protonation states and binding modes? Many other factors may need to be considered. We need ways to automatically check that we are sampling the right motions, identifying correct binding modes/conformations, and so on, without having to become experts on the specific systems examined. Probably we will need to define ways to automatically specify what order parameters should be monitored to assess for adequate sampling.

### C. Modularization will be key

To achieve these goals, researchers should develop or package tools so that they take a set of well specified

inputs and provide well specified outputs in an interchangeable way. This may involve containerizing key pieces of workflows such as in Docker [49] or Singularity [114] containers, and developing standards as to what inputs and outputs are provided to each component of the workflow. Another key goal of modularization is to separate the *operator* from the *method*. Currently, binding calculations are most often done by a human expert, who makes a variety of decisions along the way (though Schrödinger’s workflow represents real progress toward automation [223]), making it difficult to separate the importance of human expertise from the merits of the methods employed. Containerizing and modularization will be key for this, allowing methods to be employed only in a well-defined way which is reproducible. It is this type of science – coupled with benchmark tests – which is needed to advance the field.

## VII. THE FUTURE OF THIS WORK

This work has so far presented a small group of benchmarks for binding free energy calculations, highlighted some of the ways in which they have already proven their utility, and has suggested some potential future benchmark systems. We hope that the community will become involved in identifying, characterizing, and helping to select additional benchmark systems, both from those proposed here as well as from systems which are currently being studied or which we have overlooked. We seek community input to help characterize, identify and share such systems. We also expect that there may be community interest in test systems specifically selected to challenge sampling algorithms, without reference to experimental data.

In order to provide for updates of this material as new benchmark systems are defined, and to enable community input into the process of choosing them, we have made the LaTeX source for this article on GitHub at <http://www.github.com/mobleylab/benchmarksets>, with each version having a permanent DOI assigned by Zenodo. We encourage use of the GitHub issue tracker for discussion, comments, and proposed updates. We are incorporating new material via GitHub as one would for a coding project, and then making releases available both on GitHub and in more permanent form at eScholarship at <http://escholarship.org/uc/item/9p37m6bq>. Given substantial changes to this initial version of the paper, it may ultimately be appropriate to make it available as a “perpetual review” [146] via another forum allowing versioned updates of publications.

Ideally, we might also update this work in the future with results from “standard” calculations on the benchmark systems discussed here, along with links to code to allow reproduction of those calculations.

## VIII. CONCLUSIONS AND OUTLOOK

Binding free energy calculations are a promising tool for predicting and understanding molecular interactions and appear to have enough accuracy to provide substantial benefits in a pharmaceutical drug discovery context. However, progress is needed to improve these tools so that they can achieve their potential. To achieve steady progress, and to avoid potentially damaging cycles of enthusiasm and disillusionment, we need to understand and be open and honest about key challenges. Community adoption of well-chosen benchmark systems is vital for this, as it will allow researchers to rigorously test and compare methods, arrive at a shared understanding of problems, and measure progress on well-characterized yet challenging systems. It is also worth emphasizing the importance of sharing information about apparently well thought-out and even promising methods that do *not* work, rather than sharing only what does appear to work. Identifying and addressing failure cases and problems is critically important to advancing this technology, but failures can be harder to publish, and may even go unpublished, even though they serve a unique role in advancing the field. We therefore strongly encourage that such results be shared and welcomed by the research community. Potentially, the GitHub repository connected with this perpetual review paper could serve as a place to deposit input files and summary results of tests on these benchmark systems, with summary information perhaps being included in this work itself or topical sub-reviews within the same repository. Host-guest input files are now available there.

Here, we have proposed several benchmark systems for binding free energy calculations. These embody a subset of the key challenges facing the field, and we plan to expand the set as consensus emerges. Hopefully, these systems will serve as challenging standard test cases for

new methods, force fields, protocols, and workflows. Our desire is that these benchmarks will advance the science and technology of modeling and predicting molecular interactions, and that other researchers in the field will contribute to identifying new benchmark sets and updating the information provided about these informative systems.

## DISCLOSURE STATEMENT

D.L.M. is a member of the Scientific Advisory Board for OpenEye Scientific Software. M.K.G. is a cofounder and has equity interest in the company VeraChem LLC.

## ACKNOWLEDGMENTS

DLM appreciates financial support from the National Institutes of Health (NIH; 1R01GM108889-01) and the National Science Foundation (NSF; CHE 1352608). MKG thanks the NIH for partial support of this work through grants R01GM061300, R01GM070064 and U01GM111528. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIH or the NSF.

We also appreciate helpful discussions with a huge number of people in the field, including a wide variety of participants at recent meetings such as the 2016 Workshop on Free Energy Methods in Drug Discovery. Conversations with John Chodera (MSKCC), Chris Oostenbrink (BOKU), Julien Michel (Edinburgh), Robert Abel (Schrödinger), Bruce Gibb (Tulane), Matt Sullivan (Tulane), and Lyle Isaacs (Maryland) were particularly helpful. We thank David Slochower (UCSD) for a critical reading of the manuscript, and others listed on our GitHub repository for their contributions.

- [1] R. Abel. Accelerating drug discovery with free energy calculations. [http://www.alchemy.org/wiki/images/e/eb/Vertex\\_talk\\_5\\_15\\_2016\\_clean3.pdf](http://www.alchemy.org/wiki/images/e/eb/Vertex_talk_5_15_2016_clean3.pdf), May 2016.
- [2] M. Aldeghi, A. Heifetz, M. J. Bodkin, S. Knapp, and P. C. Biggin. Accurate calculation of the absolute free energy of binding for drug molecules. *Chem. Sci.*, 7(1):207–218, 2016.
- [3] M. Aldeghi, A. Heifetz, M. J. Bodkin, S. Knapp, and P. C. Biggin. Predictions of Ligand Selectivity from Absolute Binding Free Energy Calculations. *J. Am. Chem. Soc.*, 139(2):946–957, Jan. 2017.
- [4] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford Science Publications. Oxford University Press, New York, NY, June 1989.
- [5] K. I. Assaf and W. M. Nau. Cucurbiturils: From synthesis to high-affinity binding and catalysis. *Chem Soc Rev*, 44(2):394–418, Jan. 2015.
- [6] P. Bamborough, C.-w. Chung, R. C. Furze, P. Grandi, A.-M. Michon, R. J. Sheppard, H. Barnett, H. Diallo, D. P. Dixon, C. Douault, E. J. Jones, B. Karamshi, D. J. Mitchell, R. K. Prinjha, C. Rau, R. J. Watson, T. Werner, and E. H. Demont. Structure-Based Optimization of Naphthyridones into Potent ATAD2 Bromodomain Inhibitors. *J. Med. Chem.*, 58(15):6151–6178, Aug. 2015.
- [7] S. Banba and C. L. Brooks III. Free energy screening of small ligands binding to an artificial protein cavity. *The Journal of Chemical Physics*, 113(8):3423–3433, Aug. 2000.
- [8] S. Banba, Z. Guo, and C. L. Brooks III. Efficient Sampling of Ligand Orientations and Conformations in Free Energy Calculations Using the  $\lambda$ -Dynamics Method. *J. Phys. Chem. B*, 104(29):6903–6910, July 2000.
- [9] C. C. Bannan, K. H. Burley, M. Chiu, M. R. Shirts, M. K. Gilson, and D. L. Mobley. Blind prediction of cyclohexane–water distribution coefficients from the SAMPL5 challenge. *J. Comput. Aided Mol. Des.*, pages

- 1–18, Sept. 2016.
- [10] B. Baum, L. Muley, M. Smolinski, A. Heine, D. Hangauer, and G. Klebe. Non-additivity of Functional Group Contributions in Protein–Ligand Binding: A Comprehensive Study by Crystallography and Isothermal Titration Calorimetry. *Journal of Molecular Biology*, 397(4):1042–1054, Apr. 2010.
- [11] I. Beà, C. Jaime, and P. Kollman. Molecular recognition by  $\beta$ -cyclodextrin derivatives: Molecular dynamics, free-energy perturbation and molecular mechanics/Poisson–Boltzmann surface area goals and problems. *Theor Chem Acc*, 108(5):286–292, Nov. 2002.
- [12] A. F. Bell, L. Hecht, and L. D. Barron. New Evidence for Conformational Flexibility in Cyclodextrins from Vibrational Raman Optical Activity. *Chemistry - A European Journal*, 3(8):1292–1298, Aug. 1997.
- [13] A. P. Bento, A. Gaulton, A. Hersey, L. J. Bellis, J. Chambers, M. Davies, F. A. Krüger, Y. Light, L. Mak, S. McGlinchey, M. Nowotka, G. Papadatos, R. Santos, and J. P. Overington. The ChEMBL bioactivity database: An update. *Nucl Acids Res*, 42(D1):D1083–D1090, Jan. 2014.
- [14] S. Bhakat and P. Söderhjelm. Resolving the problem of trapped water in binding cavities: Prediction of host-guest binding free energies in the SAMPL5 challenge by funnel metadynamics. *J Comput Aided Mol Des*, 2016.
- [15] F. Biedermann, W. M. Nau, and H.-J. Schneider. The Hydrophobic Effect RevisitedStudies with Supramolecular Complexes Imply HighEnergy Water as a Noncovalent Driving Force. *Angewandte Chemie International Edition*, 53(42):11158–11171, Oct. 2014.
- [16] S. Boresch, F. Tettinger, M. Leitgeb, and M. Karplus. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *The Journal of Physical Chemistry B*, 107(35):9535–9551, Sept. 2003.
- [17] S. Bosisio, A. S. J. S. Mey, and J. Michel. Blinded predictions of host-guest standard free energies of binding in the SAMPL5 challenge. *J Comput Aided Mol Des*, 2016.
- [18] G. Bouvignies, P. Vallurupalli, D. F. Hansen, B. E. Correia, O. Lange, A. Bah, R. M. Vernon, F. W. Dahlquist, D. Baker, and L. E. Kay. Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature*, 477(7362):111–114, Sept. 2011.
- [19] S. E. Boyce, D. L. Mobley, G. J. Rocklin, A. P. Graves, K. A. Dill, and B. K. Shoichet. Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. *J. Mol. Biol.*, 394(4):747–763, Dec. 2009.
- [20] R. Brenk, S. W. Vetter, S. E. Boyce, D. B. Goodin, and B. K. Shoichet. Probing Molecular Docking in a Charged Model Binding Site. *Journal of Molecular Biology*, 357(5):1449–1470, Apr. 2006.
- [21] W. Cai, T. Sun, P. Liu, C. Chipot, and X. Shao. Inclusion Mechanism of Steroid Drugs into  $\beta$ -Cyclodextrins. Insights from Free Energy Calculations. *J. Phys. Chem. B*, 113(22):7836–7843, June 2009.
- [22] G. Calabò, C. J. Woods, F. Powlesland, A. S. J. S. Mey, A. J. Mulholland, and J. Michel. Elucidation of Nonadditive Effects in Protein–Ligand Binding Energies: Thrombin as a Case Study. *J. Phys. Chem. B*, June 2016.
- [23] L. Cao and L. Isaacs. Absolute and relative binding affinity of cucurbit[7]uril towards a series of cationic guests. *Supramolecular Chemistry*, 26(3-4):251–258, Mar. 2014.
- [24] L. Cao, M. Šekutor, P. Y. Zavalij, K. Mlinarić-Majerski, R. Glaser, and L. Isaacs. Cucurbit[7]uril-Guest Pair with an Attomolar Dissociation Constant. *Angew. Chem. Int. Ed.*, 53(4):988–993, Jan. 2014.
- [25] R. S. Carnegie, C. L. D. Gibb, and B. C. Gibb. Anion Complexation and The Hofmeister Effect. *Angew. Chem.*, 126(43):11682–11684, Oct. 2014.
- [26] J. Carrazana, A. Jover, F. Meijide, V. H. Soto, and J. Vázquez Tato. Complexation of Adamantyl Compounds by  $\beta$ -Cyclodextrin and Monoaminoderivatives. *J. Phys. Chem. B*, 109(19):9719–9726, May 2005.
- [27] C. Cézard, X. Trivelli, F. Aubry, F. Djedaini-Pilard, and F.-Y. Dupradeau. Molecular dynamics studies of native and substituted cyclodextrins in different media: 1. Charge derivation and force field performances. *Physical Chemistry Chemical Physics*, 13(33):15103–15121, 2011.
- [28] W. Chen, C.-E. Chang, and M. K. Gilson. Calculation of Cyclodextrin Binding Affinities: Energy, Entropy, and Implications for Drug Design. *Biophysical Journal*, 87(5):3035–3049, Nov. 2004.
- [29] J. D. Chodera, D. L. Mobley, M. R. Shirts, R. W. Dixon, K. Branson, and V. S. Pande. Alchemical free energy methods for drug discovery: Progress and challenges. *Curr Opin Struct Biol*, 21(2):150–160, Feb. 2011.
- [30] C. D. Christ. Binding affinity prediction from molecular simulations: A new standard method in structure-based drug design? [dx.doi.org/10.7490/f1000research.1112651.1](https://dx.doi.org/10.7490/f1000research.1112651.1), May 2016.
- [31] C. D. Christ and T. Fox. Accuracy Assessment and Automation of Free Energy Calculations for Drug Design. *J. Chem. Inf. Model.*, 54(1):108–120, Jan. 2014.
- [32] C. D. Christ, A. E. Mark, and W. F. van Gunsteren. Basic ingredients of free energy calculations: A review. *J. Comput. Chem.*, 31(8):1569–1582, June 2010.
- [33] C.-w. Chung. *Progress in Medicinal Chemistry*, volume 51. Elsevier, Cambridge, MA, 2012.
- [34] C.-w. Chung, H. Coste, J. H. White, O. Mirquet, J. Wilde, R. L. Gosmini, C. Delves, S. M. Magny, R. Woodward, S. A. Hughes, E. V. Boursier, H. Flynn, A. M. Bouillot, P. Bamborough, J.-M. G. Brusq, F. J. Gellibert, E. J. Jones, A. M. Riou, P. Homes, S. L. Martin, I. J. Uings, J. Toum, C. A. Clément, A.-B. Boullay, R. L. Grimley, F. M. Blandel, R. K. Prinjha, K. Lee, J. Kirillovsky, and E. Nicodeme. Discovery and Characterization of Small Molecule Inhibitors of the BET Family Bromodomains. *J. Med. Chem.*, 54(11):3827–3838, June 2011.
- [35] D. J. Cole, J. Tirado-Rives, and W. L. Jorgensen. Molecular dynamics and Monte Carlo simulations for protein–ligand binding and inhibitor design. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1850(5):966–971, May 2015.
- [36] M. D. Collins, M. L. Quillin, G. Hummer, B. W. Matthews, and S. M. Gruner. Structural Rigidity of a Large Cavity-containing Protein Revealed by High-pressure Crystallography. *Journal of Molecular Biology*, 367(3):752–763, Mar. 2007.
- [37] J. Comer, K. Schulten, and C. Chipot. Calculation of Lipid-Bilayer Permeabilities Using an Average Force. *J. Chem. Theory Comput.*, 10(2):554–564, Feb. 2014.
- [38] H. Cong, X. L. Ni, X. Xiao, Y. Huang, Q.-J. Zhu, S.-F.

- Xue, Z. Tao, L. F. Lindoy, and G. Wei. Synthesis and separation of cucurbit[n]urils and their derivatives. *Org. Biomol. Chem.*, 14(19):4335–4364, May 2016.
- [39] K. A. Connors. The Stability of Cyclodextrin Complexes in Solution. *Chem. Rev.*, 97(5):1325–1358, Aug. 1997.
- [40] E. S. Cotner and P. J. Smith. Phosphotyrosine Binding by Ammonium- and Guanidinium-Modified Cyclodextrins. *J. Org. Chem.*, 63(5):1737–1739, Mar. 1998.
- [41] G. Cui. Affinity Predictions with FEP+: A Different Perspective on Performance and Utility. [dx.doi.org/10.7490/f1000research.1112773.1](https://dx.doi.org/10.7490/f1000research.1112773.1), May 2016.
- [42] T. G. Davies, R. E. Hubbard, and J. R. Tame. Relating structure to thermodynamics: The crystal structures and binding affinity of eight OppA-peptide complexes. *Protein Science*, 8(7):1432–1444, Jan. 1999.
- [43] A. de Ruiter and C. Oostenbrink. Efficient and Accurate Free Energy Calculations on Trypsin Inhibitors. *J. Chem. Theory Comput.*, 8(10):3686–3695, Oct. 2012.
- [44] A. de Ruiter and C. Oostenbrink. Protein–Ligand Binding from Distancefield Distances and Hamiltonian Replica Exchange Simulations. *J. Chem. Theory Comput.*, 9(2):883–892, Feb. 2013.
- [45] V. Deepak, B. Wang, D. Koot, A. Kasonga, X. X. Stander, M. Coetzee, and A. Stander. In silico design and bioevaluation of selective benzotriazepine BRD4 inhibitors with potent antiosteoclastogenic activity. *Chem. Biol. Drug. Des.*, 90(1):97–111, Feb. 2017.
- [46] N. Deng, S. Forli, P. He, A. Perryman, L. Wickstrom, R. S. K. Vijayan, T. Tiefenbrunn, D. Stout, E. Gallicchio, A. J. Olson, and R. M. Levy. Distinguishing Binders from False Positives by Free Energy Calculations: Fragment Screening Against the Flap Site of HIV Protease. *J. Phys. Chem. B*, 119(3):976–988, Jan. 2015.
- [47] Y. Deng and B. Roux. Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant. *Journal of Chemical Theory and Computation*, 2(5):1255–1273, Sept. 2006.
- [48] B. M. Dickson, P. W. de Waal, Z. H. Ramjan, H. E. Xu, and S. B. Rothbart. A fast, open source implementation of adaptive biasing potentials uncovers a ligand design strategy for the chromatin regulator BRD4. *J. Chem. Phys.*, 145:154113, Oct. 2016.
- [49] Docker Inc. What is Docker? <https://www.docker.com/what-docker>, 2015-05-14T16:17:40-07:00.
- [50] H. Dodziuk. *Cyclodextrins and Their Complexes: Chemistry, Analytical Methods, Applications*. John Wiley & Sons, 2006.
- [51] J. B. Dunbar, R. D. Smith, C.-Y. Yang, P. M.-U. Ung, K. W. Lexa, N. A. Khazanov, J. A. Stuckey, S. Wang, and H. A. Carlson. CSAR Benchmark Exercise of 2010: Selection of the Protein–Ligand Complexes. *J. Chem. Inf. Model.*, 51(9):2036–2046, Sept. 2011.
- [52] A. Eklund, T. E. Nichols, and H. Knutsson. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci. U.S.A.*, 113(28):7900–7905, July 2016.
- [53] A. E. Eriksson, W. A. Baase, J. A. Wozniak, and B. W. Matthews. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature*, 355(6358):371–373, Jan. 1992.
- [54] J. Ewell, B. C. Gibb, and S. W. Rick. Water Inside a Hydrophobic Cavitand Molecule. *The Journal of Physical Chemistry B*, 112(33):10272–10279, Aug. 2008.
- [55] P.-A. Faugeras, B. Boëns, P.-H. Elchinger, F. Brouillette, D. Montplaisir, R. Zerrouki, and R. Lucas. When Cyclodextrins Meet Click Chemistry. *Eur. J. Org. Chem.*, 2012(22):4087–4105, Aug. 2012.
- [56] A. T. Fenley, N. M. Henriksen, H. S. Muddana, and M. K. Gilson. Bridging Calorimetry and Simulation through Precise Calculations of Cucurbituril–Guest Binding Enthalpies. *Journal of Chemical Theory and Computation*, 10(9):4069–4078, Sept. 2014.
- [57] P. Filippakopoulos and S. Knapp. Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat. Rev. Drug Discov.*, 13:337–356, May 2014.
- [58] P. Filippakopoulos, S. Picaud, O. Fedorov, M. Keller, M. Wrobel, O. Morgenstern, F. Bracher, and S. Knapp. Benzodiazepines and benzotriazepines as protein interaction inhibitors targeting bromodomains of the bet family. *Bioorg. Med. Chem.*, 20(6):1878–1886, Mar. 2012.
- [59] P. Filippakopoulos, S. Picaud, M. Mangos, T. Keates, J.-P. Lambert, D. Barsyte-Lovejoy, I. Felletar, R. Volkmer, S. Muller, T. Pawson, A.-C. Gingras, C. H. Arrowsmith, and S. Knapp. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell*, 149(1):214–231, Mar. 2012.
- [60] P. Filippakopoulos, J. Qi, S. Picaud, Y. Shen, W. B. Smith, O. Fedorov, E. M. Morse, T. Keates, T. T. Hickman, I. Felletar, M. Philpott, S. Munro, M. R. McKeown, Y. Wang, A. L. Christie, N. West, M. J. Cameron, B. Schwartz, T. D. Heightman, N. La Thangue, C. A. French, O. Wiest, A. L. Kung, S. Knapp, and J. E. Bradner. Selective inhibition of BET bromodomains. *Nature*, 468(7327):1067–1073, Dec. 2010.
- [61] P. V. Fish, P. Filippakopoulos, G. Bish, P. E. Brennan, M. E. Bunnage, A. S. Cook, O. Federov, B. S. Gerstenberger, H. Jones, S. Knapp, B. Marsden, K. Nocka, D. R. Owen, M. Philpott, S. Picaud, M. J. Primiano, M. J. Ralph, N. Sciammetta, and J. D. Trzupek. Identification of a chemical probe for bromo and extra C-terminal bromodomain inhibition through optimization of a fragment-derived hit. *J. Med. Chem.*, 55(22):9831–9837, Oct. 2012.
- [62] M. M. Fitzgerald, R. A. Musah, D. E. McRee, and D. B. Goodin. A ligand-gated, hinged loop rearrangement opens a channel to a buried artificial protein cavity. *Nat Struct Mol Biol*, 3(7):626–631, July 1996.
- [63] M. M. Fitzgerald, M. L. Trester, G. M. Jensen, D. E. Mcree, and D. B. Goodin. The role of aspartate-235 in the binding of cations to an artificial cavity at the radical site of cytochrome c peroxidase. *Protein Science*, 4(9):1844–1850, Sept. 1995.
- [64] H. Flyvbjerg and H. G. Petersen. Error estimates on averages of correlated data. *The Journal of Chemical Physics*, 91(1):461, July 1989.
- [65] W. A. Freeman, W. L. Mock, and N. Y. Shih. Cucurbituril. *J. Am. Chem. Soc.*, 103(24):7367–7368, Dec. 1981.
- [66] K.-H. Frömming and J. Szejthi. Cyclodextrin Derivatives. In *Cyclodextrins in Pharmacy*, pages 19–32. 1994.
- [67] H. Fujitani, Y. Tanida, M. Ito, G. Jayachandran, C. D. Snow, M. R. Shirts, E. J. Sorin, and V. S. Pande. Direct calculation of the binding free energies of FKBP ligands. *The Journal of Chemical Physics*, 123(8):084108, Aug. 2005.
- [68] H. Fujitani, Y. Tanida, and A. Matsuura. Massively par-

- allel computation of absolute binding free energy with well-equilibrated states. *Phys. Rev. E*, 79(2):021914, Feb. 2009.
- [69] E. Gallicchio, H. Chen, H. Chen, M. Fitzgerald, Y. Gao, P. He, M. Kalyanikar, C. Kao, B. Lu, Y. Niu, M. Pethe, J. Zhu, and R. M. Levy. BEDAM binding free energy predictions for the SAMPL4 octa-acid host challenge. *J Comput Aided Mol Des*, 29(4):315–325, Mar. 2015.
- [70] E. Gallicchio, N. Deng, P. He, L. Wickstrom, A. L. Perryman, D. N. Santiago, S. Forli, A. J. Olson, and R. M. Levy. Virtual screening of integrase inhibitors by large scale binding free energy calculations: The SAMPL4 challenge. *J Comput Aided Mol Des*, 28(4):475–490, Feb. 2014.
- [71] E. Gallicchio, M. Lapelosa, and R. M. Levy. Binding Energy Distribution Analysis Method (BEDAM) for Estimation of ProteinLigand Binding Affinities. *Journal of Chemical Theory and Computation*, 6(9):2961–2977, Sept. 2010.
- [72] H. Gan, C. J. Benjamin, and B. C. Gibb. Nonmonotonic Assembly of a Deep-Cavity Cavitand. *Journal of the American Chemical Society*, 133(13):4770–4773, Apr. 2011.
- [73] K. Gao, J. Yin, N. M. Henriksen, A. T. Fenley, and M. K. Gilson. Binding Enthalpy Calculations for a Neutral Host–Guest Pair Yield Widely Divergent Salt Effects across Water Models. *Journal of Chemical Theory and Computation*, 11(10):4555–4564, Oct. 2015.
- [74] S. Gathiaka, S. Liu, M. Chiu, H. Yang, J. A. Stuckey, Y. N. Kang, J. Delproposto, G. Kubish, J. B. Dunbar, H. A. Carlson, S. K. Burley, W. P. Walters, R. E. Amaro, V. A. Feher, and M. K. Gilson. D3R grand challenge 2015: Evaluation of protein–ligand pose and affinity predictions. *J Comput Aided Mol Des*, 30(9):651–668, Sept. 2016.
- [75] J. Gebhardt and N. Hansen. Calculation of binding affinities for linear alcohols to  $\alpha$ -cyclodextrin by twin-system enveloping distribution sampling simulations. *Fluid Phase Equilibria*, 422:1–17, Aug. 2016.
- [76] V. S. Gehling, M. C. Hewitt, R. G. Vaswani, Y. Leblanc, A. Cote, C. G. Nasveschuk, A. M. Taylor, J.-C. Harmane, J. E. Audia, E. Pardo, S. Joshi, P. Sandy, J. A. Mertz, R. J. Sims III, L. Bergeron, B. M. Bryant, S. Bellon, F. Poy, H. Jayaram, R. Sankaranarayanan, S. Yellapantula, N. B. Srinivasamurthy, S. Birudukota, and B. K. Albrecht. Discovery, design, and optimization of isoxazole azepine BET inhibitors. *ACS Med. Chem. Lett.*, 4(9):835–840, July 2013.
- [77] C. L. D. Gibb and B. C. Gibb. Well-Defined, Organic Nanoenvironments in Water: The Hydrophobic Effect Drives a Capsular Assembly. *J. Am. Chem. Soc.*, 126(37):11408–11409, Sept. 2004.
- [78] C. L. D. Gibb and B. C. Gibb. Guests of differing polarities provide insight into structural requirements for templates of water-soluble nano-capsules. *Tetrahedron*, 65(35):7240–7248, Aug. 2009.
- [79] C. L. D. Gibb and B. C. Gibb. Anion Binding to Hydrophobic Concavity Is Central to the Salting-in Effects of Hofmeister Chaotropes. *Journal of the American Chemical Society*, 133(19):7344–7347, May 2011.
- [80] C. L. D. Gibb and B. C. Gibb. Binding of cyclic carboxylates to octa-acid deep-cavity cavitand. *J Comput Aided Mol Des*, 28(4):319–325, Nov. 2013.
- [81] M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCommon. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys J*, 72(3):1047–1069, Mar. 1997.
- [82] M. K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, and J. Chong. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucl. Acids Res.*, 44(D1):D1045–D1053, Apr. 2016.
- [83] R. Gosmini, V. L. Nguyen, J. Toum, C. Simon, J.-M. G. Brusq, G. Krysa, O. Mirguet, A. M. Riou-Eymard, E. V. Boursier, L. Trottet, P. Bamforth, H. Clark, C.-w. Chung, L. Cutler, E. H. Demont, R. Kaur, A. J. Lewis, M. B. Schilling, P. E. Soden, S. Taylor, A. L. Walker, M. D. Walker, R. K. Prinjha, and E. Nicodème. The Discovery of I-BET726 (GSK1324726A), a Potent Tetrahydroquinoline ApoA1 Up-Regulator and Selective BET Bromodomain Inhibitor. *J. Med. Chem.*, 57(19):8111–8131, Oct. 2014.
- [84] A. P. Graves, R. Brenk, and B. K. Shoichet. Decoys for Docking. *Journal of Medicinal Chemistry*, 48(11):3714–3728, June 2005.
- [85] A. P. Graves, D. M. Shivakumar, S. E. Boyce, M. P. Jacobson, D. A. Case, and B. K. Shoichet. Rescoring Docking Hit Lists for Model Cavity Sites: Predictions and Experimental Testing. *Journal of Molecular Biology*, 377(3):914–934, Mar. 2008.
- [86] J. C. Gumbart, B. Roux, and C. Chipot. Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? *J. Chem. Theory Comput.*, 9(1):794–802, Jan. 2013.
- [87] O. Guvench, S. S. Mallajosyula, E. P. Raman, E. Hatcher, K. Vanommeslaeghe, T. J. Foster, F. W. Jamison, and A. D. MacKerell. CHARMM Additive All-Atom Force Field for Carbohydrate Derivatives and Its Utility in Polysaccharide and Carbohydrate–Protein Modeling. *J. Chem. Theory Comput.*, 7(10):3162–3180, Oct. 2011.
- [88] G. Heinzelmann, N. M. Henriksen, and M. K. Gilson. Attach-pull-release calculations of ligand binding and conformational changes on the first BRD4 bromodomains. *J. Chem. Theory Comput.*, ASAP, May 2017.
- [89] N. M. Henriksen, A. T. Fenley, and M. K. Gilson. Computational Calorimetry: High-Precision Calculation of Host–Guest Binding Thermodynamics. *Journal of Chemical Theory and Computation*, 11(9):4377–4394, Sept. 2015.
- [90] N. M. Henriksen and M. K. Gilson. Evaluating Force Field Performance in Thermodynamic Calculations of Cyclodextrin HostGuest Binding: Water Models, Partial Charges, and Host Force Field Parameters. *Journal of Chemical Theory and Computation*, July 2017.
- [91] J. Hermans and S. Subramaniam. The free energy of xenon binding to myoglobin from molecular dynamics simulation. *Isr. J. Chem.*, 27:225–227, Jan. 1986.
- [92] M. B. Hillyer, C. L. D. Gibb, P. Sokkalingam, J. H. Jordan, S. E. Ioup, and B. C. Gibb. Synthesis of Water-Soluble Deep-Cavity Cavitands. *Org. Lett.*, 18(16):4048–4051, Aug. 2016.
- [93] N. Homeyer, F. Stoll, A. Hillisch, and H. Gohlke. Binding Free Energy Calculations for Lead Optimization: Assessment of Their Accuracy in an Industrial Drug Design Context. *J. Chem. Theory Comput.*, 10(8):3331–3344, Aug. 2014.
- [94] Y.-W. Hsiao and P. Söderhjelm. Prediction of SAMPL4

- host–guest binding affinities using funnel metadynamics. *J Comput Aided Mol Des*, 28(4):443–454, Feb. 2014.
- [95] L. Isaacs. Cucurbit[n]urils: From mechanism to structure and function. *Chem. Commun.*, (6):619–629, 2009.
- [96] L. Isaacs. Personal communication, Sept. 2016.
- [97] G. Jayachandran, M. R. Shirts, S. Park, and V. S. Pande. Parallelized-over-parts computation of absolute binding free energy with docking and molecular dynamics. *The Journal of Chemical Physics*, 125(8):084901, Aug. 2006.
- [98] W. Jiang and B. Roux. Free Energy Perturbation Hamiltonian Replica-Exchange Molecular Dynamics (FEP/H-REMD) for Absolute Ligand Binding Free Energy Calculations. *Journal of Chemical Theory and Computation*, 6(9):2559–2565, Sept. 2010.
- [99] D. Jiao, P. A. Golubkov, T. A. Darden, and P. Ren. Calculation of protein–ligand binding free energy by using a polarizable potential. *PNAS*, 105(17):6290–6295, Apr. 2008.
- [100] D. Jiao, J. Zhang, R. E. Duke, G. Li, M. J. Schnieders, and P. Ren. Trypsin-ligand binding free energies from explicit and implicit solvent simulations with polarizable potential. *J. Comput. Chem.*, 30(11):1701–1711, Aug. 2009.
- [101] Jindřich and I. Tišlerová. Simple Preparation of 3I-O-Substituted  $\beta$ -Cyclodextrin Derivatives Using Cinnamyl Bromide. *J. Org. Chem.*, 70(22):9054–9055, Oct. 2005.
- [102] W. L. Jorgensen. Quantum and statistical mechanical studies of liquids. 12. Simulation of liquid ethanol including internal rotation. *Journal of the American Chemical Society*, 103(2):345–350, Jan. 1981.
- [103] W. L. Jorgensen, J. K. Buckner, S. Boudon, and J. Tirado-Rives. Efficient computation of absolute free energies of binding by computer simulations. Application to the methane dimer in water. *The Journal of Chemical Physics*, 89(6):3742–3746, Sept. 1988.
- [104] M. Karplus and J. A. McCammon. Molecular dynamics simulations of biomolecules. *Nat Struct Mol Biol*, 9(9):646–652, Sept. 2002.
- [105] J. W. Kaus, E. Harder, T. Lin, R. Abel, J. A. McCammon, and L. Wang. How to deal with multiple binding poses in alchemical relative protein–ligand binding free energy calculations. *J. Chem. Theory Comput.*, 11(6):2670–2679, May 2015.
- [106] J. W. Kaus, E. Harder, T. Lin, R. Abel, J. A. McCammon, and L. Wang. How To Deal with Multiple Binding Poses in Alchemical Relative Protein–Ligand Binding Free Energy Calculations. *J. Chem. Theory Comput.*, 11(6):2670–2679, June 2015.
- [107] P. S. Kerry, S. Mohan, R. J. M. Russell, N. Bance, M. Niikura, and B. M. Pinto. Structural basis for a class of nanomolar influenza A neuraminidase inhibitors. *Sci Rep*, 3, Oct. 2013.
- [108] I. V. Khavrutskii and A. Wallqvist. Improved Binding Free Energy Predictions from Single-Reference Thermodynamic Integration Augmented with Hamiltonian Replica Exchange. *Journal of Chemical Theory and Computation*, 7(9):3001–3011, Sept. 2011.
- [109] W. Khuntawee, T. Rungrotmongkol, P. Wolschann, P. Pongsawasdi, N. Kungwan, H. Okumura, and S. Hannongbua. Conformation study of -cyclodextrin: Replica exchange molecular dynamics simulations. *Carbohydrate Polymers*, 141:99–105, May 2016.
- [110] S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, J. Wang, B. Yu, J. Zhang, and S. H. Bryant. PubChem Substance and Compound databases. *Nucleic Acids Res.*, 44(D1):D1202–D1213, Jan. 2016.
- [111] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, and R. J. Woods. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.*, 29(4):622–655, Mar. 2008.
- [112] R. Kitahara and F. A. A. Mulder. Is pressure-induced signal loss in NMR spectra for the Leu99Ala cavity mutant of T4 lysozyme due to unfolding? *PNAS*, 112(9):E923–E923, Mar. 2015.
- [113] M. Kuang, J. Zhou, L. Wang, Z. Liu, J. Guo, and R. Wu. Binding kinetics versus affinities in BRD4 inhibition. *J. Chem. Inf. Model.*, 55(9):1926–1935, 2015.
- [114] G. M. Kurtzer. Singularity 2.1.2 - Linux application and environment containers for science, Dec. 2016.
- [115] C. T. Lee, J. Comer, C. Herndon, N. Leung, A. Pavlova, R. V. Swift, C. Tung, C. N. Rowley, R. E. Amaro, C. Chipot, Y. Wang, and J. C. Gumbart. Simulation-Based Approaches for Determining Membrane Permeability of Small Compounds. *J. Chem. Inf. Model.*, 56(4):721–733, Apr. 2016.
- [116] J. W. Lee, H. H. L. Lee, Y. H. Ko, K. Kim, and H. I. Kim. Deciphering the Specific High-Affinity Binding of Cucurbit[7]uril to Amino Acids in Water. *The Journal of Physical Chemistry B*, 119(13):4628–4636, Apr. 2015.
- [117] J. W. Lee, S. Samal, N. Selvapalam, H.-J. Kim, and K. Kim. Cucurbituril Homologues and Derivatives: New Opportunities in Supramolecular Chemistry. *Accounts of Chemical Research*, 36(8):621–630, 2003.
- [118] M. S. Lee and M. A. Olson. Calculation of Absolute Protein-Ligand Binding Affinity Using Path and Endpoint Approaches. *Biophysical Journal*, 90(3):864–877, Feb. 2006.
- [119] G. Leonis, T. Steinbrecher, and M. G. Papadopoulos. A Contribution to the Drug Resistance Mechanism of Darunavir, Amprenavir, Indinavir, and Saquinavir Complexes with HIV-1 Protease Due to Flap Mutation I50V: A Systematic MM-PBSA and Thermodynamic Integration Study. *J. Chem. Inf. Model.*, 53(8):2141–2153, Aug. 2013.
- [120] M. T. Lerch, C. J. López, Z. Yang, M. J. Kretzman, J. Horwitz, and W. L. Hubbell. Structure-relaxation mechanism for the response of T4 lysozyme cavity mutants to hydrostatic pressure. *PNAS*, 112(19):E2437–E2446, Dec. 2015.
- [121] N. M. Lim, L. Wang, R. Abel, and D. L. Mobley. Sensitivity in binding free energies due to protein reorganization. *Journal of Chemical Theory and Computation*, 12(9):4620–4631, July 2016.
- [122] Y.-L. Lin, A. Aleksandrov, T. Simonson, and B. Roux. An Overview of Electrostatic Free Energy Computations for Solutions and Proteins. *J. Chem. Theory Comput.*, 10(7):2690–2709, July 2014.
- [123] S. Liu, S. Cao, K. Hoang, K. L. Young, A. S. Paluch, and D. L. Mobley. Using MD Simulations To Calculate How Solvents Modulate Solubility. *Journal of Chemical Theory and Computation*, 12(4):1930–1941, Feb. 2016.
- [124] S. Liu, C. Ruspic, P. Mukhopadhyay, S. Chakrabarti, P. Y. Zavalij, and L. Isaacs. The Cucurbit[n]uril Family: Prime Components for Self-Sorting Systems. *Journal of the American Chemical Society*, 127(45):15959–15967,

- Nov. 2005.
- [125] S. Liu, Y. Wu, T. Lin, R. Abel, J. P. Redmann, C. M. Summa, V. R. Jaber, N. M. Lim, and D. L. Mobley. Lead optimization mapper: Automating free energy calculations for lead optimization. *J Comput Aided Mol Des*, 27(9):755–770, Sept. 2013.
- [126] T. Liu, Y. Lin, X. Wen, R. N. Jorissen, and M. K. Gilson. BindingDB: A web-accessible database of experimentally determined protein–ligand binding affinities. *Nucl. Acids Res.*, 35(suppl 1):D198–D201, Jan. 2007.
- [127] X. Lucas, D. Wohlwend, M. Hugle, K. Schmidtkunz, S. Gerhardt, R. Schule, M. Jung, O. Einsle, and S. Guntner. 4-Acyl pyrroles: Mimicking acetylated lysines in histone code reading. *Angew. Chem. Int. Ed.*, 52(52):14055–14059, Dec. 2013.
- [128] V. Luzhkov and J. Åqvist. Free-energy perturbation calculations of binding and transition-state energies: Hydrolysis of phenyl esters by  $\beta$ -cyclodextrin. *Chemical Physics Letters*, 302(3–4):267–272, Mar. 1999.
- [129] A. Maeno, D. Sindhikara, F. Hirata, R. Otten, F. W. Dahlquist, S. Yokoyama, K. Akasaka, F. A. A. Mulder, and R. Kitahara. Cavity as a Source of Conformational Fluctuation and High-Energy State: High-Pressure NMR Study of a Cavity-Enlarged Mutant of T4Lysozyme. *Biophysical Journal*, 108(1):133–145, Jan. 2015.
- [130] A. E. Mark, S. P. van Helden, P. E. Smith, L. H. M. Janssen, and W. F. van Gunsteren. Convergence Properties of Free Energy Calculations: .alpha.-Cyclodextrin Complexes as a Case Study. *J. Am. Chem. Soc.*, 116(14):6293–6302, July 1994.
- [131] C. Márquez, R. R. Hudgins, and W. M. Nau. Mechanism of Host-Guest Complexation by Cucurbituril. *J. Am. Chem. Soc.*, 126(18):5806–5816, May 2004.
- [132] M. Maurer, S. B. A. de Beer, and C. Oostenbrink. Calculation of Relative Binding Free Energy in the Water-Filled Active Site of Oligopeptide-Binding Protein A. *Molecules*, 21(4):499, Apr. 2016.
- [133] M. Merski, M. Fischer, T. E. Balias, O. Eidam, and B. K. Shoichet. Homologous ligands accommodated by discrete conformations of a buried cavity. *PNAS*, 112(16):5039–5044, Apr. 2015.
- [134] M. Mic, A. Piřnáu, M. Bogdan, and I. Turcu. Inclusion complex of benzocaine and  $\beta$ -cyclodextrin:  $^1\text{H}$  NMR and isothermal titration calorimetry studies. In *AIP Conference Proceedings*, volume 1565, pages 63–66. AIP Publishing, Nov. 2013.
- [135] J. Michel and J. W. Essex. Hit Identification and Binding Mode Predictions by Rigorous Free Energy Simulations. *J. Med. Chem.*, 51(21):6654–6664, Nov. 2008.
- [136] J. Michel and J. W. Essex. Prediction of protein–ligand binding affinity by free energy simulations: Assumptions, pitfalls and expectations. *J Comput Aided Mol Des*, 24(8):639–658, May 2010.
- [137] J. Michel, M. L. Verdonk, and J. W. Essex. Protein-Ligand Binding Affinity Predictions by Implicit Solvent Simulations: A Tool for Lead Optimization? *J. Med. Chem.*, 49(25):7427–7439, Dec. 2006.
- [138] P. Mikulskis, D. Cioloboc, M. Andrejić, S. Khare, J. Brorsson, S. Genheden, R. A. Mata, P. Söderhjelm, and U. Ryde. Free-energy perturbation and quantum mechanical study of SAMPL4 octa-acid host–guest binding energies. *J Comput Aided Mol Des*, 28(4):375–400, Apr. 2014.
- [139] P. Mikulskis, S. Genheden, and U. Ryde. A Large-Scale Test of Free-Energy Simulation Estimates of Protein–Ligand Binding Affinities. *J. Chem. Inf. Model.*, 54(10):2794–2806, Oct. 2014.
- [140] D. L. Mobley, J. D. Chodera, and K. A. Dill. On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.*, 125:084902, Jan. 2006.
- [141] D. L. Mobley, J. D. Chodera, and K. A. Dill. Confine-and-Release Method: Obtaining Correct Binding Free Energies in the Presence of Protein Conformational Change. *Journal of Chemical Theory and Computation*, 3(4):1231–1235, July 2007.
- [142] D. L. Mobley, J. D. Chodera, L. Isaacs, and B. C. Gibb. Advancing predictive modeling through focused development of new systems to drive new modeling innovations. <http://doi.org/10.5281/zenodo.163963>, Oct. 2016.
- [143] D. L. Mobley, A. P. Graves, J. D. Chodera, A. C. McReynolds, B. K. Shoichet, and K. A. Dill. Predicting absolute ligand binding free energies to a simple model site. *J. Mol. Biol.*, 371(4):1118–1134, Aug. 2007.
- [144] D. L. Mobley and P. V. Klimovich. Perspective: Alchemical free energy calculations for drug discovery. *J. Chem. Phys.*, 137(23):230901, Jan. 2012.
- [145] D. L. Mobley, S. Liu, N. M. Lim, K. L. Wymer, A. L. Perryman, S. Forli, N. Deng, J. Su, K. Branson, and A. J. Olson. Blind prediction of HIV integrase binding from the SAMPL4 challenge. *J Comput Aided Mol Des*, 28(4):327–345, Mar. 2014.
- [146] D. L. Mobley and D. M. Zuckerman. A proposal for regularly updated review/survey articles: “Perpetual Reviews”. *arXiv:1502.01329 [cs]*, Feb. 2015.
- [147] W. L. Mock and N. Y. Shih. Host-guest binding capacity of cucurbituril. *The Journal of Organic Chemistry*, 48(20):3618–3619, Oct. 1983.
- [148] S. Moghaddam, Y. Inoue, and M. K. Gilson. Host-guest complexes with protein-ligand-like affinities: Computational analysis and design. *Journal of the American Chemical Society*, 131(11):4012–4021, Mar. 2009.
- [149] S. Moghaddam, C. Yang, M. Rekharsky, Y. H. Ko, K. Kim, Y. Inoue, and M. K. Gilson. New Ultrahigh Affinity Host-Guest Complexes of Cucurbit[7]uril with Bicyclo[2.2.2]octane and Adamantane Guests: Thermodynamic Analysis and Evaluation of M2 Affinity Calculations. *Journal of the American Chemical Society*, 133(10):3570–3581, Mar. 2011.
- [150] J. I. Monroe and M. R. Shirts. Converging free energies of binding in cucurbit[7]uril and octa-acid host–guest systems from SAMPL4 using expanded ensemble simulations. *J Comput Aided Mol Des*, 28(4):401–415, Mar. 2014.
- [151] A. Morton, W. A. Baase, and B. W. Matthews. Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. *Biochemistry*, 34(27):8564–8575, July 1995.
- [152] A. Morton and B. W. Matthews. Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: Linkage of dynamics and structural plasticity. *Biochemistry*, 34(27):8576–8588, July 1995.
- [153] H. S. Muddana, A. T. Fenley, D. L. Mobley, and M. K. Gilson. The SAMPL4 host–guest blind prediction challenge: An overview. *J Comput Aided Mol Des*, 28(4):305–317, Mar. 2014.

- [154] H. S. Muddana and M. K. Gilson. Prediction of SAMPL3 host–guest binding affinities: Evaluating the accuracy of generalized force-fields. *J Comput Aided Mol Des*, 26(5):517–525, Jan. 2012.
- [155] H. S. Muddana, C. D. Varnado, C. W. Bielawski, A. R. Urbach, L. Isaacs, M. T. Geballe, and M. K. Gilson. Blind prediction of host–guest binding affinities: A new SAMPL3 challenge. *J Comput Aided Mol Des*, 26(5):475–487, Feb. 2012.
- [156] H. S. Muddana, J. Yin, N. V. Sapra, A. T. Fenley, and M. K. Gilson. Blind prediction of SAMPL4 curcubit[7]uril binding affinities with the mining minima method. *J Comput Aided Mol Des*, 28(4):463–474, Feb. 2014.
- [157] R. A. Musah, G. M. Jensen, S. W. Bunte, R. J. Rosenfeld, and D. B. Goodin. Artificial protein cavities as specific ligand-binding templates: Characterization of an engineered heterocyclic cation-binding site that preserves the evolved specificity of the parent protein1. *Journal of Molecular Biology*, 315(4):845–857, Jan. 2002.
- [158] C. Muvva, E. R. A. Singam, S. S. Raman, and V. Subramanian. Structure-based virtual screening of novel, high-affinity BRD4 inhibitors. *Mol. Biosyst.*, 10(9):2384–2397, July 2014.
- [159] J. Newman, O. Dolezal, V. Fazio, T. Caradoc-Davies, and T. S. Peat. The DINGO dataset: A comprehensive set of data for the SAMPL challenge. *J Comput Aided Mol Des*, 26(5):497–503, Dec. 2011.
- [160] C. N. Nguyen, T. K. Young, and M. K. Gilson. Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *The Journal of Chemical Physics*, 137(4):044101, July 2012.
- [161] N. V. Nucci, B. Fuglestad, E. A. Athanasoula, and A. J. Wand. Role of cavities and hydration in the pressure unfolding of T4 lysozyme. *PNAS*, 111(38):13846–13851, Sept. 2014.
- [162] R. K. Pal, K. Haider, D. Kaur, W. Flynn, J. Xia, R. M. Levy, T. Taran, L. Wickstrom, T. Kurtzman, and E. Gallicchio. A combined treatment of hydration and dynamical effects for the modeling of host-guest binding thermodynamics: The SAMPL5 blinded challenge. *Journal of Computer-Aided Molecular Design*, 2016.
- [163] J. Park, I. Nessler, B. McClain, D. Macikenas, J. Baltrušaitis, and M. J. Schnieders. Absolute Organic Crystal Thermodynamics: Growth of the Asymmetric Unit into a Crystal via Alchemy. *J. Chem. Theory Comput.*, 10(7):2781–2791, July 2014.
- [164] T. S. Peat, O. Dolezal, J. Newman, D. Mobley, and J. J. Deadman. Interrogating HIV integrase for compounds that bind– a SAMPL challenge. *J Comput Aided Mol Des*, 28(4):347–362, Feb. 2014.
- [165] S. Picaud, C. Wells, I. Felletar, D. Brotherton, S. Martin, P. Savitsky, B. Diez-Dacal, M. Philpott, C. Bountra, H. Lingard, O. Fedorov, S. Muller, P. E. Brennan, S. Knapp, and P. Filippakopoulos. RVX-208, an inhibitor of BET transcriptional regulators with selectivity for the second bromodomain. *PNAS*, 110(49):19754–19759, Dec. 2013.
- [166] N. Plattner and F. Noé. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nature Communications*, 6:7653, July 2015.
- [167] S. Putta, G. A. Landrum, and J. E. Penzotti. Conformation Mining: An Algorithm for Finding Biologically Relevant Conformations. *J. Med. Chem.*, 48(9):3313–3318, May 2005.
- [168] Q. I. Qu, E. Tucker, and S. D. Christian. Sulfoalkyl Ether  $\beta$ -Cyclodextrin Derivatives: Synthesis and Characterizations. *Journal of Inclusion Phenomena*, 43(3–4):213–222, Aug. 2002.
- [169] U. Raj, H. Kumar, and P. K. Varadwaj. Molecular docking and dynamics simulation study of flavonoids as BET bromodomain inhibitors. *J. Biomol. Struct. Dyn.*, pages 1–12, Aug. 2016.
- [170] T. Ran, Z. Zhang, K. Liu, Y. Lu, H. Li, J. Xu, X. Xiong, Y. Zhang, A. Xu, S. Lu, H. Liu, T. Lu, and Y. Chen. Insight into the key interactions of bromodomain inhibitors based on molecular docking, interaction fingerprinting, molecular dynamics and binding free energy calculation. *Mol. Biosyst.*, 11(5):1295–1304, May 2015.
- [171] X. Ran, Y. Zhao, L. Liu, L. Bai, C.-Y. Yang, B. Zhou, J. L. Meagher, K. Chinnaswamy, J. A. Stuckey, and S. Wang. Structure-Based Design of  $\gamma$ -Carboline Analogues as Potent and Specific BET Bromodomain Inhibitors. *J. Med. Chem.*, 58(12):4927–4939, June 2015.
- [172] M. M. Reif and C. Oostenbrink. Net charge changes in the calculation of relative ligand-binding free energies via classical atomistic molecular dynamics simulation. *J. Comput. Chem.*, 35(3):227–243, Jan. 2014.
- [173] M. V. Rekharsky, R. N. Goldberg, F. P. Schwarz, Y. B. Tewari, P. D. Ross, Y. Yamashoji, and Y. Inoue. Thermodynamic and Nuclear Magnetic Resonance Study of the Interactions of - and -Cyclodextrin with Model Substances: Phenethylamine, Ephedrines, and Related Substances. *Journal of the American Chemical Society*, 117(34):8830–8840, Aug. 1995.
- [174] M. V. Rekharsky and Y. Inoue. Complexation Thermodynamics of Cyclodextrins. *Chem. Rev.*, 98(5):1875–1918, July 1998.
- [175] M. V. Rekharsky, M. P. Mayhew, R. N. Goldberg, P. D. Ross, Y. Yamashoji, and Y. Inoue. Thermodynamic and Nuclear Magnetic Resonance Study of the Reactions of - and -Cyclodextrin with Acids, Aliphatic Amines, and Cyclic Alcohols. *The Journal of Physical Chemistry B*, 101(1):87–100, Jan. 1997.
- [176] M. V. Rekharsky, T. Mori, C. Yang, Y. H. Ko, N. Selvapalam, H. Kim, D. Sobransingh, A. E. Kaifer, S. Liu, L. Isaacs, W. Chen, S. Moghaddam, M. K. Gilson, K. Kim, and Y. Inoue. A synthetic host-guest system achieves avidin-biotin affinity by overcoming enthalpy–entropy compensation. *PNAS*, 104(52):20737–20742, Dec. 2007.
- [177] G. J. Rocklin, S. E. Boyce, M. Fischer, I. Fish, D. L. Mobley, B. K. Shoichet, and K. A. Dill. Blind Prediction of Charged Ligand Binding Affinities in a Model Binding Site. *J. Mol. Biol.*, 425(22):4569–4583, Nov. 2013.
- [178] G. J. Rocklin, D. L. Mobley, K. A. Dill, and P. H. Hünenerger. Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An accurate correction scheme for electrostatic finite-size effects. *J. Chem. Phys.*, 139(18):184103, Jan. 2013.
- [179] A. I. Rodriguez-Perez, C. Rodriguez-Tenreiro, C. Alvarez-Lorenzo, P. Taboada, A. Concheiro, and J. J. Torres-Labandeira. Sertaconazole/hydroxypropyl- $\beta$ -cyclodextrin complexation: Isothermal titration

- calorimetry and solubility approaches. *J. Pharm. Sci.*, 95(8):1751–1762, Aug. 2006.
- [180] K. E. Rogers, J. M. Ortiz-Sánchez, R. Baron, M. Fajer, C. A. F. de Oliveira, and J. A. McCammon. On the Role of Dewetting Transitions in Host–Guest Binding Free Energy Calculations. *Journal of Chemical Theory and Computation*, 9(1):46–53, Jan. 2013.
- [181] R. J. Rosenfeld, A.-M. A. Hays, R. A. Musah, and D. B. Goodin. Excision of a proposed electron transfer pathway in cytochrome c peroxidase and its replacement by a ligand-binding channel. *Protein Science*, 11(5):1251–1259, May 2002.
- [182] M. J. Schnieders, J. Baltrusaitis, Y. Shi, G. Chattree, L. Zheng, W. Yang, and P. Ren. The Structure, Thermodynamics, and Solubility of Organic Crystals from Simulation with a Polarizable Force Field. *J. Chem. Theory Comput.*, 8(5):1721–1736, May 2012.
- [183] P. Schreiner. Theoretical prediction of affinities to cucurbiturils – the blind prediction hydrophobe challenge. <https://www.uni-giessen.de/fbz/fb08/dispersion/projects/HydrophobeChallenge>, 2016.
- [184] B. Sellner, G. Zifferer, A. Kornherr, D. Krois, and U. H. Brinker. Molecular Dynamics Simulations of  $\beta$ -Cyclodextrin-Aziadamantane Complexes in Water. *J. Phys. Chem. B*, 112(3):710–714, Jan. 2008.
- [185] B. Sherborne, V. Shanmugasundaram, A. C. Cheng, C. D. Christ, R. L. DesJarlais, J. S. Duca, R. A. Lewis, D. A. Loughney, E. S. Manas, G. B. McGaughey, C. E. Peishoff, and H. van Vlijmen. Collaborating to improve the use of free-energy and other quantitative methods in drug discovery. *J Comput Aided Mol Des*, 30(12):1139–1141, Dec. 2016.
- [186] M. Shi, C. Zhang, Y. Xie, and D. Xu. Stereoselective inclusion mechanism of ketoprofen into  $\beta$ -cyclodextrin: Insights from molecular dynamics simulations and free energy calculations. *Theor Chem Acc*, 133(10):1556, Oct. 2014.
- [187] M. R. Shirts. *Calculating Precise and Accurate Free Energies in Biomolecular Systems*. Stanford University, Stanford, CA, 2004.
- [188] M. R. Shirts and J. D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics*, 129(12):124105, Sept. 2008.
- [189] M. R. Shirts, C. Klein, J. M. Swails, J. Yin, M. K. Gilson, D. L. Mobley, D. A. Case, and M. R. Shirts. Lessons learned from comparing molecular dynamics engines on the SAMPL5 dataset. *J Comput Aided Mol Des*, 2016.
- [190] M. R. Shirts and D. L. Mobley. An Introduction to Best Practices in Free Energy Calculations. In *Biomolecular Simulations*, volume 924. Methods in Molecular Biology, Jan. 2013.
- [191] M. R. Shirts, D. L. Mobley, and S. P. Brown. Free-energy calculations in structure-based drug design. In J. Merz, Kenneth M, D. Ringe, and C. H. Reynolds, editors, *Drug Design: Structure and Ligand-Based Approaches*. Cambridge University Press, Jan. 2010.
- [192] H.-J. Shu, C.-M. Zeng, C. Wang, D. F. Covey, C. F. Zorumski, and S. Mennerick. Cyclodextrins sequester neuroactive steroids and differentiate mechanisms that rate limit steroid actions. *British Journal of Pharmacology*, 150(2):164–175, Jan. 2007.
- [193] T. Simonson and B. Roux. Concepts and protocols for electrostatic free energies. *Molecular Simulation*, 42(13):1090–1101, Sept. 2016.
- [194] A. G. Skillman. SAMPL3: Blinded prediction of host–guest binding affinities, hydration free energies, and trypsin inhibitors. *J Comput Aided Mol Des*, 26(5):473–474, May 2012.
- [195] S. H. Sleigh, P. R. Seavers, A. J. Wilkinson, J. E. Ladbury, and J. R. H. Tame. Crystallographic and Calorimetric Analysis of Peptide Binding to OppA Protein. *Journal of Molecular Biology*, 291(2):393–415, Aug. 1999.
- [196] P. W. Smith, S. L. Sollis, P. D. Howes, P. C. Cherry, I. D. Starkey, K. N. Cobley, H. Weston, J. Scicinski, A. Merritt, A. Whittington, P. Wyatt, N. Taylor, D. Green, R. Bethell, S. Madar, R. J. Fenton, P. J. Morley, T. Pateman, and A. Beresford. Dihydropyran-carboxamides Related to Zanamivir: A New Series of Inhibitors of Influenza Virus Sialidases. 1. Discovery, Synthesis, Biological Activity, and Structure-Activity Relationships of 4-Guanidino- and 4-Amino-4H-pyran-6-carboxamides. *J. Med. Chem.*, 41(6):787–797, Mar. 1998.
- [197] P. Sokkalingam, J. Shrberg, S. W. Rick, and B. C. Gibb. Binding Hydrated Anions with Hydrophobic Pockets. *Journal of the American Chemical Society*, 138(1):48–51, Jan. 2016.
- [198] D. Spiliotopoulos and A. Caflisch. Molecular dynamics simulations of bromodomains reveal binding-site flexibility and multiple binding modes of the natural ligand acetyl-lysine. *Isr. J. Chem.*, 54(9):1084–1092, Aug. 2014.
- [199] P. Štefanič Anderluh, M. Anderluh, J. Iláš, J. Mravljak, M. Sollner Dolenc, M. Stegnar, and D. Kikelj. Toward a Novel Class of Antithrombotic Compounds with Dual Function. Discovery of 1,4-Benzoxazin-3(4H)-one Derivatives Possessing Thrombin Inhibitory and Fibrinogen Receptor Antagonistic Activities. *J. Med. Chem.*, 48(9):3110–3113, May 2005.
- [200] S. Steiner, A. Magno, D. Huang, and A. Caflisch. Does bromodomain flexibility influence histone recognition? *FEBS Lett.*, 587(14):2158–2163, July 2013.
- [201] J. Su, X. Liu, S. Zhang, F. Yan, Q. Zhang, and J. Chen. A computational insight into binding modes of inhibitors XD29, XD35, and XD28 to bromodomain-containing protein 4 based on molecular dynamics simulations. *J. Biomol. Struct. Dyn.*, pages 1–13, May 2017.
- [202] M. R. Sullivan, P. Sokkalingam, T. Nguyen, J. P. Donahue, and B. C. Gibb. Binding of carboxylate and trimethylammonium salts to octa-acid and TEMOA deep-cavity cavitands. *J Comput Aided Mol Des*, pages 1–8, July 2016.
- [203] H. Sun, C. L. D. Gibb, and B. C. Gibb. Calorimetric Analysis of the 1:1 Complexes Formed between a Water-soluble Deep-cavity Cavitand, and Cyclic and Acyclic Carboxylic Acids. *Supramolecular Chemistry*, 20(1-2):141–147, Jan. 2008.
- [204] L. Szente and J. Szejtli. Highly soluble cyclodextrin derivatives: Chemistry, properties, and trends in development. *Advanced Drug Delivery Reviews*, 36(1):17–28, Mar. 1999.
- [205] K. Tai. Conformational sampling for the impatient. *Bioophysical Chemistry*, 107(3):213–220, Feb. 2004.
- [206] R. Talhout, A. Villa, A. E. Mark, and J. B. F. N. Engberts. Understanding Binding Affinity: A Com-

- bined Isothermal Titration Calorimetry/Molecular Dynamics Study of the Binding of a Series of Hydrophobically Modified Benzamidinium Chloride Inhibitors to Trypsin. *J. Am. Chem. Soc.*, 125(35):10570–10579, Sept. 2003.
- [207] J. R. Tame, E. J. Dodson, G. Murshudov, C. F. Higgins, and A. J. Wilkinson. The crystal structures of the oligopeptide-binding protein OppA complexed with tripeptide and tetrapeptide ligands. *Structure*, 3(12):1395–1406, Dec. 1995.
- [208] N. Taulier and T. V. Chalikian. Hydrophobic Hydration in Cyclodextrin Complexation. *The Journal of Physical Chemistry B*, 110(25):12222–12224, June 2006.
- [209] B. L. Tembe and J. A. McCammon. Ligand Receptor Interactions. *Comput. Chem.*, 8(4):281–283, Jan. 1984.
- [210] F. Tofoleanu, J. Lee, F. C. Pickard IV., G. König, J. Huang, M. Baek, C. Seok, and B. R. Brooks. Absolute binding free energy calculations for octa-acids and guests. *J Comput. Aided Mol. Des.*, 2016.
- [211] H. Ueno, K. Yokota, J.-i. Hoshi, K. Yasue, M. Hayashi, I. Uchida, K. Aisaka, Y. Hase, S. Katoh, and H. Cho. Discovery of novel tetrahydroisoquinoline derivatives as potent and selective factor Xa inhibitors. *Bioorganic & Medicinal Chemistry Letters*, 15(1):185–189, Jan. 2005.
- [212] C. Velez-Vega and M. K. Gilson. Force and Stress along Simulated Dissociation Pathways of Cucurbituril–Guest Systems. *J. Chem. Theory Comput.*, 8(3):966–976, Mar. 2012.
- [213] C. Velez-Vega and M. K. Gilson. Overcoming dissipation in the calculation of standard binding free energies by ligand extraction. *J. Comput. Chem.*, 34(27):2360–2371, Oct. 2013.
- [214] A. Verras. Free Energy Perturbation at Merck: Benchmarking against Faster Methods. [http://www.alchemy.org/wiki/images/c/c3/Vertex\\_FreeEnergyWorkshop2016\\_AV.pdf](http://www.alchemy.org/wiki/images/c/c3/Vertex_FreeEnergyWorkshop2016_AV.pdf), May 2016.
- [215] L. R. Vidler, P. Filippakopoulos, O. Fedorov, S. Picaud, S. Martin, M. Tomsett, H. Woodward, N. Brown, S. Knapp, and S. Hoelder. Discovery of novel small-molecule inhibitors of BRD4 using structure-based virtual screening. *J. Med. Chem.*, 56(20):8073–8088, Oct. 2013.
- [216] A. Villa, R. Zangi, G. Pieffet, and A. E. Mark. Sampling and convergence in free energy calculations of protein–ligand interactions: The binding of triphenoxypyridine derivatives to factor Xa and trypsin. *J. Comput. Aided Mol. Des.*, 17(10):673–686, 2003.
- [217] B. Vinciguerra, P. Y. Zavalij, and L. Isaacs. Synthesis and Recognition Properties of Cucurbit[8]uril Derivatives. *Org. Lett.*, 17(20):5068–5071, Oct. 2015.
- [218] A. J. Wand and N. V. Nucci. Reply to Kitahara and Mulder: An ensemble view of protein stability best explains pressure effects in a T4 lysozyme cavity mutant. *PNAS*, 112(9):E924–E924, Mar. 2015.
- [219] J. Wang, Y. Deng, and B. Roux. Absolute Binding Free Energy Calculations Using Molecular Dynamics Simulations with Restraining Potentials. *Biophysical Journal*, 91(8):2798–2814, Oct. 2006.
- [220] K. Wang, J. D. Chodera, Y. Yang, and M. R. Shirts. Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. *J. Comput. Aided Mol. Des.*, 27(12):989–1007, Dec. 2013.
- [221] K. Wang, P. Sokkalingam, and B. C. Gibb. ITC and NMR analysis of the encapsulation of fatty acids within a water-soluble cavitand and its dimeric capsule. *Supramolecular Chemistry*, 28(1-2):84–90, Feb. 2016.
- [222] L. Wang, B. J. Berne, and R. A. Friesner. On achieving high accuracy and reliability in the calculation of relative protein–ligand binding affinities. *PNAS*, 109(6):1937–1942, July 2012.
- [223] L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyán, S. Robinson, M. K. Dahlgren, J. Greenwood, D. L. Romero, C. Masse, J. L. Knight, T. Steinbrecher, T. Beuming, W. Damm, E. Harder, W. Sherman, M. Brewer, R. Wester, M. Murcko, L. Frye, R. Farid, T. Lin, D. L. Mobley, W. L. Jorgensen, B. J. Berne, R. A. Friesner, and R. Abel. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.*, 137(7):2695–2703, Feb. 2015.
- [224] B. Q. Wei, W. A. Baase, L. H. Weaver, B. W. Matthews, and B. K. Shoichet. A Model Binding Site for Testing Scoring Functions in Molecular Docking. *Journal of Molecular Biology*, 322(2):339–355, Sept. 2002.
- [225] B. Q. Wei, L. H. Weaver, A. M. Ferrari, B. W. Matthews, and B. K. Shoichet. Testing a Flexible-receptor Docking Algorithm in a Model Binding Site. *Journal of Molecular Biology*, 337(5):1161–1182, Apr. 2004.
- [226] L. Wickstrom, N. Deng, P. He, A. Mentes, C. Nguyen, M. K. Gilson, T. Kurtzman, E. Gallicchio, and R. M. Levy. Parameterization of an effective potential for protein–ligand binding from host–guest affinity data. *J. Mol. Recognit.*, 29(1):10–21, Jan. 2016.
- [227] L. Wickstrom, P. He, E. Gallicchio, and R. M. Levy. Large Scale Affinity Calculations of Cyclodextrin Host–Guest Complexes: Understanding the Role of Reorganization in the Molecular Recognition Process. *J. Chem. Theory Comput.*, 9(7):3136–3150, July 2013.
- [228] H.-J. Woo and B. Roux. Calculation of absolute protein–ligand binding free energy from computer simulations. *PNAS*, 102(19):6825–6830, Oct. 2005.
- [229] M. Wszelaka-Rylik and P. Gierycz. Isothermal titration calorimetry (ITC) study of natural cyclodextrins inclusion complexes with drugs. *J. Therm. Anal. Calorim.*, 111(3):2029–2035, Mar. 2013.
- [230] I. W. Wyman and D. H. Macartney. Cucurbit[7]uril host–guest complexes with small polar organic guests in aqueous solution. *Organic & Biomolecular Chemistry*, 6(10):1796, 2008.
- [231] X. Xiong, Z. Chen, B. P. Cossins, Z. Xu, Q. Shao, K. Ding, W. Zhu, and J. Shi. Force fields and scoring functions for carbohydrate simulation. *Carbohydrate Research*, 401:73–81, Jan. 2015.
- [232] M. Xu, A. Unzue, J. Dong, D. Spiliotopoulos, C. Nevado, and A. Caflisch. Discovery of CREBBP bromodomain inhibitors by high-throughput docking and hit optimization guided by molecular dynamics. *J. Med. Chem.*, 59(4):1340–1349, Feb. 2016.
- [233] X. Xue, Y. Zhang, Z. Liu, M. Song, , Y. Xing, Q. Xiang, Z. Wang, Z. Tu, Y. Zhou, K. Ding, and Y. Xu. Discovery of benzo[cd]indol-2(1H)ones as potent and specific BET bromodomain inhibitors: Structure-based virtual screening, optimization, and biological evaluation. *J. Med. Chem.*, 59(4):1565–1579, Jan. 2016.
- [234] Y. Yang, L. Zhao, B. Xu, L. Yang, J. Zhang, H. Zhang,

- and J. Zhou. Design, synthesis and biological evaluation of dihydroquinoxalinone derivatives as BRD4 inhibitors. *Bioorganic Chemistry*, 68:236–244, Oct. 2016.
- [235] J. Yin, A. T. Fenley, N. M. Henriksen, and M. K. Gilson. Toward Improved Force-Field Accuracy through Sensitivity Analysis of Host-Guest Binding Thermodynamics. *The Journal of Physical Chemistry B*, 119(32):10145–10155, Aug. 2015.
- [236] J. Yin, N. M. Henriksen, D. R. Slochower, and M. K. Gilson. The SAMPL5 Host-Guest Challenge: Computing Binding Free Energies and Enthalpies from Explicit Solvent Simulations by the Attach-Pull-Release (APR) Method. *J Comput Aided Mol Des*, (31):133–145, 2017.
- [237] J. Yin, N. M. Henriksen, D. R. Slochower, M. R. Shirts, M. W. Chiu, D. L. Mobley, and M. K. Gilson. Overview of the SAMPL5 host–guest challenge: Are we doing better? *J Comput Aided Mol Des*, (31):1–19, 2017.
- [238] F. M. Ytreberg. Absolute FKBP binding affinities obtained via nonequilibrium unbinding simulations. *The Journal of Chemical Physics*, 130(16):164906, Apr. 2009.
- [239] G. Zhang, A. N. Plotnikov, E. Rusinova, T. Shen, K. Morohashi, J. Joshua, L. Zeng, S. Mujtaba, M. Ohlmeyer, and M.-M. Zhou. Structure-Guided Design of Potent Diazobenzene Inhibitors for the BET Bromodomains. *J. Med. Chem.*, 56(22):9251–9264, Nov. 2013.
- [240] H. Zhang, T. Tan, and D. van der Spoel. Generalized Born and Explicit Solvent Models for Free Energy Calculations in Organic Solvents: Cyclodextrin Dimerization. *J. Chem. Theory Comput.*, 11(11):5103–5113, Nov. 2015.
- [241] H. Zhang, C. Yin, H. Yan, and D. van der Spoel. Evaluation of Generalized Born Models for Large Scale Affinity Prediction of Cyclodextrin HostGuest Complexes. *Journal of Chemical Information and Modeling*, Sept. 2016.
- [242] H. Zhao, L. Gartenmann, J. Dong, D. Spiliotopoulos, and A. Caflisch. Discovery of BRD4 bromodomain inhibitors by fragment-based high-throughput docking. *Bioorg. Med. Chem. Lett*, 24(11):2493–2496, June 2014.