Published in final edited form as:

Annu Rev Biophys. 2017 May 22; 46: 531–558. doi:10.1146/annurev-biophys-070816-033654.

Predicting binding free energies: Frontiers and benchmarks

David L. Mobley[1] and **Michael K. Gilson**[2]

[1]Departments of Pharmaceutical Sciences and Chemistry, University of California, Irvine, CA, USA, 92697; dmobley@mobleylab.org

[2]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, CA, USA, 92092; mgilson@ucsd.edu

Abstract

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 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.

Keywords

binding free energy; molecular simulation; alchemical; benchmark; biomolecular interactions; binding affinity

1. 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 movie of 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

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.

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.

1.1. 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 (83, 19, 27, 109, 128, 18, 25, 123), solubility predictions beginning to come online (107, 99, 70), and predicted drug resistance/selectivity also apparently tractable (67, 67), with some headway apparent on membrane permeability (62, 23). A considerable amount of science and engineering still remains to make this vision a reality, but, given recent progress, the question now seems more one of *when* rather than *whether*.

1.2. Increasing accuracy will yield increasing payoffs

Recent progress in computational power, especially the widespread availability of graphics processing units (GPUs) and advances in automation (72) 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 (78, 53, 109, 128, 18, 25, 123). 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 ~ 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 (113, 83). Thus, one can gain substantial benefit from simulations that are good yet still quite imperfect.

More broadly, 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.

1.3. 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 (59), with the system effectively coupled to a heat bath at constant temperature, but Monte Carlo methods may also be used (75, 76, 21). 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 (76, 20, 17, 112) 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 free energy of binding of two molecules by computing 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 (133, 138, 122, 50, 54, 8). In addition, artificial restraints may be useful to avoid sampling problems in the face of often complex barriers along the pathway (133, 122, 50, 54, 8). By contrast, alchemical pathway methods artificially decouple the ligand from the binding site and then recouple it to solution from the protein (58, 51, 45, 9, 80). Although alchemical decoupling methods may avoid clashes of the ligand with the protein that might be problematic in pathway methods for a tight 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 (65, 49, 26, 136).

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 artificially converting one ligand into another, first in the bound state and then free in solution (119, 76, 20, 17). 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 (72). 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 (83, 72, 128).

Importantly, the accuracy and precision of all of these methods are controlled by the same 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 highly sensitive to initial conditions (2), repeated calculations, using different random number seeds or initial states, will yield 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.

1.4. 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 or 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 — 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 (103). 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 (81, 12, 82, 56, 37, 127); and larger-scale protein motions induced by some ligands also posing challenges (12, 68).

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 (30) 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.

1.5. 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 (?), 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 (6, 136?)) 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, detailed below.

2. THE NEED FOR WELL-CHOSEN BENCHMARK SYSTEMS

Although tests of individual free energy methods are not uncommon today (78, 128, 18, 25, 123), 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 3). These selections result in part from two series of blinded prediction challenges (SAMPL (91), and CSAR (29) followed by D3R (40)), 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.

2.1. 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.

2.1.1. Hard benchmarks

2.1.1.1. 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 (111), 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 the 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 (111).

Similarly, 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 ≈ 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 (110, 35), For this purpose, the correctly computed values need not agree with experiment; indeed, experimental results are unnecessary.

2.1.1.2. 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 (118, 112), 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 barriers 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, and the point of comparison may be, at least in part, sampling measures.

2.1.1.3. 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 (57); 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 (50), 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 or new experimental conditions (such as temperature and solvent composition) explored.

2.1.2. Soft benchmarks

2.1.2.1. Systems to challenge conformational sampling techniques: Enhanced sampling techniques (Section 2.1.1.2), 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.

2.1.2.2. 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 in real-world challenges. Appropriate soft benchmarks are likely to be cases which are still relatively tractable, involving small proteins and simple binding sites. We need a series of benchmark protein-ligand systems that introduce various challenges in a well-understood manner. Systems should introduce 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 receptor 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
    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

2.1.2.3. 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.

2.2. Applications and limitations of benchmark systems

Standard benchmark systems along the lines sketched above will allow potential solutions 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 with experiment—with the understanding that force field and sampling also play a role. Such studies will be greatly facilitated by well-characterized standard benchmarks.

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.

3. BENCHMARK SYSTEMS FOR BINDING PREDICTION

No molecular systems have been explicitly accepted by the field as benchmarks for free energy calculations, but certain host molecules (see below) and designed binding sites in the enzyme T4 lysozyme have emerged as particularly helpful and widely studied test cases. Here, we describe these artificial receptors and propose specific host-guest and T4 lysozyme-ligand 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 4). The present focus is on cases where experimental data are available and add value, rather than ones chosen specifically to test conformational sampling methods, where experimental data are not required (Section 2.1).

3.1. 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 de- fined above (Section 2.1). 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 2.1.1.3). 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 (50), so they provide an additional check of the validity of simulations. A variety of curated 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 nonaqueous environment within a lipid membrane. Here, we focus on two host families, the cucurbiturils (36, 85); and the octa-acids (more generally, Gibb deep cavity cavitands) (41, 52), which have already been the subject of concerted attention from the simulation community, due in part to their use in the SAMPL blinded prediction challenges (93, 91, 136).

3.1.1. Cucurbiturils—The cucurbiturils (Figure 1) are achiral rings of glycoluril monomers (36). 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) (71), which have been characterized to different extents. Of note, 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 (134, 63)) that forms stabilizing interactions with the oxygens of the carbonyl groups fringing both portals of the host (71). Although derivatives of these parent compounds have been made (64, 124, 3, 24), 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, Tables 1 and 2). 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 (71, 102, 86, 15). Finally, CB7 is already familiar to a number of computational chemistry groups, as it figured in two of the three SAMPL challenges that included host-guest components (93, 91), and it is currently the focus of the “hydrophobe challenge” (108).

3.1.1.1. 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 (86, 87) groups, do not fit easily through the constrictive portals (121). 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 (122, 50). 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 (50), 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 (45), 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 (105).

3. **Salt concentration and buffer conditions**: Binding thermodynamics are sensitive to the composition of dissolved salts, both experimentally (87, 86, 91) and computationally (94, 54). 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 (104, 69, 100, 114).

3.1.1.2. 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 (71, 15). Each series is based on a common chemical scaffold, making it amenable to not only absolute but also alchemical relative free energy calculations (Section 1.3). One set is based on an adamantane core (Table 1), and the other on an aromatic ring (Table 2). These systems can be run to convergence to allow detailed comparisons among methods and with experiment. Their binding free energies range from -5.99 to -17.19 kcal/mol, with the adamantane series spanning a particularly large range of free energies.

3.1.1.3. 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 (50, 88). Due to its top-bottom symmetry, flips of guests from "head-in" to "head-out" configurations are not necessary to obtain convergence (33). However, sampling of the guest geometry can be a challenge, with transitions between binding modes as slow as 0.07 flips/ns (88), and flexible guests also presenting challenges (88). As noted above, water sampling can also be an issue, with wetting/dewetting transitions occurring on the 50 ns timescale (105).

Salt and buffer conditions are also key. In addition to the strong salt-dependence of binding (87), acetic acid (such as in a sodium acetate buffer) can compete with guests for the binding site (86). This may partially explain systematic errors in some computational studies (94, 54). 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 (94, 91). Cationic guests could also have substantial and differing interactions with the counterions in solution as well, potentially lowering affinity relative to zero-salt conditions (91). Thus, one

group found a 6.4–6.8 kcal/mol dependence on salt concentration (54), possibly impacting other studies as well (88)

Despite these issues, CB7 appears to be at the point where careful studies can probe the true accuracy of our force fields (50, 39, 135), and the results can be sobering, with RMS errors in the binding free energies as high as 8 kcal/mol (50, 88). More encouragingly, the values of R^2 values can be as high as 0.92 (50). Some force fields appear relatively worse than others (54, 92). Calculated values are in many cases quite sensitive to details of force field parameters (88, 87, 92). For example, modest modification of some Lennard-Jones parameters yielded dramatic improvements in calculated values (135), and host-guest binding data has, accordingly, been suggested as an input for force field development (135, 50, 39). Water structure around CB7 and calculated binding enthalpies also appear particularly sensitive to the choice of water model (105, 33, 39), and water is clearly important for modulating binding (97). The water model also impacts the number of sodium ions which must be displaced (in sodium-based buffer) on binding (39, 50).

Despite its apparent simplicity, 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. 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 (91). Until the origin of such discrepancies is clear, it is difficult to know how accurate our methods truly are.

3.1.2. Gibb Deep Cavity Cavitands (GDCC)—The octa-acids (OA) (Figure 1) are synthetic hosts with deep, basket-shaped, hydrophobic binding sites (41). 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) (41) and for a derivative with four added methyl groups at equivalent locations in the entryway, where they can contact a bound guest (TEMOA) (38, 116). (Note that OA and TEMOA have also been called OAH and OAMe, respectively (136).) Additional family members with other substituents around the portal have been reported, as has a new series in which the eponymic carboxylic groups are replaced by various other groups, including a number of basic amines (52). However, we are not aware of binding data for these derivatives. In view of these other hosts, however, 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(77); and, in some studies, the benzoic acid “aps” around the entry occasionally ip upward and into contact with the guest (137, 120), though this motion has not been verified experimentally. Additionally, the four priopionate groups protruding into solution 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 (42, 44). Compounds with adamantane or noradamantane groups display perhaps the highest affinities observed so far, with binding free energies ranging to about -8 kcal/mol (117). Much 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, chlorate and isothiocyanate can shift binding enthalpies by up to about 10 kcal/mol and free energies by around 2 kcal/mol (43). 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 (115), 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 (42).

3.1.2.1. The proposed GDCC benchmark sets are drawn from SAMPL: As a core benchmark series for this family, we propose two sets which formed part of the SAMPL4 and SAMPL5 challenges, based on adamantane derivatives (Table 3) and cyclic (aromatic and saturated) carboxylic acids (Table 4) binding to hosts OA and TEMOA with free energies of −3.7 to −7.6 kcal/mol. These cases offer aqueous binding data with a reasonably broad range of binding free energies, frequently along with binding enthalpies; the hosts and many or all of their guests are small and rigid enough to allow convincing convergence 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 (91, 136).

3.1.2.2. OA introduces 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 methyls groups can additionally hinder sampling of guest poses in the bound state, leading to convergence problems (136) specific to TEMOA.
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 (77). Similarly, benzoic acid flips (137, 120) 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 (32). 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 (8), 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 (43, 115) and computationally (98, 120). 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. (104, 69, 100, 114).

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 (91, 32, 120), particularly given experimental evidence for extreme binding-driven pKa shifts of 3-4 log units for some carboxylate compounds (126, 115). Thus, attention should be given to ionization states and their modulation by binding.

3.1.2.3. 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 proprionate groups occurring on 1–2 ns timescales (77)); motions of the benzoic acid aps were also relatively slow (137, 120) though perhaps thermodynamically unimportant. Guest sampling can also be an issue, at least in TEMOA (136), and this hosts's tight cavity may also have implications for binding entropy (137).

Salt concentration strongly modulates binding affinity, at least for anions, and the nature of the salt also plays an important role (16). Co-solvent anions can also increase *or* decrease binding depending on their identity (43). Some salts even bind to OA themselves, with perchlorate (43) and trichloroacetate (115) 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 (120).

Naively, protonation states of the guests might seem clear and unambiguous. But since OA can bind guests of diverse net charges, the protonation state may not always be clear. One study used absolute binding free energy calculations for different guest charge states, coupled with pKa calculations, and found that inclusion of pKa corrections and the possibility of alternate charge states of the guests affected calculated binding free energies by up to 2 kcal/mol (120). As noted above, experimental evidence also indicates major pKa shifts on binding so that species such as acetate, formate and others would bind in neutral form at neutral pH (126, 115). Even the host protonation state may be unclear; while OA is often 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 (32). Thus, there are uncertainties as to the host

protonation state (91, 32), which perhaps also could be modulated by guest binding. Several groups used different methods but the same force field and water model in SAMPL5, with rather varied levels of success because of discrepancies in calculated free energies (136, 10, 8). However, some of these issues were resolved in follow-up work (8), bringing the methods into fairly good agreement for the majority of cases (137, 10).

3.2. 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 (89, 90) and L99A/M102Q (130, 47) for point mutations which create the cavities of interest, are created in artificial mutants of phage T4 lysozyme, and 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 are absent at room temperature and ambient pressure, allowing calculated binding free energies to apparently converge relatively easily. However, like host-guest systems, these binding sites are still surprisingly challenging (80, 81, 82, 12, 56, 37, 68). In addition, 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 2.1). The importance of the lysozyme model sites is also driven by the relative wealth of experimental data. It is relatively easy to identify new ligands and obtain high quality crystal structures and affinity measurements, allowing two different rounds of blind predictions testing free energy calculations (82, 12).

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 (96, 66, 22), yet the protein can undergo pressure-induced filling (22, 66) or denaturation (96), which can be inhibited by binding of ligand (96, 66). Pressure may also cause the protein to populate an excited state (73, 61) (but see (129)) which is already present to a very limited extent at equilibrium (11). 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 (96, 66, 73), though as we discuss below, some ligands do induce a protein conformational change which affects the same helix as the proposed excited state (74). It seems likely that the conformational hetereogeneity 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 the pressure.

3.2.1. The apolar and polar cavities and their ligands—The L99A site is also called the “apolar” cavity. It is relatively at and elongated, and binds mostly nonpolar 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 (89, 82), and about −4 to −5.5 kcal/mol in the polar site (12). 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 tests for free energy calculations.

For both sites, fixed charge force fields seem to yield reasonably accurate free energies, with RMS errors between 1–2 kcal/mol, and some level of correlation with experiment, despite limited dynamic range (28, 82, 12, 37, 125). System composition/preparation issues also do not seem to be a huge factor. Instead, sampling issues predominate:

1. Ligand binding mode/orientational sampling: The binding sites are buried and roughly oblong, with ligands which 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 (48, 12) but 3-chlorophenol has at least four, three of which appear to have some population in simulations (37), 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 rearrangements: Some sidechains are known to reorganize when binding certain ligands. The smallest ligands tend not to induce conformational changes, but larger ligands may induce sidechain rearrangements – often, rotamer flips – around the binding site region. These can be slow in the tightly packed binding site. This especially occurs for Val111 in the L99A site (90, 81, 56) and Leu118, Val11, and Val103 in L99A/M102Q (130, 131, 48, 12). These sidechain motions typically present sampling problems for standard MD simulations (81, 82, 12, 56, 127).
3. Backbone sampling: Larger ligands induce shifts of the F helix, residues 107 or 108 to 115, adjacent to the binding site, allowing the site to enlarge. This occurs in both binding sites (131, 12, 74), but is best characterized for L99A (74). 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; thisaffects the same region of helix F that undergoes a conformational change in the proposed excited state which is partially populated at equilibrium (11)

Tables 5 and 6 introduce proposed benchmark sets for the apolar and polar cavities, giving ligands potentially amenable to both absolute and relative free energy calculations, and spanning the range of available affinities. Co-crystal structures are available in most cases, and the PDB IDs are provided in the tables. The selected ligands span a range of challenges and levels of difficulty, ranging from fairly simple to including 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 non-binders are available, with binding affinities available for 19 compounds in the L99A site (31, 89, 82) and 16 in L99A/M102Q (130, 47, 12). 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.

3.2.2. Computational challenges posed by the T4 lysozyme benchmarks—Early work on the lysozyme sites focused on the difficulty of predicting binding modes (80, 82, 12) because of the slow interconversions noted above. Docking methods often can generate reasonable poses spanning most of the important possibilities (80, 82, 12, 48) but do not accurately predict the binding mode of individual compounds (82, 12, 48). Thus it appears necessary to consider the possibility of multiple binding modes; this is also important since some ligands actually populate multiple binding modes (12). 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 (80) and combined to find the population of each binding mode and determine the overall binding free energy. However, this is costly since each candidate binding mode requires a full binding free energy calculation.

Relative binding free energy calculations do not dramatically simplify the situation. Introduction of a ligand modification can leave the binding mode uncertain (e.g., introducing a chlorine onto phenol leaves at least two possible binding modes even if the binding mode of phenol is known) (12). A naïve solution is to consider multiple possible binding modes in relative free energy calculations (12), but this generates multiple results; determining the true relative binding free energy requires additional information (83). Enhanced sampling approaches provide one possible solution to the binding mode problem. Particularly, with λ or Hamiltonian exchange techniques, ligands can easily switch between binding modes when they are non-interacting unless they are restrained, and then moves in λ space can allow transitions back to the interacting state. Thus, approaches employing this strategy can naturally sample multiple binding modes (37, 125).

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 (81). However, this is not a general solution, since it requires knowing what sidechains might rearrange on binding and then expending substantial computational power on sampling free energy landscapes for these rearrangements. An apparently better general strategy is including sidechains in enhanced sampling regions selected for Hamiltonian exchange (56, 60) or REST (127), allowing sidechains to be alchemically softened or torsion barriers lowered (or both), to enhance sampling at alchemical intermediate states. With swaps between λ values, enhanced sidechain sampling at intermediate states can propagate to all states, improving convergence (56, 127).

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 (12). However, recent work noted above highlighted how a helix in the apolar cavity can open to accommodate larger ligands (74). 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 (68). Including part of the protein in the enhanced sampling region via REST2 provides some benefits, but sampling these motions will likely prove a valuable test for enhanced sampling methods.

4. THE FUTURE OF BENCHMARKS AND OF THIS REVIEW

This work has so far presented a small set of benchmark systems for binding free energy calculations, and has highlighted some of the ways in which they have already proven their utility. However, the scope of these sets is still quite limited. More, increasingly diverse, host-guest systems will help probe the strengths and weaknesses of force fields, and to drive their improvement. At the other end of the spectrum, we need more complex and challenging benchmark sets for proteins including simple models, like T4 lysozyme as well as candidate drug targets. And there may be community interest in test systems specifically selected to challenge sampling algorithms, without reference to experimental data.

Several candidate hosts and proteins are worth mentioning in this regard. Among host-guest systems, there is a particularly extensive experimental literature on cyclodextrins (46, 101), and they are tractable computationally (50, 132). As to artificial protein binding sites, the two variants of the CCP protein model binding site (34, 4, 5, 103, 95, 106) offer a modest increase in difficulty relative to the T4 lysozyme sites discussed above. And thrombin and the bromodomains appear to be promising examples of candidate drug targets for inclusion in a growing set of benchmark systems. Thrombin is a serine protease that has received prior attention from free energy studies (127, 128, 14). Experimental data exhibits interesting trends (7) that can partly be explained by simulations (14); but challenges remain (13). Bromodomains may also be interesting, especially given that relatively high accuracies have been reported, relative to experiment. At the same time, binding modes may be non-obvious and the diversity of ligands could pose problems for relative free energy calculations (1). Other systems will undoubtedly emerge as promising benchmarks as well, and we seek community input to help identify these.

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. We encourage use of the issue tracker for discussion, comments, and proposed updates. We plan to incorporate new material via GitHub as one would for a coding project, then make it available as preprints via bioRxiv. Given substantial changes to this initial version of the paper, it may ultimately be appropriate to make it available as a “perpetual review” (84) via another forum allowing versioned updates of publications.

5. 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. Benchmarks are vital for this, as they allow researchers in the field to rigorously test their 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.

Here, we 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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 grant R01GM061300. 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.

LITERATURE CITED

1. Aldeghi M, Heifetz A, Bodkin MJ, Knapp S, Biggin PC. Accurate calculation of the absolute free energy of binding for drug molecules. Chem. Sci. 2016; 7:207–218. [PubMed: 26798447]
2. Allen, MP., Tildesley, DJ. Computer simulation of liquids. Oxford Science Publications. New York, NY: Oxford University Press; 1989.
3. Assaf KI, Nau WM. Cucurbiturils: From synthesis to high-affinity binding and catalysis. Chem Soc Rev. 2015; 44:394–418. [PubMed: 25317670]

4. Banba S, Brooks CL III. Free energy screening of small ligands binding to an artificial protein cavity. The Journal of Chemical Physics. 2000; 113:3423–3433.
5. Banba S, Guo Z, Brooks CL III. Efficient Sampling of Ligand Orientations and Conformations in Free Energy Calculations Using the λ-Dynamics Method. J. Phys. Chem. B. 2000; 104:6903–6910.
6. Bannan C, Burley K, Chiu M, Shirts M, Gilson M, Mobley D. Blind prediction of cyclohexane-water distribution coefficients from the SAMPL5 challenge. J Comput Aided Mol Des. 2016
7. Baum B, Muley L, Smolinski M, Heine A, Hangauer D, Klebe G. Non-additivity of Functional Group Contributions in Protein–Ligand Binding: A Comprehensive Study by Crystallography and Isothermal Titration Calorimetry. Journal of Molecular Biology. 2010; 397:1042–1054. [PubMed: 20156458]
8. Bhakat S, Söderhjelm P. 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
9. Boresch S, Tettinger F, Leitgeb M, Karplus M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. The Journal of Physical Chemistry B. 2003; 107:9535–9551.
10. Bosisio S, Mey ASJS, Michel J. Blinded predictions of host-guest standard free energies of binding in the SAMPL5 challenge. J Comput Aided Mol Des. 2016
11. Bouvignies G, Vallurupalli P, Hansen D, Correia B, Lange O, Bah A, Vernon R, Dahlquist F, Baker D, Kay L. Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. Nature. 2011; 7362:111–114.
12. Boyce SE, Mobley DL, Rocklin GJ, Graves AP, Dill KA, Shoichet BK. Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. J. Mol. Biol. 2009; 394:747–763. [PubMed: 19782087]
13. Calabrò G. Accelerating molecular simulations implication for rational drug design. 2015
14. Calabrò G, Woods CJ, Powlesland F, Mey ASJS, Mulholland AJ, Michel J. Elucidation of Nonadditive Effects in Protein–Ligand Binding Energies: Thrombin as a Case Study. J. Phys. Chem. B. 2016
15. Cao L, Šekutor M, Zavalij PY, Mlinarić-Majerski K, Glaser R, Isaacs L. Cucurbit[7]uril·Guest Pair with an Attomolar Dissociation Constant. Angew. Chem. Int. Ed. 2014; 53:988–993.
16. Carnegie RS, Gibb CLD, Gibb BC. Anion Complexation and The Hofmeister Effect. Angew. Chem. 2014; 126:11682–11684.
17. Chodera JD, Mobley DL, Shirts MR, Dixon RW, Branson K, Pande VS. Alchemical free energy methods for drug discovery: Progress and challenges. Curr Opin Struct Biol. 2011; 21:150–160. [PubMed: 21349700]
18. Christ CD. Binding affinity prediction from molecular simulations: A new standard method in structure-based drug design? 2016
19. Christ CD, Fox T. Accuracy Assessment and Automation of Free Energy Calculations for Drug Design. J. Chem. Inf. Model. 2014; 54:108–120. [PubMed: 24256082]
20. Christ CD, Mark AE, van Gunsteren WF. Basic ingredients of free energy calculations: A review. J. Comput. Chem. 2010; 31:1569–1582. [PubMed: 20033914]
21. Cole DJ, Tirado-Rives J, Jorgensen WL. Molecular dynamics and Monte Carlo simulations for protein–ligand binding and inhibitor design. Biochimica et Biophysica Acta (BBA) - General Subjects. 2015; 1850:966–971. [PubMed: 25196360]
22. Collins M, Quillin M, Hummer G, Matthews B, Gruner S. Structural rigidity of a large cavity-containing protein revealed by high-pressure crystallography. J. Mol. Biol. 2007; 367:752–763. [PubMed: 17292912]
23. Comer J, Schulten K, Chipot C. Calculation of Lipid-Bilayer Permeabilities Using an Average Force. J Chem. Theory Comput. 2014; 10:554–564. [PubMed: 26580032]
24. Cong H, Ni XL, Xiao X, Huang Y, Zhu QJ, et al. Synthesis and separation of cucurbit[n]urils and their derivatives. Org. Biomol. Chem. 2016; 14:4335–4364. [PubMed: 26991738]
25. Cui G. Affinity Predictions with FEP+: A Different Perspective on Performance and Utility. 2016

26. de Ruiter A, Oostenbrink C. Protein–Ligand Binding from Distance field Distances and Hamiltonian Replica Exchange Simulations. J. Chem. Theory Comput. 2013; 9:883–892. [PubMed: 26588732]
27. Deng N, Forli S, He P, Perryman A, Wickstrom L, et al. Distinguishing Binders from False Positives by Free Energy Calculations: Fragment Screening Against the Flap Site of HIV Protease. J. Phys. Chem. B. 2015; 119:976–988. [PubMed: 25189630]
28. Deng Y, Roux B. Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant. Journal of Chemical Theory and Computation. 2006; 2:1255–1273. [PubMed: 26626834]
29. Dunbar JB, Smith RD, Yang CY, Ung PMU, Lexa KW, et al. CSAR Benchmark Exercise of 2010: Selection of the Protein-Ligand Complexes. J. Chem. Inf. Model. 2011; 51:2036–2046. [PubMed: 21728306]
30. Eklund A, Nichols TE, Knutsson H. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proc. Natl. Acad. Sci. U.S.A. 2016; 113:7900–7905. [PubMed: 27357684]
31. Eriksson AE, Baase WA, Wozniak JA, Matthews BW. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. Nature. 1992; 355:371–373. [PubMed: 1731252]
32. Ewell J, Gibb BC, Rick SW. Water Inside a Hydrophobic Cavitand Molecule. The Journal of Physical Chemistry B. 2008; 112:10272–10279. [PubMed: 18661937]
33. Fenley AT, Henriksen NM, Muddana HS, Gilson MK. Bridging Calorimetry and Simulation through Precise Calculations of Cucurbituril-Guest Binding Enthalpies. Journal of Chemical Theory and Computation. 2014; 10:4069–4078. [PubMed: 25221445]
34. Fitzgerald MM, Musah RA, McRee DE, Goodin DB. A ligand-gated, hinged loop rearrangement opens a channel to a buried artificial protein cavity. Nat Struct Mol Biol. 1996; 3:626–631.
35. Flyvbjerg H, Petersen HG. Error estimates on averages of correlated data. The Journal of Chemical Physics. 1989; 91:461.
36. Freeman WA, Mock WL, Shih NY. Cucurbituril. J. Am. Chem. Soc. 1981; 103:7367–7368.
37. Gallicchio E, Lapelosa M, Levy RM. Binding Energy Distribution Analysis Method (BEDAM) for Estimation of Protein-Ligand Binding Affinities. Journal of Chemical Theory and Computation. 2010; 6:2961–2977. [PubMed: 21116484]
38. Gan H, Benjamin CJ, Gibb BC. Nonmonotonic Assembly of a Deep-Cavity Cavitand. Journal of the American Chemical Society. 2011; 133:4770–4773. [PubMed: 21401093]
39. Gao K, Yin J, Henriksen NM, Fenley AT, Gilson MK. Binding Enthalpy Calculations for a Neutral Host-Guest Pair Yield Widely Divergent Salt Effects across Water Models. Journal of Chemical Theory and Computation. 2015; 11:4555–4564. [PubMed: 26574247]
40. Gathiaka S, Liu S, Chiu M, Yang H, Stuckey J, et al. D3R Grand Challenge 2015: Evaluation of Protein-Ligand Pose and Affinity Predictions. J. Comput. Aided Mol. Des. 2016 (In press).
41. Gibb CLD, Gibb BC. Well-Defined, Organic Nanoenvironments in Water: The Hydrophobic Effect Drives a Capsular Assembly. J. Am. Chem. Soc. 2004; 126:11408–11409. [PubMed: 15366865]
42. Gibb CLD, Gibb BC. Guests of differing polarities provide insight into structural requirements for templates of water-soluble nano-capsules. Tetrahedron. 2009; 65:7240–7248. [PubMed: 20606762]
43. Gibb CLD, Gibb BC. Anion Binding to Hydrophobic Concavity Is Central to the Salting-in Effects of Hofmeister Chaotropes. Journal of the American Chemical Society. 2011; 133:7344–7347. [PubMed: 21524086]
44. Gibb CLD, Gibb BC. Binding of cyclic carboxylates to octa-acid deep-cavity cavitand. J Comput Aided Mol Des. 2013; 28:319–325. [PubMed: 24218290]
45. Gilson MK, Given JA, Bush BL, McCammon JA. The statistical-thermodynamic basis for computation of binding affinities: A critical review. Biophys J. 1997; 72:1047–1069. [PubMed: 9138555]
46. Godínez LA, Schwartz L, Criss CM, Kaifer AE. Thermodynamic Studies on the Cyclodextrin Complexation of Aromatic and Aliphatic Guests in Water and Water-Urea Mixtures. Experimental Evidence for the Interaction of Urea with Arene Surfaces. J. Phys. Chem. B. 1997; 101:3376–3380.

47. Graves AP, Brenk R, Shoichet BK. Decoys for Docking. Journal of Medicinal Chemistry. 2005; 48:3714–3728. [PubMed: 15916423]
48. Graves AP, Shivakumar DM, Boyce SE, Jacobson MP, Case DA, Shoichet BK. Rescoring Docking Hit Lists for Model Cavity Sites: Predictions and Experimental Testing. Journal of Molecular Biology. 2008; 377:914–934. [PubMed: 18280498]
49. Gumbart JC, Roux B, Chipot C. Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? J. Chem. Theory Comput. 2013; 9:794–802. [PubMed: 23794960]
50. Henriksen NM, Fenley AT, Gilson MK. Computational Calorimetry: High-Precision Calculation of Host-Guest Binding Thermodynamics. Journal of Chemical Theory and Computation. 2015; 11:4377–4394. [PubMed: 26523125]
51. Hermans J, Subramaniam S. The free energy of xenon binding to myoglobin from molecular dynamics simulation. Isr. J. Chem. 1986; 27:225–227.
52. Hillyer MB, Gibb CLD, Sokkalingam P, Jordan JH, Ioup SE, Gibb BC. Synthesis of Water-Soluble Deep-Cavity Cavitands. Org. Lett. 2016; 18:4048–4051. [PubMed: 27500699]
53. Homeyer N, Stoll F, Hillisch A, Gohlke H. Binding Free Energy Calculations for Lead Optimization: Assessment of Their Accuracy in an Industrial Drug Design Context. J. Chem. Theory Comput. 2014; 10:3331–3344. [PubMed: 26588302]
54. Hsiao YW, Söderhjelm P. Prediction of SAMPL4 host-guest binding affinities using funnel metadynamics. J Comput Aided Mol Des. 2014; 28:443–454. [PubMed: 24535628]
55. Isaacs, L. Personal communication. 2016.
56. Jiang W, Roux B. Free Energy Perturbation Hamiltonian Replica-Exchange Molecular Dynamics (FEP/H-REMD) for Absolute Ligand Binding Free Energy Calculations. Journal of Chemical Theory and Computation. 2010; 6:2559–2565. [PubMed: 21857813]
57. Jorgensen WL. Quantum and statistical mechanical studies of liquids. 12. Simulation of liquid ethanol including internal rotation. Journal of the American Chemical Society. 1981; 103:345–350.
58. Jorgensen WL, Buckner JK, Boudon S, Tirado-Rives J. Efficient computation of absolute free energies of binding by computer simulations. Application to the methane dimer in water. The Journal of Chemical Physics. 1988; 89:3742–3746.
59. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nat Struct Mol Biol. 2002; 9:646–652.
60. Khavrutskii IV, Wallqvist A. Improved Binding Free Energy Predictions from Single-Reference Thermodynamic Integration Augmented with Hamiltonian Replica Exchange. Journal of Chemical Theory and Computation. 2011; 7:3001–3011. [PubMed: 22046108]
61. Kitahara R, Mulder F. Is pressure-induced signal loss in NMR spectra for the Leu99Ala cavity mutant of T4 lysozyme due to unfolding? Proc. Nat. Acad. Sci. 2015; 112:E923. [PubMed: 25630507]
62. Lee CT, Comer J, Herndon C, Leung N, Pavlova A, et al. Simulation-Based Approaches for Determining Membrane Permeability of Small Compounds. J. Chem. Inf. Model. 2016; 56:721–733. [PubMed: 27043429]
63. Lee JW, Lee HHL, Ko YH, Kim K, Kim HI. Deciphering the Specific High-Affinity Binding of Cucurbit[7]uril to Amino Acids in Water. The Journal of Physical Chemistry B. 2015; 119:4628–4636. [PubMed: 25757499]
64. Lee JW, Samal S, Selvapalam N, Kim HJ, Kim K. Cucurbituril homologues and derivatives: New opportunities in supramolecular chemistry. Acc. Chem. Res. 2003; 36:621–630. [PubMed: 12924959]
65. Lee MS, Olson MA. Calculation of Absolute Protein-Ligand Binding Affinity Using Path and Endpoint Approaches. Biophysical Journal. 2006; 90:864–877. [PubMed: 16284269]
66. Lerch M, Lopez C, Yang Z, Kreitman M, Horwitz J, Hubbell W. Structure-relaxation mechanism for the response of T4 lysozyme cavity mutants to hydrostatic pressure. Proceedings of the National Academy of Sciences. 2015; 112:E2437–E2446.
67. Leonis G, Steinbrecher T, Papadopoulos MG. 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. 2013; 53:2141–2153. [PubMed: 23834142]

68. Lim NM, Wang L, Abel R, Mobley DL. Sensitivity in binding free energies due to protein reorganization. Journal of Chemical Theory and Computation. 2016
69. Lin YL, Aleksandrov A, Simonson T, Roux B. An Overview of Electrostatic Free Energy Computations for Solutions and Proteins. J. Chem. Theory Comput. 2014; 10:2690–2709. [PubMed: 26586504]
70. Liu S, Cao S, Hoang K, Young KL, Paluch AS, Mobley DL. Using MD Simulations To Calculate How Solvents Modulate Solubility. Journal of Chemical Theory and Computation. 2016; 12:1930–1941. [PubMed: 26878198]
71. Liu S, Ruspic C, Mukhopadhyay P, Chakrabarti S, Zavalij PY, Isaacs L. The Cucurbit[n]uril Family: Prime Components for Self-Sorting Systems. Journal of the American Chemical Society. 2005; 127:15959–15967. [PubMed: 16277540]
72. Liu S, Wu Y, Lin T, Abel R, Redmann JP, et al. Lead optimization mapper: Automating free energy calculations for lead optimization. J Comput Aided Mol Des. 2013; 27:755–770. [PubMed: 24072356]
73. Maeno A, Sindhikara D, Hirata F, Otten R, Dahlquist F, Yokoyama S, Akaska K, Mulder F, Kitahara R. Cavity as a Source of Conformational Fluctuation and High-Energy State: High-Pressure NMR Study of a Cavity-Enlarged Mutant of T4Lysozyme. Biophys. J. 2015; 108:133–145. [PubMed: 25564860]
74. Merski M, Fischer M, Balius TE, Eidam O, Shoichet BK. Homologous ligands accommodated by discrete conformations of a buried cavity. PNAS. 2015; 112:5039–5044. [PubMed: 25847998]
75. Michel J, Essex JW. Hit Identification and Binding Mode Predictions by Rigorous Free Energy Simulations. J. Med. Chem. 2008; 51:6654–6664. [PubMed: 18834104]
76. Michel J, Essex JW. Prediction of protein-ligand binding affinity by free energy simulations: Assumptions, pitfalls and expectations. J Comput Aided Mol Des. 2010; 24:639–658. [PubMed: 20509041]
77. Mikulskis P, Cioloboc D, Andrejić M, Khare S, Brorsson J, et al. Free-energy perturbation and quantum mechanical study of SAMPL4 octa-acid host-guest binding energies. J Comput Aided Mol Des. 2014; 28:375–400. [PubMed: 24700414]
78. Mikulskis P, Genheden S, Ryde U. A Large-Scale Test of Free-Energy Simulation Estimates of Protein-Ligand Binding Affinities. J. Chem. Inf. Model. 2014; 54:2794–2806. [PubMed: 25264937]
79. Mobley, D., Chodera, J., Isaacs, L., Gibb, B. Advancing predictive modeling through focused development of new systems to drive new modeling innovations: An NIH proposal http://doi.org/10.5281/zenodo.163963. 2016.
80. Mobley DL, Chodera JD, Dill KA. On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. J. Chem. Phys. 2006; 125:084902. [PubMed: 16965052]
81. Mobley DL, Chodera JD, Dill KA. Confine-and-Release Method: Obtaining Correct Binding Free Energies in the Presence of Protein Conformational Change. Journal of Chemical Theory and Computation. 2007; 3:1231–1235. [PubMed: 18843379]
82. Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill KA. Predicting absolute ligand binding free energies to a simple model site. J. Mol. Biol. 2007; 371:1118–1134. [PubMed: 17599350]
83. Mobley DL, Klimovich PV. Perspective: Alchemical free energy calculations for drug discovery. J. Chem. Phys. 2012; 137:230901. [PubMed: 23267463]
84. Mobley DL, Zuckerman DM. A proposal for regularly updated review/survey articles: ”Perpetual Reviews”. arXiv:1502.01329[cs]. 2015
85. Mock WL, Shih NY. Host-guest binding capacity of cucurbituril. The Journal of Organic Chemistry. 1983; 48:3618–3619.
86. Moghaddam S, Inoue Y, Gilson MK. Host-Guest Complexes with Protein-Ligand-like Affinities: Computational Analysis and Design. Journal of the American Chemical Society. 2009; 131:4012–4021. [PubMed: 19133781]
87. Moghaddam S, Yang C, Rekharsky M, Ko YH, Kim K, et al. 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. 2011; 133:3570–3581. [PubMed: 21341773]

88. Monroe JI, Shirts MR. Converging free energies of binding in cucurbit[7]uril and octaacid host-guest systems from SAMPL4 using expanded ensemble simulations. J Comput Aided Mol Des. 2014; 28:401–415. [PubMed: 24610238]
89. Morton A, Baase WA, Matthews BW. Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. Biochemistry. 1995; 34:8564–8575. [PubMed: 7612598]
90. Morton A, Matthews BW. Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: Linkage of dynamics and structural plasticity. Biochemistry. 1995; 34:8576–8588. [PubMed: 7612599]
91. Muddana HS, Fenley AT, Mobley DL, Gilson MK. The SAMPL4 host-guest blind prediction challenge: An overview. J Comput Aided Mol Des. 2014; 28:305–317. [PubMed: 24599514]
92. Muddana HS, Gilson MK. Prediction of SAMPL3 host-guest binding affinities: Evaluating the accuracy of generalized force-fields. J Comput Aided Mol Des. 2012; 26:517–525. [PubMed: 22274835]
93. Muddana HS, Varnado CD, Bielawski CW, Urbach AR, Isaacs L, et al. Blind prediction of host-guest binding affinities: A new SAMPL3 challenge. J Comput Aided Mol Des. 2012; 26:475–487. [PubMed: 22366955]
94. Muddana HS, Yin J, Sapra NV, Fenley AT, Gilson MK. Blind prediction of SAMPL4 cucurbit[7]uril binding affinities with the mining minima method. J Comput Aided Mol Des. 2014; 28:463–474. [PubMed: 24510191]
95. Musah RA, Jensen GM, Bunte SW, Rosenfeld RJ, Goodin DB. 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. 2002; 315:845–857. [PubMed: 11812152]
96. Nucci N, Fuglestad B, Athanasoula E, Wand A. Role of cavities and hydration in the pressure unfolding of T4 lysozyme. Proceedings of the National Academy of Sciences. 2014; 111:13846–13851.
97. Nguyen CN, Young TK, Gilson MK. Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. The Journal of Chemical Physics. 2012; 137:044101. [PubMed: 22852591]
98. Pal RK, Haider K, Kaur D, Flynn W, Xia J, et al. 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
99. Park J, Nessler I, McClain B, Macikenas D, Baltrusaitis J, Schnieders MJ. Absolute Organic Crystal Thermodynamics: Growth of the Asymmetric Unit into a Crystal via Alchemy. J. Chem. Theory Comput. 2014; 10:2781–2791. [PubMed: 26586507]
100. Reif MM, Oostenbrink C. Net charge changes in the calculation of relative ligand-binding free energies via classical atomistic molecular dynamics simulation. J. Comput. Chem. 2014; 35:227–243. [PubMed: 24249099]
101. Rekharsky MV, Inoue Y. Complexation Thermodynamics of Cyclodextrins. Chem. Rev. 1998; 98:1875–1918. [PubMed: 11848952]
102. Rekharsky MV, Mori T, Yang C, Ko YH, Selvapalam N, et al. A synthetic host-guest system achieves avidin-biotin affinity by overcoming enthalpy-entropy compensation. PNAS. 2007; 104:20737–20742. [PubMed: 18093926]
103. Rocklin GJ, Boyce SE, Fischer M, Fish I, Mobley DL, et al. Blind Prediction of Charged Ligand Binding Affinities in a Model Binding Site. J. Mol. Biol. 2013; 425:4569–4583. [PubMed: 23896298]
104. Rocklin GJ, Mobley DL, Dill KA, Hünenberger PH. 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. 2013; 139:184103. [PubMed: 24320250]

105. Rogers KE, Ortiz-Sánchez JM, Baron R, Fajer M, de Oliveira CAF, McCammon JA. On the Role of Dewetting Transitions in Host-Guest Binding Free Energy Calculations. Journal of Chemical Theory and Computation. 2013; 9:46–53. [PubMed: 23316123]
106. Rosenfeld RJ, Hays AMA, Musah RA, Goodin DB. Excision of a proposed electron transfer pathway in cytochrome c peroxidase and its replacement by a ligand-binding channel. Protein Science. 2002; 11:1251–1259. [PubMed: 11967381]
107. Schnieders MJ, Baltrusaitis J, Shi Y, Chattree G, Zheng L, et al. The Structure, Thermodynamics, and Solubility of Organic Crystals from Simulation with a Polarizable Force Field. J. Chem. Theory Comput. 2012; 8:1721–1736. [PubMed: 22582032]
108. Schreiner, P. Theoretical prediction of affinities to cucurbiturils -the blind prediction hydrophobe challenge. 2016. https://www.uni-giessen.de/fbz/fb08/dispersion/projects/HydrophobeChallenge
109. Sherborne, Bradley. Opening the lid on FEP. J Comput Aided Mol Des. 2016
110. Shirts MR, Chodera JD. Statistically optimal analysis of samples from multiple equilibrium states. The Journal of Chemical Physics. 2008; 129:124105. [PubMed: 19045004]
111. Shirts MR, Klein C, Swails JM, Yin J, Gilson MK, et al. Lessons learned from comparing molecular dynamics englines on the SAMPL5 dataset. J Comput Aided Mol Des. 2016
112. Shirts MR, Mobley DL. An Introduction to Best Practices in Free Energy Calculations. Biomolecular Simulations, vol. 924 Methods in Molecular Biology. 2013
113. Shirts, MR., Mobley, DL., Brown, SP. Free-energy calculations in structure-based drug design. In: Merz Kenneth M, J.Ringe, D., Reynolds, CH., editors. Drug Design: Structure and Ligand-Based Approaches. Cambridge University Press; 2010.
114. Simonson T, Roux B. Concepts and protocols for electrostatic free energies. Molecular Simulation. 2016; 42:1090–1101.
115. Sokkalingam P, Shraberg J, Rick SW, Gibb BC. Binding Hydrated Anions with Hydrophobic Pockets. Journal of the American Chemical Society. 2016; 138:48–51. [PubMed: 26702712]
116. Sullivan MR, Sokkalingam P, Nguyen T, Donahue JP, Gibb BC. Binding of carboxylate and trimethylammonium salts to octa-acid and TEMOA deep-cavity cavitands. J Comput Aided Mol Des. 2016:1–8.
117. Sun H, Gibb CLD, Gibb BC. Calorimetric Analysis of the 1:1 Complexes Formed between a Water-soluble Deep-cavity Cavitand, and Cyclic and Acyclic Carboxylic Acids. Supramolecular Chemistry. 2008; 20:141–147.
118. Tai K. Conformational sampling for the impatient. Biophysical Chemistry. 2004; 107:213–220. [PubMed: 14967236]
119. Tembe BL, McCammon JA. Ligand Receptor Interactions. Comput Chem. 1984; 8:281–283.
120. Tofoleanu F, Lee J, Pickard FC IV, König G, Huang J, et al. Absolute binding free energy calculations for octa-acids and guests. J Comput Aided Mol Des. 2016
121. Velez-Vega C, Gilson MK. Force and Stress along Simulated Dissociation Pathways of Cucurbituril-Guest Systems. J. Chem. Theory Comput. 2012; 8:966–976. [PubMed: 22754402]
122. Velez-Vega C, Gilson MK. Overcoming dissipation in the calculation of standard binding free energies by ligand extraction. J. Comput. Chem. 2013; 34:2360–2371. [PubMed: 24038118]
123. Verras A. Free Energy Perturbation at Merck: Benchmarking against Faster Methods. 2016
124. Vinciguerra B, Zavalij PY, Isaacs L. Synthesis and Recognition Properties of Cucurbit[8]uril Derivatives. Org. Lett. 2015; 17:5068–5071. [PubMed: 26405845]
125. Wang K, Chodera JD, Yang Y, Shirts MR. Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. J Comput Aided Mol Des. 2013; 27:989–1007. [PubMed: 24297454]
126. Wang K, Sokkalingam P, Gibb BC. ITC and NMR analysis of the encapsulation of fatty acids within a water-soluble cavitand and its dimeric capsule. Supramolecular Chemistry. 2016; 28:84–90. [PubMed: 26997853]
127. Wang L, Berne BJ, Friesner RA. On achieving high accuracy and reliability in the calculation of relative protein-ligand binding affinities. PNAS. 2012; 109:1937–1942. [PubMed: 22308365]
128. Wang L, Wu Y, Deng Y, Kim B, Pierce L, et al. 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. 2015; 137:2695–2703. [PubMed: 25625324]

129. Wand A, Nucci N. Reply to Kitahara and Mulder: An ensemble view of protein stability best explains pressure effects in a T4 lysozyme cavity mutant. Proc. Nat. Acad. Sci. 2015; 112:E924. [PubMed: 25630509]
130. Wei BQ, Baase WA, Weaver LH, Matthews BW, Shoichet BK. A Model Binding Site for Testing Scoring Functions in Molecular Docking. Journal of Molecular Biology. 2002; 322:339–355. [PubMed: 12217695]
131. Wei BQ, Weaver LH, Ferrari AM, Matthews BW, Shoichet BK. Testing a Flexiblereceptor Docking Algorithm in a Model Binding Site. Journal of Molecular Biology. 2004; 337:1161–1182. [PubMed: 15046985]
132. Wickstrom L, Deng N, He P, Mentes A, Nguyen C, et al. Parameterization of an effective potential for protein-ligand binding from host-guest affinity data. J. Mol. Recognit. 2016; 29:10–21. [PubMed: 26256816]
133. Woo HJ, Roux B. Calculation of absolute protein-ligand binding free energy from computer simulations. PNAS. 2005; 102:6825–6830. [PubMed: 15867154]
134. Wyman IW, Macartney DH. Cucurbit[7]uril host-guest complexes with small polar organic guests in aqueous solution. Organic & Biomolecular Chemistry. 2008; 6:1796. [PubMed: 18452015]
135. Yin J, Fenley AT, Henriksen NM, Gilson MK. Toward Improved Force-Field Accuracy through Sensitivity Analysis of Host-Guest Binding Thermodynamics. The Journal of Physical Chemistry B. 2015; 119:10145–10155. [PubMed: 26181208]
136. Yin J, Henriksen NM, Slochower DR, Chiu MW, Mobley DL, Gilson MK. Overview of the SAMPL5 Host-Guest Challenge: Are We Doing Better? J Comput Aided Mol Des. 2016
137. Yin J, Henriksen NM, Slochower DR, Gilson MK. The SAMPL5 Host-Guest Challenge: Binding Free Energies and Enthalpies from Explicit Solvent Simulations. J Comput Aided Mol Des. 2016
138. Ytreberg FM. Absolute FKBP binding affinities obtained via nonequilibrium unbinding simulations. The Journal of Chemical Physics. 2009; 130:164906. [PubMed: 19405629]

Glossary

alchemical	Nonphysical or nonchemical; here, usually used in the context of calculations involving transitions between physical thermodynamic states via a nonphysical, “alchemical” pathway
test system	A system (here, often a binding system) employed for the purposes of testing methods, force fields, and other aspects of simulations
benchmark system	A standard test system used to evaluate, assess, compare, or explore the performance of methods, often quantitatively

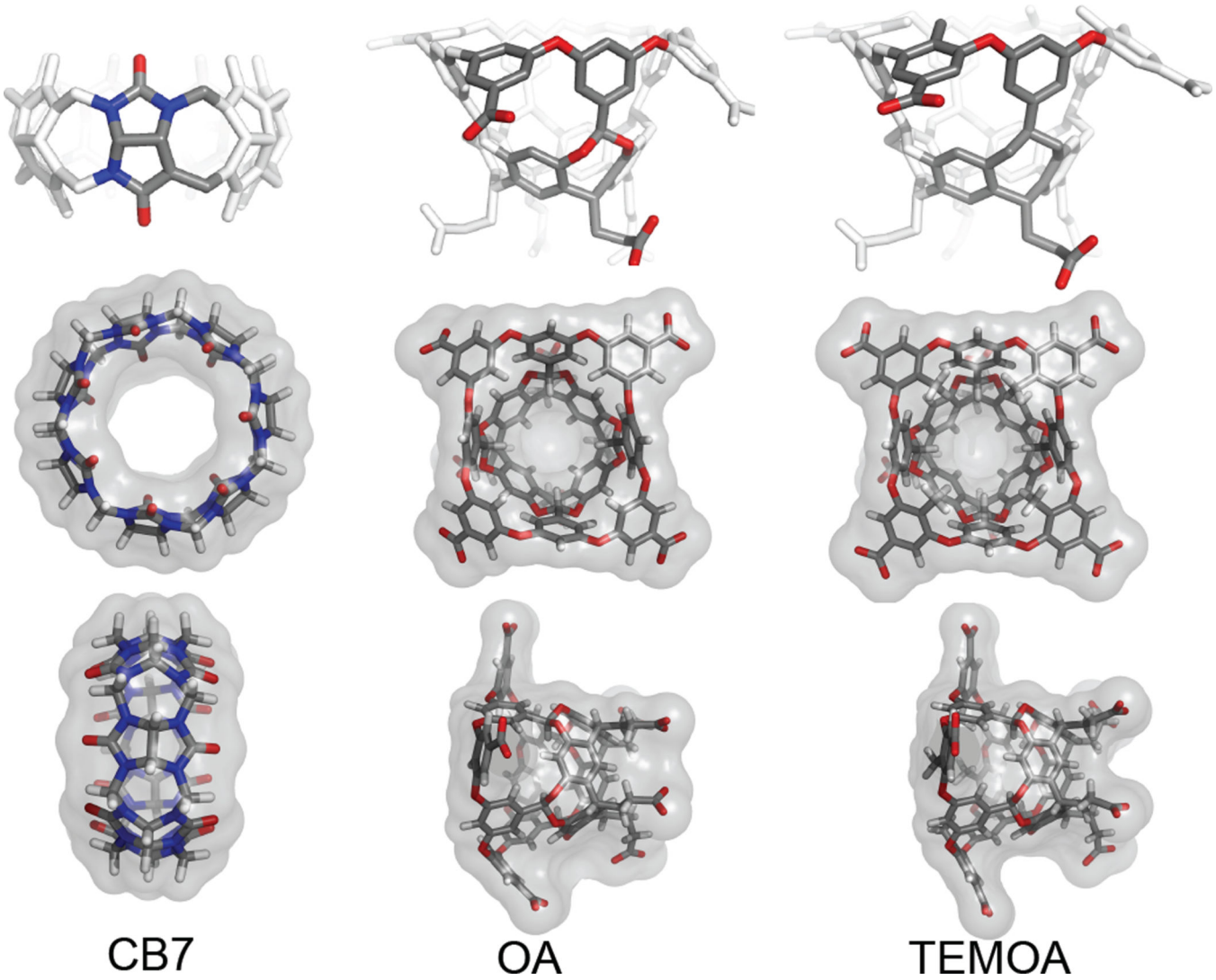

Figure 1.
OA, TEMOA, and CB7 hosts. Shown are the hosts which are the focus of our host-guest benchmark sets – two variants of the octa-acid GDCC, and CB7, a cucurbituril. Guest structures are available in the supplemental material.

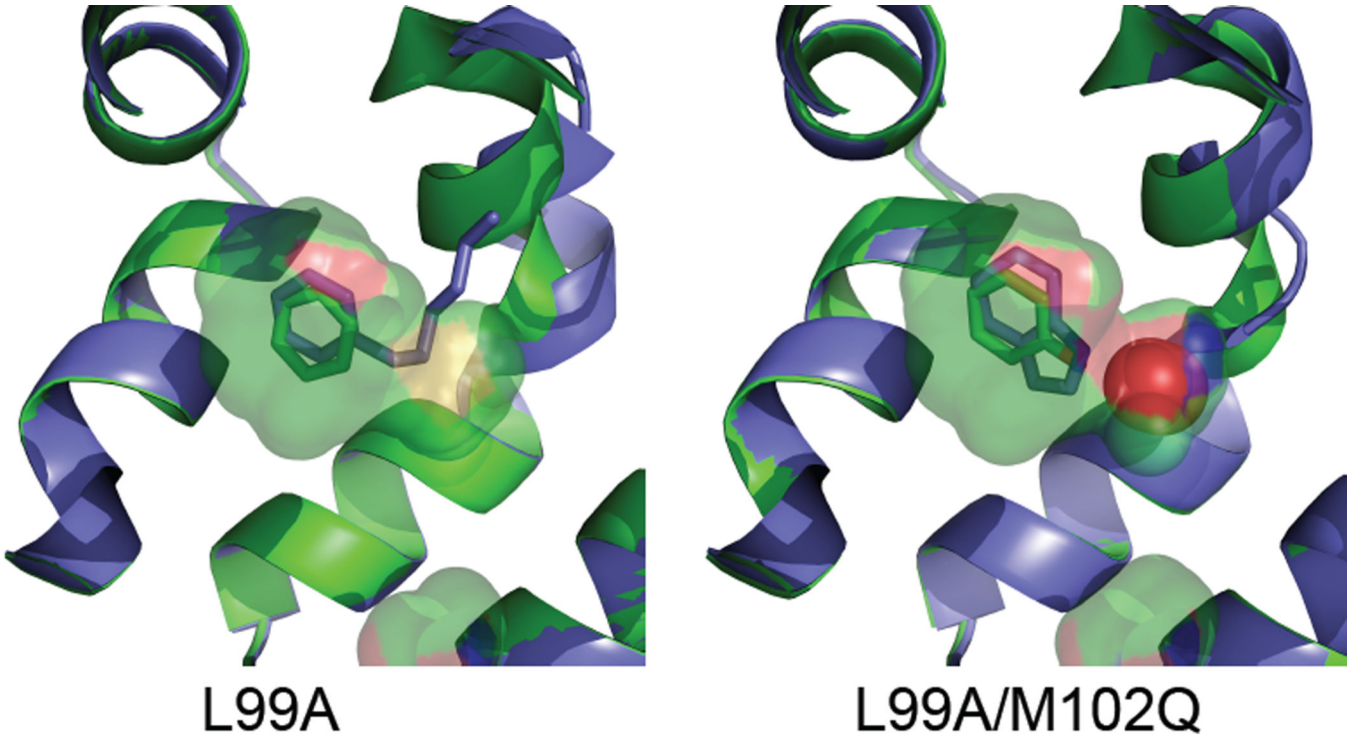

Figure 2.

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 shown 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, and 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, though this does not occur for all ligands in the polar L99A/M102Q site.

Table 1

Proposed CB7 Set 1 benchmark data

ID[a]	name	PC CID[b]	SMILES	ΔG^c (*kcal/mol*)
1	Memantine	4054	CC12CC3CC(C1)(CC(C3)(C2)N)C	−5.99 ± 0.05 [d]
3	1,3-Bis(trimethylaminio)adamantine	101379195	C[N+](C)(C)C12CC3CC(C1)CC(C3)(C2)[N+](C)(C)C	−6.55 ± 0.05 [d]
5	N-(1-Adamantyl)ethylenediamine	303798	C1C2CC3CC1CC(C2)(C3)NCCN	−18.22 ± 0.09 [e]
17	Adamantane-1,3-diamine	213512	C1C2CC3(CC1CC(C2)(C3)N)N	−11.33 ± 0.05 [d]
18	1-Adamantanecarboxylic acid	13235	C1C2CC3CC1CC(C2)(C3)C(=O)O	−11.59 ± 0.06 [d]
22	1-Adamantyltrimethylaminium	3010127	C[N+](C)(C)C12CC3CC(C1)CC(C3)C2	−16.66 ± 0.08 [d]
23	amantadine	2130	C1C2CC3CC1CC(C2)(C3)N	−17.19 ± 0.08 [d]
24	N-(1-Adamantyl)pyridinium	3848257	C1C2CC3CC1CC(C2)(C3)[N+]4=CC=CC=C4	−16.75 ± 0.07 [d]

[a] Compound ID from original paper;

[b] PubChem Compound ID (structures in supporting info);

[c] 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 (55) and dividing by $\sqrt{3}$;

[d] drawn from (71);

[e] drawn from (15).

Table 2

Proposed CB7 Set 2 benchmark data

ID[a]	name	PC CID[b]	SMILES	$\Delta G^{c,d}$ (*kcal/mol*)
2	Dopamine	681	C1=CC(=C(C=C1CCN)O)O	−6.31 ± 0.05
4	O-phenylenediamine	7243	C1=CC=C(C(=C1)N)N	−6.68 ± 0.05
5	m-Phenylenediamine	7935	C1=CC(=CC(=C1)N)N	−6.69 ± 0.02
7	4-(Aminomethyl)pyridine	77317	C1=CN=CC=C1CN	−7.56 ± 0.06
8	p-Phenylenediamine	7814	C1=CC(=CC=C1N)N	−8.60 ± 0.06
9	P-toluidine	7813	CC1=CC=C(C=C1)N	−9.43 ± 0.05
20	P-Xylylenediamine	68315	C1=CC(=CC=C1CN)CN	−12.62 ± 0.06

[a] Compound ID from original paper;

[b] PubChem Compound ID (structures in supporting info);

[c] 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 (55) and dividing by $\sqrt{3}$;

[d] drawn from (71).

Table 3

Proposed GDCC Set 1 benchmark data

ID[a]	name	PC CID[b]	SMILES	ΔG^c (*kcal/mol*)	ΔH^d (*kcal/mol*)
Octa Acid binders					
3 / OA-G1	5-Hexynoic acid	143036	C#CCCCC(=O)O	−5.40 ± 0.003	−7.71 ± 0.05
4 / OA-G6	3-nitrobenzoic acid	8497	C1=CC(=CC(=C1)[N+](=O)[O-])C(=O)O	−5.34 ± 0.005	−5.67 ± 0.01
5 / OA-G2	4-cyanobenzoic acid	12087	C1=CC(=CC=C1C#N)C(=O)O	−4.73 ± 0.01	−4.45 ± 0.08
6 / OA-G4	4-bromoadamantane-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	CCCCCC[N+](C)(C)C	−4.49 ± 0.01	−5.91 ± 0.10
8 / OA-G5	trimethylphenethylaminium	14108	C[N+](C)(C)CCC1=CC=CC=C1	−3.72 ± 0.01	−9.96 ± 0.11
TEMOA/OAMe binders					
3 / OA-G1	5-Hexynoic acid	143036	C#CCCCC(=O)O	−5.476 ± 0.006	−9.961 ± 0.006
4 / OA-G6	3-nitrobenzoic acid	8497	C1=CC(=CC(=C1)[N+](=O)[O-])C(=O)O	−4.52 ± 0.02	−9.1 ± 0.1
5 / OA-G2	4-cyanobenzoic acid	12087	C1=CC(=CC=C1C#N)C(=O)O	−5.26 ± 0.01	−7.6 ± 0.1
6 / OA-G4	4-bromoadamantane-1-carboxylic acid	12598766	C1C2CC3CC(C2)(CC1C3Br)C(=O)O	ND[e]	ND[e]
7 / OA-G3	N,N,N-trimethylhexan-1-aminium	84774	CCCCCC[N+](C)(C)C	−5.73 ± 0.06	−6.62 ± 0.2
8 / OA-G5	trimethylphenethylaminium	14108	C[N+](C)(C)CCC1=CC=CC=C1	ND[e]	ND[e]

[a] Compound ID from (116) and SAMPL5 ID from (136);

[b] PubChem Compound ID (structures in supporting info);

[c] Standard binding free energy from (116), 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 ΔG and ~6% for ΔH;

[d] measured binding enthalpy (116), subject to the same conditions/caveats as [c].

[e] not done.

Table 4

Proposed GDCC Set 2 benchmark data

ID[a]	name	PC CID[b]	SMILES	Method	ΔG^c (*kcal/mol*)
1	Benzoic acid	243	C1=CC=C(C=C1)C(=O)O	NMR	−3.72 ± 0.03
2	4-Methylbenzoic acid	7470	CC1=CC=C(C=C1)C(=O)O	NMR	−5.85 ± 0.06
3	4-ethylbenzoic acid	12086	CCC1=CC=C(C=C1)C(=O)O	ITC	−6.27 ± 0.01
4	4-Chlorobenzoic acid	6318	C1=CC(=CC=C1C(=O)O)Cl	ITC	−6.72 ± 0.01
5	3-chlorobenzoic acid	447	C1=CC(=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](CC1)C(=O)O[H])C	ITC	−7.61 ± 0.04

[a] Compound ID from original paper (44);

[b] PubChem Compound ID (structures in supporting info);

[c] Standard binding free energy from (44), 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 communication with the authors, uncertainties of perhaps 10% may be more appropriate.

Table 5

Proposed Lysozyme L99A Set benchmark data

name	ID[h]	SMILES	ΔG^a (*kcal/mol*)	PDB code	reference
benzene[b]	241	c1ccccc1	−5.19 ± 0.16	181L (90), 4W52 (74)	(89)
toluene[b]	1140	Cc1ccccc1	−5.52 ± 0.04	4W53 (74)	(89)
ethylbenzene[b]	7500	CCc1ccccc1	−5.76 ± 0.07	1NHB (90), 4W54 (74)	(89)
propylbenzene[b]	7668	CCCc1ccccc1	−6.55 ± 0.02	4W55 (74)	(89)
butylbenzene[b]	7705	CCCCc1ccccc1	−6.70 ± 0.02	186L (90), 4W57 (74)	(89)
hexylbenzene[b]	14109	CCCCCCc1ccccc1	UNK[c]	4W59 (74)	(89)
p-xylene[d]	7809	Cc1ccc(cc1)C	−4.67 ± 0.06	187L (90)	(89)
benzofuran	9223	c1ccc2c(c1)cco2	−5.46 ± 0.03	182L (90)	(89)
thieno[2,3-c]pyridine	9224	c1cncc2c1ccs2	NB[e]	ND[f]	(82)
phenol[g]	996	c1ccc(cc1)O	NB[e]	ND[f]	(89, 82)

[a] T=302K, with compounds from (89) 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;

[b] part of the series of (74), so larger ligands in the series induce conformational change;

[c] unknown due to solubility limitations, but likely binds strongly;

[d] L99A sidechain undergoes rotation;

[e] nonbinder;

[f] not done;

[g] included since it is a binder in the polar cavity;

[h] PubChem ID (structures in supporting info).

Table 6

Proposed Lysozyme L99A/M102Q Set benchmark data

ligand	ID[k]	SMILES	ΔG[a] (*kcal/mol*)	PDB code	reference
toluene[b]	1140	Cc1ccccc1	−4.93	ND[c]	(130)
phenol	996	c1ccc(cc1)O	−5.24	1LI2 (130)	(130)
catechol[h]	289	c1ccc(c(c1)O)O	−4.16 ± 0.03	1XEP (47)	(130)
2-ethoxyphenol[d]	66755	CCOc1ccccc1O	−4.02 ± 0.03	3HU8 (12)	(12)
benzyl acetate[e,f]	8785	CC(=O)OCc1ccccc1	−4.48 ± 0.16	3HUK (12)	(12)
4,5,6,7-tetrahydroindole[f]	57452536	c1c[nH]c2c1CCCC2	−4.61 ± 0.09	3HUA (12)	(12)
n-phenylglycinonitrile[g]	76372	c1ccccc1NCC#N	−5.52 ± 0.18	2RBN (12)	(12)
3-chlorophenol	7933	c1cc(cc(c1)Cl)O	−5.51	1LI3 (130)	(130)
2-methoxyphenol	460	COc1ccccc1O	NB[i]	ND[c]	(12)
4-vinylpyridine	7502	C=Cc1ccncc1s	NB[i]	ND[c]	(130)

[a] T=283K, with measurements done at pH 6.8 in 50 mM potassium phosphate, 200 mM potassium chloride buffer in the case of (12);

[b] included for symmetry with the L99A site since this (unlike phenol and benzene) binds in both;

[c] not determined;

[d] fails to make crystallographic hydrogen bond (12);

[e] multiple binding modes;

[f] induces helix F motion;

[g] induces flip of Val111 sidechain;

[h] induces flip of Leu118 sidechain;

[j] nonbinder;

[k] PubChem ID (structures in supporting info).