Taming Multiple Binding Poses in Alchemical Binding Free Energy Prediction: the β-cyclodextrin Host-Guest SAMPL9 Blinded Challenge

Sheenam Khuttan · Solmaz Azimi · Joe Z. Wu · Sebastian Dick[†] · Chuanjie Wu[‡] · Huafeng Xu[§] · Emilio Gallicchio[*]

Abstract We apply the Alchemical Transfer Method (ATM) and a bespoke fixed partial charge force field to the SAMPL9 bCD host-guest binding free energy prediction challenge that comprises a combination of complexes formed between five phenothiazine guests and two cyclodextrin hosts. Multiple chemical forms, competing binding poses, and computational modeling challenges pose significant obstacles to obtaining reliable computational predictions for these systems. The phenothiazine guests exist in solution as racemic mixtures of enantiomers related by nitrogen inversions that bind the hosts in various binding poses, each requiring an individual free energy analysis. Due to the large size of the guests and the conformational reorganization of the hosts, which prevent a direct absolute binding free energy route, binding free energies are obtained by a series of absolute and relative binding alchemical steps for each chemical species in each binding pose. Metadynamics-accelerated conformational sampling was found to be necessary to address the poor convergence of some numerical estimates affected by conformational trapping. Despite these challenges, our blinded predictions quantitatively reproduced the experimental affinities for the β-cyclodextrin host and, to a lesser extent, those with a methylated derivative. The work illustrates the challenges of obtaining reliable free energy data in in-silico drug design for even seemingly simple systems and introduces some of the technologies available to tackle them.

Keywords Binding free energy, ligand Binding, alchemical method, alchemical transfer, SAMPL, binding modes, conformational sampling, metadynamics, β-cyclodextrin

[†]Present affiliation: D.E. Shaw Research, New York, NY

[‡]Present affiliation: Schrödinger Inc, New York, NY

[§]Present affiliation: Atomapper Inc, New York, NY

S. Khuttan
Department of Chemistry, Brooklyn College of the City University of New York
PhD Program in Biochemistry, Graduate Center of the City University of New York

S. Azimi
Department of Chemistry, Brooklyn College of the City University of New York
PhD Program in Biochemistry, Graduate Center of the City University of New York

J. Z. Wu
Department of Chemistry, Brooklyn College of the City University of New York
PhD Program in Chemistry, Graduate Center of the City University of New York

S. Dick
Roivant Sciences, New York, NY

C. Wu
Roivant Sciences, New York, NY

H. Xu
Roivant Sciences, New York, NY

E. Gallicchio
Department of Chemistry, Brooklyn College of the City University of New York
PhD Program in Chemistry, Graduate Center of the City University of New York
PhD Program in Biochemistry, Graduate Center of the City University of New York
E-mail: egallicchio@brooklyn.cuny.edu

1 Introduction

Developing in-silico methodologies capable of consistent and reliable binding free energy estimates would be a major breakthrough for drug design and other areas of chemical research.[1, 2, 3, 4] With several advanced simulation software packages now routinely used in industry and academia to model binding affinities of

protein-drug complexes,[5, 6] the field has made significant strides toward this goal. Alchemical methods have emerged as the industry standard partly because they can target the changes of binding affinities upon specific chemical modifications of the ligands directly.[7, 8, 9] Theoretical and methodological aspects of free energy are continuously refined and improved.[10, 11, 12, 13] However, many challenges remain about the quality of potential energy models[14] and the correct representation of all of the relevant conformations of the molecular systems.[15]

The validation of computational predictions with respect to experimental binding affinities has given the community an understanding of the pitfalls of the models, with indications of ways in which they can be avoided.[16, 9] Blinded validations, such as the Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL),[17] have played a particularly important role in this process.[18, 19, 20, 21] Because computational predictions are formulated without prior knowledge of experimental results, the evaluation of the models' relative performance is free of implicit biases and reflects more faithfully the expected reliability of the computational models in actual research and discovery settings.

Many SAMPL challenges include host-guest systems are considered to be straightforward, as well as more approachable, theoretically and experimentally, than macromolecular systems in terms of testing for reliability in free energy prediction tools.[22, 23] In this work, we report our findings in tackling the SAMPL9 bCD challenge set, which includes the binding of five phenothiazine-based drugs[24] to the β-cyclodextrin host and its methylated derivative.[25, 26] Molecular complexes of β-cyclodextrin (bCD) are well-known and are used in a variety of biomedical and food science applications.[27] They are extensively modeled[28, 29, 30, 31, 32, 33] and thus provide a familiar testing ground for computational models. However, as we will show, the binding equilibrium between phenothiazines and cyclodextrin hosts is far from straightforward and requires deploying the most advanced computational tools and methods in our arsenal. As also discussed in later sections, handling conformational heterogeneity in the form of chirality and multiple binding poses has been the greatest challenge in our computational protocol.

This paper is organized as follows. We first describe the above molecular systems in detail to illustrate how these exist in equilibrium as a mixture of many conformations, each with its distinct binding characteristics. We then review the Alchemical Transfer Method (ATM)[33, 34] and the FFEngine bespoke force field parameterization used here to estimate the binding free energies of the cyclodextrin complexes. We describe the extensive alchemical process involving absolute as well as relative binding free energy calculations to obtain the binding constants of each pose in the host-guest systems and how these constants are combined[35] to yield values comparable to the experimental readouts. Reaching convergence for some complexes involving slow intramolecular degrees of freedom required advanced metadynamics-based conformational sampling strategies,[36, 37] which we incorporated into the alchemical binding free energy calculations. This significant intellectual and computational effort resulted in converged binding free energy estimates with a very good experimental agreement for the bCD host. The effort also illustrates the major challenges inherent in modeling complex molecular binding phenomena as well as the theories and technologies available to tackle these challenges.

2 Molecular Systems

The bCD SAMPL9 challenge concerned the binding of five phenothiazine drugs (Figure 1)[38] to β-cyclodextrin (hereafter bCD) and a modified β-cyclodextrin (hereafter mCD) (Figure 2). The guests[24] share a 3-ring phenothiazine scaffold with a variable alkylamine sidechain on the nitrogen atom of the central ring. Unlike the other guests, PMT's alkylamine sidechain is branched at a chiral center. The CPZ, TDZ, and TFP guests also have a small substituent (chlorine, thiomethyl, and trifluoromethyl, respectively) on one of the aromatic rings of the phenothiazine scaffold.

The β-cyclodextrin host (Figure 2) is a cyclic oligosaccharide of seven D-glucose monomers forming a binding cavity with a wide opening surrounded by secondary hydroxyl groups (top in Figure 2) and a narrower opening (bottom) surrounded by primary hydroxyl groups. Hence, we will refer to the wide opening as the secondary face of bCD and the narrow opening as the primary face. The second host, heptakis-2,6-dimethyl-β-cyclodextrin (mCD, Figure 2), is a derivative of β-cyclodextrin in which all of the primary hydroxyl groups and half of the secondary ones are methylated, affecting the size, accessibility, and hydrophobicity of the binding cavity. Although mCD does not have secondary and primary hydroxyl groups, for simplicity, we will refer to the two openings of mCD as secondary and primary faces by analogy with bCD. Being composed of chiral monomers, bCD and mCD are themselves chiral molecules with potentially different affinities for the enantiomers of optically active guests.[39]

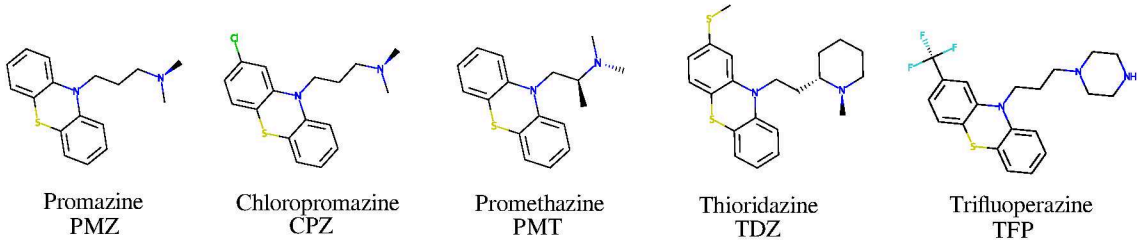

Fig. 1 The phenothiazine molecular guests included in the SAMPL9 β-cyclodextrin challenge.

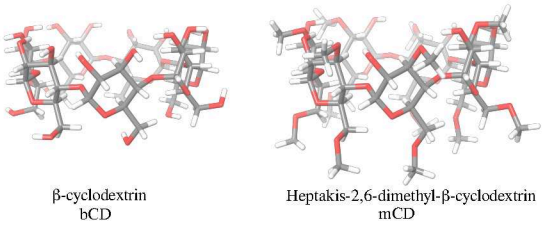

Fig. 2 The β-cyclodextrin (left) and the heptakis-2,6-dimethyl-β-cyclodextrin (right) molecular hosts included in the SAMPL9 β-cyclodextrin challenge. The top face β-cyclodextrin is surrounded by primary hydroxyl groups and the bottom face by secondary hydroxyl groups. The corresponding faces of heptakis-2,6-dimethyl-β-cyclodextrin are partially or totally methylated relative to β-cyclodextrin.

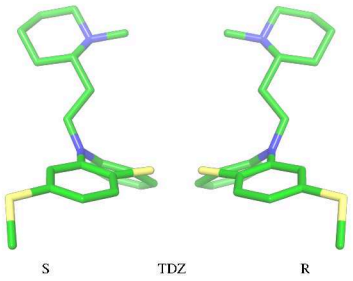

Fig. 3 Illustration of the two conformational enantiomers of the TDZ guest. Similarly to the CPZ and TFP guests, chirality is induced by the phenothiazine substituent (a thioether here). The unsubstituted guests PMZ and PMT do not possess conformational chirality.

2.1 Multiple Chemical Species of the Guests

The amine group of the alkylamine sidechain is expected to be largely protonated in solution and the host-guest complex at pH 7.4 of the experiment. However, the two tautomers of the protonated piperazine group of the TFP guest are likely to exist at appreciable concentrations and can contribute to host binding to different extents. Similarly, in the case of TDZ, protonation of the alkyl nitrogen produces two enantiomers that can interact differently with the cyclodextrin hosts. Moreover, rather than being planar, the phenothiazine ring system is bent at the connecting central ring with conformations with both positive and negative bends present in equal amounts in solution (Figure 3). As illustrated for TDZ in Figure 3, when a substituent of the phenothiazine moiety is present, the species with positive and negative bend are conformational enantiomers, each with the potential to interact differently with the cyclodextrin hosts.[39, 40]

The experimental binding assay reports an average over the contributions of the various chemical species of the guests. However, because interconversions between species cannot occur in molecular mechanics simulations or occur too slowly relative to the molecular dynamics timescales, to obtain a binding affinity estimate comparable with the experimental observations, it is necessary to model the binding of each relevant species individually and combine the results.[41] In this work, we have considered the two conformational enantiomers for each guest (including those of PMZ and PMT with the unsubstituted phenothiazine scaffold compounds to test for convergence), plus the two chiral forms of the protonated alkyl nitrogen of TDZ and the two forms of TFP protonated at the distal and proximal alkyl nitrogens, for a total of 14 guest species. We labeled the seven species with R chirality of the phenothiazine scaffold as PMZ1R, CPZ1R, PMT1R, TDZNR1R, TDZNS1S, TFP1aR, and TFP1bR, where the first part of the label identifies the compound, followed by the net charge (+1 for all the species considered) with 'a' and 'b' label identifying the distal and proximal protonated forms of TFP respectively, plus 'NR' and 'NS' labels to distinguish the R and S chiral forms of the protonated alkyl nitrogen of TDZ. The last letter identifies the chirality of the phenothiazine scaffold so that the seven species with S chirality are named PMZ1S, CPZ1S, etc.

2.2 Multiple Binding Poses

We modeled the guests binding to the cyclodextrin hosts in four distinct binding modes (Figure 4) in which the phenothiazine moiety goes through the bCD cavity.[24])

To identify the binding modes, we will refer to the narrow opening of the β-cyclodextrin circled by primary hydroxyl groups as the primary face of the host (the bottom opening in Figure 2). Similarly, the wider opening (top in Figure 2) surrounded by secondary hydroxyl groups will be referred to as the secondary face of the hosts. The guests can bind the cyclodextrin hosts with the alkylamine sidechain pointing towards the host's secondary (denoted by 's') or primary (denoted by 'p') faces (Figure 4). Each of these poses is further classified in terms of the position of the small substituent of the phenothiazine moiety, which can be at either the secondary or primary faces of the host. Hence the binding modes of the guest/cyclodextrin complexes are labeled: 'ss', 'sp', 'ps', and 'pp', where the first letter refers to the position of the alkylamine sidechain and the second to the position of the small substituent (Figure 4).

The binding mode labels are combined with the labels discussed above that identify the chemical form of the guest to obtain the labels for each form of the guest in each binding mode. For example, the guest PMT with +1 charge with R chirality in the 'ss' binding mode is labeled as PMT1Rss.

For the purpose of the alchemical calculations, the binding modes are defined geometrically in terms of the polar angle θ and the azimuthal angle ψ illustrated in Figure 6. θ is the angle between the molecular axes of the host and the guest and determines the orientation of the alkylamine sidechain relative to the host. The molecular axis of the cyclodextrin host (labeled z in Figure 6) is oriented from the primary to the secondary faces going through the centroid of the atoms lining the faces (see Computational Details). The molecular axis of the guests goes from the sulfur and nitrogen atoms of the central phenothiazine ring. The angle ψ describes the rotation around the molecular axis of the guest and determines the position of the of the phenothiazine substituent. The 'sp' binding mode, for example, is defined by $0 \leq \theta \leq 90°$ and $90° \leq \psi \leq 180°$ (see Figure 4 and Computational Details).

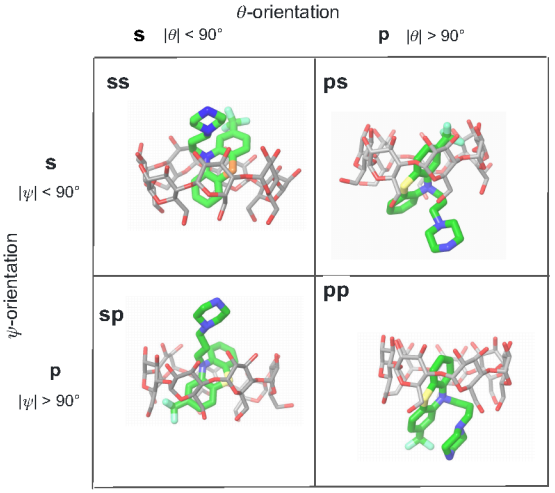

Fig. 4 Illustration of the classification of the four binding poses of the phenothiazine/cyclodextrin complexes based on the polar and twist angles introduced in Figure 6. Poses are labeled as 'ss', 'sp', etc. where 's' refers to the secondary face of the host and 'p' to the primary face of the host. The first letter of the label refers to the orientation of the alkylamine sidechain that can protrude out from the secondary face (poses 'ss' and 'sp') or from the primary face of the host (poses 'ps' and 'pp'). Similarly, the second letter refers to the position of the small substituent of the phenothiazine moiety protruding out from either the primary or secondary faces of the host.

3 Theory and Methods

3.1 Design of the Alchemical Process

The alchemical calculations aim to estimate the guests' absolute binding free energies (ABFEs) to each host. Direct alchemical ABFE calculations failed to reach convergence for these systems partly due to the relatively large sizes of the guests and partly because of the slow conformational reorganization of the cyclodextrin hosts from a closed apo state to an open guest-bound state.[42, 31] To overcome these obstacles, we adopted a step-wise process made of a series of relative binding free energy calculations (RBFE) starting from the ABFE of a small guest that could be reliably estimated. Specifically, we obtained the ABFE of trans-4-methylcyclohexanol (Figure 5)–the G1 guest of the SAMPL7 bCD challenge[43]–for the secondary and primary binding modes to each host. We defined the 'G1s' binding mode of the G1 guest as the one in which the hydroxyl group points toward the secondary face of the cyclodextrin host, while it points to the opposite face in the 'G1p' mode (Figure 7).

Each binding mode of the complex with G1 was then alchemically converted to an intermediary phenothiazine guest (N-methylphenothiazine, or MTZ, in Figure 5), similar to the SAMPL9 phenothiazine derivatives with a small methyl group replacing the large alkylamine sidechains. Even though MTZ does not have conformational chirality (Figure 3), we treated its S and R enantiomers individually to test the convergence of the RBFE estimates for each binding pose. We used atom indexes to distinguish the S and R enantiomers of these symmetric guests. Calculations were conducted to obtain the RBFEs from the G1s to the MTZRsp, MTZRss, MTZSsp, and MTZSss binding poses of the complexes of MTZ with bCD and mCD, and from G1p to the MTZRps, MTZRpp, MTZSps, and MTZSpp of

Fig. 5 The structures and abbreviations of the molecular guests used as intermediate compounds in the alchemical process.

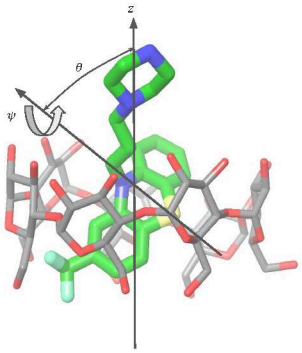

Fig. 6 Illustration of the geometrical definition of the binding poses of the phenothiazine/cyclodextrin complexes. The definition is based on the polar (θ) and twist (ψ) angles of the molecular axis of the guest with respect to the coordinate frame of the host, which includes the z-axis that runs from the primary to the secondary face of the host. See Computational Details for the specific definition of the guests' and hosts' coordinate frames.

the same complexes, all independently and from different starting conformations. The MTZRsp, MTZRss, MTZSsp, and MTZSss complexes are equivalent and should yield the same RBFE values within uncertainty. Similarly, the MTZRps, MTZRpp, MTZSps, and MTZSpp should yield equivalent RBFEs but distinguishable from those of the MTZRsp, MTZRss, MTZSsp, MTZSss group.

Next, RBFEs were obtained for each complex of MTZ to the corresponding complex of PMZ. For example, the MTZRsp binding pose of the MTZ complex with bCD and mCD were converted to the PMZ1Rsp binding pose of the corresponding complexes between PMZ and the hosts. Finally, the RBFEs between each pose of PMZ and the corresponding binding poses of the other guests were obtained. During this process, we monitored convergence by looking at the discrepancy between the RBFEs corresponding to the equivalent symmetric poses of the achiral PMZ and PMT guests. The overall alchemical process to obtain the ABFEs of the SAMPL9 phenothiazine guests is summarized in Figure 8.

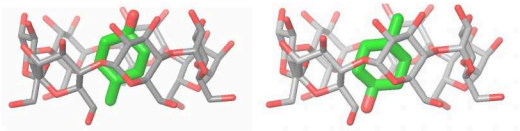

Fig. 7 The 's' and 'p' binding modes of the G1/β-cyclodextrin complex. The 's' mode, in which the hydroxyl group points towards the secondary face of the host, is used as a starting species for the 'ss' and 'sp' binding modes of the phenothiazine/cyclodextrin complexes. The 'p' mode, which points towards the primary face, is the starting species for the 'ps' and 'pp' modes (Figure 4).

3.1.1 Free Energy of Binding for Complexes with Multiple Binding Modes

The observed binding constant K_b of the complex RL of a receptor R with a ligand L present in forms or poses L_i, $i = 1, 2, \ldots$ is the weighted average of the binding constant $K_b(i)$ for each form with weights equal to the relative population $P_0(i)$ of each form in solution[41, 35]

$$K_b = \sum_i P_0(i) K_b(i) \tag{1}$$

When expressed in terms of binding free energies, Eq. (1) becomes

$$\Delta G_b = \ln \left[\sum_i P_0(i) e^{-\Delta G_b(i)/k_B T} \right] \tag{2}$$

where ΔG_b is the overall binding free energy and $\Delta G_b(i)$ the binding free energy of mode i. Statistical mechanics-based derivations of the latter formula, which we refer to as the Free Energy Combination Formula, are available.[41, 35] The Free Energy Combination Formula can also be derived using elementary notions as follows: $K_b = C^\circ \frac{[RL]}{[R][L]} = \sum_i C^\circ \frac{[RL_i]}{[R][L]} = \sum_i \frac{[L_i]}{[L_i]} C^\circ \frac{[RL_i]}{[R][L]}$ and $C^\circ \frac{[RL_i]}{[R][L]} = K_b(i)$ and $\frac{[L_i]}{[L]} = P_0(i)$, where $C^\circ = 1$ mol/L, $[RL] = \sum_i [RL_i]$ is the total molar concentration of the complex and $[RL_i]$ is the concentration of mode i of the complex. Similar definitions apply to the concentrations of the ligand $[L]$ and $[L_i]$, and $P_0(i) = [L_i]/[L]$ is the population of mode i of the ligand in solution.

Moreover, as also shown in the Appendix, the fractional contribution of binding mode i to the overall binding constant is the population, $P_1(i)$ of mode i of the complex:[35]

$$P_1(i) = \frac{P_0(i) K_b(i)}{K_b} \tag{3}$$

Below we used this property to infer the probability of occurrence of each mode of the host-guest complexes.

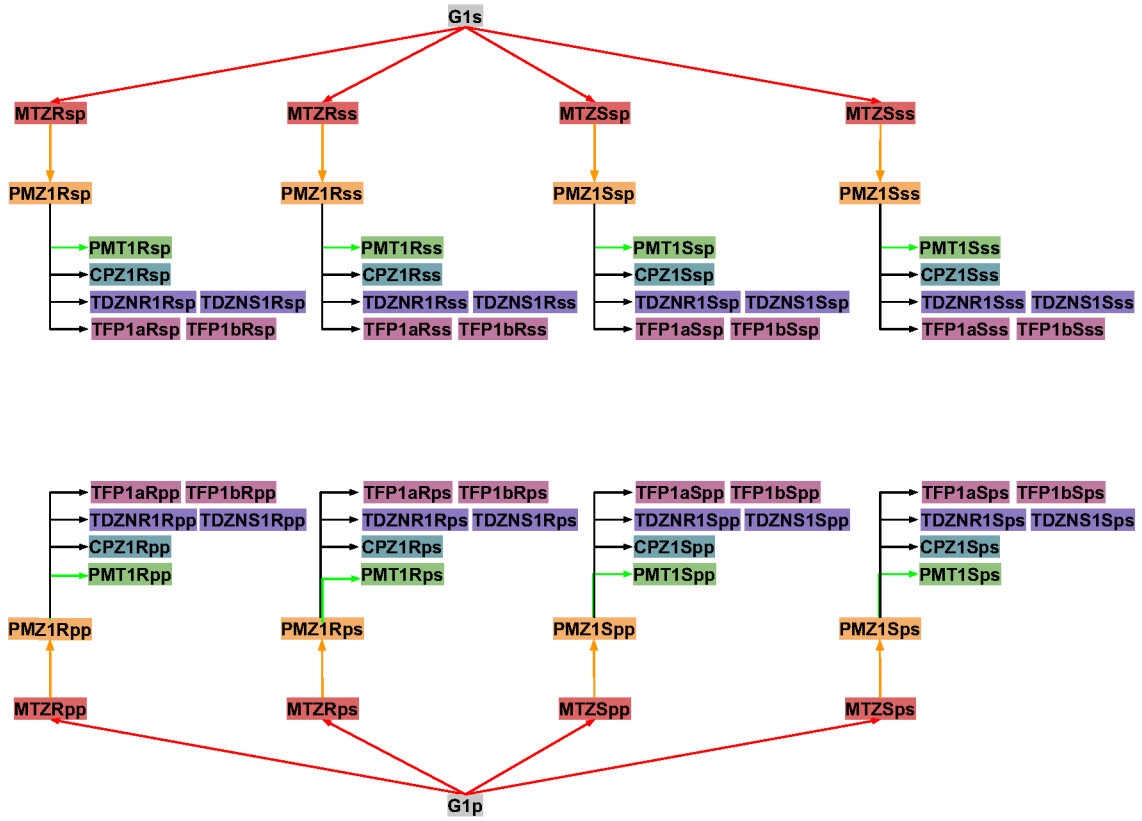

Fig. 8 The map of relative binding free energy calculations to obtain the binding free energies of each pose of each guest starting from the absolute binding free energy of the corresponding poses of the G1 guest. Nodes of the same color contribute to the binding free energy estimate of one of the five guests: PMZ (yellow), PMT (green), CPZ (cyan), TDZ (violet), and TFP (purple).

In this specific application, the binding modes refer to the 'ss', 'sp', etc. orientations of the R and S enantiomers of each guest. We individually obtained the binding constants $K_b(i)$ for each binding mode. In the corresponding alchemical simulations, the orientation of the ligand in the binding site is set by restraining potentials based on the θ and ψ angles (see Figure 6 and Computational Details). These orientations are equally likely in solutions. We also assume an equal likelihood of the R and S conformational enantiomers of the guests in solution, leading to $P_0(i) = 1/8$ for each pose of each guest. The TDZ and TFP guests have twice as many poses due to point chirality and multiple protonation states of their alkylamine sidechain that are approximately equally likely in solution based on pKa analysis with Epik.[44] Hence, for simplicity, we set $P_0(i) = 1/16$ for each state of the TDZ and TFP guests.

3.2 The Alchemical Transfer Method

The Alchemical Transfer Method (ATM) is based on a coordinate displacement perturbation of the ligand between the receptor-binding site and the explicit solvent bulk and a thermodynamic cycle connected by a symmetric intermediate in which the ligand interacts with the receptor and solvent environments with equal strength.[33, 20] The perturbation energy u for transferring the ligand from the solution to the binding site is incorporated into a λ-dependent alchemical potential energy function of the form

$$U_\lambda(x) = U_0(x) + W_\lambda(u) \tag{4}$$

where x represents the system's coordinates, $U_0(x)$ is the potential energy function that describes the unbound state, and W_λ is the softplus alchemical potential[13, 33, 20] such that the system's potential energy function is transformed from that of the unbound state to that of the bound state as λ goes from 0 to 1. This alchemical process computes the excess component of the free energy of binding. The ideal term $-k_BT \ln C^\circ V_{\text{site}}$, where $C^\circ = 1$ mol/L and V_{site} is the volume of the binding site is added in post-processing to obtain the standard free energy of binding.

In this work, we used the strategy above to compute the absolute binding free energy (ABFE) of the G1

guest in two different poses. The binding free energies of the phenothiazine guests are obtained by a series of relative binding free energy (RBFE) calculations starting from G1 (Figure 8). The alchemical RBFE implementation of the ATM method[34] is similar to ABFE implementation except that one ligand of the pair under investigation is transferred from the solution to the binding site while the other ligand is transferred in the opposite direction. The receptor and the two ligands are placed in a single solvated simulation box in such a way that one ligand is bound to the receptor and the other is placed in an arbitrary position in the solvent bulk. Molecular dynamics simulations are then conducted with a λ-dependent alchemical potential energy function that connects, in an alchemical sense, the state of the system where one guest is bound to the receptor and the other in solution, to the state where the positions of the two guests are reversed. The free energy change of this process yields the relative binding free energy of the two guests. To enhance convergence, the two ligands are kept in approximate alignment to prevent the one in solution to reorient away from the orientation of the bound pose. We have shown mathematically that the alignment restraints implemented in ATM do not bias the binding free energy estimates.[34]

In this work, we employed metadynamics conformational sampling to obtain converged RBFE estimates for the PMT guest. Well-tempered metadynamics[45] is a well-established enhanced sampling technique to sample rare events during MD simulations when they are separated from other metastable states by large energy barriers. The technique uses a bias potential, U_{bias}, that lowers energy barriers along a slow degree of freedom. In this work, the metadynamics biasing potential is obtained along a dihedral angle, φ, of the alkylamine sidechain of PMT (see Computational Details) from a simulation in a water solution, using OpenMM's well-tempered metadynamics implementation by Peter Eastman.[37] The alchemical binding free energy calculation is then conducted with the biasing potential, $U_{\text{bias}}(\varphi)$, added to the alchemical potential energy function in Eq. (4). The resulting binding free energy estimate is then unbiased using a book-ending approach[46] by computing the free energy differences of the system without the biasing potential from samples collected with the biasing potential at the endpoints of the alchemical path. In this work, we used a simple unidirectional exponential averaging formula

$$-k_B T \ln\langle \exp(U_{\text{bias}}/k_B T)\rangle_{\text{bias}}$$

to evaluate the free energy corrections for unbiasing. Due to the larger excursions of the dihedral angle with metadynamics, the unbiased ensemble is a subset of the biased ensemble and the exponential averaging estimator converges quickly in this case.

3.3 Force Field Parametrization

Force field parameters were assigned to the hosts and the guests using an in-house development FFEngine workflow at Roivant Discovery. FFEngine is a workflow for the bespoke parametrization of ligands with the Amber force field functional form.[47] The partial charges were derived from GFN2-xTB/BCC with pre-charges from semi-empirical QM method GFN2-xTB,[48] and bond charge correction (BCC) parameters fitted to the HF/6-31G* electrostatic potential (ESP) with the COSMO implicit solvation model from a 50,000 drug-like compounds dataset. The ESP with an implicit solvation model was deemed necessary for these highly polar and charged host-guest systems even though it is expected to yield a fixed charge model slightly more polarized than the default GAFF2/Amber one.

3.4 Computational Details

The guests were manually docked to the hosts in each binding pose using Maestro (Schrödinger, Inc.) in each of the four binding poses. A flat-bottom harmonic positional restraint with a force constant $k_c = 25$ kcal/mol/Å^2 and a tolerance of 4 Å was applied to define the binding site region.[49, 35] For this purpose, the centroid of the guest was taken as the center of the central ring of the phenothiazine core, and the centroid of the cyclodextrin host was defined as the center of the oxygen atoms forming the ether linkages between the sugar monomers. Boresch-style[50] orientational restraints were imposed to keep each complex in the chosen binding pose. These were implemented as flat-bottom restraints acting on the θ and ϕ angles in Figure 6 with a force constant of $k_a = 0.05$ kcal/mol/deg^2, and centers and tolerances tailored for each pose. For example, the orientational restraints for the 'sp' pose are centered on $\theta = 0°$ and $\phi = 180°$ with $\pm 90°$ tolerances for both. The θ angle is defined as the angle between the z-axis of the host, defined as the axis going through the centroid of the oxygen atoms of the primary hydroxyl groups and the centroid of the oxygen atoms of the secondary hydroxyl groups, and the molecular axis of the guest, defined as the axis going through the S and N atoms of the central ring of the phenothiazine core. The ϕ angle is defined as the dihedral angle between the plane formed by the C1-N-S triplet of atoms of the phenothiazine core of the guest and the plane spanned by the z-axis of the host

and the molecular axis of the guest, where C1 represents the carbon atom of the phenothiazine host with the substituent of the phenothiazine moiety. Very loose flat-bottom harmonic positional restraints (4 Å tolerance and 25 kcal/mol/Å^2 force constant) were applied to the ether linkages oxygen atoms of the cyclodextrins to keep the hosts from wandering freely in the simulation box. The ATM displacement vector was set to $(-30, 0, 0)$ Å .

Force field parameters were assigned as described above. In RBFE calculations, the second ligand in the pair was placed in the solvent by translating it along the displacement vector. The resulting system was then solvated using tleap[51] in a TIP3P box with a 10 Å solvent buffer and sodium and chloride counterions to balance the negative and positive charges, respectively. The systems were energy minimized, thermalized, and relaxed in the NPT ensemble at 300 K and 1 atm pressure. Annealing of the systems to $\lambda = 1/2$ in the nVT ensemble followed to obtain initial structures to initiate the parallel replica exchange ATM calculations. Alchemical calculations were conducted with the OpenMM 7.7 MD engine on the CUDA platform with a 2 fs timestep, using the AToM-OpenMM software.[52] Asynchronous Hamiltonian Replica Exchange[53] in λ space was attempted every 20 ps and binding energy samples were collected at the same frequency. The ATM-RBFE employed 22 λ states distributed between $\lambda = 0$ to $\lambda = 1$ using the non-linear alchemical soft-plus potential and the soft-core perturbation energy with parameters $u_{\max} = 200$ kcal/mol, $u_c = 100$ kcal/mol, and $a = 1/16$.[13] The input files of the simulations are provided in our lab's GitHub repository (https://github.com /GallicchioLab/SAMPL9-bCD). We collected approximately 20 ns trajectories for each replica corresponding to approximately 440 ns for each of the 64 alchemical steps for each host (Figure 8). Overall, we simulated the systems for over 6 μs. UWHAM[54] was used to compute binding free energies and the corresponding uncertainties after discarding the first half of the trajectory.

To obtain the torsional flattening biasing potential, we simulated the PMT guest in solution with metadynamics over the (C-N-C-C) alkylamine sidechain torsional angle highlighted in Figure 10. A well-tempered metadynamics bias factor of 8 was used, with a Gaussian height of 0.3 kcal/mol and width of 10°.[45] The bias potential was collected for 20 ns, updating it every 100 ps. The resulting potential of mean force is shown in Figure 10. The metadynamics-derived biasing potential was used in all the RBFE calculations involving the PMT guest to accelerate the sampling of the slow torsional degree of freedom in question.

4 Results

4.1 Binding Free Energy Predictions

The calculated binding free energies of the cyclodextrin/phenothiazine complexes obtained by combining the pose-specific binding free energies are listed in Table 1 compared to the experimental measurements. We provide the results of each individual free calculation in the Supplementary Information. The second column of Table 1 reports the blinded computational predictions submitted to the SAMPL9 organizers and the results of revised predictions (third column) obtained subsequently to correct setup errors and resolve unconverged calculations. Specifically, we uncovered cases where binding poses were misidentified and where centers of ligands and the hosts had reversed chirality during energy minimizations due to close initial atomic overlaps. As discussed below, in the initial predictions, we were also unable to obtain consistent binding free energy predictions for symmetric poses. In the binding mode analysis reported below we used exclusively data from the corrected molecular simulations.

The predictions for the bCD complexes are in reasonable agreement with the experiments. The revised predictions, in particular, are all within 1.5 kcal/mol of the experimental measurements and within 1 kcal/mol for four of the five bCD complexes. Although the range of the binding affinities is small, some trends are reproduced and the weakest binder (PMT) is correctly identified. The quality of the predictions for the mCD host is not as good, and it worsened upon revision. The experiments show that the phenothiazine guests bind slightly more strongly to mCD than bCD. However, except for CPZ, the calculations predict significantly weaker binding to mCD relative to bCD. The computed free energies of the mCD complexes are on overage over 2 kcal/mol less favorable than the experimental ones. The revised prediction for the mCD-TDZ complex is particularly poor and fails to identify this complex as the most stable in the set. While a detailed investigation of the sources of the poor prediction for mCD has not been carried out, our model could not have identified the best possible binding poses for this more flexible host. mCD is also more hydrophobic and the energy model may overpredict the reorganization free energy to go from the apo to the holo conformational ensemble for this host.

4.2 Binding Mode Analysis

We used the binding mode-specific binding constants we obtained (see Supplementary Information) to infer

Table 1 The binding free energy predictions submitted to the SAMPL9 challenge compared to the revised predictions and the experimental measurements.

Name	ΔG_b(SAMPL9)^{a,b,c}	ΔG_b(ATM)^{a,b,d}	ΔG_b(expt)^{a,e}
bCD-TDZ	-4.28 ± 0.90	-4.56 ± 0.47	-5.73
bCD-TFP	-6.51 ± 1.11	-5.42 ± 0.99	-5.09
bCD-PMZ	-3.73 ± 0.48	-4.03 ± 0.45	-5.00
bCD-PMT	-2.53 ± 0.70	-3.00 ± 0.47	-4.50
bCD-CPZ	-7.28 ± 0.92	-4.64 ± 0.70	-5.45
mCD-TDZ	-5.16 ± 1.40	-2.96 ± 0.68	-6.50
mCD-TFP	-4.14 ± 0.62	-3.98 ± 0.70	-5.57
mCD-PMZ	-2.37 ± 0.54	-2.34 ± 0.55	-5.08
mCD-PMT	-1.80 ± 0.99	-1.58 ± 0.60	-5.39
mCD-CPZ	-5.22 ± 0.90	-5.13 ± 0.88	-5.43

[a]In kcal/mol. [b]Errors are reported as twice the standard deviation. [c]Blinded computational predictions submitted to the SAMPL9 challenge organizers. [d]Revised ATM computational predictions. [e]SAMPL9 blinded experimental isothermal calorimetry data.[25]

the population of each binding mode for each complex shown in Figure 9. The results indicate that the complexes visit many poses with appreciable population. The only exceptions are TFP binding to bCD and CPZ binding to mCD which are predicted to exist with over 75% population in only one pose ('sp' in the R configuration in the case of the TFP-bCD complex and 'sp' in the S configuration in the case of the CPZ-mCD complex). In general, the guests bind the hosts preferentially in the 'sp' and 'ss' modes with the alkylamine sidechain placed near the primary face of the hosts (Figure 4). This trend is less pronounced for the complexes between PMT, PMZ, and TDZ with bCD, which occur in the 'sp'/'ss' and 'pp'/'ps' modes with similar frequency, and it is more pronounced for all complexes with mCD which strongly prefer the alkylamine sidechain towards the primary face. Unlike the alkylamine sidechain, the substituents of the phenothiazine aromatic ring of the CPZ, TDZ, and TFP guests are preferentially placed towards the secondary face of the cyclodextrin hosts. This is evidenced by the higher probability of the 'sp' binding modes (red and green bars in Figure 9) over the 'ss' binding modes (blue and yellow).

Reassuringly, the calculations predict that the populations of the symmetric binding modes of the complexes with the PMZ and PMT guests are more evenly distributed than for the other complexes. Lacking a substituent of the phenothiazine moiety (Figure 1), the PMZ and PMT guests do not display conformational chirality (Figure 3). Hence, their 'ssS', 'spS', 'ssR', and 'spR' binding modes are chemically equivalent and should have the same population. Similarly, the binding modes 'psS', 'ppS', 'psR', and 'ppR' of these guests are mutually equivalent. Still, they are distinguishable from the 'ssS', 'spS', 'ssR', 'spR' group by the position of the alkylamine sidechain (Figure 4). We used these equivalences to assess the level of convergence of the binding free energy estimates. Although redundant for symmetric poses, we simulated each binding mode of these guests individually, starting from different initial configurations, and checked how close the pose-specific binding free energies varied within each symmetric group. For example, the computed populations of the 'ssS', 'spS', 'ssR', and 'spR' poses of the PMZ-bCD complex vary in a narrow range between 7.5 and 15.9%, indicating good convergence. However, the corresponding populations for the complex with mCD are not as consistent–the 'ssS' mode predicted to be significantly less populated (4%) than the other modes (20-40%)–reflecting poorer convergence.

The pose-specific binding free energy estimates probe the chiral binding specificity of the hosts. Except for the TFP guest that is predicted to bind predominantly in the R chiral form (88% population), bCD shows little chiral preference. mCD displays a slightly stronger chiral specificity, with CPZ predicted to bind predominantly in the S form and TFP in the R form.

4.3 Comparison between Predicted Binding Pose Populations and NMR Experiments

In addition to providing the Isothermal Calorimetry (ITC) binding affinity data for the SAMPL9 bCD blinded challenge, Gilson and collaborators probed the conformational propensities of the phenothiazine complexes with bCD and mCD by proton Nuclear Magnetic Resonance (NMR) Nuclear Overhauser Effect (NOE) measurements.[25] Our calculated binding pose population distributions generally agree with the experimental observations.

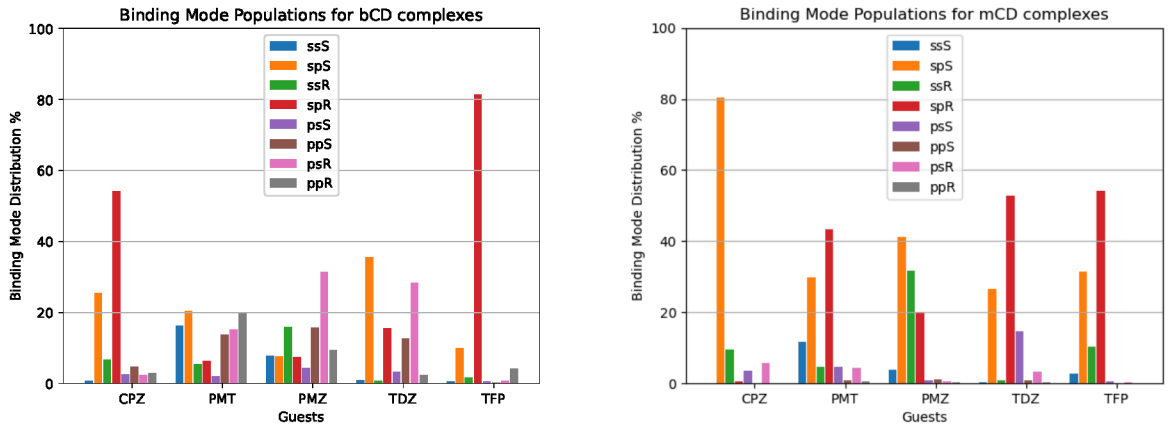

Fig. 9 Binding mode populations of the complexes with bCD (left) and mCD (right).

Consistent with the significant conformational variability that we predicted, only a sparse set of NOEs of the complexes of bCD and mCD with the symmetric guests PMT and PMZ are observed.[25] NOE signals for these complexes include only a few interactions between the phenothiazine moieties and protons at either face of the host, as seen in the molecular simulations. No NOE interactions are observed for the mCD-PMT complex. The high conformational variability of the complexes with the TDZ guest (Figure 9A) is similarly confirmed by the lack of NOEs involving the alkylamine sidechain and the observed NOEs with the phenothiazine core, indicating that it binds in both orientations with the substituted end near both the primary and secondary faces of the host.

In agreement with our predictions (Figure 9A), the observed NOEs indicate binding of the CPZ and TFP guests in a single dominant binding pose, but not always the predicted pose. The set of NOEs of the bCD-CPZ complex,[25] is in agreement with the prediction that the bCD-CPZ complex has an aggregate 'sp' binding pose population of over 80% (Figure 9A). However, the computational prediction that the mCD-CPZ complex exists predominantly in the 'spS' binding pose (Figure 9B) does not appear to be well supported by the experimental NOEs between the secondary face of mCD and the protons of the phenothiazine moiety closest to the substituent.[25] Finally, the predictions that the TFP guest binds the bCD and mCD hosts mainly in the 'sp' pose, in which the alkylamine sidechain is oriented toward the secondary face of the host and the trifluoromethyl group toward the primary face, is not born out in the experimental NOEs that indicate a strong preference for 'ps' and 'pp' poses. The contrast between these structural inconsistencies and the good alignment between the computed and experimental binding free energies for these complexes (Table 1) are potentially a further indication that binding pose populations can be very sensitive to minute shifts in interatomic interaction energies.

4.4 Enhanced Conformational Sampling of the PMT Guest

As discussed above, the 'ssS', 'spS', 'ssR', and 'spR' binding poses of the PMT guest, which lacks phenothiazine substituent, are chemically indistinguishable and should yield equivalent pose-specific ABFE estimates. Similarly, the 'psS', 'ppS', 'psR', and 'ppR' should yield the same binding free energy within statistical uncertainty. Yet, in our first attempt submitted to SAMPL, our predictions did not achieve the expected consistency (Table 2, second column). In Table 2 we show the binding free energy estimates for each PMT binding pose relative to the same pose of PMZ, whose poses are equivalent in the same way as for PMT. For instance, while the four top poses for bCD are expected to yield the same RBFEs, the actual estimates show a scatter of more than 4 kcal/mol. The other groups of equivalent binding poses of bCD and mCD also show significant scatter, indicating poor convergence.

The molecular dynamics trajectories analysis later revealed that the observed scatter of relative binding free energy estimates was due to the conformational trapping of the PMT guest in the starting conformation, which was randomly set during the system setup. Simulations with PMT trapped in some conformations overestimated the RBFE while those in the other underestimated it. We pin-pointed the conformational trapping to the branched alkylamine side chain of PMT which showed hindered rotation around one of its torsional angles (Figure 10) caused by a large free energy

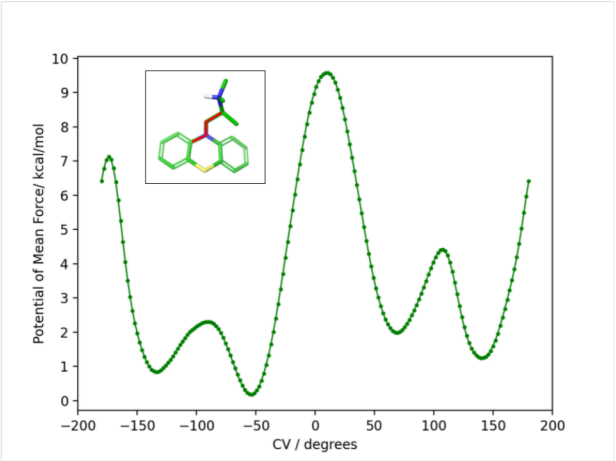

Fig. 10 The potential of mean force (PMF) in water solution along the highlighted torsional angle, φ, of PMT1 computed by well-tempered metadynamics sampling.[45] The PMF identifies two major gauche conformational states at positive and negative angles around 60° and −120° separated by a large free energy barrier at 180° of more than 7 kcal/mol from the global minimum at −50°. The free energy barrier is sufficiently high that interconversions between the two stable conformational states are not observed in the time-scale of the alchemical calculations without the metadynamics landscape-flattening potential.

Table 2 Relative binding free energy estimates of the binding poses of PMT relative to the same binding pose of PMZ for the two cyclodextrin hosts bCD and mCD and with and without metadynamics enhanced sampling.

Pose	$\Delta\Delta G_b$(ATM)^{a,b,c}	$\Delta\Delta G_b$(ATM+MetaD)^{a,b,d}
	bCD	
spS	3.94 ± 0.39	0.44 ± 0.25
ssS	2.80 ± 0.39	0.58 ± 0.25
spR	0.28 ± 0.34	1.12 ± 0.24
ssR	4.62 ± 0.45	1.65 ± 0.25
psS	1.98 ± 0.36	1.49 ± 0.24
ppS	2.03 ± 0.39	1.10 ± 0.24
psR	0.77 ± 0.29	1.46 ± 0.24
ppR	0.28 ± 0.39	0.57 ± 0.24
	mCD	
spS	1.93 ± 0.37	0.96 ± 0.25
ssS	0.57 ± 0.44	0.11 ± 0.26
spR	1.59 ± 0.41	0.30 ± 0.25
ssR	0.25 ± 0.42	1.90 ± 0.25
psS	1.26 ± 0.39	-0.14 ± 0.24
ppS	0.20 ± 0.42	0.95 ± 0.25
psR	1.54 ± 0.41	-0.41 ± 0.25
ppR	0.10 ± 0.38	0.10 ± 0.24

[a]In kcal/mol. [b]Errors are reported as twice the standard deviation. [c]Estimates computational predictions submitted to the SAMPL9 challenge organizers. [d]Revised ATM estimates with metadynamics conformational sampling.

barrier separating the gauche(+) and gauche(-) conformers (Figure 10). The variations of conformers in the alchemical calculation broke the symmetry between equivalent poses and caused the scatter in the observed RBFEs.

To correct these inconsistencies, we modified our alchemical binding protocol to include a metadynamics-derived flattening potential bias that reduced the magnitude of the free energy barrier separating the conformers of the alkylamine sidechain of PMT (see Methods and Computational Details). We confirmed that the biasing potential successfully induced rapid conformational transitions between these conformers, which were rarely achieved with the conventional ATM protocol. Consequently, integrating metadynamics-enhanced sampling with ATM (ATM-MetaD) indeed produced much better convergence of binding free energy estimates of symmetric poses starting from different initial conformers (Table 2). For example, the large discrepancy of RBFE estimates between the 'spS' and 'spR' binding poses was reduced to less than 1 kcal/mol with ATM-metaD and in closer consistency with statistical uncertainties. With only one exception, improved convergence was also achieved for the equivalent binding poses of bCD and mCD, falling within a 1 kcal/mol range of each other (Table 2).

5 Discussion

Molecular binding equilibria are central to applications ranging from pharmaceutical drug design to chemical engineering. Obtaining reliable estimates of binding affinities by molecular modeling is one of the holy grails of computational science. Enabled by recent developments in free energy theories and models and increased computing power, early static structure-based virtual screening tools, such as molecular docking, are increasingly complemented by more rigorous dynamical free energy models of molecular recognition representing the conformational diversity of molecules at atomic resolution. However, many challenges still remain to achieve a sufficient level of usability and performance for free energy models to apply them to chemical research widely. By offering a platform to assess and validate computational models against high-quality experimental datasets in an unbiased fashion, the SAMPL series of blinded challenges have significantly contributed to the advancement of free energy models.[55] By participating in SAMPL challenges we have refined and improved our models against host-guest and protein-ligand datasets[56, 57, 19, 58, 22, 20] and built an appreciation for the complexities of molecular recognition phenomena and the level of detail required to model them accurately.

The present SAMPL9 bCD challenge highlights the importance of properly treating conformational heterogeneity to obtain reliable quantitative descriptions of binding equilibria. We undertook this challenge with the mindset that host-guest systems are simpler surrogates of more challenging and conformationally diverse protein-ligand complexes and, hence, more suitable to assess computational methodologies. However, as later confirmed by the NMR NOE experimental analysis,[25] we realized that the phenothiazine/cyclodextrin complexes could be far more chemically and conformationally diverse than expected. Most of the guests exist in solution as mixtures of enantiomers related by nitrogen inversion (Figure 3) which are distinctly recognized by the chiral hosts. As a result, one enantiomer could be significantly enriched relative to the other when bound to the host. In addition, each enantiomer binds the host in four generally distinct binding orientations that differ in the placement of the alkylamine sidechain and the the substituent of the phenothiazine moiety relative to the host (Figure 4). While in the experimental setting the guests and the complexes rapidly transition from one pose to another, this level of conformational heterogeneity poses serious challenges for standard molecular dynamics conformational sampling algorithms, which are generally limited to one binding pose.

When facing these complexities, it is tempting to limit the modeling to the most important binding pose. While it is true that often the most favorable pose contributes the most to the binding affinity and that neglecting minor poses results in small errors, binding pose selection remains an unresolved issue. Clearly, identifying the major pose cannot be carried out by binding free energy analysis of each pose because that is precisely what one seeks to avoid in such a scenario. Whichever approach is adopted, it must be capable of identifying the most stable pose of each complex among many competing poses. The present results illustrate this challenge. For example, the populations derived from our free energy analysis (9) indicate that the 'spR' binding pose is often one of the most populated (red in Figure 9). However, CPZ is predicted to strongly prefer the 'spS' pose when binding to mCD (orange in Figure 9B), and limiting the modeling to the 'spR' pose would result in a gross underestimation of the binding free energy. Similarly, the TDZ-bCD complex is predicted to exist in a variety of poses (Figure 9A), including, for instance, the 'psR' pose with the alkylamine sidechain pointing towards the primary face of bCD, with the 'spR' pose contributing only a small fraction of the observed binding affinity. Clearly, at least in this case, limiting the modeling to one carefully selected predetermined pose would lead to significant mispredictions for individual complexes.

To obtain an estimate of the observed binding constants, in this work, we opted to compute the binding free energies of all of the relevant binding poses of the system and to integrate them using the free energy combination formula [Eq. (1)]. The combination formula requires the populations of the conformational states of the guest in solution that, in this case, are easily obtained by symmetry arguments. Still, the work involved 64 individual alchemical free energy calculations (Figure 8) and hundreds of nanoseconds of simulation on GPU devices. While we attempted to automate the process as much as possible, setup errors were made and it is likely that some yet undiscovered defects are still affecting our revised results. We assessed the convergence of the pose-specific binding free energy estimates by monitoring the consistency between the results for symmetric poses. As a result of this assessment, we realized that one guest (PMT) was affected by slow conformational reorganization that required metadynamics treatment. This best-effort attempt resulted in good quantitative predictions for the complexes with β-cyclodextrin host. However, our model failed to properly describe the binding free energies of the complexes with the methylated derivative (mCD). Force field limitations that cause excessive reorganization of the host in solution and the existence of alternative binding modes not considered in our analysis are some of the possible explanations of why our free energy predictions consistently underestimated the binding affinities of the complexes with mCD (Table 1).

6 Conclusions

In this work, we describe our effort to obtain alchemical computational estimates of the binding constants of a set of phenothiazine guests to cyclodextrin hosts as part of the SAMPL9 bCD challenge using the Alchemical Transfer Method. The free energy modeling of these systems proved significantly more challenging than expected due to the multiple conformational states of the guests and the multiple binding poses of the complexes which had to be treated individually. Overall, 64 individual alchemical calculations were employed to obtain binding free energy estimates comparable to the experimental observations. The predictions were quantitative for the β-cyclodextrin host but failed to accurately describe the observed binding affinities to the methylated derivative. The work shows that even simple molecular systems can require extensive modeling efforts to treat conformational heterogeneity appropriately and it illustrates the role that multiple binding poses can play in

protein-ligand binding prediction and, ultimately, drug design.

7 Acknowledgements

We acknowledge support from the National Science Foundation (NSF CAREER 1750511 to E.G.). Molecular simulations were conducted on Roivant's computational cluster and on the Expanse GPU supercomputer at the San Diego Supercomputing Center supported by NSF XSEDE award TG-MCB150001. We are grateful to Mike Gilson for providing experimental data for the SAMPL9 bCD challenge. We appreciate the National Institutes of Health for its support of the SAMPL project via R01GM124270 to David L. Mobley.

8 Data Availability

Input files of the AToM-OpenMM simulations are available on the GitHub repository `github.com/GallicchioLab/SAMPL9-bCD`. The AToM-OpenMM software is freely available for download on GitHub.[52] A detailed list of the results and their analysis are provided in the Supplementary Information. Simulation MD trajectories are available from the corresponding author upon reasonable request.

9 Supplementary information

Spreadsheets SAMPL9-bCDpc-FFEngine and SAMPL9-mCDpc-FFEngine containing: (i) the absolute binding free energy of host G1 in the 's' and 'p' binding modes, (ii) the relative binding free energies between G1 in the 'p' and 's' poses and MTZ in the 'ss', 'sp', etc. binding modes, (iii) the relative binding free energy between MTZ and PMZ in each of the eight binding modes, (iv) the relative binding free energies between PMZ and the other guests in each of the eight binding poses, (v) the binding mode specific binding constants for each complex in each binding mode, and (vi) the calculated populations of each binding mode for each complex.

A Proof of Eq. (3)

Consider the potential energy function $U_0(x)$ of the unbound state of the receptor-ligand complex and $U_1(x)$ the one corresponding to the bound state. Here x represents, collectively, the degrees of freedom of the system. The probability that the complex to in binding mode i is

$$P_1(i) = \frac{\int_i e^{-\beta U_1(x)} dx}{\int e^{-\beta U_1(x)} dx} \tag{5}$$

where the denominator is the configurational partition function of the complex in the bound state, and the numerator, where the integration is restricted to regions of conformational space corresponding to binding mode i, is the configurational partition function of binding mode i in the bound state. Next, multiply and divide Eq. (5) by the partition function $\int_i \exp[-\beta U_0(x)]dx$ of binding mode i in the unbound state, noting that:

$$\frac{\int_i e^{-\beta U_1(x)} dx}{\int_i e^{-\beta U_0(x)} dx} = K_b(i) \tag{6}$$

where $K_b(i)$ is the binding mode-specific binding constant.

To obtain an expression for the reminder ratio of the partition function of binding mode i in the unbound ensemble to the partition function of the complex in the bound ensemble, multiply and divide by the partition function of the system in the unbound ensemble $\int \exp[-\beta U_0(x)]dx$ noting that:

$$\frac{\int_i e^{-\beta U_0(x)} dx}{\int e^{-\beta U_0(x)} dx} = P_0(i) \tag{7}$$

where $P_0(i)$ is the population of binding mode i in the unbound ensemble and

$$\frac{\int e^{-\beta U_1(x)} dx}{\int e^{-\beta U_0(x)} dx} = K_b \tag{8}$$

is the overall binding constant. Collecting the terms above yields (3).

References

1. William L Jorgensen. The many roles of computation in drug discovery. *Science*, 303:1813–1818, 2004.
2. Zoe Cournia, Bryce K Allen, Thijs Beuming, David A Pearlman, Brian K Radak, and Woody Sherman. Rigorous free energy simulations in virtual screening. *J. Chem. Inf. Model.*, 60(9):4153–4169, 2020.
3. Charles D Griego, John R Kitchin, and John A Keith. Acceleration of catalyst discovery with easy, fast, and reproducible computational alchemy. *Int. J. Quantum Chem.*, 121(1):e26380, 2021.
4. Huafeng Xu. The slow but steady rise of binding free energy calculations in drug discovery. *J. Comp.-Aid. Mol. Des.*, pages 1–8, 2022.
5. Robert Abel, Lingle Wang, Edward D Harder, BJ Berne, and Richard A Friesner. Advancing drug discovery through enhanced free energy calculations. *Acc. Chem. Res.*, 50(7):1625–1632, 2017.
6. Abir Ganguly, Hsu-Chun Tsai, Mario Fernández-Pendás, Tai-Sung Lee, Timothy J Giese, and Darrin M York. Amber drug discovery boost tools: Automated workflow for production free-energy simulation setup and analysis (professa). *Journal of Chemical Information and Modeling*, 62(23):6069–6083, 2022.
7. Zoe Cournia, Bryce Allen, and Woody Sherman. Relative binding free energy calculations in drug discovery: recent advances and practical considerations. *J. Chem. Inf. Model.*, 57(12):2911–2937, 2017.
8. Kira A Armacost, Sereina Riniker, and Zoe Cournia. Novel directions in free energy methods and applications, 2020.
9. Christina EM Schindler, Hannah Baumann, Andreas Blum, Dietrich Böse, Hans-Peter Buchstaller, Lars Burgdorf, Daniel Cappel, Eugene Chekler, Paul

Czodrowski, Dieter Dorsch, Eguida Merveille K I, Bruce Follows, Thomas Fuchs, Ulrich Grädler, Jakub Gunera, Theresa Johnson, Catherine Jorand Lebrun, Srinivasa Karra, Markus Klein, Tim Knehans, Lisa Koetzner, Mireille Krier, Matthias Leiendecker, Birgitta Leuthner, Liwei Li, Igor Mochalkin, Djordje Musil, Constantin Neagu, Friedrich Rippmann, Kai Schiemann, Robert Schulz, Thomas Steinbrecher, Eva-Maria Tanzer, Andrea Unzue Lopez, Ariele Viacava Follis, Ansgar Wegener, and Daniel Kuhn. Large-scale assessment of binding free energy calculations in active drug discovery projects. *J. Chem. Inf. Model.*, 60(11):5457–5474, 2020.

10. Antonia S. J. S. Mey, Bryce K. Allen, Hannah E. Bruce Macdonald, John D. Chodera, David F. Hahn, Maximilian Kuhn, Julien Michel, David L. Mobley, Levi N. Naden, Samarjeet Prasad, Andrea Rizzi, Jenke Scheen, Michael R. Shirts, Gary Tresadern, and Huafeng Xu. Best practices for alchemical free energy calculations [article v1.0]. *Living Journal of Computational Molecular Science*, 2(1):18378, 12 2020.
11. Tai-Sung Lee, Bryce K Allen, Timothy J Giese, Zhenyu Guo, Pengfei Li, Charles Lin, T Dwight McGee Jr, David A Pearlman, Brian K Radak, Yujun Tao, Hsu-Chun Tsai, Huafeng Xu, Woody Sherman, and Darrin M York. Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery. *J. Chem. Inf. Model.*, 60(11):5595–5623, 2020.
12. Marina Macchiagodena, Marco Pagliai, Maurice Karrenbrock, Guido Guarnieri, Francesco Iannone, and Piero Procacci. Virtual double-system single-box: A nonequilibrium alchemical technique for absolute binding free energy calculations: Application to ligands of the sars-cov-2 main protease. *J. Chem. Theory Comput.*, 16(11):7160–7172, 2020.
13. Sheenam Khuttan, Solmaz Azimi, Joe Z Wu, and Emilio Gallicchio. Alchemical transformations for concerted hydration free energy estimation with explicit solvation. *J. Chem. Phys.*, 154(5):054103, 2021.
14. Yudong Qiu, Daniel GA Smith, Simon Boothroyd, Hyesu Jang, David F Hahn, Jeffrey Wagner, Caitlin C Bannan, Trevor Gokey, Victoria T Lim, Chaya D Stern, Andrea Rizzi, Bryon Tjanaka, Gary Tresadern, Xavier Lucas, Michael R. Shirts, Michael K. Gilson, John D. Chodera, Christopher I. Bayly, David L. Mobley, and Lee-Ping Wang. Development and benchmarking of open force field v1. 0.0—the Parsley small-molecule force field. *J. Chem. Theory Comput.*, 17(10):6262–6280, 2021.
15. David L Mobley. Lets get honest about sampling. *J. Comp. Aided Mol. Des.*, 26:93–95, 2012.
16. Lingle Wang, Yujie Wu, Yuqing Deng, Byungchan Kim, Levi Pierce, Goran Krilov, Dmitry Lupyan, Shaughnessy Robinson, Markus K Dahlgren, Jeremy Greenwood, Donna L. Romero, Craig Mass, L. Jennifer Knight, Thomas Steinbrecher, Thijs Beuming, Wolfgang Damm, Ed Harder, Woody Sherman, Mark Brewer, Ron Wester, Mark Murcho, Leah Frye, Ramy Farid, Teng Lin, David L. Mobley, William L. Jorgensen, Bruce J. Berne, Richard A. Friesner, and Robert Abel. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.*, 137:2695–2703, 2015.
17. Matthew T Geballe, A Geoffrey Skillman, Anthony Nicholls, J Peter Guthrie, and Peter J Taylor. The SAMPL2 blind prediction challenge: introduction and overview. *J. Comp. Aided Mol. Des.*, 24(4):259–279, 2010.
18. David L Mobley, Shuai Liu, Nathan M Lim, Karisa L Wymer, Alexander L Perryman, Stefano Forli, Nanjie Deng, Justin Su, Kim Branson, and Arthur J Olson. Blind prediction of hiv integrase binding from the SAMPL4 challenge. *J. Comp. Aided Mol. Des.*, pages 1–19, 2014.
19. Emilio Gallicchio, Nanjie Deng, Peng He, Alexander L. Perryman, Daniel N. Santiago, Stefano Forli, Arthur J. Olson, and Ronald M. Levy. Virtual screening of integrase inhibitors by large scale binding free energy calculations: the SAMPL4 challenge. *J. Comput.-Aided Mol. Des.*, 28:475–490, 2014.
20. Solmaz Azimi, Joe Z Wu, Sheenam Khuttan, Tom Kurtzman, Nanjie Deng, and Emilio Gallicchio. Application of the alchemical transfer and potential of mean force methods to the sampl8 host-guest blinded challenge. *J. Comp. Aid. Mol. Des.*, 36(1):63–76, 2022.
21. Martin Amezcua, Jeffry Setiadi, Yunhui Ge, and David L Mobley. An overview of the SAMPL8 host–guest binding challenge. *J. Comp.-Aided Mol. Des.*, 36(10):707–734, 2022.
22. Rajat Kumar Pal, Kamran Haider, Divya Kaur, William Flynn, Junchao Xia, Ronald M. Levy, Tetiana Taran, Lauren Wickstrom, Tom Kurtzman, and Emilio Gallicchio. A combined treatment of hydration and dynamical effects for the modeling of host-guest binding thermodynamics: The SAMPL5 blinded challenge. *J. Comp. Aided Mol. Des.*, 31:29–44, 2016.
23. Jian Yin, Niel M Henriksen, David R Slochower, Michael R Shirts, Michael W Chiu, David L Mobley, and Michael K Gilson. Overview of the sampl5 host–guest challenge: Are we doing better? *J. Comp.-Aided Mol. Des.*, 31:1–19, 2017.
24. Andrés Guerrero-Martínez, Teresa Montoro, Montserrat H Viñas, and Gloria Tardajos. Complexation and chiral drug recognition of an amphiphilic phenothiazine derivative with β-cyclodextrin. *J. Pharm. Sci.*, 97(4):1484–1498, 2008.
25. Brenda Andrade, Ashley Chen, and Michael K Gilson. Binding of phenothiazine drugs to heptakis-methylated β-cyclodextrin derivatives: Thermodynamics and structure. *Phys. Chem. Chem. Phys.*, In press., 2023.
26. The SAMPL9 blind prediction challenges for computational chemistry.
27. Gary L Bertrand, James R Faulkner Jr, Soon M Han, and Daniel W Armstrong. Substituent effects on the binding of phenols to cyclodextrins in aqueous solution. *The Journal of Physical Chemistry*, 93(18):6863–6867, 1989.
28. Wei Chen, Chia-En Chang, and Michael K Gilson. Calculation of cyclodextrin binding affinities: energy, entropy, and implications for drug design. *Biophysical journal*, 87(5):3035–3049, 2004.
29. Lauren Wickstrom, Peng He, Emilio Gallicchio, and Ronald M. Levy. Large scale affinity calculations of cyclodextrin host-guest complexes: Understanding the role of reorganization in the molecular recognition process. *J. Chem. Theory Comput.*, 9:3136–3150, 2013.
30. Niel M Henriksen and Michael K Gilson. Evaluating force field performance in thermodynamic calculations of cyclodextrin host–guest binding: Water models, partial charges, and host force field parameters. *Journal of chemical theory and computation*, 13(9):4253–4269, 2017.
31. Peng He, Sheila Sarkar, Emilio Gallicchio, Tom Kurtzman, and Lauren Wickstrom. Role of displacing confined solvent in the conformational equilibrium of β-

cyclodextrin. *J. Phys. Chem. B*, 123(40):8378–8386, 2019.
32. Andrea Rizzi, Travis Jensen, David R Slochower, Matteo Aldeghi, Vytautas Gapsys, Dimitris Ntekoumes, Stefano Bosisio, Michail Papadourakis, Niel M Henriksen, Bert L De Groot, Zoe Cournia, Alex Dickson, Julien Michel, Michael K. Gilson, Michael R. Shirts, David L. Mobley, and John D. Chodera. The SAMPL6 SAMPLing challenge: Assessing the reliability and efficiency of binding free energy calculations. *J. Comp. Aid. Mol. Des.*, pages 1–33, 2020.
33. Joe Z Wu, Solmaz Azimi, Sheenam Khuttan, Nanjie Deng, and Emilio Gallicchio. Alchemical transfer approach to absolute binding free energy estimation. *Journal of Chemical Theory and Computation*, 17(6):3309–3319, 2021.
34. Solmaz Azimi, Sheenam Khuttan, Joe Z Wu, Rajat K Pal, and Emilio Gallicchio. Relative binding free energy calculations for ligands with diverse scaffolds with the alchemical transfer method. *J. Chem. Inf. Model.*, 62(2):309–323, 2022.
35. Emilio Gallicchio and Ronald M Levy. Recent theoretical and computational advances for modeling protein-ligand binding affinities. *Adv. Prot. Chem. Struct. Biol.*, 85:27–80, 2011.
36. Giovanni Bussi, Alessandro Laio, and Pratyush Tiwary. Metadynamics: A unified framework for accelerating rare events and sampling thermodynamics and kinetics. *Handbook of Materials Modeling: Methods: Theory and Modeling*, pages 1–31, 2018.
37. Peter Eastman, Jason Swails, John D Chodera, Robert T McGibbon, Yutong Zhao, Kyle A Beauchamp, Lee-Ping Wang, Andrew C Simmonett, Matthew P Harrigan, Chaya D Stern, Rafal P Wiewiora, Bernard R Brooks, and Vijay S Pande. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.*, 13(7):e1005659, 2017.
38. Svein G Dahl, Edward Hough, and Petter-Arnt Hals. Phenothiazine drugs and metabolites: molecular conformation and dopaminergic, alpha adrenergic and muscarinic cholinergic receptor binding. *Biochemical pharmacology*, 35(8):1263–1269, 1986.
39. Mikhail Rekharsky and Yoshihisa Inoue. Chiral recognition thermodynamics of β-cyclodextrin: The thermodynamic origin of enantioselectivity and the enthalpy-entropy compensation effect. *J. Am. Chem. Soc.*, 122(18):4418–4435, 2000.
40. Cassandre Quinton, Lambert Sicard, Nicolas Vanthuyne, Olivier Jeannin, and Cyril Poriel. Confining nitrogen inversion to yield enantiopure quinolino [3, 2, 1-k] phenothiazine derivatives. *Advanced Functional Materials*, 28(39):1803140, 2018.
41. Guha Jayachandran, Michael R Shirts, Sanghyun Park, and Vijay S Pande. Parallelized-over-parts computation of absolute binding free energy with docking and molecular dynamics. *J. Chem. Phys.*, 125:084901, 2006.
42. Lauren Wickstrom, Nanjie Deng, Peng He, Ahmet Mentes, Crystal Nguyen, Michael K Gilson, Tom Kurtzman, Emilio Gallicchio, and Ronald M Levy. Parameterization of an effective potential for protein–ligand binding from host–guest affinity data. *J. Mol. Recognition*, 29(1):10–21, 2016.
43. Martin Amezcua, Léa El Khoury, and David L Mobley. SAMPL7 host–guest challenge overview: assessing the reliability of polarizable and non-polarizable methods for binding free energy calculations. *J. Comp.-Aid. Mol. Des.*, 35(1):1–35, 2021.
44. John C Shelley, Anuradha Cholleti, Leah L Frye, Jeremy R Greenwood, Mathew R Timlin, and Makoto Uchimaya. Epik: a software program for pk a prediction and protonation state generation for drug-like molecules. *Journal of computer-aided molecular design*, 21(12):681–691, 2007.
45. Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100(2):020603, 2008.
46. Phillip S Hudson, H Lee Woodcock, and Stefan Boresch. Use of interaction energies in qm/mm free energy simulations. *J. Chem. Theory Comput.*, 15(8):4632–4645, 2019.
47. Junmei Wang, Wei Wang, Peter A Kollman, and David A Case. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.*, 25(2):247–260, 2006.
48. Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.*, 15(3):1652–1671, 2019.
49. M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.*, 72:1047–1069, 1997.
50. S Boresch, F Tettinger, M Leitgeb, and M Karplus. Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. B*, 107:9535–9551, 2003.
51. D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York, S. Zhao, and P.A. Kollman. Amber 2019, 2019.
52. AToM-OpenMM. `https://github.com/Gallicchio-Lab/AToM-OpenMM`, 2022.
53. Emilio Gallicchio, Junchao Xia, William F Flynn, Baofeng Zhang, Sade Samlalsingh, Ahmet Mentes, and Ronald M Levy. Asynchronous replica exchange software for grid and heterogeneous computing. *Comput. Phys. Commun.*, 196:236–246, 2015.
54. Zhiqiang Tan, Emilio Gallicchio, Mauro Lapelosa, and Ronald M. Levy. Theory of binless multi-state free energy estimation with applications to protein-ligand binding. *J. Chem. Phys.*, 136:144102, 2012.
55. David L Mobley and Michael K Gilson. Predicting binding free energies: frontiers and benchmarks. *Ann. Rev. Bioph.*, 46:531–558, 2017.
56. E. Gallicchio and R. M. Levy. Prediction of SAMPL3 host-guest affinities with the binding energy distribution analysis method (BEDAM). *J. Comp. Aided Mol. De-*

sign., 25:505–516, 2012.
57. Emilio Gallicchio, Haoyuan Chen, He Chen, Michael Fitzgerald, Yang Gao, Peng He, Malathi Kalyanikar, Chuan Kao, Beidi Lu, Yijie Niu, Manasi Pethe, Jie Zhu, and Ronald M Levy. BEDAM binding free energy predictions for the SAMPL4 octa-acid host challenge. *J. Comp. Aided Mol. Des.*, 29:315–325, 2015.
58. Nanjie Deng, William F Flynn, Junchao Xia, RSK Vijayan, Baofeng Zhang, Peng He, Ahmet Mentes, Emilio Gallicchio, and Ronald M Levy. Large scale free energy calculations for blind predictions of protein–ligand binding: the d3r grand challenge 2015. *J. Comp.-Aided Mol. Des.*, 30(9):743–751, 2016.