

# PopShift: A Thermodynamically Sound Approach to Estimate Binding Free Energies by Accounting for Ligand-Induced Population Shifts from a Ligand-Free Markov State Model

Louis G. Smith,<sup>⊥</sup> Borna Novak,<sup>⊥</sup> Meghan Osato, David L. Mobley, and Gregory R. Bowman<sup>\*</sup>



Cite This: *J. Chem. Theory Comput.* 2024, 20, 1036–1050



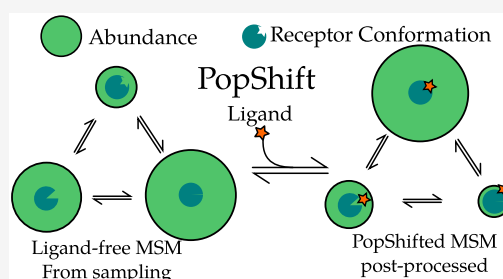
Read Online

ACCESS |

Metrics & More

Article Recommendations

**ABSTRACT:** Obtaining accurate binding free energies from *in silico* screens has been a long-standing goal for the computational chemistry community. However, accuracy and computational cost are at odds with one another, limiting the utility of methods that perform this type of calculation. Many methods achieve massive scale by explicitly or implicitly assuming that the target protein adopts a single structure, or undergoes limited fluctuations around that structure, to minimize computational cost. Others simulate each protein–ligand complex of interest, accepting lower throughput in exchange for better predictions of binding affinities. Here, we present the PopShift framework for accounting for the ensemble of structures a protein adopts and their relative probabilities. Protein degrees of freedom are enumerated once, and then arbitrarily many molecules can be screened against this ensemble. Specifically, we use Markov state models (MSMs) as a compressed representation of a protein’s thermodynamic ensemble. We start with a ligand-free MSM and then calculate how addition of a ligand shifts the populations of each protein conformational state based on the strength of the interaction between that protein conformation and the ligand. In this work we use docking to estimate the affinity between a given protein structure and ligand, but any estimator of binding affinities could be used in the PopShift framework. We test PopShift on the classic benchmark pocket T4 Lysozyme L99A. We find that PopShift is more accurate than common strategies, such as docking to a single structure and traditional ensemble docking—producing results that compare favorably with alchemical binding free energy calculations in terms of RMSE but not correlation—and may have a more favorable computational cost profile in some applications. In addition to predicting binding free energies and ligand poses, PopShift also provides insight into how the probability of different protein structures is shifted upon addition of various concentrations of ligand, providing a platform for predicting affinities and allosteric effects of ligand binding. Therefore, we expect PopShift will be valuable for hit finding and for providing insight into phenomena like allostery.



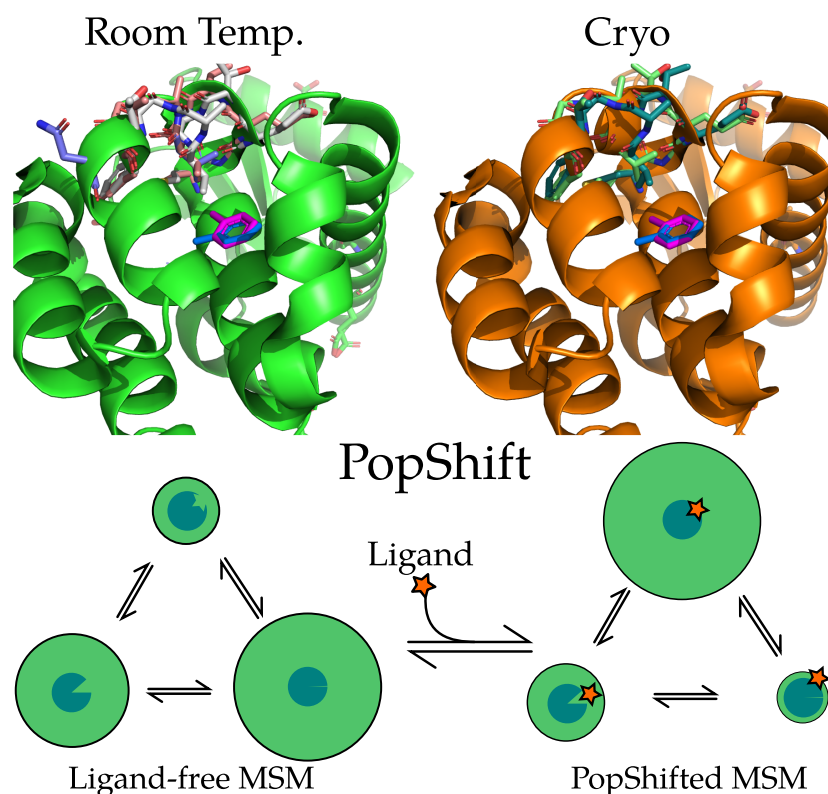
## I. INTRODUCTION

Developing strategies to accelerate and simplify hit discovery in drug development is one of the core foci of computational chemistry. Because huge arcs of chemical space must be subtended, methods that scale well per ligand predominate.<sup>1</sup> Most of these methods are based on docking a set of compounds to a single protein structure as rapidly as possible to maximize the chemical space that can be considered. The scores predicted by these methods correlate so poorly with true binding affinities that they are typically judged by how much the high scoring compounds are enriched for tight binders compared to randomly selected compounds.<sup>2,3</sup> Of course, a wide range of methods have been developed to make different trade-offs between speed and accuracy. Of these, alchemical free energy calculations are some of the most physically rigorous and should, in principle, be capable of quantitatively accurate predictions.<sup>4,5</sup> However, routinely achieving quantitative predictions with any method remains difficult.<sup>6</sup>

One striking feature of all of these methods is the extent to which they assume proteins adopt a limited set of highly similar structures. Many docking algorithms do not include any protein conformational heterogeneity. The cross docking problem highlights the limitations this assumption imposes (i.e., docking a library of compounds against a protein structure obtained by removing a ligand from a ligand-bound structure is more predictive than docking against a structure obtained in the absence of ligand).<sup>7</sup> To address this, some docking algorithms allow limited protein flexibility such as rotations of side-chains. However, many do not find that incorporating conformational

**Received:** August 8, 2023  
**Revised:** November 28, 2023  
**Accepted:** November 29, 2023  
**Published:** January 31, 2024





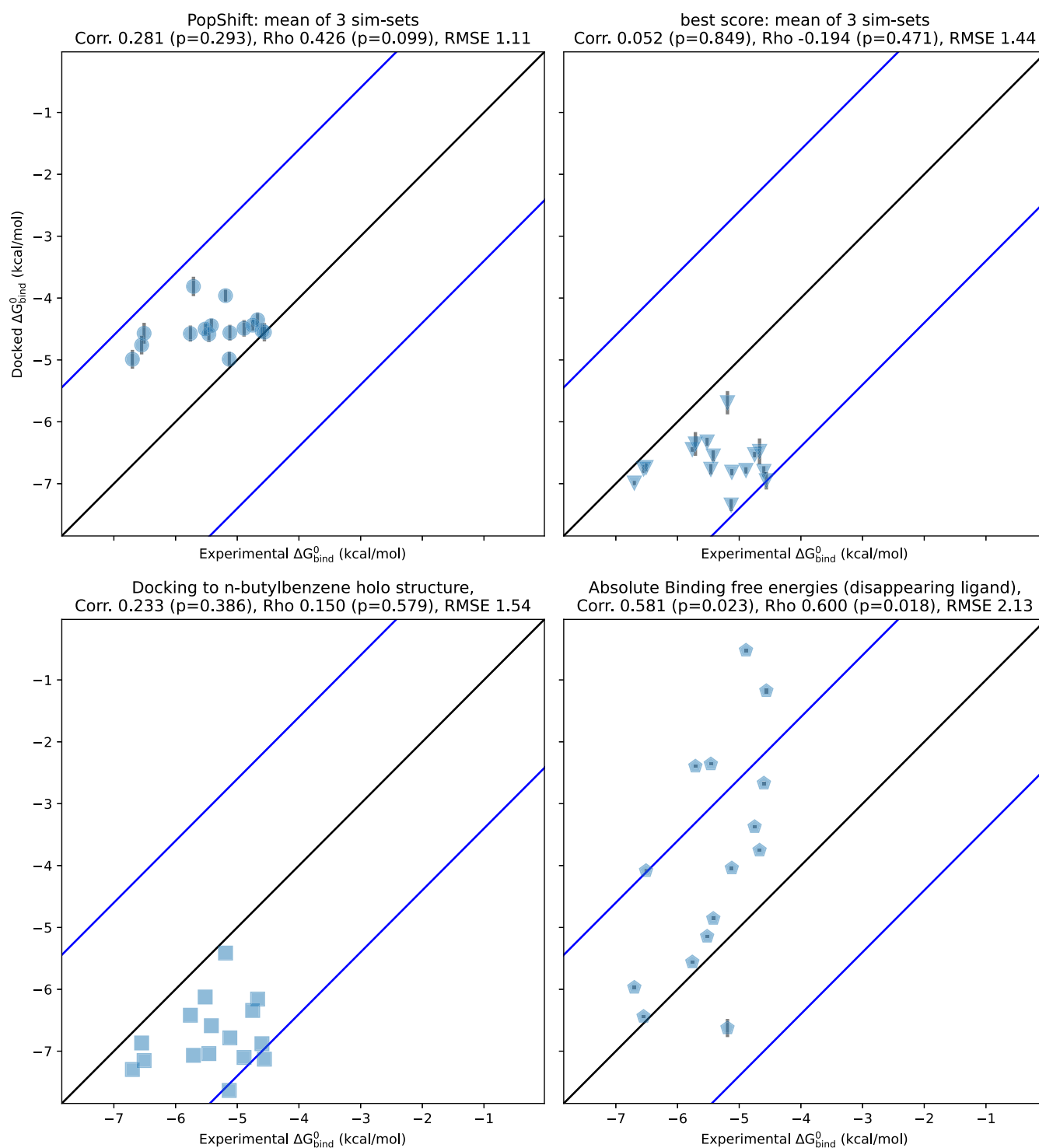
**Figure 1.** Examples of conformational heterogeneity in the T4 lysozyme and a schematic of how PopShift accounts for this heterogeneity. These renders show the multiple conformations even the L99A pocket bound to toluene is capable of accessing under crystallographic study. The top section shows the room temperature structure (PDB 7L39) and a cryogenic structure from the same study (PDB 7L3A). All residues with alternative locations in the F-helix, and also toluene, are shown in sticks. Extensive alternative locations are present in both, even though this protein is reckoned to be rigid and to bind simple, largely rigid, fragments. Note that the two alternative locations for the ligand are nearly identical at both temperatures. Nearly every residue in the critical F-helix shows heterogeneity, centered on valine 111, which extends down toward the toluene. The lower panel shows a schematic of the PopShift method, showing MSM populations from a ligand free ensemble being biased by varying degrees of ligand affinity to those states to approximate the ligand-bound ensemble. The sea-blue pac-man represents the protein, with three states in equilibrium, green circle sizes indicating relative abundance, and the shape cut out of the pac-man representing varying degrees of pocket accessibility to the ligand, which is schematically represented by a star.

heterogeneity in this way is worth the additional computational cost.<sup>8</sup> In principle, alchemical free energy calculations should be able to deal with protein conformational heterogeneity, as every degree of freedom is allowed to move as dictated by the force field. However, in practice, alchemical free energy simulations are so short that the protein only undergoes limited fluctuations around the starting structure.<sup>9</sup> Phrased differently, ignoring receptor conformational heterogeneity for the sake of computational performance is one of the key approximations of most digital screening campaigns.

Ensemble docking has emerged as a strategy to address protein conformational heterogeneity but still faces significant limitations.<sup>10</sup> In ensemble docking, one generates a set of protein structures (often via molecular dynamics simulations) and then docks a library of compounds against each of these structures. Typically, one then ranks the compounds based on their best score against any protein structure, though there are other flavors of ensemble docking. While this ensemble docking approach recognizes that there is uncertainty in which protein structure is relevant, it still essentially assumes that a single structure is relevant in the end. It also throws out thermodynamic information from the simulations, instead giving all of the protein structures equal weight. These methods are still generally incapable of quantitative predictions and suffer from some strange pathologies. For example, it has been

reported that ensemble docking against short simulations outperforms docking to a single structure but that adding more simulation data often hurts performance rather than helping.<sup>10–13</sup> Other efforts to include conformational heterogeneity into docking have included the existence of multiple conformations using some assessment of their relative abundance, but have done so in an ad-hoc fashion.<sup>14,15</sup>

Here, we propose a reweighting approach called PopShift, schematically depicted in Figure 1, that uses Markov state models (MSMs) of a ligand-free protein to account for the populations of different protein structures and how they are shifted upon binding to a ligand. PopShift builds on the numerous successes of simulations of apo proteins in capturing rare conformational changes. For example, we have predicted and experimentally confirmed cryptic pockets formed by motions ranging from side chain displacements to displacements of entire secondary structural elements.<sup>14,16–18</sup> MSMs can be viewed as a compressed representation of the system's thermodynamic ensemble.<sup>19</sup> Thus, MSMs representing the ligand-free protein ensemble contain all the receptor information needed to estimate ligand binding.<sup>20</sup> In order to weight the contribution of state populations in the apo context versus how tightly they bind ligand, we estimate binding to representative conformations from each state, obtaining a per-state binding free energy by averaging them. We then combine these using an



**Figure 2.** PopShift compares well to alternative predictors, such as docking to a single crystal structure, traditional “best score” ensemble docking, and alchemical free energy calculations. The x values are the binding free energy estimated by each in silico method given in the panel title. Correlation is the Pearson’s correlation coefficient, rho is Spearman’s ranking coefficient, and RMSE is the root-mean-squared error in kcal/mol. The error bars on the top two panels are the standard uncertainty in the mean across the three replica data sets. The error bars on the ABFE points are those provided by alchemlyb/pymbar. Black and blue lines are visual guides and represent the 1–1 line and  $2k_B T$  deviated values, respectively.

“over parts” approach analogous to that used in Jayachandran, Shirts, Park, and Pande<sup>21</sup> to combine these per-state affinities with weights from a ligand free MSM. Thus, instead of taking the best score against any structure, as in traditional ensemble docking, we take a correctly weighted average over all structural states. This is made tractable by the MSM, since it means we

only need one or, to be more confident, a handful of binding estimates per MSM state. A related idea that treats affinity per conformation using the same math, but does not leverage an MSM to index conformational heterogeneity, is the Implicit Ligand Theory.<sup>22,23</sup>

By capturing how ligands shift the relative probabilities of different protein conformations, PopShift also provides an opportunity to understand how ligands remodel their binding sites or even allosterically impact distant sites. Population-shift in response to ligand binding is, by definition, allostery. If the ligand-free MSM is a compressed representation of the perturbed model's ensemble, then the reweighted MSM is a compressed representation of the receptor's liganded ensemble. Thus, observables of interest can be estimated with the reweighted state probabilities to understand the allosteric mechanism. Because the expression for reweighted state probabilities includes ligand concentration, the impact on these averages can also be used to compute an EC50.

To test PopShift, we compare its performance to several other candidates on a simple-yet-subtle benchmark for protein–ligand binding, T4 Lysozyme L99A. Because this system has received extensive study over the years the field has accepted it as a first-pass benchmark for many computational protein–ligand binding methods, despite certain drawbacks (notably the poor dynamic range inherent in the ligands that have actually received ITC affinity validation).<sup>23–25</sup> Indeed the literature on this receptor is deep, with quality work on the heterogeneity of the receptor and the impact this has on binding reiterated especially in the last 10 years by the Minh, Mobley, Fischer, and Shoichet groups, among others.<sup>1,23,26,27</sup>

From a set of apo simulations we build an MSM. We then sample conformations from each MSM state, and dock to them, using the customary organic fragments from Morton, Baase, and Matthews.<sup>28</sup> We use the docking score as a heuristic for the free energy of binding to a particular conformation. We recognize that docking has severe limitations, especially for ligands with rotatable bonds. However, docking provides a simple and highly relevant starting point given its widespread use in drug discovery, and the fragments we consider here are not subject to the known issues with rotatable bonds. In the future, it will be interesting to try alternatives to docking in the PopShift framework. In PopShift, the free energies of binding are then incorporated into an affinity estimate and reweighted state probabilities. We compare these strategies to best-score docking, docking to holo crystal structures with the ligand removed, and absolute binding free energy calculations performed in the customary style with docked and hand-adjusted starting poses. We also explore how the conformational preferences of the protein are altered by the addition of ligand.

## II. RESULTS AND DISCUSSION

**II.A. PopShift Performs Well Compared to Alternative In Silico Estimators of Binding Free Energies.** We reasoned that modern simulations are sufficiently predictive that both structures from these simulations and their populations can inform a successful hit finding strategy.<sup>29</sup> In particular, MSMs provide a powerful and quantitatively predictive map of a protein's conformational ensemble and therefore approximate its partition function. Thus, we hypothesized that using the populations from an MSM in an “over parts”<sup>21,30</sup> approach—with the MSM supplying the state populations—would allow correct incorporation of docking scores from across this sample.

To test this hypothesis, we collected three replica simulation data sets of L99A, made MSMs from them, and estimated binding affinities using PopShift and other popular alternatives. Each replica consisted of 10 simulations,  $5 \times 4 \mu\text{s}$  and  $5 \times 8 \mu\text{s}$ , started from PDB 187L with the ligand (*p*-xylene) removed. One MSM was made for each replica data set using TICA<sup>31</sup> on

the pocket residue backbone and side chain torsions, and VAMP-2<sup>32</sup> to validate the number of clusters for *k*-means as has been done in Meller, Lotthammer, Smith, Novak, Lee, Kuhn, Greenberg, Leinwand, Greenberg, and Bowman.<sup>18</sup> We docked ligands from the classic Morton, Baase, and Matthews<sup>28</sup> set against structures from each MSM state using the SMINA docking algorithm.<sup>33</sup> Macroscopic binding affinities were estimated using the PopShift framework; see methods (Section IV). For comparison to extant approaches, we used the conventional ensemble docking approach of taking the best score across a set of samples, and of docking compounds of interest to a holo crystal structure with ligand removed. We also performed absolute binding free energy simulations using a vanishing ligand transformation from initial hand-selected ligand poses.

We find that PopShift performs well compared to alternative docking approaches and even showed some advantages compared to alchemical free energy calculations (Figure 2). Docking each ligand to a single holo structure (the *n*-butylbenzene structure, PDB 186L) and then taking the minimized score as an affinity estimate gives a poor correlation to experimentally measured binding affinities and poor accuracy, as measured by the root mean squared error (RMSE) from experimental results. PopShift also outperforms simply taking the best score, a traditional ensemble docking approach, most notably for ranking. The best-score approach systematically predicts affinities that are too favorable. Although using docking to a ligand-removed crystal structure's scores as affinity estimates exhibited slightly better correlation with experiment on this data set than the best-score approach, these scores emanating from a lone structure exhibited a similar pattern of overly favorable affinity estimates.

Absolute binding free energy estimates produced the strongest correlation and ranking results but struggled with accuracy. This is likely related to initial poses and receptor conformations failing to relax fully in the course of the windowed simulations. This interpretation is complicated by our use of the docking energy function to make affinity estimates, which is very different from the force field we used to obtain the MSM we dock to. Because docking to many samples from an MSM allows us to estimate affinity to many receptor conformations, it sidesteps the issue of having to choose a “most relevant” one to start from. This is especially important if there may be multiple thermodynamically relevant poses for the ligand.<sup>27</sup>

The dynamic range of our reference data presented in Figure 2 is a concern. Given the lower RMSE for PopShift and the small dynamic range of the experimental data, it is not surprising that the RMSE when using the average PopShift prediction as an estimate for each ligand is also low (1.12). However, the value of Spearman- $\rho$  for PopShift's predictions suggests the method has ranking power, whereas guessing a single value for each ligand would give a  $\rho$  of zero, which is essentially what we see for the best score and holo crystal docking approaches.

Making predictions for other ligands also confirms that PopShift is not merely predicting a constant plus noise that happens to be in the right ballpark for every ligand. If experimentally measured affinities were available, we would have added these ligands ITC data taken at a particular temperature is limited for this ligand—the largest collection of these coming from the data set presented in Figure 5, from Morton, Baase, and Matthews.<sup>28</sup> The change in melting temperature induced by a broader array of ligands has been collected from several other studies by Xie, Nguyen, and Minh.<sup>23</sup>



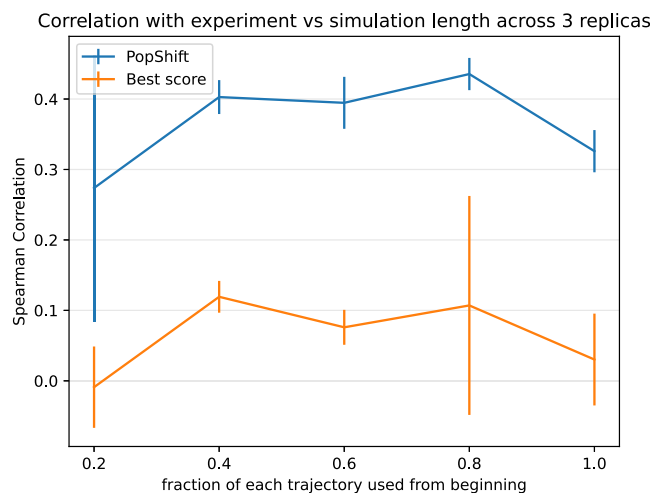
While we cannot directly compare our predicted affinities to these changes in melting temperature, we did predict affinities for this set of ligands and obtained predictions ranging from  $-5.45$  to  $-1.15$  kcal/mol. These results demonstrate that PopShift does not merely predict the same result for every ligand.

As a further test of whether PopShift is appropriately averaging the per-state estimates of binding free energies, we also tried two additional approaches: (1) the population weighted average docking score ( $\Delta G = \sum_i \pi_i \Delta G_i$ , which we have previously called Boltzmann docking<sup>14</sup>), and (2) using eq 2, but with every state given equal probability instead of using the populations from our MSMs. The first reduction of the docking scores results in absurd values with an RMSE of 6.8 kcal/mol. This is because most states yield a positive docking score, reflecting the abundance of closed states in the apo ensemble. If one fails to account for the fact that binding makes these states exceedingly rare, then these rare complexes unrealistically dominate the prediction. The second reduction yields an RMSE of 1.40 kcal/mol and ablates the correlation to 0.29 with a p-value of 0.28, demonstrating that the equilibrium probabilities from our MSM contribute to the predictive power of PopShift.

Application of PopShift to estimate the affinity of an inhibitor called blebbistatin to a set of different myosin motor domains also supports our claim that the method is predictive.<sup>18</sup> Our estimates of the affinity of that ligand to isoforms of myosin was between  $-6$  and  $-9$  kcal/mol, with an R2 of 0.82 to previous experimental measurements. PopShift also successfully predicted the result of an experiment that we performed after making our prediction. These results again show that PopShift is capable of covering a larger dynamic range than for the small set of lysozyme ligands we focus on in this work and has the potential to match experiment well.

Based on reports of poorer performance of best score aggregation on longer simulations, we were curious to know how sensitive our method is to data set size. We reasoned that larger data sets will typically include some incredibly rare conformation that, when docked to, will give a higher score than anything in a smaller data set. Without correctly accounting for such a conformation's rarity, this will trend toward worse results with increasing sampling. Phrased differently, the best score approach is an outlier detector that is only in the correct ballpark when—by happenstance—no outlier conformations have been sampled yet. In contrast, more data from longer initial simulations should cause PopShift's estimate to converge as the simulations do. The aforementioned outlier conformations, when correctly weighted by their rarity, will simply contribute to the overall picture of the ensemble, instead of dominating the prediction.

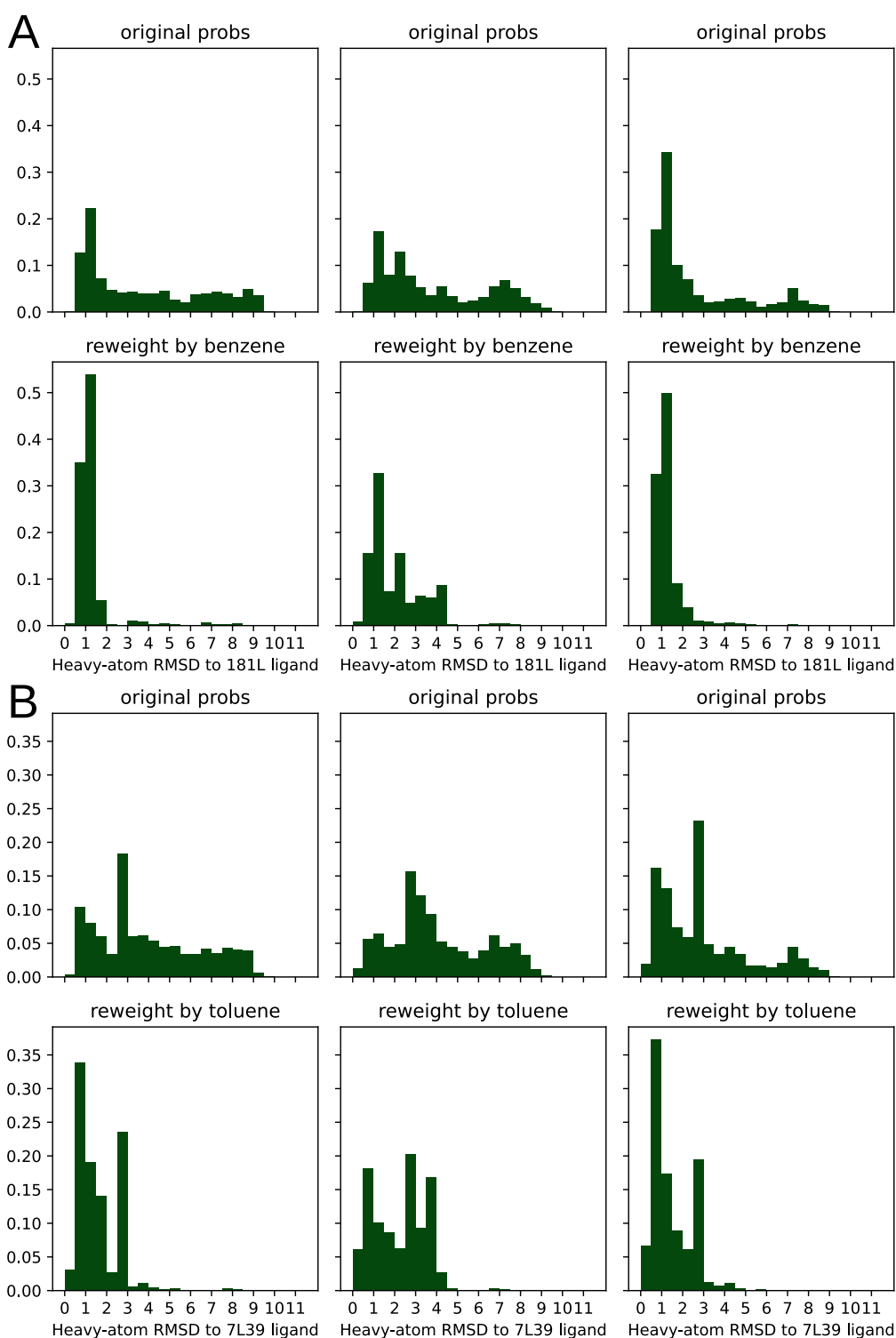
To test this notion, we truncated our data set as a series of fractions (that is, we took the first X% of each trajectory, where X corresponds to the fraction shown) and reran our analysis. Each result based on truncated data was generated by reworking that data set as though it were the full length data set, including featurization and clustering. We then inspected Pearson and Spearman correlations as a function of data set size (Figure 3). As before, we selected 20 frames per state to generate these truncated data sets. We also held cluster count fixed at 75, so that the total number of structures to dock to was not changing: only their diversity as the underlying feature trajectories became more mature and the quality of the MSMs providing the equilibrium population estimates.



**Figure 3.** PopShift performs well as one varies the amount of simulation data, whereas traditional “best score” ensemble docking gives worse performance and has greater statistical variation as data is added. The data set used for MSM construction was truncated by taking the indicated fraction from the beginning of each trajectory. This can be viewed as asking the question, “what would happen if simulations had been stopped early?” The error bars arise from the standard uncertainty in the mean across the three replica data sets.

We found that PopShift is less sensitive to the extent of input data than best-score ensemble docking. Because the number of structures docked to is constant across this sweep, it highlights how additional structural heterogeneity is not correctly indexed by simply looking for the most favorable score. As we suspected, the best score gets worse with more data because it detects outliers. In other words, if an ensemble is scored by its most favorable possible interaction with a ligand, the strain or unfavorability of that conformation on the protein is neglected. If one could create an ideal binding site for a ligand by moving residues out of the way such that it has ideal contacts, it would get very favorable docking scores when that conformation was docked to, but in fact the affinity for this site would be quite low because of how badly strained the protein would be by such rearrangements. In contrast, PopShift benefits from having more data, both in terms of the mean correlation with experiment and the statistical certainty in the results. These results also emphasize that at least for macroscopic binding constant estimation, our results for PopShift are not particularly sensitive to the length of input simulations. This is consistent with prior results suggesting that thermodynamic properties of MSMs converge quickly.<sup>34</sup> Taken together, this implies that adding conformational heterogeneity to a docking campaign is best done by including a correctly weighted sample of receptor conformations if the objective is ranked estimated affinity prediction.

**II.B. PopShift Retrodicts Ligand Poses and Their Relative Abundance.** Given the low RMSE between PopShift's predicted binding free energies and experimental measurements, we hypothesized that the approach also accurately predicts the pose the ligand adopts. Specifically, we reasoned that any ligand likely adopts a wide variety of different poses in different protein conformations from the ligand-free MSM. If PopShift works as intended, protein–ligand structures where the ligand resembles ligand-bound crystal structures should have significant increases in their equilibrium probabilities compared to the same protein structure in the ligand-free

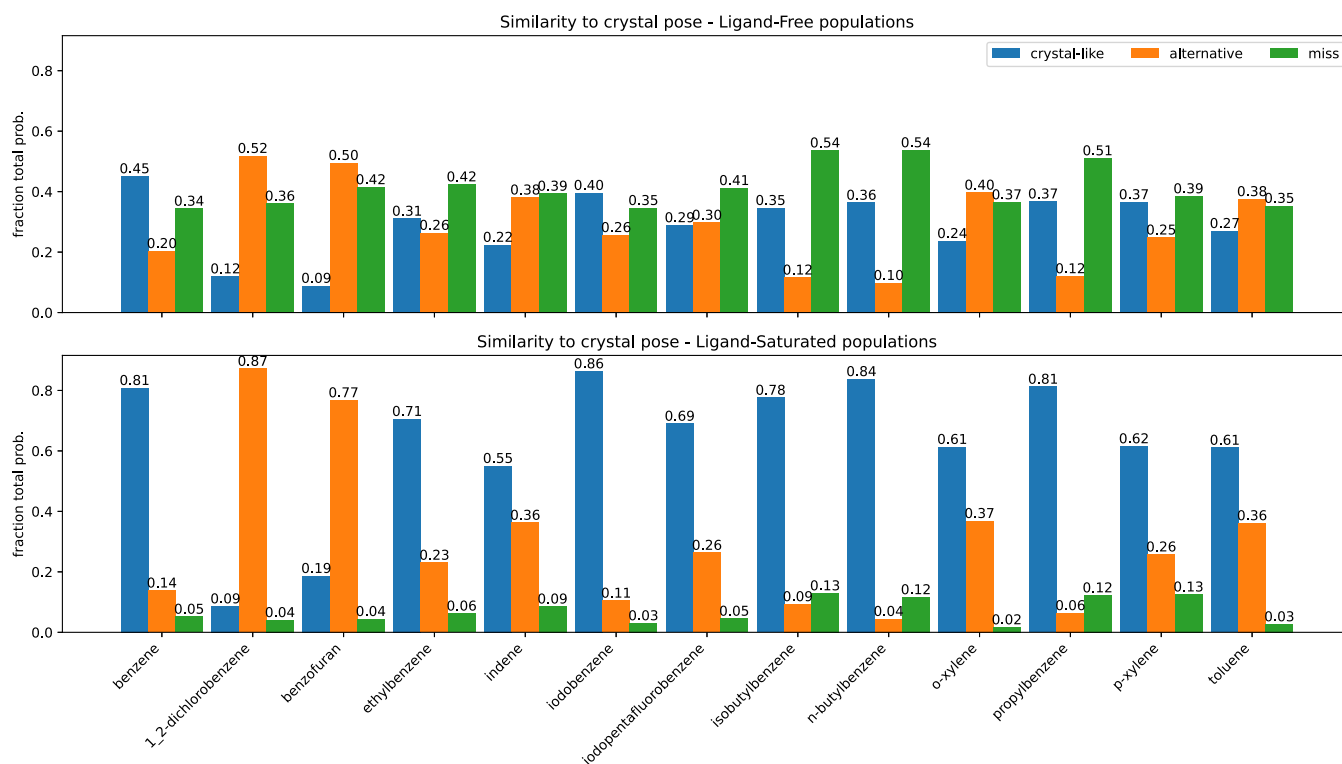


**Figure 4.** Population shift calculated by PopShift correctly favors ligand poses with a low RMSD to the crystal structure. The data histogrammed is the symmetry corrected RMSD<sup>35</sup> of the predicted pose to the holo crystal structure, where the structures are superimposed according to an alignment of their pocket atoms but not any ligand atoms. The RMSD histogram here is across all of the heavy atoms after this alignment. For each main panel, the three subpanel columns represent individual replica data sets. The top row provides the ligand-free equilibrium probabilities, and the lower row shows how the population is redistributed in the presence of the ligand. Main panel A provides the data for benzene, while the second panel B provides the data for toluene.

ensemble. In this case, the distribution of RMSDs from the reweighted ensemble should be more favorable than the distribution from the original ensemble (i.e., using the state populations from the ligand-free MSM instead of updating the

populations based on the strength of the interaction between protein and ligand).

To test our hypothesis, we compared the distribution of RMSDs to the ligand-bound crystal structure before and after reweighting the states based on the interaction with ligand



**Figure 5.** Summary of pose accuracy across all ligands studied. Each grouped bar represents the fraction of ligand poses that fall into the categories “crystal-like”, for within 2 Å symmetry-corrected RMSD<sup>35</sup> of the holo crystal structure, “alternative”, for between 2 and 4 Å RMSD from the holo crystal, and “miss”, for poses above 4. As before, the top panel represents the poses with an “apo” ensemble weighting. The lower panel provides pop-shifted reweights for the same data set.

(Figure 4). For each sample from each MSM bin,  $i$ , we weighted its apo probability as being  $\pi_i/n$ , where  $\pi_i$  is the equilibrium probability for that bin, and  $n$  is the number of samples drawn from each bin. We used eq 5 to estimate each sample’s probability in the presence of a saturating ligand based on the docking score for that particular sample. We aligned based on pocket residue heavy atoms (residues within 5 Å of *p*-xylene in PDB 186L). RMSDs were then computed across all heavy atoms in the ligands. We plot both histograms for each replica in Figure 4 to convey how reproducible the results are with different sets of simulations.

Our results for benzene and toluene show that PopShift does indeed favor low RMSD states compared with the broad heterogeneity in pose RMSDs from the original ensemble (Figure 4, panels for ligand-free populations.) Many states from the ligand-free ensemble are not compatible with the experimentally observed binding pose, resulting in RMSDs between the best scoring pose and the ligand-bound crystal structure over 4 Å. When pose RMSDs were reweighted using pop-shifted equilibrium probabilities at saturating ligand concentrations, the distribution collapses and poses become holo-like. Interestingly, for some ligands, such as toluene (Figure 4, panel B), alternative conformations appear to be present. Given the way crystal structures solved at cryogenic temperatures are known to favor low energy structures and underestimate structural heterogeneity, it is interesting to consider the possibility that the heterogeneity in poses that PopShift predicts for some ligands is real.

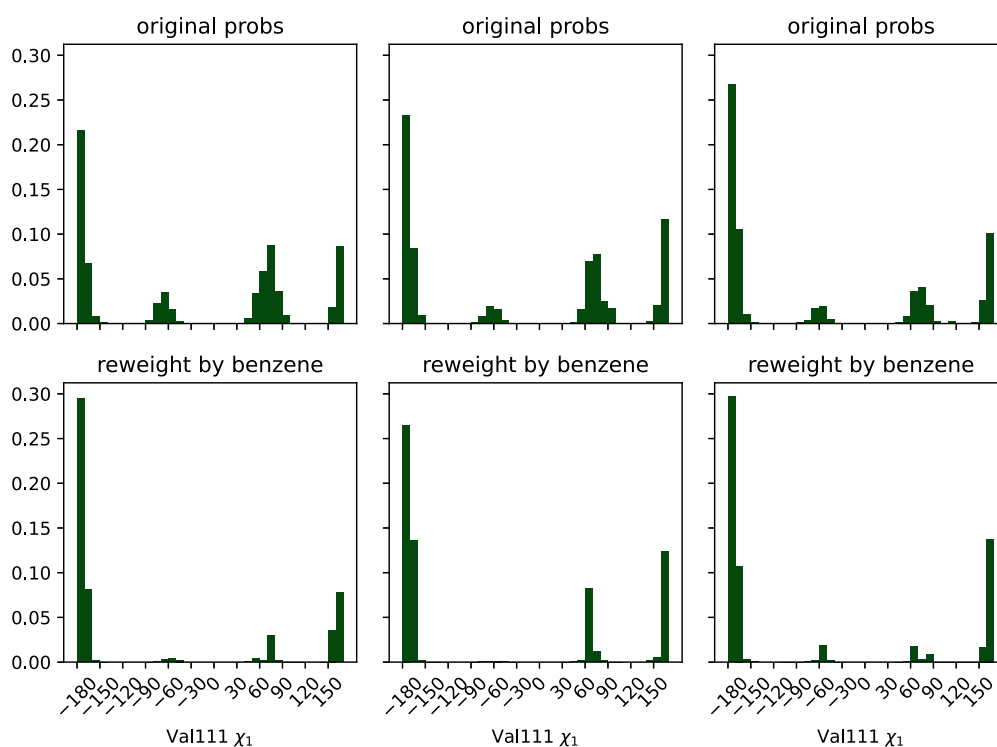
To test the generality of our results for the two ligands from Figure 4, we devised a means to judge how closely our predictions agree with experiments across multiple compounds. Because experimental techniques have a hard time describing

conformational heterogeneity, it is possible that poses dissimilar to experiment have relevance for the thermodynamic ensemble of the complex. Thus, we chose to use three categories: one for configurations similar to the crystal pose, one for conformations that were dissimilar but likely still in the pocket, and one for conformations with clashes or completely alternative ligand placements. We reasoned this would be reasonably measured by aligning the receptor pockets but transforming the ligands by that alignment.

We binned our histograms into three categories—fraction of samples that are similar to the crystal structure’s pose (RMSD < 2 Å), others that are in some alternative pose but probably still in the binding site ( $2 \leq \text{RMSD} < 4$  Å), and ones that are likely in a very different pose or outside the binding site altogether (RMSD  $\geq 4$  Å) in Figure 5. We aligned the  $\alpha$ -carbons of the pocket residues we used to build our MSMs, and then transformed our predicted ligand poses by that alignment transform. Thus, high-RMSD scores likely emerge from poses that have significant displacements in center of mass—that is, poses that are not properly in the binding pocket. We named the three categories of poses “crystal-like”, “alternative”, and “miss”, as abbreviations of this interpretation.

The pattern we demonstrated for benzene and toluene in Figure 4 is consistent across all the ligands we tested (Figure 5). Poses we categorize as “miss” are quite common with apo weights, but become rare after reweighting with PopShift. With ligand saturated weights, we often observe alternative poses. It is hard to know whether these conformations exist in solution, but they are probable in the ligand biased ensemble, suggesting that they contribute nontrivially to our estimates of affinity.

**II.C. PopShift Predicts How Ligands Change the Abundance of Protein Conformations.** Because macro-



**Figure 6.** Three alternative conformations of the Val111  $\chi_1$  angle in the ligand-free MSM collapse to mostly trans population in the presence of the ligand, in agreement with the dominant pose seen crystallographically. The top row of plots is histograms of the  $\chi_1$  torsion across the frames sampled from the MSM states, weighted using the ligand-free MSM equilibrium probabilities. The second row displays the same data, but weighted by the benzene-saturated equilibrium probabilities. Each column represents the results from one fully independent replica.

scopic affinities estimated with correctly weighted per-state affinities seem accurate, we reasoned that reweighted state probabilities might also be usefully accurate. We knew that Valine 111's dihedral angle is able to occupy several rotameric states in apo simulations, but that the distribution is different upon ligand binding.<sup>27,36</sup> Thus, we hypothesized that the broad distribution from our ligand-free MSM should collapse to the binding-compatible one upon reweighting with ligand-saturated populations.

We tested this hypothesis by histogramming valine 111 angles from the receptor structures we sampled, weighted by both apo- and ligand saturated state probabilities (Figure 6). As in Figure 4, we plot both histograms for each replica to convey the reproducibility of our results.

We find that the ligand-free protein broadly populates several different structures, but the ligand shifts the population to favor the trans state, with some population of gauche+. This angle is noted as having many different distributions in RT crystal structures for liganded T4 lysozyme L99A protein.<sup>27</sup> Trans is the angle modeled into cryo-structures from previous efforts (PDB ID 181L, 4W52).<sup>28,37</sup> Our plots suggest that, like room-temperature X-ray structures, the ligand-saturated ensemble is heterogeneous but does prefer certain angles, the primary of these being shared with the cryo X-ray structures. Thus, the receptor population has shifted through conformational selection to a binding-compatible ensemble.

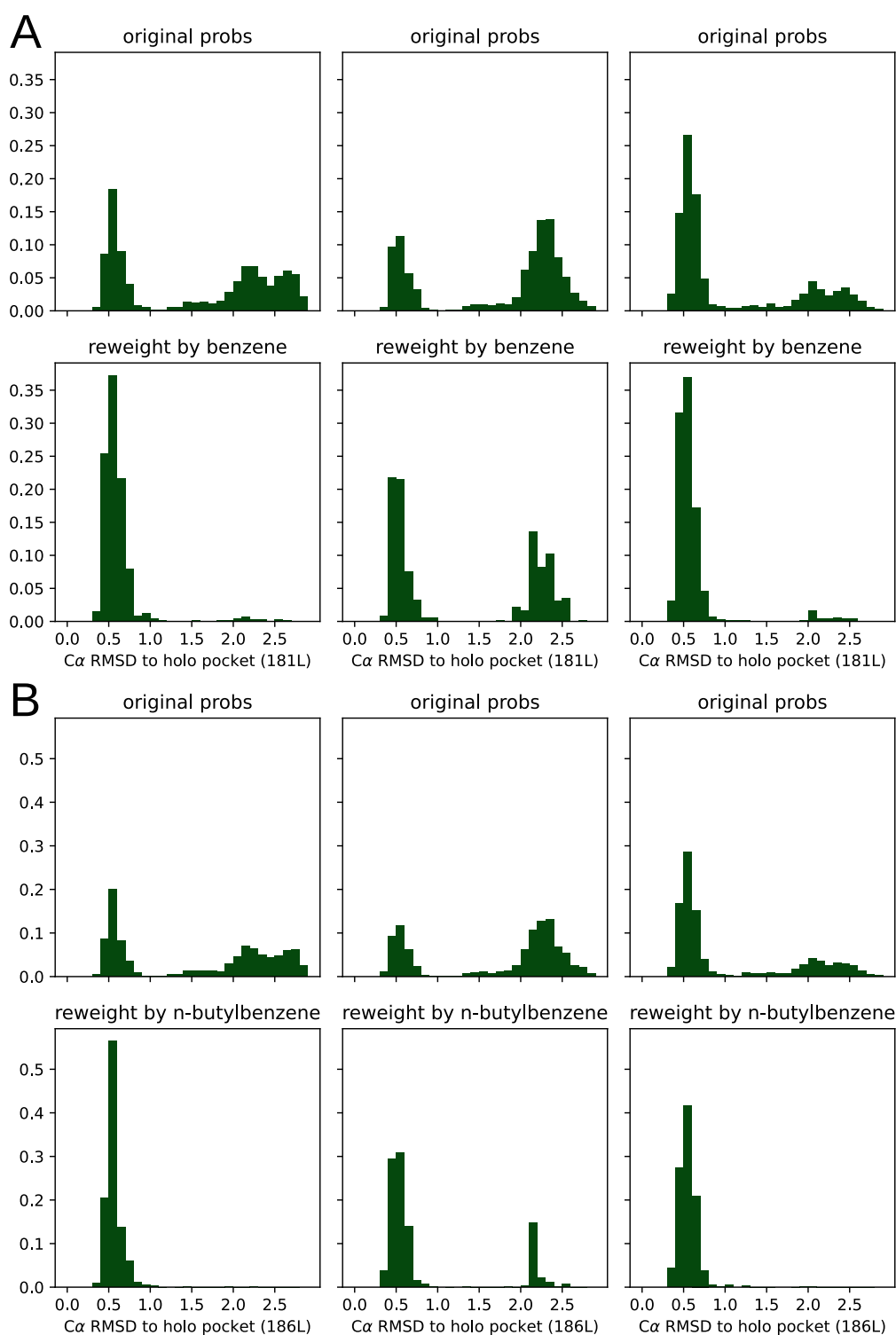
To assay this another way, we also calculated receptor  $\alpha$ -carbon RMSD histograms to the cognate crystal structure for pocket residues (see Figure 7). Because the F-helix is known to occupy several different conformations, we reasoned that the distance from the holo structure for a large and a small ligand would be indicative of what conformations are enriched by

ligand. We did this for benzene and *n*-butylbenzene as examples of smaller and larger ligands. We found that in both cases, as with the valine torsion histograms, our broad sampling of pocket conformations collapses to the holo-like structure in the presence of a saturating ligand.

**II.D. PopShift Can Estimate Ensemble Features as a Function of Ligand Concentration.** Because the histograms from Figure 4 represent unliganded and saturated conformational preferences, respectively, we hypothesized that inspecting the conformational preference as a function of ligand concentration might help with analyzing binding preferences and allosteric effects. We wanted to know at what ligand concentrations certain histogram populations become more prominent since structural features not directly corresponding to ligand binding are often relevant for drug development—particularly in the case of allosteric modulators. For example, we previously identified both activators and inhibitors that bind a cryptic pocket in the protein TEM  $\beta$ -lactamase.<sup>14</sup> Looking at what structures are stabilized/destabilized by a ligand could provide a facile means to predict their effects on the structural preferences of distant sites and, ultimately, on function. If a structural feature were used as a heuristic for some mechanistic action, that feature could be used to compute an EC50.

To display this transition, we computed ligand-rmsd histograms at various ligand concentrations and stacked them by descending ligand concentration as a dilution series. To do this, we recomputed the histograms from Figure 4 using eq 5 with a range of concentrations plugged in for  $x$  (see Figure 8). Each histogram-bin's probability was displayed using color, so that each row in the heatmap corresponds to a particular RMSD histogram at a particular ligand concentration. Bins, and therefore the X axis match those from Figure 4. We covered



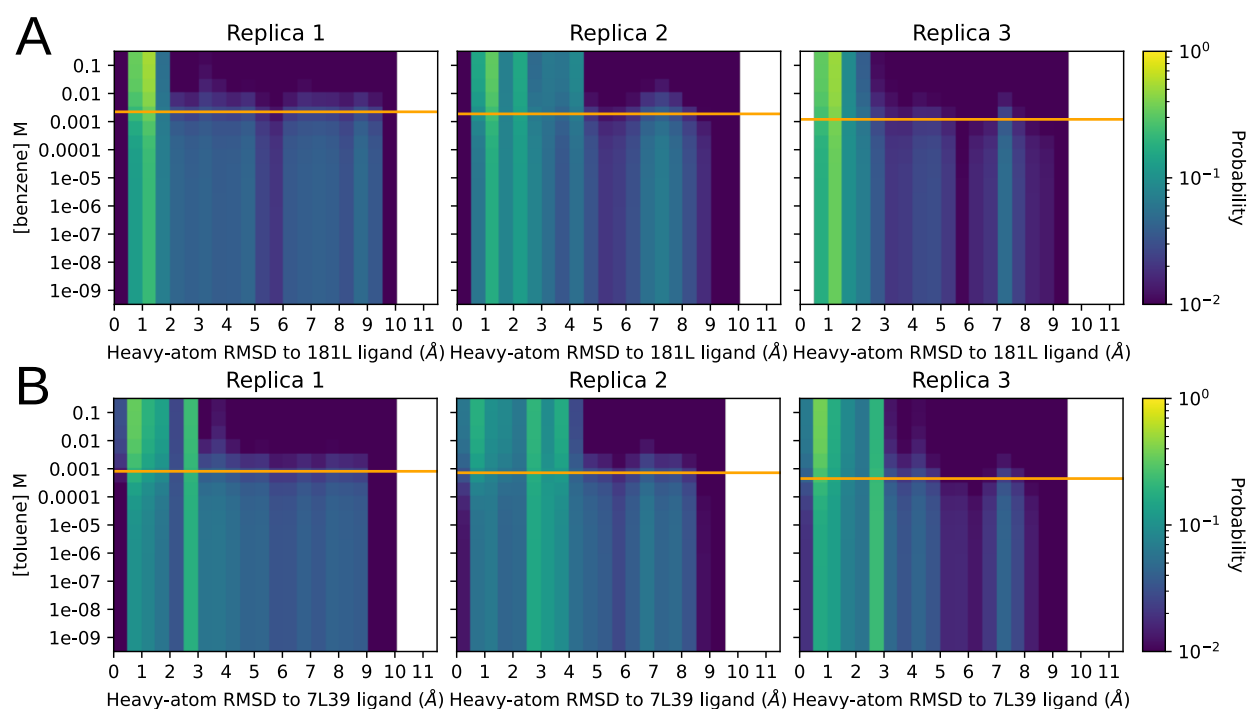


**Figure 7.** Population shift calculated by PopShift correctly favors pocket conformations with a low  $\alpha$  carbon RMSD for the holo crystal structure. The data histogrammed is the RMSD of the predicted pocket conformation to the holo crystal structure's pocket, where the structures are superimposed according to an alignment of their pocket atoms but not any ligand atoms. For each main panel, the three subpanel columns represent individual replica data sets. The top row shows the ligand-free equilibrium RMSD distributions and the lower row shows how the population is redistributed in the presence of ligand. Main panel A provides the data for benzene, while main panel B provides the data for *n*-butylbenzene.

concentrations ranging from nM—essentially no ligand for these relatively weak binders—to 0.1 M—completely saturating for all the ligands we tried. Our predicted ligand dissociation constants for each replica are marked by an orange line.

We found that the populations of conformers outside the binding site were abundant until within an order of magnitude or

so of the molecule's  $K_D$ , where a shift happened to structures that more closely resemble the ligand-bound structure. The poses closer to the  $K_D$  on the low concentration side contain a mixture of boundlike and ligand-free-like structures. They are similar to the ligand-free weighted distribution that we see in Figure 4.



**Figure 8.** Distribution of RMSDs to a ligand-bound reference structure as the concentration of the ligand is varied. Panels A and B represent the titrations for benzene and toluene, respectively. The Y axis is ligand concentration, and the X axis is the pocket-aligned symmetry corrected<sup>35</sup> Ligand RMSD to holo crystal structure, histogrammed as before (Figure 4). Each row in each heatmap is from the same raw data being histogrammed with different weights, computed by plugging the Y-value matching that row into eq 5. The orange line on each replica plot shows the  $K_D$  as calculated for that replica using eq 1 for reference. The heat, or intensity, is in logscale to aid visualization. The top row in each heatmap is holo-like, because it is at high ligand concentration, and the bottom row is apolike, because it is at low ligand concentration. Being able to titrate observables measurable from conformations in this way could provide exciting opportunities to understand ligand efficacy in systems with allosteric behavior.

### III. CONCLUSION

In this work, we have presented PopShift, a framework for estimating binding free energies in a manner that correctly weights the conformational heterogeneity present in the ligand binding sites of proteins. PopShift's estimated binding free energies from docking scores perform well compared to other common methods for the simple problem of lysozyme L99A binding to small organic compounds. Our results demonstrate that adding receptor fluctuations into docking is indeed a significant improvement over making the simplifying assumption that a single protein structure encodes all of the relevant information. Further, our approach provides an approximation of the receptor–ligand complex ensemble, which has utility in mechanistic studies, such as those focused on tuning the abundance of receptor conformations known to correlate with function.

The future directions for this approach are many. PopShift of docking scores for affinity estimation is still limited by the performance of the docking scoring function. Thus, it is likely that applying PopShift to more challenging problems, such as for ligands with many rotatable bonds or charge, will require more sophisticated estimates of per-state  $K_D$ 's. Using per-state estimates from Generalized-Born or Poisson–Boltzmann rescoring, or absolute or relative binding free energy simulations, is therefore an exciting and immediate future direction for this work.<sup>38</sup> More broadly, we see opportunities to apply this framework to other perturbations to an MSM sampled from one thermodynamic state, without having to redo the sampling in those new states—such as gracefully integrating multiple protonation states for either ligand or receptor and indexing the relative free energy changes of mutations. Expansive and

expensive sampling, done once for some reference model, can thus be reweighted to solve a host of important problems facing modern Biophysics.

### IV. METHODS

**IV.A. Free Energy Formalism with MSMs.** It has been shown previously that one can break up a binding free energy calculation across states discretized from simulations.<sup>21,30,39</sup> From Gallicchio, Lapelosa, and Levy,<sup>30</sup> (eq 22 in that work) we have

$$K = \sum_{i=1}^n \pi_i K_i \quad (1)$$

where  $K$  is the macroscopic binding constant, the index of the sum  $i$  runs over the  $n$  states of the discretization of state space for the receptor,  $\pi_i$  is the equilibrium population of that state from the ligand-free model, and  $K_i$  is the binding constant estimated for the ligand binding just to state  $i$ . Therefore, the free energy of binding becomes

$$\Delta G = -RT \log \sum_i \pi_i K_i \quad (2)$$

Critically, those states need only be sampled in one ensemble; for ease of sampling, because misestimation of both binding compatible and incompatible state abundances would lead to large errors, the apo ensemble is often preferred. This is also helpful when several or possibly many binding modes may contribute relevant amounts to the overall favorability of binding.

Gallicchio, Lapelosa, and Levy<sup>30</sup> also demonstrate that one can write the probability of a state in the ligand bound ensemble as the product of its original probability and the fractional contribution it makes to the equilibrium constant defined in eq 1.

$$\pi_i^* = \pi_i \frac{K_i}{K} \quad (3)$$

where  $\pi_i^*$  is the population of state  $i$  in the ligand bound ensemble and  $\pi_i$ ,  $K_i$ , and  $K$  are the equilibrium population from the unbound ensemble, the affinity of the ligand to state  $i$ , and the binding constant, respectively. We can use this to compute the extent to which population shifts as a function of ligand concentration, for example,  $[X]$ , as the sum of the unbound and bound population at that concentration.

$$w_i^*([X]) = \pi_i + \pi_i K_i [X] \quad (4)$$

Here  $w_i^*$ , the expression for the weight of the state, is not normalized; doing so gives

$$\pi_i^*([X]) = \frac{\pi_i(1 + K_i[X])}{\sum_i \pi_i(1 + K_i[X])} \quad (5)$$

Equation 5 is the expression for the reweighted (shifted) population of each state as a function of the free ligand concentration. In this work we are estimating  $K_i$  by docking to each state, but other means of estimating that equilibrium constant could be used.<sup>21</sup> Note that this equation has the correct limiting behavior. At high  $[X]$ , the populations of the states are dominated by the favorability of binding to each state because the right-hand numerator term and the denominator both grow with  $[X]$ , but the left-hand term does not. For any realistically measurable association constant (say,  $K > 1000$ ) with high ligand concentrations eq 5 becomes approximately equal to eq 3. That is to say, the probability of a state in an increasingly ligand-saturated ensemble approaches the probability of that state in the ligand bound ensemble. Conversely, at very low  $[X]$  the state probabilities are very nearly the apo ones, as the right-hand term in both the numerator and denominator becomes small relative to the left-hand term. (These same expressions can be obtained from Wyman and Gill,<sup>40</sup> see for example chapter 4, eqs 12 and 20 through 24.)

#### IV.B. Lysozyme MSM Simulations and Construction.

**IV.B.1. Simulations.** Simulations were run with Gromacs.<sup>41</sup> Three sets of ten simulations were run starting from protein coordinates taken from PDB ID 187L using the Amber03 force field.<sup>42</sup> Five of the trajectories in each set totaled 4  $\mu$ s of sampling, while the other five totaled 8  $\mu$ s. The protein was solvated with TIP3P explicit water in a dodecahedral box that extended one nm beyond the protein in any dimension and eight chloride ions were added to neutralize the charge.<sup>43,44</sup> This system was energy minimized with the steepest descent algorithm until the maximum force fell below 10 kJ/(mole\*nm) using a step size of 0.01 nm and a cutoff distance of 1.2 nm for the neighbor list, Coulomb interactions, and van der Waals interactions.

The system was then equilibrated at 298 K in a 1 ns NVT simulation followed by 1 ns NPT simulation with a position restraint on all protein heavy atoms (spring constant 1,000 kJ mol<sup>-1</sup> nm<sup>-2</sup>). A long-range dispersion correction was employed for both the energy and pressure. All bonds were constrained with the LINCS algorithm.<sup>45</sup> Cut-offs of 1.2, 0.9, and 0.9 nm were used for the neighbor list, Coulomb interactions, and van

der Waals interactions, respectively. The Verlet cutoff scheme was used for the neighbor list and particle mesh Ewald was employed for the electrostatics (with a grid spacing of 0.12 nm, PME order 4, and tolerance of  $1 \times 10^{-6}$ ).<sup>46</sup> The v-rescale thermostat (with a time constant of 0.1 ps) was used to hold the temperature at 298 K and the Berendsen barostat was used to bring the system to 1 bar pressure.<sup>47,48</sup> For the production runs, the position restraint was removed and the Parrinello–Rahman barostat was employed.<sup>49</sup> Snapshots were stored every 10 ps. Structures were visualized with PyMOL, and trajectories with both PyMOL and VMD.<sup>50,51</sup>

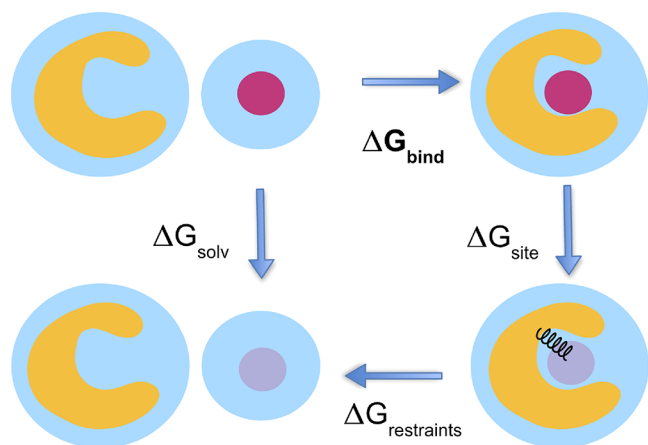
**IV.B.2. MSM Construction.** MSMs were constructed using Deeptime independently for each set of 10 simulations.<sup>52</sup> Clustering data was managed using the `RaggedArray` class from `enspara`.<sup>53</sup> Backbone and all  $\chi$  dihedrals for any residues with heavy atoms within 5 Å of *p*-xylene in PDB entry 187L were selected as input features. This feature space was reduced using TICA,<sup>31</sup> with lag times of 1, 2, and 5 ns and with a kinetic variance cutoff of 0.9 using commute mapping. Our final models were built from the 5 ns lag TICA features. We used k-means to cluster this reduced feature set, choosing our number of states using the cross-validation approach taken by Meller, Lotthammer, Smith, Novak, Lee, Kuhn, Greenberg, Leinwand, Greenberg, and Bowman.<sup>18</sup> Briefly, the reduced features were clustered by splitting features into 10 train-test pairs, where k-means with a range of  $k$  was used to cluster only the training set. Test set trajectories were then assigned to clusters using euclidean distance to the  $k$  centroids resulting from the “training” clustering. MSMs were fit to the train and test pairs using the MLE method.<sup>54</sup> The first 10 eigenmodes of both models were then VAMP-2 scored<sup>32</sup> using the train model, to estimate how overfit the model was to cluster count.<sup>55,56</sup> The number of clusters chosen for final model fitting was the point at which the VAMP-2 score of the test data starts to decline, which was  $k = 75$  in this case. The complete set of input features was reclustered using k-means with 75 clusters, and then an MLE model was fit with a lag time of 20 ns, after scrutinizing the implied time scales of the data with various lag times.

**IV.C. PopShift Workflow.** The workflow used here to do the PopShift postprocessing of our ensemble docking run is as follows:

1. Obtain a satisfactory ligand-free MSM. (Note that for thermodynamic observables all that is really needed for step 1 is a discretized state space or clustering of input features, and an associated collection of equilibrium probabilities for each state. MSM construction, especially with a collection of shorter trajectories as we have, is a sensible path to obtaining this association, but others are possible.)
2. Sample a number of receptor conformations from each state of the MSM using the assignment trajectories from clustering (Frame-picking).
3. Align these conformations so that they will fit neatly in a docking box.
4. Dock to each sample, saving the ligand pose and docking score.
5. Compile docking scores into free energies of binding using eq 2 and the equilibrium probabilities from the ligand-free MSM.
6. Compute reweighted state populations from apo weights, ligand concentration of interest, and docking scores using eq 5.

**IV.C.1. PopShift Implementation.** To pick frames, the assignment trajectories used for model selection were sorted into lists of frames corresponding to each cluster center, and then several of these (20 for the data in Figures 2 and 3) were selected by picking indices at random. These frames were extracted from the coordinate trajectories and iteratively aligned<sup>57</sup> by the  $\alpha$ -carbons of their pocket residues (defined in the same fashion as for MSM construction in Sec. IV.B.2), using LOOS.<sup>58,59</sup> Ligands and receptors were prepared using `prepare_ligand.py` and `prepare_receptor.py` from AutoDock tools.<sup>60</sup> We parallelized the preparation process by using GNU `parallel`.<sup>61</sup> Docking was performed using a box with 12-Å sides centered on the centroid of the average structure of the aligned frames using SMINA.<sup>33</sup> Each docking run targeting each extracted conformation was performed as an independent single CPU task using Jug.<sup>62</sup> We used the SMINA and Jug versions hosted on a conda forge. For SMINA the binary we used returned the following version statement (from calling `smina --version`): `Smina Nov 9 2017`. Based on AutoDock Vina 1.1.2. For Jug the version statement returned by the python module was `2.2.2`. No modifications were made to the docking energy model in this study. Analysis of extracted receptor conformations and docked ligand poses was performed using python scripts involving PyLOOS.<sup>59</sup> Ligand RMSDs were obtained via a symmetry corrected graph theoretic algorithm implemented in the `sprmsd` python package as downloaded from conda forge (version 0.6.0).<sup>35</sup> Docking scores were extracted and collated into ligand-indexed JSON associative arrays by using scripts provided in the PopShift package. OpenBabel 3.0.1 was used for ligand preparation for extended docking and to extract atomic numbers for `sprmsd`.<sup>63</sup> PopShift is available as open-source software and can be found on the Bowman Lab Github: <https://github.com/bowman-lab/PopShift>

**IV.D. Disappearing Ligand Absolute Binding Free Energy Simulations.** **IV.D.1. Starting Pose Selection.** The binding modes of the ligand for free energy calculations were selected using 5 methods, all based on the thermodynamic cycle depicted in Figure 9. For ligands with a known crystal structure bound to T4-Lysozyme, we selected the MSM pose with an



**Figure 9.** This cartoon depicts the thermodynamic cycle and transformation used for the absolute binding free energy calculations from Figure 2. We take  $\Delta G_{\text{bind}} = \Delta G_{\text{solv}} - \Delta G_{\text{site}} - \Delta G_{\text{restraints}}$ . In its weakly and noninteracting state, the ligand is free to leave the binding site. We use orientational Boresch-style restraints to reduce the phase space that must be sampled.<sup>64</sup>

RMSD closest to that of the crystal structure. All poses had an RMSD < 2 Å to the crystal pose and thus would be considered the same binding mode.<sup>65</sup> The exception to this was 1,2-dichlorobenzene which did not have an MSM pose that closely matched the known crystal structure. In the case of 1,2-dichlorobenzene, we used the exact pose from the crystal structure.

For ligands without a crystal structure, we took a known crystal structure most similar to the ligand and posed our ligand accordingly in several poses. Each pose for each ligand was simulated for 2 ns, and an RMSD analysis of the ligand throughout the trajectory as compared to the starting pose was run. The pose with the smallest change in RMSD and the smallest variance in RMSD was chosen as the most stable pose and was the pose used for the remainder of calculations. We overlaid 2-ethyltoluene with the crystal pose of *o*-xylene and then flipped 2-ethyltoluene so the ethyl group and methyl group would align first with the 1-methyl and 2-methyl of *o*-xylene as pose 1, and the 2-methyl and 1-methyl as pose 2. We overlaid 3-ethyltoluene with the crystal pose of *o*-xylene. We aligned the ethyl group of 3-ethyltoluene with each methyl group of *o*-xylene, resulting in 4 different poses. We also aligned the methyl group of 3-ethyltoluene with each methyl group of *o*-xylene resulting in 4 more different poses. We overlaid 4-ethyltoluene with the crystal pose of *p*-xylene, and then flipped 4-ethyltoluene so the ethyl group and methyl group would align first with the 1-methyl and 4-methyl of *o*-xylene as pose 1, and the 4-methyl and 1-methyl as pose 2. We overlaid thianaphthene with the crystal pose of indene as pose 1, and flipped the thianaphthene across its length so the sulfur would be on the opposite side as pose 2.

Lastly, for *m*-xylene, we began with a process similar to that of 3-ethyltoluene. We overlaid *m*-xylene with the crystal pose of *o*-xylene. We aligned the methyl groups of *m*-xylene with each methyl group of *o*-xylene resulting in 4 different poses. Again, an RMSD was used to choose the most stable pose to use for the remainder of calculations. However, upon free running free energy calculations, we observed *m*-xylene switching to a different stable pose after 1.5 ns, impacting the free energy calculation. For *m*-xylene, the stable pose found at 1.5 ns in the free energy calculation was chosen.

**IV.D.2. Ligand and Protein Parameterization.** The ligands were parametrized with Open Force Field version 2.0.0 and charged with AM1-BCC charges.<sup>66</sup> The protein (PDB 7I38) was prepared using an OpenEye Spruce to add hydrogen atoms at pH 7.0. The protonated protein was then parametrized using AMBER ff14SB and the TIP3P water model was used for the waters. GROMACS was used to solvate and add a salt concentration of 150 mM to the ligand and protein–ligand systems. Each ligand system was energy minimized and NVT equilibrated, and then a 2 ns NPT production run was performed. Each protein–ligand system was energy minimized and NVT equilibrated, then a 2 ns NPT production run was performed. The trajectory of the production run was used to select the atoms and dihedrals for the Boresch restraints to restrain the ligand to the binding site during simulation.<sup>64</sup>

**IV.D.3. Running Absolute Binding Free Energy Calculations in GROMACS.** Simulations were run by using GROMACS 2021.2. For binding site simulations, we used 20 lambda windows. In this protocol, we first restrained the ligand to the binding site, turned off the Coulomb interactions, and then turned off the vdW interactions. For unbound ligand simulations, we performed absolute hydration free energies. In



0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
restrain																				
		coulombs off																		
						van der waals off														

**Figure 10.** This  $\lambda$  schedule depicts the schedule for restraining and scaling down ligand interactions (coulomb and vdW) for the protein–ligand protocol discussed in the text.

this protocol, we first turned off the Coulomb interactions and then turned off the vdW interactions.

Prior to running production simulations, every lambda window was energy minimized for 5000 steps using the steepest descent and equilibrated at constant volume for 10 ps at 298.15 K. Production simulations were run for 15 ns per lambda window with an NPT ensemble. During production, replica exchange was attempted every 200 steps. See Figure 10 for a graphical representation of this schedule.

**IV.D.4. Analysis of Absolute Binding Simulation Results.** We obtained the free energy difference using the alchemlyb/pymbar package MBAR estimator.<sup>67</sup> The first nanosecond of the 15 ns of production simulations was discarded as equilibration. Each ligand was inspected for symmetry and the trajectory of the protein–ligand system in its unrestrained state was inspected to determine whether all symmetries were equally sampled.<sup>68</sup> The free energy difference of ligands with at least one axis of symmetry without adequate sampling of symmetries was corrected using the equation:

$$\Delta G_{\text{symmetry condition}} = \frac{1}{k_B T} \log \sigma_L \quad (6)$$

where  $k_B$  is the Boltzmann constant,  $T$  is the temperature, and  $\sigma_L$  is the ligand symmetry.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

Data files and scripts used to analyze results and produce figures, as well as trajectory data, will be available upon request. The data is hosted (possibly via links or other download means, depending on which data it is) at <https://github.com/bowman-lab/popshift-ms-data>. Software: Python command line tools written to perform the PopShift calculations discussed here are distributed as free software at: <https://github.com/bowman-lab/PopShift>.

## ■ AUTHOR INFORMATION

### Corresponding Author

**Gregory R. Bowman** – Departments of Biochemistry & Biophysics and Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, United States; [orcid.org/0000-0002-2083-4892](https://orcid.org/0000-0002-2083-4892); Email: [grbowman@seas.upenn.edu](mailto:grbowman@seas.upenn.edu)

### Authors

**Louis G. Smith** – Departments of Biochemistry & Biophysics and Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, United States; [orcid.org/0000-0001-9576-0421](https://orcid.org/0000-0001-9576-0421)

**Borna Novak** – Department of Biochemistry and Molecular Biophysics, Washington University in St. Louis, St. Louis, Missouri 63130, United States; Medical Scientist Training

Program, Washington University in St. Louis, St. Louis, Missouri 63130, United States

**Meghan Osato** – School of Pharmacy and Pharmaceutical Sciences, University of California, Irvine, Irvine, California 92697, United States

**David L. Mobley** – School of Pharmacy and Pharmaceutical Sciences, University of California, Irvine, Irvine, California 92697, United States

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jctc.3c00870>

### Author Contributions

<sup>†</sup>(L.G.S. and B.N.) These authors contributed equally to this work.

### Funding

This work was funded in part by NSF MCB-2218156, NIH NIA RF1AG067194, and the Bassett Center for BRCA at the University of Pennsylvania, to G.R.B., and NIH R35GM148236 and R01GM132386 to D.L.M. G.R.B. holds a Packard Fellowship for Science and Engineering from The David & Lucile Packard Foundation.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank Alan Grossfield for galvanizing discussions and encouragement. We also thank Artur Meller and Matt Cruz for insightful comments over the course of the work. We thank Justin J. Miller for helpful discussion of ITC reference data.

## ■ REFERENCES

- (1) Bender, B. J.; Gahbauer, S.; Lutten, A.; Lyu, J.; Webb, C. M.; Stein, R. M.; Fink, E. A.; Balius, T. E.; Carlsson, J.; Irwin, J. J.; Shoichet, B. K. A Practical Guide to Large-Scale Docking. *Nat. Protoc.* **2021**, *16*, 4799–4832.
- (2) Gentile, F.; Oprea, T. I.; Tropsha, A.; Cherkasov, A. Surely You Are Joking, Mr Docking. *Chem. Soc. Rev.* **2023**, *52*, 872–878.
- (3) Ban, F.; Dalal, K.; Li, H.; LeBlanc, E.; Rennie, P. S.; Cherkasov, A. Best Practices of Computer-Aided Drug Discovery: Lessons Learned from the Development of a Preclinical Candidate for Prostate Cancer with a New Mechanism of Action. *J. Chem. Inf. Model.* **2017**, *57*, 1018–1028.
- (4) Pohorille, A.; Jarzynski, C.; Chipot, C. Good Practices in Free-Energy Calculations. *J. Phys. Chem. B* **2010**, *114*, 10235–10253.
- (5) Mey, A. S. J. S.; Allen, B. K.; Bruce MacDonald, H. E.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; Rizzi, A.; Scheen, J.; Shirts, M. R.; Tresadern, G.; Xu, H. Best Practices for Alchemical Free Energy Calculations [Article v1.0]. *Living Journal of Computational Molecular Science* **2020**, *2*, 18378–18378.
- (6) Amezcua, M.; El Khoury, L.; Mobley, D. L. SAMPL7 Host-Guest Challenge Overview: Assessing the Reliability of Polarizable and Non-Polarizable Methods for Binding Free Energy Calculations. *J. Comput. Aided Mol. Des.* **2021**, *35*, 1–35.

- (7) Wierbowski, S. D.; Wingert, B. M.; Zheng, J.; Camacho, C. J. Cross-Docking Benchmark for Automated Pose and Ranking Prediction of Ligand Binding. *Protein Sci.* **2019**, *29*, 298–305.
- (8) Barril, X.; Morley, S. D. Unveiling the Full Potential of Flexible Receptor Docking Using Multiple Crystallographic Structures. *J. Med. Chem.* **2005**, *48*, 4432–4443.
- (9) Lim, N. M.; Wang, L.; Abel, R.; Mobley, D. L. Sensitivity in Binding Free Energies Due to Protein Reorganization. *J. Chem. Theory Comput.* **2016**, *12*, 4620–4631.
- (10) Amaro, R. E.; Baudry, J.; Chodera, J.; Demir, Ö.; McCammon, J. A.; Miao, Y.; Smith, J. C. Ensemble Docking in Drug Discovery. *Biophys. J.* **2018**, *114*, 2271–2278.
- (11) Korb, O.; Olsson, T. S. G.; Bowden, S. J.; Hall, R. J.; Verdonk, M. L.; Liebeschuetz, J. W.; Cole, J. C. Potential and Limitations of Ensemble Docking. *J. Chem. Inf. Model.* **2012**, *52*, 1262–1274.
- (12) Xu, M.; Lill, M. A. Utilizing Experimental Data for Reducing Ensemble Size in Flexible-Protein Docking. *J. Chem. Inf. Model.* **2012**, *52*, 187–198.
- (13) Mohammadi, S.; Narimani, Z.; Ashouri, M.; Firouzi, R.; Karimi-Jafari, M. H. Ensemble Learning from Ensemble Docking: Revisiting the Optimum Ensemble Size Problem. *Sci. Rep.* **2022**, *12*, 410.
- (14) Hart, K. M.; Moeder, K. E.; Ho, C. M. W.; Zimmerman, M. I.; Frederick, T. E.; Bowman, G. R. Designing Small Molecules to Target Cryptic Pockets Yields Both Positive and Negative Allosteric Modulators. *PLoS One* **2017**, *12*, e0178678.
- (15) Kamenik, A. S.; Singh, I.; Lak, P.; Balius, T. E.; Liedl, K. R.; Shoichet, B. K. Energy Penalties Enhance Flexible Receptor Docking in a Model Cavity. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, e2106195118.
- (16) Porter, J. R.; Moeder, K. E.; Sibbald, C. A.; Zimmerman, M. I.; Hart, K. M.; Greenberg, M. J.; Bowman, G. R. Cooperative Changes in Solvent Exposure Identify Cryptic Pockets, Switches, and Allosteric Coupling. *Biophys. J.* **2019**, *116*, 818–830.
- (17) Cruz, M. A.; Frederick, T. E.; Mallimadugula, U. L.; Singh, S.; Vithani, N.; Zimmerman, M. I.; Porter, J. R.; Moeder, K. E.; Amarasinghe, G. K.; Bowman, G. R. A Cryptic Pocket in Ebola VP35 Allosterically Controls RNA Binding. *Nat. Commun.* **2022**, *13*, 2269.
- (18) Meller, A.; Lotthammer, J. M.; Smith, L. G.; Novak, B.; Lee, L. A.; Kuhn, C. C.; Greenberg, L.; Leinwand, L. A.; Greenberg, M. J.; Bowman, G. R. Drug Specificity and Affinity Are Encoded in the Probability of Cryptic Pocket Opening in Myosin Motor Domains. *eLife* **2023**, *12*, e83602.
- (19) *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*; Bowman, G. R., Pande, V. S., Noé, F., Eds.; Springer Netherlands: Dordrecht, 2014; <https://link.springer.com/10.1007/978-94-007-7606-7>.
- (20) Changeux, J.-P.; Edelstein, S. Conformational Selection or Induced Fit? 50 Years of Debate Resolved. *F1000Prime Rep* **2011**, *3*, 19.
- (21) Jayachandran, G.; Shirts, M. R.; Park, S.; Pande, V. S. Parallelized-over-Parts Computation of Absolute Binding Free Energy with Docking and Molecular Dynamics. *J. Chem. Phys.* **2006**, *125*, 084901.
- (22) Minh, D. D. L. Implicit Ligand Theory: Rigorous Binding Free Energies and Thermodynamic Expectations from Molecular Docking. *J. Chem. Phys.* **2012**, *137*, 104106.
- (23) Xie, B.; Nguyen, T. H.; Minh, D. D. L. Absolute Binding Free Energies between T4 Lysozyme and 141 Small Molecules: Calculations Based on Multiple Rigid Receptor Configurations. *J. Chem. Theory Comput.* **2017**, *13*, 2930–2944.
- (24) Mobley, D. L.; Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annual Review of Biophysics* **2017**, *46*, 531–558.
- (25) Mobley, D. L.; Heinzlmann, G.; Henriksen, N. M.; Gilson, M. K. *Predicting Binding Free Energies: Frontiers and Benchmarks (a Perpetual Review)*. Aug. 21, 2017. <https://escholarship.org/uc/item/9p37m6bq> (accessed 2023).
- (26) Baumann, H. M.; Gapsys, V.; de Groot, B. L.; Mobley, D. L. Challenges Encountered Applying Equilibrium and Nonequilibrium Binding Free Energy Calculations. *J. Phys. Chem. B* **2021**, *125*, 4241–4261.
- (27) Bradford, S. Y. C.; El Khoury, L.; Ge, Y.; Osato, M.; Mobley, D. L.; Fischer, M. Temperature Artifacts in Protein Structures Bias Ligand-Binding Predictions. *Chem. Sci.* **2021**, *12*, 11275–11293.
- (28) Morton, A.; Baase, W. A.; Matthews, B. W. Energetic Origins of Specificity of Ligand Binding in an Interior Nonpolar Cavity of T4 Lysozyme. *Biochemistry* **1995**, *34*, 8564–8575.
- (29) Knoverek, C. R.; Amarasinghe, G. K.; Bowman, G. R. Advanced Methods for Accessing Protein Shape-Shifting Present New Therapeutic Opportunities. *Trends Biochem. Sci.* **2019**, *44*, 351–364.
- (30) Gallicchio, E.; Lapelosa, M.; Levy, R. M. Binding Energy Distribution Analysis Method (BEDAM) for Estimation of Protein–Ligand Binding Affinities. *J. Chem. Theory Comput.* **2010**, *6*, 2961–2977.
- (31) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of Slow Molecular Order Parameters for Markov Model Construction. *J. Chem. Phys.* **2013**, *139*, 015102.
- (32) Wu, H.; Noé, F. Variational Approach for Learning Markov Processes from Time Series Data. *J. Nonlinear Sci.* **2020**, *30*, 23–66.
- (33) Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons Learned in Empirical Scoring with Smina from the CSAR 2011 Benchmarking Exercise. *J. Chem. Inf. Model.* **2013**, *53*, 1893–1904.
- (34) Huang, X.; Bowman, G. R.; Bacallado, S.; Pande, V. S. Rapid Equilibrium Sampling Initiated from Nonequilibrium Data. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 19765–19769.
- (35) Meli, R.; Biggin, P. C. Spyrmsd: Symmetry-Corrected RMSD Calculations in Python. *Journal of Cheminformatics* **2020**, *12*, 49.
- (36) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. Predicting Absolute Ligand Binding Free Energies to a Simple Model Site. *J. Mol. Biol.* **2007**, *371*, 1118–1134.
- (37) Merski, M.; Fischer, M.; Balius, T. E.; Eidam, O.; Shoichet, B. K. Homologous Ligands Accommodated by Discrete Conformations of a Buried Cavity. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 5039–5044.
- (38) Plattner, N.; Noé, F. Protein Conformational Plasticity and Complex Ligand-Binding Kinetics Explored by Atomistic Simulations and Markov Models. *Nat. Commun.* **2015**, *6*, 7653.
- (39) Gallicchio, E.; Levy, R. M.; Christov, C. Recent theoretical and computational advances for modeling protein–ligand binding affinities. *Advances in Protein Chemistry and Structural Biology* **2011**, *85*, 27–80.
- (40) Wyman, J.; Gill, S. J. *Binding and Linkage: Functional Chemistry of Biological Macromolecules*; University Science Books: 1990; p 358.
- (41) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19.
- (42) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (43) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (44) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.
- (45) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.
- (46) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (47) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (48) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (49) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

- (50) Schrödinger, LLC. *The PyMOL Molecular Graphics System*, Version 2.5.4; 2022.
- (51) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (52) Hoffmann, M.; Scherer, M.; Hempel, T.; Mardt, A.; de Silva, B.; Husic, B. E.; Klus, S.; Wu, H.; Kutz, N.; Brunton, S. L.; Noé, F. Deeptime: A Python Library for Machine Learning Dynamical Models from Time Series Data. *Mach. Learn.: Sci. Technol.* **2022**, *3*, 015009.
- (53) Porter, J. R.; Zimmerman, M. I.; Bowman, G. R. Enspara: Modeling Molecular Ensembles with Scalable Data Structures and Parallel Computing. *J. Chem. Phys.* **2019**, *150*, 044108.
- (54) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011**, *134*, 174105.
- (55) McGibbon, R. T.; Pande, V. S. Variational Cross-Validation of Slow Dynamical Modes in Molecular Kinetics. *J. Chem. Phys.* **2015**, *142*, 124105.
- (56) Chen, S.; Wiewiora, R. P.; Meng, F.; Babault, N.; Ma, A.; Yu, W.; Qian, K.; Hu, H.; Zou, H.; Wang, J.; Fan, S.; Blum, G.; Pittella-Silva, F.; Beauchamp, K. A.; Tempel, W.; Jiang, H.; Chen, K.; Skene, R. J.; Zheng, Y. G.; Brown, P. J.; Jin, J.; Luo, C.; Chodera, J. D.; Luo, M. The Dynamic Conformational Landscape of the Protein Methyltransferase SETD8. *eLife* **2019**, *8*, e45403.
- (57) Grossfield, A.; Feller, S. E.; Pitman, M. C. Convergence of Molecular Dynamics Simulations of Membrane Proteins. *Proteins: Struct., Funct., Bioinf.* **2007**, *67*, 31–40.
- (58) Romo, T. D.; Grossfield, A. LOOS: An Extensible Platform for the Structural Analysis of Simulations. In *Engineering in Medicine and Biology Society, 2009. Annual International Conference of the IEEE; IEEE: 2009*; pp 2332–2335.
- (59) Romo, T. D.; Leioatts, N.; Grossfield, A. Lightweight Object Oriented Structure Analysis: Tools for Building Tools to Analyze Molecular Dynamics Simulations. *J. Comput. Chem.* **2014**, *35*, 2305–2318.
- (60) Forli, S.; Huey, R.; Pique, M. E.; Sanner, M. F.; Goodsell, D. S.; Olson, A. J. Computational Protein-Ligand Docking and Virtual Drug Screening with the AutoDock Suite. *Nat. Protoc.* **2016**, *11*, 905–919.
- (61) Tange, O. GNU Parallel 20220922 ('Elizabeth') Released. *Zenodo*, Sept. 22, 2022. <https://zenodo.org/record/7105792>.
- (62) Coelho, L. P. Jug: Software for Parallel Reproducible Computation in Python. *Journal of Open Research Software* **2017**, *5*, 30.
- (63) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *Journal of Cheminformatics* **2011**, *3*, 33.
- (64) Baumann, H. M.; Dybeck, E.; McClendon, C. L.; Pickard, F. C. I.; Gapsys, V.; Pérez-Benito, L.; Hahn, D. F.; Tresadern, G.; Mathiowetz, A. M.; Mobley, D. L. Broadening the Scope of Binding Free Energy Calculations Using a Separated Topologies Approach. *J. Chem. Theory Comput.* **2023**, *19*, 5058–5076.
- (65) Gill, S. C.; Lim, N. M.; Grinaway, P. B.; Rustenburg, A. S.; Fass, J.; Ross, G. A.; Chodera, J. D.; Mobley, D. L. Binding Modes of Ligands Using Enhanced Sampling (BLUES): Rapid Decorrelation of Ligand Binding Modes via Nonequilibrium Candidate Monte Carlo. *J. Phys. Chem. B* **2018**, *122*, 5579–5598.
- (66) Boothroyd, S.; Behara, P. K.; Madin, O.; Hahn, D.; Jang, H.; Gapsys, V.; Wagner, J.; Horton, J.; Dotson, D.; Thompson, M.; Maat, J.; Gokey, T.; Wang, L.-P.; Cole, D.; Gilson, M.; Chodera, J.; Bayly, C.; Shirts, M.; Mobley, D. Development and Benchmarking of Open Force Field 2.0.0 — the Sage Small Molecule Force Field. *ChemRxiv*, Nov. 21, 2022, <https://chemrxiv.org/engage/chemrxiv/article-details/637938cbe70b0a110aa33b8b>.
- (67) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105.
- (68) Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the Use of Orientational Restraints and Symmetry Corrections in Alchemical Free Energy Calculations. *J. Chem. Phys.* **2006**, *125*, 084902.