JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

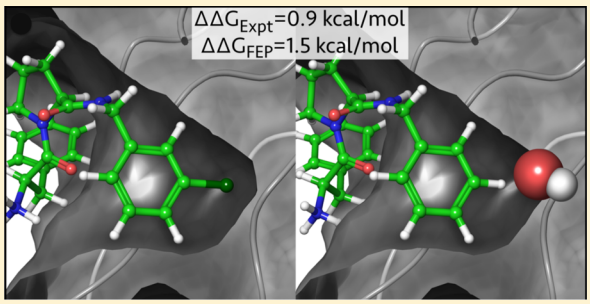
# Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field

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 Masse,[‡] Jennifer L. Knight,[†] Thomas Steinbrecher,[†] Thijs Beuming,[†] Wolfgang Damm,[†] Ed Harder,[†] Woody Sherman,[†] Mark Brewer,[†] Ron Wester,[‡] Mark Murcko,[†] Leah Frye,[†] Ramy Farid,[†] Teng Lin,[†] David L. Mobley,[⊥] William L. Jorgensen,[∥] Bruce J. Berne,[§] Richard A. Friesner,[§] and Robert Abel*[,†]

[†]Schrödinger, Inc., 120 West 45th Street, New York, New York 10036, United States
[‡]Nimbus Discovery, 25 First Street, Suite 404, Cambridge, Massachusetts 02141, United States
[§]Department of Chemistry, Columbia University, 3000 Broadway, New York, New York 10027, United States
[∥]Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States
[⊥]Departments of Pharmaceutical Sciences and Chemistry, University of California—Irvine, Irvine, California 92697, United States

Supporting Information

**ABSTRACT:** Designing tight-binding ligands is a primary objective of small-molecule drug discovery. Over the past few decades, free-energy calculations have benefited from improved force fields and sampling algorithms, as well as the advent of low-cost parallel computing. However, it has proven to be challenging to reliably achieve the level of accuracy that would be needed to guide lead optimization (~5× in binding affinity) for a wide range of ligands and protein targets. Not surprisingly, widespread commercial application of free-energy simulations has been limited due to the lack of large-scale validation coupled with the technical challenges traditionally associated with running these types of calculations. Here, we report an approach that achieves an unprecedented level of accuracy across a broad range of target classes and ligands, with retrospective results encompassing 200 ligands and a wide variety of chemical perturbations, many of which involve significant changes in ligand chemical structures. In addition, we have applied the method in prospective drug discovery projects and found a significant improvement in the quality of the compounds synthesized that have been predicted to be potent. Compounds predicted to be potent by this approach have a substantial reduction in false positives relative to compounds synthesized on the basis of other computational or medicinal chemistry approaches. Furthermore, the results are consistent with those obtained from our retrospective studies, demonstrating the robustness and broad range of applicability of this approach, which can be used to drive decisions in lead optimization.




## INTRODUCTION

Protein−ligand binding is central to both biological function and pharmaceutical activity. Some ligands simply inhibit protein function, while others induce protein conformational changes and hence can modulate key cell-signaling pathways. In either case, achieving a desired therapeutic effect is dependent upon the magnitude of the binding affinity of ligand to target receptor. Designing tight-binding ligands while maintaining the other ligand properties required for safety and biological efficacy is a primary objective of small-molecule drug discovery projects.

A principal goal of computational chemistry and computer-aided drug design (CADD) is therefore the accurate prediction of protein−ligand free energies of binding (i.e., binding affinities).[1,2] The most rigorous approach to this problem is free-energy simulation. A variety of free-energy simulation methods, such as free-energy perturbation (FEP), thermodynamic integration (TI), and $\lambda$ dynamics, employ an analysis of atomistic molecular dynamics or Monte Carlo simulations to determine the free-energy difference between two related ligands via either a chemical or alchemical path.[3−9] In drug discovery lead optimization applications, the calculation of relative binding affinities (i.e., the relative difference in binding energy between two compounds) is generally the quantity of interest and is thought to afford significant reduction in

Received: December 15, 2014
Published: January 27, 2015



DOI: 10.1021/ja512751q
J. Am. Chem. Soc. 2015, 137, 2695−2703

computational effort as compared to absolute binding free-energy calculations.

Nearly three decades have passed since the initial applications of free-energy methods to the calculation of protein−ligand binding affinities were first reported by the Jorgensen, McCammon, and Kollman groups.[10−15] Subsequent efforts since that seminal work have reported anecdotal results for a small number of protein−ligand complexes, but these have suffered from a lack of computing power and inadequacies in both sampling algorithms and molecular mechanics force fields.[1,5,8] As a result, use of free-energy calculations was limited in an industrial drug discovery setting, where high throughput, predictive accuracy, and robustness are required to make a significant impact.

In recent years, FEP calculations have benefitted from improved force fields, new sampling algorithms, and the emergence of low-cost parallel computing, which have resulted in the level of accuracy and turnaround time needed to impact lead optimization efforts, as demonstrated in several academic projects.[1,5,8,16−20] However, it has not been demonstrated that highly accurate results can be achieved reliably across a wide range of ligands and protein targets, as would be needed for the method to be useful in industrial pharmaceutical research programs.

Here, we report an FEP protocol that enables highly accurate affinity predictions across a broad range of ligands and target classes (over 200 ligands and 10 targets). The ligand perturbations include a wide range of chemical modifications that are typically seen in medicinal chemistry efforts, with modifications of up to 10 heavy atoms routinely included. Critically, we have applied the method in eight prospective discovery projects to date, with the results from two of those projects disclosed in this work. The high level of accuracy obtained in the prospective studies demonstrates the ability of this approach to drive decisions in lead optimization.

## FREE-ENERGY PERTURBATION TECHNOLOGY AND METHODOLOGY

The achievement of the results mentioned above is the consequence of an improved force field (OPLS2.1), enhanced sampling, and an automated workflow to ensure that all results are reproducible and realizable with minimal user interaction. Over the past decades, force fields for proteins, nucleic acids, lipids, and other biological molecules have improved substantially via fitting parameters to quantum chemical and experimental data;[21−25] however, adequate parametrical coverage for druglike molecules has lagged behind. For example, MMFF,[26] a widely used force field, is trained against just 140 fragment-sized compounds representing typical organic moieties found in druglike molecules. Our analysis of 1 million purchasable druglike compounds indicates that on the order of tens of thousands of such compounds are required to represent the diversity of even this limited chemical space. Using the OPLS force field[23−25] as a starting point, we have developed a new force field, OPLS2.1,[27] that incorporates a robust model for nonbonded interactions (van der Waals parameters and partial charges) in conjunction with extensive training of torsional and covalent parameters against more than 10,000 representative organic compounds. In addition, missing parameters for any molecule can be generated via an automated algorithm that performs the appropriate quantum mechanics calculations and torsion fitting. The torsional parameters are obtained by constructing model compounds containing the relevant torsional structures and fitting the parameters to quantum chemical data computed at the LMP2/cc-pVTZ(-f) level of theory, which has been shown to yield accurate relative conformational energies for the systems being modeled.[28,29] Ligand atomic partial charges are computed via CM1A-BCC methodology,[30,31] where a substantial number of bond charge corrections for challenging chemistries have been developed. A comparison of the performance of OPLS2.1 relative to MMFF in reproducing quantum chemical torsional profiles and conformational energies is presented in Figure 1. Performance of the nonbonded interaction model has been

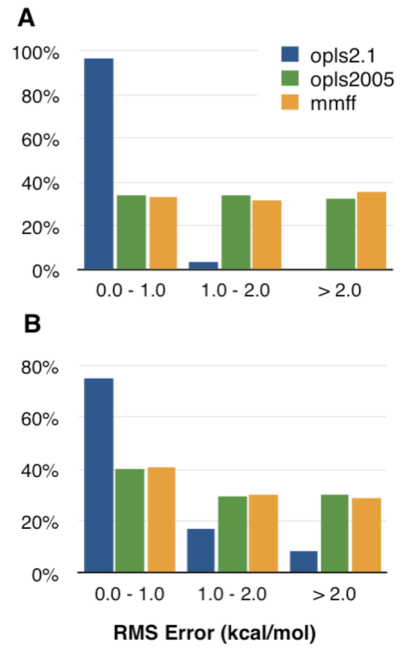



**Figure 1.** Histograms of root-mean-square error in force field relative energies, evaluated (A) over one-dimensional torsional angle scans and (B) between conformational minima, on a set of 8365 compounds. Errors are established with respect to quantum mechanical LMP2/cc-pVTZ(-f) energies evaluated on B3LYP/6-31G* optimized structures. The compound set is generated from a 6 million compound repository of druglike compounds where selected molecules are subsequently fragmented about rotatable torsion bonds retaining key proximate substituents. Selected fragment molecules are chosen such that their constituent rotatable bonds retain sufficient similarity to the OPLS2.1 training set. A rotatable bond is deemed sufficiently similar if the set of atom typed quartets across the bond match a member of the OPLS2.1 training set.

initially evaluated in the prediction of aqueous solvation free energies, the results of which were reported in a prior publication[32] and are summarized in Table 1, along with a comparison to other widely used force fields. These calibration results suggest that OPLS2.1 provides robust force field coverage in the space of druglike ligands and represents a significant advance in this regard as compared to previous general organic ligand force fields.

In addition to the development of accurate potential energy functions, a significant challenge in FEP calculations is ensuring that

**Table 1. Error Statistics for Solvation Free-Energy Results[a]**

| force field | MUE (kcal/mol) | RMSE (kcal/mol) | % > 2 kcal/mol error |
|---|---|---|---|
| ChelpG/CharmM | 1.93 | 2.28 | 44.7 |
| AM1-BCC/ GAFF | 1.17 | 1.39 | 15.0 |
| OPLS 2005 | 1.10 | 1.33 | 8.5 |
| OPLS2.1 | 0.73 | 0.88 | 2.1 |

[a]Reported by Shivakumar et al.[32] using OPLS2.1, OPLS2005, AM1-BCC/GAFF, and ChelpG/CHARMM-MSI for the test set of 239 small molecules. MUE, mean unsigned error; RMSE, root-mean-square error.



DOI: 10.1021/ja512751q
J. Am. Chem. Soc. 2015, 137, 2695−2703

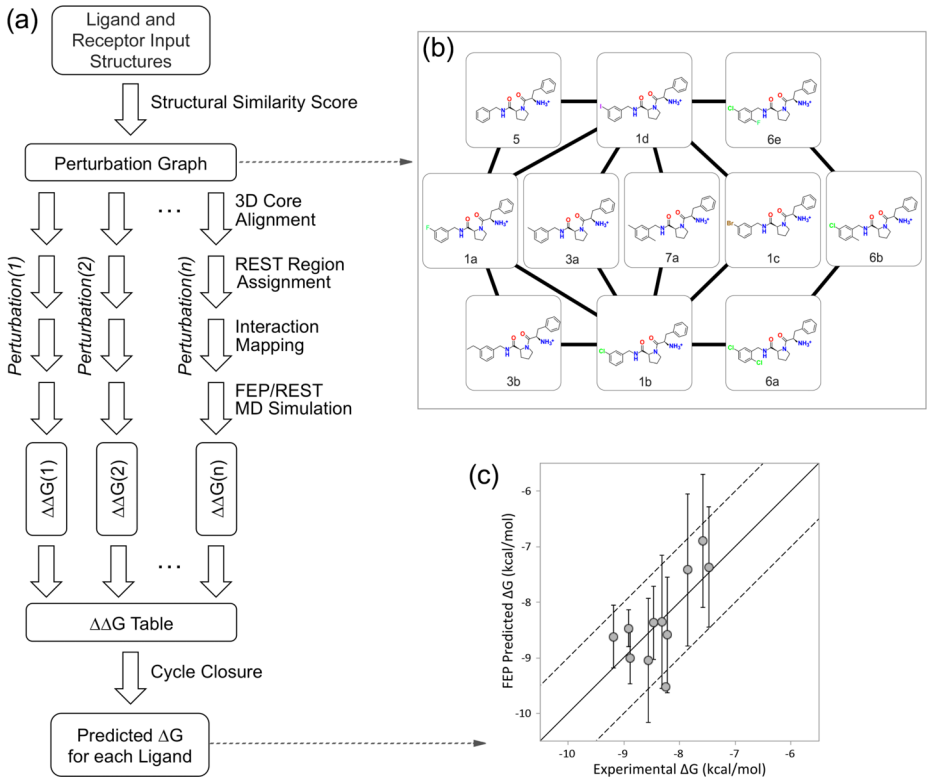



**Figure 2.** (a) Whole FEP workflow for protein−ligand binding-affinity calculations. Key steps in the workflow, including mutation graph generation, REST region assignment, cycle closure convergence, and error estimates, are explicitly shown in the figure. Note that the FEP/REST enhanced sampling method and cycle closure error estimate were reported in prior publications,[37,38] and important related work about perturbation graph generation was also reported in a prior publication;[41] the remaining components of the workflow are novel. (b) Example of mapping of a perturbation space onto a set of pathways for thrombin ligands generated from the workflow. Each line represents two full FEP calculations, one conducted in the receptor and one in solution, each perturbing between two connected ligands. (c) Correlation plot of FEP-predicted and experimental binding affinities for thrombin ligands, as generated from the automated workflow.

the molecular dynamics simulations provide a sufficiently complete sampling of the phase space of the system, while at the same time retaining computational tractability using currently available hardware. A wide range of approaches have been investigated in the literature with the objective of improving the ability to surmount energy barriers and/or reducing computation time in large-scale biomolecular simulations.[16,33] The methodology we have developed retains a full treatment of all degrees of freedom in the system, using a number of methods to substantially enhance the efficiency of phase space exploration, and is readily automated and applicable to a significant fraction of practical drug discovery projects. The key elements of our solution to this problem are outlined below.

First, we employ the Desmond program to run FEP simulations. Desmond provides good single-node and parallel performance, yielding superior performance/cost ratios as compared to alternatives.[34]

Second, we have augmented the molecular dynamics/replica exchange capabilities in Desmond with the newly developed FEP/REST (free energy perturbation/replica exchange with solute tempering) algorithm.[35−37] FEP/REST enables simulations of a selected subsystem with replicas in a higher effective temperature regime than the remainder of the system, and thus precisely focuses sampling efforts where needed to properly traverse the relevant phase space.

Prior results of FEP/REST computations demonstrated a notable improvement in predicting relative ligand binding affinities for two ligands that bind to thrombin. This improvement was accomplished by effectively locally heating the binding region yet retaining rigorous Boltzmann sampling.[37] Additional work demonstrated similarly improved binding predictions for CDK2 ligands[38] and HIV-1 reverse transcriptase inhibitors.[39] The REST methodology thus enables problematic torsional barriers, which can limit ergodicity, to be surmounted in a routine fashion.

A crucial aspect of the use of FEP/REST in practical calculations is the selection of the REST region. Heuristic rules must be developed to determine which atoms in the ligand and the protein environment should be included in the enhanced sampling region. We have developed an automated algorithm to select the REST region, and this algorithm was employed in a uniform fashion in all the studies reported here. Details about the REST region selection algorithm are described in the Methods section in Supporting Information.

Third, Desmond with FEP/REST has been implemented to run on graphics processing units (GPUs). The GPU implementation provides 50−100× speedup over a single central processing unit (CPU) and approximately 5−10× performance/cost improvement as compared to a commodity PC cluster.[40] For a typical FEP calculation (~6000 atoms in the protein) with the protocol described in this work, four perturbations per day can be completed by use of eight commodity Nvidia GTX-780 GPUs, making it feasible to evaluate thousands of molecules per year in the context of a drug discovery program with compute resources that are well within the reach of both academic institutions and commercial enterprises. We also note here, consistent with the experiences reported in ref 40, that GeForce cards do require a significant information technology (IT) commitment for effective use in a production setting.

Another critical aspect of the FEP protocol described here is ease of use, which is essential in order to have a broad impact on drug discovery projects. When considering the modification of a lead molecule, one generally explores a space of possible perturbations at different positions, using a variety of substituents. Prior implementations of FEP methods have generally required large amounts of human time to set up the calculations, and this manual setup is error-prone. In


DOI: 10.1021/ja512751q
J. Am. Chem. Soc. 2015, 137, 2695−2703

**Table 2. Relative Binding Free-Energy Calculation Results**[a]

| | system | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | BACE | CDK2 | JNK1 | MCL1 | p38 | PTP1B | thrombin | Tyk2 |
| no. of compds | 36 | 16 | 21 | 42 | 34 | 23 | 11 | 16 |
| binding affinity range (kcal/mol) | 3.5 | 4.2 | 3.4 | 4.2 | 3.8 | 5.1 | 1.7 | 4.3 |
| crystal structure | 4DJW | 1H1Q | 2GMX | 4HW3 | 3FLY | 2QBS | 2ZFF | 4GIH |
| series ref | 46 | 47 | 48 | 49 | 50 | 51 | 45 | 52,53 |
| no. of perturbations | 58 | 25 | 31 | 71 | 56 | 49 | 16 | 24 |
| MUE FEP | 0.84 ± 0.08 | 0.91 ± 0.12 | 0.78 ± 0.12 | 1.16 ± 0.10 | 0.80 ± 0.08 | 0.89 ± 0.12 | 0.76 ± 0.13 | 0.75 ± 0.11 |
| RMSE FEP | 1.03 ± 0.08 | 1.11 ± 0.12 | 1.00 ± 0.15 | 1.41 ± 0.12 | 1.03 ± 0.09 | 1.22 ± 0.17 | 0.93 ± 0.15 | 0.93 ± 0.12 |
| avg $\sigma_{cc}$ | 0.65 | 0.57 | 0.30 | 0.91 | 0.76 | 0.94 | 0.93 | 0.46 |
| obs *R*-value FEP | 0.78 ± 0.07 | 0.48 ± 0.19 | 0.85 ± 0.07 | 0.77 ± 0.05 | 0.65 ± 0.09 | 0.80 ± 0.08 | 0.71 ± 0.24 | 0.89 ± 0.07 |
| *P*-value FEP | $3.9 \times 10^{-5}$ | $1.2 \times 10^{-2}$ | $7.0 \times 10^{-8}$ | $2.2 \times 10^{-7}$ | $1.6 \times 10^{-7}$ | $7.8 \times 10^{-6}$ | $1.1 \times 10^{-2}$ | $2.3 \times 10^{-7}$ |
| obs *R*-value, MW | 0.14 | −0.48 | −0.39 | −0.55 | −0.46 | −0.84 | −0.48 | 0.00 |
| obs *R*-value, MM-GB/SA | −0.40 | −0.53 | 0.65 | 0.42 | 0.66 | 0.67 | 0.93 | 0.79 |
| obs *R*-value, Glide SP | 0.00 | −0.56 | 0.24 | 0.59 | 0.14 | 0.55 | 0.53 | 0.79 |
| anticip FEP *R*-value | 0.64 ± 0.09 | 0.73 ± 0.11 | 0.64 ± 0.12 | 0.71 ± 0.07 | 0.67 ± 0.08 | 0.79 ± 0.07 | 0.37 ± 0.26 | 0.74 ± 0.10 |
| anticip exptl *R*-value | 0.88 ± 0.03 | 0.92 ± 0.03 | 0.88 ± 0.04 | 0.91 ± 0.02 | 0.89 ± 0.03 | 0.94 ± 0.02 | 0.68 ± 0.15 | 0.92 ± 0.03 |

[a]Eight different receptors, covering a broad range of protein types, were studied. The number of ligands, experimental binding affinity range of ligands, crystal structure used in the simulation, original publication reporting the experimental binding affinity, and number of perturbations for each system are reported. Details about how the data set was selected, and how the experimental binding free energies were obtained, are included in Supporting Information. Several different metrics to assess the performance of FEP results including mean unsigned error (MUE) and root mean square error (RMSE) for all perturbations, correlation coefficient (*R*) between FEP-predicted binding affinities and experimental results, and average error for predictions calculated by cycle closure algorithm (avg $\sigma_{cc}$) are also reported. For comparison, MM-GB/SA and Glide SP scoring results are also reported. The FEP scoring weighted average *R*-value obtained is 0.75, for MM-GB/SA it is 0.35, and for Glide SP it is 0.29. Expected correlation coefficient between FEP-predicted binding affinities and experimental results (anticip FEP *R*-value) and expected correlation coefficient between two experimental measurements of binding affinities (anticip exptl *R*-value), with assumed RMSEs of 1.1 and 0.4 kcal/mol for FEP-predicted binding affinities and experimental data, respectively, are also shown (see details in Supporting Information). Errors for MUE, RMSE, and *R* values by use of the bootstrapping method are also reported. Free energies are in units of kilocalories per mole.

the protocol described here, the user inputs the molecules of interest (in any supported standard format) into a graphical interface, and the perturbation pathways are automatically generated by a variant of the LOMAP mapping algorithm.[41] In the LOMAP algorithm, the maximum common substructure (MCS) between any pair of compounds is generated and their similarity is measured. Then ligand pairs with high similarity scores are connected by edges, where each edge represents one FEP calculation that will be performed between the two ligands. The perturbation graph topology is also optimized such that (1) each edge will, if possible, be nested within at least one closed cycle; and (2) there will be at least one path containing fewer than five edges between any pair of compounds.

Figure 2a shows the automated FEP workflow for protein−ligand binding free-energy calculations, and an example mapping of a ligand series onto a set of pathways is shown in Figure 2b. The 16 separate calculations shown in Figure 2b can be prepared in approximately 30 min, whereas manual setup without a graphical user interface and automated mapping protocols would take significantly longer.

Finally, our approach includes an assessment of the reliability of the calculations, previously a notorious weak point of free-energy methods. The use of multiple pathways, via a cycle closure analysis, enables more reasonable sampling error estimates for the calculations.[38] The estimated error provides an approximation of calculation precision, which is particularly important for the prospective use of the method. Note that force field errors cannot be addressed by any such approach; cycle closure analysis error estimates analyze sampling problems only, that is, they estimate the minimal error in the free-energy results based on the conformational space sampled in all simulations.

## RESULTS AND DISCUSSION

**Validation on Eight Retrospective Data Sets.** We have tested the FEP/REST methodology described above on a diverse set of pharmaceutically relevant targets and ligands (see Table 2). We note that, of the eight data sets reported in the table, one of them (CDK2) was also used in a previous study with the OPLS 2005 force field and a manual setup,[38] and the remaining seven data sets were first studied here. Structures of the individual ligands and the target perturbations used as starting points for the FEP calculations in each data set, as well as other methodological details, are given in Supporting Information. A summary of the performance for all the pairs of perturbations is also provided in Table 2. The combination of high correlations with experimental binding affinity for each system and a low root-mean-square error (RMSE) for all 330 perturbations implies results of sufficient quality to drive decisions in the hit-to-lead and lead-optimization phases of drug discovery projects. Table 3 reports a binned error distribution for all 330 perturbations, indicating a roughly Gaussian distribution with a standard deviation of 1.1 kcal/mol.

**Table 3. Error Distribution for All 330 Perturbations**[a]

| absolute error (kcal/mol) | anticip % | obs % |
|---|---|---|
| <0.5 | 33.9 | 35.5 |
| <1.0 | 62.0 | 63.3 |
| <1.5 | 81.2 | 81.5 |
| <2.0 | 92.1 | 92.4 |
| <2.5 | 97.2 | 96.7 |
| >2.5 | 2.8 | 3.3 |

[a]The distribution is approximately Gaussian with some fattening of the tail of the distribution beyond 2.5 kcal/mol. Fitting of the error distribution by a Gaussian function with the same RMSE is given in Supporting Information, Figure S1. Obs % is the percentage of FEP perturbations found to be accurate within the specified absolute error. Anticip % is the percentage of FEP perturbations that would be expected given an underlying root-mean-square error of 1.1 kcal/mol and an ideal Gaussian error distribution.


DOI: 10.1021/ja512751q
J. Am. Chem. Soc. 2015, 137, 2695−2703

MM-GB/SA[42] and Glide SP[43] scoring results are also reported in Table 2. The FEP scoring weighted average *R*-value obtained for the series reported in Table 2 is 0.75, for MM-GB/SA it is 0.35, and for Glide SP it is 0.29. As expected, the Glide SP scoring results fail to capture many of the structure−actibity relationship (SAR) trends, since the Glide SP scoring function was developed primarily for virtual screening applications rather than lead optimization. Perhaps more surprising is the significantly worse MM-GB/SA scoring results, where the MM-GB/SA scoring results are actually anticorrelated with experimental data for two of the series. Several of the series reported here require flips of side chains lining the active site and the ligand R-groups to fully describe the SAR, which poses a severe challenge for rigid-receptor MM-GB/SA and docking calculations.

The scatter plot of predicted versus experimental binding energies for the entire data set is shown in Figure 3. Some of

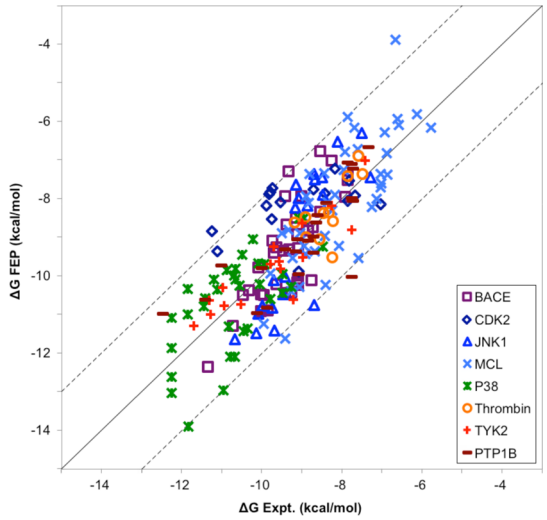



**Figure 3.** Correlation between FEP-predicted binding free energies and experimental data for all eight systems studied. FEP-predicted binding free energies for most of the ligands are within 1.0 kcal/mol of their experimental values, and only nine of 199 studied ligands deviate from their experimental free energies by more than 2 kcal/mol.

the variance can be attributed to experimental noise, which is typically on the order of 0.3−0.5 kcal/mol for high-quality binding free-energy measurements.[44] If it is assumed that the experimental measurements for binding free energies of two compounds are independent, that is, $\sigma_{ij}^2 = \sigma_i^2 + \sigma_j^2$, then the experimental error for relative binding free energies between the two compounds is about 0.4−0.7 kcal/mol. Thus, the RMSEs for any prediction method compared with experimental data cannot be less than the experimental noise of 0.4−0.7 kcal/mol. Here, the RMSE of the reported FEP/REST method is about 1.1 kcal/mol, implying the residual error to be about 0.85−1.0 kcal/mol. Hence, the experimental noise is expected to be a significant contribution to the dispersion seen in Figure 3, possibly as great as 30% of the total observed error. Interestingly, predictions that approach this expected practical accuracy limit are obtained for tyk2 and thrombin,[45] likely due to the rigidity of the systems, the small size of the perturbations, and the high quality of the experimental data (for thrombin, via isothermal calorimetry). The scatter plot of predicted versus experimental binding free energy for the thrombin data set is depicted in Figure 2c, including the cycle closure convergence error bars. As can be seen from the plot, the convergence error for this system is very small, and all predictions are uniformly within ∼1 kcal/mol of the experimental values. Nevertheless, the *R*-value obtained is a relatively modest value of 0.71. This underscores an important point regarding data analysis in general: a correlation coefficient *R* as a measure of computational model performance performs poorly if the dynamic range of the data is narrow. In particular, if an RMSE for the experimental technique of 0.4 kcal/mol is assumed, the correlation between repeated experimental measurements would be expected to produce *R*-values of only 0.68 ± 0.3 at a 95% confidence interval.[44] As an extreme example, in the limit of all the compounds being equipotent, obtaining a predictive *R*-value becomes impossible, no matter what modeling accuracy is obtained.

As mentioned in the Introduction, FEP is among the most rigorous methods available to calculate the relative binding free energies between congeneric ligands, and as such, it should capture changes in binding free energy driven by variable physical factors, including hydrophobic interactions, hydrogen-bonding interactions, solvent effects such as the displacement of water molecules, motional entropic effects, and so on, up to the limits of those interactions being well-described by the employed classical force field and the simulations being converged. Representative examples of the diversity of structure−activity relationships successfully captured by FEP scoring are shown in Figure 4.

A representative example of hydrophobic interactions contributing to binding can be observed for the pair of MCL1 ligands depicted in Figure 4A. The naphthalene ring in the first ligand (left panel) is found to make favorable hydrophobic interactions with surrounding protein residues, including Val249, Leu246, Leu235, Leu290, Val274, Ile182, and Phe270. When the naphthalene ring is mutated into a quinoline ring in the second ligand (right panel), the loss of favorable hydrophobic interactions results in a decrease in the experimental binding free energy by 1.3 kcal/mol. FEP captures this effect, resulting in a computed free energy difference between these two ligands of 1.6 kcal/mol.

A representative example of protein−ligand electrostatic interactions driving more favorable binding potency can be found for the pair of PTP1B ligands depicted in Figure 4B. Here the aniline group in the first ligand (left panel) is found to form a favorable hydrogen-bonding interaction with Gln262. When the amino group is mutated into a methoxy group in the second ligand (right panel), the experimental binding free energy is observed to decrease by 0.5 kcal/mol, presumably due to the loss of hydrogen-bonding interactions. The FEP calculation here correctly captures this binding affinity decrease ($\Delta\Delta G_{FEP}$ = 0.4 kcal/mol).

A representative example of the expulsion of a single water molecule from the active site contributing to binding is depicted for thrombin ligands in Figure 4C. In the bound complex structure of the second ligand (right panel), the water molecule occupying the S1 pocket is energetically unfavorable due to steric confinement and the lack of opportunities to make favorable polar interactions with the surrounding protein residues or the ligand. With the addition of a chloro group at the meta position of the phenyl ring in the first ligand (left panel), the high-energy water molecule in the S1 pocket is displaced, resulting in a more favorable binding free energy for the second ligand. FEP calculations appear to correctly capture the binding affinity increase due to displacement of the high


DOI: 10.1021/ja512751q
J. Am. Chem. Soc. 2015, 137, 2695−2703

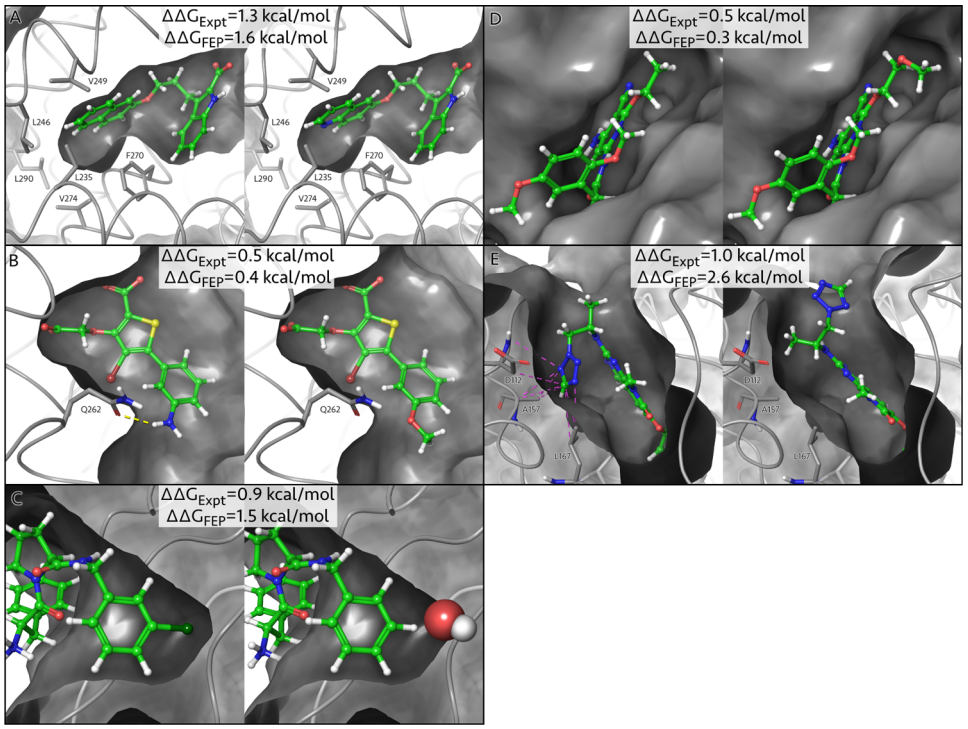



**Figure 4.** Representative examples of different types of interactions captured by FEP (see text for details).

free-energy water in the S1 pocket, in good agreement with experiment ($\Delta\Delta G_{FEP}$ = 1.5 kcal/mol versus $\Delta\Delta G_{expt}$ = 0.9 kcal/mol).

A representative example of the addition of a large flexible solvent-exposed functional group decreasing the binding potency is depicted for the JNK1 kinase ligands in Figure 4D. The solvent-exposed methoxy group in the second ligand (right panel) makes no obvious direct favorable interactions with the protein but can sample less volume when bound to the protein. As such, we believe the loss of the mobility of this group upon binding to the receptor may lead to entropic penalties that weaken the binding of ligand to receptor. FEP simulations capture both the direction and magnitude of the effect ($\Delta\Delta G_{expt}$ = 0.5 kcal/mol and $\Delta\Delta G_{FEP}$ = 0.3 kcal/mol).

A representative example of the inversion of a stereocenter, changing both the affinity and binding mode of a molecule, is depicted for p38 ligands in Figure 4E (PDB codes 3FMH and 3FMK). Although the unbound ensembles of the two ligands are necessarily nearly identical, the *S*-variant (left panel) is able to bind such that seemingly favorable interactions are formed between the ligand tetrazole functional group and nearby Ala111, Asp112, and Ala157. In contrast, the *R*-variant (right panel) is forced to present the tetrazole to a more solvent-exposed region, diminishing the binding potency of that species. The basis for the favorability of interactions of the tetrazole with the backbone amide of Ala111 and Asp112 may be somewhat counterintuitive, but we suspect the hydrophobic desolvation of $\beta$-sheets is favorable above and below the plane of the interacting amides, leading to favorable hydrophobic binding of the tetrazole to the amide in this particular case. Although the magnitude of the effect is somewhat overestimated by the calculation, the tighter binding of the *S*-variant is captured by the calculation, where $\Delta\Delta G_{expt}$ = 1.0 kcal/mol and $\Delta\Delta G_{FEP}$ = 2.6 kcal/mol.

This final case illustrates another point regarding the value of FEP scoring in practical applications: the maximum size of the perturbations that can be reliably treated is of equal significance to obtaining predictive correlation and small RMS errors. Much of the FEP literature is devoted to examining the effect of small perturbations: often only a single atom is changed, such as hydrogen to halogen, since it is assumed that such a small change should lead to only very small changes in the relevant ensemble of states. Such perturbations may in many cases make sampling much less challenging, but in an active drug discovery project, much more substantial changes to the ligand will be desired, and as the preceding case demonstrates, significant sampling challenges may still exist for perturbations of only one or a few atoms. Therefore, we have deliberately chosen ligand series containing chemical modifications across a range of sizes that are of interest in typical drug design studies. For example, representative ligand perturbations for p38 MAP kinase are shown in Supporting Information Table S1, and the input files for each of the perturbations for each series are available in Supporting Information. We find that our methodology is robust up to perturbations of approximately 10 heavy atoms.

**Prospective Studies on Two Active Drug Design Projects.** To further substantiate the above findings, we recently applied this technology prospectively in eight active drug discovery projects. We present here results from two of these projects. The first is focused on developing selective, druglike inhibitors for IRAK4 (project I), and the second is focused on developing inhibitors for TYK2 (project II). In project I, FEP was introduced into the project at a stage where several hundred compounds had already been synthesized and the target biochemical affinity ($K_i$ ~10 nM) and a number of the absorption, distribution, metabolism, and excretion (ADME) properties had been achieved. What remained was further optimization of some of the ADME properties while maintaining affinity at 10 nM or better ($pK_i$ >8). At this stage of


DOI: 10.1021/ja512751q
J. Am. Chem. Soc. 2015, 137, 2695−2703

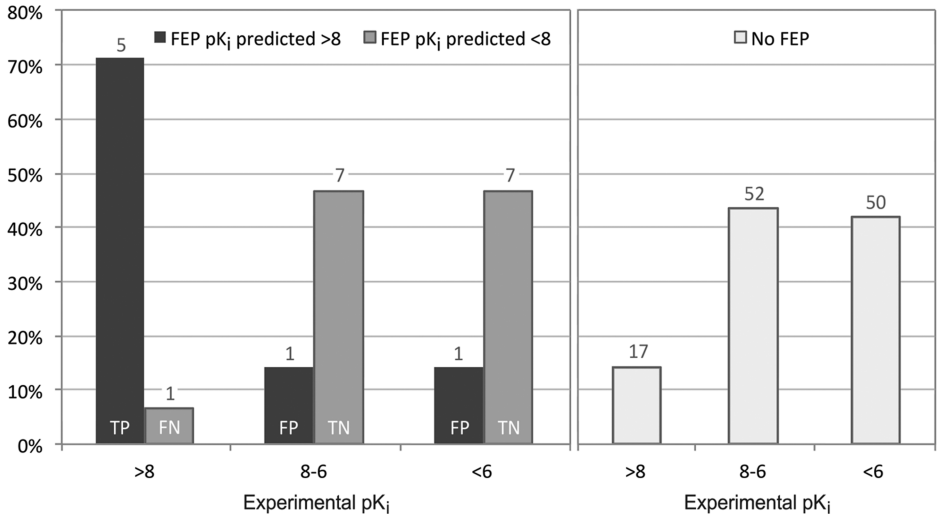



**Figure 5.** Histograms showing distribution of experimental values for compounds in project I that were predicted by FEP to have p$K_i$ > 8 (dark gray), predicted by FEP to have p$K_i$ < 8 (medium gray), and those that were not computationally predicted prior to being assayed (light gray). Numbers above the bars correspond to the actual number of compounds that were assayed. Approximately 14% of the compounds that were chosen to be synthesized and assayed without guidance from FEP had p$K_i$ > 8, while 71% of the compounds predicted by FEP to have p$K_i$ > 8 had an experimental p$K_i$ > 8. Key to labels: TP = true positive, FN = false negative, FP = false positive, TN = true negative.

the project and over a period of several months, 195 compounds were prospectively scored with FEP, and 22 were synthesized and assayed. The results of these predictions are shown in Figure 5. In total, 156 of the 195 compounds were predicted to have p$K_i \leq 8$, and the other 39 were predicted to have p$K_i$ > 8. Fifteen of the compounds predicted to have p$K_i \leq 8$ were, despite the predictions, synthesized and assayed to test various ADME hypotheses. As shown in Figure 5, 14 of these compounds (93%) turned out to have p$K_i \leq 8$, as predicted (only one false negative was found). Given this true negative rate of 93%, it is expected that ∼145 of the 156 compounds predicted to have p$K_i \leq 8$ would have been true negatives, while only ∼11 (7%) would have been false negatives (p$K_i$ > 8). Of the seven synthesized compounds predicted to have p$K_i$ > 8, five did have an experimental p$K_i$ > 8 and two did not (71% true positive rate). This constitutes a 6-fold enrichment in the synthesis of tight-binding molecules: only 12% of the compounds that were not prioritized by the FEP calculations were found to have p$K_i$ > 8, while 71% (5 out of 7) of those compounds that were predicted by FEP were correctly predicted to fall in this range. Note that a significant fraction of the 119 compounds synthesized without the benefit of FEP scoring were expected to be potent on the basis of more conventional analyses. Thus, the observed 6-fold enrichment in the synthesis of tight-binding molecules provides suggestive evidence that FEP scoring provides a substantial reduction in false positives relative to compound synthesized on the basis of other approaches.

The accuracy of the results in project II was similar to those observed in project I. In project I, the average error between predicted and experimental affinity for the 22 compounds was 1.1 kcal/mol (average p$K_i$ error of 0.8). In project II, the average error for 20 compounds that were prospectively predicted and experimentally assayed was 0.9 kcal/mol (average p$K_i$ error of 0.7). Thirty-seven compounds with p$K_i$ predictions $\leq 8$ were not synthesized; the true negative rate in project II was 75% based on results for four compounds predicted to have p$K_i \leq 8$ that were subsequently synthesized.

These accurate affinity predictions in both projects allowed the teams to reliably deprioritize a large number of proposed compounds and to focus synthesis and assay resources on efficiently achieving project potency and ADME goals.

## CONCLUSIONS

The work described here addresses several major challenges to using FEP in drug discovery programs. The OPLS2.1 force field has considerably greater chemical space coverage than other widely used force fields and provides sufficient energetic accuracy for meaningful prospective free-energy calculations. Furthermore, the efficient FEP implementation reported herein extends the Desmond/GPU molecular dynamics engine to incorporate REST enhanced sampling, thereby improving simulation convergence. The combination of improved force field and superior sampling method has contributed to improved accuracy of the FEP protocol. Additionally, the fully automated FEP calculation setup and simulation quality analysis reduce human error and workload, thus making the approach accessible to a broad population of researchers in drug discovery. For a typical-sized drug target, four perturbations can be completed by use of eight commodity Nvidia GTX-780 GPUs. The aggregate effect of these advances now positions free-energy calculations to play a guiding role in the hit-to-lead and lead-optimization phases, as indicated by encouraging results in the two active drug discovery projects presented here.

The preceding notwithstanding, a highly accurate and robust FEP methodology is not, in any way, a replacement for a creative and technically strong medicinal chemistry team; it is necessary to generate the ideas for optimization of the lead compound that are synthetically tractable and have acceptable values for a wide range of druglike properties (e.g., solubility, membrane permeability, metabolism, etc.). Rather, the computational approach described here can be viewed as a tool to enable medicinal chemists to pursue modifications and new synthetic directions that would have been considered too risky without computational validation or to eliminate compounds that would be unlikely to meet the desired target


DOI: 10.1021/ja512751q
J. Am. Chem. Soc. 2015, 137, 2695−2703

affinity. This is particularly significant when considering whether to make an otherwise highly attractive molecule that may be synthetically challenging. If such a molecule is predicted to achieve the project potency targets by reliable FEP calculations, this substantially reduces the risk of taking on such synthetic challenges. In addition, the elimination of compounds unlikely to meet project potency targets frees resources to focus on more promising compounds. Thus, extensive deployment of FEP in a drug discovery project not only will reduce the number of compounds that are made with inadequate activity but also may facilitate significant leaps in chemical space that otherwise would not have been taken, leading to more rapid completion of difficult projects, with potentially superior molecules as an end result.

## ■ ASSOCIATED CONTENT

### Ⓢ Supporting Information

Additional text and equations with details of REST region selection algorithm, simulations, OPLS2.1 force field, conversion of calculated $\Delta\Delta G$ values to $\Delta G$ values, expected FEP prediction accuracy in prospective studies, expected correlation coefficient between FEP-predicted binding affinities and experimental values and between two independent experimental measurements, experimental binding affinity data, and input structures for FEP calculations; three tables listing representative p38 ligand pairs used in FEP/REST calculations, comparison of RMSE for FEP-predicted binding affinities for all 330 pairs of ligands and RMSE assuming all ligands are equally potent, and number of OPLS2.1 missing torsions identified for three ligand sets; and two figures showing error distribution of FEP-predicted relative binding free energies compared to experimental data and its fitting by a Gaussian function and histograms of the RMSE in force-field relative energies (pdf). Experimental and predicted $\Delta G$ values and errors for all eight data sets (xls). This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION


### Corresponding Author

*robert.abel@schrodinger.com


### Notes

The authors declare the following competing financial interest(s): D.L.M., W.L.J., and B.J.B. are consultants to Schrodinger, Inc. and are on its Scientific Advisory Board. R.A.F. has a significant financial stake in, is a consultant for, and is on the Scientific Advisory Board of Schrodinger, Inc.

## ■ REFERENCES


(1) Jorgensen, W. L. *Acc. Chem. Res.* **2009**, *42*, 724.
(2) Jorgensen, W. L. *Science* **2004**, *303*, 1813.
(3) Homeyer, N.; Stoll, F.; Hillisch, A.; Gohlke, H. *J. Chem. Theory Comput.* **2014**, *10*, 3331.
(4) *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Chipot, C., Pohorille, A., Eds.; Springer Series in Chemical Physics, Vol. *86*; Springer: Berlin and Heidelberg, Germany, 2007.
(5) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150.
(6) Knight, J. L.; Brooks, C. L. *J. Comput. Chem.* **2009**, *30*, 1692.
(7) Zheng, L.; Chen, M.; Yang, W. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20227.
(8) Gallicchio, E.; Levy, R. M. *Curr. Opin. Struct. Biol.* **2011**, *21*, 161.
(9) Hansen, N.; van Gunsteren, W. F. *J. Chem. Theory Comput.* **2014**, *10*, 2632.
(10) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585.
(11) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050.
(12) Bash, P.; Singh, U.; Brown, F.; Langridge, R.; Kollman, P. *Science* **1987**, *235*, 574.
(13) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395.
(14) Wong, C. F.; McCammon, J. A. *J. Am. Chem. Soc.* **1986**, *108*, 3830.
(15) Merz, K. M.; Kollman, P. A. *J. Am. Chem. Soc.* **1989**, *111*, 5649.
(16) Deng, Y.; Roux, B. *J. Phys. Chem. B* **2009**, *113*, 2234.
(17) Gallicchio, E.; Lapelosa, M.; Levy, R. M. *J. Chem. Theory Comput.* **2010**, *6*, 2961.
(18) Durrant, J.; McCammon, J. *BMC Biol.* **2011**, *9*, 1.
(19) Riniker, S.; Christ, C.; Hansen, H.; Hünenberger, P.; Oostenbrink, C.; Steiner, D.; van Gunsteren, W. *J. Phys. Chem. B* **2011**, *115*, 13570.
(20) Michel, J.; Essex, J. W. *J. Med. Chem.* **2008**, *51*, 6654.
(21) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
(22) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.
(23) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
(24) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
(25) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474.
(26) Halgren, T. A.; Nachbar, R. B. *J. Comput. Chem.* **1996**, *17*, 587.
(27) *OPLS 2.1*; Schrodinger, Inc.: New York, 2014.
(28) Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A. *Int. J. Quantum Chem.* **2013**, *113*, 2110.
(29) Murphy, R. B.; Beachy, M. D.; Friesner, R. A.; Ringnalda, M. N. *J. Chem. Phys.* **1995**, *103*, 1481.
(30) Storer, J.; Giesen, D.; Cramer, C.; Truhlar, D. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87.
(31) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623.
(32) Shivakumar, D.; Harder, E.; Damm, W.; Friesner, R. A.; Sherman, W. *J. Chem. Theory Comput.* **2012**, *8*, 2553.
(33) Jorgensen, W. L.; Tirado−Rives, J. *J. Comput. Chem.* **2005**, *26*, 1689.
(34) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. In *Proceedings of the 2006 ACM/IEEE conference on Supercomputing*, Tampa, FL, 2006; p 84.
(35) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749.
(36) Wang, L.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem. B* **2011**, *115*, 9431.
(37) Wang, L.; Berne, B. J.; Friesner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 1937.
(38) Wang, L.; Deng, Y.; Knight, J. L.; Wu, Y.; Kim, B.; Sherman, W.; Shelley, J. C.; Lin, T.; Abel, R. *J. Chem. Theory Comput.* **2013**, *9*, 1282.
(39) Cole, D. J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2014**, 565.
(40) Bergdorf, M.; Kim, E. T.; Rendleman, C. A.; Shaw, D. E. D. E. Shaw Research, Technical Report DESRES/TR--2014-01, 2014.
(41) Liu, S.; Wu, Y.; Lin, T.; Abel, R.; Redmann, J.; Summa, C.; Jaber, V.; Lim, N.; Mobley, D. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 755.
(42) *Prime*, version 3.8; Schrodinger Inc., New York, 2014.
(43) *Glide*, version 6.5; Schrodinger Inc., New York, 2014.



DOI: 10.1021/ja512751q
*J. Am. Chem. Soc.* 2015, 137, 2695−2703

(44) Brown, S. P.; Muchmore, S. W.; Hajduk, P. J. *Drug Discovery Today* **2009**, *14*, 420.
(45) Baum, B.; Mohamed, M.; Zayed, M.; Gerlach, C.; Heine, A.; Hangauer, D.; Klebe, G. *J. Mol. Biol.* **2009**, *390*, 56.
(46) Cumming, J. N.; Smith, E. M.; Wang, L.; Misiaszek, J.; Durkin, J.; Pan, J.; Iserloh, U.; Wu, Y.; Zhu, Z.; Strickland, C.; Voigt, J.; Chen, X.; Kennedy, M. E.; Kuvelkar, R.; Hyde, L. A.; Cox, K.; Favreau, L.; Czarniecki, M. F.; Greenlee, W. J.; McKittrick, B. A.; Parker, E. M.; Stamford, A. W. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 2444.
(47) Hardcastle, I. R.; Arris, C. E.; Bentley, J.; Boyle, F. T.; Chen, Y.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R. J.; Jewsbury, P.; Menyerol, J.; Mesguiche, V.; Newell, D. R.; Noble, M. E.; Pratt, D. J.; Wang, L. Z.; Whitfield, H. J. *J. Med. Chem.* **2004**, *47*, 3710.
(48) Szczepankiewicz, B. G.; Kosogof, C.; Nelson, L. T. J.; Liu, G.; Liu, B.; Zhao, H.; Serby, M. D.; Xin, Z.; Liu, M.; Gum, R. J.; Haasch, D. L.; Wang, S.; Clampit, J. E.; Johnson, E. F.; Lubben, T. H.; Stashko, M. A.; Olejniczak, E. T.; Sun, C.; Dorwin, S. A.; Haskins, K.; Abad-Zapatero, C.; Fry, E. H.; Hutchins, C. W.; Sham, H. L.; Rondinone, C. M.; Trevillyan, J. M. *J. Med. Chem.* **2006**, *49*, 3563.
(49) Friberg, A.; Vigil, D.; Zhao, B.; Daniels, R. N.; Burke, J. P.; Garcia-Barrantes, P. M.; Camper, D.; Chauder, B. A.; Lee, T.; Olejniczak, E. T.; Fesik, S. W. *J. Med. Chem.* **2012**, *56*, 15.
(50) Goldstein, D. M.; Soth, M.; Gabriel, T.; Dewdney, N.; Kuglstatter, A.; Arzeno, H.; Chen, J.; Bingenheimer, W.; Dalrymple, S. A.; Dunn, J.; Farrell, R.; Frauchiger, S.; La Fargue, J.; Ghate, M.; Graves, B.; Hill, R. J.; Li, F.; Litman, R.; Loe, B.; McIntosh, J.; McWeeney, D.; Papp, E.; Park, J.; Reese, H. F.; Roberts, R. T.; Rotstein, D.; San Pablo, B.; Sarma, K.; Stahl, M.; Sung, M.-L.; Suttman, R. T.; Sjogren, E. B.; Tan, Y.; Trejo, A.; Welch, M.; Weller, P.; Wong, B. R.; Zecic, H. *J. Med. Chem.* **2011**, *54*, 2255.
(51) Wilson, D. P.; Wan, Z.-K.; Xu, W.-X.; Kirincich, S. J.; Follows, B. C.; Joseph-Mccarthy, D.; Foreman, K.; Moretto, A.; Wu, J.; Zhu, M.; Binnun, E.; Zhang, Y.-L.; Tam, M.; Erbe, D. V.; Tobin, J.; Xu, X.; Leung, L.; Shilling, A.; Tam, S. Y.; Mansour, T. S.; Lee, J. *J. Med. Chem.* **2007**, *50*, 4681.
(52) Liang, J.; Tsui, V.; Van Abbema, A.; Bao, L.; Barrett, K.; Beresini, M.; Berezhkovskiy, L.; Blair, W. S.; Chang, C.; Driscoll, J.; Eigenbrot, C.; Ghilardi, N.; Gibbons, P.; Halladay, J.; Johnson, A.; Kohli, P. B.; Lai, Y.; Liimatta, M.; Mantik, P.; Menghrajani, K.; Murray, J.; Sambrone, A.; Xiao, Y.; Shia, S.; Shin, Y.; Smith, J.; Sohn, S.; Stanley, M.; Ultsch, M.; Zhang, B.; Wu, L. C.; Magnuson, S. *Eur. J. Med. Chem.* **2013**, *67*, 175.
(53) Liang, J.; van Abbema, A.; Balazs, M.; Barrett, K.; Berezhkovsky, L.; Blair, W.; Chang, C.; Delarosa, D.; DeVoss, J.; Driscoll, J.; Eigenbrot, C.; Ghilardi, N.; Gibbons, P.; Halladay, J.; Johnson, A.; Kohli, P. B.; Lai, Y.; Liu, Y.; Lyssikatos, J.; Mantik, P.; Menghrajani, K.; Murray, J.; Peng, I.; Sambrone, A.; Shia, S.; Shin, Y.; Smith, J.; Sohn, S.; Tsui, V.; Ultsch, M.; Wu, L. C.; Xiao, Y.; Yang, W.; Young, J.; Zhang, B.; Zhu, B.-y.; Magnuson, S. *J. Med. Chem.* **2013**, *56*, 4521.


DOI: 10.1021/ja512751q
*J. Am. Chem. Soc.* 2015, 137, 2695−2703