

UC Irvine

UC Irvine Previously Published Works

Title

Triazole-dithiocarbamate based selective lysine specific demethylase 1 (LSD1) inactivators inhibit gastric cancer cell growth, invasion, and migration.

Permalink

<https://escholarship.org/uc/item/0bk4n198>

Journal

Journal of medicinal chemistry, 56(21)

ISSN

0022-2623

Authors

Zheng, Yi-Chao
Duan, Ying-Chao
Ma, Jin-Lian
[et al.](#)

Publication Date

2013-11-01

DOI

10.1021/jm401002r

Peer reviewed

Accurate and Reliable Prediction of Relative Protein-Ligand Binding Affinities via Free Energy Calculations: Validation in Prospective Drug Discovery

Authors: Lingle Wang,¹ Yujie Wu,¹ Yuqing Deng,¹ Byungchan Kim,¹ Levi Pierce,¹ Goran Krilov,¹ Dmitry Lupyan,¹ Shaughnessy Robinson,¹ Markus Dahlgren,¹ Jeremy Greenwood,¹ Donna L. Romero,² Craig Masse,² Jennifer L. Knight,¹ Thomas Steinbrecher,¹ Thijs Beuming,¹ Wolfgang Damm,¹ Ed Harder,¹ Woody Sherman,¹ Mark Brewer,¹ Ron Wester,² Mark Murcko,¹ Leah Frye,¹ Ramy Farid,¹ Teng Lin,¹ David L. Mobley,⁵ William L. Jorgensen,⁴ Bruce J. Berne,³ Richard A. Friesner,³ Robert Abel^{1,*}

¹Schrödinger, Inc., 120 West 45th Street, New York, NY 10036, United States

²Nimbus Discovery, 25 First Street, Suite 404 Cambridge, MA 02141, United States

³Department of Chemistry, Columbia University, 3000 Broadway, New York, NY, 10027, United States

⁴Department of Chemistry, Yale University, New Haven, CT 06520, United States

⁵Department of Pharmaceutical Sciences and Chemistry, University of California—Irvine, Irvine, CA 92697, United States

*Corresponding author

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 (~5X 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, 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 based on 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 the ligand to the 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.³⁻⁷ 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 affords significant reduction in computational effort as compared to absolute binding free energy calculations.

The initial applications of free energy methods to the calculation of protein-ligand binding affinities appeared shortly after the first molecular dynamics simulation of a protein by McCammon, Karplus, and coworkers in 1977.⁸⁻¹⁰ Subsequent efforts over the next two decades reported anecdotal results for a small number of protein-ligand complexes, but suffered from a lack of computing power and inadequacies in both sampling algorithms and molecular mechanics force fields.^{1,4,7} 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 works.^{1,4,7,11-15} 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 ten heavy atoms routinely included. Critically, we have applied the method in 8 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.

FEP 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,¹⁶⁻²⁰ however, adequate parametrical

coverage for drug-like molecules has lagged behind. For example, MMFF,²¹ a widely used force field, is trained against just 140 fragment sized compounds representing typical organic moieties found in drug-like molecules. Our analysis of one million purchasable drug-like 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 field¹⁸⁻²⁰ as a starting point, we have developed a new force field, OPLS2.1, that incorporates a robust model for non-bonded 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.^{22,23} Ligand atomic partial charges are computed via a CM1A-BCC methodology^{24,25} 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 Fig. 1. The performance of the non-bonded interaction model has been initially evaluated in the prediction of aqueous solvation free energies, the results of which are reported in a prior publication²³ and 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 drug-like 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 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.^{11,26} 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.²⁷

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.²⁸⁻³⁰ 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 binding to thrombin. This improvement was accomplished by effectively locally heating the binding region, yet retaining rigorous Boltzmann sampling.³⁰ Additional work demonstrated similarly improved binding predictions for CDK2 ligands³¹ and HIV-

1 reverse transcriptase inhibitors.³² 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.

Third, Desmond with FEP/REST has been implemented to run on graphics processing units (GPUs). The GPU implementation provides 50-100x speedup over a single CPU and approximately a 5-10x performance/cost improvement as compared to a commodity PC cluster. For a “typical” FEP calculation (~6,000 atoms in the protein) with the protocol described in this work, 4 perturbations per day can be completed using 8 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.

Another critical aspect of the FEP protocol described here is the 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 the protocol described here, the user inputs the molecules of interest (in any of the standard formats) into a graphical interface and the perturbation pathways are automatically generated using a variant of the LOMAP mapping algorithm.³³ Fig. 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 Fig. 2b. The 16 separate calculations shown in Fig. 2b can be prepared in approximately 30 minutes, 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.³¹ The estimated error provides an approximate 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, i.e. they estimate the minimal error in the free energy results based on the conformational space sampled in all simulations.

Results and discussion

Validation on 8 Retrospective Datasets

We have tested the FEP/REST methodology described above on a diverse set of pharmaceutically relevant targets and ligands (see Table 2). 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 the 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 squared 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.

The scatter plot of predicted versus experimental binding energies for the entire data set is shown in Fig. 3. Some of 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.³⁴ This suggests a practical accuracy limit for relative binding free energy calculations on the order of 0.4-0.7 kcal/mol, which is the average experimental noise when comparing pairs of compounds in common experimental assays. Hence, this noise is expected to be a significant contribution to the dispersion seen in Fig. 3. Interestingly, predictions to within this expected practical accuracy limit are actually obtained for the case of thrombin,³⁵ likely due to the rigidity of the systems, the small size of the perturbations, and the high quality of the experimental data, which was obtained via isothermal calorimetry (ITC). The scatter plot of predicted versus experimental binding free energy for the thrombin data set is depicted in Fig. 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, assuming an RMSE of the experimental technique of 0.4 kcal/mol, the correlation between repeated experimental measurements would be expected to produce R values of only 0.68 ± 0.3 at a 95% confidence interval.³⁴ 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 to calculate the relative binding free energies between congeneric ligands, and as such it should capture the binding free energy difference resulting from all different types of interactions, 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 Fig. 4.

A representative example of hydrophobic interactions contributing to binding can be observed for the pair of MCL1 ligands depicted in Fig. 4A. The naphthalene ring in 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 mutating the naphthalene ring into a quinolone ring in 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 appears to capture this effect, resulting in a computed free energy difference between these two ligands of 1.6 kcal/mol.

A similar representative example of protein-ligand electrostatic interactions driving a more favorable binding potency can be found the PTP1B ligands depicted in Fig. 4B. Here the aniline group in the first ligand (left panel) is found to form a favorable hydrogen bonding interaction with Gln262. When mutating the aniline into a methoxy-group in 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 ($\Delta\Delta G_{\text{FEP}} = 0.4$ kcal/mol).

A representative example of the expulsion of a single water molecule from the active site contributing to binding is depicted for thombin ligands in Fig. 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. The FEP calculations appear to correctly capture the binding affinity increase due to the displacement of the high free energy water in the S1 pocket in good agreement with experiment ($\Delta\Delta G_{\text{FEP}} = 1.5$ kcal/mol versus $\Delta\Delta G_{\text{Expt}} = 0.9$ kcal/mol).

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

A representative example of the inversion of a stereo-center both changing the affinity and binding mode of a molecule is depicted for p38 ligands in Fig. 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 of the favorability of the 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 affect is somewhat over estimated by the calculation, the tighter binding of the S-variant is captured by the calculation, where $\Delta\Delta G_{\text{Expt}} = 1.0$ kcal/mol and $\Delta\Delta G_{\text{FEP}} = 2.6$ 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 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 the range of sizes that are of interest in typical drug design studies. For example, some ligand perturbations for p38 MAP kinase are shown in SI Table S1 and a detailed description of the perturbations for each series is presented in the Supplementary Information. We find that our methodology is robust up to perturbations of approximately 10 heavy atoms.

Prospective Studies on 2 Active Drug Design Projects

To further substantiate the above findings, we recently applied this technology prospectively in 8 active drug discovery projects. We present here results from two of these projects. The first is focused on developing selective, drug-like 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 ($K_i \sim 10$ nM) and a number of the ADME properties had been achieved, and 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 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 Fig. 5. 156 of the 195 compounds were predicted to have a $pK_i \leq 8$, and the other 39 were predicted to have a $pK_i > 8$. 15 of the compounds predicted to have a $pK_i \leq 8$ were, despite the predictions, synthesized and assayed to test various hypotheses. As shown in Fig. 5, 14 of these compounds (93%) turned out to have a $pK_i \leq 8$, as predicted (only 1 false negative was found). Given this true negative rate of 93%, it is expected that ~ 145 of the 156 compounds predicted to have a $pK_i \leq 8$ would have been true negatives, while only ~ 11 (7%) would have been false negatives ($pK_i > 8$). Of the 7 compounds predicted to have a $pK_i > 8$, 5 did have an experimental $pK_i > 8$ and 2 did not (71% true positive rate). This constitutes a 6-fold enrichment in the synthesis of tight binding molecules; where only 12% of the compounds that were not prioritized by the FEP calculations were found to have a $pK_i > 8$, while 71% (5 out of 7) of the compounds that were predicted by FEP were correctly predicted to fall in this range.

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 pK_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 pK_i error of 0.7). 37 compounds with pK_i predictions of ≤ 8 were not synthesized; the true negative rate in Project II was 75% based on results for 4 compounds predicted to have a $pK_i \leq 8$ and 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. Further, the efficient FEP implementation reported herein extends the Desmond/GPU molecular dynamics engine to incorporate REST enhanced sampling, thereby improving simulation convergence. Additionally, the fully automated FEP calculation set up and simulation quality analysis reduces human error and workload, thus making the approach accessible to a broad population of researchers in drug discovery. 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 the 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 drug-like 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 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, it 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.

Supporting Information

Details of the OPLS2.1 force field, the Desmond molecular dynamics engine, the FEP/REST sampling algorithm, the cycle closure algorithm and the details of the simulations are available in the Supporting Information.

Figures and Tables

Table 1. Statistics of the errors for solvation free energy results reported by Shivakumar et al.¹ with OPLS2.1, OPLS2005, AM1-BCC/GAFF, and ChelpG/CHARMM-MSI for the Shivakumar test set of 237 small molecules.

Force Field	MUE (kcal/mol)	RMSE (kcal/mol)	% > 2 kcal/mol err.
ChelpG/CharmM	1.93	2.28	44.7%
AM1-BCC/GAFF	1.17	1.39	15.0%
OPLS2005	1.10	1.33	8.5%
OPLS2.1	0.73	0.88	2.1%

Table 2. Relative binding free energy calculation results. A total of 8 different receptors, covering a broad range of protein types, were studied. The number of ligands, the experimental binding affinity range of the ligands, the crystal structure used in the simulation, the original publication reporting the experimental binding affinity, and the number of perturbations for each system are reported. The details about how the data set was selected, and how the experimental binding free energies were obtained, are included in the Supplemental Information. Several different metrics to assess the performance of FEP results, including the mean unsigned error (MUE) and the root mean square error (RMSE) for all the perturbations, the correlation coefficient (R) between the FEP predicted binding affinities and the experimental results, the average error for the predictions calculated using cycle closure algorithm (average σ_{cc}) are also reported. The expected correlation coefficient between the FEP predicted binding affinities and the experimental results (Anticipated FEP R-value), and the expected correlation coefficient between two experimental measurements of the binding affinities (Anticipated Expt. R-value), assuming RMSEs of 1.1 kcal/mol and 0.4 kcal/mol for FEP predicted binding affinities and experimental data respectively, are also shown in the table. (See details in SI). The errors for the MUE, RMSE and R values are also reported using the bootstrapping method. Free energies are in units of kcal/mol.

System	BACE	CDK2	JNK1	MCL1	p38	PTP1B	Thrombin	Tyk2
Num. Compounds	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 reference	36	37	38	39	40	41	35	42,43
Num. 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
Average σ_{cc}	0.65	0.57	0.30	0.91	0.76	0.94	0.93	0.46
Observed 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	3.9x10 ⁻⁵	1.2 x10 ⁻²	7.0x10 ⁻⁸	2.2x10 ⁻⁷	1.6x10 ⁻⁷	7.8x10 ⁻⁶	1.1x10 ⁻²	2.3x10 ⁻⁷
Observed R-value MW	0.14	-0.48	-0.39	-0.55	-0.46	-0.84	-0.48	0.00
P-value MW	2.1x10 ⁻¹	3.0x10 ⁻²	4.0x10 ⁻²	8.1x10 ⁻⁵	5.3x10 ⁻³	4.0x10 ⁻⁸	6.7x10 ⁻²	0.5
Anticipated 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
Anticipated Expt. 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

Table 3. Error distribution for all 330 perturbations. The distribution is approximately Gaussian with some fattening of the tail of the distribution beyond 2.5 kcal/mol. The fitting of the error distribution by a Gaussian function with the same RMSE is given in SI Fig. S1. The observed percentage is the percentage of FEP perturbations found to be accurate within the specified absolute error. The anticipated percentage 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.

Abs. Err. (kcal/mol)	Anticipated Percentage	Observed Percentage
< 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%

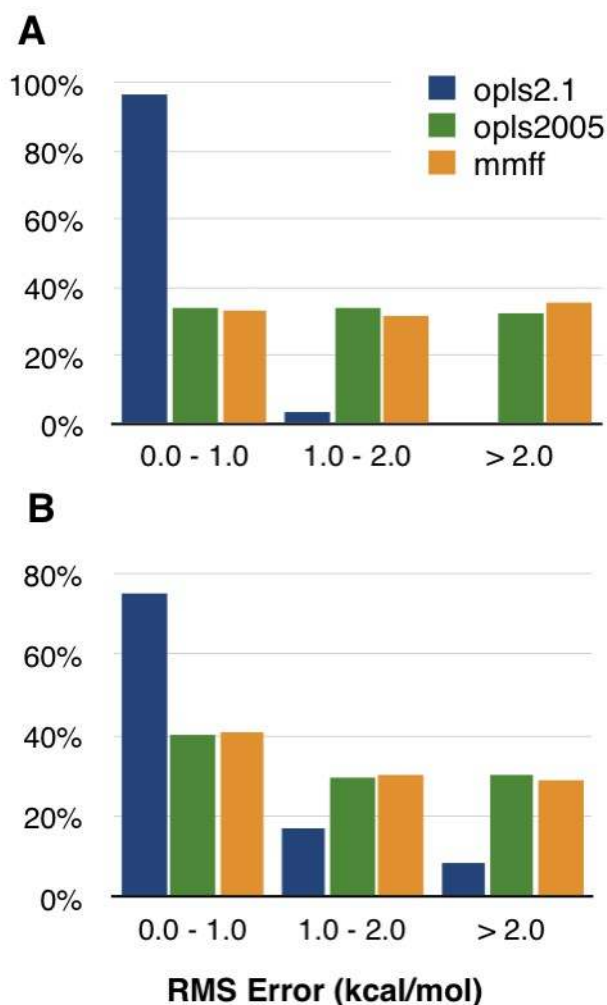


Figure 1. Histograms of the root mean square error in force-field relative energies evaluated over one-dimensional torsional angle scans (panel A) and between conformational minima (panel B) on a set of 8,365 compounds. Errors are established with respect to quantum mechanical LMP2/cc-pVTZ(-f) energies evaluated on B3LYP/631G* optimized structures. The compound set is generated from a 6 million compound repository of drug like 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.

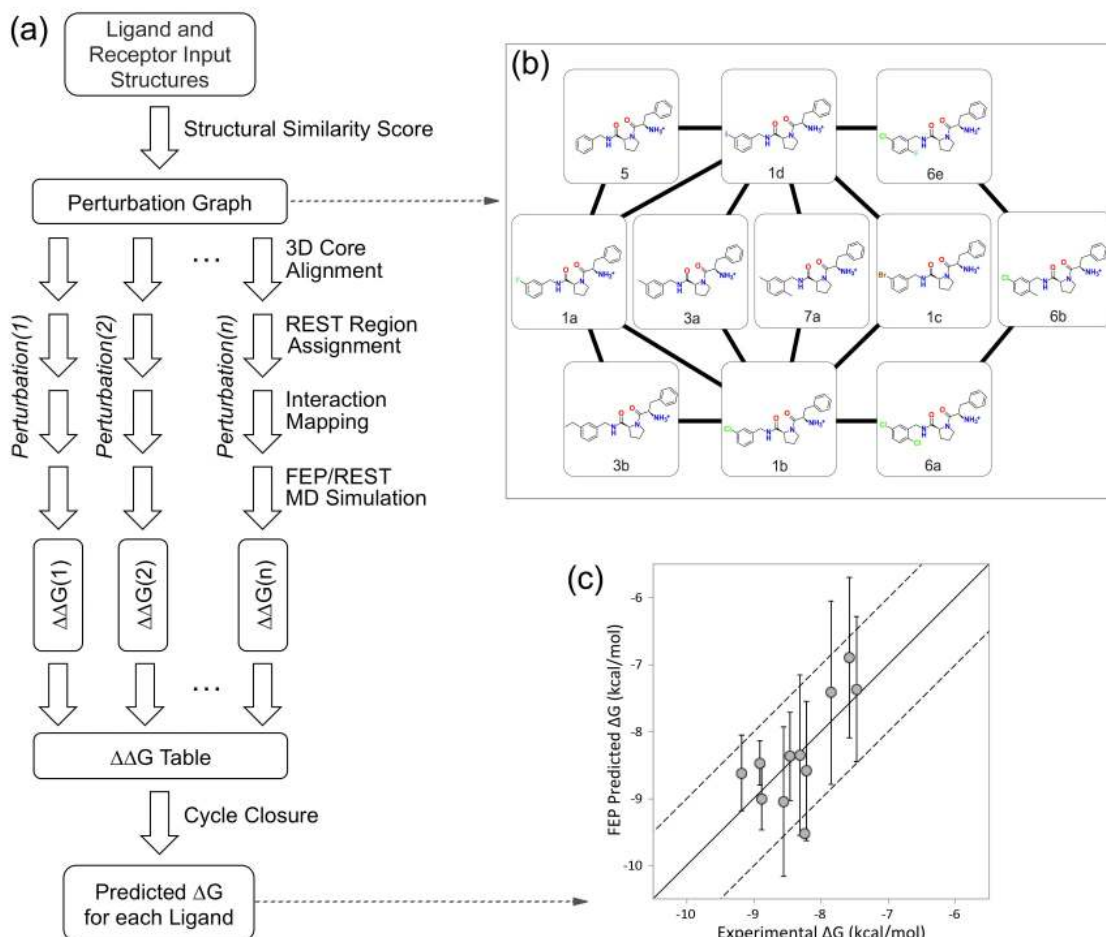


Figure 2: (a) The whole FEP workflow for protein-ligand binding affinities calculations. The key steps in the workflow, including the mutation graph generation, the REST region assignment, cycle closure convergence and error estimates and other key steps are explicitly shown in the figure. (b) An example of a 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 the two connected ligands. (c) Correlation plot of FEP predicted and experimental binding affinities for thrombin ligands, as generated from the automated workflow.

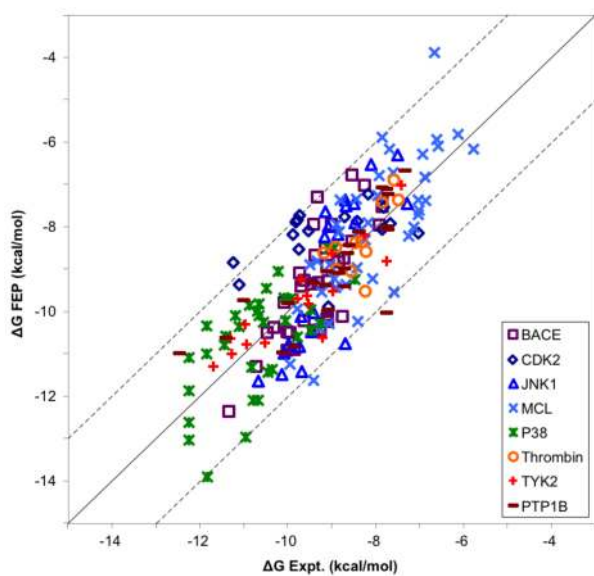


Figure 3. Correlation between the FEP predicted binding free energies and the experimental data for all 8 systems studied. FEP predicted binding free energies for most of the ligands are within 1.0 kcal/mol from their experimental values, and only 9 of 199 studied ligands deviate from their experimental free energies by more than 2 kcal/mol.

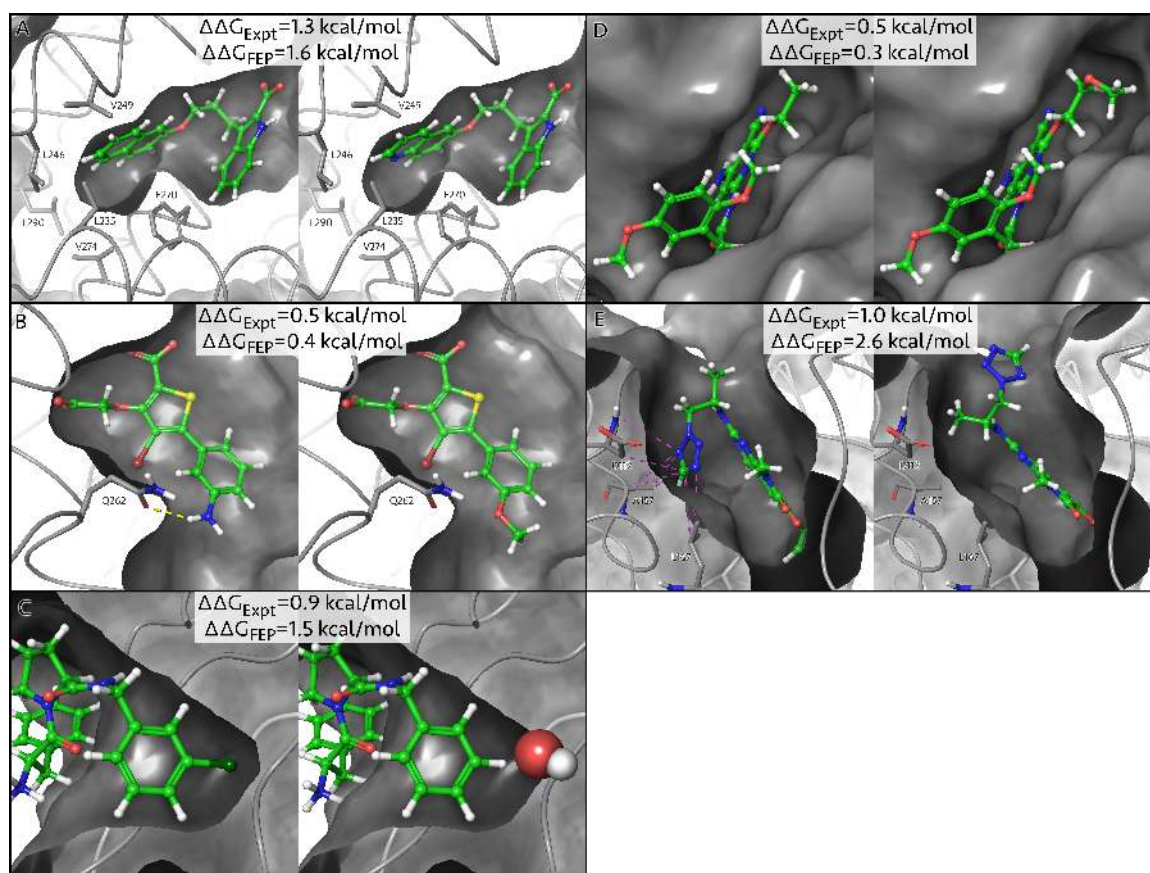


Figure 4. Representative examples of different types of interactions captured by FEP (see text for details).

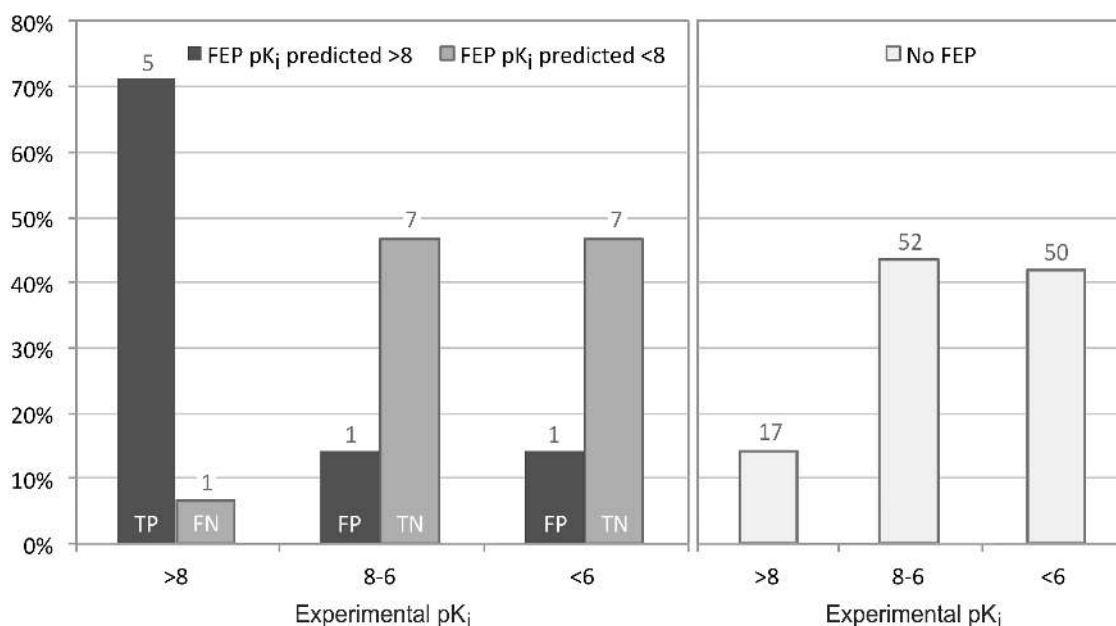


Figure 5: Histograms showing the distribution of experimental values for compounds in Project I that were predicted by FEP to have a pK_i greater than 8 (dark gray), predicted by FEP to have a pK_i less than 8 (medium gray), and those that were not computationally predicted prior to being assayed (light gray). The 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 a pK_i >8, while 71% of the compounds predicted by FEP to have pK_i >8 had an experimental pK_i >8. Key to labels: TP=true positive, FN=false negative, FP=false positive, TN=true negative.

References

- (1) Jorgensen, W. L. *Acc. Chem. Res.* **2009**, *42*, 733.
- (2) Jorgensen, W. L. *Science* **2004**, *303*, 1813.
- (3) *Free Energy Calculations: Theory and Applications in Chemistry and Biology* Berlin, 2007; Vol. 86.
- (4) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. *S. Curr. Opin. Struct. Biol.* **2011**, *21*.
- (5) Knight, J. L.; Brooks, C. L. *J. Comput. Chem.* **2009**, *30*, 1692.
- (6) Zheng, L.; Chen, M.; Yang, W. *Proc. Nat. Acad. Sci. USA* **2008**, *105*, 20227.
- (7) Gallicchio, E.; Levy, R. M. *Curr. Opin. Struct. Biol.* **2011**, *21*, 161.
- (8) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585.
- (9) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050.
- (10) Bash, P.; Singh, U.; Brown, F.; Langridge, R.; Kollman, P. *Science* **1987**, *235*, 574.
- (11) Deng, Y.; Roux, B. *J. Phys. Chem. B* **2009**, *113*, 2234.
- (12) Gallicchio, E.; Lapelosa, M.; Levy, R. M. *J. Chem. Theory Comput.* **2010**, *6*, 2961.
- (13) Durrant, J.; McCammon, J. *BMC Biol* **2011**, *9*, 1.
- (14) Riniker, S.; Christ, C.; Hansen, H.; Hünenberger, P.; Oostenbrink, C.; Steiner, D.; van Gunsteren, W. *J. Phys. Chem. B* **2011**, *115*, 13570.
- (15) Michel, J.; Essex, J. W. *J. Med. Chem.* **2008**, *51*, 6654.
- (16) MacKerell, A. D.; Bashford, D.; Bellott; Dunbrack, R. L.; 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.; 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.
- (17) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.
- (18) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
- (19) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
- (20) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474.
- (21) Halgren, T. A.; Nachbar, R. B. *J. Comput. Chem.* **1996**, *17*, 587.
- (22) Murphy, R. B.; Beachy, M. D.; Friesner, R. A.; Ringnalda, M. N. *The Journal of Chemical Physics* **1995**, *103*, 1481.
- (23) Shivakumar, D.; Harder, E.; Damm, W.; Friesner, R. A.; Sherman, W. *J. Chem. Theory Comput.* **2012**, *8*, 2553.
- (24) Storer, J.; Giesen, D.; Cramer, C.; Truhlar, D. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87.
- (25) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623.
- (26) Jorgensen, W. L.; Tirado-Rives, J. *J. Comput. Chem.* **2005**, *26*, 1689.
- (27) 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*; ACM: Tampa, Florida, 2006, p 84.

- (28) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Nat. Acad. Sci. USA* **2005**, *102*, 13749.
- (29) Wang, L.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem. B* **2011**, *115*, 9431.
- (30) Wang, L.; Berne, B. J.; Friesner, R. A. *Proc. Nat. Acad. Sci. USA* **2012**, *109*, 1937.
- (31) 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.
- (32) Cole, D. J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2014**.
- (33) 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.
- (34) Brown, S. P.; Muchmore, S. W.; Hajduk, P. J. *Drug Discovery Today* **2009**, *14*, 420.
- (35) Baum, B.; Mohamed, M.; Zayed, M.; Gerlach, C.; Heine, A.; Hangauer, D.; Klebe, G. *J. Mol. Biol.* **2009**, *390*, 56.
- (36) 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.
- (37) 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.
- (38) 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.
- (39) 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.
- (40) 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.
- (41) 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.
- (42) 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.

- (43) 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.