JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Binding Free Energy Calculations for Lead Optimization: Assessment of Their Accuracy in an Industrial Drug Design Context

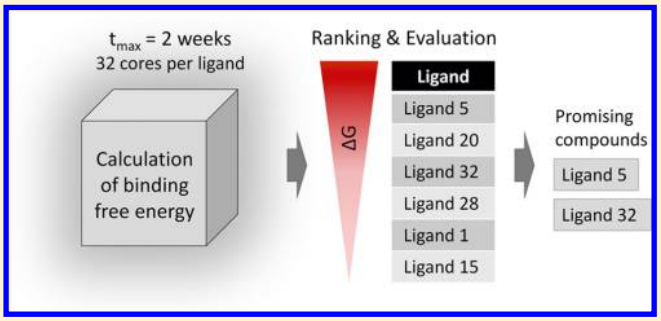
Nadine Homeyer,[†] Friederike Stoll,[§] Alexander Hillisch,[§] and Holger Gohlke*,[†]

[†]Institute for Pharmaceutical and Medicinal Chemistry, Department of Mathematics and Natural Sciences, Heinrich-Heine-University Düsseldorf, Universitätsstr.1, 40225 Düsseldorf, Germany

[§]Global Drug Discovery, Medicinal Chemistry, Bayer Pharma AG, Aprather Weg 18A, 42113 Wuppertal, Germany

Supporting Information

ABSTRACT: Correctly ranking compounds according to their computed relative binding affinities will be of great value for decision making in the lead optimization phase of industrial drug discovery. However, the performance of existing computationally demanding binding free energy calculation methods in this context is largely unknown. We analyzed the performance of the molecular mechanics continuum solvent, the linear interaction energy (LIE), and the thermodynamic integration (TI) approach for three sets of compounds from industrial lead optimization projects. The data sets pose challenges typical for this early stage of drug discovery. None of the methods was sufficiently predictive when applied out of the box without considering these challenges. Detailed investigations of failures revealed critical points that are essential for good binding free energy predictions. When data set-specific features were considered accordingly, predictions valuable for lead optimization could be obtained for all approaches but LIE. Our findings lead to clear recommendations for when to use which of the above approaches. Our findings also stress the important role of expert knowledge in this process, not least for estimating the accuracy of prediction results by TI, using indicators such as the size and chemical structure of exchanged groups and the statistical error in the predictions. Such knowledge will be invaluable when it comes to the question which of the TI results can be trusted for decision making.



INTRODUCTION

The costs spent per approved drug have increased tremendously over the last decades and were recently estimated to amount to 4 billion U.S. dollars.[1,2] Lead optimization alone costs approximately 150 million U.S. dollars per new molecular entity[3] such that the pharmaceutical industry will greatly benefit from methods that help increasing the efficiency of this early phase of drug development. Predicting binding affinities of compounds to their primary target (or at least correctly ranking them) would be one such approach as it allows concentrating synthesis efforts to more promising compounds then. However, although various approaches are available today,[4] which have proved useful for computing relative binding affinities of small molecules in certain studies,[5,6] these methods are still not routinely applied in industrial drug development.[7,8] Besides the computational expenses and the complex setup of such calculations, it is the lack of studies that demonstrate their value for industrial drug discovery projects that leads to reluctance to use these methods more widely. In fact, while binding free energy calculations have been successfully applied in numerous cases to study individual ligand data sets, with the ligands often originating from different sources, or a few interesting drug-candidate molecules, information about the accuracy of binding free energy calculations on series of ligands that are typical for later stages of drug design is largely missing. This provided the incentive for us to test the performance of binding free energy calculations on protein−ligand data sets that originate from drug design pipelines and pose "day-to-day" challenges such as the lack of experimentally determined binding modes, unusual protein−ligand interactions, large structural diversity of the ligands, ligands with varying total charges, uncertainty of protonation states, unspecified/inferred stereochemistry, and modeled and/or mobile target structures.

In terms of computing time, end-point free energy methods are the most efficient approaches for computing binding free energies because they only consider the free and the bound states of the system. The molecular mechanics Poisson−Boltzmann surface area (MM-PBSA), the molecular mechanics generalized Born surface area (MM-GBSA),[9,10] and the linear interaction energy (LIE)[11] approaches are likely the most well-known and widely applied methods in this class. MM-PBSA and MM-GBSA estimate the binding free energy from changes in the gas-phase energy, solvation free energy, and configurational entropy upon complex formation. The solvation free energy is determined by a continuum solvent model and a term accounting for the nonpolar contribution to solvation. In contrast, the LIE approach considers water interactions

Received: January 14, 2014
Published: June 6, 2014

   dx.doi.org/10.1021/ct5000296 | J. Chem. Theory Comput. 2014, 10, 3331−3344

Table 1. Studied Protein−Ligand Systems

target protein	template complex structure	no. of ligands	no. of charged ligands	challenges of data set[a]
factor Xa (FXa)	X-ray structure of FXa−F5[b] complex of Roehrig et al.[15]	25[15]	0	I, II, VI
cyclin-dependent kinase 2 (CDK2)	X-ray structure of CDK2−C37S[b] complex of Jautelat et al.[16]	29[16]	10	I, III, IV
mineralocorticoid receptor (MR)	MR−M38[b,c] complex structure modeled based on X-ray data	29 (+12)[d,18]	(2)[e]	I, IV, V, VII, VIII

[a]Challenges of the data set for binding free energy predictions: (I) experimentally determined binding modes missing for the majority of the ligands; (II) uncommon protein−ligand interactions; (III) ligands structurally diverse; (IV) ligands with different total charges; (V) protonation state not known for all ligands; (VI) orientation of central ring moiety of ligands undetermined; (VII) stereochemistry of some ligands uncertain; (VIII) partially modeled receptor structure. [b]Investigated ligands and their aliases are given in Tables S1−S3. [c]The structure of the MR bound to M38 was modeled based on MR agonist bound X-ray structures[84,85] and an active antagonist/partial agonist conformation of helix 12 from the glucocorticoid receptor.[86] [d]A subset of 12 MR antagonists was investigated in a blind test, without a priori knowledge of the IC_{50} values. [e]Compounds 8 and 24 were considered in two alternative protonation (charge) states.

Table 2. Overview of MD Simulations and Free Energy Calculations

computation parameters/conditions	studied system				
	FXa		CDK2		MR
	FXa_S	FXaRestr	CDK2_S	CDK2Wat	MR_S
MD Simulations					
crystal water considered in complex simulation	7 H_2O within 6 Å of ligand		–	added 36 H_2O of 2R3I to CDK2_S	5 H_2O within 7 Å of ligand
distance restraints	–	+	–	–	–
simulations with AM1-BCC charges for ligand	+	–	+	–	–
simulations with RESP charges for ligand	+	+	+	+	+
separate simulations for complex, receptor, and ligand	20 ns	–	20 ns	–	30 ns
simulations for complex only	–	6 ns	–	1 ns	–
protein structure used for receptor simulation	1C5M[87]	–	1PW2[88]	–	same as for the complex
MM-PB(GB)SA[a]					
(A) 1-trajectory approach:					
Interval considered	4−20 ns	2−6 ns	4−20 ns	0−1 ns	20−30 ns
Structural ensemble from RESP simulations[b]	GB, PB	PB2	GB, PB; ε_{int} varied	PB1	PB, GB
Structural ensemble from AM1-BCC simulations	GB, PB	–	GB, PB	–	–
(B) 3-trajectory approach:					
interval considered	10−20 ns	–	10−20 ns	–	20−30 ns
structural ensemble from RESP simulations	GB, PB	–	GB, PB	–	PB, GB
structural ensemble from AM1-BCC simulations	GB, PB	–	GB, PB	–	–
(C) minimized structures[c]					
ε_{int} = 1	GB, PB	–	GB, PB	–	PB,GB
ε_{int} = 2 or 4	GB, PB	–	GB, PB	–	–
LIE Calculations					
interval from RESP simulations considered	4−20 ns	–	4−20 ns	–	20−30 ns
α, β, and γ optimized	+	–	+	–	+
β = 0.5, α and γ optimized	+	–	+	–	+
approach of Hansson et al.[49] with γ optimized	–	–	+	–	–
Thermodynamic Integration					
no. of transformations	5	–	28	–	42
maximum length of TI simulations	10 or 20 ns	–	10 ns	–	10 ns

[a]If no specific information for the calculation is provided, then all types of MM-PB(GB)SA calculations mentioned in Table S4 were performed. [b]Taken as default MM-PB(GB)SA calculation protocol. [c]The structures were taken from the initial minimizations of the complexes in the equilibration step.

explicitly and estimates the binding free energy from changes in the electrostatic and van der Waals interactions of the ligand with its environment upon transfer from solution to the receptor binding site.[6,12] Both approaches commonly consider conformational effects by averaging over the energies of structural ensembles of the end-points generated by molecular dynamics (MD) simulations. Pathway methods such as thermodynamic integration (TI) and free energy perturbation (FEP) are usually more accurate but are also more time-consuming than the end-point approaches. This is because they require sampling of unphysical intermediate states in order to compute free energies along a transition path from an initial to a final state. With respect to efficiency, TI has the advantage that the precision of binding free energy predictions can be increased by subsequently including additional intermediate states.[7] Thus, this method offers the opportunity to start

dx.doi.org/10.1021/ct5000296 | J. Chem. Theory Comput. 2014, 10, 3331−3344

calculations at the lower limit of precision and only conduct additional sampling where necessary. This is especially valuable because of the inverse relationship between accuracy and required computing time usually found for binding free energy calculations,[12] which requires one to find an optimal trade-off between prediction quality and computational demand.

We investigated the performance of MM-PB(GB)SA, LIE, and TI on data sets from the lead optimization phase of industrial drug design projects with a special focus on efficiency. As lead optimization cycles typically cover only a few weeks,[13] we set ourselves a time limit of 2 weeks for the binding free energy calculations on a state of the art compute cluster with 32 cores available per ligand. Within this limit we calculated the relative binding free energies of three medium-size sets of ligands binding to the serine protease factor Xa (FXa),[14,15] the cyclin-dependent kinase 2 (CDK2),[16] and the mineralocorticoid receptor (MR).[17,18] All of these data sets pose one or multiple of the above-mentioned challenges (Table 1). Performing binding free energy calculations on these data sets will provide information about how to cope with these challenges and the accuracy of predictions that can be reached in such cases. Based on these results, we propose rules for binding free energy calculations that shall be of value for the process of lead optimization in an industrial context.

METHODS

Structure Preparation. Three data sets were used in this study (Table 1). For each system a complex structure of the target protein with a representative ligand was obtained either from X-ray crystallography or modeled based on available structural and experimental information. The other compounds of the respective data set were modeled into the binding pocket by a combination of docking and manual adjustment of the ligand molecule by an expert modeler.

The structures were prepared for free energy analyses using the programs Sybyl,[19] SwissPDBViewer,[20,21] Pymol,[22] and AMBER11.[23] In addition to the protein−ligand complex structures, we also prepared *apo* structures of FXa and CDK2 obtained from the Protein Data Bank.[24] These structures were used as models for the ligand free proteins to account for structural changes taking place upon binding.

In order to ensure that water molecules located close to the binding site are placed within this region in the starting structure of a bound system, these water molecules were retained from the complex structures of FXa and MR. For CDK2 it has been reported that it is important to consider water interactions between the protein and the ligands.[25] Thus, in addition to a bound-state CDK2 structure without water, a structure with water molecules was prepared, referred to in the following as CDK2Wat. A detailed description of the preparation of all systems is provided in the Supporting Information.

Regarding ligand structures, two alternative starting configurations were generated for FXa inhibitors F4, F6, F8, F10, and F12 because the position of the fluorine atom at the central ring of these inhibitors was not known from experimental studies; in the alternative structures the fluorine atom at the central ring pointed either toward or away from the protein surface. As it was found that the central ring turned around freely during the MD simulations, the free energies calculated for the alternative starting configurations were averaged before evaluation.

MD Simulations. Conformational ensembles for the MM-PB(GB)SA calculations and LIE analyses were generated by MD simulations in explicit water. A summary of all conducted simulations is provided in Table 2. Force field parameters for the ligands were taken from the general AMBER force field.[26] Ligand atomic charges were either determined by the AM1-BCC method[27,28] with SQM and the AMBER tool Antechamber or calculated by the restraint electrostatic potential fit procedure[29] using Gaussian03[30] and Antechamber. Protein systems and structurally bound ions were described by the ff99SB force field.[31,32] Model systems of complexes, receptors, and ligands were generated based on the prepared structures with LEaP. After neutralization by adding Na^+ or Cl^- counterions, the systems were placed in a box of TIP3P[33] water molecules extending at least 11 Å in each direction from the solute. The solvated systems served as input structures for the simulations. All simulations were conducted with AMBER11.[23] A detailed simulation protocol is provided in the Supporting Information.

MM-PBSA and MM-GBSA. Binding free energy calculations were carried out with the mm_pbsa.pl script of AMBER11[23] using structures extracted in 20 ps intervals from the MD simulations, or structures of the minimized systems, from which water and counterions had been stripped (Table 2). The sampling frequency of 20 ps is well above a correlation time of 1−5 ps of MM-PB(GB)SA effective energies found in previous studies.[34,35]

In the MM-PB(GB)SA approach the free energy of a molecule is obtained from the sum of the molecular mechanics energy, solvation free energy, and configurational entropy (eq 1). If the free energy is determined from a conformational ensemble, the individual energetic contributions are averaged over all structures of the ensemble.

$$G_{\text{molecule}} = E_{\text{MM}} + G_{\text{solv}} - \text{TS} \quad (1)$$

The binding free energy is then computed from the free energies of the complex, receptor, and ligand according to eq 2.

$$\Delta G_{\text{binding}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}} \quad (2)$$

In this study solvation free energies were determined using five flavors of implicit solvation models (Table S4). The GBSA method I of Onufriev et al.[36] (GB2) and PBSA calculations with mbondi[37] and Parse[38] radii (PB1&2) yielded overall the best results. Only these three methods will thus be considered below.

Previous performance studies of the MM-PB(GB)SA method have revealed that taking into account entropic contributions calculated by the quasiharmonic (QH) or the normal mode (NM) approach introduces a significant uncertainty in the prediction of the binding free energy.[34,39−41] Furthermore, it has been reported that considering entropic contributions does not always lead to a gain in accuracy for relative binding free energy predictions.[40,42,43] Consistent with these findings we observed no improvement in the relative binding free energy predictions for the FXa inhibitors when entropies calculated by the QH or NM approaches based on the FXa_S simulations were taken into account. In the case of the QH analyses, even the entropies of some of the ligands were found to be not converged after 20 ns of simulation time, let alone the entropies of the receptor structures. Therefore, we decided to neglect contributions due to changes in the configurational entropy as is often done if relative binding free energies are computed.[44−46] Consistent with previous definitions,[34] the binding energies computed by MM-PB(GB)SA as the sum of gas-phase

energies and solvation free energies are referred to as "effective binding energies" below.

Snapshots for the generation of conformational ensembles can either be obtained from a single trajectory of the complex (1-trajectory approach) or from individual trajectories of the complex, receptor, and ligand (3-trajectory approach). Both possibilities were tested in our study. Probably due to a larger noise caused by the usage of different conformations for complex, receptor, and ligand, which precludes the cancellation of internal energies,[40] the performance of the 3-trajectory approach was inferior compared to the one of the 1-trajectory approach for all three data sets. Hence, only the 1-trajectory alternative is considered here.

In order to test the influence of the atomic charges of the ligand on the binding free energy predictions, MM-PB(GB)SA calculations were carried out using conformational ensembles from simulations with AM1-BCC[27,28] or RESP[29] charges for the ligand. Overall slightly better predictions were obtained in the case of RESP charges, and only these results are reported below. As determining such charges is no longer hampered by the computational demand of the calculations, and automated tools facilitate the computation of charges for sets of multiple ligands,[47] we recommend to use RESP charges in general.

In summary, unless denoted otherwise, default MM-PB(GB)SA analyses were conducted using conformational ensembles extracted from the complex simulations with RESP charges and employing the 1-trajectory approach.

LIE Analyses. Binding free energy calculations according to the LIE approach were carried out using snapshots from simulations of the solvated complex and the solvated ligand (Table 2). First, for each snapshot single point molecular mechanics energy calculations were conducted for the total system as well as for its individual components using Sander of AMBER11.[23] Then the electrostatic and van der Waals (vdW) interaction energies of the ligand were determined as the difference between the total energies and the individual energy contributions (eq 3).

$$E_i = E_i^{total} - E_i^{ligand} - E_i^{rest} \quad i = \text{electrostatic|vdW} \tag{3}$$

Electrostatic and vdW interaction energies were averaged over all snapshots of the respective conformational ensemble and then used to estimate the binding free energy ΔG by a multiple linear regression to experimental binding affinities according to eq 4 employing the program R.[48]

$$\Delta G_{binding} = \alpha(\langle E_{vdW}^{bound}\rangle - \langle E_{vdW}^{free}\rangle) + \beta(\langle E_{elec}^{bound}\rangle - \langle E_{elec}^{free}\rangle) + \gamma \tag{4}$$

It has been a matter of discussion what scaling factors are to be used to weight the differences in the electrostatic and the vdW interactions, and multiple values have been suggested.[6,49−51] In order to investigate this influence three different settings were tested in this study (Table 2).

Thermodynamic Integration. TI transformations of one ligand into another were conducted by performing simulations at discrete λ steps, where λ couples the potential functions of the two states, V_0 and V_1 (eq 5).

$$V(\lambda) = (1 - \lambda)V_0 + \lambda V_1 \tag{5}$$

The free energy difference ΔG for the transformation was computed by integration over the average $dV/d\lambda$ values obtained at each λ step (eq 6).

$$\Delta G = \int_0^1 \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{6}$$

To determine the difference in the binding free energy $\Delta\Delta G$ between two ligands, transformations were performed for both the complex-bound ligands and the solvated ligands. $\Delta\Delta G$ was then calculated as the difference between the respective free energies (eq 7).

$$\Delta\Delta G = \Delta G_{bound} - \Delta G_{solvated} \tag{7}$$

Pairs of ligands for TI transformations were selected based on their similarity. For each set of ligands a similarity matrix was computed employing the TanimotoCombo score of ROCS.[52,53] Then, by means of the algorithm of Kruskal,[54] those transformations were identified that require overall the smallest structural changes. For all data sets except the FXa set, transformations between all identified "optimal" pairs of compounds were carried out (Table 2). For the FXa set, only transformations starting from the compound present in the crystal complex structure were performed (see below).

As all TI calculations for a data set were to be conducted within the scope of 2 weeks, transformation simulations were run at four λ values (λ = 0.2, 0.4, 0.6, and 0.8) applying the dual topology, 1-step softcore approach.[7,55] This approach has recently been successfully used in an investigation of FXa inhibitors.[56] As this study reported a mean absolute deviation <1 kJ mol^{-1} for relative binding free energies determined with three λ steps versus those determined with nine λ steps and concluded that it is usually sufficient to consider three λ steps in such TI calculations, we expected to obtain reasonable relative binding free energy predictions with four λ steps. However, we note that the employed number of λ steps is at the lower limit. Thus, we tested whether including additional λ values improves the accuracy of the $\Delta\Delta G$ predictions by also performing transformation simulations at λ = 0.1, 0.3, 0.5, 0.7, and 0.9 for the 20 ligand transformations with the highest similarity within the CDK2 data set.

In transformations involving a change in the total ligand charge, the overall non-neutral systems were not explicitly neutralized. In these cases a uniform neutralizing plasma, present in periodic boundary particle mesh Ewald simulations in AMBER, ensured the neutrality of the systems.[57] Under these conditions, no correction for the self-energy, i.e., the energy caused by the interaction of the charged site with its own periodic image and with the neutralizing plasma, needs to be applied. Further effects due to the finite system size have been found to be marginal if, as in all our TI simulations, more than 922 water molecules are present within the periodic box.[57] In addition, any inaccuracies caused by changing the net charge of the system should largely cancel because we calculate only relative binding free energies, and the effect of the charge change should be similar in the transformation of the ligand in the bound state and in the transformation of the ligand free in solution.

ΔG (eq 6) was computed by numerical integration employing the calculation approaches with and without linear extrapolation to the end states at λ = 0 and 1 available in the free energy workflow tool FEW.[47] We found that significantly better $\Delta\Delta G$ predictions (eq 7) were obtained without extrapolation. Likely, this results from the uncertainty in the extrapolation, which generally increases for nonlinear $dV/d\lambda$ curves with the λ step size. In addition, it has been reported for one-step softcore TI calculations that kinks can appear in the

$dV/d\lambda$ curve.[55] If such kinks occur at the terminal λ steps used for the extrapolation this can decrease the accuracy of the binding free energy predictions further. Consequently, we only report $\Delta\Delta G$ predictions obtained without extrapolation below.

A detailed description of the TI simulation procedure is given in the Supporting Information.

Bennett Acceptance Ratio (BAR) and Multistate Bennett Acceptance Ratio (MBAR) Calculations. Production simulations were conducted with Sander of AMBER11[23] as specified above for the TI calculations, and potential energies at the specific λ value as well as potential energies of a configuration at all other considered λ values (λ = 0.2, 0.4, 0.6, and 0.8) were recorded. The potential energies were extracted from the Sander output files and converted into the format required for the BAR calculations using a modified version of the mdout2pymbar.pl script provided with AmberTools12.[58] Calculations according to the BAR and the MBAR methods were conducted with the pyMBAR program, version 2.0 beta, of Shirts and Chodera[59] obtained from https://simtk.org/home/pymbar and using a temperature of T = 300 K.

Error Estimation. Standard errors of the mean (σ_{MSE}) of computed effective binding energies ($\Delta G_{effective}$) were calculated according to eq 8

$$\sigma_{MSE} = \frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\Delta G_{effective,i} - \overline{\Delta G_{effective}})^2}}{\sqrt{n}} \quad (8)$$

where n is the number of effective binding energies considered, i.e., the number of snapshots taken into account in the calculation.

For LIE the statistical error in ΔG_{LIE} was calculated from the σ_{MSE} of the electrostatic (ele) and van der Waals (vdW) energy contributions of the bound and solvated ligands according to eq 9.

$$\sigma_{LIE} = \sqrt{(\sigma_{ele,bound}^2 + \sigma_{ele,solvated}^2 + \sigma_{vdW,bound}^2 + \sigma_{vdW,solvated}^2)} \quad (9)$$

For TI the statistical error in ΔG (eq 6) denoted $\sigma_{\Delta G}$ was estimated according to reference,[55] and the error in $\Delta\Delta G$ (eq 7) denoted $\sigma_{\Delta\Delta G}$ was calculated according to eq 10.

$$\sigma_{\Delta\Delta G} = \sqrt{(\sigma_{\Delta G,solvated}^2 + \sigma_{\Delta G,bound}^2)} \quad (10)$$

The respectively largest error in $\Delta G_{effective}$, ΔG_{LIE}, or $\Delta\Delta G$ found for a data set is denoted as σ_{max} below.

Quality Measures. The degree of linear dependence between experimental and computed binding free energies was evaluated by Pearson's correlation coefficient (r^2), and the significance of the linear correlation was determined from the p-value (p) of a two-tailed Student's t test. We consider a correlation with $p < 0.05$ significant. Both r^2 and p were computed with the R program.[48] As a quality measure for the correct relative ranking of ligands by binding free energy, we used the predictive index (PI) of Pearlman.[60] The best r^2 obtained in a specific analysis is denoted r^2_{max} below, and the corresponding PI is denoted PI_{max}. Bootstrapped 95% confidence intervals for r^2_{max} were computed with the boot[61] package of R[48] performing 10,000 bootstrap replications and employing bias-corrected, accelerated percentile intervals. For evaluating the accuracy of the computed $\Delta\Delta G$ values we used the root-mean-square deviation (RMSD) computed according to eq 11

$$RMSD = \sqrt{\frac{1}{n}\sum(\Delta\Delta G_{experiment} - \Delta\Delta G_{calculated})^2} \quad (11)$$

where n is the number of ligands in a data set, and $\Delta\Delta G_{experiment}$ was calculated from $\Delta\Delta G = -RT\ln(IC_{50}^{ligand\ 1} / IC_{50}^{ligand\ 2})$ with T = 300 K. Detailed listings of all computed quality measures are provided in the Supporting Information.

RESULTS AND DISCUSSION

MM-PB(GB)SA Calculations. Although the continuum solvent molecular mechanics approach has shown a good performance in several relative binding free energy predictions,[41,46,62−64] it did not result in very good binding affinity rankings for our data sets if the default calculation protocol was used. In neither case did the uncertainty in the calculations limit the performance of the approach, as indicated by σ_{max} values given below that are at least 12-fold smaller than differences in the experimental binding affinities across the considered data sets. Rather, detailed investigations of the failures revealed several aspects that are essential for good binding free energy predictions by this approach.

FXa Data Set. Although inhibitors of FXa have already been addressed in multiple MM-PB(GB)SA based binding affinity studies,[56,65−67] this data set was selected because it poses challenges that are often faced in lead optimization such as unusual protein−ligand contacts and missing experimentally determined ligand binding modes. While no correlation between experimentally determined binding affinities and computed effective binding energies was observed when all 25 FXa inhibitors were considered (r^2_{max} = 0.08, p = 0.180, Tables 3 and S5), a significant correlation was found for the subset of compounds with IC_{50} values < 150 nM (r^2_{max} = 0.54, bootstrapped 95% confidence interval: 0.21 < r^2_{max} < 0.79, p < 0.001, σ_{max} = 0.15 kcal mol^{-1}, Tables 3 and S5). A detailed inspection of MD trajectories of FXa-ligand complexes revealed that many of the ligands partially and repeatedly move out of the protein's S1 pocket in the course of the simulations (Figure

Table 3. Best Prediction Results of the MM-PB(GB)SA Calculations

system	no. of ligands and calculation details	structures[a]	r^2_{max}	p[b]	PI_{max}
FXa_S	25 ligands	MD ensembles	0.08	0.180	0.32
	18 ligands with IC_{50} < 150 nM	MD ensembles	0.54	<0.001	0.81
	25 ligands	minimized structures	0.36	0.002	0.61
	18 ligands with IC_{50} < 150 nM	minimized structures	0.62	<0.001	0.81
CDK2_S	29 ligands	MD ensembles	0.05	0.225	0.21
	29 ligands	minimized structures	0.00	0.857	0.00
	28 ligands[c]; ε_{int} = 4 for positively charged, ε_{int} = 1 for other	MD ensembles	0.46	<0.001	0.69
MR_S	28 ligands	MD ensembles	0.30	0.003	0.61
	28 ligands	minimized structures	0.04	0.339	0.27

[a]Structures used for MM-PB(GB)SA calculations. [b]p-value of a two-tailed Student's t test to test the null hypothesis that no correlation exists. [c]Ligand C25 was not considered. See text for details.

dx.doi.org/10.1021/ct5000296 | J. Chem. Theory Comput. 2014, 10, 3331−3344

1), which is at variance with the well-defined binding modes of the chlorothiophene moiety found for several ligands in complex crystal structures. In contrast to earlier FXa inhibitors whose dominant interaction with the S1 pocket is a salt bridge, the compounds studied here mainly interact with this binding site region via a perfect steric fit, leading to the displacement of entropically unfavorable water molecules and a halogen−π interaction with Tyr228. We thus anticipated that the observed ligand movements are caused by too weak interactions between the ligand's halogen atom and the aromatic ring of Tyr228 as modeled based on the GAFF and ff99SB force fields.[68] Consequently, one should expect a better correlation for effective binding energies computed for complex structures that are more similar to the input structures; this was indeed found for the minimized input structures (r^2_{max} = 0.36, bootstrapped 95% confidence interval: 0.10 < r^2_{max} < 0.62, p = 0.002 for all FXa ligands; r^2_{max} = 0.62, bootstrapped 95% confidence interval: 0.23 < r^2_{max} < 0.86, p < 0.001 for the subset of ligands with IC_{50} < 150 nM). As an alternative, we set out to correct for the too weak halogen−π interaction by defining a distance restraint between the ligand's halogen atom and the center of Tyr228 (FXa_Restr simulations) as proposed by Wallnoefer et al.[68] However, no significant overall correlation was found (r^2_{PB2} = 0.14, p = 0.064), which is probably due to the oversimplification of the correction. In contrast, parametrizations of chlorine, bromine, and iodine interactions for the OPLSaa[69] and the AMBER[70,71] force field that take into account the σ-hole effect[72] at the halogens were shown to improve the description of halogen bonding in molecular mechanics calculations and to yield the expected trend for relative binding free energies of halogenated compounds. Consistent with previous reports that a force field accurately describing the properties of the studied system is essential,[40,73] these findings strongly call for an appropriate parametrization of nonbonded interactions as a prerequisite to obtain good binding free energy estimates, particularly if "non-conventional" interactions are formed between protein and ligand. We thus recommend to thoroughly investigate the binding mode and the motions of the ligands in the course of the simulations to identify variations from available crystal complex structures, which may be indicative of force field inaccuracies.

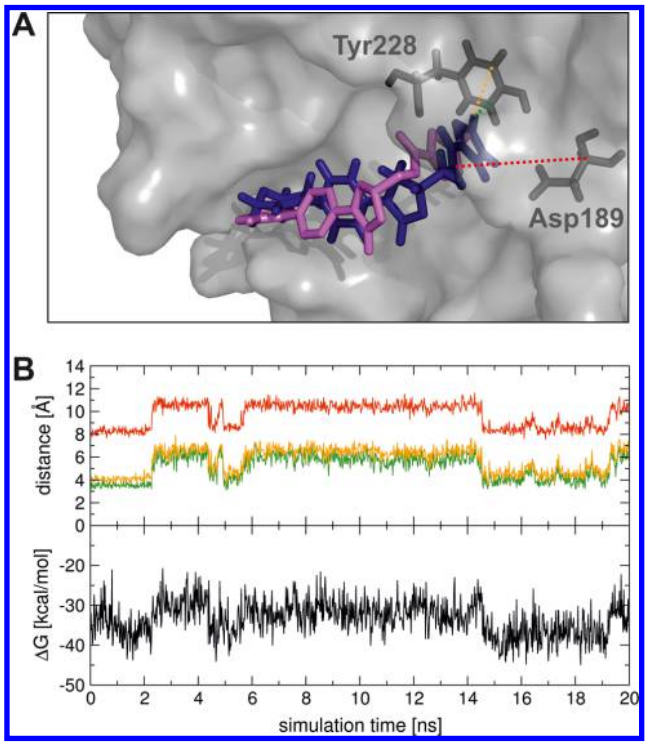


Figure 1. Movements of FXa inhibitor F5 during MD simulations. (A) Position of the inhibitor in the starting structure (blue) and after 10 ns of MD simulations (magenta) relative to FXa (crystal structure; gray surface). All atoms of the complex were considered in the structural overlay. Distances between atoms of the chlorothiophene group of the inhibitor and atoms of Tyr228 and Asp189, respectively, of FXa shown in B are indicated by dotted lines. (B) Interatomic distances in the course of the MD simulations showing the shifting of the inhibitor's chlorothiophene group in and out of the S1 pocket (top) and effective binding energies calculated by the MM-PBSA approach that reflect the movements of the inhibitor (bottom).

CDK2 Data Set. The CDK2 inhibitors belong to the well-known group of indirubine derivatives and exhibit conventional protein−ligand interactions.[16,74] However, for these ligands no correlation between calculated effective binding energies and IC_{50} values was found, neither with conformational ensembles from MD simulations (r^2_{max} = 0.05, p = 0.225, Tables 3 and S6) nor with minimized complex structures (r^2_{max} = 0.00, p = 0.857, Tables 3 and S6). Thus, inaccuracies in the force field are unlikely to be the reason for this. Next, we investigated whether explicitly incorporating crystal water molecules into the starting structures improved the binding free energy predictions (Table 2, CDK2Wat). This is based on reports according to which structural water molecules are important for maintaining the binding mode of CDK2 ligands.[25] However, no significant improvement of the prediction was observed (r^2_{max} = 0.11, p = 0.083), suggesting that the performance of the initial MM-PB(GB)SA calculations had not been deteriorated by the absence of these water molecules.

While all studied CDK2 inhibitors share a common indirubine scaffold, the substituents at the 5- and 3′-positions of the scaffold are highly diverse (Table S2). This holds particularly for the substituents at the 5-position, which carry different charges and influence the polarity of the protein−ligand interaction. It has been reported that MM-PB(GB)SA binding free energy predictions are sensitive to the polarity of interactions between the binding partners and that the dielectric constant of the solute (ε_{int}) should be increased for polar interactions to account for an increased screening effect.[40,66] In agreement, we observed that the binding free energies of the positively charged ligands were significantly underestimated in the default MM-PB(GB)SA calculations with ε_{int} = 1 due to overestimating a repulsive interaction between the ligands and a positively charged lysine residue. In contrast, when we set ε_{int} = 4 for the positively charged ligands, the prediction quality improved considerably, considering all ligands but C25 (r^2_{max} = 0.46, bootstrapped 95% confidence interval: 0.22 < r^2_{max} < 0.64, p < 0.001, σ_{max} = 0.33 kcal mol^{-1}; Tables 3 and S6, Figure 2). Structural analyses revealed that, in contrast to all other inhibitors with a diamine moiety, C25 does not adopt a conformation in which the diamine chain points into the ribose pocket of CDK2; thus, the chain strongly interacts with the positively charged lysine. This is probably due to steric hindrance of the diamine group of C25 by the large sugar moiety at the 3′-position (Figure 2). This finding emphasizes that in MM-PB(GB)SA calculations ε_{int} must be carefully chosen such as to reflect the polarity of the protein−ligand interaction and that even within a data set of structurally

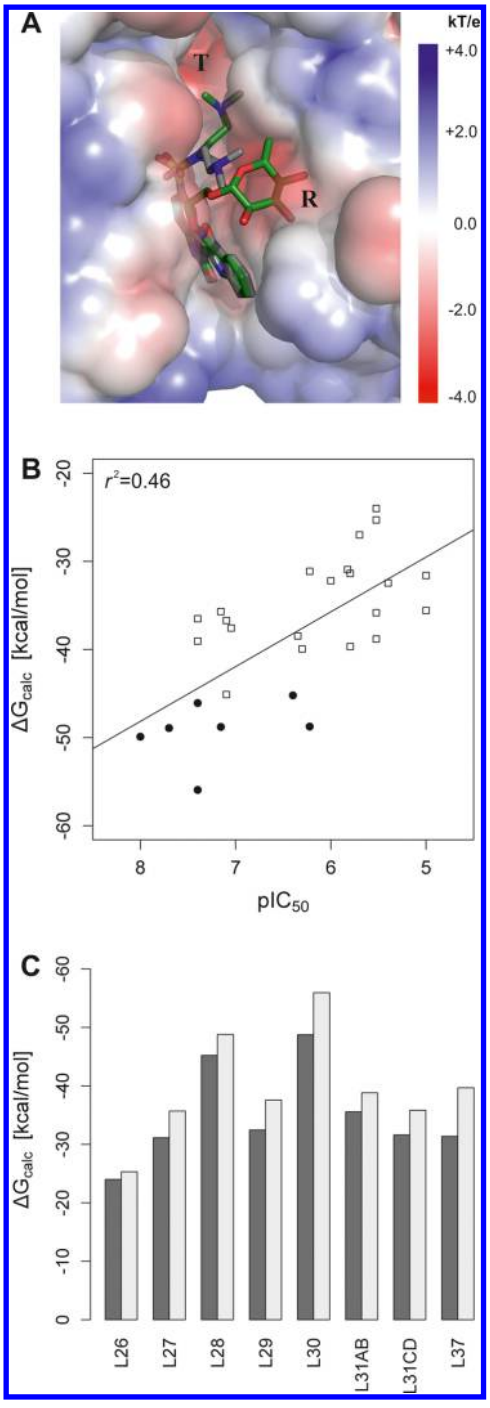


Figure 2. (A) Comparison of the relative orientation of the diamine side chain of inhibitors C22 (gray) and C25 (green) after 20 ns of MD simulations. The binding site of CDK2 with the ribose pocket (R) and the triphosphate channel (T) is depicted in surface representation and color-coded according to the electrostatic potential calculated by APBS.[89] (B) Effective binding energy (ΔG_{calc}) calculated by the MM-PBSA approach (PB1) versus pIC_{50} values for all CDK2 inhibitors except C25. For positively charged ligands ε_{int} = 4 (●) and for all other ligands ε_{int} = 1 (□) was used; σ_{max} (eq 8) ≤ 0.33 kcal mol^{-1}. (C) Effective binding energies computed by the MM-PBSA approach (PB1) for CDK2 inhibitors with *R* (black bars) or *S* (gray bars) configuration at position 3′. The standard error of the mean (which is < 0.25 kcal mol^{-1}) is not depicted.

close ligands different values for ε_{int} may be required. Finally, the effective binding energies of CDK2 ligands with a (3′*S*)-hydroxyl group were computed to be more favorable than the ones of ligands with a (3′*R*)-hydroxyl group (Figure 2); this trend has already been indicated by Jautelat et al.[16] Thus, the MM-PB(GB)SA method is capable to correctly rank the binding affinities of enantiomers in this case.

MR Data Set. The MR complexes with bound nonsteroidal, dihydropyridine-based antagonists showed a considerable structural variability in the MD simulations, particularly in the region of the receptor around helix 12 (Figure 3A). Although the simulations of the MR complexes were extended to 30 ns to allow for more sampling, and only snapshots from the last 10 ns were considered in the MM-PB(GB)SA calculations, a correlation of r^2_{max} = 0.30 (bootstrapped 95% confidence interval: 0.08 < r^2_{max} < 0.54, p = 0.003, σ_{max} = 0.18 kcal mol^{-1}; Tables 3 and S7) between predicted binding energies and experimentally determined IC_{50} values was obtained. Still, the relative ranking of the inhibitors by effective binding energy was good, as indicated by PI_{max} = 0.61 (Figure 3B). The inaccuracies of the computed binding free energies are likely caused by insufficiently equilibrated ensembles of MR-ligand structures due to the large MR movements. This finding cautions against the use of MM-PB(GB)SA in cases of highly mobile protein−ligand systems.

LIE Analyses. The LIE approach proposed by Åqvist et al.[11] was applied to snapshots from the MD simulations of the complex and the free ligand in solution (Table 2). Although the method has proven to be capable to predict binding free energies with an accuracy that allows to yield good relative binding affinity rankings,[6,49] no satisfactory correlations (r^2 < 0.30) between predicted and experimentally determined binding affinities were obtained for any of the data sets investigated here (Table S8). In addition, the uncertainty in the calculated ΔG_{LIE} energies was σ_{max} = 0.52, 0.97, and 0.49 kcal mol^{-1} for the factor Xa, CDK2, and MR data sets, respectively, and, thus, considerably larger than the one found in the MM-PB(GB)SA analyses. The σ_{max} values were only 4- to 9-fold smaller than the range in the experimental binding affinities across the data sets, so that the uncertainty partially limited the performance of the LIE calculations. In part this is very likely due to the factors discussed above, i.e., the inappropriate description of the halogen−π interaction in the case of FXa and the uncertainties in the MR complex structures. Furthermore, using a large box of explicit waters with counterions, instead of the commonly employed water sphere around the ligand binding site,[12] might introduce additional noise and, thus, might cause a slow convergence of the energy components. The latter is especially relevant for the charged ligands of the CDK2 data set, where counterions had to be added to the ligand simulations. Accordingly, compared to the overall correlation for the CDK2 data set (r^2_{max} = 0.04, p = 0.291, PI_{max} = 0.08, σ_{max} = 0.97 kcal mol^{-1}), a much better correlation (r^2_{max} = 0.31, p = 0.013, PI_{max} = 0.49, σ_{max} = 0.79 kcal mol^{-1}) was obtained when only the neutral inhibitors of this set were considered. In contrast, employing the scaling approach of Hansson et al.[49] that accounts for differences in the electrostatic response between charged and uncharged ligands and takes the effect of hydroxyl groups into account did not improve the overall correlation. This finding again points to inaccuracies in the energy components of charged ligands required to compute the binding free energy (eq 4). In summary, the unsatisfactory performance of LIE can primarily be attributed to the challenges of the data sets investigated here, which is in line

dx.doi.org/10.1021/ct5000296 | J. Chem. Theory Comput. 2014, 10, 3331−3344

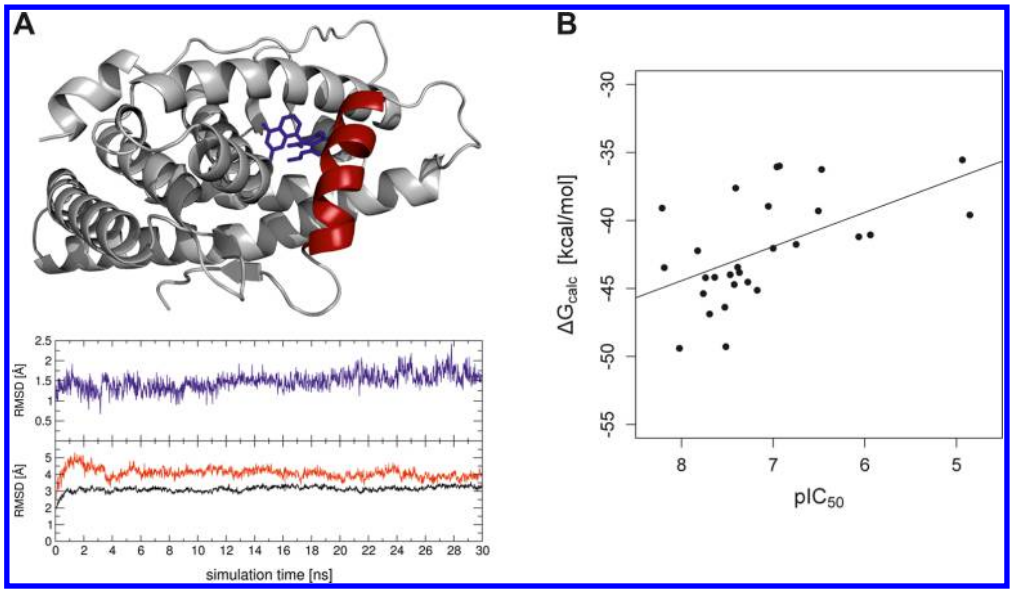


Figure 3. (A) Model structure of MR−M38 complex (top) and RMSD during MD simulations relative to this structure (bottom). RMSDs of the whole complex, helix α12, and ligand M38 are depicted in black, red, and blue, respectively. The RMSD values of helix α12 and the ligand were determined after optimal superposition of the protein part. (B) Effective binding energies (ΔG_{calc}) computed by the MM-GBSA method (GB2) for 28 MR inhibitors versus pIC_{50} values; σ_{max} (eq 8) $\leq$ 0.18 kcal mol^{-1}.

Table 4. Accuracy of $\Delta\Delta G$ Computed by the TI Method[a]

system	considered transformations	simulation interval[b]	r^2	p[c]	RMSD[d]
FXa	5 starting from F5	λ = 0.2, 0.4, 0.6: 2−5 ns; λ = 0.8: 2−10 ns[e]	0.58	0.134	2.1
CDK2	28 (complete ligand set covered)	λ = 0.2, 0.4, 0.6: 2−5 ns; λ = 0.8: 2−10 ns[e]	0.12	0.065	2.1
	20 with highest similarity between end states	λ = 0.2, 0.4, 0.6: 2−5 ns; λ = 0.8: 2−10 ns[e]	0.36	0.005	1.7
	20 with highest similarity between end states	λ = 0.1, 0.2, ..., 0.6: 2−5 ns; λ = 0.7, 0.8, 0.9: 2−10 ns	0.35	0.006	2.3
MR	23 with given IC_{50} values	λ = 0.2, 0.4, 0.6: 2−5 ns; λ = 0.8: 2−10 ns[e]	0.02	0.480	1.8
	23 with given IC_{50} values	λ = 0.2, 0.4, 0.6, 0.8: 0−2 ns	0.32	0.005	1.3
	22 not including M24[f]	λ = 0.2, 0.4, 0.6, 0.8: 0−2 ns	0.44	<0.001	1.1

[a]Accuracy of computed $\Delta\Delta G$ with respect to $\Delta\Delta G$ calculated from IC_{50} values according to $\Delta\Delta G_{experiment} = -RT \ln(IC_{50}^{A}/IC_{50}^{B})$ with T = 300 K. [b]Interval of the TI simulations taken into account in the calculation of $\Delta\Delta G$. [c]p-value of a two-tailed Student's t test to test the null hypothesis that no correlation exists. [d]In kcal mol^{-1} [e]Time interval taken into account for determining $\Delta\Delta G$ by the default calculation procedure. [f]The protonation state of M24 upon binding is not unambiguously known.

with other reports on a system-dependent prediction quality of this method.[75−77]

TI Calculations. *Overall Assessment.* Relative binding free energies ($\Delta\Delta G$) between pairs of compounds were computed by the TI approach using identified "optimal" transformations. "Optimal" refers to as small as possible overall structural changes based on a structural similarity assessment (see Methods for details). When all transformations were conducted with the default calculation procedure (Table 4), neither one of the three data sets yielded satisfactory results: Although RMSD values of 2.1, 2.1, and 1.8 kcal mol^{-1} were obtained for the FXa, CDK2, and MR data sets, respectively (Table 4), a fair correlation between experimentally determined and computed $\Delta\Delta G$ values was only found for the FXa set (r^2 = 0.58, p = 0.134, σ_{max} = 0.08 kcal mol^{-1}). The large p value is a result of the small number of transformations considered.

The fair performance of TI for the FXa data set (Figure 4A, Table S9) is surprising at a first glance because we rather expected that the inappropriate description of the halogen−π interaction mentioned above will deteriorate the prediction quality. The fact that the same initial state was used in all transformations probably led to a cancellation of errors. Furthermore, as the binding mode of the initial state has been characterized by X-ray crystallography, uncertainties arising from modeling the bound states of the other ligands are less pronounced in this case.

Transformations Involving Large Structural Changes. A detailed inspection of the results for the CDK2 data set (Table S10) revealed that large differences between computed and experimentally determined $\Delta\Delta G$ values mainly occurred for transformations involving large structural changes. Increasing the sampling period for those transformations for which the $\Delta\Delta G$ predictions were not satisfactory did not always improve the results, however. When the performance analysis was repeated considering only those 20 transformations with the highest similarity between the end states, a better prediction was found (r^2 = 0.36, bootstrapped 95% confidence interval: 0.01 < r^2_{max} < 0.69, p = 0.005, σ_{max} = 0.21 kcal mol^{-1}, RMSD = 1.7 kcal mol^{-1}, Figure 4B). Thus, in most of the other transformations the changes in the ligand structures were presumably too large as to be accurately accounted for by TI with only 4 λ steps. Furthermore, for dissimilar ligands systematic errors in the TI calculations are less likely to cancel in contrast to what has been reported for structurally similar compounds.[7] Exchanges of long side chains were particularly problematic, e.g., the exchanges of the substituents in the 5-

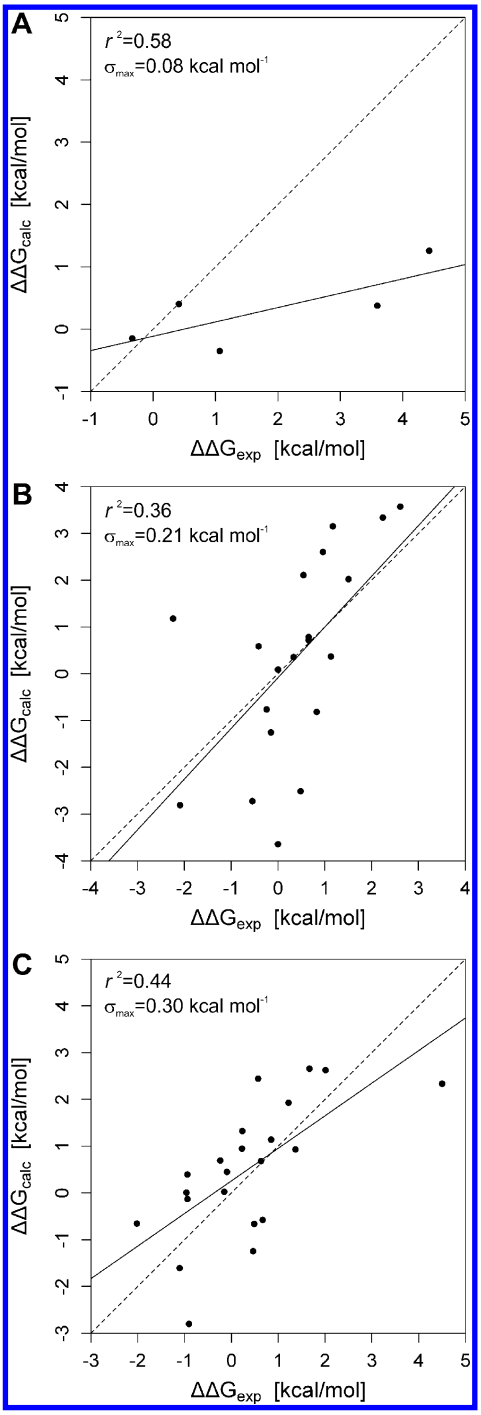


Figure 4. Correlation between $\Delta\Delta G_{exp}$ determined from IC_{50} values by $\Delta\Delta G_{exp} = -RT \ln(IC_{50}^{A}/IC_{50}^{B})$ with T = 300 K and differences in the relative binding free energies computed by the TI approach ($\Delta\Delta G_{calc}$) for 5 transformations of FXa inhibitors (A), 20 transformations with the smallest overall structural changes of CDK2 inhibitors (B), and 22 transformations of MR antagonists (C). The correlation line obtained by a linear least-squares fit of the actual data (solid) and the line of ideal correlation (dashed) are shown in addition. σ_{max} was determined according to eq 10.

position as required for transitions C22 → C35 (replacement of NMe-$(CH_2)_2$-$NH(Me)_2^+$ with NMe-CH_2-$(CHOH)_4$-CH_2OH), C27R → C33 (mostly governed by a transformation of $N(Me)_2$ into $NH(CH_2)_2OH$), and C37R → C27R (change of a hydroxypyridine moiety into a $N(Me)_2$ moiety). These substituents show a considerable degree of mobility in the complex simulations. Therefore, obtaining ensembles that are representative of the conformational space of these ligands will require additional sampling. For charged ligands involved in the transformations, no higher inaccuracy in the determined $\Delta\Delta G$ was observed in general, on the contrary to what was expected.[7,13] In fact, for the C27R → C28R and C8 → C20 transformations rather good predictions were obtained. In summary, consistent with previous reports,[7] the findings for the CDK2 data set show that for similar ligands better $\Delta\Delta G$ estimates can be expected from TI than for diverse ligand sets.

Influence of Sampling. Sampling is known to be a critical factor for the outcome of TI calculations.[13] However, because of the high computational demand of these calculations, we intended to keep the simulation time per λ step as short as possible. For individual transformations of the FXa data set, we found that $\Delta\Delta G$ changed only marginally when $dV/d\lambda$ data collected in the time interval 10−20 ns was considered in addition to data from the first 2−10 ns (Figure 5). Thus, little

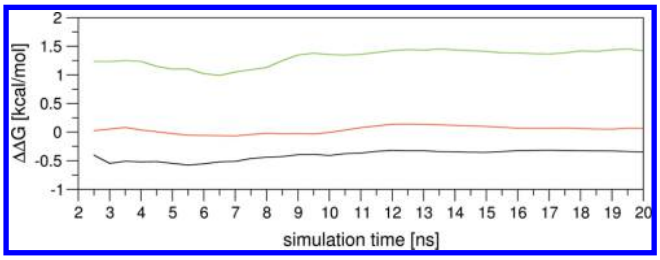


Figure 5. Relative binding free energies $\Delta\Delta G$ calculated considering increasing simulation time intervals in steps of 0.5 ns starting from 2 ns for transformations F5 → F18 (black), F5 → F17 (red), and F5 → F21 (green).

additional information could be gained from sampling more than 10 ns in these cases. As the simulations were started from structures pre-equilibrated for the initial state, the simulation time for the three smallest λ steps could be further reduced to 5 ns without a significant loss in accuracy. In contrast, a further reduction of the sampling time at λ = 0.8 led to a considerable decrease in r^2, indicative of a worse correlation between experimentally determined and calculated $\Delta\Delta G$ values (data not shown). Hence, considering $dV/d\lambda$ values collected in the 2−5 ns interval for the three smallest λ steps and in the interval 2−10 ns for λ = 0.8 resulted in good predictions for the FXa data set. For the CDK2 data set, good predictions were also obtained with these sampling intervals. In contrast, for MR, a plot of the coefficient of the correlation between computed and experimentally determined $\Delta\Delta G$ against the simulation time showed that the prediction accuracy decreased when data sampled beyond 2 ns was taken into account (Figure 6). This behavior probably originates from structural changes in the receptor observed in the course of the simulations; these structural changes are likely a result from the uncertainties in the MR model structure. Therefore, sampling for several nanoseconds is only beneficial for TI analyses if the initial complex structure is of good quality.

Comparison of TI with BAR and MBAR. Some recent studies in which the performance of free energy calculation methods was investigated showed a better performance of the BAR method compared to conventional TI.[78−80] Furthermore, it has been reported that the BAR and MBAR methods decrease the variance of the determined free energies by using

dx.doi.org/10.1021/ct5000296 | J. Chem. Theory Comput. 2014, 10, 3331−3344

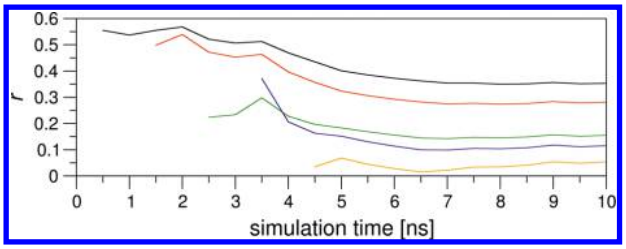


Figure 6. Pearson's correlation coefficient for computed versus experimentally determined $\Delta\Delta G$ for 23 transformations of MR antagonists as a function of increasing simulation data considered in the $\Delta\Delta G$ calculation. The colors of the curves depict when the data consideration started: ≥0 ns: black, ≥1 ns: red, ≥2 ns: green, ≥3 ns: blue, and ≥4 ns: orange. Intervals taken into account in the $\Delta\Delta G$ calculation were increased with a step size of 0.5 ns. Up to 5 ns simulation time, data were sampled for all λ steps, after that only data sampled at λ = 0.8 were additionally taken into account.

the simulation data more efficiently.[59,80,81] Thus, we investigated whether the BAR and MBAR methods result in a more accurate $\Delta\Delta G$ prediction than TI on the 20 smallest transformations of the CDK2 data set. This data set was chosen because a crystal complex structure was available that could be used as a template to build other complex models of a putatively high quality. Furthermore, a fair prediction had already been obtained for the CDK2 data set using TI calculations (Table 4). However, as the RMSD is still above the chemical accuracy of 1 kcal mol^{-1}, which is the expected accuracy in an optimal case,[7] an improvement of the prediction accuracy by the BAR or MBAR methods could be possible. Yet, the performance analysis revealed that the BAR (r^2 = 0.34, RMSD = 2.1 kcal mol^{-1}, Table S12) and MBAR (r^2 = 0.34, RMSD = 2.1 kcal mol^{-1}, Table S12) methods provide results of equivalent quality as TI (r^2 = 0.36, RMSD = 1.7 kcal mol^{-1}, Table 4).

Influence of the Number of Considered λ Values. The prediction accuracy that can be achieved by TI is highly dependent on the shape of the $dV/d\lambda$ curves, which in turn is highly system and transformation dependent. Smooth, linear $dV/d\lambda$ curves (Figure S1A) allow a good extrapolation to the end states and enable an accurate determination of $\Delta\Delta G$ with only few λ values; in contrast, kinks in a $dV/d\lambda$ curve at the terminal λ values (Figure S1B) can deteriorate the extrapolation and the $\Delta\Delta G$ prediction. In addition, minima or maxima in the middle of the curve (Figure S1C) lead to integration inaccuracies, which become more severe with a decreasing number of considered λ values. Finally, also the magnitude of the $dV/d\lambda$ energies has a significant influence on the accuracy of the $\Delta\Delta G$ prediction: For large $dV/d\lambda$ energies even small errors or insufficient sampling at individual λ's can have a large influence on the calculated $\Delta\Delta G$.

To investigate the effect of the number of considered λ values on the $\Delta\Delta G$ prediction accuracy for a data set with transformations with various shapes of $dV/d\lambda$ curves and different magnitudes of $dV/d\lambda$ energies, we conducted for the 20 most similar transformations of the CDK2 data set, in addition to TI calculations with 4 λ values and $\Delta\lambda$ = 0.2, also calculations considering 9 λ values and $\Delta\lambda$ = 0.1. The correlation between experimentally determined and computed binding affinities that was found with the 9 λ values (r^2 = 0.35, RMSD = 2.3 kcal mol^{-1}, Table 4) is very similar to the one that was obtained with the 4 λ values (r^2 = 0.36, RMSD = 1.7 kcal mol^{-1}, Table 4). The unexpected increase in the RMSD in the former case was mainly caused by a kink in the $dV/d\lambda$ curve for the C36 → C37S transformation. When this transformation is not considered the RMSD drops to 1.9 kcal mol^{-1}. In summary, no improvement of the prediction accuracy was obtained for the 20 most similar transformations of the CDK2 data set if the number of considered λ values was increased from 4 to 9. However, if the available computing time permits, higher numbers of λ values are to be recommended in those cases where the $dV/d\lambda$ curves deviate from linearity and/or the magnitude of the $dV/d\lambda$ values is large.

Influence of Protonation States. For two of the MR antagonists, M8 and M24, the protonation state upon binding was not known. TI calculations for the transformations from the protonated to the unprotonated species showed that the unprotonated, neutral species bind with at least 3 kcal mol^{-1} higher affinity than the protonated ones (Table S11). To that an additional contribution must be added if the protonation state of the ligand changes upon binding. Such a contribution can significantly influence $\Delta\Delta G$, as it can be of the same magnitude.[82] Indeed, the $\Delta\Delta G$ predicted for the M23 → $M24_B$ transformation showed a large difference from the experimentally determined $\Delta\Delta G$ value (3.6 kcal mol^{-1}), which is much higher than what has been observed for other transformations involving equally large structural exchanges (<1.9 kcal mol^{-1}). When this transformation was omitted, the accuracy of the prediction increased considerably (r^2 = 0.44, bootstrapped 95% confidence interval: 0.14 < r^2_{max} < 0.60, p < 0.001, σ_{max} = 0.30 kcal mol^{-1}, RMSD = 1.1 kcal mol^{-1}, Figure 4C). This highlights the importance of using ligands in the correct protonation state for binding free energy calculations and/or appropriately accounting for changes in the protonation state upon binding of a ligand.

Uncertainty of the Predictions. The maximum statistical error in $\Delta\Delta G$ from the TI calculations, reported as σ_{max} above, did for neither data set limit the performance of the calculations: In all three cases the difference in experimental binding affinities was at least 21-fold larger than the respective σ_{max} value. Furthermore, uncertainties in the TI calculations were of a similar magnitude as the ones found in the MM-PB(GB)SA analyses.

Blind Predictions. In order to test whether the knowledge gained in the TI performance study allows us to correctly estimate the accuracy of $\Delta\Delta G$ values calculated by TI, relative binding free energies were predicted for 12 additional MR antagonists for which no IC_{50} values were initially provided. To mimic what would have been done in a "real-life scenario", optimal transformations for these ligands were determined taking all other MR ligands into account. We then wanted to see if an experienced person is able to estimate the accuracy of the $\Delta\Delta G$ prediction using indicators he or she has evaluated beforehand on related protein−ligand systems. A combination of the size and chemical structures of the exchanged groups and the statistical error in the computed $\Delta\Delta G$ was considered for this. Based on the estimated accuracy of the predicted $\Delta\Delta G$ values the transformations were classified into four groups from high (group 1) to low accuracy (group 4). For example, the M10 → M30 transformation was assigned to group 1 because the structural changes between start and end state involved only one methyl group, and the standard error in the computed $\Delta\Delta G$ was small (0.07 kcal mol^{-1}). In contrast, it was expected that the accuracy of the $\Delta\Delta G$ predicted for the M2 → M17 transformation (group 4) is low because the structural changes are large (they involve the exchange of a whole phenyl ring),

Table 5. Average RMSD of Accuracy Groups Identified for Blind Test Set

group[a]	transformations assigned to group	average RMSD[b]
1	M11 → M33, M10 → M30, M12 → M31, M4 → M20	0.83
2	M41 → M3, M29 → M18, M1 → M13, M26 → M4, M6 → M41, M29 → M28, M32 → M29	0.91
3	M5 → M18, M25 → M1, M38 → M7, M36 → M3	2.27
4	M2 → M17	5.41

[a]The accuracy of the computed ΔΔG was predicted to decrease from group 1 to 4. [b]In kcal mol^{-1}.

and the standard error is among the eight highest computed for all MR ligand transformations. When the predicted ΔΔG values were compared with experimentally determined values provided thereafter, we found that the average RMSD per group reflected very well the predicted accuracy trend (Table 5). This suggests that it is possible for an experienced person to correctly estimate the accuracy of ΔΔG values computed by TI.

CONCLUSION

In this study, the performance of the MM-PB(GB)SA, the linear interaction energy (LIE), and the thermodynamic integration (TI) approach was investigated with respect to computing relative binding free energies. What makes this study different to others is that three challenging, "real-life" data sets of compounds from industrial lead optimization projects were considered, allowing to ask to what extent the above approaches could have contributed to decision making in that stage of drug development.

While the default calculation protocols led to ligand rankings that were moderate at best, binding free energy predictions of valuable accuracy could be obtained with MM-PB(GB)SA and TI by taking into account data set-specific features. With respect to MM-PB(GB)SA, an overall better performance was found with PB than with GB solvation models. Therefore, we recommend to use MM-PBSA with Parse[38] or mbondi[36] radii for the calculation of effective binding energies if (computing) time permits it. When MM-GBSA calculations shall be performed, we recommend to use GB model I of Onufriev et al.[36] Note, however, that these recommendations are based on our investigations of three data sets and, thus, bear some uncertainty due to system-dependent influences.

Detailed investigations of failures of the MM-PB(GB)SA and TI calculations furthermore revealed dos and do nots that are essential for good binding free energy predictions. These are (i) use only input (complex) structures of high quality, the stability of which has been tested in MD simulations; (ii) carefully test the ability of the force field to properly describe protein−ligand interactions, in particular in the presence of "non-conventional" interactions; (iii) thoroughly investigate properties (e.g., polarity) and potential challenges (e.g., exchange of large substituents) of the ligand sets.

As to the latter, our findings lead to a clear recommendation with respect to which of the above approaches can be applied, and in what order (Figure 7): (i) For data sets of highly similar, neutral compounds, MM-PB(GB)SA analyses can already yield useful effective binding energy predictions; from such an initial ranking, interesting drug candidates can then be selected for further, more detailed TI studies. (ii) For data sets of structurally divergent compounds, we suggest to first determine which transformations require overall the smallest structural changes. Then, TI calculations should be conducted for pairs of sufficiently similar compounds only, while transformations involving large, flexible groups should be avoided. Finally, MM-PBSA calculations can be performed for subgroups of the

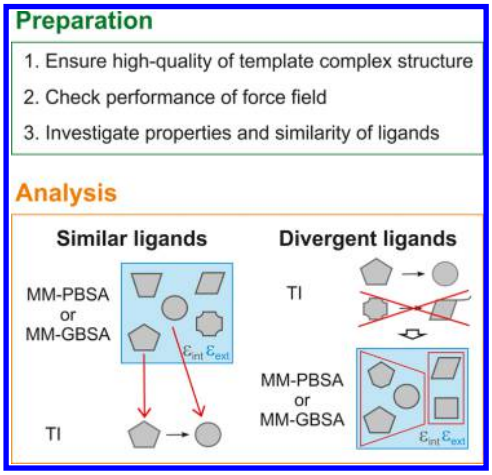


Figure 7. Scheme of the recommended procedure for binding free energy calculations. Ligands are represented as gray shapes.

compounds with similar properties, e.g., with respect to polarity, chemical, or structural composition, to get a rough estimate of binding affinities of compounds not considered so far. Special care must be taken to use appropriate solute dielectric constants for compounds of different polarity in the MM-PB(GB)SA calculations.

The above stresses that early decisions need to be made on a case-by-case basis as to what calculations are reasonable and can be beneficial for the lead optimization process, emphasizing the important role of an expert in this process. However, even when carefully set up, the calculations may not have the desired impact. Thus, whenever possible, we recommend test calculations by an expert for a training set to identify the best procedure for binding free energy calculations in a specific case. Phrased differently, a fully automated computational prediction of relative binding free energies for ligands in the lead optimization phase is not yet feasible, even if tools have become available recently that assist users in the complex and time-consuming setup of such calculations.[47,83] Along these lines, perhaps the most encouraging result of this study is that it is possible for experts to estimate the accuracy of the relative binding free energy predictions by TI calculations using indicators such as the size and chemical structures of the exchanged groups and the statistical error in the predictions. Such knowledge will be invaluable when it comes to the question which of the TI results can be trusted for decision making.

ASSOCIATED CONTENT

Supporting Information

Additional tables showing the investigated ligands, their aliases, and experimentally determined IC_{50} values (Tables S1−S3), applied flavors of MM-PB(GB)SA free energy calculations (Table S4), detailed quality assessments of binding free energies computed with the MM-PB(GB)SA and the LIE

method (Tables S5−S8), relative binding free energies computed by the TI (Tables S9−S11), and BAR/MBAR (Table S12) approaches, sample $dV/d\lambda$ curves (Figure S1), and detailed protocols of the protein system preparation and the MD simulations are provided. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: gohlke@uni-duesseldorf.de. Phone: (+49) 211-81-13662. Fax: (+49) 211-81-13847.

Notes

The authors declare the following competing financial interest(s): F. Stoll and A. Hillisch are employees of Bayer Pharma AG.

ACKNOWLEDGMENTS

We thank B. Schmitz for technical assistance and H. Kopitz and L. Bärfacker for valuable scientific discussions. We are grateful to Bayer Pharma AG for financial support and to the "Zentrum für Informations und Medientechnologie" (ZIM) at the Heinrich Heine University for computational support.

ABBREVIATIONS

MD, molecular dynamics; MM-PBSA, molecular mechanics Poisson−Boltzmann surface area; MM-GBSA, molecular mechanics generalized Born surface area; LIE, linear interaction energy; TI, thermodynamic integration; BAR, Bennett acceptance ratio; FXa, factor Xa; CDK2, cyclin-dependent kinase 2; MR, mineralocorticoid receptor

REFERENCES

(1) Herper, M. The Truly Staggering Cost Of Inventing New Drugs. *Forbes*, **2012**, http://www.forbes.com/sites/matthewherper/2012/02/22/the-truly-staggering-cost-of-inventing-new-drugs-the-print-version/ (accessed May 24, 2014).

(2) Morgan, S.; Grootendorst, P.; Lexchin, J.; Cunningham, C.; Greyson, D. The cost of drug development: A systematic review. *Health Policy* **2011**, *100*, 4−17.

(3) Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discovery* **2010**, *9*, 203−214.

(4) Gilson, M. K.; Zhou, H. X. Calculation of Protein-Ligand Binding Affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21−42.

(5) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, *303*, 1813−1818.

(6) Brandsdal, B. O.; Österberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Åqvist, J. Free Energy Calculations and Ligand Binding. *Adv. Protein Chem.* **2003**, *66*, 123−158.

(7) Michel, J.; Essex, J. W. Prediction of protein-ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 639−658.

(8) Gallicchio, E.; Levy, R. M. Recent Theoretical and Computational Advances for Modeling Protein-Ligand Binding Affinities. *Adv. Protein Chem. Struct. Biol.* **2011**, *85*, 27−80.

(9) Srinivasan, J.; Cheatham, T. E., III; Cieplak, P.; Kolman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.

(10) Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. *Mol. Inf.* **2012**, *31*, 114−122.

(11) Åqvist, J.; Medina, C.; Samuelsson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385−391.

(12) Guitiérrez-de-Terán, H.; Åqvist, J. Linear Interaction Energy: Method and Applications in Drug Design. In *Computational Drug Discovery and Design*; Baron, R., Ed.; Humana Press Inc.: Totowa, NJ, 2012; Vol. *819*, pp 305−323.

(13) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150−160.

(14) Straub, A.; Roehrig, S.; Hillisch, A. Orale, direkte Thrombin- und Faktor-Xa-Hemmer: Kommt die Ablösung für Warfarin, Blutegel und Schweinedärme? *Angew. Chem.* **2011**, *123*, 4670−4686.

(15) Roehrig, S.; Straub, A.; Pohlmann, J.; Lampe, T.; Pernerstorfer, J.; Schlemmer, K. H.; Reinemer, P.; Perzborn, E. Discovery of the Novel Antithrombotic Agent 5-Chloro-N-({(5S)-2-oxo-3-[4-(3-oxomorpholin-4-yl)phenyl]-1,3-oxazolidin-5-yl}methyl)thiophene-2-carboxamide (BAY 59−7939): An Oral, Direct Factor Xa Inhibitor. *J. Med. Chem.* **2005**, *48*, 5900−5908.

(16) Jautelat, R.; Brumby, T.; Schäfer, M.; Briem, H.; Eisenbrand, G.; Schwahn, S.; Krüger, M.; Lücking, U.; Prien, O.; Siemeister, G. From the Insoluble Dye Indirubin towards Highly Active, Soluble CDK2-Inhibitors. *ChemBioChem* **2005**, *6*, 531−540.

(17) Fagart, J.; Hillisch, A.; Huyet, J.; Bärfacker, L.; Fay, M.; Pleiss, U.; Pook, E.; Schäfer, S.; Rafestin-Oblin, M. E.; Kolkhof, P. A New Mode of Mineralocorticoid Receptor Antagonism by a Potent and Selective Nonsteroidal Molecule. *J. Biol. Chem.* **2010**, *285*, 29932−29940.

(18) Bärfacker, L.; Kuhl, A.; Hillisch, A.; Grosser, R.; Figueroa-Pérez, S.; Heckroth, H.; Nitsche, A.; Ergüden, J. K.; Gielen-Haertwig, H.; Schlemmer, K. H.; Mittendorf, J.; Paulsen, H.; Platzek, J.; Kolkhof, P. Discovery of BAY 94−8862: A Nonsteroidal Antagonist of the Mineralocorticoid Receptor for the Treatment of Cardiorenal Diseases. *ChemMedChem* **2012**, *7*, 1385−1403.

(19) *SYBYL-X*, version 1.2; Tripos International: St. Louis, MO, 2010.

(20) Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714−2723.

(21) Guex, N.; Peitsch, M.; Schwede, T.; Diemand, A. *Swiss-PdbViewer*, v4.0.1; Swiss Institute of Bioinformatics & the Biozentrum, University of Basel, Switzerland, 2008.

(22) *The PyMOL Molecular Graphics System*, version 0.99rc6; Schrödinger, LLC: Portland, OR, 2006.

(23) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 11*; University of California: San Francisco, CA, 2010.

(24) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A Computer-based Archival File For Macromolecular Structures. *J. Mol. Biol.* **1977**, *112*, 535−542.

(25) Zhang, B.; Tan, V. B. C.; Lim, K. M.; Tay, T. E. Significance of Water Molecules in the Inhibition of Cylin-dependent Kinase 2 and 5 Complexes. *J. Chem. Inf. Model.* **2007**, *47*, 1877−1885.

(26) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(27) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.

(28) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.

(29) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(30) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2004.
(31) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(32) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(33) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(34) Gohlke, H.; Case, D. A. Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *J. Comput. Chem.* **2004**, *25*, 238−250.
(35) Genheden, S.; Ryde, U. How to Obtain Statistically Converged MM/GBSA Results. *J. Comput. Chem.* **2010**, *31*, 837−846.
(36) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 383−394.
(37) Tsui, V.; Case, D. A. Theory and Applications of the Generalized Born Solvation Model in Macromolecular Simulations. *Biopolymers* **2001**, *56*, 275−291.
(38) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(39) Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U. Ligand Affinities Predicted with the MM/PBSA Method: Dependence on the Simulation Method and the Force Field. *J. Med. Chem.* **2006**, *49*, 6596−6606.
(40) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, *51*, 69−82.
(41) Kuhn, B.; Kollman, P. A. Binding of a Diverse Set of Ligands to Avidin and Atreptavidin: An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models. *J. Med. Chem.* **2000**, *43*, 3786−3791.
(42) Genheden, S.; Kuhn, O.; Mikulskis, P.; Hoffmann, D.; Ryde, U. The Normal-Mode Entropy in the MM/GBSA Method: Effect of System Truncation, Buffer Region, and Dielectric Constant. *J. Chem. Inf. Model.* **2012**, *52*, 2079−2088.
(43) Hou, T.; Yu, R. Molecular dynamics and Free Energy Studies on the Wild-type and Double Mutant HIV-1 Protease Complexed with Amprenavir and Two Amprenavir-related Inhibitors: Mechanism for Binding and Drug Resistance. *J. Med. Chem.* **2007**, *50*, 1177−1188.
(44) Ferrari, A. M.; Degliesposti, G.; Sgobba, M.; Rastelli, G. Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors. *Bioorg. Med. Chem.* **2007**, *15*, 7865−7877.
(45) Del Rio, A.; Baldi, B. F.; Rastelli, G. Activity Prediction and Structural Insights of Extracellular Signal-Regulated Kinase 2 Inhibitors with Molecular Dynamics Simulations. *Chem. Biol. Drug Des.* **2009**, *74*, 630−635.
(46) Brown, S. P.; Muchmore, S. W. Rapid Estimation of Relative Protein-Ligand Binding Affinities Using a High-Throughput Version of MM-PBSA. *J. Chem. Inf. Model.* **2007**, *47*, 1493−1503.
(47) Homeyer, N.; Gohlke, H. FEW: A Workflow Tool for Free Energy Calculations of Ligand Binding. *J. Comput. Chem.* **2012**, *34*, 965−973.
(48) *R: A Language and Environment for Statistical Computing*; R Core Team, R Foundation for Statistical Computing: Vienna, Austria, 2010, http://www.R-project.org (accessed date October 8, 2010).
(49) Hansson, T.; Marelius, J.; Åqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27−35.
(50) Wang, W.; Wang, J.; Kollman, P. A. What Determines the van der Waals Coefficient β in the LIE (Linear Interaction Energy) Method to Estimate Binding Free Energies Using Molecular Dynamics Simulations? *Proteins: Struct., Funct., Bioinf.* **1999**, *34*, 395−402.
(51) Valiente, P. A.; Gil L, A.; Batista, P. R.; Caffarena, E. R.; Pons, T.; Pascutti, P. G. New Parameterization Approaches of the LIE Method to Improve Free Energy Calculations of PlmII-Inhibitors Complexes. *J. Comput. Chem.* **2010**, *31*, 2723−2734.
(52) *ROCS*, version 3.1.1; OpenEye Scientific Software: Santa Fe, NM, http://www.eyesopen.com.
(53) Hawkins, P. C.; Skillman, A. G.; Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, *50*, 74−82.
(54) Kruskal, J. B. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Am. Math. Soc.* **1956**, *7*, 48−50.
(55) Steinbrecher, T.; Joung, I.; Case, D. A. Soft-Core Potentials in Thermodynamic Integration: Comparing One- and Two-step Transformations. *J. Comput. Chem.* **2011**, *32*, 3253−3263.
(56) Genheden, S.; Nilsson, I.; Ryde, U. Binding Affinities of Factor Xa Inhibitors Estimated by Thermodynamic Integration and MM/GBSA. *J. Chem. Inf. Model.* **2011**, *51*, 947−958.
(57) Darden, T.; Pearlman, D. A.; Pedersen, L. G. Ionic charging free energies: Spherical versus periodic boundary conditions. *J. Chem. Phys.* **1998**, *109*, 10921−10935.
(58) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AmberTools*; University of California: San Francisco, CA, 2012.
(59) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
(60) Pearlman, D. A.; Charifson, P. S. Are Free Energy Calculations Useful in Practice? A Comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System. *J. Med. Chem.* **2001**, *44*, 3417−3423.
(61) Canty, A.; Ripley, B. *boot: Bootstrap Functions*, version 1.3.9, 2013, http://cran.r-project.org/web/packages/boot/index.html (accessed date September 9, 2013).
(62) Hou, T.; Guo, S.; Xu, X. Predictions of Binding of a Diverse Set of Ligands to Gelatinase-A by a Combination of Molecular Dynamics and Continuum Solvent Models. *J. Phys. Chem. B* **2002**, *106*, 5527−5535.
(63) Huo, S.; Wang, J.; Cieplak, P.; Kollman, P. A.; Kuntz, I. D. Molecular dynamics and Free Energy Analyses of Cathepsin D-

inhibitor Interactions: Insight into Structure-Based Ligand Design. *J. Med. Chem.* **2002**, *45*, 1412−1419.
(64) Rastelli, G.; Del Rio, A.; Degliesposti, G.; Sgobba, M. Fast and Accurate Predictions of Binding Free Energies using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **2010**, *31*, 797−810.
(65) Guimarães, C. R. W.; Mathiowetz, A. M. Addressing Limitations with the MM-GB/SA Scoring Procedure using the WaterMap Method and Free Energy Perturbation Calculations. *J. Chem. Inf. Model.* **2010**, *50*, 547−559.
(66) Yang, T.; Wu, J. C.; Yan, C.; Wang, Y.; Luo, R.; Gonzales, M. B.; Dalby, K. N.; Ren, P. Virtual screening using molecular simulations. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 1940−1951.
(67) Genheden, S.; Ryde, U. Comparison of end-point continuum-solvation methods for the calculation of protein-ligand binding free energies. *Proteins: Struct., Funct., Bioinf.* **2012**, *80*, 1326−1342.
(68) Wallnoefer, H. G.; Fox, T.; Liedl, K. R.; Tautermann, C. S. Dispersion dominated halogen−π interactions: energies and locations of minima. *Phys. Chem. Chem. Phys.* **2010**, *12*, 14941−14949.
(69) Jorgensen, W. L.; Schyman, P. Treatment of Halogen Bonding in the OPLS-AA Force Field: Application to Potent Anti-HIV Agents. *J. Chem. Theory Comput.* **2011**, *8*, 3895−3901.
(70) Ibrahim, M. A. Molecular mechanical study of halogen bonding in drug discovery. *J. Comput. Chem.* **2011**, *32*, 2564−2574.
(71) Rendine, S.; Pieraccini, S.; Forni, A.; Sironi, M. Halogen bonding in ligand-receptor systems in the framework of classical force fields. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19508−19516.
(72) Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P. Halogen bonding: the σ-hole. *J. Mol. Model.* **2007**, *13*, 291−296.
(73) Wang, J.; Hou, T.; Yu, X. Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. *Curr. Comput.-Aided Drug Des.* **2006**, *2*, 95−103.
(74) Eisenbrand, G.; Hippe, F.; Jakobs, S.; Muehlbeyer, S. Molecular mechanisms of indirubine and its derivatives: novel anticancer molecules with their origin in traditional Chinese phytomedicine. *J. Cancer Res. Clin. Oncol.* **2004**, *130*, 627−635.
(75) Stjernschantz, E.; Marelius, J.; Medina, C.; Jacobsson, M.; Vermeulen, N. P. E.; Oostenbrink, C. Are Automated Molecular Dynamics Simulations and Binding Free Energy Calculations Realistic Tools in Lead Optimization? An Evaluation of the Linear Interaction Energy (LIE) Method. *J. Chem. Inf. Model.* **2006**, *46*, 1972−1983.
(76) Mikulskis, P.; Genheden, S.; Rydberg, P.; Sandberg, L.; Olsen, L.; Ryde, U. Binding affinities in the SAMPL3 trypsin and host-guest blind tests estimated with the MM/PBSA and LIE methods. *J. Comput.-Aided Mol. Des.* **2011**, *26*, 527−541.
(77) Genheden, S. MM/GBSA and LIE estimates of host-guest affinities: dependence on charges and solvation model. *J. Comput.-Aided Mol. Des.* **2011**, *25*, 1085−1093.
(78) Bruckner, S.; Boresch, S. Efficiency of Alchemical Free Energy Simulations. I. A Practical Comparison of the Exponential Formula, Thermodynamic Integration, and Bennett''s Acceptance Ratio Method. *J. Comput. Chem.* **2011**, *32*, 1303−1319.
(79) Bruckner, S.; Boresch, S. Efficiency of Alchemical Free Energy Simulations. II. Improvements for Thermodynamic Integration. *J. Comput. Chem.* **2011**, *32*, 1320−1333.
(80) Shirts, M. R.; Pande, V. S. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *J. Chem. Phys.* **2005**, *122*, 144107.
(81) Kaus, J. W.; Pierce, L. T.; Walker, R. C.; McCammont, J. A. Improving the Efficiency of Free Energy Calculations in the Amber Molecular Dynamics Package. *J. Chem. Theory Comput.* **2013**, *9*, 4131−4139.
(82) Onufriev, A. V.; Alexov, E. Protonation and pK changes in protein-ligand binding. *Q. Rev. Biophys.* **2013**, *46*, 181−209.
(83) Wang, L.; Deng, Y.; Knight, J. L.; Wu, Y.; Kim, B.; Sherman, W.; Shelley, J. C.; Lin, T.; Abel, R. Modeling Local Structural Rearrangements Using FEP/REST: Application to Relative Binding Affinity Predictions of CDK2 Inhibitors. *J. Chem. Theory Comput.* **2013**, *9*, 1282−1293.
(84) Huyet, J.; Pinon, G. M.; Fay, M. R.; Fagart, J.; Rafestin-Oblin, M. E. Structural Basis of Spirolactone Recognition by the Mineralocorticoid Receptor. *Mol. Pharmacol.* **2007**, *72*, 563−571.
(85) Li, Y.; Suino, K.; Daugherty, J.; Xu, H. E. Structural and Biochemical Mechanisms for the Specificity of Hormone Binding and Coactivator Assembly by Mineralocorticoid Receptor. *Mol. Cell* **2005**, *19*, 367−380.
(86) Schoch, G. A.; D'Arcy, B.; Stihle, M.; Burger, D.; Bär, D.; Benz, J.; Thoma, R.; Ruf, A. Molecular Switch in the Glucocorticoid Receptor: Active and Passive Antagonist Conformations. *J. Mol. Biol.* **2010**, *395*, 568−577.
(87) Katz, B. A.; Mackman, R.; Luong, C.; Radika, K.; Martelli, A.; Sprengeler, P. A.; Wang, J.; Chan, H.; Wong, L. Structural basis for selectivity of a small molecule, S1-binding, submicromolar inhibitor of urokinase-type plasminogen activator. *Chem. Biol. (Oxford, U. K.)* **2000**, *7*, 299−312.
(88) Wu, S. Y.; McNae, I.; Kontopidis, G.; McClue, S. J.; McInnes, C.; Stewart, K. J.; Wang, S.; Zheleva, D. I.; Marriage, H.; Lane, D. P.; Taylor, P.; Fischer, P. M.; Walkinshaw, M. D. Discovery of a Novel Family of CDK Inhibitors with the Program LIDAEUS: Structural Basis for Ligand-Induced Disordering of the Activation Loop. *Structure* **2003**, *11*, 399−410.
(89) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.