JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Improving the LIE Method for Binding Free Energy Calculations of Protein−Ligand Complexes

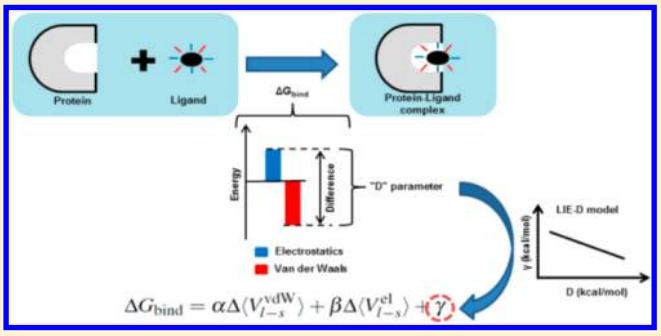
Williams E. Miranda,*[,†] Sergei Yu. Noskov,[‡] and Pedro A. Valiente*[,†]

[†]Computational Biology and Biomolecular Dynamics Laboratory, Center for Protein Studies, Faculty of Biology, University of Havana, Havana, Cuba

[‡]Centre for Molecular Simulations and Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada

Supporting Information

ABSTRACT: In this work, we introduced an improved linear interaction energy (LIE) method parameterization for computations of protein−ligand binding free energies. The protocol, coined LIE-D, builds on the linear relationship between the empirical coefficient γ in the standard LIE scheme and the D parameter, introduced in our work. The D-parameter encompasses the balance (difference) between electrostatic (polar) and van der Waals (nonpolar) energies in protein−ligand complexes. Leave-one-out cross-validation showed that LIE-D reproduced accurately the absolute binding free energies for our training set of protein−ligand complexes (<|error|> = 0.92 kcal/mol, SD_{error} = 0.66 kcal/mol, R^2 = 0.90, Q^2_{LOO} = 0.89, and s^{LOO}_{PRESS} = 1.28 kcal/mol). We also demonstrated LIE-D robustness by predicting accurately the binding free energies for three different protein−ligand systems outside the training data set, where the electrostatic and van der Waals interaction energies were calculated with different force fields.



1. INTRODUCTION

Computational chemistry and molecular modeling have become an indispensable part of the modern drug design process. However, the calculation of absolute binding affinities for protein−ligand complexes remains as one of the main challenges in computational structure-based ligand design. Several approaches to direct evaluation of binding affinities from molecular modeling have been developed, ranging from empirical and "'knowledge-based"' scoring functions to those based on free energies calculations, such as the rigorous free energy perturbation (FEP) and thermodynamic integration (TI) methods.[1,2] Although FEP and TI approaches have shown to be accurate for binding affinity calculation, they are exceedingly demanding in computational power if applied to large molecular data sets.[3,4] Thus, further development of fast and accurate methods for structure-based drug design is still needed.

Åqvist et al.[5] developed a semi-empirical method, coined the linear interaction energy (LIE) method, for absolute binding free energy calculation. This method is based on conformational sampling of receptor-bound and unbound drug states with molecular dynamics (MD) or Monte Carlo (MC) simulations. By virtue of sampling two endpoints of the process only, LIE is considerably faster than either FEP or TI simulations. However, it is considerably slower than single conformation scoring functions methods, thus presenting a very useful alternative (to FEP simulations) for already generated and curated sets of protein−drug states from docking simulations.[5] The LIE method is based on the linear response assumption[6] for electrostatic interactions combined with an empirical expression for nonpolar contributions to drug solvation and binding. The binding free energy in LIE is expressed according to eq 1:

$$\Delta G_{bind} = \beta\Delta\langle V^{el}_{l-s}\rangle + \alpha\Delta\langle V^{vdw}_{l-s}\rangle + \gamma \quad (1)$$

where V^{el}_{l-s} and V^{vdw}_{l-s} are MD or MC-generated interaction energy averages from the nonbonded electrostatic and van der Waals interactions of the ligand (l) with its surrounding environment (s), respectively. The symbol Δ denotes the change in average values when transferring the ligand from solution (free state) into the binding site of the solvated receptor (bound state). The coefficients α and β are scaling factors for these energy terms, while γ is an empirical constant. The linear response (LR) approximation theory provides a physical basis for treating the electrostatics contribution to the binding free energy, predicting a value of β = 0.5. The formalization of other terms related to nonelectrostatic effects in the protein−drug associations is, however, more challenging. Accordingly, the automatic workflows to facilitate the set up and execution of LIE-based binding free energy calculations for protein−ligand complexes have been developed.[7,8] The van der Waals interactions between the ligand and its environment, represented by a Lennard−Jones potential, are commonly used

Received: January 8, 2015
Published: July 16, 2015

DOI: 10.1021/acs.jcim.5b00012
J. Chem. Inf. Model. 2015, 55, 1867−1877

to calculate the nonpolar contribution to binding free energy.[5] However, an *a priori* prediction of α or γ parameters remains a challenge. The commonly used strategy builds on the empirical parameters fitting by available experimental data on a small set of receptor−ligand complexes and then extending it to the test sets of interest.[9−14] For example, the empirical α coefficient was initially calibrated by using β = 0.5 and four endothiapepsin inhibitors experimental binding data yielding the value of α = 0.161.[5] Next, series of MD simulations were performed with the Gromos 96 force field furthering the proposed value of the α.[15] The original model showed reasonable predictive power for different proteins in complexes with ligands with different structural scaffolds such as endothiapepsin,[5] HIV-1 protease,[6,16] glucose binding protein,[17] and trypsin.[18] Later, this parameterization was refined by Åqvist et al.,[11,12] who used 18 protein−ligand complexes comprising endothiapepsin,[5] HIV-1 protease,[6,16] glucose binding protein,[17] and trypsin[18] as the training set. The authors determined more accurate values of β by using FEP calculations. The deviation from the linear-response regime is especially pronounced for relatively polar compounds.[12,19] This resulted in an improved LIE model known as the *standard* model, referred here as LIE-S. This model uses β values, ranging between 0.33 and 0.5, along with α = 0.18 and γ = 0. Further studies have shown that binding free energies calculated with LIE-S are in good agreement with experimental data for several protein−ligand systems.[20−24]

Nevertheless, for proteins containing hydrophobic binding sites, a non-zero γ constant is required to reproduce observed trends in the absolute binding free energies. The commonly used value of α = 0.18 downscales the nonpolar contribution (αV_{l-s}^{vdw}), thus leading to a significant underestimation of binding energies.[9,10,13] Some notable cases where LIE fails to predict are to be found in binding of retinoids to retinol binding protein (RBP),[13] biotin analogs to avidin,[9] substrates to cytochrome P450 (P450cam),[10] and inhibitors to human thrombin.[25] For these systems, reported γ values range from −2.9 to −7.0 kcal/mol.[14] Almlöf et al.[22] found a clear relationship between the ranking of these binding sites hydrophobicity (RBP > P450cam > thrombin > trypsin) and the γ value. To some extent this is similar to the idea developed by Wang et al.,[9] who investigated variations of the nonpolar coefficient α in the absence of the constant term γ, as a way to distinguish among different binding-site types. The main outcome of their research is the linear correlation (R^2 = 0.96) obtained between the weighted nonpolar desolvation ratio (WNDR) and the α values in the LIE method.[9] Valiente et al.[26] established a linear relationship (R^2 = 0.85) between WNDR and γ using Wang's training set complexes. This novel model, termed LIE-C, was successfully applied to predict the binding free energy of five PlmII-Inhibitor complexes (<|error|> = 1.47 kcal/mol).[26] Nevertheless, the main disadvantage of LIE-C is the use of atomic desolvation parameters for WNDR calculation. Furthermore, the only six atom types (C, S, N, N^+, O, and O^-) considered in that report[26] exclude other heteroatoms (Fe, Zn, Cu, F, Br, I, etc.) that could also be present in protein−ligand complexes.

To overcome the limitations of our previous parametrizations based on the linear relationship between WNDR and γ coefficient, we now developed a new parameterization model of LIE. This model, coined LIE-D, is based on the linear correlation between the γ coefficient and the D parameter that accounts for the balance (difference) between the polar and nonpolar contributions to the binding free energy in protein−ligand complexes calculated from MD simulations. For our training set, we chose 24 protein−ligand complexes whose three-dimensional (3D) structures and binding free energies have been experimentally determined. To reproduce each complex experimental binding free energy, we optimized the γ coefficient by fixing α at 0.18 and used β coefficient determined by FEP methodology.[27] To evaluate LIE-D overfitting, we calculated the binding free energy for each training set complex by using leave-one-out cross-validation (LOO−CV) method.[28] We also assessed LIE-D robustness through binding free energy calculations for protein−ligand systems outside the training set.[29,30] According to our results, LIE-D reproduced well the experimental data for these systems.

2. METHODS

2.1. Protein−Ligand Systems. The 24 protein−ligand training set complexes comprised different enzymes and transport proteins (Table 1). They were selected based on the following criteria: (i) 3D structures already deposited at PDB,[31] (ii) protein−ligand complexes bound by noncovalent interactions, (iii) availability of experimental inhibition (K_i), affinity (K_a), or dissociation (K_d) constant, (iv) experimental ΔG_{bind} values ranging from −4 to −16 kcal/mol (noncovalent binding energies standard range),[32] and (v) no bound cofactors.

2.2. MD Simulations. All MD simulations were carried out for 10 ns, using the GROMACS (version 4.5.6)[53] software package, periodic boundary conditions (PBC), and the Amber99sb force field.[54] Protonation states of protein residues were calculated using PROPKA[55] (http://propka.ki.ku.dk/). Ligand parameters were generated from the general Amber force field (GAFF),[56] and charges were calculated using AM1-BCC, a semi-empirical approximation available in the Antechamber program.[57,58] All systems were neutralized (Na^+/Cl^-) and solvated by explicit water molecules, which were modeled by the TIP3P[59] parameter set, in a rhombic dodecahedron box. The distance between the protein−ligand complexes and the edge of the box were set to 10 Å. The LINCS algorithm was used to constrain all the covalent bonds in nonwater molecules,[60] while the SETTLE algorithm was used to constrain bond lengths and angles in water molecules.[61] Before the production run, systems were relaxed by 1000 steps using a steepest descent algorithm followed by other 1000 steps by conjugate gradient method, where the protein was held fixed by a 10 kcal/mol Å^2 constraint. Then, systems were gradually heated for 300 ps to reach the experimentally reported assay temperature (for selected systems assay temperature range between 290 and 310 K) followed by 200 ps of production run without constraints. The Leapfrog scheme,[62] with an integration time step of 2 fs, was employed to integrate the equations of motion. Temperature was controlled using a weak coupling to a bath with a time constant of 0.1 ps.[63] For pressure control, a Berendsen coupling algorithm with a time constant of 1.0 ps was employed.[64] Initial velocities were randomly generated from a Maxwell distribution at 1 K, in accordance with the atomic masses.

For the production run, Langevin dynamics,[65] with an integration time step of a 2 fs scheme, was employed to integrate the motion equations. Temperature was controlled using a weak coupling to a bath with a time constant of 2.0 ps. For pressure control, a Parrinello−Rahman[66] coupling algorithm with a time constant of 5.0 ps was employed. Long range electrostatic interactions were handled by Particle Mesh

DOI: 10.1021/acs.jcim.5b00012
J. Chem. Inf. Model. 2015, 55, 1867−1877

Table 1. Training Set Protein–Ligand Complexes and Their Experimental Binding Free Energy

protein	organism	PDB code[31]	Ligand ID[a]	ΔG^{b}_{exp} [kcal/mol]
retinol binding protein	*Bos taurus*	1ERB[33]	ETR	−10.84[33]
FK 506 binding protein	*Homo sapiens*	1fkg[34]	SB3	−10.34[34]
FK 506 binding protein	*Homo sapiens*	1fkh[34]	SBX	−10.54[34]
FK 506 binding protein	*Homo sapiens*	1FKI[34]	SB1	−9.05[34]
lysozyme S99A	Bacteriophage T4	1L83[35]	BNZ	−5.18[36]
lysozyme L99A	Bacteriophage T4	182L[36]	BZF	−5.45[36]
lysozyme L99A	Bacteriophage T4	183L[36]	DEN	−5.12[36]
lysozyme L99A	Bacteriophage T4	187L[36]	PXY	−4.66[36]
lysozyme L99A	Bacteriophage T4	188L[36]	OXE	−4.74[36]
lysozyme L99A	Bacteriophage T4	1NHB[36]	PYL	−5.75[36]
HIV-1 protease	HIV-1	3MXD[37]	K53	−12.23[38]
HIV-1 protease	HIV-1	3GI5[39]	K62	−15.53[38]
HIV-1 protease	HIV-1	2QI3[38]	MZ5	−14.12[38]
HIV-1 protease	HIV-1	2I0A[40]	MUI	−15.78[38]
plasmepsin II	*Plasmodium falciparum*	2BJU[41]	IH4	−12.09[26]
plasmepsin II	*Plasmodium falciparum*	1LF3[42]	EH5	−11.88[24]
plasmepsin II	*Plasmodium falciparum*	1LF2[43]	R37	−10.65[24]
plasmepsin IV	*Plasmodium malariae*	2ANL[44]	JE2	−9.86[44]
immunoglobulin 39-A11	*Mus musculus*	1A4K[45]	FRA	−10.89[45]
penicillin acylase	*Escherichia coli*	1k5q[46]	PAA	−4.10[46]
fatty acid binding	protein *Rattus norvegicus*	2IFB[47]	PLM	−7.71[48]
ketosteroidisomerase	*Pseudomonas testosteroni*	3NHX[49]	ASD	−5.21[50]
trypsin	*Bos taurus*	3ATL[51]	BEN	−6.30[52]
trypsin	*Bos taurus*	1UTN[52]	ABN	−4.70[52]

[a]2D Structures and IUPAC nomenclature of all ligands are shown in Figure S1. [b]Experimental binding free energies were obtained considering $\Delta G_{exp} = -RT \ln K_x$, where $x = i$, a, or d for inhibition, affinity, and dissociation constants, respectively.

Ewald (PME) summation.[67,68] The van der Waals interactions were modeled by Lennard–Jones (6–12) potential.[53] As GOMACS is not straightforward to obtain long-range electrostatic contributions from the PME algorithm, these were recalculated using the Reaction Field zero algorithm over the generated trajectory.[53] All ligands were solvated in a rhombic dodecahedron box with the same size of the one generated for its respective protein–ligand complex. Finally, 5 ns simulations were run using the above protocol.

To ensure that each system was sufficiently equilibrated before extracting the coordinates for energy calculations, the backbone RMSDs were analyzed. Also, each simulation was visually inspected to verify its correct behavior.

2.3. Development of LIE-D Model. In this work, the LIE formula takes into account the intra-ligand electrostatic interactions as suggested by Almlöf et al.[27] in an earlier work and takes the form of eq 2:

$$\Delta G_{bind} = \beta(\Delta\langle V^{el}_{l-s}\rangle + \Delta\langle V^{el}_{l-l}\rangle) + \alpha\Delta\langle V^{vdw}_{l-s}\rangle + \gamma \quad (2)$$

where $\Delta\langle V^{vdw}_{l-s}\rangle$ denotes the change in intra-ligand electrostatic interactions when it is transferred from solution (free state) into the solvated receptor binding site (bound state). The thermodynamic cycle is shown in Figure S2.

We calculated a β_{FEP} specific values for each ligand using the parameterization model E proposed by Almlöf et al.[27] (Table 2) by using eq 3:

$$\beta_{FEP} = \beta_0 + \frac{\Sigma_i w_i \Delta\beta_i}{\Sigma_i w_i} \quad (3)$$

Table 2. Optimized β Coefficients for Model E Proposed by Almlöf et al.[27]

β_0	0.43
$\Delta\beta_1$ (alcohols)	−0.06
$\Delta\beta_2$ (1,2-amines)	−0.04
$\Delta\beta_3$ (1-amides)	−0.02
$\Delta\beta_4$ (COOH)	−0.03
$\Delta\beta_5$ (anions)	0.02
$\Delta\beta_6$ (cations)	0.09
$\Delta\beta_7$ (others)	0

where w_i, β_0, and $\Delta\beta_i$ were calculated from explicit solvent FEP calculations of single chemical groups (Table 2). This model uses $w_i = 1$ for all neutral groups and one single weighting factor for anions and cations ($w_i = 11$).[27] We chose this parameterization scheme as it allows us to calculate β values for different chemical groups (Table 2), overcoming the limitations of other (less flexible) parameterization schemes.[5,12] From now on, we will refer β_{FEP} values from eq 3 as β in eq 2. We also employed $\alpha = 0.18$, as it seems to be a robust value from previous works.[11,12,20−24]

The balance (difference) between electrostatic (polar) and van der Waals (nonpolar) contributions to binding free energy in the LIE-method was defined as the D parameter eq 4:

$$D = \beta(\Delta\langle V^{ell}_{l-s}\rangle + \Delta\langle V^{el}_{l-l}\rangle) - \alpha\Delta\langle V^{vdw}_{l-s}\rangle \text{ [kcal/mol]} \quad (4)$$

The optimal γ coefficient value required to reproduce each complex experimental binding free energy was calculated from eq 2. The β and γ coefficients values, electrostatics, and van der Waals contributions and D parameter for the 24 training data set complexes are reported in Table S1.

2.4. Statistical Analysis. To measure the LIE-D model's overall performance relative to experimental binding free energies (ΔG_{calc} vs ΔG_{exp}), we calculated the multiple determination coefficient (R^2). This coefficient measures the variance proportion from the experimental binding data that is "explained" as model response to the simulation input data eq 5.

$$R^2 = 1 - \frac{SSE}{SST} \quad (5)$$

where $SSE = \sum_i(\Delta G_i^{calc} - \Delta G_i^{exp})^2$ is the residual unexplained square sum of errors, and $SST = \sum_i(\Delta G_i^{exp} - \langle\Delta G_i^{exp}\rangle)^2$ is the square sum of deviations explained. ΔG_i^{calc} and ΔG_i^{exp} are the calculated and experimental binding free energies for complex i, respectively. A value of $R^2 = 1$ represents a perfect match between calculated and experimental free energies. The average absolute error (<lerrorl>) and the standard deviation of the

DOI: 10.1021/acs.jcim.5b00012
J. Chem. Inf. Model. 2015, 55, 1867−1877

error (SD_{error}) were calculated as a measure of deviations, in kcal/mol, between the optimized models and the experimental values (Table 1).

Data overfitting was assessed by the leave-one-out (LOO) cross-validation (CV)[28] using SOLVERSTAT software.[69] Optimal parameters were found for each 24 training data set systems missing one of the protein−ligand complexes. Each resulting parameterization model was used to predict the left out ΔG_{bind} from the left out simulation data. The LOO−CV coefficient, Q^2_{LOO}, assesses the predictivity of a model and is calculated from

$$Q^2_{LOO} = 1 - \frac{PRESS}{SST} \quad (6)$$

where PRESS is the Predictive REsidual Sum of Squares, $PRESS = \sum_j (\Delta G_j^{calc} - \Delta G_j^{exp})^2$. In this case, ΔG_j^{calc} is calculated from a LIE model optimized on all data points except complex j. The cross-validated standard deviation is calculated as $s^{LOO}_{PRESS} = (PRESS/(n - p - 1))^{1/2}$ [kcal/mol], where $n = 24$ is the data set size, and p is the number of optimized parameters in the model ($p = 2$).

3. RESULTS AND DISCUSSION

3.1. Development and Assesment of LIE-D Model. The standard parameterization of LIE equation (LIE-S) with α fixed at 0.18 and γ at 0 was derived from a set of 18 protein−ligand complexes where the binding sites were all rather hydrophilic.[12] This model yielded an effective value of $\gamma = \gamma_p - \gamma_w \approx 0$, showing that the hypothetical solvation energy vs size intercept in the protein (p) and in water (w) are very close.[12] Nevertheless, an effective non-zero γ coefficient is required for hydrophobic binding sites in order to reproduce the absolute binding free energy.[22] We recently observed this trend for plasmepsin II, a flexible aspartic protease, when the protein binding site nonpolar groups were buried after binding of different inhibitors.[26] However, up to now an efficient computational approach to predict the γ coefficient in the LIE method is not available.

Here, we developed LIE-D, a novel approach based on the linear correlation between the γ coefficient and the D parameter that accounts for the balance (difference) between the polar and nonpolar binding free energy contributions (eq 4). For each complex, the D parameter value was calculated by introducing a β value based on the parameterization model E of Almlöf et al.[27] and α fixed at 0.18. Table S1 summarizes the γ, D, $\Delta\langle V^{el}_{l-s}\rangle$, $\Delta\langle V^{el}_{l-l}\rangle$, and $\Delta\langle V^{vdw}_{l-s}\rangle$ values calculated for each training set complex, together with the experimental binding free energies.

As shown in Figure 1, γ vs D displays a better correlation coefficient ($R^2 = 0.96$) than WNDR vs γ ($R^2 = 0.85$) reported in the LIE-C model.[26] We hypothesized that the D parameter is better than WNDR to predict γ based on the following: (i) D was calculated directly from the average electrostatic and van der Waals energies obtained from isothermic−isobaric ensemble averages generated from MD simulations, while WNDR is calculated from a single structure (docking model or X-ray structure)[9,26] (ii) The correlation between D and γ was obtained from a larger training set. (iii) Atomic solvation parameters were not required for D parameter calculation.

According to data in Table 3, LIE-D (<|error|> = 0.92 kcal/mol, SD_{error} = 0.66 kcal/mol, and $R^2 = 0.90$) outperformed LIE-S (<|error|> = 9.37 kcal/mol, SD_{error} = 4.64 kcal/mol ,and $R^2 = 0$) to reproduce the training set absolute binding free energies

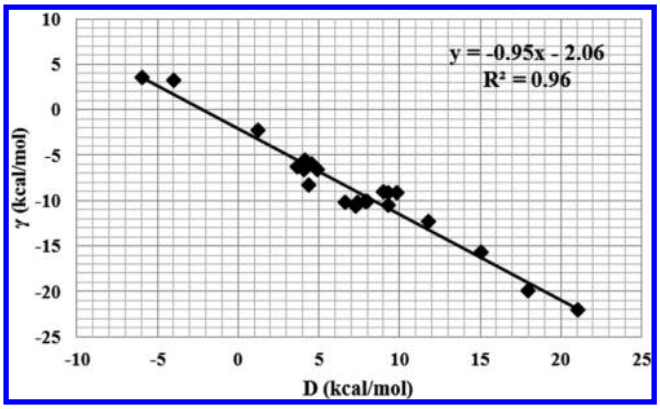


Figure 1. Scatter diagram of γ coefficient vs D parameter values for the 24 training data set protein−ligand complexes. Black solid line with black filled diamonds represents the linear relationship ($R^2 = 0.96$) found in our parameterization model LIE-D.

(Figure 2A, Table 3). Furthermore, data overfitting assessment showed a good LIE-D model performance ($Q^2_{LOO} = 0.89$ and $s^{LOO}_{PRESS} = 1.28$ kcal/mol, see Table 3). These results suggest that the D parameter is useful for γ coefficient estimation in the LIE method. The relationship between the γ coefficient and D parameter takes the form:

$$\gamma = f \times D + g \ [\text{kcal/mol}] \quad (7)$$

where f and g are the slope (−0.95) and intercept (−2.06) respectively of the linear regression in Figure 1.

We also assessed the effect of intra-ligand electrostatic interaction energies in the accuracy of LIE-D, despite this energy contribution have been neglected in most of other previous studies.[5,12,20,26] Neglecting this contribution in eq 4 did not affect the LIE-D model performance (Figure 2B and Table 3), obtaining <|error|> = 0.92 kcal/mol, $SD_{erro\ r}$= 0.69 kcal/mol, $R^2 = 0.90$, $Q^2_{LOO} = 0.89$, and $s^{LOO}_{PRESS} = 1.30$ kcal/mol.

Carlsson et al.[29] found a similar trend in LIE binding free energy calculations for non-nucleoside HIV-1 reverse transcriptase inhibitors (NNRTIs). The authors concluded that including intra-molecular energies did not improve the accuracy of the standard LIE method,[29] as also observed for our LIE-D model results (Figure 2B and Table 3).

3.2. System Dependence of γ Coefficient and D Parameter. In our work, the optimal γ coefficient values needed to reproduce the training data set binding free energies spans from −22.03 to 3.51 kcal/mol (Table S1). These extremes values correspond to HIV-1 protease (highly plastic binding site)[39] and trypsin (highly polar binding site)[52] respectively. A similar trend was observed by Almlöf et al.,[22] who found that the γ coefficient value decreases as the protein binding site hydrophobicity increases. The authors explained these results suggesting that the hypothetical solvation energy vs size intercept in the protein (p) and in water (w) are no longer the same for hydrophobic binding sites ($\gamma = \gamma_p - \gamma_w < 0$).[22]

Likewise, the γ coefficient displays different values for the same receptor when binding to different ligands. (Table S1). For HIV-1 protease−inhibitor complexes, γ ranged from −22.03 to −5.99 kcal/mol. These extreme values correspond to MUI and K53 ligands, respectively (Table S1). Although both ligands share almost the same chemical scaffold, they have one and two hydroxyl groups, respectively (Figure S1). These results suggest that the γ coefficient not only depends on the

DOI: 10.1021/acs.jcim.5b00012
J. Chem. Inf. Model. 2015, 55, 1867−1877

Table 3. Summary of Statistical Figures of Merit for Free Energy Estimation Models

model	<lerrorl> [kcal/mol]	SD_{error} [kcal/mol]	R^2	Q^2_{LOO}	s^{LOO}_{PRESS} [kcal/mol]
LIE-S	9.37	4.64	0.00	–	–
LIE-D	0.92 (0.91)*	0.66 (0.69)*	0.90 (0.90)*	0.89 (0.89)*	1.28 (1.30)*

*Reffited LIE-D model neglecting intra-ligand electrostatic contributions.

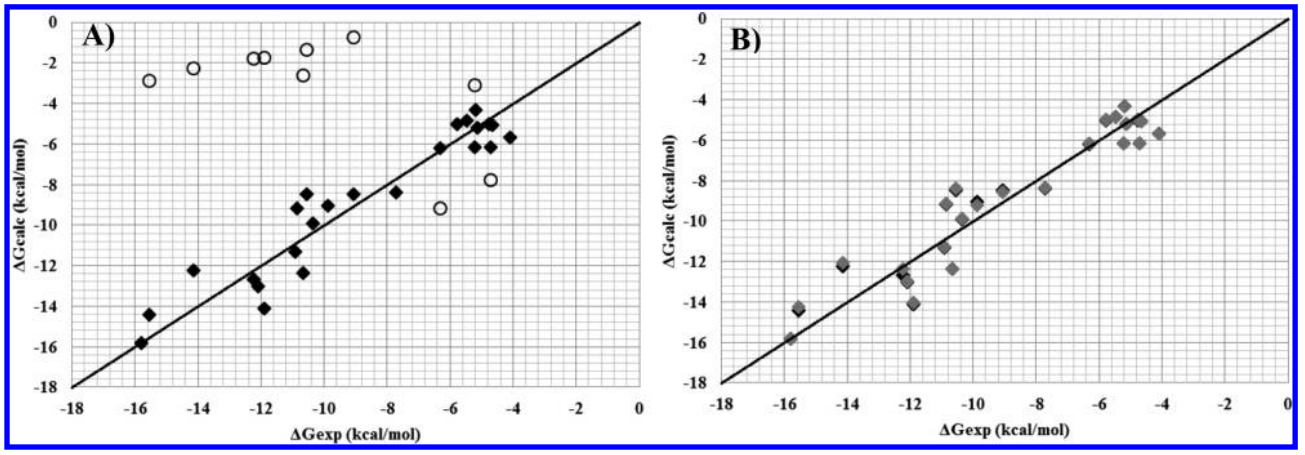


Figure 2. Scatter diagram of calculated (ΔG_{calc}) vs experimental (ΔG_{exp}) binding free energies for the 24 training data set complexes. (A) Using LIE-S (empty circles, $R^2 = 0$), and LIE-D (black solid diamonds, $R^2 = 0.90$) models. (B) LIE-D model considering (black solid diamonds, $R^2 = 0.90$) and neglecting (gray solid diamonds, $R^2 = 0.90$) intra-ligand electrostatic contribution to binding free energy.

physical–chemistry nature of the receptor binding site but also on the ligand one, as well as the binding mode between both molecules. Although narrower in range, the γ coefficient also varies for plasmepsin II (from −12.38 to −9.17 kcal/mol) and FK506 binding protein (from −10.11 to −8.30 kcal/mol, see Table S1). In both cases, the lowest γ coefficient value corresponds to the series' most hydrophobic ligand (IH4 and SB1 respectively, see Figure S1). For lsozyme complexes, the γ values spanned from −6.70 to −5.58 kcal/mol (Table S1). The small range observed for these complexes could be related to the ligands structural similarity (Figure S1). On the other hand, the γ coefficient displayed positive values for both trypsin complexes (3.18 and 3.51 kcal/mol), where ligands were small and positively charged, in other words, highly polar ligands (Figure S1).

Similar trends were obtained by Valiente et al.[26] when studying plasmepsin II complexes with ligands having different chemical scaffolds. For those systems, the optimized γ coefficient values ranged from −13.32 to 0.67 kcal/mol, where the positive value was calculated for pepstatin A, a highly polar and peptidic inhibitor and the lowest negative value for IH4, the most hydrophobic inhibitor in the series.[26] Recently, Linder et al.[70] modeled a series of esters in a lipase from *Rhizomucor miehei* and suggested that the γ coefficient must also be ligand dependent. They used an ensemble-averaged surface area term ($\gamma = f(\mathrm{SASA})$), originally proposed by Carlson and Jorgensen.[71]

So far, the γ coefficient values analysis in our work and others[22,26,70] has shown its importance for the absolute ΔG_{bind} calculation in the LIE method. Furthermore, its value is related to the prevalence of polar or nonpolar interactions in protein–ligand complexes.[22,26] On the other hand, we found a linear relationship between the γ coefficient and D parameter (Figure 1). We previously defined the D parameter as the balance (difference) difference between electrostatic (polar) and van der Waals (nonpolar) contributions to binding free energy eq 4. Then, the linear relationship in Figure 1 should help to understand the observed γ coefficient values system dependence. For example, those protein–ligand complexes where nonpolar contribution to binding free energy prevails, the $\alpha\Delta\langle V^{vdw}_{l-s}\rangle$ term in eq 4 is more negative than $\beta[\Delta\langle V^{el}_{l-s}\rangle + \Delta\langle V^{el}_{l-l}\rangle$, then $\beta[\Delta\langle V^{el}_{l-s}\rangle + \Delta\langle V^{el}_{l-l}\rangle > \alpha\Delta\langle V^{vdw}_{l-s}\rangle$. So, based on eq 4), D parameter values are positive, while (based on the linear relationship shown in Figure 1) γ coefficient values are negative. The opposite happens if polar contribution to binding prevails (D parameter values are negative while γ coefficient values are positive).

In the case of lysozyme–ligand complexes, the nonpolar contribution to binding free energy prevail ($\beta[\Delta\langle V^{el}_{l-s}\rangle + \Delta\langle V^{el}_{l-l}\rangle > \alpha\Delta\langle V^{vdw}_{l-s}\rangle$), and D parameter values are positive, while γ coefficient values are negative (Table S1). These ligands (Figure S1) and the cavity they bind in Lysozyme are nonpolar,[36] and this is reflected in the positive and negative values of the D parameter and γ coefficient, respectively (Table S1). Conversely, for both trypsin–ligand complexes, the polar contribution to ΔG_{bind} prevails ($\beta[\Delta\langle V^{el}_{l-s}\rangle + \Delta\langle V^{el}_{l-l}\rangle < \alpha\Delta\langle V^{vdw}_{l-s}\rangle$), and D parameter values are negative, while γ coefficient values are positive (Table S1). In this case, ligands are small and positive (Figure S1), and trypsin is known for its highly polar binding site.[9,22,52] That is why polar contribution to binding free energy prevail, and this is reflected in the negative and positive sign of D parameter and γ coefficient, respectively (Table S1).

In this work, the D parameter was not derived from first-principles but introduced as a fit parameter eq 4). We hypothesize that its relation to γ coefficient (Figure 1 is based on the prevalence of polar or nonpolar interactions in protein–ligand complexes eq 4. This means that both γ coefficient and D parameter reflects the physico-chemical nature of the prevalent interactions in the system. This is the strength of the LIE-D model as you can predict the γ coefficient value from the balance of polar and nonpolar contributions (D parameter, eq 4) to binding free energy calculated from the MD simulations. Why there is apparent system-dependence of the D

DOI: 10.1021/acs.jcim.5b00012
J. Chem. Inf. Model. 2015, 55, 1867–1877

Table 4. Summary of Errors When Applying Different LIE Models to External Data Sets

receptor	force field	error [kcal/mol]	LIE-S	LIE-D	LIE[a] specific
CYP1A2[30]	GROMOS45A4[79]	<lerrorl>	8.07	0.92	0.65
(13 ligands)		SD_{error}	2.10	0.58	0.39
HIV1-RT[29]	OPLS[80]	<lerrorl>	10.19	1.34 (0.94)[b]	0.83
(39 ligands)		SD_{error}	0.98	1.16 (0.60)[b]	0.61
Kv1.5[78]	AMBER99SB[54]	<lerrorl>	6.69	1.03	0.84
(5 ligands)		SD_{error}	0.74	0.76	0.61

[a]LIE model obtained using the test sets as training sets in the original reports. [b]Errors obtained neglecting the six main outliers.

parameter? The answer could be related to the physical nature of the γ coefficient in the first place. The nonpolar part of the LIE expression is driven by the dispersive component of the van der Waals interactions balanced by penalties associated with cavity formation in solvent and in the receptor, commonly considered in the γ coefficient.[22,26,70−72] The cavity formation free energy has an opposite sign and is apparently anticorrelated to the van der Waals term, which is in turn not completely independent of the electrostatic part of the expression.[73] Furthermore, the change in Lennard−Jones solute−water interaction energy brought by restructuring of the solvent in response to the electrostatic interaction is not properly taken into account in the polar part of the LIE expression.[73] We argue that the D term (e.g., difference between the electrostatic component and van der Waals) allows us to isolate and approximate the solvent and to some degree receptor cavity reorganization free energy and then correlate it to the γ coefficient.

3.3. Weight of Polar and Nonpolar Contributions in LIE-D Model. The prevalence (or weight) of polar and nonpolar contributions in the LIE-D model can be assessed in this way:

Substitution of eq 7 in eq 2 leads to

$$\Delta G_{bind} = \beta(\Delta\langle V_{l-s}^{el}\rangle + \Delta\langle V_{l-l}^{el}\rangle) + \alpha\Delta\langle V_{l-s}^{vdw}\rangle + f \times D + g \quad (8)$$

Substitution of eq 4 in eq 8 leads to

$$\Delta G_{bind} = \beta(\Delta\langle V_{l-s}^{el}\rangle + \Delta\langle V_{l-l}^{el}\rangle) + \alpha\Delta\langle V_{l-s}^{vdw}\rangle + f \times [\beta(\Delta\langle V_{l-s}^{el}\rangle + \Delta\langle V_{l-l}^{el}\rangle) - \alpha\Delta\langle V_{l-s}^{vdw}\rangle] + g \quad (9)$$

Grouping similar terms in eq 9:

$$\Delta G_{bind} = \beta(\Delta\langle V_{l-s}^{el}\rangle + \Delta\langle V_{l-l}^{el}\rangle) + f \times \beta(\Delta\langle V_{l-s}^{el}\rangle + \Delta\langle V_{l-l}^{el}\rangle) + \alpha\Delta\langle V_{l-s}^{vdw}\rangle - f \times \alpha\Delta\langle V_{l-s}^{vdw}\rangle + g \quad (10)$$

$$\Delta G_{bind} = (1 + f)\beta(\Delta\langle V_{l-s}^{el}\rangle + \Delta\langle V_{l-l}^{el}\rangle) + (1 - f)\alpha\Delta\langle V_{l-s}^{vdw}\rangle + g \quad (11)$$

$$\Delta G_{bind} = \beta'(\Delta\langle V_{l-s}^{el}\rangle + \Delta\langle V_{l-l}^{el}\rangle) + \alpha'\Delta\langle V_{l-s}^{vdw}\rangle + g \quad (12)$$

where $\beta' = (1 + f)\beta$ and $\alpha' = (1 + f)\alpha$

Taking $f = -0.95$, $\alpha = 0.18$, and for β the most common value for dipolar neutral ligands ($\beta = 0.43$), then we have $\beta' = 0.02$ and $\alpha' = 0.35$. Substitution of β', α', and g values in eq 12:

$$\Delta G_{bind} = 0.02 \times (\Delta\langle V_{l-s}^{el}\rangle + \Delta\langle V_{l-l}^{el}\rangle) + 0.35 \times \Delta\langle V_{l-s}^{vdw}\rangle - 2.06\ [\text{kcal/mol}] \quad (13)$$

The low weight of the polar (electrostatic) contribution to ΔG_{bind} ($\beta' = 0.02$) in eq 13 was expected, as for almost all complexes here analyzed, the values of the $\beta[\Delta\langle V_{l-s}^{el}\rangle + \Delta\langle V_{l-l}^{el}\rangle$ term were positive (Table S1), being unfavorable to the binding process. Most of the ligands used in our training set have more than one polar group in their structure (Figure S1). Then, they could establish favorable electrostatic interactions with protein binding site polar groups. But polar groups desolvation is energetically more expensive than for nonpolar ones.[74−76] So, if lost interactions between ligand polar groups and water are not compensated in the binding site (as it seems to happen for most of our ligands), their desolvation will oppose to the binding process. We think that is the reason for the low weight of polar contributions to ΔG_{bind}. This result also explains why considering intra-ligand electrostatic contributions does not affect the LIE-D model performance (Figure 2B). Although for some complexes $|\Delta\langle V_{l-l}^{el}\rangle| > 2$ kcal/mol (Table S1), using $\beta' = 0.02$ in eq 13 renders these contributions negligible.

On the other hand, the weight for nonpolar contributions in eq 13 is higher ($\alpha' = 0.35$). This value reflects the ligands nonpolar groups preference for protein binding site instead of bulk water. These results suggest that protein−ligand binding process is driven by the hydrophobic effect.

Now, given that

$$\Delta G_{bind} = \Delta H - T\Delta S \quad (14)$$

How are the enthalpy and entropy considered in LIE-D model? A recent study found that $\alpha\Delta\langle V_{l-s}^{vdw}\rangle$ term takes into account van der Waals interactions (predominantly dispersive part) and to a lesser degree size-dependent contributions to binding free energy such as desolvation penalties and translational/rotational entropies.[77] The three terms (and also vibrational entropy) contribute to ΔS in eq 14.[75] On the other hand, polar groups desolvation, hydrogen bonds breakage/formation (reflected in $\beta[\Delta\langle V_{l-s}^{el}\rangle + \langle V_{l-l}^{el}\rangle$ term), and van der Waals interaction energies contribute to ΔH in eq 14.[75] In this way, enthalpic and entropic contributions to ΔG_{bind} are implicit in the LIE formulation. In eq 13, $g = -2.06$ is the γ coefficient value when $D = \beta(\Delta\langle V_{l-s}^{el}\rangle + \Delta\langle V_{l-l}^{el}\rangle) - \alpha\Delta\langle V_{l-s}^{vdw}\rangle = 0$ (polar and nonpolar contributions are the same to ΔG_{bind}, see Figure 1). This (near zero) negative value could reflect a binding site where the (favorable) nonpolar groups desolvation energy is almost compensated by the (unfavorable) polar groups desolvation energy.

3.4. Assessment of LIE-D Model on Independent (external) Sets. To evaluate LIE-D model transferability, we calculated the binding free energies for three protein−ligand complexes test sets.[29,30,78] Different force fields were used for electrostatic and van der Waals interaction energies calculation. The first test set was CYP1A2, one of the most important cytochrome P450 (CYPT450) isoforms, bound to 13 differents ligands.[30] The second one was HIV-1 reverse transcriptase

DOI: 10.1021/acs.jcim.5b00012
J. Chem. Inf. Model. 2015, 55, 1867−1877

(RT) bound to 39 non-nucleoside reverse transcriptase inhibitors (NNRTI),[29] and the third one was the voltage-gated potassium ion channel Kv1.5 bound to five ligands.[78] For this validation stage, we wanted the LIE-D model to face different challenges: (i) protein−ligand systems outside the training data set, (ii) protein−ligand systems built through docking algorithms, and iii) the use of $\Delta\langle V^{el}_{l-s}\rangle$, $\Delta\langle V^{el}_{l-l}\rangle$, and $\Delta\langle V^{vdw}_{l-s}\rangle$ calculated by using different force fields.

A four steps scheme is needed for applying LIE-D model for ΔG_{bind} calculation:

(I) $\Delta\langle V^{el}_{l-s}\rangle$, $\Delta\langle V^{el}_{l-l}\rangle$, and $\Delta\langle V^{vdw}_{l-s}\rangle$ calculation from MD simulations.
(II) D parameter calculation from eq 4.
(III) Calculate the γ coefficient from D parameter using the linear model depicted in Figure 1.
(IV) ΔG_{bind} calculation using eq 2.

A summary of the results obtained for the three analyzed test sets is shown in Table 4.

CYP1A2−Inhibitor Complexes. The binding free energies for 13 CYP1A2−ligand complexes were recently reported by Vasanthanathan et al.[30] after combining docking algorithms with different LIE models and the GROMOS 45A4 force field.[79] These values were accurately predicted by us when using the LIE-D model (<|error|> = 0.92, $SD_{erro\ r}$= 0.58 kcal/mol, see Figure 3, Tables 4 and S2). This performance was

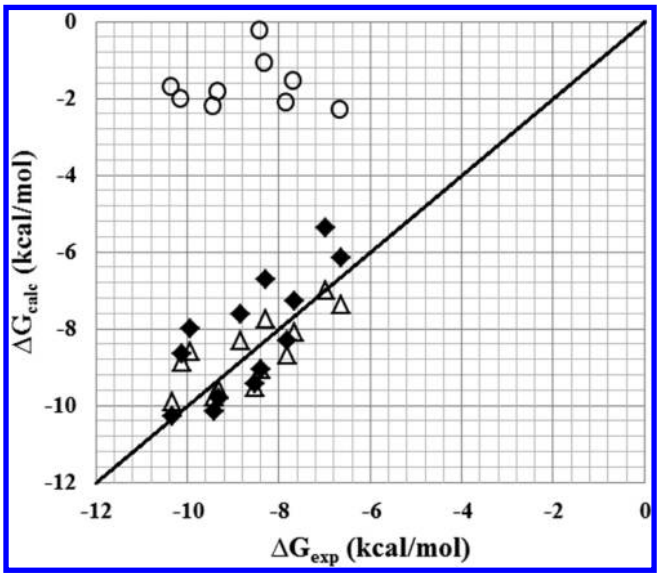


Figure 3. Scatter diagram of calculated (ΔG_{calc}) vs experimental (ΔG_{exp}) binding free energies for the 13 CYP1A2−ligand complexes from Vasanthanathan et al.[30] using LIE-S model (empty circles), LIE-D model (black solid diamonds), and model 3 of Vasanthanathan et al.[30] (empty triangles). Raw $\Delta\langle V^{el}_{l-s}\rangle$ and $\Delta\langle V^{vdw}_{l-s}\rangle$ values reported by Vasanthanathan et al.[30] as well as ΔG_{calc}, ΔG_{exp}, D parameter, and β and γ coefficients values for each complex are reported in Table S2.

similar to LIE model 3 from Vasanthanathan et al.[30] (<|error|> = 0.65, SD_{error} = 0.39 kcal/mol) that used β = 0, α = 0.21, and γ = −4.9 kcal/mol (Figure 3, Table 4 and Table S2). We also found that LIE-D clearly outperformed LIE-S model (<|error|> = 8.07, SD_{error} = 2.10 kcal/mol, see Figure 3, Table 4 and Table S2). These results look promising considering that most of previous studies aiming to predict CYTP450−ligand complexes binding free energies required LIE models previously parametrized using similar systems.[10,22,30]

The reasonable LIE-D model performance for these complexes can be (β' = 0.02, α' = 0.35, and γ = −2.06) are very similar to the previously and independently developed LIE model 3 coefficients from Vasanthanathan et al.[30] (see above). This can be related to CYP1A2 ligands hydrophobicity (Figure S1), so that electrostatic contribution to binding free energy is negligible (Table S1).

HIV-1 Reverse Transcriptase−Inhibitor Complexes. Carlsson et al.[29] previously combined docking, MD simulations, and the LIE method to predict the binding modes for 39 benzylpyridone-derivated non-nucleoside inhibitors (NNRTIs) in bound to HIV-1 reverse transcriptase (RT). The authors estimated their binding free energies using the the OPLS all-atom force field.[80] Here, we used the average interaction energies reported for the 39 complexes[29] to recalculate their binding free energies using the LIE-D model. Our model predicted the binding free energies with <|error|> = 1.34 and SD_{error} = 1.16 kcal/mol (Figure 4, Table 4 and Table S3).

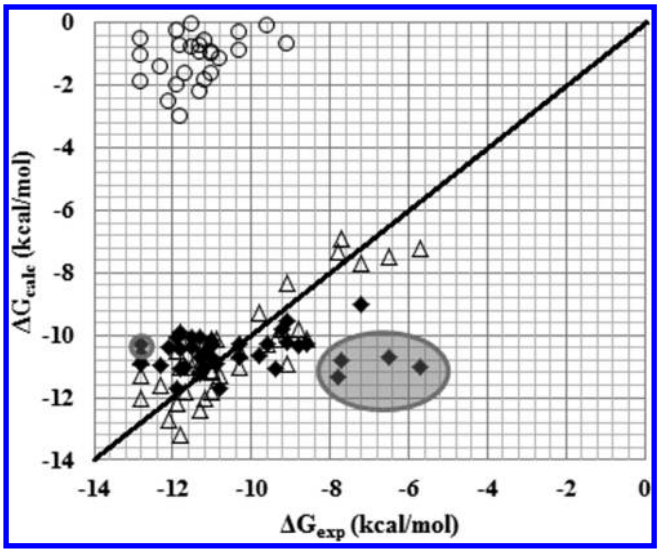


Figure 4. Scatter diagram of calculated (ΔG_{calc}) vs experimental (ΔG_{exp}) binding free energies for the 39 HIV-1 RT-NNRTI complexes from Carlsson et al.[29] using LIE-S model (empty circles), LIE-D model (black solid diamonds), and the optimized LIE standard model of Carlsson et al.[29] (empty triangles). The main six outliers for the LIE-D model are enclosed in gray circles. Raw $\Delta\langle V^{el}_{l-s}\rangle$ and $\Delta\langle V^{vdw}_{l-s}\rangle$ values reported by Carlsson et al.[29] as well as ΔG_{calc}, ΔG_{exp}, D parameter, and β and γ coefficients values for each complex are reported in Table S3.

Furthermore, LIE-D model accuracy was significantly improved (<|error|> = 0.94, SD_{error} = 0.60 kcal/mol) after six outliers (<|error|> 2.00 kcal/mol) exclusion (Figure 4, gray circles, Table 4 and Table S3). Similar figures were obtained for the LIE model from Carlsson et al.[29] (<|error|> = 0.83, SD_{error} = 0.61 kcal/mol) using the same 39 complexes as the training set (Figure 4, Table 4 and Table S3). The authors optimized their LIE model by introducing a freely optimized γ coefficient (γ = −10.20 kcal/mol), while β and α parameters were handled as in LIE-S model.[11,12] Remarkably, the LIE-D model predicted most of these complexes binding free energies (33 of 39) and clearly outperformed the LIE-S model[11,12] (<|error|> = 10.19, SD_{error} = 0.98 kcal/mol), even though none of these complexes were included in our training set (Figure 4, Table 4 and Table S3). For the six outliers, LIE-D overestimated the binding free energy (Figure 4, Table S4). It is possible for these complexes

DOI: 10.1021/acs.jcim.5b00012
J. Chem. Inf. Model. 2015, 55, 1867−1877

that the (unfavorable) polar contribution was too downscaled ($\beta' = 0.02$) and did not compensate the higher (favorable) nonpolar contribution. However, it is not clear if this is because of LIE-D model parameterization or force field dependence. The former is responsible for β, α, and γ coefficient values, while the latter influences $\Delta\langle V_{l-s}^{el}\rangle$ and $\Delta\langle V_{l-s}^{vdw}\rangle$ values. Nevertheless, we consider that LIE-D was successful for this test set as we could predict accurately the binding free energy for 85% of the complexes (Table 4 and Table S4).

Kv1.5 Channel–Inhibitor Complexes. Similar to previous test sets, the average interaction energies reported by Ander et al.[78] for seven ortho-disubstituted bisaryl compounds bound to the Kv1.5 potassium ion channel were here used for LIE-D binding free energy calculations. The authors used a three-step procedure consisting of homology modeling, automated docking, and LIE binding free energy calculations using the OPLS all-atom force field.[80] However, LIE-D failed to reproduce protein–ligand complexes binding free energies (<|error|> = 2.99 and SD_{error} = 1.58 kcal/mol, see Figure S3) for this set. To test whether these results were due to force field dependence or LIE-D model parameterization, we performed MD simulations on five complexes (structures provided by the author)[78] this time using the AMBER99SB force field.[54] The new $\Delta\langle V_{l-s}^{el}\rangle$, $\Delta\langle V_{l-l}^{el}\rangle$, and $\Delta\langle V_{l-s}^{vdw}\rangle$ values obtained improved LIE-D model ΔG_{bind} calculation accuracy (<|error|> = 1.03 and SD_{error} = 0.76 kcal/mol, see Figure 5, Table 4 and Table S4).

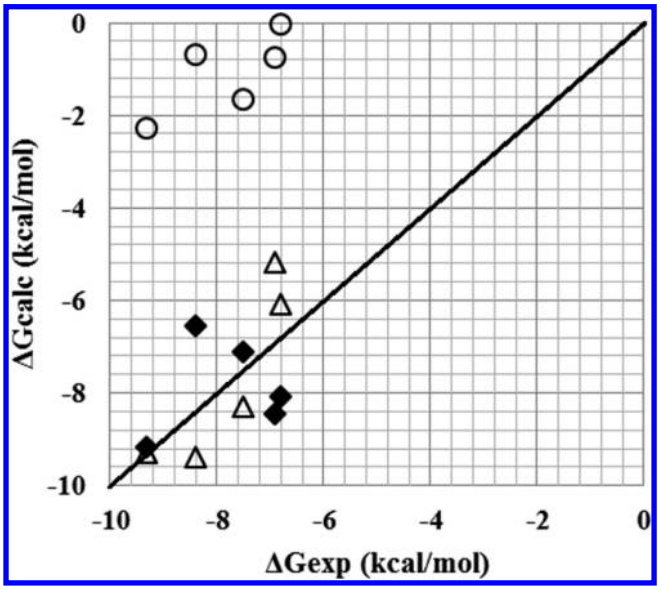


Figure 5. Scatter diagram of calculated (ΔG_{calc}) vs experimental (ΔG_{exp}) binding free energies for five Kv1.5–ligand complexes from Ander et al.[78] using the LIE-S model (empty circles), LIE-D model (black solid diamonds), and the optimized LIE standard model of Ander et al.[78] (empty triangles). ΔG_{calc}, ΔG_{exp}, D parameter, and β and γ coefficients values for each complex are reported in Table S4.

These results were very similar to the ones obtained for the Ander et al.[78] optimized LIE model (<|error|> = 0.84 and SD_{error} = 0.61 kcal/mol, see Figure 5, Table 4 and Table S4). Furthermore, LIE-D model predictions were also more accurate than LIE-S ones (<|error|> = 6.69 and SD_{error} = 0.74 kcal/mol, see Figure 5 and Table 4 and Table S4). The authors used Kv1.5 complexes as the training set to develop their optimized LIE model, comprising a freely optimized γ coefficient (γ = −3.47 kcal/mol), while β and α parameters as in the original LIE-S model. LIE-D predictions improvement by using the AMBER99SB force field[54] means that although we got good performance in our two previous test sets using different force fields (Table 4), for some notable cases as Kv1.5-ligand complexes, the LIE-D model reliability could be considerably better using the AMBER99SB force field.[54] Previously, Almlöf et al. suggested that the LIE method performance is force field independent. However, they only used CYTP450–ligand complexes as the study case,[22] and the general study of the reliability of LIE methods across a large number of systems and different force fields is a worthy future target for the field.

3.5. On the Validity of LIE-D Model. The results obtained for the three test sets highlight the LIE-D model usefulness for drug design/discovery projects; we were able to predict accurately the binding free energy for protein–ligand complexes outside our training set and obtained by docking methods. Using specific protein–ligand complexes as the training set to derive a LIE model that will be used later on similar systems could be more accurate (Table 4). But there is not always enough experimental data for specific protein–ligand complexes LIE model parameterization. Especially in those cases is where the LIE-D model would be more useful, as we have shown in this work, is its transferability to protein–ligand complexes outside our training set. Our results have been very close to the ones calculated by using the same test sets as training sets in the original reports[29,30,78] (Table 4). This transferability could be explained as the protein–ligand complexes we used in the training set comprised a broad range of enzymes and transport proteins bound to ligands belonging to different chemical scaffolds.

On the other hand, LIE-D is very straightforward to use as only $\Delta\langle V_{l-s}^{el}\rangle$, $\Delta\langle V_{l-l}^{el}\rangle$, and $\Delta\langle V_{l-s}^{vdw}\rangle$ values from MD simulations are needed for ΔG_{bind} calculation. It also takes into account the balance between electrostatic (polar) and van der Waals (nonpolar) contributions to binding free energy, reflected in the D parameter eq 4. This allows us to calculate system-dependent γ coefficient values instead of using a constant one. Then, the LIE-D model is flexible enough to take into account different interaction patterns even though the ligands share some common chemical scaffolds and are bound to the same protein receptor.

4. CONCLUDING REMARKS

We developed LIE-D, a novel LIE parameterization model that accurately predicts the γ coefficient based on the balance between polar and nonpolar contributions to binding free energy (D parameter) extracted from MD simulations. Leave-one-out assessment showed that LIE-D accurately reproduced our training data set experimental binding free energies. The model robustness was demonstrated by reproducing the binding free energies of two of the three protein–ligand sets outside the training data set, using the reported electrostatic and van der Waals interaction energies calculated with different force fields. Thus, LIE-D can be useful for lead optimization phases where computational methodologies more accurate than scoring functions will be needed to predict absolute binding free energies of protein–ligand complexes. However, to conclude which force field is the best choice for free energy calculations with the LIE-D model, further studies will be needed.

DOI: 10.1021/acs.jcim.5b00012
J. Chem. Inf. Model. 2015, 55, 1867–1877

■ ASSOCIATED CONTENT

Supporting Information

Structure and nomenclature of the ligands employed in this work, thermodynamic cycle scheme, values of β and γ coefficients, electrostatics and van der Waals energies, and value of the D parameter for the training and test set complexes. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.5b00012.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: valiente@fbio.uh.cu.

*E-mail: wmiranda@fbio.uh.cu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

P.A.V. acknowledges Cuban Ministry of High Education (MES), IFS Grant F/5198-1 for financial support. The authors thank Dr. Tirso Pons Hernández, Dr. Rossana García Fernández, and Dr. Igor Zdravkovic for critical revision of the manuscript. This work was supported by grants from the Natural Sciences and Engineering Resarch Council (Canada) to S.Y.N. (RGPIN-315019). S.Y.N. is an Alberta Innovates Technology Futures (AITF) New Faculty, Canadian Institute for Health Research New Investigator, and an Alberta Innovates Health Solutions (AIHS) Scholar. The research activities in Calgary were also supported by the Emerging Leaders in the Americas Program (ELAP) and the AITF Futures Strategic Chair in (Bio)Molecular Simulations to S.Y.N. and W.E.M. Some of the computations were performed on the West-Grid/Compute Canada facilities and the University of Calgary TNK cluster acquired with direct support by the Canada Foundation for Innovation.

■ ABBREVIATIONS

LIE, linear interaction energy; FEP, free energy perturbation; TI, thermodynamic integration; MD, molecular dynamics; MC, Monte Carlo; WNDR, weighted nonpolar desolvation ratio; LR, linear response; K_i, inhibition constant; K_a, affinity constant; K_d, dissociation constant; ΔG, binding free energy difference; <|error|>, mean absolute error; α, scaling factor for nonbonded van der Waals interaction energies; β, scaling factor for electrostatic interaction energies; γ, correction coefficient to reproduce the experimental binding free energies; CYPA12, an isoform of cytochrome P450; HIV-1 RT, reverse transcriptase HIV-1; Kv1.5, voltage-gated potassium ion channel

■ REFERENCES

(1) Kollman, P. A. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.

(2) Brandsdal, B. O.; Osterberg, F.; Almlof, M.; Feierberg, I.; Luzhkov, V. B.; Aqvist, J. Free Energy Calculations and Ligand Binding. *Adv. Protein Chem.* **2003**, *66*, 123−158.

(3) Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 2234−2246.

(4) Zhao, C.; Caplan, D. A.; Noskov, S. Y. Evaluations of the Absolute and Relative Free Energies for Antidepressant Binding to the Aminoacid Membrane Transporter LeuT with Free Energy Simulations. *J. Chem. Theory Comput.* **2010**, *6*, 1900−1914.

(5) Aqvist, J.; Medina, C.; Samuelsson, J. E. A New Method for Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Eng., Des. Sel.* **1994**, *7*, 385−391.

(6) Hulten, J.; Bonham, N. M.; Nillroth, U.; Hansson, T.; Zuccarello, G.; Bouzide, A.; Aqvist, J.; Classon, B.; Danielson, U. H.; Karlen, A.; Kvarnstrom, I.; Samuelsson, B.; Hallberg, A. Cyclic HIV-1 Protease Inhibitors Derived From Mannitol: Synthesis, Inhibitory Potencies, and Computational Predictions of Binding Affinities. *J. Med. Chem.* **1997**, *40*, 885−897.

(7) Homeyer, N.; Gohlke, H. FEW: A Workflow Tool for Free Energy Calculations of Ligand Binding. *J. Comput. Chem.* **2013**, *34*, 965−973.

(8) Marelius, J.; Kolmodin, K.; Feierberg, I.; Aqvist, J. Q: An MD program for Free Energy Calculations and Empirical Valence Bond Simulations in Biomolecular Systems. *J. Mol. Graphics Modell.* **1998**, *16*, 213−225.

(9) Wang, W.; Wang, J.; Kollman, A. P. What Determines the van der Waals Coefficient β in the LIE Method to Estimate Binding Free Energies Using Molecular Dynamics Simulations? *Proteins: Struct., Funct., Genet.* **1999**, *34*, 395−402.

(10) Paulsen, M. D.; Ornstein, R. L. Binding Free Energy Calculations for P450cam-Substrate Complexes. *Protein Eng., Des. Sel.* **1996**, *9*, 567−571.

(11) Marelius, J.; Hansson, T.; Aqvist, J. Calculation of Ligand Binding Free Energies from Molecular Dynamics Simulations. *Int. J. Quantum Chem.* **1998**, *69*, 77−88.

(12) Hansson, T.; Marelius, J.; Aqvist, J. Ligand Binding Affinity Prediction by Linear Interaction Energy Methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27−35.

(13) Aqvist, J.; Marelius, J. The Linear Interaction Energy Method for Predicting Ligand Binding Free Energies. *Comb. Chem. High Throughput Screening* **2001**, *4*, 613−626.

(14) Aqvist, J.; Luzhkov, V. B.; Brandsal, B. O. Ligand Binding Affinities from MD Simulations. *Acc. Chem. Res.* **2002**, *35*, 358−365.

(15) van Gunsteren, W. F.; Berendsen, H. J. C. *Groningen Molecular Simulation (GROMOS) Library Manual, Biomos*; Nijenborgh: Groningen, The Netherlands, 1987.

(16) Hansson, T.; Aqvist, J. Estimation of Binding Free Energies for HIV Proteinase Inhibitors by Molecular Dynamics Simulations. *Protein Eng., Des. Sel.* **1995**, *8*, 1137−1144.

(17) Aqvist, J.; Mowbray, S. L. Sugar Recognition by a Glucose/Galactose Receptor: Evaluation of Binding Energetics from Molecular Dynamics Simulations. *J. Biol. Chem.* **1995**, *270*, 9978−9981.

(18) Aqvist, J. Calculation of Absolute Binding Free Energies for Charged Ligands and Effects of Long-Range Electrostatic Interactions. *J. Comput. Chem.* **1996**, *17*, 1587−1597.

(19) Aqvist, J.; Hansson, T. On the Validity of Electrostatic Linear Response in Polar Solvents. *J. Phys. Chem.* **1996**, *100*, 9512−9521.

(20) Ersmark, K.; Feierberg, I.; Bjelic, S.; Hulten, J.; Samuelsson, B.; Aqvist, J.; Hallberg, A. C2-Symmetric Inhibitors of *Plasmodium falciparum* Plasmepsin II: Synthesis and Theoretical Predictions. *Bioorg. Med. Chem.* **2003**, *11*, 3723−3733.

(21) Ersmark, K.; Feierberg, I.; Bjelic, S.; Hamelink, E.; Hackett, F.; Blackman, M. J.; Hulten, J.; Samuelsson, B.; Aqvist, J.; Hallberg, A. Potent Inhibitors of the *Plasmodium falciparum* Enzymes Plasmepsin I and II Devoid of Cathepsin D Inhibitory Activity. *J. Med. Chem.* **2004**, *47*, 110−122.

(22) Almlof, M.; Brandsdal, B. O.; Aqvist, J. Binding Affinity Prediction with Different Force Fields: Examination of the Linear Interaction Energy Method. *J. Comput. Chem.* **2004**, *25*, 1242−1254.

(23) Marelius, J.; Graffner-Nordberg, M.; Hansson, T.; Hallberg, A.; Aqvist, J. Computation of Affinity and Selectivity: Binding of 2,4-diaminopteridine and 2,4-diaminoquinazoline Inhibitors to Dihydrofolate Reductases. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 119−131.

(24) Ersmark, K.; Samuelsson, B.; Hallberg, A. Plasmepsins as Potential Targets for New Antimalarial Therapy. *Med. Res. Rev.* **2006**, *26*, 626−666.

(25) Ljungberg, K. B.; Marelius, J.; Musil, D.; Svensson, P.; Norden, B.; Aqvist, J. Computational Modelling of Inhibitor Binding to Human Thrombin. *Eur. J. Pharm. Sci.* **2001**, *12*, 441−446.
(26) Valiente, P. A.; Gil, 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.
(27) Almlof, M.; Carlsson, J.; Aqvist, J. Improving the Accuracy of the Linear Interaction Energy Method for Solvation Free Energies. *J. Chem. Theory Comput.* **2007**, *3*, 2162−2175.
(28) Wold, S. Validation of QSAR's. *Quant. Struct.-Act. Relat.* **1991**, *10*, 191−193.
(29) Carlsson, J.; Boukharta, L.; Aqvist, J. Combining Docking, Molecular Dynamics and the Linear Interaction Energy Method to Predict Binding Modes and Affinities for Non-Nucleoside Inhibitors to HIV-1 Reverse Transcriptase. *J. Med. Chem.* **2008**, *51*, 2648−2656.
(30) Vasanthanathan, P.; Olsen, L.; Jorgensen, F. S.; Vermeulen, N. P. E.; Oostenbrink, C. Computational Prediction of Binding Affinity for CYP1A2-Ligand Complexes Using Empirical Free Energy Calculations. *Drug Metab. Dispos.* **2010**, *38*, 1347−1354.
(31) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(32) Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. The Maximal Affinity of Ligands. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9997−10002.
(33) Zanotti, G.; Malpeli, G.; Berni, R. The Interaction of N-ethyl Retinamide with Plasma Retinol-Binding Protein (RBP) and the Crystal Structure of the Retinoid-RBP Complex at 1.9-Å Resolution. *J. Biol. Chem.* **1993**, *268*, 24873−24879.
(34) Holt, D. A.; Luengo, J. I.; Yamashita, D. S.; Oh, H. J.; Konialian, A. L.; Yen, H. K.; Rozamus, L. W.; Brandt, M.; Bossard, M. J. Design, Synthesis, and Kinetic Evaluation of High-Affinity FKBP Ligands and the X-Ray Crystal-Structures of their Complexes with FKBP12. *J. Am. Chem. Soc.* **1993**, *115*, 9925−9938.
(35) Eriksson, A. E.; Baase, W. A.; Wozniak, J. A.; Matthews, B. W. A Cavity-Containing Mutant of T4 Lysozyme is Stabilized by Buried Benzene. *Nature* **1992**, *355*, 371−373.
(36) Morton, A.; Baase, W. A.; Matthews, B. W. Energetic Origins of Specificity of Ligand Binding in an Interior Nonpolar Cavity of T4 Lysozyme. *Biochemistry* **1995**, *34*, 8564−8575.
(37) Ali, A.; Reddy, G. S. K. K.; Nalam, M. N. L.; Anjum, S. G.; Cao, H.; Schiffer, C.; Rana, T. M. Structure-Based Design, Synthesis, and Structure-Activity Relationship Studies of HIV-1 Protease Inhibitors Incorporating Phenyloxazolidinones. *J. Med. Chem.* **2010**, *53* (21), 7699−7708.
(38) Altman, M. D.; Ali, A.; Kumar Reddy, G. S. K.; Nalam, M. N. L.; Anjum, S. G.; Cao, H.; Chellappan, S.; Kairys, V.; Fernandes, M. X.; Gilson, M. K.; Schiffer, C.; Rana, T. M.; Tidor, B. HIV-1 Protease Inhibitors from Inverse Design in the Substrate Envelope Exhibit Subnanomolar Binding to Drug-Resistant Variants. *J. Am. Chem. Soc.* **2008**, *130* (19), 6099−6113.
(39) Nalam, M. N. L.; Ali, A.; Altman, M. D.; Reddy, G. S. K. K.; Chellappan, S.; Kairys, V.; Ozen, A.; Cao, H.; Gilson, M. K.; Tidor, B.; Rana, T. M.; Schiffer, C. Evaluating the Substrate-Envelope Hypothesis: Structural Analysis of Novel HIV-1 Protease Inhibitors Designed to be Robust Against Drug Resistance. *J. Virol.* **2010**, *84* (10), 5368−5378.
(40) Ali, A.; Reddy, G. S. K. K.; Cao, H.; Anjum, S. G.; Nalam, M. N.; Schiffer, C.; Rana, T. M. Discovery of HIV-1 Protease Inhibitors with Picomolar Affinities Incorporating N-Aryl-oxazolidinone-5-carboxamides as Novel P2 Ligands. *J. Med. Chem.* **2006**, *49*, 7342−7356.
(41) Prade, L.; Jones, A. F.; Boss, C.; Richard-Bildstein, S.; Meyer, S.; Binkert, C.; Bur, D. X-RAY Structure of Plasmepsin II Complexed with a Potent Achiral Inhibitor. *J. Biol. Chem.* **2005**, *280*, 23837−23843.
(42) Asojo, O. A.; Gulnik, S. V.; Afonina, E.; Yu, B.; Ellman, J. A.; Haque, T. S.; Silva, A. M. Novel Uncomplexed and Complexed Structures of Plasmepsin II, an Aspartic Protease from *Plasmodium falciparum*. *J. Mol. Biol.* **2003**, *327*, 173−181.
(43) Asojo, O. A.; Afonina, E.; Gulnik, S. V.; Yu, B.; Erickson, J. W.; Randad, R.; Medjahed, D.; Silva, A. M. Structures of Ser205 Mutant Plasmepsin II from *Plasmodium Falciparum* at 1.8A in Complex with the Inhibitors RS367 and RS370. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58*, 2001−2008.
(44) Clemente, J. C.; Govindasamy, L.; Madabushi, A.; Fisher, S. Z.; Moose, R. E.; Yowell, C. A.; Hidaka, K.; Kimura, T.; Hayashi, Y.; Kiso, Y.; Agbandje-McKenna, M.; Dame, J. B.; Dunn, B. M.; McKenna, R. Structure of the Aspartic Protease Plasmepsin 4 from the Malarial Parasite *Plasmodium malariae* Bound to an Allophenylnorstatine-Based Inhibitor. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62*, 246−252.
(45) Romesberg, F. E.; Spiller, B.; Schultz, P. G.; Stevens, R. C. Immunological Origins of Binding and Catalysis in a Diels-Alderase Antibody. *Science* **1998**, *279*, 1929−1933.
(46) Alkema, W. B. L.; Hensgens, C. M. H.; Snijder, H. J.; Keizer, E.; Dijkstra, B. W.; Janssen, D. B. Structural and Kinetic Studies on Ligand Binding in Wild-Type and Active-Site Mutants of Penicillin Acylase. *Protein Eng., Des. Sel.* **2004**, *17*, 473−480.
(47) Sacchettini, J. C.; Gordon, J. I.; Banaszak, L. J. Crystal Structure of Rat Intestinal Fatty-Acid-Binding Protein. Refinement and Analysis of the *Escherichia coli*-Derived Protein with Bound Palmitate. *J. Mol. Biol.* **1989**, *208*, 327−339.
(48) Lowe, J. B.; Sacchettini, J. C.; Laposata, M.; McQuillan, J. J.; Gordon, J. I. Expression of Rat Intestinal Fatty Acid-Binding Protein in *Escherichia coli*. Purification and Comparison of Ligand Binding Characteristics with that of *Escherichia coli*-Derived Rat Liver Fatty Acid-Binding Protein. *J. Biol. Chem.* **1987**, *262*, 5931−5937.
(49) Schwans, J.; Ruben, E.; Sunden, F.; Gonzalez, A.; Tsai, Y.; Herschlag, D., Crystal Structure of Ketosteroid Isomerase D99N from *Pseudomonas Testosteroni* (tKSI) with 4-Androstene-3,17-dione Bound. To be published.
(50) Hawkinson, D. C.; Eames, T. C. M.; Pollack, R. M. Energetics of 3-Oxo-Δ5-steroid Isomerase: Source of the Catalytic Power of the Enzyme. *Biochemistry* **1991**, *30*, 10849−10858.
(51) Yamane, J.; Yao, M.; Zhou, Y.; Hiramatsu, Y.; Fujiwara, K.; Yamaguchi, T.; Yamaguchi, H.; Togame, H.; Tsujishita, H.; Takemoto, H.; Tanaka, I. In-Crystal Affinity Ranking of Fragment Hit Compounds Reveals a Relationship with their Inhibitory Activities. *J. Appl. Crystallogr.* **2011**, *44*, 798−804.
(52) Leiros, H. K. S.; Brandsdal, B. O.; Andersen, O. A.; Os, V.; Leiros, I.; Helland, R.; Otlewski, J.; Willasen, N. P.; SMALÅS, A. O. Trypsin Specificity as Elucidated by LIE Calculations, X-ray Structures, and Association Constant Measurements. *Protein Sci.* **2004**, *13*, 1056−1070.
(53) van der Spoel, D.; Lindahl, E.; Hess, B.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; van Drunenand, R.; Berendsen, H. J. C. *Gromacs User Manual*, version 4.5.6; Royal Institute of Technology and Uppsala University: Sweden, 2010.
(54) 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., Genet.* **2006**, *65*, 712−725.
(55) Li, H.; Robertson, A. D.; Jensen, J. H. Very Fast Empirical Prediction and Rationalization of Protein pKa Values. *Proteins: Struct., Funct., Genet.* **2005**, *61*, 704−721.
(56) Wang, J. M.; 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.
(57) 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.
(58) 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.
(59) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Klein, M.; Impey, R. W. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

DOI: 10.1021/acs.jcim.5b00012
J. Chem. Inf. Model. 2015, 55, 1867−1877

(60) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(61) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952−962.
(62) Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159*, 98−103.
(63) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(64) Berendsen, H. J. C.; Postma, J. P. M.; DiNola, A.; Haak, J. R.; van Gunsteren, W. F. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(65) van Gunsteren, W. F.; Berendsen, H. J. C. A Leap-Frog Algorithm for Stochastic Dynamics. *Mol. Simul.* **1988**, *1*, 173−185.
(66) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(67) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An W log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10093.
(68) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8592.
(69) Comuzzi, C.; Polese, P.; Melchior, A.; Portanova, R.; Tolazzi, M. "SOLVERSTAT: A New Utility for Multipurpose Analysis. An Application to the Investigation of Dioxygenated Co(II) Complex Formation in Dimethylsulfoxide Solution.". *Talanta* **2003**, *59*, 67−80.
(70) Linder, M.; Ranganathan, A.; Brinck, T. Adapted Linear Interaction Energy: A Structure-Based LIE Parameterization for Fast Prediction of Protein-Ligand Affinities. *J. Chem. Theory Comput.* **2013**, *9*, 1230−1239.
(71) Carlson, H. A.; Jorgensen, W. L. An Extended Linear Response Method for Determining Free Energies of Hydration. *J. Phys. Chem.* **1995**, *99*, 10667−10673.
(72) Lamb, M. L.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of the Binding Affinities of FKBP12 Inhibitors Using a Linear Response Method. *Bioorg. Med. Chem.* **1999**, *7*, 851−860.
(73) Westergren, J.; Lindfors, L.; Hoglund, T.; Luder, K.; Nordholm, S.; Kjellander, R. In Silico Prediction of Drug Solubility: 1. Free Energy of Hydration. *J. Phys. Chem. B* **2007**, *111*, 1872−1882.
(74) Mobley, D. L.; Dill, D. A. Binding of Small-Molecule Ligands to Proteins: "What You See" Is Not Always "What You Get". *Structure* **2009**, *17*, 489−498.
(75) Kawasaki, Y.; Freire, E. Finding a Better Path to Drug Selectivity. *Drug Discovery Today* **2011**, *16*, 985−990.
(76) Barratt, E.; Bronowska, A.; Vondrasek, J.; Cerny, J.; Bingham, R.; Phillips, S.; Homans, S. W. Thermodynamic Penalty Arising from Burial of a Ligand Polar Group within a Hydrophobic Pocket of a Protein Receptor. *J. Mol. Biol.* **2006**, *362*, 994−1003.
(77) Carlsson, J.; Aqvist, J. Calculations of Solute and Solvent Entropies from Molecular Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5385−5395.
(78) Ander, M.; Luzhkov, V. B.; Aqvist, J. Ligand Binding to the Voltage-Gated Kv1.5 Potassium Channel in the Open-State: Docking and Computer Simulations of a Homology Model. *Biophys. J.* **2008**, *94*, 820−831.
(79) Lins, R. D.; Hunenberger, P. A New GROMOS Force Field for Hexopyranose-Based Carbohydrates. *J. Comput. Chem.* **2005**, *26*, 1400−1412.
(80) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS all-atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

DOI: 10.1021/acs.jcim.5b00012
J. Chem. Inf. Model. 2015, 55, 1867−1877