JCTC Journal of Chemical Theory and Computation

Article

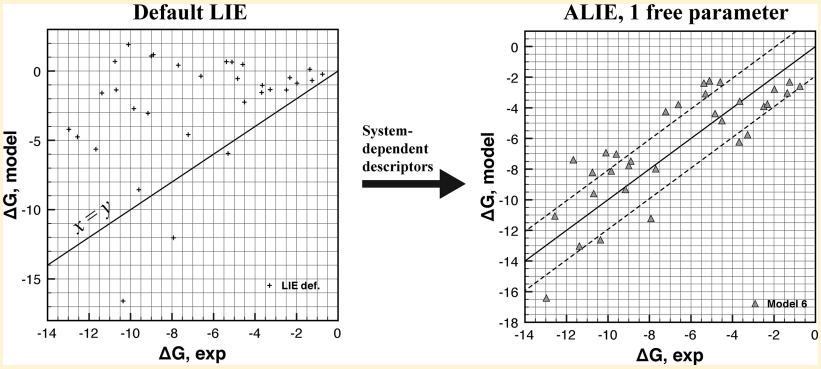
pubs.acs.org/JCTC

"Adapted Linear Interaction Energy": A Structure-Based LIE Parametrization for Fast Prediction of Protein−Ligand Affinities

Mats Linder, Anirudh Ranganathan, and Tore Brinck*

Applied Physical Chemistry, KTH Royal Institute of Technology, Teknikringen 30, S-100 44 Stockholm, Sweden

Supporting Information



ABSTRACT: We present a structure-based parametrization of the Linear Interaction Energy (LIE) method and show that it allows for the prediction of absolute protein−ligand binding energies. We call the new model "Adapted" LIE (ALIE) because the α and β coefficients are defined by system-dependent descriptors and do therefore not require any empirical γ term. The best formulation attains a mean average deviation of 1.8 kcal/mol for a diverse test set and depends on only one fitted parameter. It is robust with respect to additional fitting and cross-validation. We compare this new approach with standard LIE by Åqvist and co-workers and the LIE + γSASA model (initially suggested by Jorgensen and co-workers) against in-house and external data sets and discuss their applicabilities.

1. INTRODUCTION

In silico prediction of protein−ligand binding energies is a central problem in computational biophysics, with important implications for, e.g., drug discovery and enzyme design.[1] Many methods have been proposed for this notoriously difficult task. The major efforts have been based on ensemble analysis of Molecular Dynamics/Monte Carlo (MD/MC) trajectories, like the Linear Interaction Energy method (LIE),[2−4] Molecular Mechanics-Poisson−Boltzmann (Generalized Born)-Surface Area (MM-PB(GB)SA),[5,6] or the more rigorous Free Energy Perturbation/Thermodynamic Integration (FEP/TI) methods.[4,7] Knowledge-based approaches such as molecular docking[8] or QSAR[9] have enjoyed increasing popularity mainly due to their speed, but docking algorithms rely heavily on fitting and can fail for systems falling outside their respective training sets.

The LIE method, first proposed by Åqvist et al. in 1994,[2] has attained popularity due to its reasonably wide applicability, computational ease, and relative accuracy, making it well suited for desktop calculations. It is based on a linear response between free energy and the relative changes in the solute's interactions when it moves from one environment to another, such as bulk water and protein environments, the expression for which is given in eq 1

$$\Delta G_{\mathrm{bind}} = \alpha\Delta\langle V_{\mathrm{l-s}}^{\mathrm{vdW}}\rangle + \beta\Delta\langle V_{\mathrm{l-s}}^{\mathrm{el}}\rangle + \gamma \qquad (1)$$

$\Delta\langle V_{\mathrm{l-s}}^{\mathrm{vdW}}\rangle$ and $\Delta\langle V_{\mathrm{l-s}}^{\mathrm{el}}\rangle$ are the differences in van der Waals (vdW) and electrostatic interactions between the ligand−receptor and free ligand systems, taken as ensemble averages. In the original formulation of LIE, the additive γ term was set to zero. The β coefficient was initially set to 0.5 from the standard linear response approximation (LRA).[2] It was later given values between 0.33 and 0.50, depending on the chemical nature of the ligand, to better fit experimental results.[10] There is no analytic way of deriving α; it was originally fitted to 0.161 and later changed to 0.18,[10] but other workers have obtained different optimal values.[11] In practice, a small α often leads to an underestimation of ΔG_{bind}, which is offset by taking γ as a constant, enzyme-dependent correction,[12] found by fitting to experimental data. This however means that for a system without available experimental data, γ is unknown, and hence only relative ΔG_{bind} can be predicted with acceptable accuracy. Setting $\gamma = 0$ only works in a limited number of cases;[10,12] the binding strength typically is underestimated without a negative γ term.

Instead of using a system-dependent γ, one could let α vary to incorporate the nonpolar energy contributions. We recalculated ΔG_{bind} with the simple assumption that $\alpha = \beta = 0.5$

Received: September 9, 2012
Published: December 18, 2012

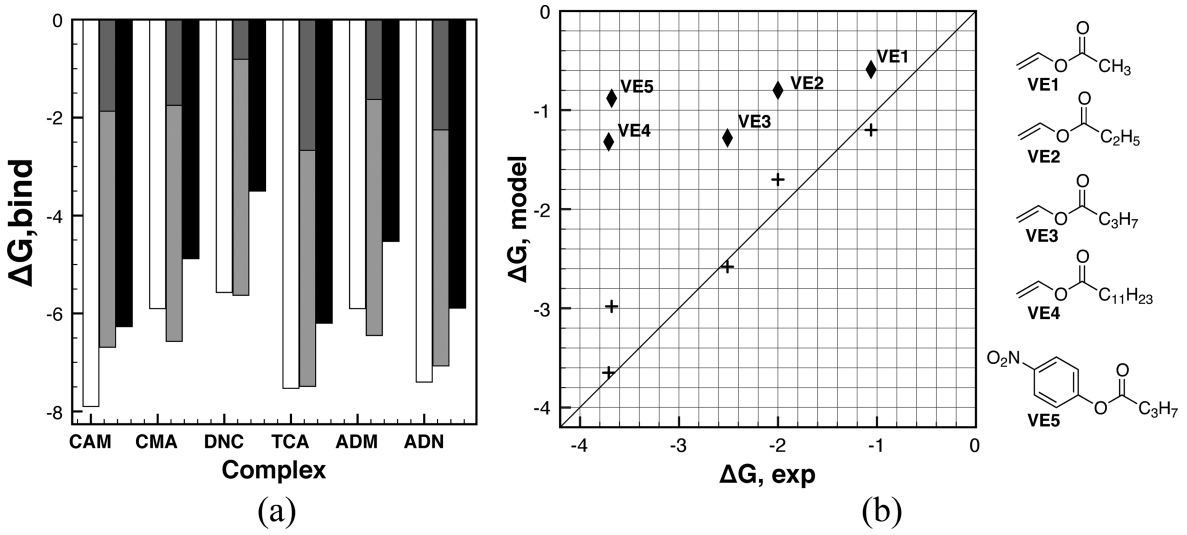


Figure 1. (a) Comparison of the standard LIE parametrization (dark gray), plus a fitted constant γ term (light gray) and the LIE "half-and-half" parametrization (black) of a series of camphor derivates in cytochrome p450.[12] The white bar shows the experimental values. (b) Binding energies of a series of ligands and lipase from *R. meihei* (PDB entry 4TGL[13]). The diagram shows calculated ΔG_{bind} vs experimental[14] values of LIE with $\gamma = 0$ (diamonds) and the effect of adding the surface area term (ΔG_{SA}) from the MM-PBSA calculation to the LIE energy ($\gamma = \Delta G_{SA}^{PB}$) (crosses). The simulations were performed using AMBER10[15] and the FF99SB[16] force field, producing 2 ns trajectories.

for a previously studied system with cytochrome p450 and mostly nonpolar ligands,[12] which was found to have a large and negative value of γ. Because nonpolar interactions dominate, this parametrization performs almost as well as the best one proposed in ref 10, as seen in Figure 1a. The results are similar to those found by Paulsen and Ornstein for the same system[11] and indicate that with $\alpha = 0.18$, the role of γ becomes, in practice, to compensate for the downscaled nonpolar interactions. Indeed, Åqvist and colleagues have stated that γ can be described as a measure of "hydrophobicity" of the enzyme.[3,12]

With access to experimental data, γ has been determined to be values between 0.0 and −7.0 kcal/mol in a range of studies.[10,12,17,18] However, the constant γ approximation fails for data series where ligand sizes vary considerably since a small α leads to an underestimation of size effects. For example, for a series of esters modeled in a lipase from *Rhizomucor miehei*, we found that γ must also be ligand-dependent for acceptable accuracy (0b). We tried representing γ with the ensemble-averaged surface area-dependent term ($\gamma = f(\mathrm{SASA})$), originally proposed by Carlson and Jorgensen[19]

$$\gamma = \Delta G_{SA}^{PB} = \gamma'(\mathrm{SASA}_{compl} - \mathrm{SASA}_{prot} - \mathrm{SASA}_{lig}) + c \quad (2)$$

where SASA = "solvent accessible surface area." Using eq 2 led to better correlated binding energies (Figure 1b). Importantly, the severe deviation with increasing ligand size was amended. This problem could clearly not have been solved with a constant value of γ. Jorgensen and co-workers have used the $\gamma = f(\mathrm{SASA})$ augmentation in a number of protein−ligand systems.[20,21] The model was later extended to the so-called LIE-SGB models,[22] in which an ensemble average of the area differences was used along with the use of an explicit solvent or a continuum. Similar applications of surface area terms have been used in a number of other works;[23,24] we will refer to them as LIE-γSASA models.

Albeit more flexible than the original LIE method, the LIE-γSASA approach has still been applied using considerable reparametrization aided by experimental data. In the everyday work of a computational chemist in applications like drug design or discovery, however, experimental values are generally unavailable. The aim of this work was therefore to explore if there are some system-dependent properties that can be exploited in order to develop a more general yet variable parametrization of eq 1, with the goal of being able to perform qualified estimates of binding energies without any *a priori* knowledge. Additional data for system-dependent parameters should be readily available from crystal/NMR structures and/or from carrying out MD simulations. A variable, yet systematic determination of LIE coefficients would render the method more powerful as well as improve understanding of the factors contributing to the more elusive aspects of the protein−ligand binding energy.

An extension of LIE to include more separable properties resembles the ideas behind qualitative structure−activity relationship (QSAR)[9] methods as well as scoring functions in molecular docking.[25] Jorgensen and colleagues have developed an expanded linear response (ELR) approach,[26−28] in which ΔG_{bind} is computed according to eq 3.

$$\Delta G_{bind} = \sum_i c_i E_i + c_0 \quad (3)$$

E_i can be both a standard interaction energy as well as some qualitative property, e.g., the number of hydrogen bonds formed. Although this method has been applied in numerous systems,[29−32] it is still dependent on extensive fitting and thus requires experimental data.[33] Similariy, Åqvist and co-workers recently studied how the β coefficient could be varied with respect to discrete, chemical properties to improve calculations of solvation free energies.[34] In this work, we seek neither to introduce additional energy terms in the ΔG_{bind} expression nor to base coefficient determination on some discrete property. We rather investigate if the LIE coefficients can be varied by system-derived descriptors.[35] The results obtained herein indicate that a simple expression, dependent on only one fitted parameter, is surprisingly accurate and applicable to a broad range of systems.

2. METHODS

2.1. Protein−Ligand Systems. We have observed, both from the literature[36,37] and by experience, that LRA approaches have problems treating the electrostatics of ionic or zwitterionic ligands, so we have limited the studied systems to proteins with uncharged ligands and no cofactors. We searched the literature for such systems that had (i) determined binding constants (as K_S or K_M) and (ii) protein structures deposited in the PDB.[38]

Since the goal was to develop a general augmentation to the LIE method, we used a training set covering ΔG_{bind} values from −1 to −13 kcal/mol, which is a standard range of noncovalent binding energies.[39] Both enzymes and binding proteins have been considered.[40] In the case of enzymes, we have approximated the binding constant K_S with the Michaelis constant K_M, since it can be assumed that the binding and release steps are faster than the reaction rate except in the diffusion limit. There is always a correlation between ligand size and ΔG^{exp}_{bind}, and a too strong one could potentially lead to biased results. The selected data set has a linear correlation coefficient (R^2) of 0.27 (Figure S1), from which we conclude that the free energy-molecular weight relationship does not override other possible correlations.

A comprehensive summary of all proteins and their associated ligands is given in Table 1. Structures of all ligands are given in Figure S3.

Table 1. Summary of Systems Included in the Data Set and Their Binding Energy Ranges

protein[a]	PDB code	ligands[b]	$-\Delta G^{exp}_{bind}$[c] [kcal/mol]
R. meihei lipase (RML)	4TGL[13]	**VE1−VE5**	1.3−3.7[14]
T. lanuginosa lipase (TLL)	1GT6[41]	**VE1−VE4**	0.8−3.4[14]
C. antarctica lipase B (CALB)	1LBT[42]	**VE5**	4.5[43]
bacteriophage T4 lysozyme S99A (LYS)	1L83[44]	**BN1−BN4**	4.5−6.6[25,45]
retinol binding protein (RBP)	1RBP[46]	**ret**	9.2[25,46]
purine nucleoside phosphorylase (PLP)	1ULB[47]	**gun**	7.2[25,48]
P. testisteroni ketosteroid isomerase (KSI)	1QJG[49]	**ans**	4.8[50]
HIV-1 protease (HP)	1HVR[51]	**HP1−HP3**	11.4−13.0[25,52−54]
S. typhimurium glucose binding protein (GlBP)	3GBP[55]	**glc**, **meg**	9.6, 5.3[56]
E. coli galactose binding protein (GaBP)	2GBP[57]	**gal**	7.9[25,58]
S. rubiginosus xylolise omerase (XIS)	2XIS[59]	**xyl**	10.4[25,48]
C. parasitica endothiapepsin (EP)	4ER1[60]	**EP1−EP3**	9.8−11.7[2,10]
FK 506 binding protein (FKBP12)	1FKG[61]	**FK1−FK5**	7.7−10.8[21]
B. taurus trypsin (TRP)[d]	3PTB[62]	**BA1−BA4**	3.8−6.5[25,63]
fatty acid binding protein (FBP)[d]	2IFB[64]	**fat**	7.4[25,65]
HIV-1 protease (HP)[d]	1HVR[51]	**HP4**	8.4−13.0[25,66]

[a]Organism is *Homo sapiens* if not stated otherwise. [b]Ligand structures are given in Figure S2. [c]Binding energies vs ligand size for each protein are plotted in Figure S1. [d]Ionic ligands, not included in the model fitting (see below).

2.2. MD Simulations. All MD simulations were carried out for 2−4 ns, using the Amber 10[15] package, periodic boundary conditions (PBC), and the FF99SB[16] force field. Ligand parameters were generated from the general Amber force field (GAFF),[67] and charges were calculated using the Antechamber module at the AM1-BCC level.[68] Ligand placement (in cases where the ligand is not found in the initial protein structure) was derived either from modifying cocrystallized molecules, using the LEaP software,[15] or by molecular docking.[69]

All systems were neutralized (Na^+/Cl^-) and solvated by an 8 Å truncated octahedron shell of explicit water molecules, modeled by the TIP3P[70] parameter set. The production runs were preceded by system relaxation and 20 ps of gradual heating to 300 K. The relaxation was conducted by 1000 steps where the protein was held fixed, followed by 5000 steps without constraints. The first half of both minimization steps was done using a steepest descent algorithm, and the second by the conjugate gradient method. The gradual heating was conducted while slowly releasing a mild constraint on the protein atoms (10 kcal/mol Å). Long range electrostatics were handled by Particle Mesh Ewald (PME) summation.[71] The time step was set to 2 fs, and consequently hydrogen movements were constrained by the SHAKE algorithm.[72] All solvated ligands were simulated for 1 ns in 16 Å water boxes, using an otherwise identical protocol.

Backbone RMSDs were analyzed for each trajectory to ensure that each system was sufficiently equilibrated before extracting coordinates for energy calculations, and each simulation was visually inspected to verify that it was well-behaved. We extracted 300−500 snapshots for computation of $\langle V^{el}\rangle$ and $\langle V^{vdW}\rangle$ in protein and solvent, respectively.

2.3. LIE Calculations and System-Dependent Coefficients. $\langle V^{el}\rangle$ and $\langle V^{vdW}\rangle$ were calculated using the ANAL script from AMBER 9,[73] using a cutoff distance of 14 Å (see Table S1 for raw data). The dependence on cutoff distance and the number of snapshots were examined for a subset of entries, as described in the Supporting Information, and no significant deviations from the default truncations were observed.

The trial models for ΔG_{bind} in this work take on the standard form of the LIE formula (eq 1), i.e.

$$\Delta G^{new}_{bind} = \alpha_{l,p}\Delta\langle V^{vdW}_{l-s}\rangle + \beta_{l,p}\Delta\langle V^{el}_{l-s}\rangle + \gamma'_{l,p}\langle \Delta SASA\rangle \quad (4)$$

where $x_{l,p}$ indicates that the coefficient is a function of one or several ligand and/or protein properties $p_{l,p}$. For example, $\alpha_{l,p} = c_\alpha p_{l,p}$ where c_α is a free parameter and $p_{l,p}$ is one of the descriptors given in Table 2. Each considered property is designed to be dimensionless and vary over a relatively small range (ideally 0 to 1), and all are derived from the systems' molecular structure. For comparison, we also evaluate models with nonvariable coefficients (see below).

Table 2. Definition of Considered Ligand- and Protein-Dependent Properties

property	$p_{l,p}$	definition[a]
ligand sphericity	a_l	$(36\pi v^2_{lig})^{1/3}/SA_{lig}$
reduced ligand volume	v_{red}	v_{lig}/v_{cav}
cavity polarity	η	$PSA_{cav}/SA^{total}_{cav}$
ligand polarity	π	$PSA_{lig}/SA^{total}_{lig}$

[a]SA = surface area, PSA = polar SA.

With a robust, system-dependent parametrization, the binding energy prediction used in the LIE method has potential for substituting scoring functions in molecular (ensemble) docking.

Table 3. Fitted Coefficients and Errors for the Main Models Considered[a]

	definitions			fitted parameters					
model	α	β	γ'	c_1	c_2	c_3	MAD	$\lvert\varepsilon_{max}\rvert$	R^2
1	0.18	0.5	c_1	0.0074			2.65	7.3	−0.02
2	0.18	0.5	$c_1(1-\eta)$	0.012			2.96	7.5	−0.93
3	0.18	0.5	$c_1(1-\pi)$	0.010			2.57	8.6	−0.06
4	$c_1(1-\eta)$	$c_1\eta$	0	0.67			2.05	5.4	0.50
5	$c_1(1-\pi)$	$c_1\pi$	0	0.43			1.95	4.9	0.59
6	$c_1(2-\eta-\pi)$	$c_1(\eta+\pi)$	0	0.27			1.92	4.3	0.63
7	$c_1(1-\eta)(1-\pi)$	$c_1(\eta\pi)$	0	0.74			2.06	6.1	0.53
8	$c_1(2-\eta-\pi)$	$c_2(\eta+\pi)$	c_3	0.19	0.15	0.0046	1.66	3.4	0.72
9[b]	0.5	0.5	0				2.93	15.9	−0.43
10[c]	c_1	c_2	c_3	0.12	0.15	0.0072	1.94	4.6	0.61

[a]Raw data and descriptors are provided in Table S1. [b]No fitting was conducted. [c]Complete refit of constant α, β, γ.

With this in mind, we chose to treat each property in Table 2, as well as the ΔSASA terms,[74] as single point values derived from crystal or minimized structures, although we recognize that they can in principle be treated as ensemble averages.

Ligand volumes and surface areas were calculated using the "3V" tool on the molmovpdb.org server.[75,76] Ligand polar SA was estimated using the topological polar surface area (TPSA) approach of Ertl et al.[77] Cavity volume is a property that cannot be uniquely defined and must ultimately be based on human decision. In this work, we have used the CavBase entries in the Relibase+[78−80] database as the cavity definition; in cases of ambiguity we examined each entry. The cavity polarity was also extracted from CavBase data. We note that there is no uniquely valid way of defining these properties, and different methods can be used to calculate SAs, volumes, etc.

On the basis of physical arguments, a range of expressions were tested against the data set presented above. Depending on the expression, 1−3 coefficients were fitted using (multi)linear regression, and the results were evaluated on the basis of mean absolute deviation (MAD), maximum error ($\lvert\varepsilon_{max}\rvert$), and the correlation coefficient (R^2). The latter is calculated by eq 5:

$$R^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} \tag{5}$$

where SSE is the sum of squared errors and is given by $\sum_i(\Delta G_{i,\mathrm{calc}} - \Delta G_{i,\mathrm{exp}})^2$, and SST is the total sum of squares, $\sum_i(\Delta G_{i,\mathrm{exp}} - \langle\Delta G_{\mathrm{exp}}\rangle)^2$. Note that R^2 may assume values below zero if a full optimization based on regression analysis is not conducted, which is the case in several of the trial models (see below).

3. RESULTS

3.1. Fitted Models. Among the new expressions for ΔG_{bind} examined, only a few gave MAD values of <2 kcal/mol when fitted to the 32-entry data set. We will concentrate on the best performing models or those otherwise elucidating some key point in the main text, with additional results given in Table S2. The models discussed here are presented in Table 3 along with resulting coefficients and errors. Model 9 is the LIE "half-and-half" model used in Figure 1a, while model 10 is the LIE + γSASA model with three refitted coefficients. Fitting parameters are denoted c_i.

While the shape descriptors a_l and v_{red} did not add any significant accuracy to the tested formulas (Table S2), the cavity and ligand polarity (η and π) provided a remarkable improvement compared to all models where α and β were given fixed values (models 1−3, 9). With only one fitted parameter, the data set was fitted to a MAD as low as 1.9 kcal/mol (model 6). The use of cavity and ligand polarities to scale the LIE coefficients is reminiscent of the "weighted nonpolar desolvation ratio" by Kollman and co-workers,[35] although we use a simpler expression and scale both α and β. We also note that the free parameter fitting conducted in model 10 gave values close to those obtained by Jones-Hertzog and Jorgensen,[20] who used a different data set, a different force field (OPLS-AA),[81] Monte Carlo simulations, and a different code to generate their results. We interpret this similarity as an indication that despite inevitable deviations in absolute values of energies, our simulation protocol seems to generate sound figures.

Intriguingly, the protein and ligand descriptors are most effective when used together (model 6), which seems to take care of the most severe outliers present in both model 4 and model 5. Taking the product of both descriptors (model 7) or using π alone (model 5) yields decent MADs but is problematic since in some instances (e.g., benzene) $\pi = 0$. A small improvement is provided by including the γSASA term in the best performing set of descriptors and varying all three coefficients (model 8).

Models 4−7 have only one fitted parameter and are not significantly improved by letting α and β be fitted independently (MAD = 2.08, 1.96, 1.91, 1.87 and $\lvert\varepsilon_{max}\rvert$ = 4.6, 4.9, 4.3, 6.1 for the respective two-parameter models). Interestingly, the fitting parameters used in the two-parameter variants of models 5 and 6 become virtually equal upon regression (Table S2, entries 13 and 14). Hence, the introduction of the system-dependent descriptors η and π has effectively reduced the LIE model from depending on two or three empirical parameters to only one. We propose to denote this novel LIE treatment "Adapted LIE" (ALIE) since α and β are adapted to each given system by intrinsic properties. We will focus on these models in the following.

3.2. Dependence of Different Ways to Derive π. In practice, α and β are scaled between the values displayed in Table 4. There are no dramatic changes compared to previous

Table 4. Extreme Values of α and β in the ALIE Models

	α		β	
model	min	max	min	max
model 4	0.28	0.48	0.19	0.40
model 5	0.17	0.43	0.00	0.27
model 6	0.21	0.44	0.10	0.32
model 7	0.12	0.48	0.00	0.27

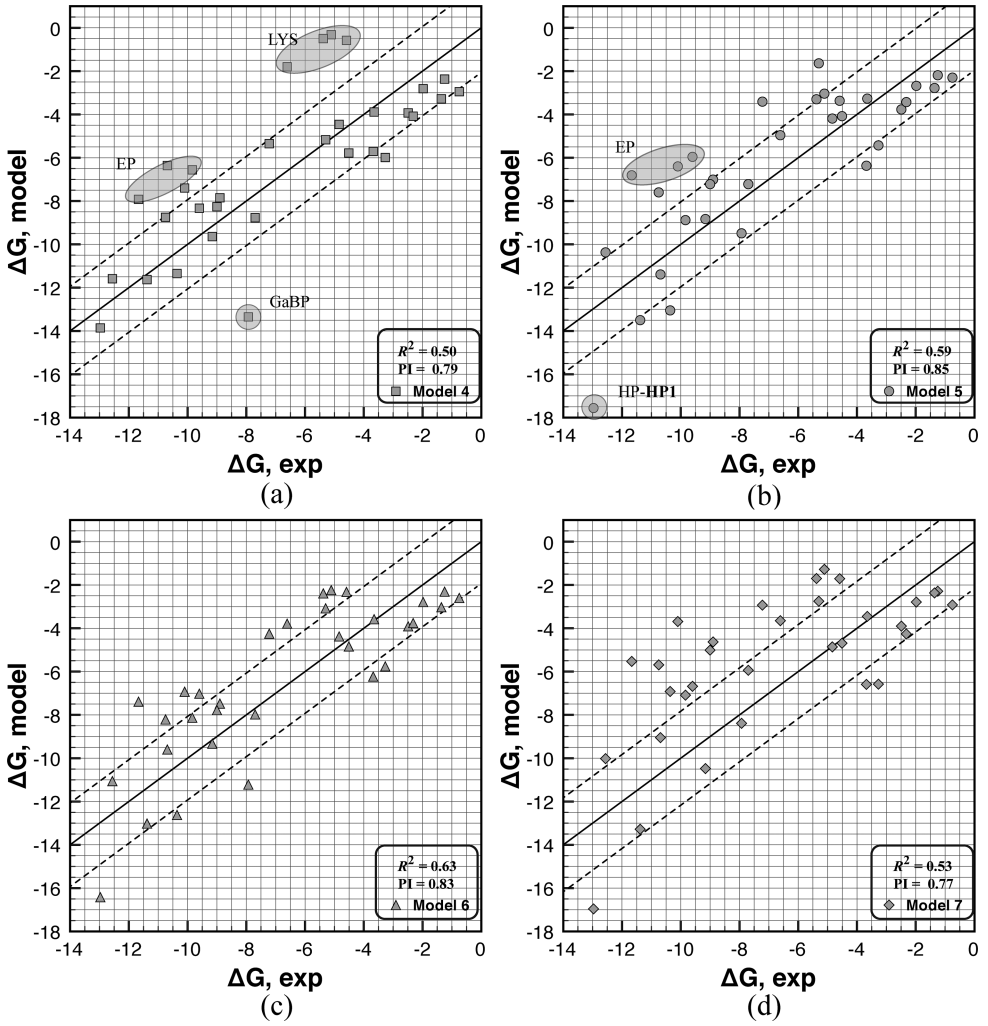


Figure 2. Model vs experimental ΔG_{bind} of (a–d) models 4–7. All energies are in kcal/mol. Correlation coefficients (R^2) and predictive indexes (PI) are given in each graph. The unbroken line denotes $x = y$. MADs are indicated as dashed lines. Selected outlying systems are indicated in shaded regions.

LIE parametrizations, but it is interesting to note that all four ALIE models provide larger α and smaller β coefficients than the standard models.[10] As mentioned above, it can be considered problematic that models 5 and 7 yield $\beta = 0$ in cases where the simple TPSA method[77] (and other similar approaches) predicts a completely nonpolar SA. This coarseness could possibly be amended with a more advanced description of the polar SA.

Consequently, we refitted models 5–7 using the "total local polarity" of Brinck et al.,[82] which is derived from the electrostatic potential of an *ab initio* calculation (and is always >0 except in spherically symmetric systems). This description provided very similar MADs but did not lead to an improved correlation. Hence, the ALIE approach does not seem overly sensitive to the choice of definition of the properties used (see the Supporting Information for details).

3.3. Correlation and Outliers. Calculated versus experimental ΔG_{bind} values for models 4–7 are plotted in Figure 2, along with their R^2 values and predictive indices (PIs) as defined by Pearlman and Charifson.[83,84] The PI is a measure of how capable a model is of ranking pairs of data points correctly, which is of course an important attribute for a model aiming at predicting binding energies without calibration. The modest R^2 values of the ALIE models can be attributed to the clusters of outliers present in all four cases, as exemplified by the shaded regions. However, they are superior to the correlations seen in Figure 3 for models without system-dependent coefficients.

From the selected SASA-dependent and unfitted models displayed in Figure 3, the sugar molecules are identified as the most severe outliers. Their common feature is a very large and negative $\Delta\langle V^{el}_{l-s}\rangle$ (see Table S1), which with the usual LIE scaling[10] leads to an overestimation of ΔG_{bind}. A strictly negative γSASA value does obviously not improve the situation in such cases. In contrast, the binding energies of these ligands are fairly well reproduced by the ALIE models.

The above example highlights the benefits of the system-dependent scaling of the α and β coefficients: It makes ALIE flexible enough to handle extreme cases, from nonpolar ligands with variable size (which LIE + γSASA also handles well) to small molecules where electrostatic interactions dominate the binding energy. The PI values further demonstrate that they are reasonably successful at ranking the data points correctly. In comparison, the PIs are 0.66, 0.73, and 0.78 for models 1, 9, and 10, respectively.[84]

The outliers in Figure 2 are typically clustered in groups of a single protein with its series of ligands. Therefore, when performing a leave-one-out cross-validation of model 6 (see Table S3), we excluded all ligands for one protein at a time. We obtained a predictive MAD of 2.05 kcal/mol and a q^2 value of

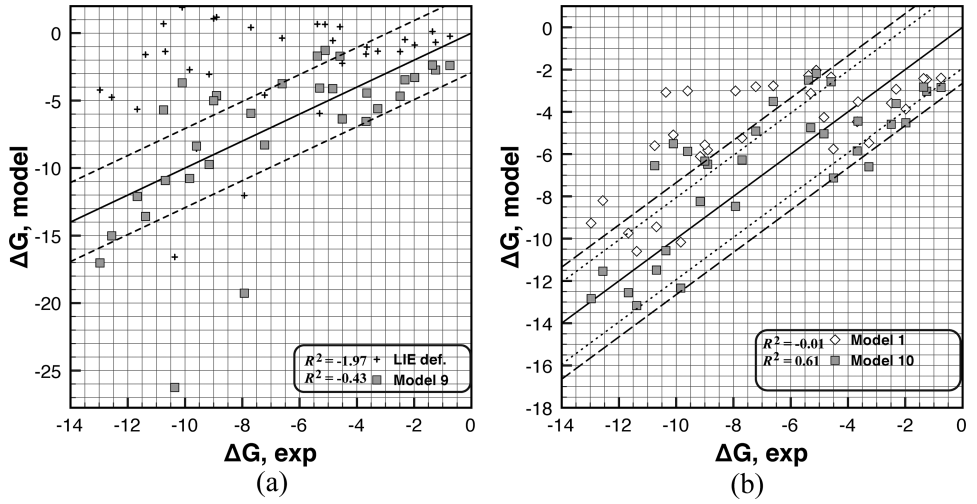


Figure 3. Model vs experimental ΔG_{bind} of (a) model 9 and the standard LIE model ("LIE def.," α = 0.18, β = 0.43) and (b) models 1 and 10. All energies are in kcal/mol. Correlation coefficients are given in each graph. The unbroken line denotes $x = y$. MADs are indicated as dashed lines, where the larger spacing is used for model 1 in b.

0.57. The average cross-validated standard deviation (s_{PRESS}) was 2.44 kcal/mol. Although no set of ligands stand out as dominant outliers, the best refitted MADs were found when removing LYS (1.77) and EP (1.84). With both removed simultaneously (25 ligands remaining), c_1 changed slightly to 0.25 while the MAD improved to 1.66 (Table S3).

3.4. Charged Systems. Although the ALIE models were fitted against a set of neutral ligands, it is interesting to probe their applicability on charged systems. A multitude of systems relevant for medicinal research contain ionic ligands, so it is vital for an effective model to treat these accurately. Thus, we selected a small set of ionic ligands, namely **BA1**−**BA4**, **fat**, and **HP4** (see Tables 1 and S1 as well as Figure S2). The system setup, simulation, and calculation of descriptors was done as described above, and the fitting parameters from Table 3 were used to calculate ΔG_{bind}.

Figure 4 displays the results employing models 4−7 as well as standard LIE. The accuracy is surprisingly good, more so for the more π-dependent models 5 and 7. This result indicates that the ALIE approach could indeed be extended to charged ligands. Instead, the general convergence problems in calculating the electrostatic linear response energy probably has more influence on the actual accuracy.

3.5. External Systems. As a final challenge of the robustness of the ALIE methods, we recalculated binding energies for three ligand series based on data from two other groups. The first series is the cytochrome p450−camphor derivate series described in Figure 1a,[12] the second is avidin with a series of biotin derivatives, and the third is the human factor Xa with a series of inhibitors. The raw data of the latter two are taken from a study by Genheden and Ryde,[37] where a variety of linear approaches was used for the electrostatic and nonpolar contributions. We selected the data obtained from the standard LIE approach for calculating the nonpolar energy and MMGB with ε = 4 for electrostatics, to be as consistent as possible with the original study.[85] Although this is not strictly LIE, we let this related method serve as an approximation. For comparison, we include a pure LIE treatment of the avidin−biotin system conducted by the same authors.[36] The **Btn1**−**Btn3** ligands were excluded due to the notoriously deviant results obtained for these strong binders.

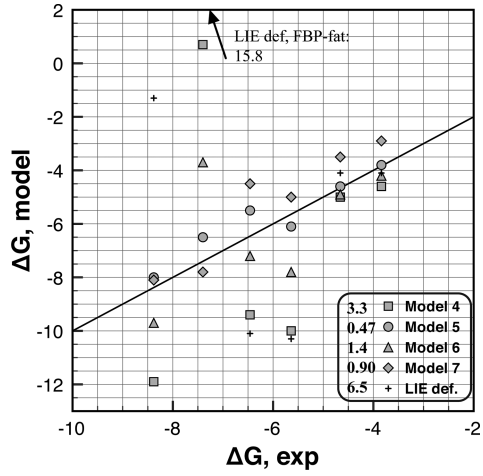


Figure 4. Model vs experimental ΔG_{bind} of a series of six ionic systems using ALIE models and standard LIE (LIE def.) All energies are in kcal/mol and the solid line denotes $y = x$. The numbers next to the legend indicate mean absolute deviations (MADs) for each model. The arrow indicates one LIE data point outside the graph.

To apply the ALIE models, we determined η and π for each protein and ligand and used the parameters listed in Table 3 to scale the raw data. In addition, we computed binding energies using the methods from the original references ("LIE def.") as well as model 1, using single-point areas to compute ΔSASA. The results are summarized in Table 5. ΔG_{bind} of models 1, 4, and 6 compared to experimental and originally computed values are given in Figure 5. For these ligand series, model 1 performs better than the ALIE models in all cases except the avidin ligands, consistent with the good performance of "standard" LIE + γSASA when applied to data sets with small contributions from $\Delta\langle V^{el}_{l-s}\rangle$, c.f. Figures 1 and 3). However, ALIE significantly lowers the MADs compared to the original models for the Avidin and Factor Xa series. It can be seen from Figure 4 that models 4 and 6 underestimate ΔG_{bind} in both the p450 and factor Xa series, which is reminiscent of the "bands" of outliers observed in Figure 2. Nevertheless, the results are surprisingly accurate considering the fact that the electrostatic energies were obtained by different protocols than the method used for fitting the ALIE models.

dx.doi.org/10.1021/ct300783e | J. Chem. Theory Comput. 2013, 9, 1230−1239

Table 5. Errors When Applying Fitted Models to External Data Sets[a]

system	error	LIE def.[b]	model 1	model 4	model 6
p450−LIE[12]	MAD	0.5[c]	1.4	2.5	2.3
(6 ligands)	$\|\varepsilon_{max}\|$	1.2[c]	2.3	2.8	2.7
factor Xa−MMGB[37]	MAD	5.7	0.9	2.4	2.1
(9 ligands)	$\|\varepsilon_{max}\|$	7.0	3.3	3.4	3.0
avidin−MMGB[37]	MAD	4.1	0.8	0.7	0.6
(4 ligands)	$\|\varepsilon_{max}\|$	5.0	1.5	1.3	1.6
avidin−LIE[36]	MAD	3.5	1.6	1.2	1.1
(4 ligands)	$\|\varepsilon_{max}\|$	5.9	2.6	1.9	1.8

[a]For all external systems, ΔG_{bind} essentially consists of a polar and nonpolar part (model 1 has an additional SA term), scaled according to Table 3. [b]Using the parametrization from the original models (i.e., α = 0.18, β = 0.43 for the p450 series and 0.5 for the factor Xa and avidin series.) [c]The default energy includes a constant γ = −4.82 kcal/mol.

4. DISCUSSION

The ALIE models developed in this study overall perform better than most other methods tested for this diverse data set, the exception being one with three refitted parameters. In models 4−7, the number of fitted parameters is reduced to one. It thus seems that the inclusion of system-dependent descriptors adds the flexibility other LRA approaches lack. We observe that for smaller data sets with more homogeneous electrostatics (such as **VE1−VE5** or the external fXa inhibitor series), a LIE + γSASA approach can be more precise. Nevertheless, the simplicity and broad range of applicability of ALIE is encouraging and may have interesting applications in enzyme design and drug discovery. The MAD of 1.92 kcal/mol for model 6 can for example be compared with the scoring function SCORE2 by Böhm[25] (fitted MAD = 1.75 kcal/mol, seven free parameters) and to ChemScore[86] (fitted MAD = 1.85 kcal/mol, multiple free parameters). Both cavity and ligand polarity (η and π) are readily obtained from structural information, and we have attained good accuracy with quite limited sampling. This makes the ALIE approach a good method for initial prediction of ΔG_{bind} when no experimental data are available.

At this point, it is appropriate to ask what physical meaning the system-dependent scaling of α and β has. Ideally, one would wish for the FEP-derived β = 0.5 to work universally, but as has been pointed out previously, this is usually not the case.[34] As shown for solvation energies in ref 34, the accuracy of FEP solvation energies is improved considerably by letting β vary with the chemical nature of the ligand. The nonpolar term in the LIE model is essentially size-dependent and does in principle contain entropic terms as well.[87] In the ALIE models, η and π effectively scale the LIE coefficients up or down depending on the polar nature of the ligand and cavity, so that a highly nonpolar system will have an enhanced contribution to ΔG_{bind} from $\langle V_{l-s}^{vdW}\rangle$. In contrast, a ligand with high polarity (high π) in a cavity with low polarity (low η) will get an even spread of contributions from $\langle V_{l-s}^{el}\rangle$ and $\langle V_{l-s}^{vdW}\rangle$ in model 6. In a way, these scalings are reminiscent of chemistry-based determination of the coefficients,[12,34] although the polarity measure disregards the nature of the contributing functional groups. One could also speculate that the scaling of the electrostatic energies to some extent models the nonlinear effects of the Coulombic interactions due to polarization. However, it is not appropriate to draw any lengthy conclusions about the physical implication of variable coefficients; the whole family of LIE approaches (including LIE + γSASA, ELR, and ALIE) are (semi)-empirical by nature.

A persistent issue with approximate free energy calculations is the dependence on choice of force field, sampling, and statistical convergence, which means that in principle, each unique protocol needs a unique set of parameters fitted to experimental data. For example, we note that the unscaled energies produced by our protocol (see Table S1) are quite different for several ligands from those reported by Åqvist and co-workers.[2,10] However, refitting the single free parameter in the ALIE models for the external data sets (not shown) yields good accuracy, suggesting that the approach can be implemented on an arbitrary simulation protocol.

In this work, we have had no intentions of comparing simulation methods, truncation errors, or statistical convergence. We have primarily been interested in improving the predictive power of the LIE method, so that quantitative estimates can be obtained from MD simulations without extensive simulation time or any *a priori* knowledge of the system (save structural information). To this end, we tested a number of system-dependent parameters that scale the LIE energy on a diverse

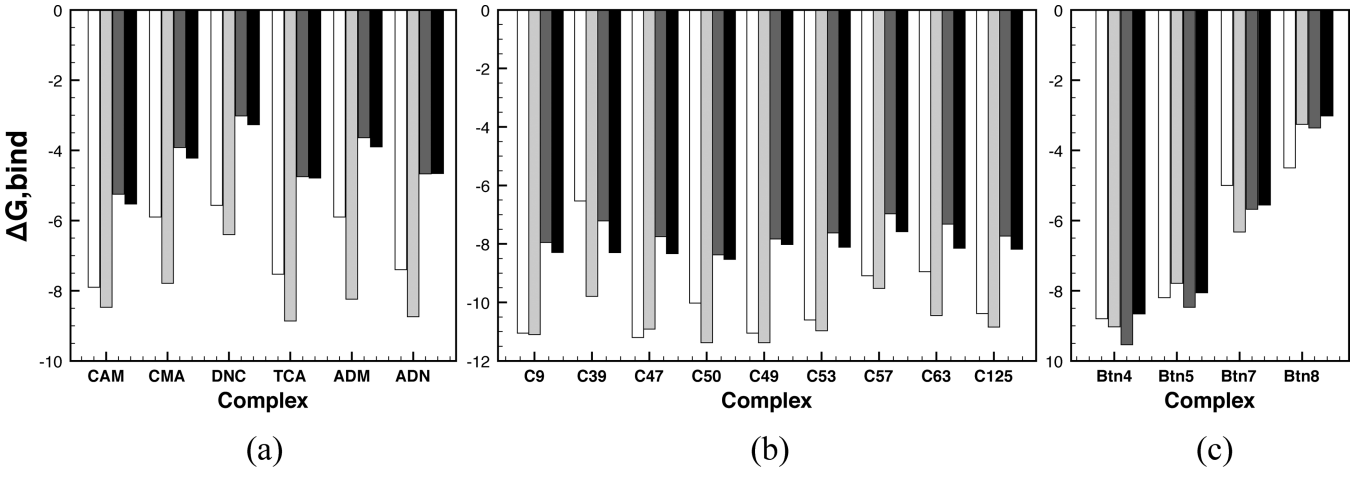


Figure 5. Data sets taken from the literature and recalculated using models 1 (light gray), 4 (dark gray), and 6 (black). White bars represent experimental ΔG_{bind} values. (a) Cytochrome P450 with a series of camphor derivates.[12] (b) A series of inhibitors to fXa.[37] (c) Biotin derivates binding to avidin (MMGB energies).[37]

protein−ligand data set, which led to the one-parameter ALIE models. The ligand and cavity polarity descriptors were clearly singled out as important, but there are probably more. We surmise that productive additions should be Coulombic by nature (e.g., polarizability, "softness," etc.), and perhaps size- or shape-related, although the latter should implicitly be included in the nonpolar energy.

Introducing system-dependent parameters also raises the issue of how to best calculate them. In this work, we have used very approximate definitions to keep additional processing at a minimum. The finding that two other formulations of π, one classical and one quantum chemical (see the Supporting Information), did not lead to increased accuracy supports this choice. However, as the charged systems appear to be better described by the more π-dominated models 5 and 7, it may be a good idea to search further for a description that does not return "0" each time a ligand lacks a heteroatom. We also note that by the way the coefficients are defined in the one-parameter models, $\alpha + \beta$ always equal a constant $2c_1$. This constraint is not necessary, although we have seen that it is close to optimal, in particular in model 6.

This study does not answer the question of whether this approach is universally applicable, or whether ligand and cavity polarity are the optimal properties to use for scaling the LIE coefficients (or if they are described appropriately). However, the results presented in this work raise interesting questions about the dominant interactions for different protein−ligand systems and open up intriguing possibilities for designing more general energy functions for both force field methods and molecular docking. The decent accuracy obtained when rescaling the energies from the three external data sets is an encouraging result indicating that ALIE is a general approach, although it requires further analysis and benchmarking.

5. CONCLUSION

We have designed an "Adapted" LIE (ALIE) method that scales the α and β coefficients using system-dependent polarity descriptors. We have shown that ALIE works well for a broad range of protein−ligand binding energies and can be used to calculate absolute binding affinities with an expected accuracy of ±2 kcal/mol. While it is certainly possible to improve the accuracy for a given subset of protein−ligand systems using either the LIE-γSASA or ELR approach, the one-parameter models presented herein show a broad applicability while requiring only limited knowledge of the system.

The standardized simulation methods used in this work suggest that the model can be used as a computationally efficient way to obtain good estimates of the ensemble-average based binding free energy directly from MD studies of protein−ligand interactions. As in the original LIE model, no additional simulations are required except for the solvated ligand, and the energy sampling is conveniently obtained along with all other trajectory information. This, in turn, might have applications in computational protein design as well as the design of energy-based scoring functions.

■ ASSOCIATED CONTENT

Supporting Information

Additional figures, data, and a discussion on two alternate ways to calculate the π descriptor. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: tore@physchem.kth.se.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work has been supported by the Cambridge Crystallographic Data Centre (CCDC). The authors thank Dr. John Liebeschuetz and Dr. Tjelvar Olsson at the CCDC for helpful critical discussions.

■ REFERENCES

(1) Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. *Curr. Med. Chem.* **2004**, *26*, 3093−3118.
(2) Åqvist, J.; Medina, C.; Samuelsson, J. *Protein Eng.* **1994**, *7*, 385−391.
(3) Åqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. *Acc. Chem. Res.* **2002**, *35*, 358−365.
(4) Brandsdal, B. O.; Österberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Åqvist, J. Protein Simulations. *Adv. Protein Chem.*; Daggett, V., Ed.; Academic Press: New York, 2003; Vol. *66*; pp 123−158.
(5) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.
(6) Massova, I.; Kollman, P. *Perspect. Drug Discovery Des.* **2000**, *18*, 113−135.
(7) Chipot, C.; Pohorille, A. Free Energy Calculations: Theory and Applications in Chemistry and Biology. In *Springer Series in Chemical Physics*; Springer: New York, 2007; Vol. *86*.
(8) Huang, S.-Y.; Zou, X. *Int. J. Mol. Sci.* **2010**, *11*, 3016−3034.
(9) Selassie, C. In Drug Discovery. *Burger's Medicinal Chemistry and Drug Discovery*; Abraham, D., Ed.; John Wiley & Sons: New York, 2003; Vol. *1*; pp 1−48.
(10) Hansson, T.; Marelius, J.; Åqvist, J. *J. Comput.-Aided. Mol. Des.* **1998**, *12*, 27−35.
(11) Paulsen, M. D.; Ornstein, R. L. *Protein Eng.* **1996**, *9*, 567−571.
(12) Almlöf, M.; Brandsdal, B. O.; Åqvist, J. *J. Comput. Chem.* **2004**, *25*, 1242−1254.
(13) Derewenda, U.; Brzozowski, A. M.; Lawson, D. M.; Derewenda, Z. S. *Biochemistry* **1992**, *31*, 1532−1541.
(14) Chahinian, H.; Ali, Y. B.; Abousalham, A.; Petry, S.; Mandrich, L.; Manco, G.; Ccanaan, S.; Sarda, L. *Biochim. Biophys. Acta* **2005**, *1738*, 29−36.
(15) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. *AMBER 10*; University of California: San Francisco, CA, 2008.
(16) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712−725.
(17) Åqvist, J.; Marelius, J. *Comb. Chem. High Throughput Screening* **2001**, *35*, 613−626.
(18) Carlsson, J.; Andér, M.; Nervall, M.; Åqvist, J. *J. Phys. Chem. B* **2006**, *110*, 12034−12041.
(19) Carlson, H. A.; Jorgensen, W. L. *J. Phys. Chem.* **1995**, *99*, 10667−10673.
(20) Jones-Hertzog, D. K.; Jorgensen, W. L. *J. Med. Chem.* **1997**, *40*, 1539−1549.
(21) Lamb, M. L.; Tirado-Rives, J.; Jorgensen, W. L. *Bioorg. Med. Chem.* **1999**, *7*, 851−860.
(22) Zhou, R.; Friesner, R. A.; Ghosh, A.; Rizzo, R. C.; Jorgensen, W. L.; Levy, R. M. *J. Phys. Chem. B* **2001**, *105*, 10388−10397.

(23) Su, Y.; Gallicchio, E.; Das, K.; Arnold, E.; Levy, R. M. *J. Chem. Theory Comput.* **2007**, *3*, 256−277.
(24) Bortolato, A.; Moro, S. *J. Chem. Inf. Model.* **2007**, *47*, 572−582.
(25) Böhm, H.-J. *J. Comput.-Aided. Mol. Des.* **1998**, *22*, 309−323.
(26) Duffy, E. M.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2000**, *122*, 2878−2888.
(27) Rizzo, R. C.; Tirado-Rives, J.; Jorgensen, W. L. *J. Med. Chem.* **2001**, *44*, 145−154.
(28) Pierce, A. C.; Jorgensen, W. L. *J. Med. Chem.* **2001**, *44*, 1043−1050.
(29) Rizzo, R. C.; Udier-Blagović, M.; Wang, D.-P.; Watkins, E. K.; Kroeger Smith, M. B.; Smith, R. H.; Tirado-Rives, J.; Jorgensen, W. L. *J. Med. Chem.* **2002**, *45*, 2970−2987.
(30) Wesolowski, S. S.; Jorgensen, W. L. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 267−270.
(31) Udier-Blagović, M.; Watkins, E. K.; Tirado-Rives, J.; Jorgensen, W. L. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3337−3340.
(32) Ostrovsky, D.; Udier-Blagović, M.; Jorgensen, W. L. *J. Med. Chem.* **2003**, *46*, 5691−5699.
(33) The sum in eq 3 is typically reduced to 2−3 terms plus a constant correction, i.e., the same level of parametrization as in LIE.[29−32]
(34) Almlöf, M.; Carlsson, J.; Åqvist, J. *J. Chem. Theory Comput.* **2007**, *3*, 2162−2175.
(35) Wang, W.; Wang, J.; Kollman, P. A. *Proteins* **1999**, *34*, 395−402.
(36) Genheden, S.; Ryde, U. *J. Chem. Theory Comput.* **2011**, *7*, 3768−3778.
(37) Genheden, S.; Ryde, U. *Proteins* **2012**, *80*, 1326−1342.
(38) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids. Res.* **2000**, *28*, 235−242.
(39) Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9997−10002.
(40) In the case of enzymes, we have approximated the binding constant K_S with the Michaelis constant K_M, since it can be assumed that the binding and release steps are faster than the reaction rate except in the diffusion limit.
(41) Yapoudjian, S.; Ivanova, M. G.; Brzozowski, A. M.; Patkar, S. A.; Vind, J.; Svendsen, A.; Verger, R. *Eur. J. Biochem.* **2002**, *269*, 1613−1621.
(42) Uppenberg, J.; Öhrner, N.; Norin, M.; Hult, K.; Kleywegt, G. J.; Patkar, S.; Waagen, V.; Anthonsen, T.; Jones, T. A. *Biochemistry* **1995**, *34*, 16838−16851.
(43) Skjot, M.; De Maria, L.; Chatterjee, R.; Svendsen, A.; Patkar, S. A.; Ostergaard, P. R.; Brask, J. *ChemBioChem* **2009**, *10*, 520−527.
(44) Eriksson, A. E.; Baase, W. A.; Wozniak, J. A.; Matthews, B. W. *Nature* **1992**, *355*, 371−373.
(45) Morton, A.; Matthews, B. W. *Biochemistry* **1995**, *34*, 8576−8588.
(46) Cowan, S. W.; Newcomer, M. E.; Jones, T. A. *Proteins* **1990**, *8*, 44−61.
(47) Ealick, S. E.; Babu, Y. S.; Bugg, C. E.; Erion, M. D.; Guida, W. C.; Montgomery, J. A.; Secrist, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 11540−11544.
(48) Zollner, H. *Handbook of Enzyme Inhibitors*; VHC Publishers: Weinheim, Germany, 1993.
(49) Cho, H.-S.; Ha, N.-C.; Choi, G.; Kim, H.-J.; Lee, D.; Oh, K. S.; Kim, K. S.; Lee, W.; Choi, K. Y.; Oh, B.-H. *J. Biol. Chem.* **1999**, *274*, 32863−32868.
(50) Hawkinson, D. C.; Eames, T. C. M.; Pollack, R. M. *Biochemistry* **1991**, *30*, 10849−10858.
(51) Lam, P.; Jadhav, P.; Eyermann, C.; Hodge, C.; Ru, Y.; Bacheler, L.; Meek, J.; Otto, M.; Rayner, M.; Wong, Y.; et al. *Science* **1994**, *263*, 380−384.
(52) Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. *J. Am. Chem. Soc.* **1995**, *117*, 1181−1182.
(53) Lam, P.; Jadhav, P.; Eyermann, C.; Hodge, C.; Ru, Y.; Bacheler, L.; Meek, J.; Otto, M.; Rayner, M.; Wong, Y. *Science* **1994**, *263*, 380−384.
(54) Erickson, J.; Neidhart, D.; VanDrie, J.; Kempf, D.; Wang, X.; Norbeck, D.; Plattner, J.; Rittenhouse, J.; Turon, M.; Wideburg, N. *Science* **1990**, *249*, 527−533.
(55) Mowbray, S.; Smith, R.; Cole, L. *Receptor* **1990**, *1*, 41−53.
(56) Åqvist, J.; Mowbray, S. L. *J. Biol. Chem.* **1995**, *270*, 9978−9981.
(57) Vyas, N.; Vyas, M.; Quiocho, F. *Science* **1988**, *242*, 1290−1295.
(58) Miller, D. M.; Olson, J. S.; Pflugrath, J. W.; Quiocho, F. A. *J. Biol. Chem.* **1983**, *258*, 13665−13672.
(59) Whitlow, M.; Howard, A. J.; Finzel, B. C.; Poulos, T. L.; Winborne, E.; Gilliland, G. L. *Proteins* **1991**, *9*, 153−173.
(60) Pearl, L.; Blundell, T. *FEBS Lett.* **1984**, *174*, 96−101.
(61) 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. a. *J. Am. Chem. Soc.* **1993**, *115*, 9925−9938.
(62) Marquart, M.; Walter, J.; Deisenhofer, J.; Bode, W.; Huber, R. *Acta Crystallogr., Sect. B* **1983**, *39*, 480−490.
(63) Mares-Guia, M.; Shaw, E. *J. Biol. Chem.* **1965**, *240*, 1579−1585.
(64) Sacchettini, J. C.; Gordon, J. I.; Banaszak, L. J. *J. Mol. Biol.* **1989**, *208*, 327−339.
(65) Lowe, J. B.; Sacchettini, J. C.; Laposata, M.; McQuillan, J. J.; Gordon, J. I. *J. Biol. Chem.* **1987**, *262*, 5931−5937.
(66) Miller, M.; Schneider, J.; Sathyanarayana, B.; Toth, M.; Marshall, G.; Clawson, L.; Selk, L.; Kent, S.; Wlodawer, A. *Science* **1989**, *246*, 1149−1152.
(67) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(68) (a) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132−146. (b) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623−1641.
(69) GOLD 5.0. http://www.ccdc.cam.ac.uk/products/life_sciences/gold/ (accessed Dec. 2012).
(70) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926−935.
(71) (a) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10093. (b) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pederson, L. *J. Chem. Phys.* **1995**, *103*, 8577−8592.
(72) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(73) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S. R.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W.; Kollman, P. A. *AMBER 9*; University of California: San Francisco, CA, 2006.
(74) Shrake, A.; Rupley, J. A. *J. Mol. Biol.* **1973**, *79*, 351−371.
(75) Voss, N.; Gerstein, M. *Nucleic Acids Res.* **2010**, *38*, W555−W562.
(76) Gerstein, M.; Krebs, W. *Nucleic Acids Res.* **1998**, *26*, 4280−4290.
(77) Ertl, P.; Rohde, B.; Selzer, P. *J. Med. Chem.* **2000**, *43*, 3714−3717.
(78) Hendlich, M.; Bergner, A.; Günther, J.; Klebe, G. *J. Mol. Biol.* **2003**, *326*, 607−620.
(79) Schmitt, S.; Hendlich, M.; Klebe, G. *Angew. Chem., Int. Ed.* **2001**, *40*, 3141−3144.
(80) Schmitt, S.; Kuhn, D.; Klebe, G. *J. Mol. Biol.* **2002**, *323*, 387−406.
(81) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(82) Brinck, T.; Murray, J. S.; Politzer, P. *Mol. Phys.* **1992**, *76*, 609−617.
(83) Pearlman, D. A.; Charifson, P. S. *J. Med. Chem.* **2001**, *44*, 3417−3423.
(84) Pearlman and Charifson obtained PIs of 0.84 and 0.043 using thermodynamic integration and the ChemScore scoring function, respectively, for a series of 16 ligands.[83] While their set is arguably more challenging because of its homogeneity, these benchmark values give a hint on what defines a "good" predictive index.

(85) We recognize that the authors of ref 37 found different optimal values of ϵ in the two protein−ligand systems. We selected a value that in our view minimizes the differences in average errors for the MMGB electrostatics.
(86) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−445.
(87) Carlsson, J.; Åqvist, J. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5385−5395.