JCTC Journal of Chemical Theory and Computation

LETTER

pubs.acs.org/JCTC

Improving MM-GB/SA Scoring through the Application of the Variable Dielectric Model

Krishna Ravindranathan,[†] Julian Tirado-Rives,[†] William L. Jorgensen,[†] and Cristiano R. W. Guimarães*[,‡]

[†]Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States

[‡]Worldwide Medicinal Chemistry Department, Pfizer Global Research and Development, Eastern Point Road, Groton, Connecticut 06340, United States

Supporting Information

ABSTRACT: A variable dielectric model based on residue types for better description of protein−ligand electrostatics in MM-GBSA scoring is reported. The variable dielectric approach provides better correlation with binding data and reduces the score dynamic range, typically observed in the standard MM-GB/SA method. The latter supports the view that exaggerated enthalpic separation between weak and potent compounds due to the lack of shielding effects in the model is greatly responsible for the wide scoring spread.

INTRODUCTION

Molecular mechanics based scoring methods using all atom force fields coupled with Poisson−Boltzmann (MM-PB/SA)[1] or generalized Born calculations (MM-GB/SA)[2] to model solvation have recently seen an upsurge in popularity. When compared to docking scoring functions, the physics-based methods provide improved enrichment in the virtual screening of databases and better correlation between calculated binding affinities and experimental data.[3] Their main purpose is to rescore docking poses in order to circumvent limitations of the docking scoring functions, in particular the ones associated with desolvation, intramolecular, and entropy penalties for the ligands upon binding.

We investigated the performance of our own flavor of MM-GB/SA when rescoring docking poses of congeneric series for pharmaceutically relevant targets.[4] The correlations with experimental results obtained with the physics-based scoring were far superior to the ones obtained with the Glide XP scoring function and competitive with the computationally intensive FEP methods.[4,5] Despite showing good accuracy when applied within a series, much work is necessary to improve the MM-GB/SA method and gain greater efficiency in drug design. For example, MM-GB/SA suffers from poor estimation of protein desolvation and a large dynamic range observed in the scoring when compared to the experimental range.

In the case of the GB/SA protein desolvation, substituting this term with the free energy associated with displacing binding-site waters upon ligand binding estimated by the WaterMap method,[6] which treats the solvent explicitly, provides superior results.[4b,7] As for the large theoretical dynamic range for the binding energies, that seems to be a direct result of the degree of sampling since FEP simulations with restricted flexibility as well as MM-GB/SA approaches that make use of a single configuration for the protein−ligand complexes are plagued by this.[4a] Computational van't Hoff analysis suggests that the wider scoring spread is affected by not only missing entropic contributions due to restricted sampling but also exaggerated enthalpic separation between compounds.[4a]

One plausible cause for the exaggerated enthalpic gap is the application of an internal dielectric constant (ε_{in}) of 1 in a model where protein motions and polarization are not taken into account. Hence, electrostatic interactions are not shielded enough, and protein−ligand electrostatic attractions and repulsions are overestimated, causing the large separation between potent and weak compounds.[4a,b] As previously described, when the protein permanent dipoles are included explicitly but their relaxation, i.e., the protein reorganization, and the protein induced dipoles are considered implicitly, the value of ε_{in} is not well-defined.[8] Warshel and co-workers suggest that ε_{in} should be between 4 and 6 for dipole−charge interactions and 10 for charge−charge interactions.[9,10] More recently, a variable dielectric model has been developed to increase the accuracy in protein side chain and loop predictions.[11] The authors introduced an energy model where ε_{in} is allowed to vary as a function of the interacting residues.

In this work, we explore the use of a variable dielectric model based on residue types to alleviate the overestimation of electrostatic effects between protein residues and ligands for improved MM-GBSA scoring. Since a poor description of protein−ligand electrostatic interactions could not only result in a wider scoring spread but also affect the correlation with experimental results, we decided to use binding data to derive the set of variable dielectric constants. Specifically, the pharmaceutically relevant targets CDK2, FactorXa, p38, and PDE10A and respective congeneric series were considered in the optimization process that led to the set of variable dielectric constants, subsequently tested on two additional data sets, the human carbonic anhydrase (hCAII) and a second p38 chemical series.

Received: August 12, 2011
Published: November 14, 2011

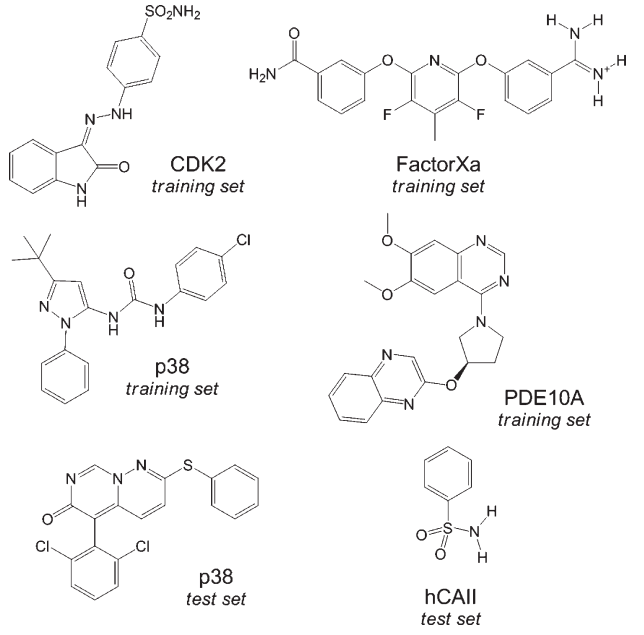

Figure 1. Representative structures for the congeneric series used in this work. Tables with all derivatives and binding data can be found in the Supporting Information.

METHODS

Data Sets and System Setup. The crystal structures for CDK2 (PDB ID: 1E9H), FactorXa (PDB ID: 1FJS), PDE10A (unpublished structure), both p38 chemical series (PDB IDs: 1KV1 and 1OUY), and hCAII (PDB ID: 2WEJ) were selected. The 1E9H CDK2 structure was modified as described before,[4c] as it is not consistent with the conditions for the biological assay. All complexes were submitted to a series of restrained, partial minimizations using the OPLS_2005 force field[12] within Macromodel.[13] Prior to the docking calculations using Glide XP,[14] the selected congeneric series of CDK2,[15] FactorXa,[16] PDE10A,[17] p38,[18,19] and hCAII[20] were submitted to a pre-energy minimization using the OPLS_2005 force field and the GB/SA method as the implicit water model.[21] Representative structures for each chemical series are illustrated in Figure 1. Tables with all derivatives and binding data can be found in the Supporting Information. In order to accommodate for the fact that the protein structure used for docking will not in general be optimized to fit a particular ligand, the van der Waals radii for nonpolar protein atoms were scaled by a factor of 0.8, while those for the ligands were not scaled.

MM-GB/SA Rescoring. Our implementation of the MM-GB/SA rescoring procedure (eq 1) has been described in detail elsewhere.[4] In eq 1, ΔE_{intra} and ΔG_{solv} are the intramolecular strain and desolvation penalty for each ligand upon binding. The conformational entropies (S_{conf}) in solution were computed from the probabilities (P_i) assuming a Boltzmann distribution (eq 2), where k_B is the Boltzmann constant. In the bound state, it was assumed that there was only one conformation accessible to each ligand; its conformational entropy is therefore zero. Thus, $-T\Delta S_{conf}$ is the ligand conformational entropy penalty, multiplied by the temperature to convert it into energy. E_{VDW} and E_{Elect} are the protein−ligand intermolecular van der Waals and electrostatic interaction energies, respectively. E_{PTN}, the protein energy in the bound state, describes the protein deformation or strain imposed by each ligand. E_{GB} is the solvent shielding of protein−ligand electrostatic interactions estimated by the GB model. This term, generally not included in the scoring as it has no significant impact on the MM-GB/SA results,[4] is necessary here as it depends on ε_{in}. Finally, ΔG^{ptn}_{solv} is the protein desolvation term calculated by the continuum model. The final ranking is obtained by calculating relative binding energies, $\Delta\Delta G_{bind}$, using the top-scoring inhibitor as a reference.

$$\Delta G_{bind} = \Delta E_{intra} + \Delta G_{solv} - T\Delta S_{conf} + E_{VDW} + E_{Elect} + E_{GB} + E_{PTN} + \Delta G^{ptn}_{solv} \quad (1)$$

$$S_{conf} = -k_B \sum_{i=1}^{n} P_i \ln P_i \quad (2)$$

The conformational search for the inhibitors in the unbound state and energy minimization for the complexes were performed with BOSS and MCPRO,[22] respectively, instead of Macromodel.[13] This was done with the purpose of facilitating the implementation of the variable dielectric approach. The **Z** matrices for the complexes obtained from the docking calculations, and ligands in the unbound state were prepared using the pepz program.[22] The proteins were considered in their entirety. The protonation states of histidine side chains were assigned with the assistance of the software PROPKA 2.0.[23] Charge neutrality for the protein systems in MCPRO was imposed by assigning normal protonation states at physiological pH to basic and acidic residues near the active site and making the adjustments for neutrality to the most distant residues. The OPLS-AA force field was used for the protein.[12a] The energetics for the ligands were represented with the OPLS/CM1A force field.[24] The CM1A atomic charges for the neutral ligands were scaled by 1.14.[25] The GB/SA solvation model implemented in BOSS and MCPRO was used.[26]

The Variable Dielectric Formalism and Implementation. The focus here is to obtain a set of dielectric constants based on residue types that improves MM-GB/SA scoring. Protein intramolecular electrostatic interactions were not considered in the optimization process as the set of optimal dielectric constants for each residue−residue pair is not necessarily the best that can be obtained to improve the description of ligand−residue electrostatic interactions. In addition, the E_{PTN} values within a congeneric series generally fall in a very narrow range due to the constraints applied in the energy minimizations for the complexes.[4] As this term behaves almost like a constant, it was excluded from the scoring equation used to develop the variable dielectric formalism. Although ΔG^{ptn}_{solv}, like E_{PTN}, depends on ε_{in}, it was also excluded from the scoring equation, as it is usually very noisy and deteriorates the correlation with experimental data.[4b,c] The energy minimizations for the complexes, however, are performed using GB/SA.

Equations 3 and 4 describe the protein−ligand electrostatic interactions and the solvent shielding estimated by the GB model, respectively, where q_i and q_j are charges on atom i belonging to the protein and j belonging to ligand, r_{ij} is their distance, ε_{in} is the internal dielectric constant, ε_{sol} is the dielectric constant in water, and α_{ij} is the geometric average of the Born radii α_i and α_j.

$$E_{elect} = \sum_{i,j} \frac{q_i q_j}{\varepsilon_{in} r_{ij}} \quad (3)$$

Table 1. Optimal Set of Residue-Based Dielectric Constants ($\varepsilon_{in(k)}$) for the Targets Individually and Best Overall

res type	CDK2		FactorXa		PDE10A		p38_u		optimal
Ser	20	1	1	1	20	1	20	20	1
Thr	20	2	20	2	1	2	1	1	4
Asn	1	1	20	1	20	8	1	4	2
Gln	1	1	1	20	4	2	1	1	1
His	20	1	20	20	20	1	1	20	4
Lys	2	2	20	1	20	20	1	1	2
Arg	1	1	1	1	2	1	20	20	20
Asp	1	8	20	20	2	20	20	20	2
Glu	20	20	1	1	1	1	20	20	8
Other	1	1	4	8	20	4	20	2	4
	$R^2 = 0.56$	PI = 0.76	$R^2 = 0.68$	PI = 0.90	$R^2 = 0.53$	PI = 0.71	$R^2 = 0.58$	PI = 0.75	

$$E_{GB} = -\left(\frac{1}{\varepsilon_{in}} - \frac{1}{\varepsilon_{sol}}\right) \sum_{i,j} \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_{ij}^2 \exp\left(\frac{r_{ij}^2}{4\alpha_{ij}^2}\right)}} \quad (4)$$

If ε_{in} adopts different values depending on the residue type k interacting with the ligand, eqs 3 and 4 have to be rewritten as shown in eqs 5 and 6. The dielectric constant $\varepsilon_{in(k)}$ is assigned for the interactions between all atoms i belonging to the residue type k and the ligand atoms j. The side chains for all polar (Ser, Thr, Asn, Gln) and ionizable residues (His, Lys, Arg, Asp, Glu), which are expected to be more polarizing, were considered individually in the optimization. All remaining residue side chains and backbone atoms were bundled in a group called "other". For simplicity, neutral and protonated states of the ionizable residues were not treated separately. The dielectric constant $\varepsilon_{in(k)}$ for a given residue−ligand pair is the same whether the residue interacts with a neutral or charged ligand.

$$E_{elect} = \sum_k \frac{1}{\varepsilon_{in(k)}} \sum_{i \in k,j} \frac{q_i q_j}{r_{ij}} \quad (5)$$

$$E_{GB} = -\sum_k \left(\frac{1}{\varepsilon_{in(k)}} - \frac{1}{\varepsilon_{sol}}\right) \sum_{i \in k,j} \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_{ij}^2 \exp\left(\frac{r_{ij}^2}{4\alpha_{ij}^2}\right)}} \quad (6)$$

Five different values (1, 2, 4, 8, and 20) were considered for every residue type (Ser, Thr, Asn, Gln, His, Lys, Arg, Asp, Glu, and other). Computational efficiency was achieved by obtaining the inner summation terms in eqs 5 and 6 for every residue k through one single point calculation per complex using the geometries obtained with $\varepsilon_{in} = 1$. No cutoffs were applied to the protein−ligand electrostatic interactions. The inner summation terms were then combined with 5^{10} dielectric constant permutations outside of MCPRO to generate the MM-GB/SA scores corresponding to each set of E_{elect} and E_{GB} values. Their performance against the experimental data was judged on the basis of correlation coefficients (R^2) and predictive indices (PI).[19] The latter is a measure of how accurate the predicted rank order is compared to the experiment, with −1, 0, and +1 meaning opposite, random, or perfect predictions, respectively. The set of dielectric constants that overall provided the best R^2 and PI values for the training set targets and ligands was selected and then tested on two additional data sets.

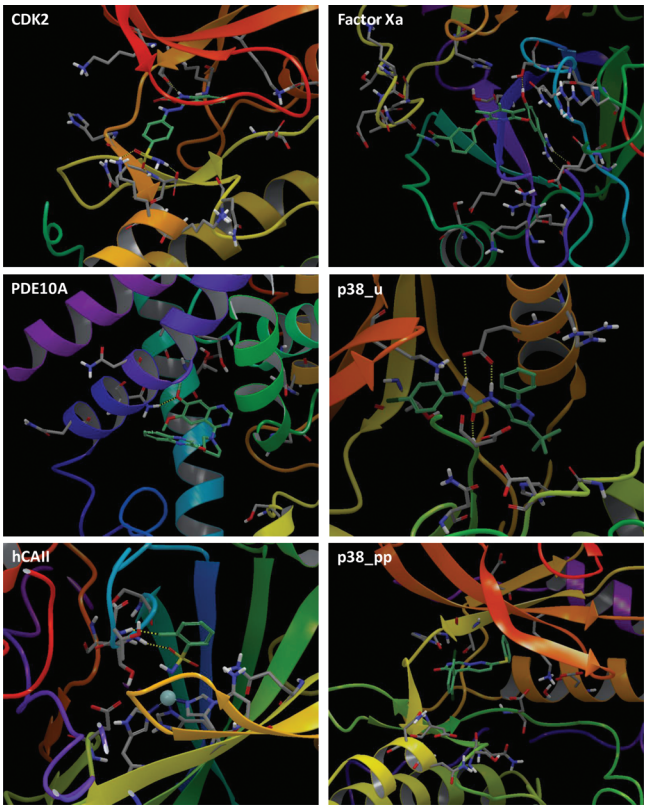

Figure 2. Ligands in their binding sites. Only the polar and ionizable residues within 6 Å from each ligand are shown for CDK2, Factor Xa, PDE10A, hCAII, and the p38 urea (p38_u) pyridazo-pyrimidinone (p38_pp) series.

RESULTS AND DISCUSSION

Residue-Based Dielectric Constant Optimization. In the optimization process, 5^{10} combinations of dielectric constants based on the residue type were generated for each system in the training set, CDK2, FactorXa, PEDE10A, and the p38 urea series (p38_u). The combinations that provided the MM-GB/SA scores with the best agreement with binding data based on R^2 and PI values for each target individually were set aside.

dx.doi.org/10.1021/ct200565u |J. Chem. Theory Comput. 2011, 7, 3859–3865

Table 2. MM-GB/SA Results Using Standard and Variable Dielectric Protein−Ligand Electrostatics

target	N^a	out^b	R^2 STD^c	R^2 NO^d	R^2 VAR^e	PI STD	PI NO	PI VAR	DR^f STD	DR NO	DR VAR
CDK2	36	7	0.48	0.41	0.52	0.69	0.66	0.73	17.6	13.7	13.1
FactorXa	22	3	0.59	0.09	0.66	0.84	0.37	0.87	14.8	6.3	9.3
PDE10A	28	0	0.47	0.52	0.51	0.65	0.69	0.67	8.3	6.7	7.1
p38_u	36	2	0.33	0.56	0.55	0.46	0.67	0.71	20.9	21.0	21.9
hCAII	17	2	0.32	0.39	0.58	0.63	0.74	0.75	11.9	10.9	7.8
p38_pp	17	3	0.41	0.49	0.43	0.68	0.70	0.70	10.5	9.4	10.5

[a] Number of total compounds in each series. [b] Number of outliers removed. [c] Standard dielectric protein−ligand electrostatics (STD). [d] No protein−ligand electrostatics. E_{elect} and E_{GB} are removed from the scoring equation (NO). [e] Variable dielectric protein−ligand electrostatics (VAR). [f] Score dynamic range (DR) in kcal/mol.

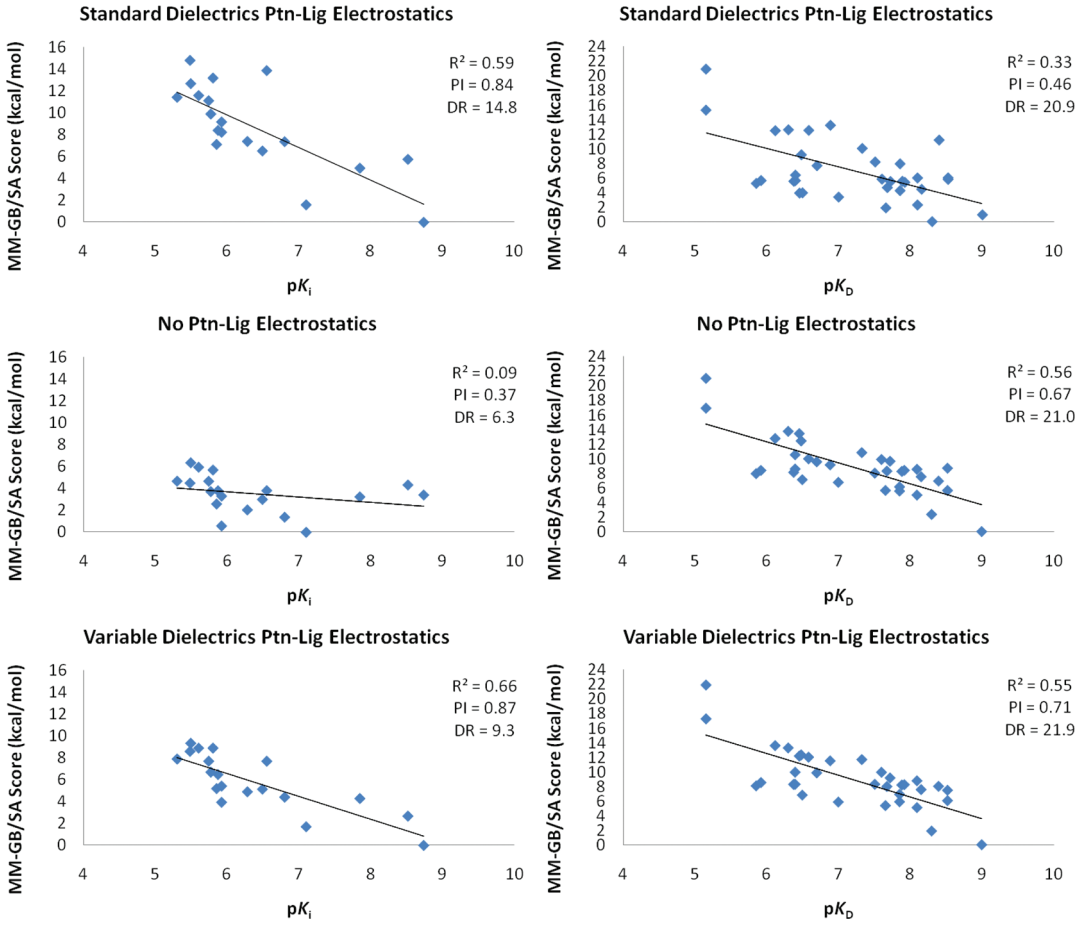

Figure 3. MM-GB/SA scoring versus experimental results using standard and variable dielectric protein−ligand electrostatics, and with the E_{elect} and E_{GB} terms removed for two systems, Factor Xa (left) and p38_u (right).

Table 1 illustrates the combinations for each target that maximized R^2 and PI separately. Although the combinations listed in Table 1 are the very best for each system according to each metric, there were few other solutions that provided just slightly worse results. The set of optimal dielectric constants for all targets simultaneously was derived by verifying the combination whose R^2 and PI values deviated the least from each individual's best (Table 1).

Figure 2 shows the polar and ionizable residues within 6 Å from the ligands depicted in Figure 1 for the systems that belong not only to the training set but also to the test set, hCAII and the p38 pyridazo-pyrimidinone (p38_pp) series. As electrostatic interactions are very long-range, it is obvious that polar and ionizable residues beyond 6 Å are also relevant; their interactions with the ligands are included in E_{elect} and E_{GB}. Although the optimal set of dielectric constants is a function of the specific residue−ligand interactions in the training set and inaccuracies in the charges of the force field of choice, it is tempting to physically interpret the $\varepsilon_{in(k)}$ values obtained. Here, Arg and Glu have the largest $\varepsilon_{in(k)}$ values, 20 and 8, respectively, which seems reasonable since those residues, very flexible and polarizing, can be more easily shielded. The optimal value for Asp, with a shorter side chain and reduced flexibility when compared to Glu, is 2.

Interestingly, a much smaller dielectric constant was obtained for Lys in contrast to the one for Arg. In a study of residue density in proteins,[27] it became evident that in spite of evolutionary relatedness, Arg is more buried and more frequently involved in salt bridges, hydrogen bonds, and cationic−aromatic contacts. It is then plausible that a larger $\varepsilon_{in(k)}$ value than that for Lys emerges for Arg in the optimization process. In the case of the polar and the ionizable His residues, found to be in its neutral state in the majority of the cases here, the $\varepsilon_{in(k)}$ values are fairly small, ranging from 1 to 4. For three of the five residues, the dielectric constants seem to be correlated with the magnitude of the side chain dipole moments, with the $\varepsilon_{in(k)}$ value for His > Asn > Ser. As for Gln and Thr, the dielectric constants obtained seem to be counterintuitive when compared to their closest analogs, Asn and Ser, respectively. The optimal $\varepsilon_{in(k)}$ value for the more flexible side chain of the pair is actually smaller; this might be a function of the specific residue distributions for the targets in the training set. Finally, the $\varepsilon_{in(k)}$ value of 4 for the set that contains all remaining side chains might be somewhat high because it also contains all backbone atoms.

Standard versus Variable Dielectric MM-GB/SA Scoring. It should be noted that the variable dielectric approach was unable to recover some outliers in the training set in the preliminary rounds of dielectric constant optimization. Those compounds had to be ultimately excluded before the optimization round that led to the $\varepsilon_{in(k)}$ values in Table 1. The number of outliers removed, in both the training and test sets, is listed in Table 2. Except for CDK2, the outliers represent between 0% and 15% of the compounds in each series and are highlighted in the tables provided in the Supporting Information. The balance between the number of remaining binding data points in the training set (110) and the number of adjustable $\varepsilon_{in(k)}$ values (10), a 11:1 ratio, seems appropriate and suggests that the set of dielectric constants obtained might be applied more widely. Table 2 illustrates the R^2 and PI values obtained using the standard and variable dielectric MM-GB/SA scoring procedures, the latter using the optimal $\varepsilon_{in(k)}$ values in Table 1. The results where E_{elect} and E_{GB} are removed from the scoring equation are also included for comparison. They provide a baseline for the cases where the standard dielectric MM-GB/SA performs poorly. In those cases, the variable dielectric approach should at least outperform the scoring equation with E_{elect} and E_{GB} removed.

It is disturbing that the scoring equation with no protein−ligand electrostatics performs better than the standard MM-GB/SA in four out of the six systems (Table 2). Factor Xa, where E_{elect} and E_{GB} are critical, and CDK2, where they marginally improve the results, are the only exceptions. In Factor Xa, the most potent compounds of the series establish a key hydrogen bond with the catalytic triad Ser residue (Figure 2), which cannot be described without the electrostatic terms. Although it is difficult to pinpoint whether the deterioration of the correlation with experimental data when E_{elect} and E_{GB} are introduced has its origin in the force field charges, the lack of explicit polarization, or the lack of dynamical screening due to the use of a single structure for the complex, it is clear that the variable dielectric approach improves the description of protein−ligand electrostatics. It performs better than standard dielectrics in all cases, including the two systems in the test set, hCAII and p38_pp. Figure 3 illustrates the results for Factor Xa and p38_u using the different MM-GB/SA scoring approaches. Plots for all systems can be found in the Supporting Information.

It is also reassuring that the performance of the variable dielectric approach is at least equivalent or superior to the scoring equation with E_{elect} and E_{GB} removed. A scoring method that describes protein−ligand electrostatic interactions properly as well as the fine balance with the desolvation penalty process is highly desirable since the introduction of polarity, which drives the compound to a better property space,[28] often kills or attenuates binding affinity. One should note that the variable dielectric approach will not necessarily provide significant improvements over the standard electrostatic treatment for all cases. This is illustrated in Table 2 for a couple of systems, PDE10A and p38_pp. In those instances, the residues around the ligand are mostly nonpolar, and the electrostatic interactions between them are not appreciably large (Figure 2). Specifically, the combined E_{elect} and E_{GB} terms for all ligands in the congeneric series obtained with ε_{in} = 1 range from −1.9 to −0.2 kcal/mol and −3.6 to 0.0 kcal/mol for PDE10A and p38_pp, respectively. In other words, protein−ligand electrostatic interactions play a minor role in the rank-ordering for the PDE10A and p38_pp, and the attenuation or exclusion of E_{elect} and E_{GB} has no impact on the results. This contrasts with p38_u, for example, with a much wider range (−11.8 to +1.9 kcal/mol) for the combined E_{elect} and E_{GB} terms. The protein conformation (DFG out) for p38, the chemical series, and binding mode are very different, with the urea group establishing hydrogen bonds with a Glu residue and a backbone NH (Figure 2). In this case, the variable dielectric approach greatly helps since E_{elect} and E_{GB} are relevant. Finally, Table 2 indicates that the variable dielectric approach reduces the score dynamic range (DR) in four of the six systems, although no attempt has been made to improve DR in the optimization process. This supports the view that the exaggerated protein−ligand electrostatic interactions due to the lack of shielding effects in the standard MM-GB/SA model are indeed a key factor in its wide scoring spread.

■ ASSOCIATED CONTENT

Supporting Information. Tables with all derivatives and binding data and plots of the multiple MM-GB/SA approaches against the experiment for all systems are available. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*Phone: (860) 686-2915. E-mail: cristiano.guimaraes@pfizer.com.

■ ACKNOWLEDGMENT

The authors thank the Computational Sciences Center of Emphasis at Pfizer and the National Institutes of Health (GM32136 and AI44616) for support.

■ REFERENCES

(1) Barril, X.; Gelpí, J. L.; López, J. M.; Orozco, M.; Luque, F. J. How accurate can molecular dynamics/linear response and Poisson−Boltzmann/solvent accessible surface calculations be for predicting relative binding affinities? Acetylcholinesterase huprine inhibitors as a test case. *Theor. Chem. Acc.* **2001**, *106*, 2–9.

(2) Kuhn, B.; Kollman, P. A. Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their

relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.* **2000**, *43*, 3786–3791.

(3) (a) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* **2011**, *51*, 69–82. (b) Haider, M. K.; Bertrand, H.-O.; Hubbard, R. E. Predicting fragment binding poses using a combined MCSS MM-GBSA approach. *J. Chem. Inf. Model.* **2011**, *51*, 1092–1105. (c) Lee, M. R.; Sun, Y. Improving docking accuracy through molecular mechanics generalized Born optimization and scoring. *J. Chem. Theory Comput.* **2007**, *3*, 1106–1119. (d) Foloppe, N.; Hubbard, R. Towards predictive ligand design with free-energy based computational methods?. *Curr. Med. Chem.* **2006**, *13*, 3583–3608. (e) Huang, N.; Kalyanaraman, C.; Irwin, J. J.; Jacobson, M. P. Physics-based scoring of protein-ligand complexes: Enrichment of known inhibitors in large-scale virtual screening. *J. Chem. Inf. Model.* **2006**, *46*, 243–253. (f) Huang, N.; Kalyanaraman, C.; Bernacki, K.; Jacobson, M. P. Molecular mechanics methods for predicting protein−ligand binding. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5166–5177. (g) Lyne, P. D.; Lamb, M. L.; Saeh, J. C. Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem.* **2006**, *49*, 4805–4808.

(4) (a) Guimarães, C. R. W. A direct comparison of the MM-GB/SA scoring procedure and free-energy perturbation calculations using Carbonic Anhydrase as a test case: Strengths and pitfalls of each approach. *J. Chem. Theory Comput.* **2011**, *7*, 2296–2306. (b) Guimarães, C. R. W.; Mathiowetz, A. M. Addressing limitations with the MM-GB/SA scoring procedure using the WaterMap method and free-energy perturbation calculations. *J. Chem. Inf. Model.* **2010**, *50*, 547–559. (c) Guimarães, C. R. W.; Cardozo, M. MM-GB/SA rescoring of docking poses in structure-based lead optimization. *J. Chem. Inf. Model.* **2008**, *48*, 958–970.

(5) (a) Guimarães, C. R. W.; Boger, D. L; Jorgensen, W. L. Elucidation of fatty acid amide hydrolase inhibition by potent α-ketoheterocycle derivatives from Monte Carlo simulations. *J. Am. Chem. Soc.* **2005**, *127*, 17377–17384. (b) Simonson, T.; Georgios, A.; Karplus, M. Free energy simulations come of age: Protein-ligand recognition. *Acc. Chem. Res.* **2002**, *35*, 430–437.(c) Jorgensen, W. L. Free energy changes in solution. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; Wiley: New York, 1998; Vol. 2, pp 1061−1070. (d) Kollman, P. A. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417. (e) Jorgensen, W. L. Free energy calculations: A breakthrough for modeling organic chemistry in solution. *Acc. Chem. Res.* **1989**, *22*, 184–189.

(6) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the active-site solvent in the thermodynamics of Factor Xa ligand binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817–2831.

(7) Abel, R.; Salam, N. K.; Shelley, J.; Farid, R.; Friesner, R. A.; Sherman, W. Contribution of explicit solvent effects to the binding affinity of small-molecule inhibitors in blood coagulation factor serine proteases. *ChemMedChem* **2011**, *6*, 1049–1066.

(8) Schutz, C. N.; Warshel, A. What are the dielectric "constants" of proteins and how to validate electrostatic models?. *Proteins* **2001**, *44*, 400–417.

(9) Sham, Y. Y.; Muegge, I.; Warshel, A. The effect of protein relaxation on charge− charge interactions and dielectric constants of proteins. *Biophys. J.* **1998**, *74*, 1744–1753.

(10) Muegge, I.; Schweins, T.; Langen, R.; Warshel, A. Electrostatic control of gtp and gdp binding in the oncoprotein p21 ras. *Structure* **1996**, *4*, 475–489.

(11) Zhu, K.; Shirts, M. R.; Friesner, R. A. Improved methods for side chain and loop predictions via the protein local optimization program: Variable dielectric model for implicitly improving the treatment of polarization effects. *J. Chem. Theory Comput.* **2007**, *3*, 2108–2119.

(12) (a) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236. (b) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. J. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(13) *MacroModel*, version 9.0; Schrödinger, LLC: New York, 2005.

(14) (a) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749. (b) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision Glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.

(15) Bramson, H. N.; Corona, J.; Davis, S. T.; Dickerson, S. H.; Edelstein, M.; Frye, S. V.; Gampe, R. T.; Harris, J. P. A.; Hassell, A.; Holmes, W. D.; Hunter, R. N.; Lackey, K. E.; Lovejoy, B.; Luzzio, M. J.; Montana, V.; Rocque, W. J.; Rusnak, D.; Shewchuk, L.; Veal, J. M.; Walker, D. H.; Kuyper, L. F. Oxindole-based inhibitors of Cyclin-Dependent Kinase 2 (CDK2): Design, synthesis, enzymatic activities, and X-ray crystallographic analysis. *J. Med. Chem.* **2001**, *44*, 4339–4358.

(16) (a) Phillips, G.; Davey, D. D.; Eagen, K. A.; Koovakkat, S. K.; Liang, A.; Ng, H. P.; Pinkerton, M.; Trinh, L.; Whitlow, M.; Beatty, A. M.; Morrissey, M. M. Design, synthesis, and activity of 2,6-diphenoxypyridine-derived Factor Xa inhibitors. *J. Med. Chem.* **1999**, *42*, 1749–1756. (b) Phillips, G.; Guilford, W. J.; Buckman, B. O.; Davey, D. D.; Eagen, K. A.; Koovakkat, S.; Liang, A.; McCarrick, M.; Mohan, R.; Ng, H. P.; Pinkerton, M.; Subramanyam, B.; Ho, E.; Trinh, L.; Whitlow, M.; Wu, S.; Xu, W.; Morrissey, M. M. Design, synthesis, and activity of a novel series of Factor Xa inhibitors: Optimization of arylamidine groups. *J. Med. Chem.* **2002**, *45*, 2484–2493.

(17) (a) Verhoest, P. R.; Chapin, D. S.; Corman, M.; Fonseca, K.; Harms, J. F.; Hou, X.; Marr, E. S.; Menniti, F. S.; Nelson, F.; O'Connor, R.; Pandit, J.; Proulx-LaFrance, C.; Schmidt, A. W.; Schmidt, C. J.; Suiciak, J. A.; Liras, S. Discovery of a novel class of Phosphodiesterase 10A inhibitors and identification of clinical candidate 2-[4-(1-Methyl-4-pyridin-4-yl-1H-pyrazol-3-yl)-phenoxymethyl]-quinoline (PF-2545920) for the treatment of schizophrenia. *J. Med. Chem.* **2009**, *52*, 5188–5196. (b) Chappie, T. A.; Humphrey, J. M.; Allen, M. P.; Estep, K. G.; Fox, C. B.; Lebel, L. A.; Liras, S.; Marr, E. S.; Menniti, F. S.; Pandit, J.; Schmidt, C. J.; Tu, M.; Williams, R. D.; Yang, F. Y. Discovery of a series of 6,7-dimethoxy-4-pyrrolidylquinazoline PDE10A inhibitors. *J. Med. Chem.* **2007**, *50*, 182–185.

(18) Regan, J.; Breitfelder, S.; Cirillo, P.; Gilmore, T.; Graham, A. G.; Hickey, E.; Klaus, B.; Madwed, J.; Moriak, M.; Moss, N.; Pargellis, C.; Pav, S.; Proto, A; Swinamer, A.; Tong, L.; Torcellini, C. Pyrazole urea-based inhibitors of p38 MAP kinase: From lead compound to clinical candidate. *J. Med. Chem.* **2002**, *45*, 2994–3008.

(19) Pearlman, D. A.; Charifson, P. S. Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system. *J. Med. Chem.* **2001**, *44*, 3417–3423.

(20) Scott, A. D.; Phillips, C.; Alex, A.; Flocco, M.; Bent, A.; Randall, A.; O'Brien, R.; Damian, L.; Jones, L. H. Thermodynamic optimization in drug discovery: A case study using Carbonic Anhydrase inhibitors. *Chem. Med. Chem.* **2009**, *4*, 1985–1989.

(21) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(22) Jorgensen, W. L.; Tirado-Rives, J. Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J. Comput. Chem.* **2005**, *26*, 1689–1700.

(23) Bas, D. C.; Rogers, D. M.; Jensen, J. H. Very fast prediction and rationalization of pK_a values for protein−ligand complexes. *Proteins* **2008**, *73*, 765–783.

(24) Jorgensen, W. L.; Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6665–6670.

(25) Udier-Blagovic, M.; Morales De Tirado, P.; Pearlman, S. A.; Jorgensen, W. L. Accuracy of free energies of hydration from CM1 and CM3 atomic charges. *J. Comput. Chem.* **2004**, *25*, 1322–1332.

(26) Jorgensen, W. L.; Ulmschneider, J. P.; Tirado-Rives, J. Free energies of hydration from a Generalized Born model and an all-atom force field. *J. Phys. Chem. B* **2004**, *108*, 16264–16270.

(27) Baud, F.; Karlin, S. Measures of residue density in protein structures. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 12494–12499.

(28) (a) Hughes, J. D.; Blagg, J.; Price, D. A.; Bailey, S.; DeCrescenzo, G. A.; Devraj, R.; Ellsworth, E.; Fobian, Y. M.; Gibbs, M. E.; Gilles, R. W.; Greene, N.; Huang, E.; Krieger-Burke, T.; Loesel, J.; Wager, T.; Whiteley, L.; Zhang, Y. Physiochemical drug properties associated with in vivo toxicological outcomes. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4872–4875. (b) Wager, T. T.; Hou, X.; Verhoest, P. R.; Villalobos, A. Moving beyond rules: The development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties. *ACS Chem. Neurosci.* **2010**, *1*, 420–434.