JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

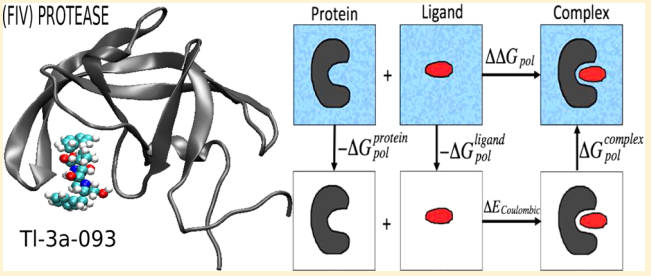
Protein−Ligand Electrostatic Binding Free Energies from Explicit and Implicit Solvation

Saeed Izadi, Boris Aguilar, and Alexey V. Onufriev*

Department of Biomedical Engineering and Mechanics, Department of Computer Science, and Departments of Computer Science and Physics, Virginia Tech, Blacksburg, Virginia 24060, United States

Supporting Information

ABSTRACT: Accurate yet efficient computational models of solvent environment are central for most calculations that rely on atomistic modeling, such as prediction of protein−ligand binding affinities. In this study, we evaluate the accuracy of a recently developed generalized Born implicit solvent model, GBNSR6 (Aguilar et al. *J. Chem. Theory Comput.* **2010**, *6*, 3613−3639), in estimating the electrostatic solvation free energies (ΔG_{pol}) and binding free energies ($\Delta\Delta G_{pol}$) for small protein−ligand complexes. We also compare estimates based on three different explicit solvent models (TIP3P, TIP4PEw, and OPC). The two main findings are as follows. First, the deviation (RMSD = 7.04 kcal/mol) of GBNSR6 binding affinities from commonly used TIP3P reference values is comparable to the deviations between explicit models themselves, e.g. TIP4PEw vs TIP3P (RMSD = 5.30 kcal/mol). A simple uniform adjustment of the atomic radii by a single scaling factor reduces the RMS deviation of GBNSR6 from TIP3P to within the above "error margin" − differences between $\Delta\Delta G_{pol}$ estimated by different common explicit solvent models. The simple radii scaling virtually eliminates the systematic deviation ($\Delta\Delta G_{pol}$) between GBNSR6 and two out of the three explicit water models and significantly reduces the deviation from the third explicit model. Second, the differences between electrostatic binding energy estimates from different explicit models is disturbingly large; for example, the deviation between TIP4PEw and TIP3P estimates of $\Delta\Delta G_{pol}$ values can be up to ∼50% or ∼9 kcal/mol, which is significantly larger than the "chemical accuracy" goal of ∼1 kcal/mol. The absolute ΔG_{pol} calculated with different explicit models could differ by tens of kcal/mol. These discrepancies point to unacceptably high sensitivity of binding affinity estimates to the choice of common explicit water models. The absence of a clear "gold standard" among these models strengthens the case for the use of accurate implicit solvation models for binding energetics, which may be orders of magnitude faster.



1. INTRODUCTION

An accurate yet efficient determination of the electrostatics in protein−ligand interactions is of profound importance in molecular design and drug discovery.[1−6] The computational prediction of binding free energies is however complex and challenging,[7,8] and its outcomes can depend strongly on the molecular modeling technique used.[9] In particular, accuracy of solvation and binding free energies calculations depends critically on the quality of the underlying solvent model.[10−12]

Extensive studies have been performed to evaluate the accuracy of solvent models in predicting solvation free energies for small molecular systems;[10−13] in many cases the desirable "chemical accuracy" of 1 kcal/mol was reported,[10,11,14−16] at least on average. Yet, the high accuracy in predicting individual solvation free energies does not necessarily translate into high accuracy in atomistic binding free energy calculations.[17−19] High accuracy and robustness of the force fields and solvent models in these calculations has proven difficult to achieve.[20−25] Errors in even some of the highly accurate calculated binding energies can represent a significant percentage of the target binding free energies, sometimes as large as ∼50% in relative error or ∼7 kcal/mol in absolute error,[26] in particular when the number of interactions in a protein−ligand complex increases.[19] Even for small and relatively rigid host−guest systems, predicting binding affinities within typical chemical accuracy of ∼1 kcal/mol remains elusive.[22,23,27] Perhaps not surprisingly, binding free energies can be sensitive to the method parameters: the relatively small binding affinities are the difference between large free energy terms corresponding to the bound and unbound states.[17] The interactions between proteins and ligands are short ranged and strong, which leads to the strong dependence of energy functions on the details of molecular conformation.[22] Effects of systematic error cancellation can also be consequential in binding free energy calculations.[17,18] Since long-range electrostatic interactions play a dominant role in biomolecular simulations, a careful treatment of electrostatic interactions is essential[28−34] and merits special attention: this is the focus of the current work.

Implicit solvent models are currently routinely employed for evaluation of electrostatic interactions in many scenarios of

Received: May 22, 2015

J. Chem. Theory Comput.
Downloaded from pubs.acs.org by UNIV OF PENNSYLVANIA on 08/22/15. For personal use only.

© XXXX American Chemical Society

DOI: 10.1021/acs.jctc.5b00483
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

J. Chem. Theory Comput.
Downloaded from pubs.acs.org by UNIV OF PENNSYLVANIA on 08/22/15. For personal use only.

biomolecular modeling.[35−52] By replacing discrete water models with a continuum medium using the average dielectric properties of water, implicit solvents provide significant decrease in the computational cost of simulations. Within the implicit solvent framework, the Generalized Born (GB) model[43,53−75] provides a relatively simple, efficient, and robust estimate to calculate the long-range electrostatic interactions in molecular simulations.[43,76] Methods based on the implicit solvent framework, such as MMPB(GB)/SA,[77,78] are extensively used in estimates of solvation free energies and protein−ligand binding interactions.[31,37,41,62,79]

Recently, a new flavor of GB model, the so-called GBNSR6,[80] was reported; unlike most of its predecessors, the model relies on the "R6"[81−84] effective Born radius, which is calculated as a single $r/|r|^6$ integral over the Lee−Richards molecular surface.[80] A good agreement of the electrostatic component of the solvation free energies (ΔG_{pol}) by the R6 flavor compared to the more fundamental Poisson−Boltzmann (PB) model for small proteins and DNA was previously reported.[84] It was also shown that GBNSR6 and the TIP3P explicit model are in close agreement for different conformations of alanine polypeptide.[80] In a recent study, it was shown that GBNSR6 (with an appropriate nonpolar contribution added) reproduces experimentally measured solvation free energies of small molecules with near "chemical" accuracy,[13] on average.

Given the high promise of the GBNSR6 method in predicting solvation free energies of some molecular systems, here we evaluate the accuracy of the model in predicting protein−ligand binding energies, crucial for rational drug design. To address this question, in this study we evaluate the binding free energies from GBNSR6 for a set of 15 small protein−ligand complexes, using explicit solvent free energies as reference.

While using explicit solvent as an accuracy reference for an implicit solvent model is natural, the question arises which of the great many[85] available explicit water models should be used. The question is nontrivial, as none of the current models can be considered as the uncontested "gold standard".[86−88] Explicit water models are built to reproduce bulk properties, but, being imperfect,[86−88] improved performance in pure water properties does not necessarily translate into better performance in solvated systems. For example, TIP4PEw is more accurate than TIP3P in predicting water bulk properties[88] but less accurate in predicting hydration energies of small molecules;[11] TIP5P is superior to both in reproducing details of water structure[89] but trails behind TIP4PEw in accuracy of predicted small molecule hydration energies.[11] Besides, transferability of the accuracy of these explicit models in predicting solvation free energies of small molecules (RMS errors slightly over 1 kcal/mol) to macromolecular systems is not guaranteed,[19] nor is it certain whether the same level of accuracy is achievable in binding free energy calculations, which is of main interest to us here. The question would be moot, however, if commonly used water models showed consistent performance in these calculations, say within 1 kcal/mol of each other. Whether or not common (and some new) explicit water models are equivalent in this respect is the second main question we address in this work. For this purpose we compare ΔG_{pol} and $\Delta\Delta G_{pol}$ computed with two highly popular, fixed-charge, rigid explicit water models for which free energy calculation protocols are well-established and their computational expense is reasonable: TIP3P[91] and TIP4PEw.[92] We also make a comparison with a recently developed 4-point rigid explicit water model, OPC,[93] which is arguably the first model of this class that predicts hydration free energies of small molecules with RMSD accuracy of less than 1 kcal/mol.[93]

The remainder of the paper is organized as follows. In Section 2 we present the specifics of the protein−ligand complexes used for the comparative studies. The details of implicit solvent calculations and explicit solvent calculations using Thermodynamics Integration are provided in Section 2. The comparative studies are presented in Section 3. A summary of our findings is discussed in Section 4.

2. METHODS

2.1. Preparation of Complexes. A set of 15 protein−ligand complexes was selected (Table 1). One feature of the selected complexes is their small size (~1635−1995 atoms), essential to ensure convergence of the (lengthy) free energy perturbation (FEP) estimates. As a result of the limitation on the structure size, the diversity of the set is limited in terms of biological function of the complexes. However, the set is diverse with respect to values of electrostatic binding free energies: it covers a wide range of those (see Section 3), which is sufficient for our purpose. Another feature of the collected set is that the ligands are neutral and the proteins are either neutral or are forced to be neutral. The neutralization is performed to avoid various uncertainties and complications[94] due to the use of Ewald summation and periodic boundary conditions in explicit solvent simulations. For each component that needed to be neutralized, the neutralization was performed as follows. First, its isoelectric point *pI* was computed. Then, protonation state and charge states of each titratable group was set according to its computed *pK* value at *pH* = *pI*, which forced overall neutrality of the structure. The calculations of *pK*, *pI*, the titration curves, and the protonation state adjustments were

Table 1. Specifics of Protein and Ligand Components for the Set of 15 Small Complexes Studied Here

PDB ID	protein name	no. of atoms in protein	ligand name	no. of atoms in ligand
1b11	feline immunodeficiency virus protease	1824	TL-3-093	66
1bkf	FK506 binding protein FKBP mutant R42K/H87V	1659	FK506	128
1f40	FKBP12	1662	GPI-1046	54
1fb7	HIV-1 protease mutant	1566	Saquinavir	99
1fkb	human immunophilin FKBP-12	1662	Rapamycin	144
1fkf	immunophilin FKBP	1662	FK506	126
1fkg	FKBP	1662	SB3	68
1fkh	FKBP	1662	SBX	74
1fkj	FKBP12	1662	FK506	128
1fkl	FKBP12	1661	Rapaymycin	146
1pbk	FKBP25	1851	RAP	144
1zp8	HIV protease	1566	inhibitor AB-2	88
2fke	FK-506-binding protein	1662	8-deethyl-8-[but-3-enyl]-ascomycin	126
3kfp	HIV protease	1569	inhibitor TL-3	66
2hah	FIV/HIV chimeric protease	1800	broad-based inhibitor, TL 3	66

DOI: 10.1021/acs.jctc.5b00483
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

J. Chem. Theory Comput.
Downloaded from pubs.acs.org by UNIV OF PENNSYLVANIA on 08/22/15. For personal use only.

performed using the H++ server[95] with the default settings; the server employs a continuum electrostatic approach to *pK* prediction. In principle, the explicit solvent box could alternatively be neutralized by adding counterions, however we did not follow this approach due to notably slow convergence of counterion distributions in MD simulations[96] (tens of nanoseconds for monovalent ions), which would make our TI-based estimates of ΔG_{pol} prohibitively expensive here.

For setting up the structures we used the H++[95] server that creates topology and coordinates files in Amber[97] format. The ff99bsc0 parameters and the GAFF force field, both part of Amber12,[98] were used for preparing the topology and coordinate files which include partial charges. We performed 500 steps of minimization on the neutral complexes (without restraint) in vacuum to relax the structure. The minimization was performed in SANDER molecular dynamics module of Amber with a 12 Å cutoff distance. After minimization, the complex structures were broken down to the protein and ligand components to be used for the binding free energy calculations. The Amber format topology and coordinate files as well as the corresponding ΔG_{pol} values are available in the Supporting Information.

2.2. Implicit Solvent Details. *GBNSR6.* GBNSR6 is an implementation of the Generalized Born (GB) model in which the effective Born radii are computed numerically, via the so-called "R6" integration,[13,80] over the Lee−Richards molecular surface.[99] The polar component of the solvation energy, ΔG_{pol}, is calculated by the ALPB model,[100] which introduces physically correct dependence on dielectric constants into the original GB model of Still et al.[54] while maintaining the efficiency of the original

$$\Delta G_{el} \approx -\frac{1}{2}\left(\frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{out}}\right)\frac{1}{1+\beta\alpha}\sum_{ij} q_i q_j\left(\frac{1}{f^{GB}} + \frac{\alpha\beta}{A}\right) \quad (1)$$

where ϵ_{in} and ϵ_{out} are the dielectric constants of the solute and the solvent respectively, $\beta = \epsilon_{in}/\epsilon_{out}$, $\alpha = 0.571412$, and A is the electrostatic size of the molecule, which is essentially the overall size of the structure, which can be computed analytically. Here, q_i is the partial charge of atom i. The most widely used functional form[76] of $f^{GB} = [r_{ij}^2 + R_i R_j \exp(-r_{ij}^2/R_i R_j)]^{1/2}$ is employed, where R_i is the effective Born radius of atom i, and r_{ij} is the distance between atoms i and j. We set $\epsilon_{in} = 1$ and $\epsilon_{out} = 80$ in eq 1. Note that to mitigate uncertainties related to conformational sampling[23,101] and to facilitate direct comparison between implicit with explicit solvent model predictions, we eliminated structural fluctuations by keeping all of the structures fixed by strong coordinate restraints in all of the explicit solvent simulations performed in this study (see Section 2.3). As a result, there was no dielectric response from the protein. This scenario is consistent with a value of unity for the solute dielectric constant ($\epsilon_{in} = 1$) in the corresponding implicit solvent modeling[62,102] eq 1. The assignment of solute dielectric constant can, however, be different for a direct comparison to experiment.

The effective Born radii R_i are calculated via

$$R_i^{-3} = \left(-\frac{1}{4\pi}\oint_{\partial V}\frac{\mathbf{r}-\mathbf{r}_i}{|\mathbf{r}-\mathbf{r}_i|^6}\cdot d\mathbf{S}\right) \quad (2)$$

where ∂V represents the molecular surface of the molecule, $d\mathbf{S}$ is the infinitesimal surface element vector, r_i is the position of atom i, and r represents the position of the infinitesimal surface element. In contrast to most GB practical models, the GBNSR6 model is essentially parameter-free in the same sense as the numerical PB framework is. Thus, accuracy of GBNSR6 relative to the PB standard is unaffected by the choice of input atomic radii. Here we use the simple, standard Bondi[103] radii set to determine the surface of the molecule. The solvent probe radius is equal to 1.4 Å. We use the same constant offset $B = 0.028$ Å^{-1} to the inverse radii as in Mongan et al.[84]

The GBNSR6 model exploits the Cartesian grid developed previously for the PBSA module of Amber,[104] to build a numerical discretization of the Lee−Richards molecular surface.[99] The spacing between two neighboring grid points is uniformly set to h = 0.3 Å, for the molecular surface resolution. The arc resolution (*arcres*), defined as the arc length between two neighboring solvent probe sites as the probe rolls over the atoms,[104] is set to 0.1 Å. This implementation of GBNSR6 is currently available as a part of Amber Tools suite of programs in Amber 2015.[105]

PB Model. The Adaptive Poisson−Boltzmann Solver (APBS) software package[106] was used for evaluating the polar part of solvation energies. The solute dielectric constant was set to 1, and the solvent dielectric constant was 80, which are consistent with the values chosen for GBNSR6. The grid spacing was set to 0.3 Å. To set the dimensions of the grids, we keep a distance equal to the size of the structure between the protein boundary and the grid boundary for the largest structure. Accordingly, the grid dimension size was set to 449 in x, y, and z directions for all of the structures. The solvent probe radius is 1.4 Å. We use APBS default values for the remaining parameters and assume no monovalent salt present, as in both the GB and the explicit solvent calculations.

2.3. Explicit Solvent Calculations. Electrostatic components ΔG_{pol} of explicit solvation free energies are computed by using the Thermodynamic Integration (TI) method of the SANDER module in Amber12.[98] Here we only compute the free energy transformations where the charges on protein−ligand complexes are removed: state 0 represents all solute atomic charges "on", and state 1 represents all solute atomic charges "off". We have used 5 values of lambda for TI calculations, λ = 0.04691, 0.23076, 0.50000, 0.76923, 0.95308. The TI values were obtained from Gaussian integration over the λ values. TI calculations were performed in water (TIP3P, TIP4PEw, or OPC explicit model) and in vacuum. Then the corresponding free energy values were subtracted to cancel out the intrasolute charge interactions as well as the restraint energies.[107] In all the simulations, the bonds to hydrogen atoms were constrained with the SHAKE algorithm using a geometrical tolerance of 0.000001 Å. The nonbonded interaction cutoff was 9 Å for simulations in water and 99 Å (effectively infinite) for simulations in vacuum. A time step of 2 fs is used (reduced to 1.8 fs if numerical instability was encountered). The following process is performed for each value of lambda: first, we run 1000 steps of minimization using the steepest decent method. Then, we run 30 ps of NVT ensemble by gradually increasing the temperature from 0 to 300 K. Then we run 1 ns of NPT ensemble at 300 K, for density equilibration. For the production we run 2 ns of NVT ensemble at 300 K. We run 200 ps MD simulation in vacuum. Protein−ligand complexes have many degrees of freedom that makes exploring all potentially relevant conformations computationally intractable.[23,101] Thus, 200 kcal/mol/Å^2 harmonic Cartesian coordinate restraints were imposed to all atoms, except during the

DOI: 10.1021/acs.jctc.5b00483
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

J. Chem. Theory Comput.
Downloaded from pubs.acs.org by UNIV OF PENNSYLVANIA on 08/22/15. For personal use only.

minimization step for which 500 kcal/mol/Å^2 harmonic Cartesian coordinate restraints were applied to all atoms.

Robustness of the Protocol and Error Estimate. The standard deviation of computed ΔG_{pol} values is smaller than ±0.7 kcal/mol for complexes and protein components and smaller than ±0.14 kcal/mol for ligands components. The standard deviations are calculated by assuming a correlation time of 1 ps, which is a conservative assumption considering that it is usually smaller than 0.8 ps.[108] The analysis is performed on the last 1.5 ns of 2 ns long simulation to ensure convergence. To test sensitivity to initial conditions we repeated the calculations for two complexes using different random seeds for the random number generator and obtained differences less than the average standard deviation of computed ΔG_{pol} values given above. To further test that the TI results have reached convergence, we extended the simulation time from 2 to 5 ns for two randomly selected complexes and noticed the resulting TI values differ by less than the standard deviation above.

2.4. The Electrostatic Component of Binding Free Energies,$\Delta\Delta G_{pol}$. The electrostatic component of the binding free energies can be calculated using the thermodynamics cycle shown in Figure 1. The first step is to transfer the individual

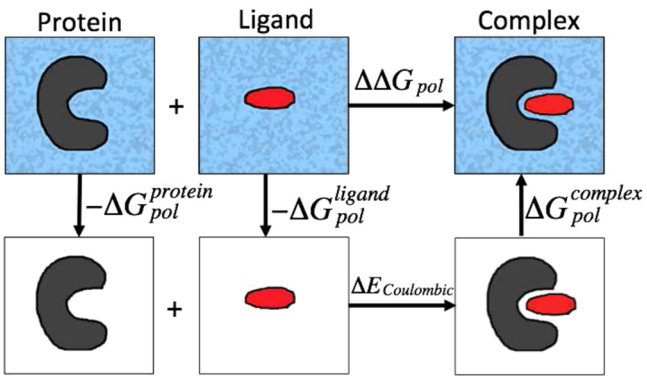


Figure 1. Illustration of the thermodynamic cycle for the decomposition of the electrostatic component of the binding free energies. The surrounding dielectric medium is shaded for water and is white for vacuum.

protein and ligand from the solvent into vacuum, with the energy costs of $-\Delta G_{pol}^{protein}$ and $-\Delta G_{pol}^{ligand}$, respectively. The second step is to combine the protein and the ligand into a complex in vacuum. The energy cost would be the difference in Coulombic energies in vacuum ($\Delta E_{Coulombic} = E_{Coulombic}^{complex} - E_{Coulombic}^{protein} - E_{Coulombic}^{ligand}$). Calculation of $\Delta E_{Coulombic}$ is the same for all models. The final step is to solvate the complex into the solvent, where the corresponding energy cost would be $\Delta G_{pol}^{complex}$. Using this cycle, the electrostatic component of binding free energies ($\Delta\Delta G_{pol}$) can be computed as

$$\Delta\Delta G_{pol} = \Delta G_{pol}^{complex} - \Delta G_{pol}^{protein} - \Delta G_{pol}^{ligand} + \Delta E_{Coulombic} \quad (3)$$

The standard deviation of computed $\Delta\Delta G_{pol}$ values from TI is smaller than ±0.58 kcal/mol.

2.5. Computational Expense. A general performance comparison of the different methods used here to compute the electrostatic solvation free energies of the protein−ligand complexes is given in Table 2. The computations of GBNSR6 and PB are performed on a commodity PC with Intel(R) Core(TM) i7-3770 CPU 3.40 GHz processor and 16 GB of RAM memory. All of the explicit solvent free energy TI calculations were performed on Virginia Tech's HokieSpeed supercomputing cluster (http://www.arc.vt.edu) on a single node that has 12 processors. Expectedly, GBNSR6 is significantly faster than the PB and explicit models studied here. Our studies show that the computational time required for grid-based calculation of the molecular surface needed by GBNSR6 is similar to that of MSMS-based calculations; a detailed and exhaustive performance analysis of GBNSR6 based on MSMS molecular surface[109] is presented in ref 13.

Table 2. Average Computational Time for Calculating ΔG_{pol} per Complex

method	computational time
explicit solvent TI	≈12 h
PB	≈15 min
GBNSR6	≈6 s

3. RESULTS AND DISCUSSION

At the moment, comparison with explicit solvent predictions is a natural way to evaluate the accuracy of implicit solvent models such as the GB. Among the most commonly used explicit water models, TIP3P is known to give better accuracy in hydration free energy calculations than many other water models tested previously;[11] TIP3P has been commonly used for benchmarking implicit solvent models. However, recent developments in building explicit water models yielded a model (OPC water model[93]) that shows better agreement with experiment in hydration free energy calculations of small molecules. To be consistent with earlier works, we first compare deviation of the implicit models relative to TIP3P as reference, although we do not imply that TIP3P is any better or worse than the other explicit models studied here. We also benchmark computed electrostatic solvation free energy ΔG_{pol} and electrostatic binding free energy $\Delta\Delta G_{pol}$ values against free energy estimates performed in OPC water. Below we give a brief summary of the agreement of GBNSR6 (and the numerical PB) compared with the explicit solvent. The correlation of ΔG_{pol} and $\Delta\Delta G_{pol}$ values relative to TIP3P and OPC are presented in Figures 2 and 4, respectively. The statistics of ΔG_{pol} and $\Delta\Delta G_{pol}$ values relative to TIP3P are given in Table 3 and Table 4, and relative to OPC they are given in Table 6 and Table 7, respectively.

3.1. Implicit Models. *ΔG_{pol} Deviations from TIP3P.* The computed values of ΔG_{pol} for the protein−ligand complexes and their components obtained from GBNSR6 are compared with the corresponding TIP3P (TI) values. Figure 2(a) and (b) shows that among different implicit and explicit models, $\Delta G_{pol}^{complex}$ and $\Delta G_{pol}^{protein}$ values from GBNSR6 agree best with TIP3P ΔG_{pol} values, with RMSD values of 10.76 and 10.53 kcal/mol, respectively (Table 3). Deviation of ΔG_{pol}^{ligand} values computed using GBNSR6 from TIP3P is comparable to that of OPC from TIP3P (Figure 2(c)). The next model to best reproduce TIP3P's $\Delta G_{pol}^{complex}$ and $\Delta G_{pol}^{protein}$ values is the PB model, although it shows the lowest agreement with TIP3P for ΔG_{pol}^{ligand}, among all the solvent models studied here. At the same time, the better agreement of GBNSR6 with TIP3P values holds for the ligands as well.

$\Delta\Delta G_{pol}$ Deviations from TIP3P. The RMSD error of PB's $\Delta\Delta G_{pol}$ relative to TIP3P is 5.14 kcal/mol, which is slightly lower than the deviation of TIP4PEw from TIP3P (RMSD =

DOI: 10.1021/acs.jctc.5b00483
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

J. Chem. Theory Comput.
Downloaded from pubs.acs.org by UNIV OF PENNSYLVANIA on 08/22/15. For personal use only.

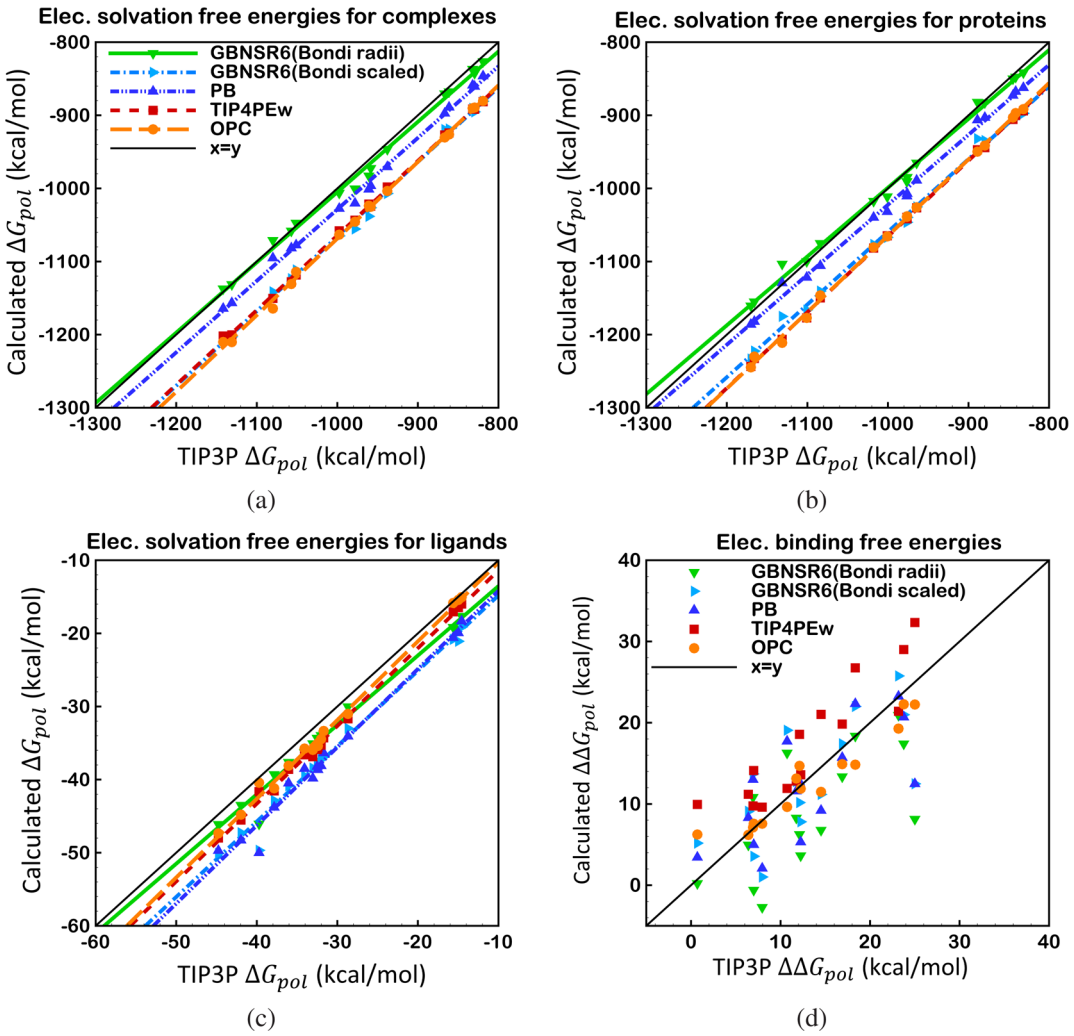


Figure 2. Correlation between ΔG_{pol} and $\Delta\Delta G_{pol}$ computed by GBNSR6, PB, TIP4PEw, and OPC solvent models relative to TIP3P for 15 small protein−ligand complexes specified in Table 1: a) ΔG_{pol} of protein−ligand complexes, b) ΔG_{pol} of protein components, c) ΔG_{pol} of ligand components, and d) electrostatic binding free energies, $\Delta\Delta G_{pol}$.

Table 3. Deviation of (ΔG_{pol}) Values (kcal/mol) from Those Computed with TIP3P Explicit Solvent Model[a]

RMSD	TIP4PEw	OPC	PB (Bondi)	GBNSR6 (Bondi)	GBNSR6 (Bondi scaled)
$\Delta G_{pol}^{complexes}$	63.9	67.3	30.2	10.8	65.8
$\Delta G_{pol}^{proteins}$	65.6	64.8	24.3	10.5	59.8
$\Delta G_{pol}^{ligands}$	2.9	2.2	5.9	2.8	5.6

[a]Atomic radii sets used in implicit solvent estimates are given in parentheses.

Table 4. Deviation of ($\Delta\Delta G_{pol}$) Values (kcal/mol) from Those Computed with TIP3P Explicit Solvent Model[a]

	TIP4PEw	OPC	PB (Bondi)	GBNSR6 (Bondi)	GBNSR6 (Bondi scaled)
RMSD	5.30	2.47	5.14	7.04	5.31
av	4.29	−0.57	−0.99	−4.37	−0.36
corr coeff (r^2)	0.81	0.91	0.52	0.47	0.50
RMS of worst 20%	8.36	4.41	9.2	12.56	9.57

[a]Atomic radii sets used in implicit solvent estimates are given in parentheses.

5.30 kcal/mol) (Table 4). The RMSD error of GBNSR6 based on Bondi radii relative to TIP3P (7.04 kcal/mol) is slightly higher than that of the PB (5.14 kcal/mol). As can be seen in Table 4, the average error in $\Delta\Delta G_{pol}$ values of GBNSR6 is relatively large (−4.37 kcal/mol) which indicates a systematic error[19] relative to TIP3P. The systematic deviation reflects the uncertainties associated with the "best" definition of the dielectric boundary needed by the GB and PB models.[110] To confirm the boundary definition origin of the systematic component of the deviation between GBNSR6 and TIP3P electrostatic binding energies, we show that the average error can be virtually eliminated (−0.36 kcal/mol) by a uniform scaling (multiplication) of the Bondi radii[102,110] with a single coefficient of 0.968 (Table 4 and Figure 2). The resulting RMSD error of GBNSR6 based on scaled Bondi radii relative to TIP3P is reduced to 5.31 kcal/mol and becomes comparable to the RMS deviations of TIP4PEw and PB models relative to TIP3P (5.3 kcal/mol). We stress that uniform radii scaling by a single multiplicative factor[102,110] is not tantamount to full reoptimization[111,112] of the radii intended for best fit against a specific explicit solvent reference. The same scaling of Bondi radii by 0.968 also virtually eliminates the systematic deviation between $\Delta\Delta G_{pol}$ from GBNSR6 and a very different explicit

DOI: 10.1021/acs.jctc.5b00483
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

solvent model (OPC) (as we shall see later) and nearly halves the deviation between GBNSR6 and TIP4PEw (from −8.66 kcal/mol to −4.65 kcal/mol).

3.2. Explicit Models. *ΔG_{pol} Deviations from TIP3P.* The absolute ΔG_{pol} values calculated using explicit water models differ significantly between themselves, Figure 2 and Table 3. Surprisingly, TIP4PEw and OPC show the lowest agreement with TIP3P in $\Delta G_{pol}^{complex}$ and $\Delta G_{pol}^{protein}$ values compared to implicit models. As seen from Figure 2, ΔG_{pol} values from TIP4PEw and OPC deviate systematically from the TIP3P values, with overall average deviations of −43.96 and −44.45 kcal/mol and RMS deviations of 52.89 and 53.96 kcal/mol, respectively. Significant discrepancies between the values from OPC and TIP4PEw relative to TIP3P are likely due to stronger electrostatic interactions in OPC and TIP4PEw compared to TIP3P. For instance, while the dipole moment for TIP3P and TIP4PEw are close to each other (2.35 D vs 2.32 D, respectively) (Table 5), TIP4PEw's square quadrupole is significantly larger than that of TIP3P (2.16 DÅ vs 1.72 DÅ) (Table 5). OPC's dipole and square quadrupole moments are both larger (and closer to Quantum Mechanical predictions, Table 5) than those of TIP3P and TIP4PEw (Table 5). As a result of the differences in the strength of electrostatic interactions,[30] the numbers of hydrogen bonds formed by the three explicit water models differ (Figure 3). Specifically, the average number of hydrogen bonds formed between the solute and the solvent for our molecular systems in TIP4PEw is higher than that in TIP3P, and it is the highest for OPC (Figure 3). Solvation free energies increase almost linearly with the average number of hydrogen bonds, Figure 3. As a result, OPC yields the largest ΔG_{pol} values, followed by TIP4PEw and TIP3P. We did not find a correlation between the value of the static dielectric constant and electrostatic solvation free energies for the explicit solvent models studied here.

Table 5. Three Lowest Order Multipole Moments of the Water Molecule[a]

model	μ [D]	Q_0 [DÅ]	Q_T [DÅ]	Ω_0 [DÅ^2]	Ω_T [DÅ^2]
EXP[113]	2.5−3	NA	NA	NA	NA
QM[114]	2.55	0.20	2.81	−1.52	2.05
TIP3P[91]	2.35	0.23	1.72	−1.21	1.68
TIP4PEw[92]	2.32	0.21	2.16	−1.53	2.11
OPC[93]	2.48	0.20	2.3	−1.484	2.068

[a]The values found in explicit water models are compared to experiment (EXP) where available and liquid phase quantum calculations (QM). Moments are computed relative to oxygen center: dipole (μ), linear (Q_0) and square (Q_T) quadrupole, linear (Ω_0) and square (Ω_T) octupole.

$\Delta\Delta G_{pol}$ Deviations from TIP3P. The RMS deviation in $\Delta\Delta G_{pol}$ calculated with OPC and TIP4PEw water models relative to the TIP3P reference are 2.47 and 5.3 kcal/mol, respectively. The RMSD of "worst" (largest deviation) 20% of TIP4PEw's $\Delta\Delta G_{pol}$ values relative to TIP3P is as large as 8.36 kcal/mol. The deviations of $\Delta\Delta G_{pol}$ from explicit models appear much smaller than the deviations in ΔG_{pol}; however, $\Delta\Delta G_{pol}$ values are relatively much smaller (tens of kcal/mol) than ΔG_{pol} (thousands of kcal/mol), see Figure 2 and Figure 4, which results in large relative errors in $\Delta\Delta G_{pol}$. For instance, deviation of TIP4PEw from TIP3P can be up to 50% in relative difference of $\Delta\Delta G_{pol}$ values. Note that, expectedly, the correlation between $\Delta\Delta G_{pol}$ estimated by solvent models of the same class (e.g., TIP4PEw vs TIP3P) is considerably better than that between very different solvent models such as the GB and TIP3P (Table 4). At the same time, the average deviations between $\Delta\Delta G_{pol}$ computed by explicit models (e.g., TIP4PEw vs TIP3P) is still large, essentially comparable to the deviations between implicit and explicit models. The high correlation between the explicit solvent estimates suggests that the deviations between them may be systematic. This observation is further strengthened by the fact that the systematic error between the implicit and explicit solvent $\Delta\Delta G_{pol}$ can be virtually eliminated by a one-parameter adjustment of the dielectric boundary used in the implicit estimates, as discussed earlier.

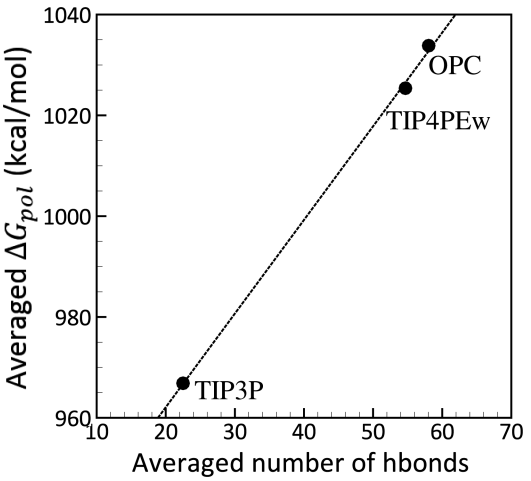


Figure 3. Correlation between the electrostatic solvation free energies ΔG_{pol} and the number of hydrogen bonds formed between the complexes and the explicit solvent models (TIP3P, TIP4PEw, and OPC). The ΔG_{pol} values shown for each model are averages over complexes, and the number of hydrogen bonds represents averages over MD trajectory and over complexes. The hydrogen bond is considered to be formed if the distance between the acceptor (A) and the donor (D) atoms is smaller than 3 Å, and angle D-H-A is greater than 135°.[115] Connecting line is shown to guide the eye.

Another interesting observation is that the ability of one explicit model to emulate estimates of ΔG_{pol} by another model can be independent of its ability to emulate $\Delta\Delta G_{pol}$; for instance, among all implicit and explicit models studied here, OPC shows the closest agreement with TIP3P in $\Delta\Delta G_{pol}$, while its ΔG_{pol} is furthest from TIP3P.

3.3. ΔG_{pol} and $\Delta\Delta G_{pol}$ Deviations from OPC. Here we investigate deviation of computed electrostatic solvation free energy ΔG_{pol} and electrostatic binding free energy $\Delta\Delta G_{pol}$ values from the values estimated with a recently developed explicit water model, OPC. The GBNSR6's ΔG_{pol} values based on Bondi radii are systematically shifted from the OPC reference values (Figure 4 (a), (b), and (c)). Yet, deviation of GBNSR6's $\Delta\Delta G_{pol}$ values from OPC is comparable to that of TIP4PEw from OPC. The same simple uniform scaling (multiplication) of all the radii in the Bondi set by 0.968 introduced earlier also virtually eliminates the systematic deviation between GBNSR6 and OPC in ΔG_{pol} and $\Delta\Delta G_{pol}$ values simultaneously (Figure 4 and Table 6). Interestingly, the radii rescaling, which amounts to the dielectric boundary adjustment, makes the deviation of GBNSR6's $\Delta\Delta G_{pol}$ from OPC even smaller than that of TIP4PEw from OPC (average error 0.2 kcal/mol vs 4.86 kcal/mol.)

J. Chem. Theory Comput.
Downloaded from pubs.acs.org by UNIV OF PENNSYLVANIA on 08/22/15. For personal use only.

DOI: 10.1021/acs.jctc.5b00483
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

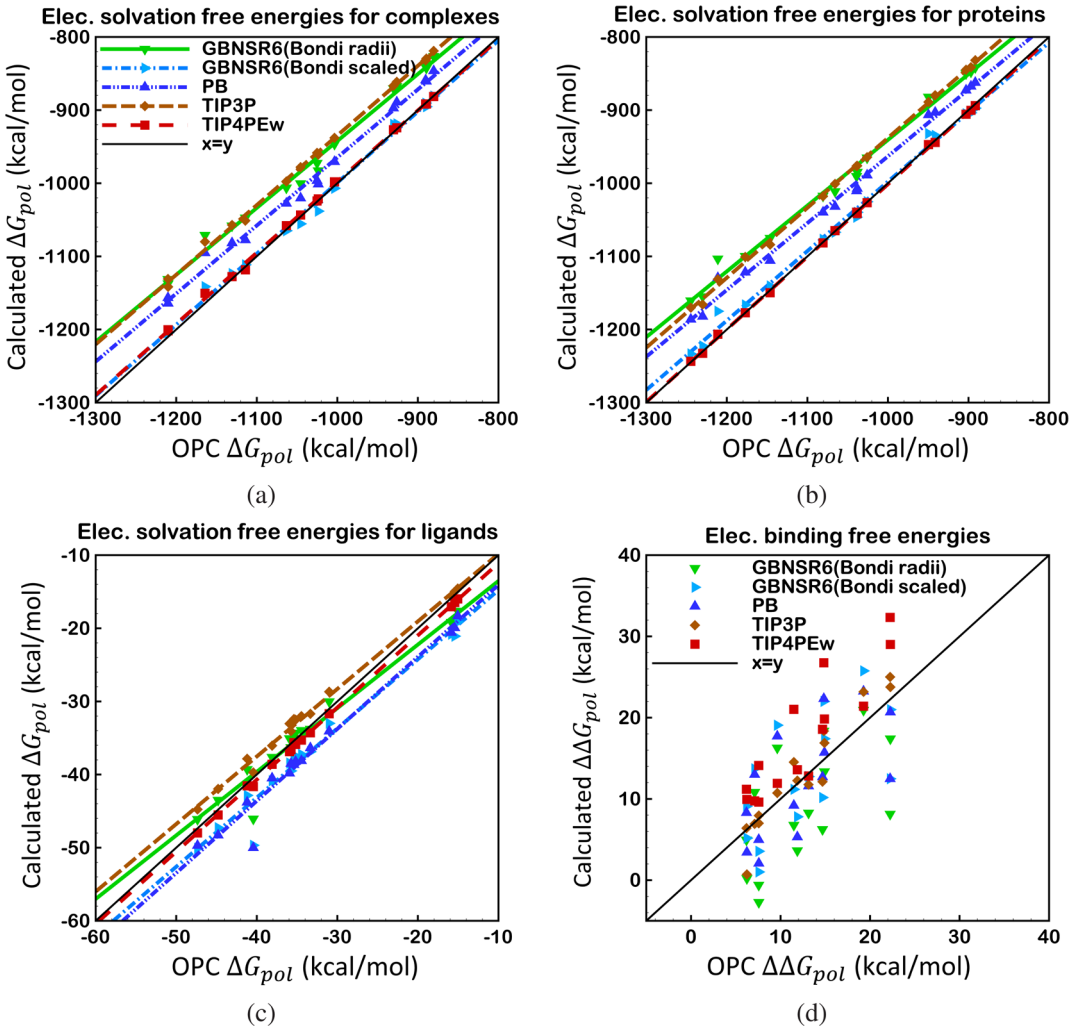


Figure 4. Correlation between ΔG_{pol} and $\Delta\Delta G_{pol}$ computed by GBNSR6, PB, TIP4PEw, and TIP3P solvent models relative to OPC for 15 small protein−ligand complexes specified in Table 1: a) ΔG_{pol} of protein−ligand complexes, b) ΔG_{pol} of protein components, c) ΔG_{pol} of ligand components, and d) electrostatic binding free energies, $\Delta\Delta G_{pol}$.

Table 6. Deviation of (ΔG_{pol}) Values (kcal/mol) from Those Computed with OPC Explicit Solvent Model[a]

RMSD	TIP3P	TIP4PEw	PB (Bondi)	GBNSR6 (Bondi)	GBNSR6 (Bondi scaled)
$\Delta G_{pol}^{complexes}$	67.3	5.62	39.38	62.14	8.92
$\Delta G_{pol}^{proteins}$	64.8	2.49	44.00	66.82	11.99
$\Delta G_{pol}^{ligands}$	2.2	0.83	4.08	2.26	3.98

[a]Atomic radii sets used in implicit solvent estimates are given in parentheses.

Optimizing atomic radii, including uniform scaling, was used earlier[102,110−112,116−118] to better reproduce solvation free energies from explicit solvent models. Here the scaling is used mainly to show that the apparent systematic deviation between the GB and explicit solvent is a consequence of a (radii-specific) definition of the dielectric boundary, the deviation can be removed by a uniform "shift" of the boundary. Still, achieving a good agreement with 3 different explicit water models simultaneously by a single-parameter uniform scaling of the Bondi radii, one of the smallest and simplest radii sets available in the literature, seems noteworthy. Obviously, transferability of the scaled Bondi radii set optimized for the limited set of protein−ligand complexes is not guaranteed, which motivates future studies.

To further illustrate the sensitivity of electrostatic binding free energies to the choice of explicit water model, we have also compared the TIP4PEw to the OPC-based estimates, (Figure 4, Table 6, and Table 7). In some of these comparisons, TIP4PEw and OPC water models appear more similar to each other than to TIP3P: both are 4-point models parametrized for use in long-range electrostatics interactions, and the polarization correction is included in calculations of heat of vaporization

Table 7. Deviation of ($\Delta\Delta G_{pol}$) Values (kcal/mol) from Those Computed with OPC Explicit Solvent Model[a]

	TIP3P	TIP4PEw	PB (Bondi)	GBNSR6 (Bondi)	GBNSR6 (Bondi scaled)
RMSD	2.47	5.92	5.00	6.80	5.38
av	−0.57	4.86	−0.42	−3.80	0.20
corr coeff (r^2)	0.91	0.80	0.44	0.37	0.41
RMS of worst 20%	4.41	10.56	8.5	11.20	8.86

[a]Atomic radii sets used in implicit solvent estimates are given in parentheses.

J. Chem. Theory Comput.
Downloaded from pubs.acs.org by UNIV OF PENNSYLVANIA on 08/22/15. For personal use only.

DOI: 10.1021/acs.jctc.5b00483
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

in the parametrization procedure.[88,92,93] It is evident from Figure 4 (a), (b), and (c) that ΔG_{pol} values estimated with TIP4PEw and OPC are highly correlated, and the RMS deviation of ΔG_{pol} calculations using TIP4PEw relative to OPC is relatively small. Yet, $\Delta\Delta G_{pol}$ values from TIP4PEw substantially deviate from that of OPC (RMSD = 5.92 kcal/mol). Despite the much smaller absolute $\Delta\Delta G_{pol}$ values compared to ΔG_{pol} values, the RMS deviation of TIP4PEw from OPC in $\Delta\Delta G_{pol}$ values is even larger than that in ΔG_{pol} (5.92 kcal/mol for $\Delta\Delta G_{pol}$ vs 5.62 kcal/mol for $\Delta G_{pol}^{complex}$). The deviation of TIP4PEw from OPC is even larger than the one between TIP4PEw and TIP3P − water models that are parametrized quite differently (Table 3). Surprisingly, the RMS deviation of TIP4PEw relative to OPC is similar to the RMSD error of implicit models (GBNSR6 based on scaled Bondi radii and PB) relative to TIP3P.

4. CONCLUSION

An accurate representation of the solvent is crucial for realistic and physically rigorous calculations of solvation and protein−ligand binding free energies. In this work, we have evaluated the accuracy of a recently developed generalized Born model, GBNSR6, in predicting the electrostatic binding free energies $\Delta\Delta G_{pol}$ and electrostatic solvation free energies ΔG_{pol} of small protein−ligand complexes and their components. The estimates from GBNSR6 (and also the standard numerical PB) were compared to the estimates based on three explicit solvent models: RMS deviations of GBNSR6 and the PB from the explicit models were found to be comparable. It was shown that RMS deviation from TIP3P of GBNSR6 (Bondi radii) is comparable to the "error margin" of the explicit models themselves − the differences between the $\Delta\Delta G_{pol}$ values obtained from the explicit models (e.g., TIP4PEw vs TIP3P). Expectedly, the r^2 correlation between either of the implicit models and the explicit solvent is lower than between different explicit solvent models. GBNSR6's $\Delta\Delta G_{pol}$ is closer to estimates based on OPC − a new 4-point rigid water model shown to give higher accuracy in estimation of solvation free energies of small molecules compared to TIP3P.[93] A simple uniform scaling of Bondi radii set was shown to bring GBNSR6 RMS deviation essentially within the "error margin" of the three explicit models. The same simple scaling of Bondi radii was shown to virtually eliminate the systematic deviation of GBNSR6 binding affinities from two out of the three explicit models and reduce the deviation from the third one by about 50%. Although the scaled Bondi radii set presented here is not guaranteed to be transferable to protein−ligand systems other than the ones studied here, the fact that a single-parameter uniform scaling of radii significantly improves the agreement of implicit solvent GBNSR6 with all three explicit models simultaneously is noteworthy.

A perhaps unexpected finding is that computed binding and solvation free energies using explicit water models can deviate significantly from each other. Also counterintuitively, the lowest RMS deviation from TIP3P's ΔG_{pol} is achieved by GBNSR6, rather than by the solvent models of the same class such as TIP4PEw and OPC, with RMS errors being up to tens of kcal/mol smaller. The results show that RMS deviations of $\Delta\Delta G_{pol}$ values obtained from different explicit models can be larger than that of ΔG_{pol}, although $\Delta\Delta G_{pol}$ is often orders of magnitude smaller than ΔG_{pol} in absolute values. The discrepancies between results from explicit models indicate the high sensitivity of electrostatic solvation and protein−ligand binding free energy calculations to the choice of explicit water models. Other studies have previously reported that Poisson-Boltzmann based approaches that are accurate in calculating ΔG_{pol} may not be equally successful at predicting $\Delta\Delta G_{pol}$.[17] The high sensitivity of $\Delta\Delta G_{pol}$ values to the choice of explicit water models observed in this work suggests that the sensitivity is not necessarily inherent to implicit solvent models.

In the absence of a "gold standard" explicit solvent model, the large discrepancy in free energy estimates from explicit water models is of paramount concern as it is unclear which of these models is most accurate in these calculations. Some of the water models most commonly used for solvation free energy calculations (e.g., TIP3P) can misrepresent key bulk properties of water by as much as 250% off the experimental values,[90] suggesting the presence of serious physical flaws in these models. At the same time, these explicit models are often treated as accuracy "gold standard" for implicit solvents such as the GB, justified by the idea that implicit solvent models are designed to mimic the effects of explicit models as their higher level predecessors in the hierarchy of approximations leading to these models. In fact, the approach of fitting GB models to explicit solvent models seems reasonable because going directly from GB models to experimental observations with multiple levels of approximation can lead to overfitting of the GB model parameters. Yet, given the significantly improved accuracy of most recently developed GB models, the strategy is called into question by the large discrepancies between free energy estimates obtained from commonly used explicit models. Efforts to develop more accurate explicit models,[93,119,120] or identifying the best among the existing ones, will ultimately help improve accuracy of implicit solvent models as well. In the meantime, adjusting implicit solvent theory to provide the best match with several explicit models simultaneously might be the best practical strategy.

A direct comparison with experiment is needed for a decisive accuracy evaluation of explicit models in protein−ligand interactions. However, these comparisons are not straightforward. Protein−ligand complexes are very flexible and have many degrees of freedom introducing large uncertainties in calculations that make a direct comparison with experiment often difficult. An appealing alternative is to compare the computed binding enthalpies with experimental enthalpies for small host−guest systems.[101,121] Many fewer degrees of freedom and relative rigidity of host−guest systems compared to protein−ligand systems make these calculations computationally more straightforward and robust, albeit being still time-consuming.[23,101]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00483.

ΔG_{pol} values for 15 small protein−ligand complexes and their components computed using TIP3P, TIP4PEw, OPC, APBS, and GBNSR6 (XLSX)

Amber format topology and coordinate files (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: alexey@cs.vt.edu.

Notes

The authors declare no competing financial interest.

J. Chem. Theory Comput.
Downloaded from pubs.acs.org by UNIV OF PENNSYLVANIA on 08/22/15. For personal use only.

DOI: 10.1021/acs.jctc.5b00483
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

J. Chem. Theory Comput.
Downloaded from pubs.acs.org by UNIV OF PENNSYLVANIA on 08/22/15. For personal use only.

■ ACKNOWLEDGMENTS

We would like to thank Ekaterina Katkova for the help with the preparation of the small protein−ligand complexes used in this work. We are also grateful to Marcia O. Fenley and Robert C. Harris for valuable comments and suggestions. This work was supported by the NIH GM076121 and in part by NSF grant CNS-0960081 and the HokieSpeed supercomputer at Virginia Tech.

■ REFERENCES

(1) Jorgensen, W. L. *Science* **2004**, *303*, 1813−1818.
(2) Mobley, D. L.; Dill, K. A. *Structure* **2009**, *17*, 489−498.
(3) Shirts, M. R.; Mobley, D. L.; Brown, S. P. In *Free energy calculations in structure-based drug design*, 1st ed.; Merz, K. M., Ringe, D., Reynolds, C. H., Eds.; Lecture Notes in Computer Science; Cambridge University Press: Cambridge, New York, USA, 2010; pp 61−85.
(4) Skandani, A. A.; Zeineldin, R.; Al-Haik, M. *Langmuir* **2012**, *28*, 7872−7879. PMID: 22545729.
(5) Lau, A. Y.; Roux, B. A. *Nat. Struct. Mol. Biol.* **2011**, *18*, 283−287.
(6) Lin, Y.-L.; Meng, Y.; Jiang, W.; Roux, B. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 1664−1669.
(7) Shirts, M.; Mobley, D. *Biomolecular Simulations*; Humana Press: 2013; Vol. *924*, pp 271−311.
(8) Gallicchio, E.; Levy, R. M. Recent theoretical and computational advances for modeling protein-ligand binding affinities. In *Computational chemistry methods in structural biology*; Christov, C., Ed.; Academic Press: 2011; Vol. *85*, pp 27−80.
(9) Rocklin, G. J.; Mobley, D. L.; Dill, K. A. *J. Chem. Theory Comput.* **2013**, *9*, 3072−3083.
(10) Mobley, D. L.; Dill, K. A.; Chodera, J. D. *J. Phys. Chem. B* **2008**, *112*, 938−946.
(11) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. *J. Chem. Theory Comput.* **2009**, *5*, 350−358.
(12) Shivakumar, D.; Deng, Y.; Roux, B. *J. Chem. Theory Comput.* **2009**, *5*, 919−930.
(13) Aguilar, B.; Onufriev, A. V. *J. Chem. Theory Comput.* **2012**, *8*, 2404−2411.
(14) Bartlett, R. J.; Musiał, M. *Rev. Mod. Phys.* **2007**, *79*, 291−352.
(15) Helgaker, T.; Klopper, W.; Tew, D. P. *Mol. Phys.* **2008**, *106*, 2107−2143.
(16) Fennell, C. J.; Kehoe, C. W.; Dill, K. A. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3234−3239.
(17) Harris, R. C.; Mackoy, T.; Fenley, M. O. *J. Chem. Theory Comput.* **2015**, *11*, 705−712.
(18) Harris, R. C.; Mackoy, T.; Fenley, M. O. *Mol. Based Math. Biol.* **2013**, *1*, 63−74.
(19) Merz, K. M. *J. Chem. Theory Comput.* **2010**, *6*, 1769−1776.
(20) Friesner, R. A.; Repasky, M.; Farid, R. In *Small Molecule Docking. Computational Structural Biology*; Schwede, T., Peitsch, M. C., Eds.; World Scientific: Switzerland, 2008; pp 469−500.
(21) Kolb, P.; Irwin, J. *Curr. Top. Med. Chem.* **2009**, *9*, 755−770.
(22) Gilson, M. K.; Zhou, H. X. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21−42.
(23) Chen, W.; Chang, C.-E.; Gilson, M. K. *Biophys. J.* **2004**, *87*, 3035−3049.
(24) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. *J. Mol. Biol.* **2007**, *371*, 1118−1134.
(25) Onufriev, A. V.; Alexov, E. *Q. Rev. Biophys.* **2013**, *46*, 181−209.
(26) Deng, Y.; Roux, B. *J. Phys. Chem. B* **2009**, *113*, 2234−2246. PMID: 19146384.
(27) Moghaddam, S.; Inoue, Y.; Gilson, M. K. *J. Am. Chem. Soc.* **2009**, *131*, 4012−4021. PMID: 19133781.
(28) Davis, M. E.; McCammon, J. A. *Chem. Rev.* **1990**, *90*, 509−521.
(29) Jayaram, B.; Sharp, K. A.; Honig, B. *Biopolymers* **1989**, *28*, 975−993.
(30) Morokuma, K. *Acc. Chem. Res.* **1977**, *10*, 294−300.
(31) Qin, S.; Zhou, H.-X. *Biopolymers* **2007**, *86*, 112−118.
(32) Anandakrishnan, R.; Baker, C.; Izadi, S.; Onufriev, A. V. *PLoS One* **2013**, *8*, e67715.
(33) Zhang, Z.; Witham, S.; Alexov, E. *Phys. Biol.* **2011**, *8*, 035001.
(34) Kundrotas, P. J.; Alexov, E. *Biophys. J.* **2006**, *91*, 1724−1736.
(35) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161−2200.
(36) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144−1149.
(37) Beroza, P.; Case, D. A. *Methods Enzymol.* **1998**, *295*, 170−189.
(38) Madura, J. D.; Davis, M. E.; Gilson, M. K.; Wade, R. C.; Luty, B. A.; McCammon, J. A. *Rev. Comp. Chem.* **1994**, *5*, 229−267.
(39) Gilson, M. K. *Curr. Opin. Struct. Biol.* **1995**, *5*, 216−223.
(40) Scarsi, M.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem. A* **1997**, *101*, 8098−8106.
(41) Luo, R.; David, L.; Gilson, M. K. *J. Comput. Chem.* **2002**, *23*, 1244−1253.
(42) Simonson, T. *Rep. Prog. Phys.* **2003**, *66*, 737−787.
(43) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins: Struct., Funct., Genet.* **2004**, *55*, 383−394.
(44) Labute, P. *J. Comput. Chem.* **2008**, *29*, 1693−1698.
(45) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435−445.
(46) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037−10041.
(47) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219−10225.
(48) Im, W.; Beglov, D.; Roux, B. *Comput. Phys. Commun.* **1998**, *111*, 59−75.
(49) Madura, J. D.; Davist, M. E.; Gilson, M. K.; Wades, R. C.; Luty, B. A.; McCammon, J. A. In *Biological Applications of Electrostatic Calculations and Brownian Dynamics Simulations*; John Wiley and Sons, Inc.: 2007; pp 229−267.
(50) Altman, M. D.; Bardhan, J. P.; White, J. K.; Tidor, B. *J. Comput. Chem.* **2009**, *30*, 132−153.
(51) Li, B.; Cheng, X.; Zhang, Z. *SIAM J. Appl. Math.* **2011**, *71*, 2093−2111.
(52) Simonov, N. A.; Mascagni, M.; Fenley, M. O. *J. Chem. Phys.* **2007**, *127*, 185105.
(53) Feig, M.; Brooks, C. L. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217−224.
(54) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(55) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122−129.
(56) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824−19836.
(57) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578−1599.
(58) Qiu, D.; Shenkin, P.; Hollinger, F.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.
(59) Edinger, S.; Cortis, C.; Shenkin, P.; Friesner, R. *J. Phys. Chem. B* **1997**, *101*, 1190−1197.
(60) Jayaram, B.; Liu, Y.; Beveridge, D. L. *J. Chem. Phys.* **1998**, *109*, 1465−1471.
(61) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983−10990.
(62) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−152.
(63) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. *J. Chem. Phys.* **2002**, *116*, 10606−10614.
(64) Felts, A. K.; Harano, Y.; Gallicchio, E.; Levy, R. M. *Proteins: Struct., Funct., Genet.* **2004**, *56*, 310−321.
(65) Romanov, A. N.; Jabin, S. N.; Martynov, Y. B.; Sulimov, A. V.; Grigoriev, F. V.; Sulimov, V. B. *J. Phys. Chem. A* **2004**, *108*, 9323−9327.
(66) Dominy, B. N.; Brooks, C. L. *J. Phys. Chem. B* **1999**, *103*, 3765−3773.
(67) David, L.; Luo, R.; Gilson, M. K. *J. Comput. Chem.* **2000**, *21*, 295−309.
(68) Tsui, V.; Case, D. *J. Am. Chem. Soc.* **2000**, *122*, 2489−2498.
(69) Calimet, N.; Schaefer, M.; Simonson, T. *Proteins: Struct., Funct., Genet.* **2001**, *45*, 144−158.
(70) Spassov, V. Z.; Yan, L.; Szalma, S. *J. Phys. Chem. B* **2002**, *106*, 8726−8738.

DOI: 10.1021/acs.jctc.5b00483
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(71) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124*, 11258−11259.
(72) Wang, T.; Wade, R. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 158−169.
(73) Nymeyer, H.; Garcia, A. E. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13934−13949.
(74) Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2004**, *25*, 479−499.
(75) Lee, M. C.; Duan, Y. *Proteins: Struct., Funct., Genet.* **2004**, *55*, 620−634.
(76) Onufriev, A. In *Continuum Electrostatics Solvent Modeling with the Generalized Born Model*, 1st ed.; Feig, M., Ed.; Wiley: USA, 2010; pp 127−165.
(77) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.
(78) Gohlke, H.; Kiel, C.; Case, D. A. *J. Mol. Biol.* **2003**, *330*, 891−913.
(79) Dong, F.; Zhou, H.-X. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 87−102.
(80) Aguilar, B.; Anandakrishnan, R.; Ruscio, J. Z.; Onufriev, A. V. *Biophys. J.* **2010**, *98*, 872−880.
(81) Svrcek-Seiler, A. *Personal communication*, 2001.
(82) Grycuk, T. *J. Chem. Phys.* **2003**, *119*, 4817−4826.
(83) Tjong, H.; Zhou, H. X. *J. Phys. Chem. B* **2007**, *111*, 3055−3061.
(84) Mongan, J.; Svrcek-Seiler, A.; Onufriev, A. *J. Chem. Phys.* **2007**, *127*, 185101−185101.
(85) Guillot, B. *J. Mol. Liq.* **2002**, *101*, 219−260.
(86) Wu, Y.; Tepper, H. L.; Voth, G. A. *J. Chem. Phys.* **2006**, *124*, 024503.
(87) Vega, C.; Abascal, J. L. F.; Conde, M. M.; Aragones, J. L. *Faraday Discuss.* **2009**, *141*, 251−276.
(88) Vega, C.; Abascal, J. L. F. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19663−19688.
(89) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910−8922.
(90) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(91) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(92) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665−9678.
(93) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. *J. Phys. Chem. Lett.* **2014**, *5*, 3863−3871.
(94) Lin, Y.-L.; Aleksandrov, A.; Simonson, T.; Roux, B. *J. Chem. Theory Comput.* **2014**, *10*, 2690−2709.
(95) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. *Nucleic Acids Res.* **2012**, *40*, W537−W541.
(96) Kirmizialtin, S.; Elber, R. *J. Phys. Chem. B* **2010**, *114*, 8207−8220. PMID: 20518549.
(97) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(98) Case, D. et al. In *Amber 12*; University of California: San Francisco, 2012.
(99) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.
(100) Sigalov, G.; Fenley, A.; Onufriev, A. *J. Chem. Phys.* **2006**, *124*, 124902.
(101) Fenley, A. T.; Henriksen, N. M.; Muddana, H. S.; Gilson, M. K. *J. Chem. Theory Comput.* **2014**, *10*, 4069−4078.
(102) Zhang, L. Y.; Gallicchio, E.; Friesner, R. A.; Levy, R. M. *J. Comput. Chem.* **2001**, *22*, 591−607.
(103) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441−451.
(104) Cai, Q.; Ye, X.; Wang, J.; Luo, R. *J. Chem. Theory Comput.* **2011**, *7*, 3608−3619.
(105) Case, D. et al. In *Amber 2015*; University of California: San Francisco, 2015.
(106) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037−10041.
(107) Roe, D. R.; Okur, A.; Wickstrom, L.; Hornak, V.; Simmerling, C. *J. Phys. Chem. B* **2007**, *111*, 1846−1857.
(108) Tan, C.; Yang, L.; Luo, R. *J. Phys. Chem. B* **2006**, *110*, 18680−18687. PMID: 16970499.
(109) Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996**, *38*, 305−320.
(110) Onufriev, A. V.; Aguilar, B. *J. Theor. Comput. Chem.* **2014**, *13*, 1440006.
(111) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239−5248.
(112) Swanson, J. M. J.; Mongan, J.; McCammon, J. A. *J. Phys. Chem. B* **2005**, *109*, 14769−14772.
(113) Gregory, J. K.; Clary, D. C.; Liu, K.; Brown, M. G.; Saykally, R. J. *Science* **1997**, *275*, 814−817.
(114) Coutinho, K.; Guedes, R.; Cabral, B. C.; Canuto, S. *Chem. Phys. Lett.* **2003**, *369*, 345−353.
(115) Roe, D. R.; Cheatham, T., III *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.
(116) Banavali, N. K.; Roux, B. *J. Phys. Chem. B* **2002**, *106*, 11026−11035.
(117) Rashin, A. A.; Honig, B. *J. Phys. Chem.* **1985**, *89*, 5588−5593.
(118) Chocholousova, J.; Feig, M. *J. Phys. Chem. B* **2006**, *110*, 17240−17251. PMID: 16928023.
(119) Wang, L. P.; Martinez, T. J.; Pande, V. S. *J. Phys. Chem. Lett.* **2014**, *5*, 1885−1891.
(120) Fennell, C. J.; Li, L.; Dill, K. A. *J. Phys. Chem. B* **2012**, *116*, 6936−6944.
(121) Wickstrom, L.; He, P.; Gallicchio, E.; Levy, R. M. *J. Chem. Theory Comput.* **2013**, *9*, 3136−3150.

J. Chem. Theory Comput.
Downloaded from pubs.acs.org by UNIV OF PENNSYLVANIA on 08/22/15. For personal use only.

DOI: 10.1021/acs.jctc.5b00483
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX