JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Efficient Computation of the Total Solvation Energy of Small Molecules via the R6 Generalized Born Model

Boris Aguilar[†] and Alexey V. Onufriev*[,†,‡]

[†]Department of Computer Science and [‡]Department of Physics, Virginia Tech, Blacksburg, Virginia 24060, United States

Supporting Information

ABSTRACT: Efficient and accurate methodologies to compute solvation free energies of small molecules are relevant for many biological and industrial research areas including rational drug design. In this work we test the performance of a recently developed generalized Born method, GB_NSR6 (Aguilar et al. *J. Chem. Theory Comput.* **2010**, *6*, 3613−3639) on a common benchmark set of 504 small molecules. The computed solvation energies are compared with those obtained previously by explicit solvent models and experiment. The dominant polar component of the solvation energy is computed by GB_NSR6 with no adjustable parameters, producing a root mean square deviation (RMSD) of 0.89 kcal/mol with respect to explicit solvent (TIP3P). The relatively small nonpolar contribution is estimated using the Gallicchio et al. (*J. Comput. Chem.* **2005**, *25*, 479−499) approach. Our results show that GB_NSR6 offers a reasonable balance between efficiency and accuracy: the RMSD from the experiment of computed solvation energies is 1.2 kcal/mol, which is essentially the same as the accuracy of the much more computationally expensive explicit solvent treatment. The average computational time needed to compute the total solvation energy per molecule via GB_NSR6 is only tens of milliseconds on a commodity PC for a typical molecule of about 20 atoms. All of the software developed in this work is freely available from http://people.cs.vt.edu/onufriev/software.php.

1. INTRODUCTION

Accurate computation of the solvation free energy (ΔG_{solv}) of a molecule is central to numerous areas of biomedical and industrial research. In particular, availability of computationally efficient and accurate methods of ΔG_{solv} calculation is important for protein ligand binding and rational drug design,[1−4] which involves screening potential drug molecules for optimal binding affinity with the target protein in the presence of aqueous solvent. Here, the solvation energy is one of the main components of the computed binding affinity. Accurate determination of solvation energy is also important in the study of many other physical properties relevant to drug discovery, such as ionization state changes, solubility, phase transfer, and aggregation.[5,6]

Alchemical free energy calculation, in which all water molecules are explicitly incorporated in the model, is arguably the most realistic and accurate practically available procedure to compute ΔG_{solv}. Recent studies have shown that the approach is able to reproduce experimental solvation energies with a good degree of accuracy.[7,8] For instance, using molecular dynamic (MD) simulations with explicit TIP3P water model, Mobley et al.[8] have obtained an agreement with experimental solvation energies within 1.2 kcal/mol (RMSD). However, an adequate representation of a solvated small molecule (tens of atoms) typically requires hundreds of discrete water molecules; exploring all degrees of freedom in such a system is computationally too intense for many practical applications, especially to analyze large sets of molecules.

The implicit solvent model is a popular alternative framework used to compute solvation free energies of molecules. In this model, discrete water molecules are replaced by an infinite continuum medium with the average dielectric properties of water, thus considerably reducing the number of degrees of freedom of the system. The computational cost associated with the use of these models is therefore significantly smaller than the cost of representing water explicitly. Recent studies using implicit solvent models show a reasonable compromise between accuracy and computational efficiency,[6,9−12] although high-quality explicit solvent simulations provide a more accurate representation of solvation effects.

Within the implicit solvent framework the solvation free energy is typically divided into polar (ΔG_{pol}) and nonpolar (ΔG_{nonpol}) components: $\Delta G_{solv} = \Delta G_{pol} + \Delta G_{nonpol}$. The generalized Born (GB) model is often employed to compute ΔG_{pol}. This model has become quite popular due to its relative simplicity and computational efficiency compared to more standard numerical methodologies, such as solving the Poisson−Boltzmann (PB) equation numerically.[13] Many flavors of the GB model have been proposed recently; these flavors differ in the way they compute the so-called effective Born radii R_i which play a key role in the GB model, determining its accuracy and efficiency, see e.g., refs 14, 15, and 16 for reviews of the model. Many of the GB flavors are commonly used in MD simulations, due to a reasonable compromise between accuracy and speed.[13−16] However, a recent study[17] showed that the accuracy of many commonly used GB flavors in computing ΔG_{solv} of small molecules is still worse than the ∼1 kcal/mol "accuracy limit" expected from continuum models that share the same underlying physics with the more fundamental PB treatment.[10]

A new family of GB flavors have been recently developed;[11,18,19] these are based on the so-called R6 effective Born radius which is calculated as a single $\vec{r}/|\vec{r}|^6$ (R6) integral

Received: November 4, 2011
Published: May 24, 2012

over the Lee–Richards[20] molecular surface (although some authors advocate the use of the van der Waals surface (or volume) instead).[11,19] Unlike the approximations that had been the basis of the majority of the previous generation flavors, the R6 model is exact for what is, perhaps, the single most important limiting case—the sphere. Previous work demonstrated that the values of ΔG_{pol} obtained by the Lee–Richards-based R6 models were in a very close agreement with the values of ΔG_{pol} computed via the more fundamental PB model for small proteins and DNA;[21] in fact, as far as ΔG_{pol} was concerned, the R6 effective radii appear to have reached the accuracy limit of the so-called perfect effective radii based directly on the PB model. A good agreement of the Lee–Richards-based R6 with the TIP3P explicit solvent model for different conformations of alanine polypeptide was also reported.[18] Given the high promise of the R6 flavor, the natural question to ask at this point is whether the R6 is also as good for small molecules relevant to drug design. Two key ingredients are needed to answer this question: a "pure", parameter free R6 GB, and a common, diverse test set of small molecules. A recent implementation of the R6 flavor,[18] called GB_NSR6, is a good fit for the task; its crux is the numerically exact computation of the R6 effective Born radii over the Lee–Richards[20] molecular surface. Importantly, once the geometrical properties of a molecule have been specified, GB_NSR6 requires no additional parameters to compute ΔG_{pol}. Thus, GB_NSR6 can be used to test the accuracy of the different radii sets commonly used in solvation energy calculations. As an added bonus, GB_NSR6 inherits the expected efficiency of the GB models.

At the same time, a fairly large set of small molecules has recently emerged that has already been used extensively in testing of various solvent models, including several common GB flavors.[17] This is the set of 504 neutral small molecules for which experimental solvation free energies are available, conveniently along with the solvation energies from explicit solvent alchemical calculations reported by Mobley et al.[8]

The rest of this article is organized as follows. First, we present a brief description of the data set and the methodology employed here to compute total solvation energies. The Results Section contains an evaluation of the accuracy and computational efficiency of the GB_NSR6 model, compared to explicit solvent models and experimental data. A brief comparison with several other common GB flavors is also included. The results are summarized in the Conclusions Section.

2. METHODS

In this work we compute solvation energies of a data set of 504 neutral small molecules. The experimental solvation energy, the coordinates, and the topology files were obtained from the Supporting Information of Mobley et al.,[8] details of these structures can be found in the same reference.

2.1. Polar Component of ΔG_{solv}. The polar component of the solvation energy was calculated by the analytical linearized Poisson–Boltzmann (ALPB) model,[22] which reintroduces physically correct dependence on the dielectric constants into the original GB model of Still et al.,[23] while maintaining the efficiency of the original. The ALPB model approximates ΔG_{pol} using the following formula:

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

where ε_{in}, and ε_{out} are the dielectric constants of the solute and the solvent respectively, $\beta = \varepsilon_{in}/\varepsilon_{out}$, $\alpha = 0.571412$, and A is the electrostatic size of the molecule, which is essentially the overall size of the structure, that can be computed analytically.[22] Here, q_i is the partial charge of atom i. We employ the most widely used functional form[23] of f^{GB}: $f^{GB} = [r_{ij}^2 + R_i R_j \exp(-r_{ij}^2/4R_iR_j)]^{1/2}$, where R_i is the so-called effective Born radius of atom i, and r_{ij} is the distance between atoms i and j. We set $\varepsilon_{in} = 1$ and $\varepsilon_{out} = 80$ in eq 1.

In this work, the effective Born radii R_i were calculated by the following equation:

$$R_i^{-1} = \left(\frac{3}{4\pi}\int_{ext}\frac{d\mathbf{V}}{|\mathbf{r}-\mathbf{r}_i|^6}\right)^{1/3} \quad (2)$$

proposed by Svrcek-Seiler[24] and independently by Grycuk.[25] In eq 2 the integral (ext) is taken over the region outside the molecule, and r_i is the position of atom i. We use an equivalent formulation described in Mongan et al.:[21]

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

which, by Gauss–Ostrogradski theorem, is equivalent to eq 2. Here, ∂V represents the molecular surface of the molecule, and $d\mathbf{S}$ is the infinitesimal surface vector.

At the first stage of the computation of R_i, the methodology uses the MSMS package[26] to build a numerical triangulation of the Lee–Richards[20] molecular surface. There are no adjustable parameters in the GB_NSR6 model used here (the empirical constant offset B to the inverse radii used in Mongan et al.[21] was set to zero in this work), that is, no fitting of the model to the data was performed. While technically the resolution of the MSMS molecular surface triangulation is controlled by a density parameter, it was not fitted; we chose a value (6 vertex/Å^2) high enough to ensure that the molecular surface is accurate enough. Higher densities result in only marginal increases in accuracy and are not justified by the associated increase in computational costs. Equation 3 is then numerically approximated using the triangles that form the molecular surface, and their respective unit vectors that are orthogonal to the triangles. The surface of the molecule is determined by the intrinsic radius of each atom type of the molecule as well as the solvent probe radius; these parameters obviously affect the ΔG_{solv}, but are external to the GB model; they are used in the same manner in the equivalent PB treatment, to define molecular surface. Here we have tested three standard sets of intrinsic atomic radii: BONDI,[27] PARSE,[28] and ZAP9.[6] Partial atomic charges are also external to the GB model, the same set is used for the corresponding explicit solvent calculations, see below.

2.2. Nonpolar Component of ΔG_{solv}. The nonpolar component of the solvation free energy is commonly modeled as being proportional to the solvent accessible surface area (SASA). However it was shown that this strategy generates a poor correlation when compared to explicit and experimental results.[29–31] Recently, more sophisticated methodologies have been proposed.[11,29–32] Here, unless otherwise specified, the

nonpolar component of the solvation energy is computed in a manner similar to the approach proposed by Gallicchio et al.,[32] in which the nonpolar component of the solvation energy is decomposed into cavity and van der Waals (vdW) terms:

$$\Delta G_{\text{nonpol}} = \Delta G_{\text{cav}} + \Delta G_{\text{vdW}} \quad (4)$$

The cavity component is calculated by

$$\Delta G_{\text{cav}} = \gamma \cdot \text{SASA} \quad (5)$$

where γ is the surface tension parameter. In the original formulation of Gallicchio et al.,[32] γ is an atom-dependent parameter. Here we use a constant atom-independent parameter instead. The parameter γ is optimized using the procedures described further in this section. SASA and surface triangulation are computed by the MSMS[26] package on the same footing.

The solute−solvent van der Waals interaction term is calculated by

$$\Delta G_{\text{vdW}} = \sum_i \mu_i \frac{a_i}{(R_i + \rho_w)^3} \quad (6)$$

where μ_i is a dimensionless adjustable parameter that depends on the atom type, ρ_w is the water probe radius, and R_i is the effective Born radius of atom i previously computed by GB_NSR6 in the computation of polar solvation energy. The values of a_i in eq 6 are computed as[32]

$$a_i = -\frac{16}{3}\pi d_w \varepsilon_{iw} \sigma_{iw}^6 \quad (7)$$

where d_w = 0.033428 Å^{-3} is the number density of water at standard conditions, ε_{iw} and σ_{iw} are computed by

$$\sigma_{iw} = \sqrt{\sigma_i \sigma_w} \quad (8)$$

$$\varepsilon_{iw} = \sqrt{\varepsilon_i \varepsilon_w} \quad (9)$$

where σ_w = 1.7683 Å and ε_w = 0.1520 kcal/mol are the Lennard-Jones parameters of the TIP3P water oxygen, and σ_i and ε_i are the Lennard-Jones parameters for atom i. The values of σ_i and ε_i for each atom type were taken from AMBER 8 and are presented in Table 1.

To find an optimum set of values for μ_i and γ parameters, we randomly selected 104 molecules (our training set) from the main data set. Their names are specified in the Supporting Information. The remaining 400 molecules make up the test set. We used the Nelder−Mead simplex algorithm[33] for optimization. The objective function employed consists of the equally weighted sum of two terms: the root-mean-square deviation (RMSD) of the computed total solvation energies and the experimental energies, and the RMSD of the ΔG_{nonpol} values obtained by GB_NSR6 with respect to those obtained by TIP3P. This is because using only the RMSD with respect to experimental ΔG_{solv} would compensate not only for the error in ΔG_{nonpol} but for errors in the polar part as well, likely resulting in overfitting. Because the Nelder−Mead algorithm finds only local minima, we have tried different initial guesses and selected the one that produces the smallest value of the objective function.

Table 1. Lennard-Jones Parameters Used for the Computation of ΔG_{vdW} for the ZAP9 Radii Set[a]

	σ_i (Å)	ε_i (kcal/mol)	μ_i
H	1.4870	0.0157	−1.5869
C	1.9080	0.1094	4.1621
N	1.8240	0.1700	5.983
O	1.6612	0.2100	4.8457
S	2.0000	0.2500	0.4412
P	2.1000	0.2000	−1.6941
F	1.75	0.061	1.3331
Cl	1.948	0.265	2.7095
Br	2.22	0.320	0.3188
I	2.35	0.40	−0.8997

[a]The optimized value of γ is 0.01 kcal/mol/Å^2 and is the same for all atom types.

Table 1 presents the optimum values of μ_i for each atom type and the optimal γ for the ZAP9 radii set. A total of 10 atom types were considered. Optimum parameters for BONDI and PARSE radii sets are presented in the Supporting Information. In all of the calculations, we have used a water probe radius of 1.4 Å.

The negative values of μ_i can potentially produce positive values of ΔG_{vdW}, see eqs 6 and 7. This is unphysical because ΔG_{vdW} is an approximation of the attractive van der Waals part of the energy and should be negative. Moreover, the negative values of μ_i may indicate overfitting of the polar component; the objective function includes both polar and nonpolar terms. To further investigate this issue, we have optimized the μ_i parameters using only the RMSD of the computed ΔG_{nonpol} energies relative to those of TIP3P. First, we have used the unconstrained Nelder−Seilder optimization method. The resulting μ_i values are still negative, see Supporting Information. This excludes overfitting of the polar component of the solvation energy, because the new objective function is based on the nonpolar component only. The negative values of μ_i in this case are suggestive of inconsistencies in the functional form eq 6 itself. The problems are unlikely to be severe though. When the μ_i parameters were constrained to the $0 < \mu_I < 2$ range, the resulting RMSD (1.25 kcal/mol) of ΔG_{solv} relative to experiment was only marginally larger than the RMSD (1.2 kcal/mol) based on the parameters of Table 1, some of which are negative. Since the set of parameters in Table 1 still produces slightly more accurate results, and since the computed total van der Waals energy of each of the 504 molecules is still positive in this case, we decided to use the μ_i values of Table 1 throughout this work. All the sets of optimized parameters (μ_i and γ) and their corresponding RMS values relative to experiment are available in the Supporting Information.

It is important to note that the μ_i and γ parameters are optimized for small molecules and are not necessarily transferable to proteins or peptides. The set of non-negative μ_i may be interesting to explore in that respect.

2.3. PB Calculations. All the PB calculations were performed using ZAP v2.0, the PB solver from OpenEye.[34] The polar component of the solvation energy was computed using the Gaussian dielectric boundary (solute/solvent interface) option, with a dielectric constant of 80 for water and unity for the internal dielectric. We used a grid resolution of 0.25 Å and a buffer region around the molecule of 4 Å from the surface to box boundary.

3. RESULTS

3.1. Accuracy of Computed Total Solvation Free Energies. We have used GB_NSR6 to calculate the ΔG_{solv} for the data set of 504 neutral small molecules; this set was introduced in ref 12 and subsequently used by a number of

authors.[8,17,35] The computed values of total ΔG_{solv} are compared with experimental solvation energies; the polar and nonpolar components of the solvation energy are compared separately with those obtained previously by explicit solvent MD simulations.[8] We begin by comparing the accuracy of the computed ΔG_{solv} with experimental data; in the computation we have used three different intrinsic radii sets, see Table 2. Clearly, ZAP9 produces more accurate values of ΔG_{solv} with respect to experimental data, compared to the other two radii sets.

Table 2. RMSD and Correlation Coefficients (r^2) of Computed ΔG_{solv} (kcal/mol) with Respect to Experimental Data

radii set	RMSD (kcal/mol)	correlation coefficient (r^2)
PARSE	2.34	0.74
BONDI	1.79	0.70
ZAP9	1.20	0.86

Since both polar and nonpolar components of ΔG_{solv} computed with TIP3P explicit solvent model are available, we can evaluate the relative contribution of the two components to the accuracy of the computed ΔG_{solv}. Figure 1 shows that among the different radii sets, the values of ΔG_{pol} obtained using ZAP9 agree best with the TIP3P ΔG_{pol} values, with a RMSD of 0.89 kcal/mol. It is important to point out that 124 out of the 504 molecules in the data set used here were also used for the original ZAP9 parametrization.[6] ZAP9 radii set was optimized for the PB solver ZAP[34]using experimental solvation energies.

The values of nonpolar energy for typical small molecules (ranging from 0 to 3.5 kcal/mol for this particular data set) are much smaller than the corresponding polar solvation energies and thus contribute much less to the accuracy of the total solvation energy than the polar component, which ranges from 0 to −15 kcal/mol. Thus, at least for the current test set, the accuracy of the total solvation energy is determined mainly by the calculation of the polar component of the solvation energy, in which the radii set plays a key role. We have thus chosen ZAP9 to be our "benchmark" radii set throughout the paper.

The total solvation energies obtained by our GB flavor agree with experiment to the same extent as do the energies based on TIP3P explicit solvent calculations, Figure 2. Namely, GB_NSR6 augmented by the nonpolar term, as described in the Methods Section, produces a RMSD of 1.2 kcal/mol and a correlation coefficient r^2 = 0.86 relative to experimental data, while TIP3P explicit solvent yields in a RMSD of 1.26 kcal/mol and a correlation coefficient of 0.89 relative to experiment.

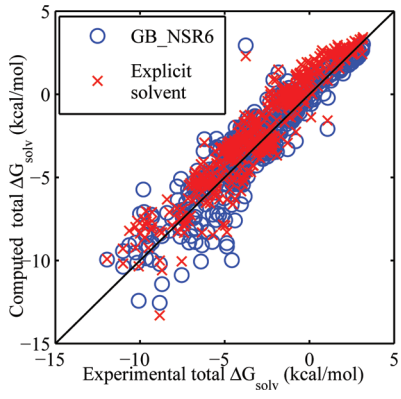

Figure 2. Correlation between the computed and the experimental solvation free energies, ΔG_{solv} for the entire set of 504 small molecules. The computed ΔG_{solv} is obtained via GB_NSR6 (using SASA + vdW as nonpolar term), as described in the Methods Section. The explicit solvent ΔG_{solv} was taken from ref 8. ZAP9 radii are used in the GB_NSR6 calculations.

Additional statistics of the GB_NSR6 model quality, evaluated by a direct comparison with experimental data, are presented in Table 3 in which the results are divided into the training and test sets (104 and 400 molecules, respectively). The RMSDs, the average unsigned errors, and the correlation

Table 3. Accuracy in ΔG_{solv} Calculation (kcal/mol) with Respect to Experimental Data[a]

	training set		test set	
	TIP3P	GB_NSR6	TIP3P	GB_NSR6
RMSD	1.21	1.16	1.27	1.21
avg. error	−0.49	0.01	0.72	−0.25
avg. |error|	1.01	0.88	1.04	0.86
corr. coef. (r^2)	0.89	0.88	0.89	0.86
%|error| > $2k_BT$	36.5%	28.9%	41%	25%
5% worst RMS	2.70	3.07	3.00	3.57

[a]The explicit (TIP3P) solvent results are from ref 8. The last two rows represent the percentage of molecules with gross errors (error > $2k_BT$ ≈ 1.2 kcal/mol at 300 K) and the RMSD of the 5% of the molecules with the largest unsigned errors, respectively. ZAP9 radii are used in the GB_NSR6 calculations.

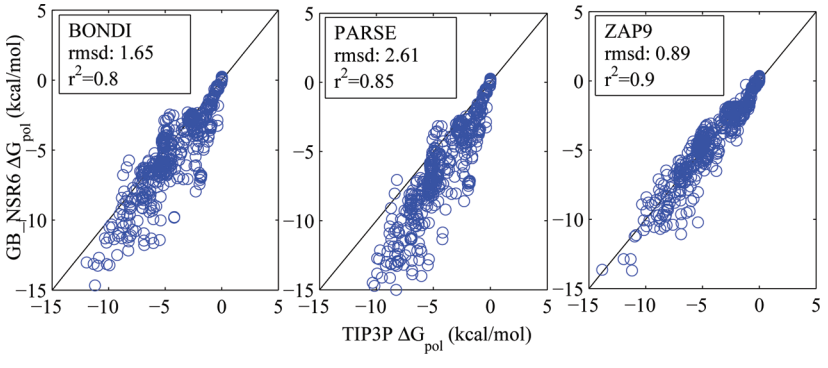

Figure 1. Correlation plots between the values of ΔG_{pol} computed by GB_NSR6 and those included in ref 8 obtained by explicit (TIP3P) solvent model, for a set of 504 neutral small molecules. Three intrinsic atomic radii sets are used in the GB_NSR6 method: BONDI, PARSE, and ZAP9.

dx.doi.org/10.1021/ct200786m | J. Chem. Theory Comput. 2012, 8, 2404−2411

coefficients obtained by GB_NSR6 in both training and test sets are similar to those obtained by the TIP3P explicit solvent model. Moreover, the average error obtained by GB_NSR6 is slightly closer to zero than that obtained by the TIP3P model. About 25% of the molecules in the test set present gross errors, defined as the unsigned errors greater than $2k_BT \sim 1.2$ kcal/mol. This percentage is somewhat smaller than that of TIP3P for which 40% of the molecules present gross errors. Finally, GB_NSR6 produces a RMSD of 3.5 kcal/mol for the 5% of the molecules in the testing set with the largest unsigned error. This number is slightly lower, 3.0 kcal/mol, for the TIP3P solvation energies.

A rigorous comparison of accuracies of different GB flavors is a difficult task, as many of them were developed in different contexts and for different applications. For example, some of the flavors were parametrized for specific radii set(s) and for MD simulations of proteins; these radii sets may not perform optimally on small molecules. Thus, comparing between GB flavors with their respective default parameters may not be as beneficial for model developers who seek to understand the root cause of the differences. Likewise, an optimal radii set may result in suboptimal performance within a "radii-dependent" GB flavor, which also complicates meaningful comparison. Nevertheless, it is important to put the results of GB_NSR6 in perspective with respect to other GB flavors. Such a comparison may be of particular interest to practitioners who want to identify the most suitable "canned" method for the specific task at hand. A recent extensive survey of implicit solvent models by Knight et al.[17] offers an excellent reference point, as the set of molecules employed in that survey is the exact same set used in this work. The RMSD relative to experiment of the GB models benchmarked in Knight et al. varies between 1.5−2.1 kcal/mol with correlations coefficients between $r^2 = 0.66$ and 0.81. GB_NSR6 with ZAP9 radii set produces a RMSD of 1.2 kcal/mol with a correlation coefficient of $r^2 = 0.86$, see Table 3. These numbers essentially do not change if we use the SASA-only model for the nonpolar term as in Knight et al., see the Nonpolar Solvation Energy Section for more details.

Overall, these results show that GB_NSR6 in combination with the ZAP9 radii set provides ΔG_{solv} values in reasonable agreement with experiment for the entire data set of 504 small molecules. The overall level of accuracy is similar to the that of the (orders of magnitude more computationally expensive) TIP3P explicit solvent approach. An analysis of the computational efficiency of the model is presented in the following subsections.

3.2. Nonpolar Solvation Energy. The nonpolar part of the total solvation free energy is calculated via the methodology developed by Gallicchio and Levy.[32] Basically, the nonpolar term is divided into two components, the cavity formation and the solute−solvent van der Waals interaction (SASA + vdW), see Methods Section. This approach requires 11 parameters which we have optimized against experimental total solvation energies and nonpolar solvation energy calculated by TIP3P. Please refer to the Methods Section for details of this strategy.

Our results (blue circles in Figure 3) show a reasonable, though far from perfect, agreement (RMSD = 0.49 kcal/mol and a correlation coefficient of $r^2 = 0.6$) between the values of ΔG_{nonpol} obtained by SASA + vdW and the explicit solvent (TIP3P) ΔG_{nonpol}. To asses potential advantages of the relatively more complicated (Gallicchio and Levy)[32] form of the ΔG_{nonpol} used here vs a cruder model in which $\Delta G_{nonpol} = \Delta G_{cav} = \gamma \cdot$SASA, we have investigated how the total solvation energy would agree with experiment if only the cavity term (SASA-only) of the nonpolar component were considered ($\Delta G_{nonpol} = \gamma \cdot$SASA). The surface tension parameter, γ, for this case has been reoptimized by minimizing the RMSD of the total solvation energy obtained by SASA-only and the experimental data available for the training set. The optimum value thus obtained for γ is 0.0051 kcal/mol/Å^2. The RMSDs and the correlation coefficients relative to experiments obtained by the SASA-only model and the full SASA + vdW are included in Table 4. The SASA-only model yields a RMSD of 1.16 kcal/mol between the computed and the experimental total solvation energies for the 504 molecules, which is essentially the same as the RMSD obtained by the full SASA + vdW approach based on 10 additional parameters. Nevertheless, the red "×" symbols in Figure 3 show that the simpler SASA-only model produces a very poor correlation (RMSD of 0.83 and correlation coefficient of 0.04) with respect to the nonpolar solvation obtained by TIP3P solvent model.

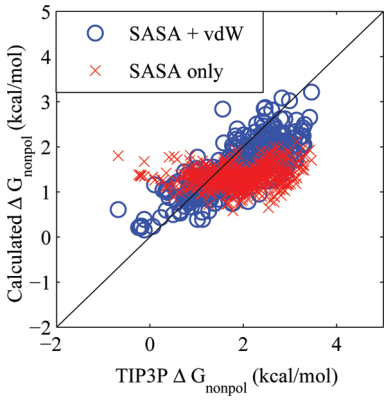

Figure 3. Correlation plots between the "SASA only" and "SASA + vdW" models of the nonpolar solvation energies (ΔG_{nonpol}) and the corresponding explicit solvent (TIP3P) results from ref 8.

Table 4. RMSDs (kcal/mol) and Correlation Coefficients (r^2) Obtained by The SASA + vdW and SASA-only Procedures with Respect to Experimental ΔG_{solv} Values and TIP3P Computed Values of ΔG_{nonpol}

	SASA + vdW		SASA-only	
	RMSD	r^2	RMSD	r^2
ΔG_{solv} Relative to Experiment				
entire data set	1.20	0.86	1.16	0.86
alkanes only	0.41	0.66	0.79	0.50
ΔG_{nonpol} Relative to TIP3P				
entire data set	0.83	0.60	0.83	0.04
alkanes only	0.78	0.37	1.23	0.19

To further investigate this trend, we have calculated solvation energies of 35 alkane molecules included in the data set. In this type of molecule the nonpolar contribution to the total solvation free energy is no longer considerably smaller than the polar component, so that any inaccuracy in ΔG_{nonpol} will produce a noticeable effect on the accuracy of the computed total solvation energy. These results, see Table 4, show that for these very nonpolar compounds SASA + vdW provides a distinctly better degree of correlation (relative to SASA-only term) with experimental and TIP3P based values of ΔG_{nonpol}. The loss of correlation of the simpler SASA-only model

reported here is consistent with previous observations,[31] in which the SASA-only model was used for the calculation of solvation energies of alkane molecules.

3.3. Computational Performance Analysis. The computation of the total solvation free energy of the 504 molecules takes approximately 20 s (on average 0.04 s per molecule), on a commodity PC with a Pentium(R) Dual-Core CPU T4200 2 GHz Intel processor and 2 GB of RAM memory. Most of the computational time is consumed by the R6 effective Born radii calculation (50%), followed by the combined processes of surface triangulation and SASA calculation (40%), which are performed by MSMS. The rest of the computational time is consumed by relatively faster processes, such as evaluation of eq 1 and computation of the electrostatic size. Note that once the computation of polar free energy is finished, the van der Waals term is obtained at virtually no additional cost because the effective Born radius of each atom, R_i, in eq 1 has been already computed.

Any detailed and exhaustive comparison of currently available methods is outside the scope of this work. Moreover, our experience suggests that when comparing speeds of very different algorithms, seeking to achieve a distinction more accurate than about a factor of 2 may not be prudent. The speed of any particular implementation of a complex algorithm can always be improved by clever optimizations at various levels. Accordingly, here we use performance data from a recent study that provide a general performance comparison sketch of different methodologies to compute solvation free energies of small molecules, see Table 5. Readers still interested in precise timings for some of the methods we tested in the course of this study are referred to the Supporting Information. We notice that, being fully implicit, GB_NSR6 is expectedly faster than fully or semi-explicit methods. Analytical GB flavors, such as the widely used GB_HTC[36] from the Amber package,[37] are several times faster than the numerical GB_NSR6 but at the expense of a substantial reduction of the accuracy, see Table 5.

Table 5. Average Computational Time Per Molecule and RMSD to Experiment for the Calculation of Total Solvation Energies of 504 Neutral Small Molecules[a]

methodology	computational time	RMSD to experiment
fully explicit solvent (TIP3P)	hours to days	1.2 kcal/mol
semi-explicit (SEA)	around a second	1.3 kcal/mol
numerical implicit solvent (GB_NSR6)	few tens of milliseconds	1.2 kcal/mol
analytical implicit solvent (GB_HCT)	few milliseconds	2.5 kcal/mol

[a]The data for TIP3P and SEA performance are from ref 35. Computations performed on a single processor commodity PC.

3.4. Comparison between GB_NSR6 and PB Methods. Here we examine the accuracy of the GB_NSR6 and the numerical PB treatment on the same footing, relative to both experiment and explicit solvent model (TIP3P). The RMSD of the computed ΔG_{solv} and ΔG_{pol} free energies relative to experiment and TIP3P, respectively, is shown in Figure 4. In all of these calculations the SASA-only model was used to compute ΔG_{nonpol}. Overall, these results show that for an optimum radii set, GB_NSR6 is at the same level of accuracy as the PB model.

Within the PB model, (squares, Figure 4) ZAP9 radii also perform better than the other two radii sets. This is expected since ZAP9 was optimized for numerical PB using experimental solvation energies.[6] Considering the optimum ZAP9 radii set, GB_NSR6 and the PB produce almost the same RMSD relative to experiment, 1.2 and 1.21 kcal/mol, respectively. Relative to explicit solvent calculations (TIP3P), GB_NSR6 appears to be slightly more accurate than the PB. However, the difference between GB_NSR6 and the PB (RMS error 0.9 kcal/mol) is smaller than the error of the TIP3P explicit solvent model relative to experiment (RMS error 1.26 kcal/mol), and so the GB_NSR6 vs PB difference here should not be viewed as significant.

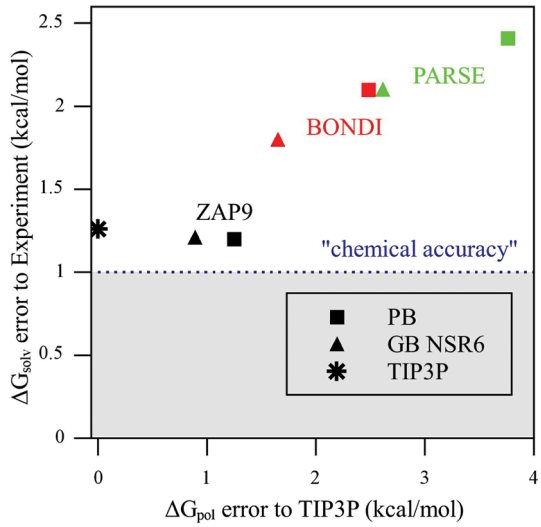

Figure 4. RMS errors of the computed total ΔG_{solv} and polar ΔG_{pol} relative to experiment and explicit solvent models (TIP3P), respectively. The PB (squares) and GB_NSR6 (triangles) methods are used with ZAP9 (black), BONDI (red), and PARSE (green) radii sets. The shaded region represent the desirable "chemical accuracy" of 1 kcal/mol relative to experiment.

For suboptimal radii sets (BONDI and PARSE), the RMS errors of both the PB and GB_NSR6 solvation energies are substantially larger than those produced by ZAP9. A noticeable difference also exists between the GB_NSR6 and the PB results, see the green and blue symbols in Figure 4. There may be many reasons for the systematic errors produced by the BONDI and PARSE radii sets; we discuss one concrete conjecture in the Supporting Information.

4. CONCLUSIONS

In this work we have evaluated the performance of a recently developed model, GB_NSR6,[18] which is based on the so-called "R6" flavor of the generalized Born implicit solvent approximation. Our main motivation was to test how well the parameter-free GB_NSR6 can perform on small molecules relevant to drug-design efforts where efficient and accurate computation of solvation energies is key. We have used a common data set of 504 small molecules with available experimental and explicit solvent total solvation energies.[8] For this particular data set, GB_NSR6 produces a RMSD in ΔG_{solv} of 1.2 kcal/mol relative to experimental solvation energies. This level of accuracy is the same as that (1.26 kcal/mol) obtained by computationally much more expensive MD simulations with explicit (TIP3P) solvent model. The ~1 kcal/mol accuracy limit for small molecules ΔG_{solv} is the current expectation for models that share the same underlying physics with the Poisson treatment.[10] While a higher accuracy may be achieved for

specific data sets at the expense of a large number of adjustable parameters,[11] transferability will likely be affected. The proposed GB_NSR6 is free from this defect as the electrostatic part of ΔG_{solv} is computed with no parameters (and the nonpolar part has little affect on the overall accuracy). Thus, we are confident that the achieved accuracy will be relevant beyond the specific test set we used.

Comparing computational efficiencies between very different methods (e.g., fully explicit solvent, semi-explicit[35] or fully implicit, such as the GB) is always difficult. However, rough order-of-magnitude comparisons can still be made, especially if based on the same test set.[35] In this respect, GB_NSR6, which is based on numerical integration to obtain effective Born radii, is secondary only to optimized analytical GB models, such as those available in AMBER. The latter is still several times faster than GB_NSR6 but at the expense of substantial accuracy loss.

The polar component of the solvation energy is computed with no adjustable parameters by GB_NSR6, producing a RMSD of 0.89 kcal/mol relative to explicit solvent treatment. This degree of accuracy is similar to those obtained by other recent implicit or hybrid methods, such as the one proposed by Fennell et al.[35] Our results show that the intrinsic radii set, which defines the molecular surface, plays a key role in the accuracy of the polar part of solvation energy calculations. The ZAP9 radii set recently proposed by Nicholls et al.[6] provides more accurate results than other standard radii set, such as BONDI and PARSE, when compared to the TIP3P explicit solvent model. For this optimal set, both the total and polar solvation energies obtained by GB_NSR6 and the PB treatment are very similar. These observations warrant further investigation to evaluate the performance of ZAP9 radii set for larger molecules such as nucleic acids or proteins, as its transferability outside of small molecules is by no means guaranteed.

In this work, the nonpolar component of the solvation free energy is calculated as the sum of two components, the cavity and vdW terms as proposed by Gallicchio and Levy.[32] Once ΔG_{pol} is computed, the vdW term is calculated with almost no cost, since it depends on the R6 Born radii obtained previously in the computation of ΔG_{pol}. Our results show that the RMSD of this methodology (relative to experimental data and TIP3P results) is similar to that obtained when ΔG_{nonpol} is calculated as being proportional to the Solvent Accessible Surface Area. However, the Gallicchio et al. methodology provides a better correlation with TIP3P based ΔG_{nonpol}, and experimental solvation free energies of alkane molecules, in which the nonpolar component is predominant, see Table 4.

Overall the GB_NSR6 model used in this work shows competitive performance when applied to a data set of 504 small molecules, providing a good balance between computational efficiency and accuracy. In our opinion, interested parties may benefit from these features especially for fast computation of large data set of molecules. Another useful feature of this GB flavor is that it is not optimized for any specific radii set and thus can be used to evaluate relative performances of different radii sets. All of the software developed in this work is freely available from http://people.cs.vt.edu/onufriev/software.php.

■ ASSOCIATED CONTENT

Supporting Information

Four additional tables containing the names of the molecules in the training set, additional optimized set of parameters (μ_i and γ), and the computational timing of some of the methods used to compute solvation energies. This material is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

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

Notes

The authors declare no competing financial interest.

■ 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.; L., M. D.; Brown, S. P. In *Structure Based Drug Design*, 1st ed.; Merz, K. M., Ringe, D., Reynolds, C. H., Eds.; Lecture Notes in Computer Science; Cambridge University Press: New York, 2010; pp 61−85.
(4) Gilson, M. K.; Zhou, H. X. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21−42.
(5) Llinàs, A.; Glen, R. C.; Goodman, J. M. *J. Chem. Inf. Model.* **2008**, *48*, 1289−1303.
(6) Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. *J. Med. Chem.* **2008**, *51*, 769−779.
(7) Shivakumar, D.; Deng, Y.; Roux, B. *J. Chem. Theory Comput.* **2009**, *5*, 919−930.
(8) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. *J. Chem. Theory Comput.* **2009**, *5*, 350−358.
(9) Nicholls, A.; Wlodek, S.; Grant, J. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 293−306.
(10) Nicholls, A.; Wlodek, S.; Grant, J. A. *J. Phys. Chem. B* **2009**, *113*, 4521−4532.
(11) Labute, P. *J. Comput. Chem.* **2008**, *29*, 1693−1698.
(12) Mobley, D. L.; Dill, K. A.; Chodera, J. D. *J. Phys. Chem. B* **2008**, *112*, 938−946.
(13) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 265−284.
(14) Onufriev, A. In *Modeling Solvent Environments*, 1st ed.; Feig, M., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2010; pp 127−165.
(15) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−152.
(16) Feig, M.; Brooks, C. L. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217−224.
(17) Knight, J. L.; Brooks, C. L. *J. Comput. Chem.* **2011**, *32*, 2909−2923.
(18) Aguilar, B.; Shadrach, R.; Onufriev, A. V. *J. Chem. Theory Comput.* **2010**, *6*, 3613−3630.
(19) Tjong, H.; Zhou, H. X. *J. Phys. Chem. B* **2007**, *111*, 3055−3061.
(20) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.
(21) Mongan, J.; Svrcek-Seiler, A.; Onufriev, A. *J. Chem. Phys.* **2007**, *127*, 185101−185101.
(22) Sigalov, G.; Fenley, A.; Onufriev, A. *J. Chem. Phys.* **2006**, *124*, 124902.
(23) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(24) Svrcek-Seiler, A. *personal communication*, 2001
(25) Grycuk, T. *J. Chem. Phys.* **2003**, *119*, 4817−4826.
(26) Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996**, *38*, 305−320.
(27) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441−451.
(28) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(29) Tan, C.; Tan, Y.-H.; Luo, R. *J. Phys. Chem. B* **2007**, *111*, 12263−12274.
(30) Wagoner, J. A.; Baker, N. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8331−8336.
(31) Zacharias, M. *J. Phys. Chem. A* **2003**, *107*, 3000−3004.
(32) Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2004**, *25*, 479−499.
(33) Nelder, J. A.; Mead, R. *Comput. J.* **1965**, *7*, 308−315.

(34) Grant, J. A.; Pickup, B. T.; Nicholls, A. *J. Comput. Chem.* **2001**, *22*, 608−640.
(35) Fennell, C. J.; Kehoe, C. W.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3234−3239.
(36) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824−19839.
(37) Case, D. A. et al. *AMBER 9*; University of California: San Francisco, CA, 2006.