J Comput Aided Mol Des (2010) 24:361–372
DOI 10.1007/s10822-010-9339-3

Predictions of hydration free energies from continuum solvent with solute polarizable models: the SAMPL2 blind challenge

Alexandre Meunier · Jean-François Truchon

Received: 5 December 2009 / Accepted: 16 March 2010 / Published online: 31 March 2010

Abstract This paper reports the results of our attempt to predict hydration free energies on the SAMPL2 blind challenge dataset. We mostly examine the effects of the solute electrostatic component on the accuracy of the predictions. The usefulness of electronic polarization in predicting hydration free energies is assessed by comparing the Electronic Polarization from Internal Continuum model and the self consistent reaction field IEF-PCM to standard non-polarizable charge models such as RESP and AM1-BCC. We also determine an optimal restraint weight for Dielectric-RESP atomic charges fitting. Statistical analysis of the results could not distinguish the methods from one another. The smallest average unsigned error obtained is 1.9 ± 0.6 kcal/mol (95% confidence level). A class of outliers led us to investigate the importance of the solute–solvent instantaneous induction energy, a missing term in PB continuum models. We estimated values between -1.5 and -6 kcal/mol for a series of halo-benzenes which can explain why some predicted hydration energies of non-polar molecules significantly disagreed with experiment.

Keywords EPIC · Poisson-Boltzmann · Implicit solvent · Polarizability · Hydration energy · Solvation energy · SAMPL2

Electronic supplementary material The online version of this article (doi:10.1007/s10822-010-9339-3) contains supplementary material, which is available to authorized users.

A. Meunier · J.-F. Truchon (✉)
Merck Frosst Canada, 16711, TransCanada Hwy, Kirkland, QC H9H 3L1, Canada
e-mail: jeanfrancois_truchon@merck.com

Introduction

This article reports the results for the prediction of free energy of hydration on the SAMPL2 blind challenge [1]. The main focus is to assess the usefulness of a solute-polarizable model in predicting hydration free energies in the context of a Poisson-Boltzmann implicit solvent approach. The ability to predict accurately free energies of hydration is of primary importance in the field of drug design. The desolvation process in protein–ligand association largely contributes to the free energy of ligand binding [2]. The hydration process, which consists of transfering a molecule from gas phase to pure water can be viewed as a simplification that decouples the solvent effects from the protein–ligand interactions. The importance of the solvent contribution to protein–ligand association often takes the form of a desolvation penalty, the loss of activity from the presence of a polar chemical moiety forced to bind into a hydrophobic protein site. Solvent effects play a major role in explaining the structure–activity relationship (SAR) in drug discovery. In spite of the apparent simplicity of the evaluation of solvent effects, theoretical approaches have a limited quantitative ability to predict hydration energy producing errors ranging from 2.4 to 3.6 kcal/mol in a blind prediction study [3].

Undertaking the challenge of improving the accuracy of theoretical methods, we evaluated the performance of a newly published model that brings theoretical advantages. The electronic polarization from the internal continuum (EPIC) model was shown to be accurate in predicting molecular polarizability tensors, liquid state refractive indices and free energy of hydration on a wide variety of bioorganic molecules. One advantage of this approach is to bring within the same theoretical framework the solvent polarization and the solute polarization although those

phenomena occur at very different time scale and have different origins.

This attempt at the SAMPL2 blind challenge aims to compare different related methods based on implicit solvation. First, because we keep the solvent model constant, we will be comparing the effects of the charging methods. We designed the calculations in order to assess the impact of: using multiple conformations, minimizing the geometry at different quantum mechanics (QM) level of theory, using a polarizable solute and QM solute polarizability versus EPIC. Given that a previous study showed that a polarizable solute was not necessary to achieve good correlation with experimental hydration energies in a purely training setup, it is of interest here to assess the usefulness of a polarizable solute model such as EPIC in a prospective blind challenge. It is important to highlight here that no parameter optimization was done in this study.

In the remainder of this article, we will give a brief overview of the methods employed, provide the computational conditions in the Sect. "Method" and report the results along with comparisons and some lessons learned. In particular, we highlight an important defect when using an implicit solvent approach together with any polarizable solute and quantify it through appropriate simulations.

Theory

Implicit solvent

The implicit solvent approach taken here has been applied to model the effect of solvent polarization by many [4–13]. The implicit solvent approach can be justified by the partitioning of the free energy of hydration ΔG_{hyd}, into a polar term (ΔG_{polar}) and a non-polar term (ΔG_{npolar}) such as

$$\Delta G_{\text{hyd}} = \Delta G_{\text{polar}} + \Delta G_{\text{npolar}} \quad (1)$$

The non-polar term is then usually approximated, by analogy with the macroscopic surface tension, as

$$\Delta G_{\text{npolar}} = \gamma \cdot S \quad (2)$$

where S is the molecule surface in Å^2 and γ is the tension surface at the molecular level in kcal/mol/Å^2. This term has the advantage of being simple but certainly constitutes an oversimplification of the physics behind the non-polar terms such as the hydrophobic non-polar effects, the solute–solvent dispersion interaction energy, etc. However, the polar component is treated with a more detailed model that includes the water polarization due mainly to water configurational averaging. With Poisson's equation, the macroscopic idea of a dielectric response is applied at the molecular level where the source for the water polarization is the solute electric field. This approach uses the experimental bulk water low frequency dielectric to calculate the average response from water. The task is then to solve Poisson's equation

$$\vec{\nabla} \cdot \left[\varepsilon(\vec{r}) \vec{\nabla} \varphi(\vec{r}) \right] = \sum_{i=1}^{N} q_i \delta(\vec{r}_i - \vec{r}) \quad (3)$$

where $\varepsilon(\vec{r})$is the dielectric function that varies through space, $\varphi(\vec{r})$ is the electrostatic potential, q_i is one of the N point charges and $\delta(\vec{r}_i - \vec{r})$ is a Dirac delta function needed since the charge density is a point. When Eq. 3 is solved, for instance with a finite difference solver, the polar hydration energy can be obtained from the classical electric energy difference equation

$$\Delta G_{\text{polar}} = \frac{1}{2} \sum_{i=1}^{N} q_i \left[\varphi^{\text{wat}}(\vec{r}_i) - \varphi^{\text{gas}}(\vec{r}_i) \right] \quad (4)$$

where $\varphi^{\text{gas}}(\vec{r}_i)$ is the gas-phase electrostatic potential at the position of the atomic charge i in the absence of solvent response, and the $\varphi^{\text{wat}}(\vec{r}_i)$ is the electrostatic potential in the presence of the solvent dielectric medium. With this use of Poisson's equation, the dielectric function plays a central role. When the dielectric function is set to the vacuum permittivity value, the solution to Eq. 3 corresponds to the gas phase state. A solvated state is defined by a solute cavity with a dielectric value outside a certain exclusion volume set to the water low frequency relative permittivity (~ 80). Most of the time, this exclusion volume is defined by hard spheres with a certain radii centered on the atoms. In this work, we used a Gaussian smooth boundary first implemented in Zap [14]. This functional form greatly empowers the efficiency of multigrid solvers to lead to very fast methods [8] and was shown to lead to accurate free energy of hydration [9, 11, 13, 15].

EPIC and the 3-zone dielectric

The correct dielectric value to assign inside the solute cavity is a matter of debate and controversy in the literature [16–19] and recent work shed a clarifying light on the subject. Truchon et al. [13] were able to use Poisson's equation to reproduce gas phase DFT molecular polarizability tensors accurately with a small number of parameters for many molecules with a large variety of functional groups. The idea was to find the appropriate dielectric function to use in Eq. 3. As expected, the inter-dependency of the atomic radii and internal dielectric value was crucial and permitted only a subset of useful values. Curiously, in the "electronic polarization from the internal continuum" (EPIC) approach, it was necessary for the dielectric function to adopt a relatively high value (>8) inside the molecular cavity and atomic radii typically much smaller than the regular contact radii used in solvent PB approaches. In contrast with solvent

polarization usually modeled with the implicit solvent paradigm, EPIC uses the dielectric function to model the electronic polarization in the presence of an external electric field. Although such parameters contradict experimental high frequency dielectrics ($\sim$2) derived from refractive indices, it was shown that sets of radii and dielectric able to accurately predict gas phase polarizability tensors. The same parameters could also accurately reproduce liquid phase refractive indices for a wide range of small molecules. They argued that the interstices in between molecules where electronic density is negligible command higher internal dielectric values to average out to the experimental values. The smaller atomic radii they found rather corresponded to covalent radii.

Of greater interest to this work, the EPIC model could be used to reproduce accurate free energies of hydration [13] by combining in the same dielectric function both the electronic and water polarization features. However, two major modifications to the usual Poisson-Boltzmann approaches were necessary. Firstly, the smaller radii obtained with EPIC cannot be used directly to define a solvent cavity. For this reason, a third zone was added to the usual 2-zone "inner" and "outer" dielectric model. Conceptually, this third zone represents a low-dielectric coating around the small high-dielectric core of the solute (responsible for solute polarization) and the high-dielectric solvent (responsible for solvent polarization). In the 3-zone systems, the dielectric constant varies from ε_{in} at the EPIC polarization boundary down to 1 in the "third-zone" interstice and then back up to the solvent ε_{out} at the solvent boundary. This way, the solute can polarize the water continuum which in turns polarizes the solute electrons all by solving the same Poisson's equation. This contrasts with previous polarization models, such as the point inducible dipoles [20], that require much complicated mathematics to be coupled with implicit solvent. An illustration of the 3-zone dielectric model applied to the isothiazole molecule is given in Fig. 1. The second delicate point concerned the atomic partial charges that needed to be derived such that in the context of an internal dielectric larger than one, the permanent electric potential and field felt by the solvent continuum correspond to the gas phase value. This significantly changes the atomic partial charges compared to conventional charges (with inner dielectric of one) due to dielectric shielding inside the solute. This is necessary to leave to the solute and solvent polarizabilities the task of appropriately enhancing the molecule polarity. Truchon et al. (see Sect. "Method") showed that by reformulating the equations in a more general way, the usual gas phase QM ESP-fit can be performed. In this paper, we will compare models with and without solute polarization. It may be useful to mention that a non-polarizable solute requires atomic charges reflecting the increased polarity due to water solvation whereas a polarizable solute needs gas-phase derived atomic charges obtained, for example, from a fit to an accurate gas-phase wavefunction. In this work, the AM1-BCC charges and HF/6-31G* derived charges (with increased polarity compared to gas phase) should be used with a non-polarizable solute since they already include the desired level of polarization. The DFT-derived charges (having only gas phase polarity) should be used with a polarizable solute model.

Method

Geometry optimization

The geometry optimization was carried out after transforming the SMILES [21–23] into three-dimensional objects with the Omega [24] software that also generated conformations. Additional conformations for diflunisal were built by hand. Each conformer was minimized with Schrödinger Macromodel software and the MMFFs force field with the use of a dielectric constant of 10 to avoid internal electrostatic interactions that would be pre-polarizing the atomic charges. Before QM ESP calculations on a grid for charge fitting, an additional full minimization was carried out with the Gaussian 03 software [25] with a 6–31+G(d, p) basis set both for the B3LYP [26–28] and Hartree–Fock (HF) level of theory. All conformations within an energy window of 0.1 kcal/mol were represented by one conformer.

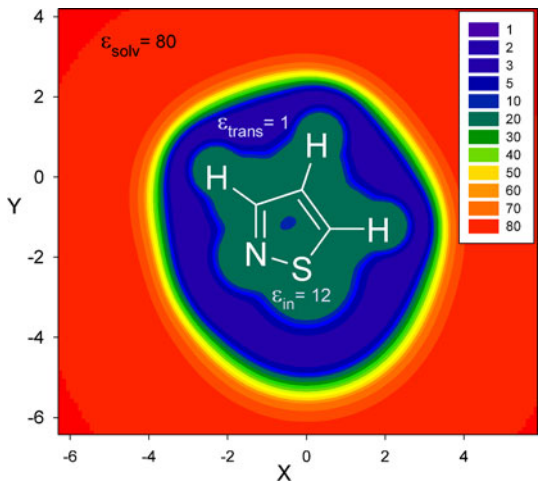

Fig. 1 The 3-zone dielectric model applied to isothiazole. The dielectric function outside the solute cavity is 80, it transitions to the vacuum dielectric value of one and reaches a dielectric value of 12 intramolecularly which introduces the electronic polarization of the EPIC model

Charging methods

AM1-BCC charges were obtained following the Jakalian et al. [29, 30] procedure with the use of in-house tools. The topologically equivalent atoms had their charges averaged and conformationally averaged whenever applicable. ESP-fitted charges were also generated. A face-centered cubic grid was used with points located outside the van der Waals (vdW) surface at 1.4–2.0 times the vdW radius of each atom [29–31]. The grid point spacing was set to 0.5 Å. For each conformer, the ESP was calculated on the grid at the HF/6-31G* and B3LYP/cc-pVTZ level of theory. From those two sets of ESPs we generated 3 charge sets: RESP, DESP and DRESP. The RESP, DESP and DRESP charges came from B3LYP while the RESP charges come from HF. In DESP and DRESP, the *D* stands for dielectric and it means that the G1-12 dielectric function was used in fitting the charges following the procedure outlined by Truchon et al. [32]. In RESP, DRESP and DESP, the charges were fitted to minimize the root mean square deviation (RMSD) from the QM ESP. Constraints to neutralize the total charge was always applied and topologically equivalent atoms were constrained to bear the same charges.

In DRESP and RESP, we applied an additional hyperbolic restraint to take advantage of the Bayly et al. [33] RESP fitting procedure. Reference charges of zero were used for the restraint. Following the usual 2-stages fitting [33] approach, we applied a 3-stages fitting workflow for both DRESP and RESP. The first stage consists in fitting all charges with a weak hyperbolic restraint weight of 0.0005 a.u. [34]. and only constraining the net charge to zero. In the second stage, all polar atoms (not alkyl carbon and not alkyl hydrogen) without topologically equivalent atom have their charges kept frozen from stage 1. Also in this stage, topologically equivalent polar atoms are constraint to have the same optimal charges. Again, a weak restraint weight of 0.0005 a.u. is applied. In the third stage, all polar atom charges are frozen and all non-polar atom charges are fit by enforcing topological equivalence. In this last stage, a stronger restraint weight of 0.001 a.u. is applied. The difference from the original Bayly et al. [33]. work is the addition of an intermediate step to include topologically equivalent polar atoms. Although this fitting strategy has been shown to be appropriate with a gas phase dielectric medium ($\varepsilon = 1$), it is not obvious that the same value of the restraining weight should apply to DRESP. We devote a section of this paper to demonstrate it along with the huge resulting partial charge magnitude reduction. Finally, the charges from the multiple conformations were simultaneously fitted [29, 30, 35].

IEF-PCM

Free energies of hydration were also computed at the B3LYP level of theory with a cc-pVTZ [36] basis set through the Gaussian 03 implementation of the IEF-PCM self consistent reaction field [37–39]. The default UA0 radii [39] in Gaussian 03 were used. Tight SCF convergence was required, with nosymmetry and performing only single points. In this approach, the electronic structure calculations provide an accurate solute description and a continuum-based water model is used to include the solvent polarization. The mutual solvent–solute polarization is brought to self consistency.

Poisson

All the finite difference solutions to Poisson's equation were obtained with in-house code [13, 40] which implements the 3-zone implicit solvent model. The reported G2-12 SA parameters [13] were used to define the solvent boundary part of the dielectric function. The previously reported G1-12 parameters [13] for the solute polarizability were used with $\varepsilon_{in} = 12$. However, the iodine and phosphorous atoms were omitted from the previous parameterization. As a necessary approximation, the inner atomic radii (G1), which determine the solute polarization volume, were set to zero for oxygen bonded to P, P and I atoms. The outer atomic radii (G2), which determine the solute/solvent boundary, were set to zero for phosphorus since the oxygen radii were sufficient to create the cavity; the iodine radius was made equal to the bromine radius. Both G1-12 and G2-12 parameter sets can be used separately or together since they each model different polarization phenomena and the G2-12 was shown to be transferable to different charge sets such as RESP [13] and AM1-BCC (results not shown). The grid resolution was set to 0.35 Å between grid points, with a buffer of 7 Å between the molecular surface and the box boundary. The solvent dielectric was set to 80. The obtained polar hydration energies ΔG_{polar} were arithmetically averaged over the different conformations with the exception of one MMFFs [41–44] calculation (identified as M4 in Table 1) where only the lowest energy conformation was used. Lastly, the surface area was calculated using the Zap toolkit [14] based on the molecular surface generated using the atomic Bondi radii [45] and a probe radius of 1.4 Å.

Average instantaneous induction energy

Molecular dynamics simulations were performed with benzene, hexafluorobenzene, hexachlorobenzene and hexabromobenzene solvated by TIP3P [46] water in a

Table 1 Average unsigned and signed error obtained with the 6 methods of this study

	M1	M2	M3	M4	M5	M6
Geom.	B3LYP	HF	MMFFs	HF	B3LYP	B3LYP
Charges	DRESP	HF-RESP	AM1-BCC	AM1-BCC	N.A.	DESP
Polarizable	EPIC	NO	NO	NO	B3LYP	EPIC
ΔG_{hyd}	P/conf[a]	P/conf	P/conf	P	IEFPCM/conf	P/conf
Explanatory (8 molecules)						
AUE	3.07	2.61	3.13	2.92	5.86	3.18
ASE	1.88	1.98	1.90	1.66	5.36	1.67
Obscure (23 molecules)						
AUE	2.00	2.99	1.57	1.51	1.97	1.73
ASE	1.90	2.95	0.81	0.79	0.51	1.45
Obscure select (10 molecules from the obscure set)						
AUE	2.40	3.54	1.43	1.35	2.74	2.06
ASE	2.40	3.53	1.06	0.97	1.78	1.79
Overall (31 molecules)						
AUE	2.27	2.89	1.97	1.87	2.98	2.10
Std	2.25	2.74	1.84	1.65	2.95	2.26
ASE	1.89	2.70	1.09	1.01	1.76	1.51
RMSD	3.20	3.99	2.70	2.50	4.20	3.09
AUE[b]	1.84	2.40	1.78	1.70	2.80	1.69
Std[b]	1.55	1.99	1.72	1.57	2.98	1.57
ASE[b]	1.44	2.19	0.84	0.79	1.50	1.05
RMSD[b]	2.42	3.12	2.47	2.32	4.09	2.30

[a] P means using Poisson implicit solvation model (EPIC) and conf means that the hydration energy averaged over the conformers

[b] Omitting the d-glucose and d-xylose, the major outliers

periodic boundary box within the NPT ensemble. A production MD run over 100 ps was carried out after a 100 ps equilibration run. The GAFF force field [47] in AMBER 8.0 software was employed. During the production run, 17 evenly distributed snapshots were kept for further analysis. All water molecules with their oxygen atom within 8 Å of any atom of the solute were kept to form a large water cluster around the solute. The EPIC G1-12 parameters were assigned on the solute only and all the solute atomic partial charges were set to zero while the TIP3P partial charges were assigned on the water molecules. The work of moving an uncharged polarizable solute into the cavity from infinity was assessed by solving Poisson's equation with a uniform dielectric of one (no solute electronic polarization) and with the EPIC solute dielectric (including the solute electronic polarization). The electrostatic interaction energy was evaluated by using Eq. 4 that required the evaluation of the ESP at the position of all water charges for both situtations. We calculate the induction energy by taking the difference between the polarized and the unpolarized systems. This quantity is time averaged, noted $\langle E^0_{ind} \rangle$ and called the average instantenous induction energy.

Results

DRESP

As previously reported [32], the use of an internal dielectric larger than one in EPIC necessitates charges sometimes significantly larger than normally expected. In the same study, Truchon et al. mentioned that restraints could also be applied on the fitted charges (see the Sect. "Method" for further details). Since we applied RESP-like restraints to the dielectric based ESP-fit in this study, it was necessary to find the restraint weight that can be used with the G1-12 polarizable solute. With the dielectric ESP fitting, two effects are combined to produce unusually large atomic charges: the inherent numerical instability of the ESP fitting problem [33] and the internal dielectric polarization that shields the atomic charges [32]. By restraining the charges towards lower magnitudes, the impact of the numerical instability is greatly reduced and the magnitude of the DRESP-fitted charges will be principally due to the higher inner dielectric. Figure 2 shows that for the 10 molecules examined, a restraint weight less than 0.001 has a minimal impact on the quality of fit. Indeed, when

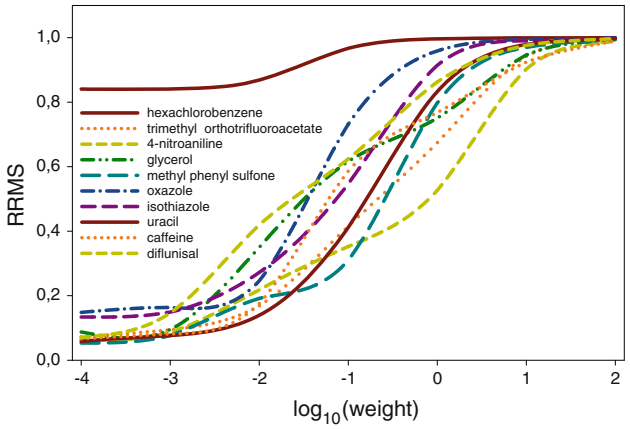

Fig. 2 Variation of the relative root mean square deviation (RRMS) as a function of the restraint weight used in a one stage DRESP fit. A value less than 0.001 seems to minimally impact the quality of fit to the B3LYP/cc-pVTZ electrostatic potential

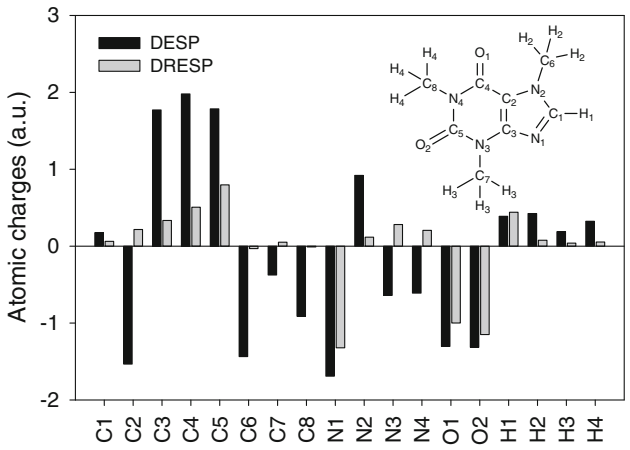

Fig. 3 Atomic charges assigned to the caffeine molecule when fitted [32] to reproduce the B3LYP/cc-pVTZ electrostatic potential with (DRESP) and without (DESP) magnitude restraints. The same EPIC G1-12 parameters [13] were applied in both cases. All buried atoms exhibit very important charge changes while the electrostatic potential on the grid is of the same quality

comparing the ESP relative root mean square deviation between DESP and DRESP, one find a R^2 value of 0.98, a slope of 1.08 and an ordinate to the origin of −0.02, showing that the quality of fit of the ESP is maintained with DRESP. However, as shown for the caffeine molecule in Fig. 3 and more generally for all the atomic charges obtained for the 10 molecules (see Fig. 4), the charge magnitudes are greatly reduced by the application of restraints. A weight of 0.001 in the final stage of the 3-stage DRESP fitting is appropriate. Surprisingly, this weight corresponds to the early findings of Bayly et al. [33]. that did not have the extra complication of a non-uniform dielectric constant bathing the charges. Finally, it is worth

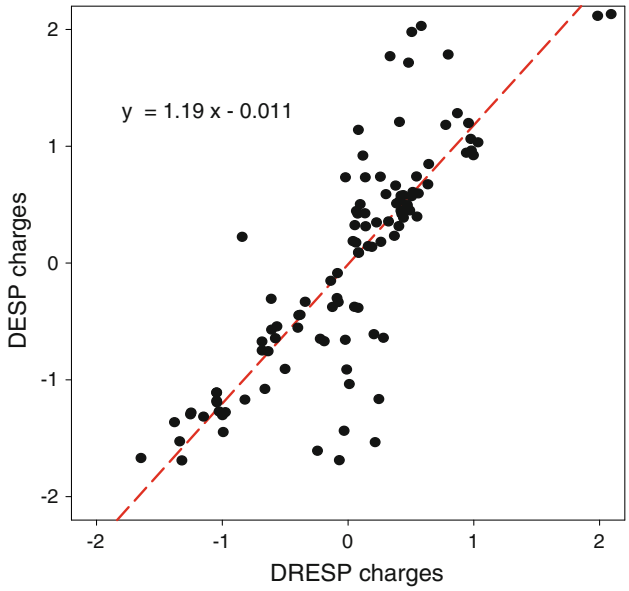

Fig. 4 Comparison of the atomic partial charge magnitude on the 10 molecules from SAMPL2 examined. The DESP charges are compared with the DRESP charges (3-stages fit with a restraint of 0.001) showing that DRESP significantly reduce the magnitude of the charges for an equivalent quality of fit

noting that a change in the grid point density of the grid on which the QM ESP is calculated could change the ideal value for the restraint weight [48].

General statistical considerations

Table 1 reports the average unsigned error (AUE) and the average signed error (ASE) for the six methods used and according to the different subsets from the challenge design. According to Student's *t*-tests for the comparison of two means, all the AUE means are statistically undistinguishable at the 95% confidence level. This is mainly due to the large standard deviation on the AUE also reported in Table 1. Figure 5 shows the overall AUE for each method together with the Student's *t*-test 85% interval for the mean prediction. At this lower confidence level, one can only state that the methods based on the AM1-BCC charges are better than the method based on the HF-RESP charges and IEF-PCM. The similitude between the results obtained with the different methods can be due to the common continuum-based solvent models. It may in fact be reassuring that the different solute electrostatic source (charges or electronic density) produce similar results. Moreover, the standard deviations also reported in Table 1 largely contributed to the large error intervals. This being said, one needs to appreciate that the AM1-BCC charges were more robust with smaller standard deviations of the unsigned error than the other methods. This difference seems to be explained mainly by the poor performances of the ESP-fit

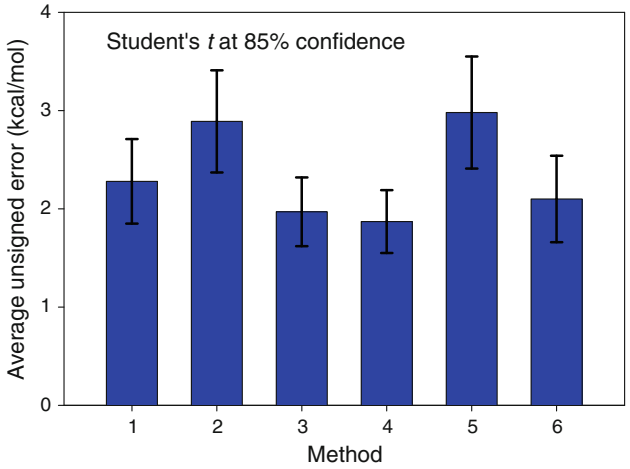

Fig. 5 The average unsigned error in kcal/mol between the experimental and calculated free energy of hydration for each of the 6 methods (M1–M6) as outlined in Table 1. The *error bars* were obtained from the single tailed Student's *t* distribution at the 85% confidence level. At 95% confidence level, the means are not statistically different

charge sets on d-glucose and d-xylose. Indeed, when comparing the errors and standard deviations of the methods without d-glucose and d-xylose, the EPIC based and AM1-BCC based solute polarity models become completely equivalent as outlined in Table 1 and shown by the correlation plots of Fig. 6. Finally, there is a systematic positive error as indicated by the ASE with the predicted hydration energy being too high. The error could be reduced by about 1 kcal/mol if all the predicted hydration energies were reduced by as much.

Training set hydration energies compared to SAMPL2

The SAMPL2 data set contains a significant number of very polar molecules not found in the Rizzo set [49] and not used in the optimization of the G2-12 radii [13] as shown in Fig. 7. This set of solvent cavity radii nevertheless extrapolates fairly well considering the results of the M3 method that uses AM1-BCC charges. In Fig. 7, the hydration energy values from EPIC (M6) seem to be systematically overestimated. This is corroborated by the systematically positive ASE reported in Table 1. We have not found any explanation, but it is likely that the extra attention paid to the quality of the experimental values used in the SAMPL2 data set may suggest that a re-optimization of the G2-12 parameters on better experimental values could be suitable.

Multiple conformations

Mobley et al. [50]. showed, with converged free energy perturbation approaches applied to an explicit solvent model, that the change in conformation during the transfer from vacuum to gas phase can be significant. However, they could not demonstrate that the inclusion of multiple conformers can improve the overall fit of an implicit solvent approach. Nicholls et al. [9]. observed the same behavior and explained their results by the compensation between the intra-molecular strain energy and the extra solvation effects of elongated conformations in water. Here, we make the same observation with different solvent cavity radii and surface area term. This can be seen in Table 1 where the single conformer approach (M4) is one of the best. It has to be noted that all the PB parameters used in previous study [9, 50] and in the current study were obtained with a single conformer based parameterization. Also, as it will be discussed below, the quality of the results may strongly depend on the chosen conformations. It may then be more robust, on average, to use multiple conformations and weight them appropriately. This still remains to be shown.

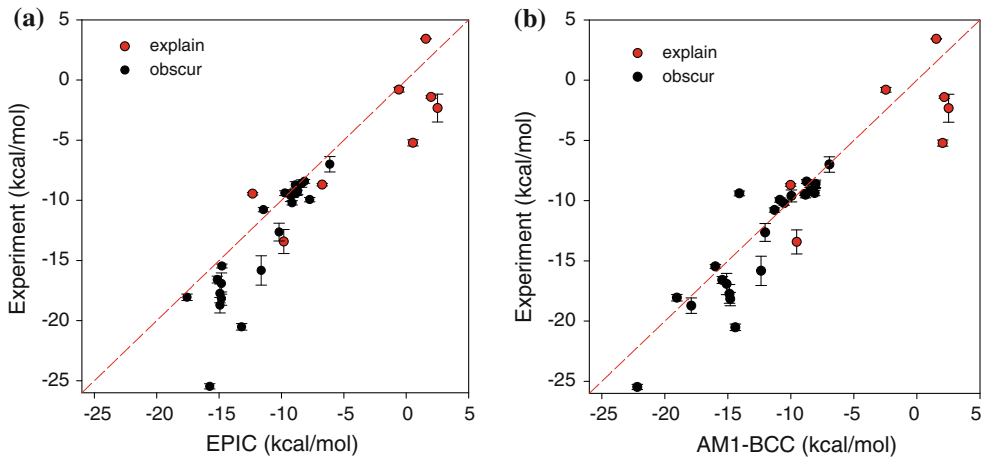

Fig. 6 Correlation plot between experimental and predicted ΔG_{hyd} for the explain and obscure sets. The M1 (Table 1) EPIC based model with DRESP charges (**a**) and the AM1-BCC charges (**b**) (M3 in Table 1) have a similar R^2 of 0.72 and 0.68 respectively

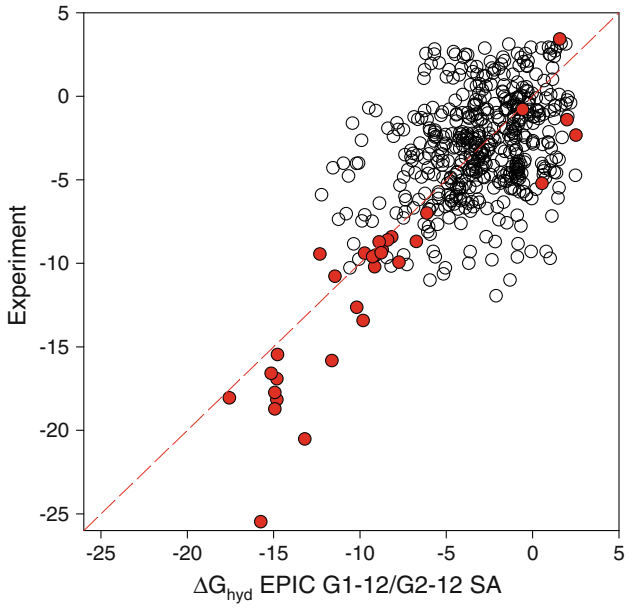

Fig. 7 Experimental ΔG_{hyd} in function of the EPIC predicted value for the original training set of the solvent cavity parameters (G2-12) shown as empty black circle and the SAMPL2 blind challenge filled in *red*. The polarity of many of the molecules in SAMPL2 goes clearly outside the G2-12 training set

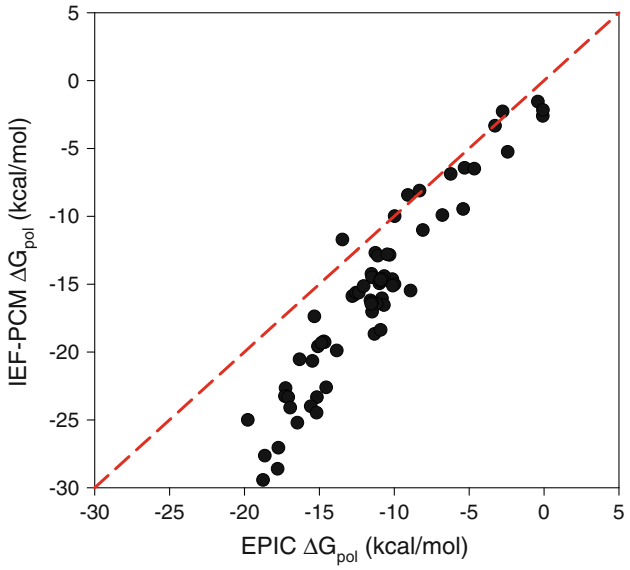

Fig. 8 Correlation between the polar hydration energies obtained with B3LYP/IEF-PCM and EPIC. The solute polarization comes from B3LYP in the former case and from EPIC with G2-12SA parameters in the latter. The implicit solvent dielectric boundaries of these approaches differ considerably. The EPIC solute polarization parameters and the atomic partial charges were derived from B3LYP

Solute polarizability

It is interesting to note, from Table 1, that the three methods based on pre-polarized atomic charges without solute polarizability (M2, M3 and M4) were not outperformed by the methods based on EPIC or IEF-PCM. They perform equally given the reported errors. The EPIC radii and atomic charges used here were obtained by fit to the B3LYP level of theory, the same used in our IEF-PCM calculations. The comparison of the solute–solvent interactions (ΔG_{polar}) energies between EPIC (M6), a purely classical solute and IEF-PCM, a purely QM solute, can lead to interesting observations. In Fig. 8, the ΔG_{polar} correlation plot includes all conformers considered in this study that are exactly the same for both methods. From Fig. 8, the solute–solvent interaction in EPIC is systematically more positive than with IEF-PCM. Even more so with hydration energies lower than −7 kcal/mol. None of these two methods was parameterized with such polar solutes. When adding the non-polar terms, IEF-PCM seems to perform better than EPIC with ΔG_{hyd} lower than −20 kcal/mol mainly because for the d-xylose and d-glucose values this in spite that the same conformers were used. However, at more positive ΔG_{hyd}, EPIC agrees slightly better with experiment. IEF-PCM uses united atom radii that omit hydrogens and this alone could explain the difference with EPIC. Unfortunately, it is hard here to isolate the cause of the difference. It is possible that a special atom typing could be needed for very polar hydrogen atoms with EPIC solvent cavity radii. The new negative hydration energies from the SAMPL2 dataset could help in improving both approaches.

Two outliers: d-glucose and d-xylose

These two molecules are the major outliers for M1, M2, M5 and M6. As shown in Table 1, withdrawing the d-glucose and the d-xylose considerably reduces the errors. Further examination of the six conformers considered showed that one was the lowest in MMFFs internal energy by at least 6 kcal/mol. The ΔG_{polar} energy ranges from −15.2 down to −18.6 kcal/mol. The conformer with the lowest internal energy forms partial intramolecular H-bonds which is an artifact from the gas phase minimization since water molecules would be efficiently competing to form H-bonds with the hydroxyl groups. If the intramolecular H-bonds are broken simply by rotating the hydroxyl hydrogens, the ΔG_{polar} energy (M1) of the right chair conformation becomes −26 kcal/mol. After the addition of the 2.2 kcal/mol from the surface area term the hydration energy becomes −23.8 kcal/mol in much better agreement with the experimental value of −25.5 kcal/mol. This rationalization reminds us of the importance of using the relevant conformation in the water phase. Because the single MMFFs ($\varepsilon = 10$) conformer used with AM1-BCC charges was not followed by gas phase QM minimization, the intramolecular H-bond network was not as well organized and lead to a better hydration energy estimates of −22 kcal/mol. Retrospectively, a better paradigm would have been to freeze all torsions involving a polar atom in the QM minimization

process. If additional computational work is not an issue, QM minimization with a reaction field like IEF-PCM can be an interesting alternative. Regarding the d-xylose, we realized that the α-D-xylopyranose form was used which constitutes less than 1% of the isomers present in water. It would have been more appropriate to use the right proportion of the 2 major furanose forms. Using the most abundant D-xylofuranose form and by preventing intramolecular H-bonds to form, the M1 predicted hydration energy becomes −21.9 kcal/mol. The experimental value is −20.52 kcal/mol and the previously predicted solvation energy was −13.2 kcal/mol. This can explain why all methods were overestimating the hydration energies by at least 3.6 kcal/mol for xylose.

Solute–solvent instantaneous induction

The hexachlorobenzene molecule raises fundamental questions about the validity of the continuum solvent models for less polar molecules and, by extension, to non-polar components of globally polar molecules. Indeed, the experimental hydration energies are −0.8 kcal/mol and −2.3 kcal/mol for benzene and hexachlorobenzene respectively in spite of the rather low polarity of these species. As shown in Fig. 9, the atomic charges on the carbon atoms vary from −0.13 e to 0.03 e. This leads to a negligible $\Delta G_{\text{polar}} = -0.08$ kcal/mol for C_6Cl_6. Added to this is the hydrophobic effect, through the surface area term, that inevitably makes ΔG_{hyd} positive. The missing attractive non polar dispersion interaction energy between the water molecules and the solute could certainly play a role. Here, we explore a term that should, in principle, exacerbate the problem with the use of gas phase atomic charges with a polarizable solute model.

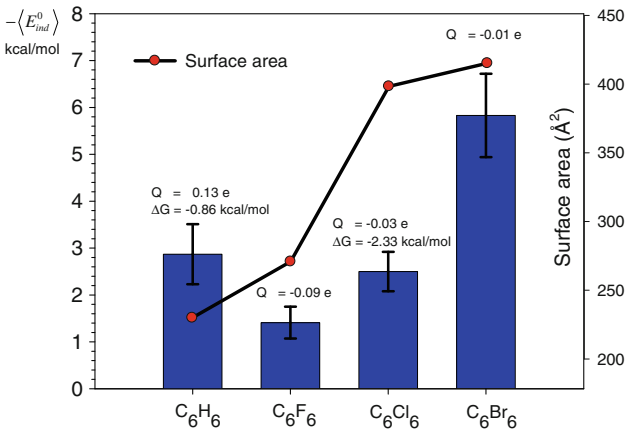

Fig. 9 The estimated average solute–solvent instantaneous induction energy $-\langle E^0_{\text{ind}}\rangle$ between the TIP3P surrounding water charges and the uncharged EPIC-polarizable benzene, hexafluorobenzene, hexachlorobenzene and hexabromobenzene. The molecular surface area of each molecule is plot and the experimental hydration energies reported above each *bar*. The carbon partial charges that best fit the gas phase ESP are also reported

A major assumption in the continuum solvent is the validity of the water configurational averaging. The configurational relaxation time of water is about 4 orders of magnitude slower than the electronic relaxation responsible for the solute polarization. Therefore, the solute is instantaneously polarized at each explicit water configuration adding an instantaneous induction energy even for a non-polar solute that we call average solute–solvent instantaneous induction that we note $\langle E^0_{\text{ind}}\rangle$. An explicit water simulation snapshot is shown in Fig. 10 with hexachlorobenzene. It is clear that momentarily, hydrogen atoms are pointing toward the solute volume and, unless the atoms have a well defined polarity, the water molecules will not exhibit strongly oriented averaged configuration. The configurational averaging of water in PB totally misses this energy since it can only capture the average configurational effects. This is reflected in implicit solvent water where solvent polarization mostly occurs in regions close to static partial charges inside the solute cavity. To understand the potential importance of $\langle E^0_{\text{ind}}\rangle$, we estimate its magnitude by performing molecular dynamic simulations with benzene, hexafluorobenzene, hexachlorobenzene and hexabromobenzene solutes in a periodic boundary water box. At different snapshots, we keep a 8 Å shell of TIP3P water molecules around the solute. The system is put in vacuum and, using Poisson's equation and EPIC parameters, we evaluate the polarization energy between the EPIC-polarizable solutes with atomic charges set to zero and the TIP3P water charges. Obviously, such an uncharged solute with continuum water would produce an hydration energy of zero. In contrast, Fig. 9 shows that this amounts to significant induction energies of −2.9 kcal/mol for benzene, −1.4 kcal/mol for hexafluorobenzene, −2.5 kcal/mol for hexachlorobenzene and −6 kcal/mol for hexabromobenzene. These numbers do not explain the difference between benzene and hexachlorobenzene, but provide a basis to understand their negative ΔG_{hyd}.

In principle, pre-polarized partial charges do not suffer as much because the instantaneous induction is taken into account in an average way in the polar region of the molecule. However, with molecules like benzene or hexachlorobenzene, the out-of-plane polarization plays an important role and the atomic partial charges may not always be pre-polarized appropriately. This is reflected in this work with the AM1-BCC charges that overestimate by 4.8 kcal/mol the hydration energy of hexachlorobenzene. The magnitude of the solute–solvent instantaneous induction energy should depend on the shape of the solute which affects the number of first shell water and the polarity of the solute which is an important factor dictating the orientation of the solvent molecules.

From Fig. 9, it is clear that the surface area term cannot be used to quickly estimate $\langle E^0_{\text{ind}}\rangle$ due to the lack of

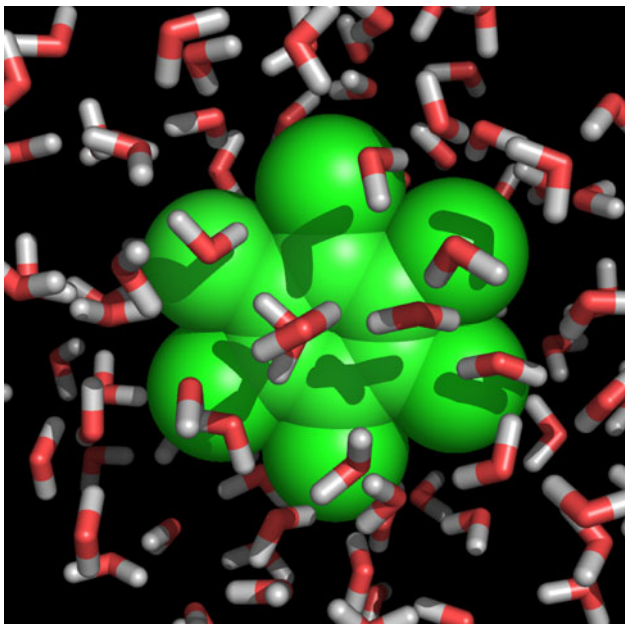

Fig. 10 Example of an explicit water snapshot from a MD simulation in which some water molecules are transiently orienting their dipole moment toward the surface of the hexachlorobenzene solute

correlation. Here, as in a previous study [13], the surface area term provides the right average shift on the prediction. A systematic inclusion of the induction energy in implicit solvent hydration energies seems important and may improve the precision of the method.

With a different computational approach, Dr. Anthony Nicholls from OpenEye Scientific Software Inc. reported that he could explain the *too* negative hydration energies of CF_4, CBr_4 and CI_4 by including the average interaction energy of a single water molecule rolled over a solute to which he assigned an internal dielectric of 1.76 (from refractive indices) [51]. He carefully adjusted the Bondi atomic radii to empirically match the ESP from the solute felt by the water molecule. He reported that the surface area tension constant he obtained by fitting had the value of 32 cal/Å^2, exactly the water to alkane experimental surface tension. This finding was before the discovery of EPIC and was including, in an empirical way, the solute induction by a water molecule. The current study shows how to only assess this term with a clear definition of the physical principles in play. We nevertheless end with conclusions very similar to Dr. Nicholls that the explicit configurations of water molecules should play a major role in the prediction of hydration energies of non polar species.

Conclusion

The SAMPL2 set was very challenging for our methods. Compared to the experimental data previously available, this challenge contained much more polar molecules and extended up to the highest positive hydration energies available. Furthermore, the highly halogenated molecules as such present in the challenge are known to be difficult for hydration energy prediction.

In this paper we showed that the use of a 3-stages restrained charge fitting protocol with DRESP drastically reduced the magnitude of the charges of buried atoms while maintaining the quality of the ESP fit and the accuracy of the predicted hydration energies. As with RESP, DRESP with hyperbolic charge restraints remove a significant amount of numerical instability.

Student's *t*-test did not confirm the superiority or inferiority of any of the 6 methods used in this study within the normative 95% confidence level. Indeed, two outliers, the d-glucose and d-xylose, contributed for a significant part of the variance of most of the methods. However, regarding these two outliers, the AM1-BCC charges showed a better robustness compared to ESP-fitted charges. The error in the predictions for d-glucose was attributed to the use inappropriate conformations resulting from the subsequent B3LYP gas phase minimization of the conformations. The error in the prediction of the hydration energy of d-xylose is attributed to the use of a minor form of the sugar in water. Finally, despite the fact that the explanatory set was known at the moment of the calculations, the obtained performances were worst on that set.

In line with other studies, we conclude that the use of multiple conformers do not impact the results significantly. However, strong intramolecular electrostatic interactions seen in gas-phase minimized molecules can impair the results as in the case of d-glucose. It is not clear from the different assessment made up to now, including the current paper, that molecules with drug-like molecular weight and flexibility would not benefit from conformational sampling for hydration energy prediction. It is remarkable that the cheapest method in human and computer time, involving no QM calculation, gave among the best performance (M3). Hence, it is fair to say that the M3 method that uses a single MMFFs ($\varepsilon = 10$) minimized low energy conformer with AM1-BCC atomic charges and the G2-12SA Poisson solvent model stood out in this study as the preferred approach.

Non-polar molecules were also important outliers in our predictions. Hexachlorobenzene and hexachloroethane have a significantly negative experimental ΔG_{hyd}. Seeking to understand why, we examined the magnitude of the average solute–solvent instantaneous induction energy, a term absent in PB calculations. Our MD simulations estimate values ranging between −1.5 and −6.0 kcal/mol for a series of halo-benzene molecules. This shows that the electronic polarization plays an important role in solute–solvent interactions that may not be sufficiently

accounted for even with pre-polarized charges. This term could potentially explain the trend of halogenated methane discussed elsewhere [9]. The solute–solvent instantaneous induction energy should play a role in explicit solvent simulations as well since the water-phase pre-polarized charges can only account for average water configurations.

The current study points toward few points for improving the calculation of hydration energies. Firstly, it is important that the appropriate conformers of highly polar molecules (or functional group) be represented. Secondly, attractive terms such as the average instantaneous solute–solvent induction energy and potentially the dispersion energy need to be considered. A specialized gas-phase AM1-BCC charging scheme could be advantageously used with EPIC for rapidity and robustness. Finally, such changes would command a re-optimization of the solvent atomic radii used to solve Poisson's equation. The incorporation of the polar molecules found in SAMPL2 should extend the validity of the obtained model.

Acknowledgments The authors thank Dr. Christopher Bayly from Merck Frosst Canada for his guidance on the 3-stage RESP procedure, for the informatics tools he gratefully provided to us and for his comments on the manuscript.

References

1. Overview paper SAMPL2 referenceGuthrie, J. P. JCAM. 2010
2. Michel J, Tirado-Rives J, Jorgensen WL (2009) Energetics of displacing water molecules from protein binding sites: consequences for ligand optimization. J Am Chem Soc 131:15403–15411
3. Guthrie JP (2009) A blind challenge for computational solvation free energies: introduction and overview. J Phys Chem B 113:4501–4507
4. Honig B, Nicholls A (1995) Classical electrostatics in biology and chemistry. Science 268:1144–1149
5. Jeancharles A, Nicholls A, Sharp K, Honig B, Tempczyk A, Hendrickson TF, Still WC (1991) Electrostatic contributions to solvation energies—comparison of free-energy perturbation and continuum calculations. J Am Chem Soc 113:1454–1455
6. Sharp K, Jean-Charles A, Honig B (1992) A local dielectric-constant model for solvation free-energies which accounts for solute polarizability. J Phys Chem 96:3822–3828
7. Sitkoff D, Sharp KA, Honig B (1994) Accurate calculation of hydration free-energies using macroscopic solvent models. J Phys Chem 98:1978–1988
8. Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks CL (2004) Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. J Comput Chem 25:265–284
9. Nicholls A, Wlodek S, Grant JA (2009) The SAMP1 solvation challenge: further lessons regarding the pitfalls of parametrization. J Phys Chem B 113:4521–4532
10. Nina M, Beglov D, Roux B (1997) Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. J Phys Chem B 101:5239–5248
11. Nicholls A, Mobley DL, Guthrie JP, Chodera JD, Bayly CI, Cooper MD, Pande VS (2008) Predicting small-molecule solvation free energies: an informal blind test for computational chemistry. J Med Chem 51:769–779
12. Shivakumar D, Deng YQ, Roux B (2009) Computations of absolute solvation free energies of small molecules using explicit and implicit solvent model. J Chem Theory Comput 5:919–930
13. Truchon J-F, Nicholls A, Roux B, Iftimie RI, Bayly CI (2009) Integrated continuum dielectric approaches to treat molecular polarizability and the condensed phase: refractive index and implicit solvation. J Chem Theory Comput 5:1785–1802
14. OpenEye Scientific Software Inc (2007) *Zap Toolkit,* version 2.2.1; Santa Fe, NM, USA
15. Grant JA, Pickup BT, Nicholls A (2001) A smooth permittivity function for Poisson-Boltzmann solvation methods. J Comput Chem 22:608–640
16. Tan YH, Luo R (2007) Continuum treatment of electronic polarization effect. J Chem Phys 126:094103
17. Naim M, Bhat S, Rankin KN, Dennis S, Chowdhury SF, Siddiqi I, Drabik P, Sulea T, Bayly CI, Jakalian A, Purisima EO (2007) Solvated interaction energy (SIE) for scoring protein-ligand binding affinities. 1. Exploring the parameter space. J Chem Inf Model 47:122–133
18. Rankin KN, Sulea T, Purisima EO (2003) On the transferability of hydration-parametrized continuum electrostatics models to solvated binding calculations. J Comput Chem 24:954–962
19. Warshel A, Sharma PK, Kato M, Parson WW (2006) Modeling electrostatic effects in proteins. Biochimica et Biophysica Acta-Proteins Proteomics 1764:1647–1676
20. Schnieders MJ, Baker NA, Ren P, Ponder JW (2007) Polarizable atomic multipole solutes in a Poisson-Boltzmann continuum. J Chem Phys 126:124114
21. Weininger D (1990) Smiles.3. Depict—graphical depiction of chemical structures. J Chem Inf Model 30:237–243
22. Weininger D, Weininger A, Weininger JL (1989) Smiles.2. Algorithm for generation of unique smiles notation. J Chem Inf Model 29:97–101
23. Weininger D (1988) Smiles, a chemical language and information-system.1. Introduction to methodology and encoding rules. J Chem Inf Model 28:31–36
24. OpenEye Scientific Software Inc (2007) *OMEGA,* version 2.2.1; Santa Fe, NM, USA
25. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery Jr JA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Menucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03, Revision, version C.02; Wallingford CT, USA
26. Becke AD (1993) Density-functional thermochemistry.3. The role of exact exchange. J Chem Phys 98:5648–5652
27. Becke AD (1993) A new mixing of Hartree-Fock and local density-functional theories. J Chem Phys 98:1372–1377
28. Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic-behavior. Phys Rev A 38:3098–3100

29. Jakalian A, Jack DB, Bayly CI (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J Comput Chem 23:1623–1641
30. Jakalian A, Bush BL, Jack DB, Bayly CI (2000) Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. J Comput Chem 21:132–146
31. Singh UC, Kollman PA (1984) An approach to computing electrostatic charges for molecules. J Comput Chem 5:129–145
32. Truchon J-F, Nicholls A, Grant JA, Iftimie RI, Roux B, Bayly CI (2009) Extending the polarizable continuum dielectric model to account for electronic polarization in intermolecular interactions. J Comput Chem 31:811–824
33. Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges—the resp model. J Phys Chem 97:10269–10280
34. The restraints are in a.u. and correspond to the a parameter in equation 10 of reference 33. 2009
35. Bush BL, Bayly CI, Halgren TA (1999) Consensus bond-charge increments fitted to electrostatic potential or field of many compounds: application to MMFF94 training set. J Comput Chem 20:1495–1516
36. Woon DE, Dunning J (1993) Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. J Chem Phys 98:1358–1371
37. Mennucci B, Cances E, Tomasi J (1997) Evaluation of solvent effects in isotropic and anisotropic dielectrics and in ionic solutions with a unified integral equation method: theoretical bases, computational implementation, and numerical applications. J Phys Chem B 101:10506–10517
38. Cances E, Mennucci B, Tomasi J (1997) A new integral equation formalism for the polarizable continuum model: theoretical background and applications to isotropic and anisotropic dielectrics. J Chem Phys 107:3032–3041
39. Mennucci B, Tomasi J (1997) Continuum solvation models: a new approach to the problem of solute's charge distribution and cavity boundaries. J Chem Phys 106:5151–5158
40. Truchon J-F (2008) Modéliser la polarisation électronique par un continuum diélectrique intramoléculaire: Vers un champ de force polarisable pour la chimie bioorganique. Ph.D. Université de Montréal
41. Halgren TA (1999) MMFF VI. MMFF94 s option for energy minimization studies. J Comput Chem 20:720–729
42. Halgren TA (1999) MMFF VII. Characterization of MMFF94, MMFF94 s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. J Comput Chem 20:730–748
43. Halgren TA (1996) Merck molecular force field.1. Basis, form, scope, parameterization, and performance of MMFF94. J Comput Chem 17:490–519
44. Halgren TA (1992) Representation of Vanderwaals (Vdw) interactions in molecular mechanics force-fields—potential form, combination rules, and Vdw parameters. J Am Chem Soc 114:7827–7843
45. Bondi A (1964) van der Waals volumes and radii. J Phys Chem 68:441–451
46. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935
47. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25:1157–1174
48. Private communication with Dr. Christopher Bayly 2009
49. Rizzo RC, Aynechi T, Case DA, Kuntz ID (2006) Estimation of absolute free energies of hydration using continuum methods: accuracy of partial, charge models and optimization of nonpolar contributions. J Chem Theory Comput 2:128–139
50. Mobley DL, Dill KA, Chodera JD (2008) Treating entropy and conformational changes in implicit solvent simulations of small molecules. J Phys Chem B 112:938–946
51. Nicholls A (2007) OpenEye CUP8 conference, Santa FE, NM, February 26–28, 2007