

Predicting hydration free energies of polychlorinated aromatic compounds from the SAMPL-3 data set with FiSH and LIE models

Traian Sulea · Enrico O. Purisima

Received: 24 October 2011 / Accepted: 8 December 2011 / Published online: 22 December 2011
© Her Majesty the Queen in Rights of Canada as represented by the National Research Council of Canada 2011

Abstract Next-generation solvation models are devised to mimic the accuracy and generality of explicit solvation models at the speed of current popular implicit solvation models. One such method is the first-shell of hydration (FiSH) continuum model that was trained on hydration energetics from LIE calculations and molecular dynamics simulations in explicit solvent. Here we tested prospectively the FiSH model on the SAMPL-3 hydration data set that zooms in the effect of chlorination on solvation. We compare these FiSH predictions with those from retrospective LIE calculations. We find that neither FiSH nor LIE can reproduce well the absolute values and the trend of hydration free energies in the biphenyl and dioxin aromatic chlorination series. Some of the hypotheses behind this performance are discussed and tested. The LIE explicit-solvent model shows some improvement relative to the FiSH continuum model, and we correct a systematic deviation in the continuum van der Waals term of FiSH associated with aromatic Cl atom type.

Keywords Continuum solvation · First hydration shell · Linear interaction energy · Prospective study

Introduction

Water plays a central role in virtually all cellular systems. This role is not limited to providing a passive embedding

medium for proteins, nucleic acids, and other biomolecules. Water also takes an active role by contributing to the dynamics and the outcome of many biophysically interesting phenomena. The thermodynamic contribution arising from the change in hydration of the evolving molecular objects is particularly important in intermolecular recognition and binding, folding and other conformational equilibria, reaction kinetics, and phase partitioning. Along with significant advances in computational biophysics in recent years, a number of solvation models presenting varying degrees of complexity and levels of theory have been developed and parametrized [1]. Explicit-solvent models of hydration, including pathway methods such as free energy perturbation (FEP) or thermodynamic integration (TI) [2–4], or end-point methods such as linear interaction energy (LIE) [5–7], address the discrete nature of water around the solute but are not practical for high-throughput applications. Implicit models of hydration (i.e., continuum models) have been precisely developed to address the speed issue and they excel in this regard, but sometimes at a cost on accuracy [8–10].

One limitation that affects the accuracy of implicit solvation models is the continuum replacement of particulate structure in the first hydration shell around the solute. The local ordering of water in the first shell is different from isotropic bulk water and varies depending on solute polarity. Around a hydrophobic solute surface, interactions within the first hydration shell itself are favoured over interactions with the solute or with bulk solvent [11]. Around polar solute surfaces, water molecules interact strongly with the solute but orient differently around positively and negatively charged atoms, a phenomenon known as the charge asymmetry of water [12]. Several recent simulations in explicit solvent uncovered significant details in the water structure and energetics [13–18]. The approximations made in current popular implicit models in

Electronic supplementary material The online version of this article (doi:10.1007/s10822-011-9522-1) contains supplementary material, which is available to authorized users.

T. Sulea · E. O. Purisima (✉)
Biotechnology Research Institute, National Research Council
Canada, 6100 Royalmount Avenue, Montreal,
QC H4P 2R2, Canada
e-mail: Enrico.Purisima@nrc.gc.ca

describing the physics of hydrophobic and polar hydration limit their ability to describe local details and to capture their effects.

The need to account for the first-shell water ordering effects prompted several recent developments within the realm of implicit solvation. These include, among other methods, the semi-explicit assembly (SEA) [19, 20] and first-shell hydration (FiSH) [21] continuum methods. The idea is to learn from the physics captured by explicit solvent models, which in principle can describe first hydration shell effects. For example, the FiSH continuum model was trained on hydration energetics from LIE calculations from molecular dynamics (MD) simulations in explicit solvent [7, 22]. These new methods also maintain the computational efficiency characteristic to continuum models in order to allow portability and tractability for applications like virtual screening and hit optimization. Therefore, these next-generation solvation models are devised to mimic the accuracy and generality of explicit solvation models at the speed of current popular implicit solvation models.

Here we tested prospectively the FiSH continuum model on the SAMPL-3 blind challenge for predicting hydration free energies. The SAMPL-3 data set focuses on the effect of chlorination on solvation, both for aliphatic compounds—the ethane series—and aromatic compounds—the biphenyl and dioxin series. Such a focused data set has the potential to generate significant insight with respect to successes and failures of the participating methods in describing the energetics of solvation associated with polychlorination. Retrospectively, we compared FiSH predictions with those from LIE calculations given that the FiSH continuum model was devised to mimic the LIE explicit-solvent based predictions. We find that neither FiSH nor LIE can reproduce well the absolute values and the trend of transfer free energies in the aromatic chlorination series. Some of the hypothesis behind this performance are discussed and tested. In addition, because the LIE model shows some improvement relative to the FiSH continuum model, we correct here a systematic deviation in the continuum van der Waals term of FiSH associated with aromatic Cl atom type.

Materials and methods

Structures were energy minimized with the MMFF94 force-field [23] in vacuum using SZYBKI (OpenEye Scientific Software, Santa Fe, NM). For all biphenyls (IDs 11–24), the starting conformation was set to a 45° angle between the phenyl rings. For polychlorinated ethane analogs 4, 6 and 8, the rotamer having the Cl atoms farthest from each other was selected as starting point for minimization. For solvation calculations, partial charges were

calculated on the MMFF-minimized structures with the AM1-BCC method [24, 25] using the single-point (SP) option in MOLCHARGE (OpenEye Scientific Software). For retrospective calculations, the MMFF-minimized structures were also charged with quantum mechanical (QM) partial charges at the MP2/cc-PVTZ SCRF level (kindly supplied by D. Mobley, see the paper by Mobley et al. in this special issue).

Continuum solvation calculations were carried out with the FiSH model on the MMFF minimized structures [21] using the parameters developed and applied previously [21, 22]. The FiSH continuum solvation model has been developed to mimic the accuracy of an explicit-solvent LIE hydration model [7] and includes an electrostatic, a van der Waals and a cavity term,

$$\Delta G_{\text{hyd}}^{\text{FiSH}}(\{r_i^{\text{Born}}\}, \gamma_{\text{cav}}) = \Delta G_{\text{hyd}}^{\text{R}}(\{r_i^{\text{Born}}\}) + U^{\text{vdw}} + \gamma_{\text{cav}} \text{MSA} + C \quad (1)$$

To restore the charge asymmetry of water into continuum electrostatics models [12, 26] the solute reaction field energy, $\Delta G_{\text{hyd}}^{\text{R}}$, is corrected by modifying the initial guess for the atomic Born radii, $\{r_i^{\text{Born}}\}$, taken from force-field van der Waals parameters, based on atom-averaged induced-surface-charge densities obtained from a boundary element calculation [27, 28]. The correction function parameters were trained on the electrostatic component from the LIE model [7, 21]. A term proportional to the molecular surface area (MSA) of the solute and trained on pseudo-experimental free energies, γ_{cav} , is employed to account for the cost of cavity formation [21].

A 2-region representation of the continuum van der Waals term, U^{vdw} , was implemented in FiSH [21]. The water molecules in the first shell are assumed to be uniformly distributed along the solvent-accessible surface (SAS). Their contribution to U^{vdw} is calculated as an integral of the Lennard-Jones interaction between the solute atoms and the SAS. The SAS is constructed using solute atom-specific solvent probe radii, which were based on the average first radial-distribution-function (RDF)-peak distances between the water oxygen and GAFF atom types determined from MD simulations in explicit water on a training set. The SAS is then generated by operationally inflating the force-field van der Waals atomic radii by the appropriate probe radii. Additional manual fine-tuning of the radii was carried out in order to reproduce the average solute-solvent van der Waals interaction energy with an effective first hydration shell calculated with the LIE approach based on MD simulations in explicit water. The second region starts at the SAS + 2.8 Å (a water diameter) and incorporates contributions from the second hydration shell outwards to infinity *via* a traditional van der Waals continuum model [29].

Retrospectively, the position of the SAS around the Cl atom was modified to better represent the first RDF peak from MD simulations and further adjusted to reproduce the average solute-solvent van der Waals interaction energy with first-shell water molecules from LIE calculations on 11 chlorinated benzene analogs [7]. Only 2 of these analogs, chlorobenzene and 2-chlorotoluene, were used in the original FiSH training set, whereas 9 polychlorinated benzenes were in the FiSH test set [21]. Here, a separate solvent probe radius was introduced specifically for the aromatic Cl atom, using the average LIE van der Waals energies for all 11 chlorinated benzene analogs. The new solvent probe radius for the aromatic Cl is now larger by 0.1 Å over that for the aliphatic Cl. This moves the SAS farther out and brings the Cl-SAS distance closer to the Lennard-Jones energy-minimum distance for the Cl-TIP3P pair. A solvent probe radius of 1.452 Å, originally used for any Cl atom [21] can be retained for the aliphatic Cl atom given the relatively good FiSH predictions obtained in this study for the chlorinated ethane congeners (*vide infra*).

Linear interaction energy (LIE) based solvation calculations were carried out retrospectively as described previously [7, 22]. Briefly, molecular dynamics (MD) calculations were performed in AMBER 9 [30] with the solute held rigid in the MMFF-minimized conformation and charged with either AM1-BCC-SP or MP2/cc-PVTZ-SCRF charges. The TIP3P water solvent in a 12-Å shell around the solute was heated and equilibrated for 50 ps and then simulated at 300 K for 2 ns in NPT ensemble. The GAFF force-field [31] and TIP3P parameters [32] were applied to the solute and water, respectively. Average solute-solvent van der Waals and Coulomb interaction energies, reaction-field energy outside the explicitly solvated system, and solute molecular surface area were calculated from the 2nd ns of MD simulation for 100 snapshots taken at 10 ps intervals.

All statistical data presented in this report were bootstrapped for 5,000 samples with replacement, using the R software [33] and are presented as the median value \pm standard error.

Results and discussion

Prospective FiSH predictions for the entire SAMPL-3 data set of 36 compounds are plotted against the experimental data in Fig. 1a (and listed in Table S1 provided as Supporting Information). We see that FiSH predictions are correlated with experimental hydration free energy data only for the subset of 10 aliphatic compounds including ethane and its chlorinated derivatives ($R^2 = 0.52 \pm 0.21$; MUE = 0.66 ± 0.15 kcal/mol). However, for the aromatic subset of 26 compounds including the biphenyl, dioxin, and their

chlorinated derivatives, FiSH predictions are not correlated with experimental hydration data ($R^2 = 0.05 \pm 0.06$), which are significantly underestimated by FiSH (MSE = 3.39 ± 0.42 kcal/mol; MUE = 3.43 ± 0.40 kcal/mol).

Taken separately, there appears to be even an anti-correlation for the 12 analogs in the biphenyl series ($R^2 = 0.30 \pm 0.18$, slope = -0.86 ± 0.34) and mostly for the 10 dioxins ($R^2 = 0.58 \pm 0.22$; slope = -2.71 ± 0.77). As seen in Fig. 1a, FiSH predicts less favorable hydration as the degree of chlorination increases (in the biphenyl series from analog 11—non-chlorinated to analog 24—fully chlorinated; in the dioxin series from analog 25—non-chlorinated to analog 36—fully chlorinated) over a wide range (6.5 kcal/mol for biphenyls and 5.3 kcal/mol for dioxins). The experimental data shows a much narrower range of variation with the degree of chlorination (2.7 kcal/mol for biphenyls and 1.4 kcal/mol for dioxins), and generally shows the opposite trend, i.e., more negative hydration free energies in the highly polychlorinated analogs. We note that this experimental trend is somewhat reversed in the aliphatic series of ethane chlorination.

The anti-correlation in the biphenyl and dioxin chlorination series is mostly due to the electrostatic component in the FiSH model, which becomes less negative as more chlorine atoms are added onto each aromatic scaffold (Table S1). However, the underestimation of absolute hydration free energies noted earlier (MSE = 3.39 ± 0.42 kcal/mol; MUE = 3.43 ± 0.40 kcal/mol), which is accentuated with the degree of chlorination, is not due to the electrostatic component, since this term is in the same range as the experimental data (Tables S1) and alone it provides a better model (MSE = 0.30 ± 0.33 kcal/mol; MUE = 1.49 ± 0.17 kcal/mol) than the overall FiSH model. Hence, the underestimation can be tracked to the non-polar term of FiSH, which arises from the competing action of the solute-solvent continuum van der Waals term, U_{vdw} , and the surface-area dependent cost of cavity formation in water, $\gamma_{\text{cav}}\text{MSA} + C$ (Eq. 1). As seen in Fig. 1b, the cavity cost systematically outstrips the van der Waals attraction leading to a positive non-polar contribution. The effect differs for the biphenyl and dioxin subsets, with larger and more chlorination-dependent effects for the biphenyl series. We note that the positive non-polar term appears to have the right magnitude for the non-chlorinated biphenyl (11) and dioxin (25), since their absolute hydration free energies (-2.23 and -3.15 kcal/mol, respectively) are quite well reproduced by FiSH (-1.93 and -3.68 kcal/mol, respectively). The magnitude of the non-polar term and its trend with the degree of chlorination also appear reasonable in the ethane series leading to FiSH predictions of acceptable agreement with the experiment. All these results point out toward a possible deficiency in the FiSH model to account for the adequate balance

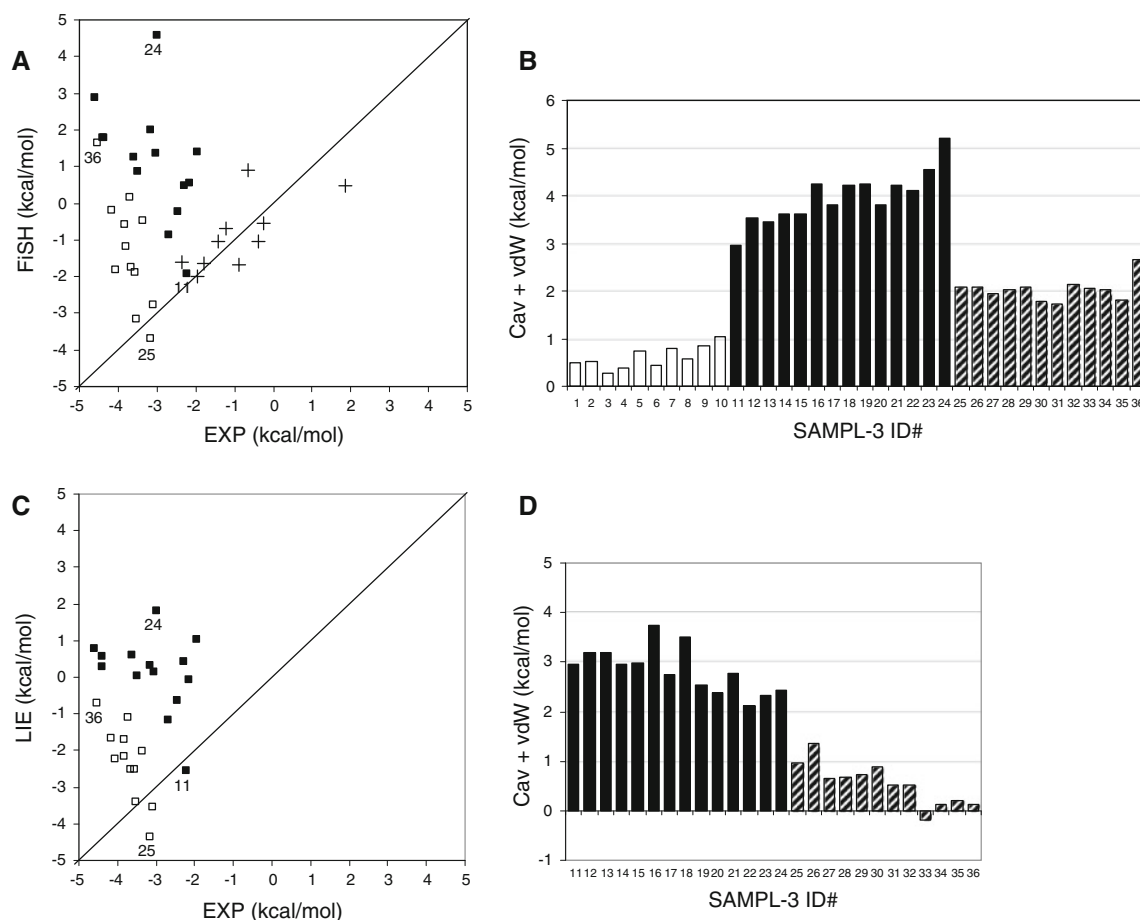


Fig. 1 Prospective FiSH and retrospective LIE predictions. MMFF94 geometries and AM1-BCC-SP partial charges were used with both methods. **a** Scatter plot of FiSH predictions versus experimental data, with the ethane series—plus symbol, biphenyl series—closed squares, dioxin series—open squares. **b** Bar-graph showing FiSH predictions of non-polar hydration contributions represented by the sum of attractive solute-solvent van der Waals interactions (vdW) and repulsive cavitation (Cav) terms. **c** Scatter plot of LIE predictions

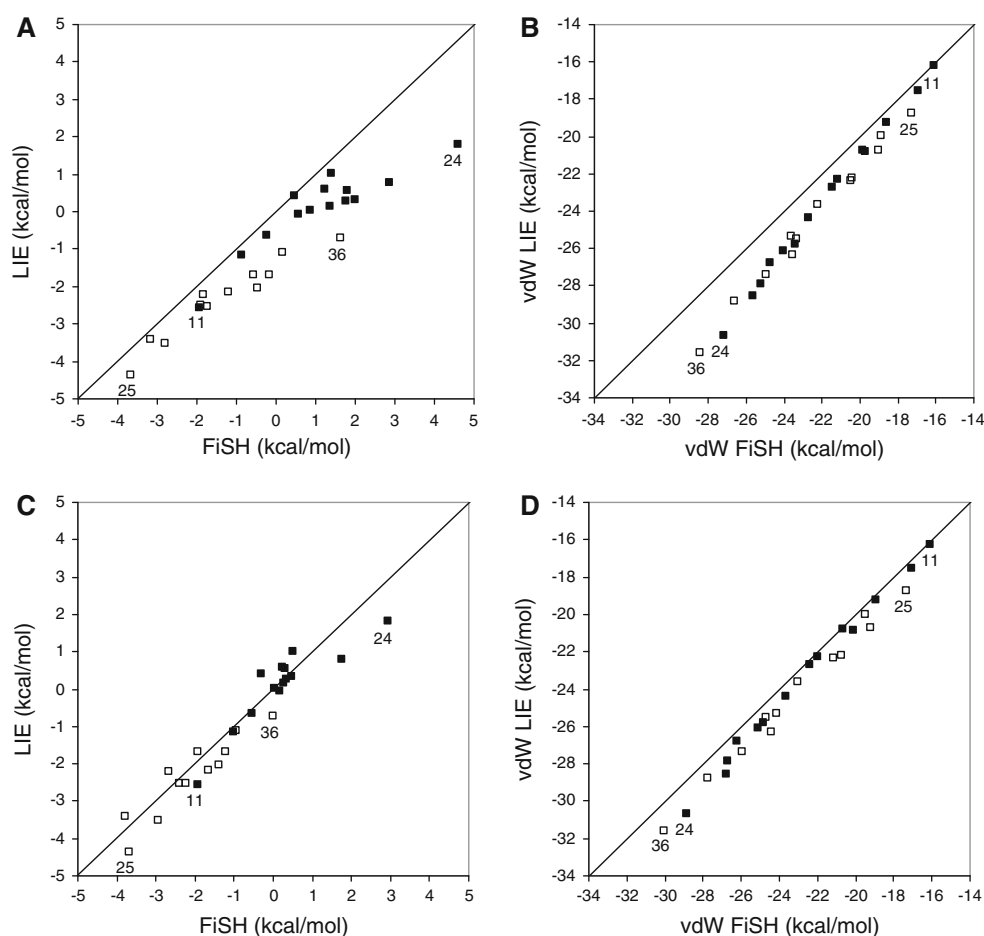
versus experimental data, biphenyl series—closed squares, dioxin series—open squares. **d** Bar-graph showing LIE predictions of non-polar hydration contributions represented by the sum of attractive solute-solvent van der Waals interactions (vdW) and repulsive cavitation (Cav) terms. Diagonal lines represent ideal predictions. Labeled points correspond to the non-chlorinated (11) and fully-chlorinated (24) biphenyl, and to the non-chlorinated (25) and fully-chlorinated (36) dioxin

between the van der Waals and cavity contributions for the aromatic Cl atom.

This concern is justified since training of the Cl atom radius was based on only a few compounds with first RDF peaks calculated from explicit solvent MD simulations [7, 21]. Therefore, to verify this hypothesis, we retrospectively calculated the hydration free energies for the biphenyl and dioxin series with the LIE method in explicit solvent. To separate various effects, these LIE calculations were carried out on the same conformations and partial charges of the solutes as in the FiSH model. The explicit-solvent LIE model showed an improvement of the absolute predictions ($\text{MUE} = 2.52 \pm 0.29$ kcal/mol for the aromatic compounds) thus reducing the underestimation seen with the FiSH model (compare Fig. 1c with Fig. 1a) due to a smaller non-polar penalty with increasing

polychlorination (compare Fig. 1d with Fig. 1b). The improvement with the LIE model is due mainly to a more attractive van der Waals term, leading to more negative hydration free energies than those calculated with the FiSH model (Fig. 2, component terms listed in Table S1). This supports our hypothesis on the underparametrized continuum van der Waals term of FiSH for the aromatic Cl atom, and motivates fine-tuning of this parameter in order to track more closely to the explicit solvent energetics (see below). Nonetheless, the still low level of agreement with these experimental data even with the explicit-solvent LIE model indicates that there are other phenomena that have yet to be more accurately captured. Among others, these may relate to the molecular mechanics force field, the level of geometry optimization, the description of electronic structure, and the non-linearity of solvent response.

Fig. 2 Improving mimicry of explicit-solvent LIE predictions by continuum FiSH model by adjusting the position of first-shell water molecules around the aromatic Cl atom. MMFF94 geometries and AM1-BCC-SP partial charges were used with both methods. **a** Scatter plot of predicted hydration free energies with the original FiSH model versus LIE model. **b** Scatter plot of solute-solvent vdW interaction energy component calculated with the original FiSH model versus LIE model. **c** Scatter plot of predicted hydration free energies with the recalibrated FiSH model versus LIE model. **d** Scatter plot of solute-solvent vdW interaction energy component calculated with the recalibrated FiSH model versus LIE model



In order to probe the effect of electronic structure alone, we repeated the explicit-solvent LIE calculations using the same conformations of the solute but with point charges derived at a higher level of theory, by replacing the AM1-BCC-SP charges with MP2/cc-PVTZ-SCRF charges (Table S1). There is only a marginal decrease in the absolute errors versus experiment ($\text{MUE} = 2.39 \pm 0.32$ kcal/mol for all aromatics, most of the improvement seen in the dioxin series), with LIE values calculated with the two charge sets tracking quite closely to each other, with $R^2 = 0.90 \pm 0.02$ and $\text{MUE} = 0.42 \pm 0.06$ kcal/mol (Figure S1). We note the excellent prediction for the non-chlorinated biphenyl (-2.32 vs. -2.23 kcal/mol experimental) and dioxin (-3.32 vs. -3.15 kcal/mol experimental) with the higher-level charges. Based on these results, we conclude that while electronic effects might be partially responsible for the relatively poor predictions for aromatic polychlorinated compounds, significant improvements may need to be sought with solvation models that go beyond the point-charge formalism.

The accuracy of geometry optimization can represent another avenue for improved predictions. Given the rigidity of the dioxin skeleton, the dihedral angle optimization may

not play a major role. However, bond lengths may be particularly important. It seems that the C-Cl bonds get shorter as more chlorines get added. With octachlorodioxin, the C-Cl bond is 1.71 Å instead of 1.74 Å [34]. This can do two things. First, the partial charges can change, e.g., as reflected in the MP2/cc-PVTZ-SCRF charge distribution. But also, this results in a smaller footprint with respect to the cavity term while increasing the van der Waals density, both effects working together to make the hydration free energy more negative.

However, even with highly-optimized structures, a recent study that applies a molecular integral equation theory to the solvation of polychlorinated benzene and biphenyl congeners points towards the need for additional empirical parameters for reproducing the experimental hydration free energies of these compounds [35]. Geometry optimization was done at the B3LYP/6-31G(d,p) level of theory, starting from crystallographic structures for some of the congeners. However, in order to improve the agreement with the experiment, the one-dimensional reference interaction site model (1D-RISM) was supplemented with a structural descriptors correction (SDC), a set of empirical correction terms in a QSAR-like approach. One structural

descriptor used as a correction term was the number of Cl atoms, which was scaled by a fitted coefficient of -1.30 kcal/mol, which effectively buffers the calculated hydration free energy at increased levels of chlorination. With this hybrid model, a good agreement with experiment (RMSE = 0.91 kcal/mol on 57 chlorinated biphenyls with available experimental data) was obtained, outperforming the predictions by other models like SM6, COSMO-SAC, and PBSA/SDC. Some of this better performance of the hybrid 1D-RISM/SDC approach was attributed to the RISM formalism that is able to properly capture non-linear effects of the solvent response, which can occur for example due to multiple Cl substituents close to each other as in the highly chlorinated analogs.

Even with the 1D-RISM/QSAR, a bias of about 1 kcal/mol was nevertheless observed and attributed to the differences in the quality of experimental data for polychlorinated aromatic compounds including biphenyls. Indeed, there is significant uncertainty on the experimental data of polychlorinated biphenyls, with several methods providing hydration free energies that can vary by several kcal/mol and even reverse the trend with increased level of chlorination (see Ref. [35] and references cited therein). This underscored the need for additional independent measurements to ascertain the reliability of experimental data.

In the meanwhile, we tried to improve the FiSH model in order to mimic the LIE model more closely with respect to the Cl parametrization with respect to the solute van der Waals interactions with the first-shell solvent. For the original FiSH parametrization, only 2 monochlorinated aromatic analogs were used in the training set, whereas 9 polychlorinated benzene congeners were left in the test set and were all underestimated by FiSH, particularly the tri-Cl and tetra-Cl benzene derivatives. These compounds are now used here to recalibrate the FiSH parameters with respect to the solvent probe radius around the Cl atom. Furthermore, we separated here the aromatic Cl atom type from aliphatic Cl, thus providing a new solvent probe radius specifically for the aromatic Cl (the original solvent probe radius can be retained for the aliphatic Cl atom type based on the reasonable results obtained here for the SAMPL-3 ethane series). This separation is justified given the different electronic effects of the aromatic and aliphatic Cl atom types, which are probably not captured well enough by point charges and molecular mechanics force fields. Similar deficiencies were also noted for other third period elements in relation to hydration free energy calculations [36].

Retrospective predictions with the recalibrated FiSH model show an improved agreement with the LIE predictions on the same geometries and partial charges. With AM1-BCC charges for example, the MUE relative to LIE hydration free energies decreases from 1.02 ± 0.13 kcal/mol with the

original FiSH parameters to 0.41 ± 0.06 kcal/mol with the new parameter for aromatic Cl (also compare Fig. 2c with Fig. 2a). This improvement is driven by a better agreement between the van der Waals components with the newer Cl parameters (compare Fig. 2d with Fig. 2b, Table S1). A slightly better FiSH-LIE agreement is obtained in the case of MP2/cc-PVTZ-SCRF charges (Figure S2). Thus, the reparametrization of the FiSH continuum model with respect to the aromatic Cl atom leads to an improved mimic of explicit-water-based LIE predictions and consequently better predictions of the SAMPL-3 experimental data at hand.

Acknowledgments We thank Dr. David L. Mobley (University of New Orleans) for providing the MP2/cc-PVTZ partial charges. This is NRC Canada publication number 53159.

References

- Purisima EO, Sulea T (2012) In: Toth G (ed) Computational structure-based drug discovery. Wiley-Blackwell, Chichester, Great Britain
- Mobley DL, Bayly CI, Cooper MD, Shirts MR, Dill KA (2009) *J Chem Theory Comput* 5:350–358
- Klimovich P, Mobley D (2010) *J Comput -Aided Mol Des* 24:307–316
- Shivakumar D, Deng Y, Roux B (2009) *J Chem Theory Comput* 5:919–930
- Carlson HA, Jorgensen WL (1995) *J Phys Chem* 99: 10667–10673
- Almlöf M, Carlsson J, Aqvist J (2007) *J Chem Theory Comput* 3:2162–2175
- Sulea T, Corbeil CR, Purisima EO (2010) *J Chem Theory Comput* 6:1608–1621
- Cramer CJ, Truhlar DG (1999) *Chem Rev* 99:2161–2200
- Tomasi J, Mennucci B, Cammi R (2005) *Chem Rev* 105:2999–3094
- Chen J, Brooks CL III, Khandogin J (2008) *Cur Opin Struct Biol* 18:140–148
- Raschke TM, Levitt M (2005) *Proc Nat Acad Sci USA* 102:6777–6782
- Mobley DL, Barber AE, Fennell CJ, Dill KA (2008) *J Phys Chem B* 112:2405–2414
- Chorny I, Dill KA, Jacobson MP (2005) *J Phys Chem B* 109:24056–24060
- Fennell CJ, Bizjak A, Vlachy V, Dill KA (2009) *J Phys Chem B* 113:6782–6791
- Cerutti DS, Baker NA, McCammon JA (2007) *J Chem Phys* 127:155101–155112
- Cheng LT, Wang Z, Setny P, Dzubiella J, Li B, McCammon JA (2009) *J Chem Phys* 131:144102
- Baron R, Setny P, McCammon JA (2010) *J Am Chem Soc* 132:12091–12097
- Setny P, Baron R, McCammon JA (2010) *J Chem Theory Comput* 6:2866–2871
- Fennell CJ, Kehoe C, Dill KA (2009) *J Am Chem Soc* 132:234–240
- Fennell CJ, Kehoe CW, Dill KA (2011) *Proc Nat Acad Sci USA* 108:3234–3239
- Corbeil CR, Sulea T, Purisima EO (2010) *J Chem Theory Comput* 6:1622–1637

22. Purisima E, Corbeil C, Sulea T (2010) *J Comput -Aided Mol Des* 24:373–383
23. Halgren TA (1996) *J Comput Chem* 17:490–519
24. Jakalian A, Bush BL, Jack DB, Bayly CI (2000) *J Comput Chem* 21:132–146
25. Jakalian A, Jack DB, Bayly CI (2002) *J Comput Chem* 23:1623–1641
26. Purisima EO, Sulea T (2009) *J Phys Chem B* 113:8206–8209
27. Purisima EO, Nilar SH (1995) *J Comput Chem* 16:681–689
28. Purisima EO (1998) *J Comput Chem* 19:1494–1504
29. Tan C, Tan YH, Luo R (2007) *J Phys Chem B* 111:12263–12274
30. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) *J Comput Chem* 26:1668–1688
31. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) *J Comput Chem* 25:1157–1174
32. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) *J Chem Phys* 79:926–935
33. R Foundation for Statistical Computing (2005) *R: A Language and Environment for Statistical Computing*. Vienna, Austria
34. Cantrell JS, Beiter TA, Tomlin D (1989) *Chemosphere* 19: 155–160
35. Ratkova EL, Fedorov MV (2011) *J Chem Theory Comput* 7:1450–1457
36. Mobley DL, Bayly CI, Cooper MD, Dill KA (2009) *J Phys Chem B* 113:4533–4537