Prediction of hydration free energies for aliphatic and aromatic chloro derivatives using molecular dynamics simulations with the OPLS-AA force field

Oliver Beckstein · Bogdan I. Iorga

Abstract All-atom molecular dynamics computer simulations were used to blindly predict the hydration free energies of a range of chloro-organic compounds as part of the SAMPL3 challenge. All compounds were parameterized within the framework of the OPLS-AA force field, using an established protocol to compute the absolute hydration free energy via a windowed free energy perturbation approach and thermodynamic integration. Three different approaches to deriving partial charge parameters were pursued: (1) using existing OPLS-AA atom types and charges with minor adjustments of partial charges on equivalent connecting atoms; (2) calculation of quantum mechanical charges via geometry optimization, followed by electrostatic potential (ESP) fitting, using Jaguar at the LMP2/cc-pVTZ(-F) level; and (3) via geometry optimization and CHelpG charges (Gaussian03 at the HF/6-31G* level), followed by two-stage RESP fitting. Protocol 3 generated the most accurate predictions with a root mean square (RMS) error of 1.2 kcal mol^{-1} for the entire data set. It was found that the deficiency of the standard OPLS-AA parameters, protocol 1 (RMS error 2.4 kcal mol^{-1} overall), was mostly due to compounds with more than three chlorine substituents on an aromatic ring. For this latter subset, the RMS errors were 1.4 kcal mol^{-1} (protocol 3) and 4.3 kcal mol^{-1} (protocol 1), respectively. We propose new OPLS-AA atom types for aromatic carbon and chlorine atoms in rings with ≥ 4 Cl-substituents that perform better than the best QM-based approach, resulting in an RMS error of 1.2 kcal mol^{-1} for these difficult compounds.
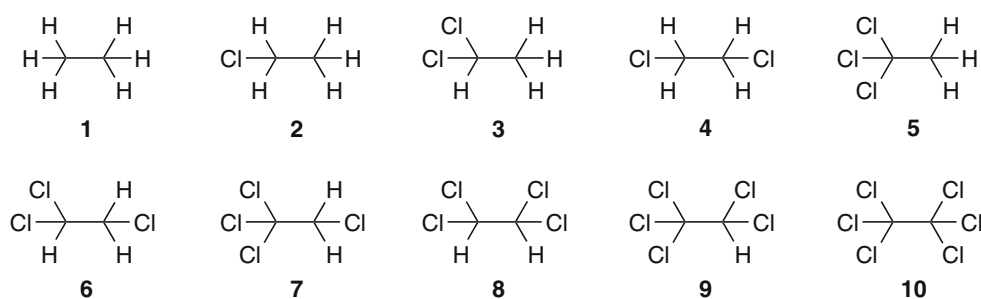
O. Beckstein
Structural Bioinformatics and Computational Biochemistry Unit, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK

Present Address:
O. Beckstein
Department of Physics, Arizona State University, P.O. Box 871504, Tempe, AZ 85287-1504, USA

B. I. Iorga (✉)
Institut de Chimie des Substances Naturelles, CNRS UPR 2301, Centre de Recherche de Gif-sur-Yvette, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette, France
e-mail: bogdan.iorga@icsn.cnrs-gif.fr

Introduction

The hydration free energy ΔG_{hyd}, i.e. the change in Gibbs free energy for the transfer of a molecule from the gas phase to aqueous solution at constant temperature and pressure, is of fundamental importance to characterize the distribution of compounds in chemical and biological systems. It has also become a standard test case for the evaluation of the predictive power of quantitative computational methods. The SAMPL challenges (SAMPL1 [1], SAMPL2 [2], SAMPL3 in this issue) in particular asked participants to blindly predict hydration free energies for compounds whose unpublished experimental values were only revealed after submission of the predictions. This has turned out to be a fruitful approach to assess quantitative computational methods without bias and point out areas where improvements are required [3]. Solvation free energies have also been used to either parameterize

Fig. 1 Chemical structures of subset 1, containing chlorinated ethane derivatives

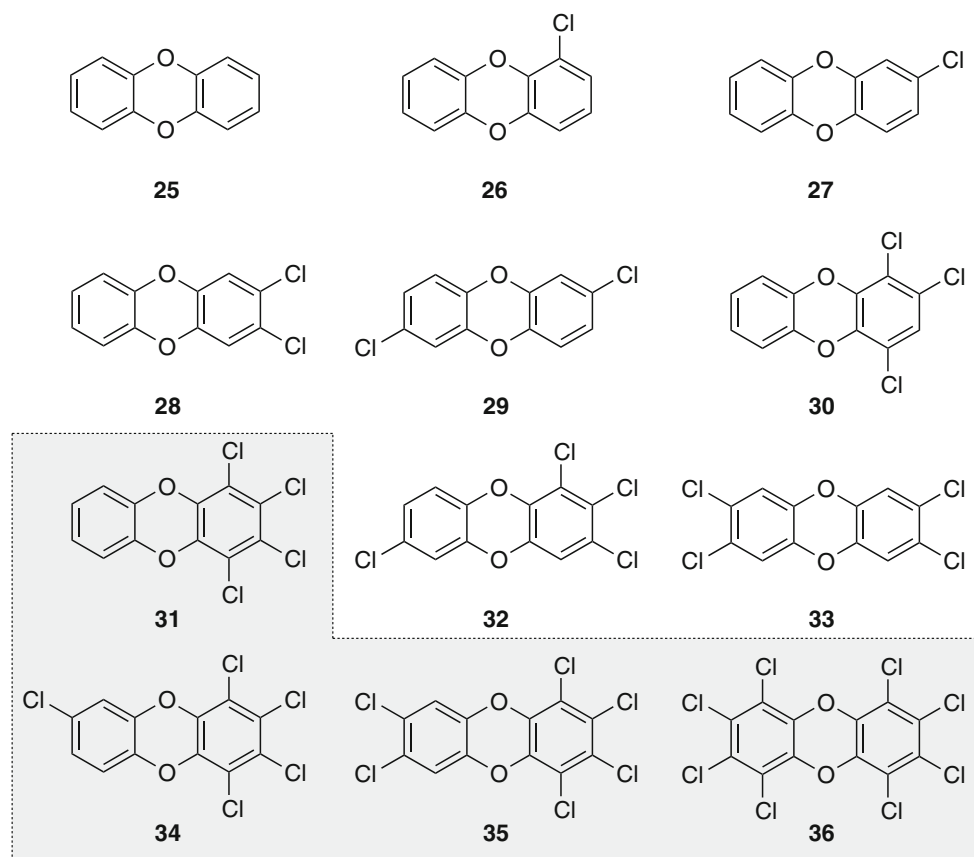**Fig. 2** Chemical structures of subset 2, containing polychlorinated biphenyl (PCB) derivatives. The *highlighted* compounds **15** and **19–24** contain aromatic ring systems with ≥4 chlorine substituents. Hydrogen atoms are omitted for clarity
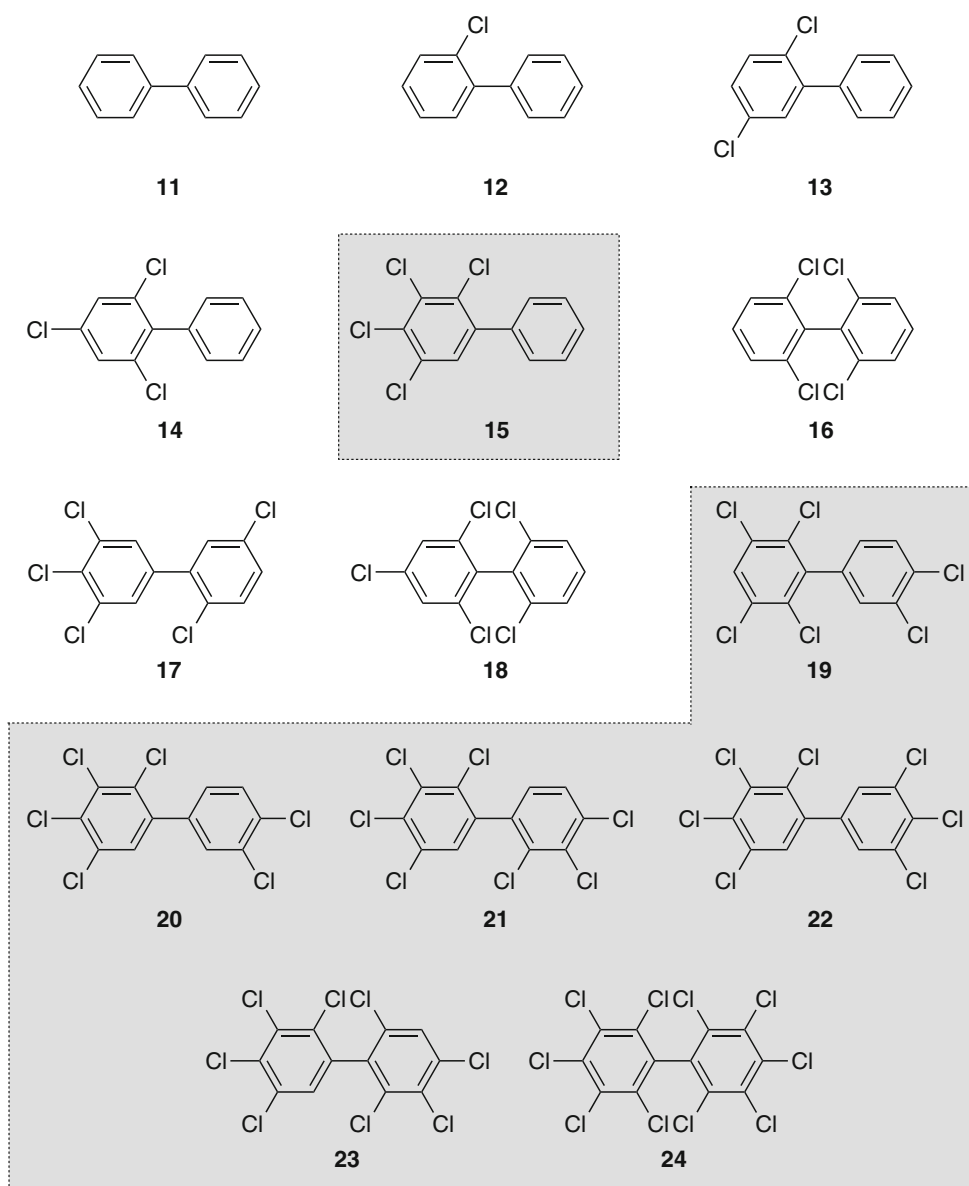
computational models such as classical additive force fields used in molecular mechanics [4–6] or to evaluate the quality of existing force fields [7, 8].

Here we are focusing on classical all-atom molecular dynamics (MD) simulations in explicit solvent with

additive and transferrable force fields. Our main interest is to assess the quality of parameterization of small molecules as such a parameterization is a crucial first step in studying the interactions of drug- or toxin-like molecules with biological systems. Methods based on classical all-atom MD

Fig. 3 Chemical structures of subset 3, containing polychlorinated dibenzo-*p*-dioxin (PCDD) derivatives. The *highlighted* compounds **31** and **34–36** contain aromatic ring systems with ≥4 chlorine substituents. Hydrogen atoms are omitted for clarity

simulations can provide precise predictions for hydration free energies [9] and current force fields are believed to be accurate to 1–1.5 kcal mol^{-1} for the prediction of hydration free energies of typical small organic molecules [3, 10–16]

The data set provided for the SAMPL3 challenge consists of 36 chloro-organic compounds in three distinct chemical subsets with varying numbers of chlorine substituents. Understanding the relationship between structure and physico-chemical properties for this family of compounds is important in order to model the spread of these persistent and often toxic organic pollutants through the environment [17]. Subset 1 contains polychlorinated derivatives of ethane with all possible substitution combinations and up to six chlorine atoms (compounds **1–10**, Fig. 1). Polychlorinated derivatives of biphenyl in subset 2 (**11–24**, Fig. 2) and of dibenzo-*p*-dioxin (subset 3, **25–36**, Fig. 3) contain aromatic ring systems with different substitution patterns for the chlorine substituents.

Computational studies predicting the hydration free energy for polychlorinated compounds are relatively rare in the literature. Previous work [18] described the prediction of hydration free energies for chlorinated ethanes using molecular dynamics simulations, in explicit solvent and

with the OPLS force field. As the original parameters gave very poor results for the prediction of absolute and relative hydration free energies, the non-bonded r^{12} coefficients for carbon, hydrogen and chlorine atoms were improved to reproduce experimental free energies of hydration [18]. More recently, hydration free energies of a data set of 57 polychlorobiphenyls (PCBs) were recently computed to a root mean square error (RMSE) of about 0.9 kcal mol^{-1} with an empirically corrected one-dimensional reaction site interaction model [19]. To our knowledge, no study dedicated to the prediction of hydration free energies of the polychlorinated dibenzo-*p*-dioxins (PCDD) family has been published to date.

Here we used all atom classical molecular dynamics (MD) simulations to compute the hydration energy from the interactions between the solute and water. These interactions are parameterized in a force field; for this work we used the OPLS-AA force field [5, 20–33] as a basis. The strongest non-bonded interaction in molecular systems is the electrostatic force, which is parameterized as a Coulomb force between point charges situated on the sites of the atoms. The point charges are typically not taken as full formal charges but as partial charges that, in the sum, reproduce the correct

Fig. 4 Correlation between
experimental and computed
hydration free energies. The
computed ΔG_{hyd} is shown
against the experimental value
for each parameterization
protocol **P1–P4**. Perfect
agreement is shown by the
diagonal line in each plot and
$\pm 1 \, \mathrm{kcal \, mol^{-1}}$ from ideal is
indicated by the shaded area.
Error bars denote the
experimental or computational
error at one standard deviation
from the mean. Computational
errors are typically smaller than
the marker symbols
($< 0.15 \, \mathrm{kcal \, mol^{-1}}$). The plot
range for **P1** is expanded
compared to the other protocols
in order to accommodate the
predictions for compounds
with ≥ 4 chlorine substituents
on aromatic rings that deviate
substantially from the
experimental values (*black filled
markers*)
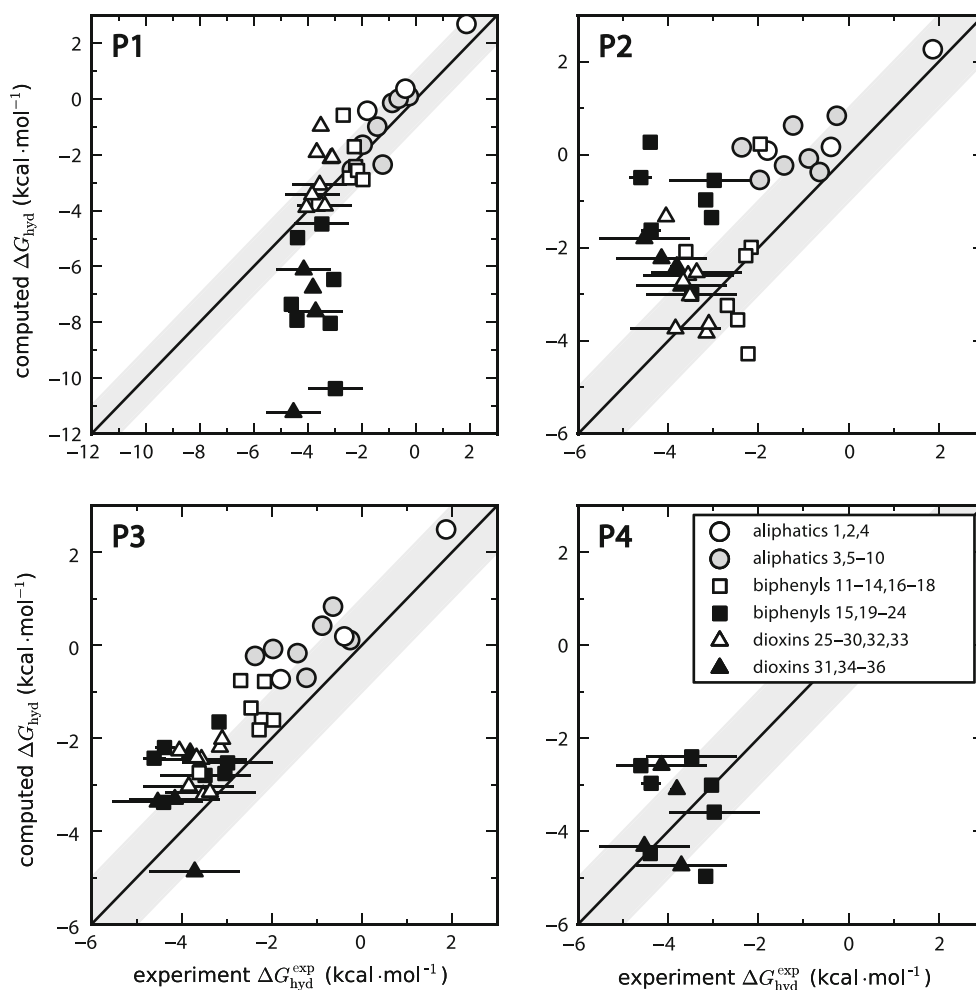
electrostatics of the molecule. In OPLS-AA, charges are
determined for small model compounds and then directly
transferred to an atom in another molecule that experiences
the same chemical environment as the atom in the model
compound. Water is a polar molecule and thus electrostatic
interactions typically dominate the hydration free energy.
Therefore, we evaluated three different protocols for generating the partial charges for the compounds **1–36**. Protocol
P1 essentially employed the native OPLS-AA charges
whereas P2 and P3 computed the charges at the quantum
mechanical (QM) level for each compound.

Comparison of the predicted hydration free energies
with the hydration free energies released in the SAMPL3
evaluation phase (Fig. 4) indicated that the QM partial
charges produced better predictions than the OPLS-AA
charges. In particular, compounds with at least 4 chlorine
substituents on aromatic rings had much too negative
hydration free energies in OPLS-AA. This observation
prompted us to re-parameterize the corresponding Cl and C
partial charges, using the hydration free energy of the
model compound hexachlorobenzene **37** (not in the test set)
as a target. We show that the resulting parameterization

protocol P4, which amounts to the introduction of two new
atom types into the standard OPLS-AA parameterization
protocol, performs better on the SAMPL3 set than any of
our other approaches P1–P3.

Methods

All compound structures were converted from SDF format to
PDB format with CORINA version 3.44 (http://www.molecular-networks.com). Analysis of molecular dynamics trajectories
was carried out with Gromacs tools (http://www.gromacs.org)
[34] and MDAnalysis [35]. The OPLS-AA force field files
that are part of Gromacs 4.5.3 [34] were used as a basis for
the parameterization of the compounds **1–37**.

Parameterization protocols

Topologies of all compounds were generated using an in-house developed script that assigns atom types based on the
OPLS-AA force field. The hydration free energy depends
strongly on the partial charges and hence three different

protocols P1–P3 were tested for charge generation. A fourth protocol, P4, was used to derive optimised partial charge parameters for chlorine and carbon atoms that were insufficiently represented in the current OPLS-AA force field. The Lennard-Jones and bonded interactions were kept the same for protocols P1–P4.

P1 OPLS-AA charges For parameterization protocol P1, the native charges from OPLS-AA were used. When these were missing, they were adapted from the existing ones (e.g. atom type *opls_152* with modified charges of 0.297 and 0.600 were used for C atoms in $RCHCl_2$ and $RCCl_3$, respectively; see "Parameterization and validation of new OPLS-AA atom types" section for a post-hoc validation of this approach).

P2 LMP2/cc-pVTZ(-F) QM charges For protocol P2, the charges were obtained after geometry optimization and electrostatic potential (ESP) fitting, using Jaguar (http://www.schrodinger.com) at the LMP2/cc-pVTZ(-F) level. This methodology has been previously used for the OPLS-AA force field development (protein amino acids and small organic molecules) [21, 26, 27]. The charges for chemically equivalent atoms obtained from the Jaguar calculations output were generally not identical, and they were adjusted manually before further use.

P3 HF/6-31G* QM charges For P3, the charges were obtained after geometry optimization and CHelpG charges calculation using Gaussian03 (http://www.gaussian.com) [36] at the HF/6-31G* level, then a two-stage RESP fitting using AmberTools (http://ambermd.org). This methodology has been previously used for heterocycles parameterization in the OPLS-AA force field [23, 24].

P4 New OPLS-AA parameters based on the hydration free energy of hexachlorobenzene New OPLS-AA atom types for chlorine and carbon were developed for the case of aromatic rings with four or more chlorine substituents, using the model compound hexachlorobenzene **37** (Fig. 5A). Lennard-Jones and bonded interactions were retained from OPLS-AA atom types *opls_264* (Cl) and *opls_263* (C), which were originally derived from chlorobenzene [37]. The partial charges of the Cl and the C atom were set to equal but opposite values ($q_{Cl} = -q_C$). They were adjusted until the computed hydration free energy matched the experimental estimate, $\Delta G_{hyd}^{exp} = -2.26$ kcal/mol, which had been calculated as the mean from two different experimental measurements [19].

Hydration free energy calculation

Hydration free energies were calculated via free energy perturbation (FEP) molecular dynamics (MD) simulations
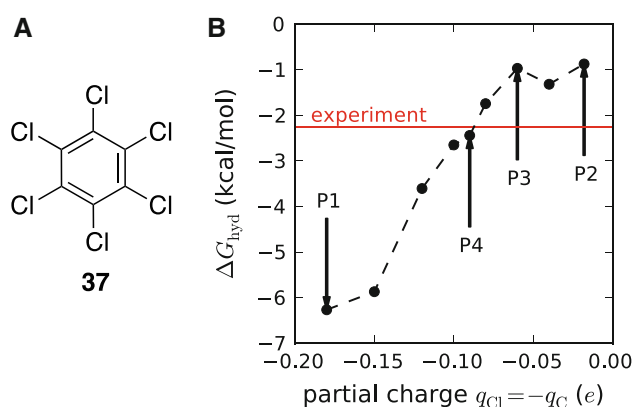
Fig. 5 Model compound hexachlorobenzene (**37**). **a** Chemical structure. **b** Calculated hydration free energy ΔG_{hyd} as a function of the partial charge q_{Cl} on the chlorine substituent or $-q_{Cl}$ on the C_{ipso} atom bearing the chlorine substituent. The *red line* shows the average of the known experimental values, -2.26 kcal/mol [19]. *Arrows* indicate values generated from the protocols P1–P3 discussed in the text. The new value q_{Cl} proposed for OPLS-AA is indicated by *P4*

of each molecule in a water box. All simulations were performed with Gromacs 4.5.3 [34]. Processing and analysis of the free energy simulations was carried out with scripts based on the GromacsWrapper Python tool kit (O. Beckstein, http://github.com/orbeckst/GromacsWrapper).

The FEP protocol follows the works of Mobley and colleagues [10], with some small differences. A molecule was solvated with TIP4P water [38] in a dodecahedral periodic simulation cell with at least 1.0 nm between the solute and the box surfaces. Initial equilibrium *NPT* MD simulation ($T = 300$ K, $P = 1$ bar) of the molecule were run for 50 ns. The simulations were run as Langevin dynamics (integration time step 2 fs) for temperature control, with the friction coefficient for each particle computed as mass/0.1 ps. The average pressure was held constant with an isotropic Berendsen barostat (relaxation time constant $\tau_p = 1$ ps and compressibility $\kappa_p = 4.6 \times 10^{-5}$ bar^{-1}). The grid-based neighbor list was updated every five time steps. Lennard-Jones interactions were calculated up to a cutoff of 1 nm and a dispersion correction (implemented in Gromacs) was applied to energy and pressure to account for van der Waals interactions beyond the cutoff in a mean field manner. Coulomb interactions were handled with the SPME method [39] (short range cutoff 1 nm, 0.12 nm Fourier grid spacing, sixth order spline interpolation). Bonds containing hydrogen atoms were constrained with the P-LINCS algorithm [40] (fourth order expansion with a single iteration).

The last frame of the *NPT* simulation was used as the input for the free energy perturbation (FEP) simulations. FEP calculations were performed in the *NVT* ensemble without a barostat but used the same parameters as the *NPT*

simulations with the exception of FEP specific alterations and a higher P-LINCS order of 12.

Simulations were run on up to eight cores. The domain decomposition algorithm in Gromacs 4.5.3 did not always handle well the simulations for these relatively small systems and therefore most of them were run with the particle-based decomposition. The simulation system sizes were fairly small (8–22 atoms for solute and 1,100–2,600 atoms for solvent) and did not substantially benefit from parallelizing a single simulation further. However, the automated FEP protocol described below requires 21 simulations that can all be carried out independently. One eight-core workstation was sufficient to calculate one hydration free energy in 1–2 days, depending on the size of the system.

The FEP protocol computes the free energy difference between the solvated molecule and the molecule in the gas phase by constructing an alchemical (non-physical) pathway during which the interactions of the molecule with the solvent are stepwise decoupled. A coupling parameter λ for each interaction is used to scale the interaction: $\lambda = 0$ indicates that the interaction is at full strength and at 1 it is switched off. Coulomb interactions were linearly reduced to zero over five windows (coupling parameter $\lambda_{Coul} \in \{0, 0.25, 0.5, 0.75, 1\}$) while the van der Waals (Lennard-Jones) interactions were maintained (i.e. $\lambda_{vdW} = 0$); sixteen windows were used to switch off the Lennard-Jones term for the uncharged solute ($\lambda_{Coul} = 1$ and $\lambda_{vdW} \in \{0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1\}$). Each window was simulated for 5 ns. The van der Waals calculations used soft core potentials with the values suggested by [10] ($\alpha = 0.5$, power 1, and $\sigma = 0.3$ nm). The calculations made use of the "couple-intramol = no" feature in Gromacs [34] (new in release 4.x), which performs a decoupling simulation that leaves intra-molecular interactions unchanged and thus avoids having to calculate gas-phase contributions separately. The derivative of the Hamiltonian $\mathscr{H}$ with respect to the coupling parameter λ, $\partial\mathscr{H}/\partial\lambda$, was saved for every time step. Thermodynamic integration (TI) over the average derivatives of all windows,

$$\Delta G = \int_0^1 \left\langle \frac{\partial\mathscr{H}}{\partial\lambda} \right\rangle d\lambda, \tag{1}$$

yielded the free energy contributions for the decharging and decoupling steps. Eq. 1 was integrated numerically with the composite Simpson's rule [41] as implemented in SciPy (http://www.scipy.org).

The total hydration free energy (transfer from gas phase to aqueous phase at the 1M/1M Ben-Naim standard state) is calculated as $\Delta G_{hyd} = -(\Delta G_{Coul} + \Delta G_{vdW})$. Note that any corrections for simulating at constant volume have been neglected (and hence technically our transfer energy

is a Helmholtz free energy). In the last stages of preparation of this manuscript for submission, these corrections have been evaluated more in depth and we realized that they can be significant in some cases. Preliminary results indicate that correct treatment of these corrections will substantially improve our results.

The error on the hydration free energy is estimated from the errors of the individual $\langle\partial\mathscr{H}/\partial\lambda\rangle$ of each FEP window. The error δ of the mean $\langle\partial\mathscr{H}/\partial\lambda\rangle$ is calculated as

$$\delta = \sqrt{2t_c C(0)\tau^{-1}} \tag{2}$$

where $C(t)$ is the autocorrelation function of the fluctuations around the mean, $\partial\mathscr{H}/\partial\lambda - \langle\partial\mathscr{H}/\partial\lambda\rangle$, t_c the correlation time (assuming an inital exponential decay of the autocorrelation function $C(t) \sim \exp(-t/t_c)$), and τ the total length of the simulation [42]. The error on each TI integral is calculated analytically via propagation of errors through Simpson's rule (implemented in GromacsWrapper as an extension of the SciPy functionality for Simpson's rule). The final error on ΔG_{hyd} is $(\delta_{Coul}^2 + \delta_{vdW}^2)^{1/2}$.

Results and discussion

The main results of the hydration free energy calculations are summarized in Fig. 4 and detailed in Table 1.

Predicted hydration free energies

For each molecule, hydration free energy calculations were carried out using the three topologies generated with protocols P1–P3. The corresponding three sets of results ($\Delta G_{hyd}^{(1)}$ to $\Delta G_{hyd}^{(3)}$ in Table 1) were submitted to the SAMPL3 challenge. The accuracy of the computed ΔG_{hyd} was quantified by computing the root mean square error (RMSE) from the experimental hydration free energies, which were released by the SAMPL3 organisers after the submission of the predictions. A RMSE to within chemical accuracy, i.e. about 1 kcal mol^{-1}, can be considered sufficient for many applications [3].

At the time of submission, experimental values for the hydration free energy were known for two compounds from this dataset: ethane **1** and hexachloroethane **10**, the latter being present in the SAMPL2 data set [2]. For both compounds, the second protocol (P2) gave the best results (RMSE 0.3 kcal mol^{-1}) and thus from this very limited data set, P2 appeared to be the most promising approach. However, comparison of the predictions for the whole data set **1–36** to the released experimental data (Fig. 4) showed that P3 (QM charges from Gaussian03) produced the best blind predictions with a RMSE of 1.2 kcal mol^{-1} compared to

Table 1 Computed and experimental hydration free energies ΔG_{hyd} with error estimate (in kcal mol^{-1}) for all compounds (denoted by "id")

id	Exp. ΔG_{hyd}^{exp}	Protocol 1 $\Delta G_{hyd}^{(1)}$	$D^{(1)}$	Protocol 2 $\Delta G_{hyd}^{(2)}$	$D^{(2)}$	Protocol 3 $\Delta G_{hyd}^{(3)}$	$D^{(3)}$	Protocol 4 $\Delta G_{hyd}^{(4)}$	$D^{(4)}$
1	1.87 (10)	2.69 (3)	0.82	2.27 (3)	0.40	2.49 (3)	0.62		
2	−0.39 (10)	0.37 (4)	0.76	0.17 (4)	0.56	0.19 (4)	0.58		
3	−0.88 (10)	−0.14 (5)	0.74	−0.08 (4)	0.80	0.42 (4)	1.30		
4	−1.80 (10)	−0.42 (7)	1.38	0.09 (6)	1.89	−0.73 (6)	1.07		
5	−0.26 (10)	0.10 (5)	0.36	0.84 (5)	1.10	0.11 (5)	0.37		
6	−1.97 (10)	−1.64 (20)	0.33	−0.54 (6)	1.43	−0.08 (7)	1.89		
7	−1.43 (10)	−0.99 (5)	0.44	−0.23 (5)	1.20	−0.17 (5)	1.26		
8	−2.37 (10)	−2.53 (14)	0.16	0.16 (8)	2.53	−0.23 (9)	2.14		
9	−1.23 (10)	−2.35 (6)	1.12	0.63 (6)	1.86	−0.70 (6)	0.53		
10	−0.64 (10)	0.01 (7)	0.65	−0.37 (7)	0.27	0.83 (6)	1.47		
RMSE **1–10**			0.76		1.39		1.26		
RMSE **1,2,4**			1.03						
RMSE **3,5–10**			0.62						
11	−2.23 (10)	−2.42 (7)	0.19	−4.28 (8)	2.05	−1.60 (7)	0.63		
12	−2.69 (10)	−0.58 (8)	2.11	−3.24 (7)	0.55	−0.76 (7)	1.93		
13	−2.46 (10)	−2.83 (8)	0.37	−3.55 (8)	1.09	−1.35 (8)	1.11		
14	−2.16 (10)	−2.56 (8)	0.40	−1.99 (8)	0.17	−0.78 (9)	1.38		
15	−3.48 (1.00)	−4.48 (9)	1.00	−3.00 (9)	0.48	−2.80 (9)	0.68	−2.40 (9)	1.08
16	−2.28 (12)	−1.71 (10)	0.57	−2.17 (9)	0.11	−1.82 (10)	0.46		
17	−3.61 (13)	−3.77 (9)	0.16	−2.08 (9)	1.53	−2.74 (9)	0.87		
18	−1.96 (10)	−2.89 (10)	0.93	0.23 (9)	2.19	−1.61 (10)	0.35		
19	−4.38 (22)	−4.97 (11)	0.59	−1.62 (10)	2.76	−2.20 (10)	2.18	−2.97 (11)	1.41
20	−3.04 (10)	−6.47 (11)	3.43	−1.35 (9)	1.69	−2.76 (10)	0.28	−3.01 (10)	0.03
21	−4.40 (10)	−7.93 (10)	3.53	0.27 (10)	4.67	−3.38 (11)	1.02	−4.48 (11)	0.08
22	−3.17 (10)	−8.04 (11)	4.87	−0.97 (11)	2.20	−1.65 (11)	1.52	−4.97 (10)	1.80
23	−4.61 (25)	−7.36 (11)	2.75	−0.49 (13)	4.12	−2.43 (10)	2.18	−2.59 (11)	2.02
24	−2.98 (1.00)	−10.38 (12)	7.40	−0.55 (11)	2.43	−2.53 (12)	0.45	−3.59 (11)	0.61
RMSE **11–24**			2.90		2.29		1.25		
RMSE **15,19–24**			4.31		3.17		1.51		1.35
25	−3.15 (10)	−2.13 (7)	1.02	−3.83 (7)	0.68	−2.19 (7)	0.96		
26	−3.52 (10)	−0.97 (7)	2.55	−3.03 (8)	0.49	−3.22 (8)	0.30		
27	−3.10 (10)	−2.10 (8)	1.00	−3.64 (8)	0.54	−2.02 (7)	1.08		
28	−3.56 (1.00)	−3.08 (9)	0.48	−2.59 (8)	0.97	−2.45 (8)	1.11		
29	−3.67 (12)	−1.91 (8)	1.76	−2.72 (9)	0.95	−2.42 (9)	1.25		

Table 1 continued

id	Exp. ΔG_{hyd}^{exp}	Protocol 1 $\Delta G_{hyd}^{(1)}$	$D^{(1)}$	Protocol 2 $\Delta G_{hyd}^{(2)}$	$D^{(2)}$	Protocol 3 $\Delta G_{hyd}^{(3)}$	$D^{(3)}$	Protocol 4 $\Delta G_{hyd}^{(4)}$	$D^{(4)}$
30	−4.05 (10)	−3.88 (9)	0.17	−1.33 (9)	2.72	−2.27 (9)	1.78		
31	−3.81 (14)	−6.77 (10)	2.96	−2.40 (10)	1.41	−2.31 (9)	1.50	−3.10 (9)	0.71
32	−3.84 (1.00)	−3.43 (9)	0.41	−3.74 (9)	1.00	−3.03 (9)	0.81		
33	−3.37 (1.00)	−3.83 (9)	0.46	−2.53 (10)	0.84	−3.16 (9)	0.21		
34	−4.15 (1.00)	−6.12 (10)	1.97	−2.23 (9)	1.92	−3.31 (10)	0.84	−2.58 (10)	1.57
35	−3.71 (1.00)	−7.62 (10)	3.91	−2.82 (10)	0.89	−4.87 (10)	1.16	−4.74 (11)	1.03
36	−4.53 (1.00)	−11.23 (12)	6.70	−1.80 (11)	2.73	−3.37 (11)	1.16	−4.32 (11)	0.21
RMSE 25–36			2.66		1.44		1.10		
RMSE 31,34–36			4.27		1.87		1.19		1.01
RMSE 1–36			2.41		1.80		1.21		
RMSE 15,19–24,31,34–36			2.49		2.72		1.39		1.22

The absolute difference between computed and experimental hydration free energy is shown for each compound and parameterization protocol i as $D_{id}^{(i)} = |\Delta G_{hyd,id}^{(i)} - \Delta G_{hyd,id}^{exp}|$. The standard error in the last significant digits is given in parentheses. The root mean square error (RMSE) $\sqrt{N^{-1}\sum_{id}^{N} D_{id}^2}$ is listed for chemically distinct classes of compounds (see text)

1.8 kcal mol^{-1} for P2 (QM charges from Jaguar). Compared to the QM-derived charges, the standard OPLS-AA approach (P1) was less accurate with RMSE 2.4 kcal mol^{-1} (Table 1).

A more detailed analysis in terms of distinct chemical groups showed a more diverse picture. In subset 1, which contains the chloro ethane derivatives **1–10**, the trend was completely reversed. In this case, protocol P1 provided the most accurate predictions, with a RMSE value of 0.8 kcal mol^{-1}. The QM-based methods did show smaller errors for the more complicated compounds in subsets 2 (**11–24**) and 3 (**25–36**) than the standard OPLS-AA charges where P3 produced more accurate results than P2 or P1 (Table 1).

The encouraging results for the standard OPLS-AA force field for subset 1 suggested that the force field is able to reach near chemical accuracy for chloro-organic compounds and its failure to do so for subsets 2 and 3 might have a cause lying in the chemical structure of the compounds. It turned out that compounds with at least four chlorine substituents on an aromatic ring (**15**, **19–24**, **31**, **34–36**) had substantially too negative hydration free energies (RMSE 4.3 kcal mol^{-1} compared to the best RMSE 1.4 kcal mol^{-1} for P3). This was particularly obvious when the correlation between experimental and computed values was plotted as in Fig. 4. These results prompted us to develop new force field parameters for these difficult compounds.

Parameterization and validation of new OPLS-AA atom types

Parameterization of aliphatic C atoms in RCHCl$_2$ *and* RCCl$_3$ In the original OPLS-AA force field, no parameters were available for aliphatic carbon atoms bearing two or three chlorine substituents, only aliphatic C atoms of RCH$_2$Cl type being parameterized (available as *opls_152*), starting from 1,2-dichloroethane [37]. As several compounds from the SAMPL3 dataset (**3**, **5–10**) contained the RCHCl$_2$ and RCCl$_3$ motifs, two new atom types were defined starting from *opls_152* and following the same OPLS-AA philosophy. Consequently, modified charges of 0.297 and 0.600 were used in protocol P1 for C atoms in RCHCl$_2$ and RCCl$_3$, respectively, all other parameters being identical with *opls_152* (Table 2). The values of these charges were chosen in order to conserve the overall neutrality of the molecule and make use of the existing atom types [37] for hydrogen and chlorine atoms in alkyl chlorides (*opls_153* and *opls_151*, respectively). The predictions for the compounds containing these new atom types (**3**, **5–10**) were in very good agreement with the experimental hydration free energy values (RMSE 0.62 kcal mol^{-1}, see Table 1), better than the predictions

Table 2 New OPLS-AA parameters for carbon atoms in RCHCl$_2$ and RCCl$_3$ (*oplsa_152A* and *oplsa_152B*) and carbon (*oplsa_263A*) and chlorine (*oplsa_264A*) atoms in aromatic rings with ≥4 chlorine substituents

name[a]	type[b]	Z[c]	m (u)[d]	$q(e)$[e]	σ (nm)[f]	ϵ(kJ mol^{-1})[g]
opls_152A	CT	6	12.011	0.297	0.350	0.276144; **C** in RCHCl2
opls_152B	CT	6	12.011	0.600	0.350	0.276144; **C** in RCCl3
opls_263A	CA	6	12.011	0.090	0.355	0.29288; **C**(Cl) perchlorobenzenes
opls_264A	Cl	17	35.453	−0.090	0.340	1.25520; **Cl** perchlorobenzenes

[a] Proposed OPLS-AA atom type name

[b] Bonded type

[c] Atomic number

[d] Atomic mass in atomic mass constants $m_u = 1.660538921 \times 10^{-27}$ kg

[e] Partial charge in elementary charges $e = 1.602176565 \times 10^{-19}$ C

[f] Length parameter of the OPLS-AA Lennard-Jones potential $V_{LJ}(r) = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6)]$[5]

[g] Energy well depth of the OPLS-AA Lennard-Jones potential $V_{LJ}(r)$

for the compounds making use of "classical" OPLS-AA parameters (**1–2**,**4**, RMSE 1.03 kcal mol^{-1}). In these conditions, we believe that these two new atom types are validated and appropriate for further use in molecular dynamics simulations with the OPLS-AA force field.

Parameterization of Cl atoms in aromatic rings with ≥4 chloro substituents Three different atom types are available in the OPLS-AA force field for fluorobenzenes (*opls_719*, *opls_728* and *opls_721*, defined for systems with one, two and six fluorine substituents on the aromatic ring, respectively). On the other hand, only one atom type (*opls_264*) is available for chlorobenzenes, corresponding to the monosubstituted aromatic rings [37]. As already stressed above, the compounds with four or more chloro substituents on the aromatic ring (**15**, **19–24**, **31**, **34–36**) contributed most to the overall prediction error for the hydration free energy in the SAMPL3 dataset (Table 1, Fig. 4). A possible explanation for these results is that the *opls_264* atom type is not appropriate for use with these highly substituted systems, possibly due to interactions between the chlorine atoms as also found in a different study [19]. Within the SAMPL3 challenge, we were able to obtain improved predictions for these difficult compounds using modified charges generated from QM calculations (protocols P2 and P3, with RMSE 2.72 kcal mol^{-1} and 1.39 kcal mol^{-1}, respectively, compared with 4.29 kcal mol^{-1} using protocol P1). However, with protocols P2 and P3 the charge on Cl atoms always differs according to their environment within the molecules. Such variable charges are not compatible with the OPLS-AA philosophy and make it impossible to transfer these charges to other, chemically similar compounds. With approaches P2 and P3, a full QM calculation would have to be carried out for each new compound of interest. The SAMPL3 data provided an opportunity to define a new transferrable OPLS-AA atom type for Cl atoms in highly substituted aromatic systems (perchlorobenzenes). We used hexachlorobenzene (**37**, Fig. 5a), which is not included in the SAMPL3

data set, but was present in the SAMPL2 data set [2], as a model compound. Starting from the standard OPLS-AA atom type *opls_264* for Cl (protocol P1), we computed ΔG_{hyd} as a function of the partial charge q_{Cl} on the Cl atom (Fig. 5b). The aromatic C$_{ipso}$ atom bearing the Cl substituent was based on *opls_263* and was assigned the same charge as the Cl atom but with opposite sign ($q_{Cipso} = -q_{Cl}$). As a target value for ΔG_{hyd} we employed the experimental estimate of −2.26 kcal mol^{-1}, which was also used by Ratkova et al [19]. The standard OPLS-AA parameters ($q_{Cl} = -0.180$) yielded a ΔG_{hyd} too negative by 4.01 kcal mol^{-1}, consistent with the general trend observed for the aromatic compounds (**11–36**) from the SAMPL3 set (P1). The QM approaches produced smaller Cl charges of −0.060 (P3) and −0.018 (P2), leading to ΔG_{hyd} being too positive by 1.29 and 1.39 kcal mol^{-1}, respectively. We then systematically varied q_{Cl} to match the target hydration free energy and finally chose $q_{Cl} = -0.090$ with a signed deviation of −0.18 kcal mol^{-1} to define new atom types *opls_263A* (C$_{ipso}$) and *opls_264A* (Cl) for perchlorobenzenes (Table 2).

The new atom types (termed protocol P4) were tested for the compounds from the SAMPL3 data set with ≥4 chlorine atoms on an aromatic ring (Table 1, Fig. 4). Gratifyingly, the RMSE for these outlier compounds improved from 4.29 kcal mol^{-1} using P1 to 1.22 kcal mol^{-1} using P4. Therefore, using the appropriate OPLS-AA parameters (the original force field parameters and the ones proposed in this work), the hydration free energies for all compounds from the SAMPL3 dataset could be calculated with a RMSE of 1.0 kcal mol^{-1}.

Conclusion

In this study, hydration free energies for the SAMPL3 dataset were predicted using molecular dynamics

simulations with the OPLS-AA force field. Three different protocols for charge generation were used (the original OPLS-AA charges and two variants of QM-derived charges). One of the QM-derived protocols gave the best results, leading to a RMSE value of $1.2 \, \text{kcal mol}^{-1}$ for the whole dataset, whereas the original OPLS-AA charges yielded a RMSE of $2.4 \, \text{kcal mol}^{-1}$. The largest contribution to this RMSE was due to compounds with four or more chlorine atoms on the aromatic ring. A new OPLS-AA atom type was developed for these difficult compounds, which allowed a substantial improvement in prediction error. Consequently, using the appropriate OPLS-AA parameters (the original force field parameters and the ones proposed in this work) we are now able to compute the hydration free energies for the whole SAMPL3 dataset with a RMSE of $1.0 \, \text{kcal mol}^{-1}$.

References

1. Guthrie JP (2009) A blind challenge for computational solvation free energies: introduction and overview. J Phys Chem B 113(14):4501–4507. doi:10.1021/jp806724u

2. Geballe MT, Skillman AG, Nicholls A, Guthrie JP, Taylor PJ (2010) The SAMPL2 blind prediction challenge: introduction and overview. J Comput Aided Mol Des 24(4):259–279. doi:10.1007/s10822-010-9350-8

3. Klimovich PV, Mobley DL (2010) Predicting hydration free energies using all-atom molecular dynamics simulations and multiple starting conformations. J Comput Aided Mol Des 24(4):307–316. doi:10.1007/s10822-010-9343-7

4. Mackerell AD (2004) Empirical force fields for biological macromolecules: overview and issues. J Comp Chem 25(13):1584–1604. doi:10.1002/jcc.20082

5. Jorgensen WL, Tirado-Rives J (2005) Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. Proc Natl Acad Sci USA 102(19):6665–6670. doi:10.1073/pnas.0408037102

6. van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, Gee P, Geerke DP, Glattli A, Hunenberger PH, Kastenholz MA, Oostenbrink C, Schenk M, Trzesniak D, van der Vegt NFA, Yu HB (2006) Biomolecular modeling: goals, problems, perspectives. Angew Chem Int Ed 45(25):4064–4092. doi:10.1002/anie.200502655

7. Helms V, Wade R (1997) Free energies of hydration from thermodynamic integration: Comparison of molecular mechanics force fields and evaluation of calculation accuracy. J Comput Chem 18(4):449–462. doi:10.1002/(SICI)1096-987X(199703)18:4<449::AID-JCC1>3.0.CO;2-T

8. Geerke DP, van Gunsteren WF (2006) Force field evaluation for biomolecular simulation: free enthalpies of solvation of polar and apolar compounds in various solvents. Chem Phys Chem 7(3):671–678. doi:10.1002/cphc.200500510

9. Shirts MR, Pande VS (2005) Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. J Chem Phys 122(13):134508. doi:10.1063/1.1877132

10. Mobley DL, Dumont E, Chodera JD, Dill KA (2007) Comparison of charge models for fixed-charge force fields: Small-molecule hydration free energies in explicit solvent. J Phys Chem B 111(9):2242–2254. doi:10.1021/jp0667442

11. Mobley DL, Bayly CI, Cooper MD, Dill KA (2009) Predictions of hydration free energies from all-atom molecular dynamics simulations. J Phys Chem B 113(14):4533–4537. doi:10.1021/jp806838b

12. Mobley DL, Bayly CI, Cooper MD, Shirts MR, Dill KA (2009) Small molecule hydration free energies in explicit solvent: an extensive test of fixed-charge atomistic simulations. J Chem Theory Comput 5(2):350–358. doi:10.1021/ct800409d

13. Sulea T, Corbeil C, Purisima E (2010) Rapid prediction of solvation free energy. 1. An extensive test of linear interaction energy (LIE). J Chem Theory Comput 6(5):1608–1621. doi:10.1021/ct9006025

14. Purisima EO, Corbeil CR, Sulea T (2010) Rapid prediction of solvation free energy. 3. Application to the SAMPL2 challenge. J Comput Aided Mol Des 24(4):373–383. doi:10.1007/s10822-010-9341-9

15. Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, Sherman W (2010) Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. J Chem Theory Comput 6(5):1509–1519. doi:10.1021/ct900587b

16. Baker C, Lopes P, Zhu X, Roux B, MacKerell A Jr (2010) Accurate calculation of hydration free energies using pair-specific Lennard-Jones parameters in the CHARMM Drude polarizable force field. J Chem Theory Comput 6(4):1181–1198. doi:10.1021/ct9005773

17. Henschler D (1994) Toxicity of chlorinated organic compounds: effects of the introduction of chlorine in organic molecules. Angew Chem Int Ed Engl 33(19):1920–1935. doi:10.1002/anie.199419201

18. Paulsen M, Straatsma T (1996) Chlorinated ethanes in aqueous solution: parameterization based on thermodynamics of hydration. Chem Phys Lett 259(1–2):142–145. doi:10.1016/0009-2614(96)00729-4

19. Ratkova E, Fedorov M (2011) Combination of RISM and cheminformatics for efficient predictions of hydration free energy of polyfragment molecules: application to a set of organic pollutants. J Chem Theory Comput 7(5):1450–1457. doi:10.1021/ct100654h

20. Kaminski G, Duffy E, Matsui T, Jorgensen W (1994) Free energies of hydration and pure liquid properties of hydrocarbons from the OPLS all-atom model. J Phys Chem 98(49):13077–13082. doi:10.1021/j100100a043

21. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118(45):11225–11236. doi:10.1021/ja9621760

22. Damm W, Frontera A, Tirado-Rives J, Jorgensen W (1997) OPLS all-atom force field for carbohydrates. J Comput Chem 18(16):1955–1970. doi:10.1002/(SICI)1096-987X(199712)18:16<1955::AID-JCC1>3.0.CO;2-L

23. Jorgensen WL, McDonald NA (1998) Development of an all-atom force field for heterocycles. Properties of liquid pyridine and diazenes. J Mol Struct Theochem 424(1-2):145–155. doi:10.1016/S0166-1280(97)00237-6

24. McDonald NA, Jorgensen WL (1998) Development of an all-atom force field for heterocycles. Properties of liquid pyrrole, furan, diazoles, and oxazoles. J Phys Chem B 102(41):8049–8059. doi:10.1021/jp981200o

25. Rizzo RC, Jorgensen WL (1999) OPLS all-atom model for amines: Resolution of the amine hydration problem. J Am Chem Soc 121(20):4827–4836. doi:10.1021/ja984106u

26. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J Phys Chem B 105(28):6474–6487. doi:10.1021/jp003919d

27. Watkins EK, Jorgensen WL (2001) Perfluoroalkanes: Conformational analysis and liquid-state properties from ab initio and Monte Carlo calculations. J Phys Chem A 105(16):4118–4125. doi:10.1021/jp004071w

28. Price M, Ostrovsky D, Jorgensen W (2001) Gas-phase and liquid-state properties of esters, nitriles, and nitro compounds with the OPLS-AA force field. J Comput Chem 22(13):1340–1352. doi:10.1002/jcc.1092

29. Kony D, Damm W, Stoll S, Van Gunsteren W (2002) An improved OPLS-AA force field for carbohydrates. J Comput Chem 23(15):1416–1429. doi:10.1002/jcc.10139

30. Kahn K, Bruice T (2002) Parameterization of OPLS-AA force field for the conformational analysis of macrocyclic polyketides. J Comput Chem 23(10):977–996. doi:10.1002/jcc.10051

31. Thomas L, Christakis T, Jorgensen W (2006) Conformation of alkanes in the gas phase and pure liquids. J Phys Chem B 110(42):21198–21204. doi:10.1021/jp064811m

32. Jorgensen W, Jensen K, Alexandrova A (2007) Polarization effects for hydrogen-bonded complexes of substituted phenols with water and chloride ion. J Chem Theory Comput 3(6):1987–1992. doi:10.1021/ct7001754

33. Xu Z, Luo HH, Tieleman DP (2007) Modifying the OPLS-AA force field to improve hydration free energies for several amino acid side chains using new atomic charges and an off-plane charge model for aromatic residues. J Comput Chem 28(3):689–697. doi:10.1002/jcc.20560

34. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4(3):435–447. doi:10.1021/ct700301q

35. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O (2011) MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. J Comput Chem 32:2319–2327. doi:10.1002/jcc.21787

36. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci 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 E.01. Gaussian, Inc., Wallingford

37. Jorgensen W, Ulmschneider J, Tirado-Rives J (2004) Free energies of hydration from a generalized born model and an all-atom force field. J Phys Chem B 108(41):16264–16270. doi:10.1021/jp0484579

38. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926–935. doi:10.1063/1.445869

39. Essman U, Perela L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577–8592. doi:10.1063/1.470117

40. Hess B (2008) P-LINCS: A parallel linear constraint solver for molecular simulation. J Chem Theory Comput 4(1):116–122. doi:10.1021/ct700200b

41. Jorge M, Garrido N, Queimada A, Economou I, Macedo E (2010) Effect of the integration method on the accuracy and computational efficiency of free energy calculations using thermodynamic integration. J Chem Theory Comput 6(4):1018–1027. doi:10.1021/ct900661c

42. Frenkel D, Smit B (2002) Understanding molecular simulations, 2nd edn. Academic Press, San Diego