A Toolkit to Fit Nonbonded Parameters from and for Condensed Phase Simulations

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

A Toolkit to Fit Nonbonded Parameters from and for Condensed Phase Simulations

Florent Hédin,[†,¶] Krystel El Hage,[†,¶] and Markus Meuwly[∗,†,‡]

†*Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel,*
Switzerland

‡*Department of Chemistry, Brown University, Providence, Rhode Island 02912, USA.*

¶*Contributed equally to this work*

E-mail: m.meuwly@unibas.ch

Abstract

The quality of atomistic simulations depends decisively on the accuracy of the underlying energy function (force field). Of particular importance for condensed-phase properties are nonbonded interactions, including the electrostatic and Lennard-Jones terms. Permanent atomic multipoles (MTPs) are an extension to common point-charge (PC) representations in atomistic simulations. MTPs are commonly determined from and fitted to an *ab initio* Electrostatic Potential (ESP), and Lennard-Jones (LJ) parameters are obtained from comparison of experimental and computed observables using Molecular Dynamics (MD) simulations. For this a set of thermodynamic observables such as density, heat of vaporization and hydration free energy is chosen, to which the parametrization is fitted. The current work introduces a comprehensive computing environment (Fitting Wizard - FW) for optimizing nonbonded interactions for atomistic force fields of different qualities. The FW supports fitting of standard PC-based force fields and more physically motivated multipolar (MTP) force fields. A broader study including 20 molecules ranging from N-methyl-acetamide and benzene

to halogenated benzenes, phenols, anilines and pyridines yields a root mean squared deviation for hydration free energies of 0.36 kcal/mol over a range of 8 kcal/mol. It is furthermore shown that PC-based force fields are not necessarily inferior compared to MTP parametrizations depending on the molecule considered.

Introduction

Consistent and convenient force field parametrization remains one of the main challenges for more widespread use and high quality atomistic simulations of complex systems. Although considerable progress has been made in implementing advanced treatments of intermolecular interactions, such as multipolar[1–3] and/or polarizable[4–6] force fields, their parametrization still presents a major impediment. Typically, force fields need to be fitted to a heterogeneous set of reference data originating from electronic structure calculations and experiment.[7–9] While fitting to reference energies from *ab initio* calculations is standard and only requires individual energy evaluations, using condensed-phase data such as diffusion coefficients or hydration free energies necessitates entire molecular dynamics (MD) runs.[10] This makes such parametrizations also computationally demanding.

Due to the fundamental importance of accurate descriptions of the inter- and intramolecular energetics, several tools have been developed which make force field parametrizations more amenable. Often, these approaches rely on databases and employ analogies between molecules or functional groups to minimize computational effort. Such tools include Param-Chem[11] and MATCH[12] for the CHARMM force field, and the Automated Topology Builder[13] web server for the GROMOS force field. The SwissParam[14] initiative assigns vdW terms by analogy to existing CHARMM atom types while all other parameters (charges, bonds, angles, dihedrals, impropers) are assigned by analogy from the Merck Molecular Force Field[15,16], and translated into the CHARMM format.

Significantly fewer tools are available for developing parameters directly from electronic structure calculations or from fitting to experimental data or both. One of them is Antechamber,[17] which is used to generate parameters for the AMBER and associated general Amber force field (GAFF) force field.[18] and another one is the Force Field Toolkit (FFTK),[19] which is linked to VMD (Visual Molecular Dynamics), provided limited functionality to derive CHARMM parameters from quantum mechanical (QM) calculations. The release of CGenFF, along with a set of procedures for parametrization made possible the development of a comprehensive tool capable of yielding a complete set of CHARMM-compatible parameters.[11,20] To the contrary, recent software solutions (e.g., CGenFF, MATCH) have focused on parameter assignment based on analogy only, although GAAMP (General Automated Atomic Model Parametrization)[21] does derive charge and dihedral parameters based on QM calculations.

With the advent of more advanced multipolar implementations, the need for robust parametrization tools has even increased. Here, we describe a versatile fitting environment which allows to determine high-quality multipole-based force fields together with suitable Lennard Jones parameters for condensed phase simulations. The environment is based on a graphical user interface (GUI) which handles computations and subsequently analyses data from electronic structure and molecular dynamics codes. In the present case this is output from Gaussian09[22] and input to/output from CHARMM[23]. The reference data consists of electronic structure information and thermodynamic properties from experiment.

Methods and Implementation

A stand-alone, convenient and accurate force field fitting environment involves handling several tasks. First, for the procedure pursued here, the electron density $\rho(\vec{x})$ is determined

from electronic structure calculations for an optimized structure at a given level of theory. Next, local reference axis (LRAs) systems need to be defined for calculating multipolar interactions. Then, atomic multipole coefficients (MTPs) are fitted to best reproduce the electrostatic potential (ESP). Next, atom types and bonded force field terms (bonds, angles, dihedrals) are assigned and LJ parameters for the particular atom types are required for the molecular dynamics (MD) simulations. Finally, MD simulations are run and analyzed from which the necessary thermodynamic observables are determined, compared with experiment and provide information about how to adjust the LJ parameters. These steps together with some formal background are described next.

Electronic structure calculations: All *ab initio* calculations in the present work were carried out with the Gaussian09 suite of codes,[22] using second-order Møller Plesset (MP2)[24] theory and the aug-cc-pVDZ[25–27] basis set. This level of theory is a good compromise between accuracy and speed. These are parameters that are easily changed in the protocol. After optimization of the molecular structure the electron density is extracted with the cubegen utility on a rectangular grid. Grid spacings ranging from 0.1 to 0.4 Å yield almost identical results. The initial atomic multipole moments are obtained from a Distributed Multipole Analysis[28] using the GDMA code. This corresponds to the first three steps in the top panel of Figure 1.

Determine LRAs: Local reference axes are required to define the static multipoles assigned to an atom relative to the global coordinate system. LRAs need to be assigned to each atom of the molecule which are treated with MTPs. The assignment has been described in detail previously.[29] Briefly, the procedure (see fourth step of the top panel of Figure 1) starts from the chemical atom type and determines the number and connectivity of the nearest neighbor atoms. From this information the "full atom type" is generated as a list of the atom type itself and its nearest and second nearest neighbors. From this, the LRA for each atom can

be determined.[30]

Fitting MTPs: To ensure consistency between the CGenFF nonbonded parameters[11,20] (PCs and LJ) and the fitted MTPs, each monopole was constrained to deviate at most by an amount λ_{PC} from the reference value (i.e., provided by CGenFF). Effectively, larger values of λ_{PC} will provide more flexibility—and thus better fits—at the expense of consistency with the reference PCs. Such an approach considers higher order multipoles as corrections to a zeroth-order PC force field.

The ESP can be approximated using MTPs (up to quadrupoles), at any grid point $\mathbf{r}^{(p)}$, from[31–34]

$$
\begin{aligned}
\Phi\left(\mathbf{r}^{(p)}\right) &= \sum_i \sum_j Q_j^{(i)} f_j^{(i)}\left(\mathbf{r}^{(p)}\right) \\
&\approx \sum_i Q_{00}^{(i)} r^{-1} + Q_{10}^{(i)} r^{-2}\hat{r}_z + Q_{11c}^{(i)} r^{-2}\hat{r}_x + Q_{11s}^{(i)} r^{-2}\hat{r}_y \\
&+ Q_{20}^{(i)} r^{-3}(3\hat{r}_z^2 - 1)/2 + Q_{21c}^{(i)} r^{-3}\sqrt{3}\hat{r}_x\hat{r}_z \\
&+ Q_{21s}^{(i)} r^{-3}\sqrt{3}\hat{r}_y\hat{r}_z + Q_{22c}^{(i)} r^{-3}\sqrt{3}(\hat{r}_x^2 - \hat{r}_y^2)/2 \\
&+ Q_{22s}^{(i)} r^{-3}\sqrt{3}\hat{r}_x\hat{r}_y
\end{aligned}
\tag{1}
$$

where i iterates over all atoms and j over all MTP coefficients, $\mathbf{r}$ is the vector from atom i to $\mathbf{r}^{(p)}$, $r = \parallel \mathbf{r} \parallel$ is the norm of $\mathbf{r}$, and $\hat{r}_a = \mathbf{r} \cdot \hat{a}/r$ is normalized using one of the three unit vectors $\mathbf{x}$, $\mathbf{y}$ or $\mathbf{z}$. $Q_{kl}^{(i)}$ is the lth MTP moment of rank k in spherical coordinates, and $f_j^{(i)}\left(\mathbf{r}^{(p)}\right)$ are geometrical factors, including distance- and angular-dependent terms for the MTP moment $Q_j^{(i)}$ at point $\mathbf{r}^{(p)}$.

MTP coefficients $Q_j^{(t)}$ are fitted (last step of the top panel of Figure 1) to the collection of

ESP grid points $\mathbf{r}^{(p)}$ by optimizing the target function

$$\chi^2 = \min \sum_p \left(\Phi_{ab\ initio} \left(\mathbf{r}^{(p)} \right) - \Phi_{MTP} \left(\mathbf{r}^{(p)} \right) \right) \tag{2}$$

which minimizes the error between the *ab initio* and MTP-derived ESPs. Because the problem is linear, we can rewrite the problem as $\mathbf{Xb} = \mathbf{y}$, and because of the sparsity of $\mathbf{X}$ we instead solve

$$\mathbf{X^T X b} = \mathbf{X^T y} \tag{3}$$

where $\mathbf{X^T}$ is the transpose of $\mathbf{X}$.

Assignment of Atom Types for MD Simulations: Next, atom types are required for assigning bonded terms between atoms and Lennard-Jones parameters. This step is automated and the methodology is related to the one used by the SwissParams web-portal.[14] Based on the connectivity of the atoms, a hybridization state (e.g. sp^2, sp^3) is assigned to each of the heavy (i.e. not hydrogen) atoms. Then, based on its hybridization and the hybridization of its neighbor atoms, a CGenFF FF atom type is assigned to each atom, e.g. "CT3" for a sp^3 carbon with 4 explicit substitutes. From this a PDB file compatible with CHARMM can be generated, together with a topology and a structure file.

With chemical atom types assigned, the force field for the compound (including bonds, angles, dihedrals, partial charges and Lennard Jones parameters) is generated according to the CGenFF force field.[20] However, because the electrostatic interactions are modified (i.e. switching from PCs to MTPs), reparametrization of the LJ coefficients is necessary as they were optimized for use with PCs. This is another reason why keeping PCs in the fitting close to the CGenFF values, namely that the CGenFF LJ parameters can be used as a consistent starting point in their refinement. This is part of the first step of the bottom panel of Figure 1.

MD Simulations: Atomistic simulations (bottom panel of Figure 1) are carried out in order to determine the necessary thermodynamic data (see below). The CHARMM-input files are assembled from the `Java` GUI. These are then submitted to a computing pipeline through the GUI, relying on a `Python` scripts engine, in order to allow users to easily customize the procedure.

Fitting the Lennard-Jones Parameters: For refining the LJ parameters, thermodynamic properties are often used as a reference. Here, they include pure liquid density ρ, heat of vaporization ΔH_{vap} and hydration free energies ΔG_{hyd}. Ideally, one would proceed by fitting the LJ radius of each atom type independently. However, this is neither practical (because for each combination of parameters an independent MD simulation is needed) nor desirable, as it would require a high-dimensional parametrization for an undetermined problem (typically considerably more parameters than observables). Furthermore, established LJ-parameters from a validated force field often have already a certain balance which would be compromised if arbitrary scaling would be allowed and retaining this balance may be advantageous. Hence, LJ parameters are rescaled by a parameter ℓ according to $\varepsilon^* = \ell\varepsilon$ and $R_{\text{min}}^*/2 = \ell R_{\text{min}}/2$. It is possible to use a separate scaling for ε and $R_{\text{min}}/2$. However, for a full grid evaluation this considerably increases the number of simulations to perform. This is part of the last step of the bottom panel of Figure 1.

For determining the pure liquid density, multipole-multipole interactions are needed. This requires the definition of local reference axes (see discussion above).[29] Since all coefficients are expressed in the atom's local frame, they are independent of orientation. The geometry of two atoms a and b relative to the orientation of their MTP sites is then determined by incorporating the unit vectors of their local axis systems $\{\mathbf{w}^a\} = \{\mathbf{x}^a, \mathbf{y}^a, \mathbf{z}^a\}$ for atom a and likewise for b. The set of $\{\mathbf{w}^a\}$ and $\{\mathbf{w}^b\}$ combined with the intersite unit vector $\hat{\mathbf{R}}$ defines the direction cosines $\mathbf{q} = \{R, \mathbf{w}^a \cdot \hat{\mathbf{R}}, \mathbf{w}^b \cdot \hat{\mathbf{R}}, \mathbf{w}^a \cdot \mathbf{w}^b\}$ that provide a geometric description
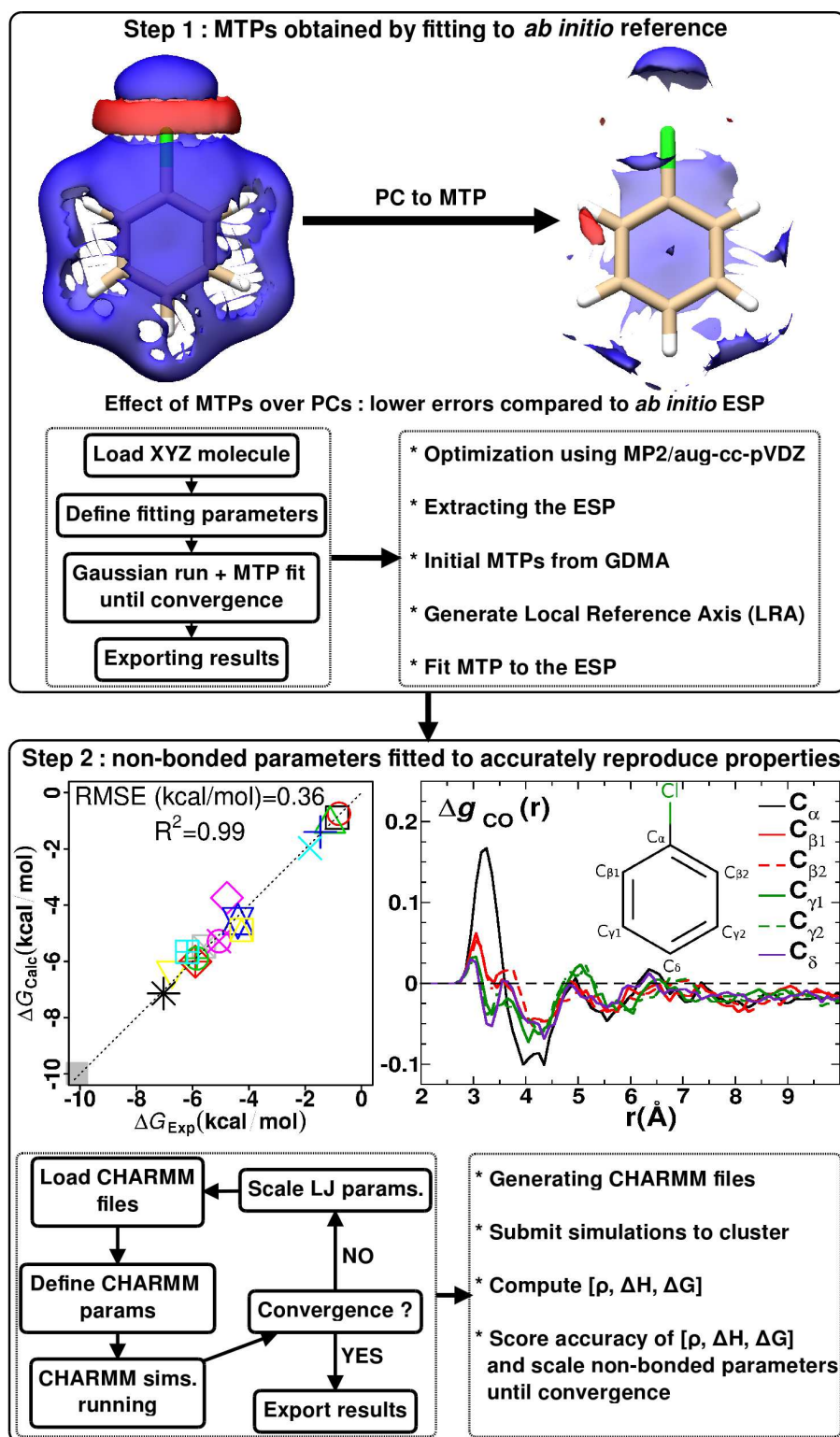
Figure 1: Flowchart illustrating the FW. The top panel corresponds to fitting the MTPs to the ESP as obtained from the electronic structure calculations. The bottom panel summarizes refinement of LJ parameters for optimal reproduction of selected thermodynamic observables. The left chart compares experimental and computed ΔG_{hyd} whereas the right chart shows atom-specific differences in the radial distribution function $\Delta g(r)$ between a PC and a MTP parametrization for Chlorobenzene.

8

of the two MTP sites. From the interaction functions $T_{tu}^{ab}(\mathbf{q})$ for two MTP moments Q_t^a and Q_u^b of order t and u on atomic sites a and b, respectively, the interaction energy is

$$U_{tu}^{ab}(\mathbf{q}) = Q_t^a \cdot Q_u^b \cdot T_{tu}^{ab}(\mathbf{q}) \tag{4}$$

This is the MTP implementation pursued in the MTPL module.[2]

The bottom panel of Figure 1 also reports concrete results from fitting studies. The left hand panel highlights the accuracy of ΔG_{hyd} for 20 compounds studied for which calculated and experimental ΔG_{hyd} agree very favorably. The right hand panel reports differences in the radial distribution functions $\Delta g(r)$ between the C atoms of PhCl and the water-oxygen atoms for PC and MTP parametrization with optimized LJ parameters.

Additional Remarks: While the GUI runs on the local machine, *ab initio* and MD calculations are carried out on a distributed computing environment, and data files are retrieved using the `ssh` transmission protocol. This approach allows to use any computing cluster and no dedicated installation procedure is required on the server side. For the LJ fit (Figure1 (bottom)), all MD simulations for estimating the thermodynamic observables are submitted at once, in order to exploit as much as possible the distributed architecture of the computing cluster. The above mentioned set of scripts currently supports the `qsub` jobs submissions, but extending the workflow for supporting other systems such as `sbatch` should be straightforward.

Computing Thermodynamic Observables

The thermodynamic observables considered here (ρ, ΔH_{vap}, ΔG_{hyd}) require entire MD simulations to be run. For automating this step, a suitable set of core input files for the MD engine used (here CHARMM) is set up. All MD simulations use a time step of $\Delta t = 1$ fs,

solvent simulations are carried out with periodic boundary conditions (PBC) with a non-bonded cutoff of 12 Å and using Particle Mesh Ewald summation[35] for the PCs, with a width of the Gaussian distribution $\kappa = 0.34$, a B-spline interpolation of fifth degree, and 32 grid points along each spatial dimension. The box size is adapted to the probe molecule's size and usually of dimension 20^3 to 25^3 Å^3, corresponding to a total number of ~ 270 to 520 water molecules. For calculating solvation free energies, the TIP3P[36] water model is used, although this is easily modified to other available water models. All simulations are carried out in the NPT ensemble, using the Leap-Frog integrator, and the Hoover algorithm is used for constant pressure and constant temperature simulations. Bonds involving hydrogens were constrained with SHAKE.[37] Further details are given below in the sections which discuss individual observables.

Heat of vaporization

Molecular dynamics simulations provide a convenient way to compute the heat of vaporization[38]

$$\Delta H_{\text{vap}}(T) = E_{\text{gas}}(T) - E_{\text{liq}}(T) + RT, \tag{5}$$

where E_{gas} and E_{liq} are the potential energies of one molecule in the gas and liquid (i.e., NPT) phases, respectively, and R is the gas constant. The gas-phase energy is computed from the minimized energy and the number of atoms, N, and constrained degrees of freedom, N_{cons}, in the molecule, according to

$$E_{\text{gas}}(T) = E_{\text{gas}}^{\text{minimized}} + \frac{1}{2}RT(3N - 6 - N_{\text{cons}}). \tag{6}$$

Thermodynamic integration

Free energies of hydration (i.e., solvation in water) are computed using thermodynamic integration (TI). TI gradually couples/decouples chemical groups from the system by applying

a scaling parameter λ to the nonbonded interactions (i.e., electrostatics and LJ). The total Hamiltonian is written as a function of λ

$$\Delta G_{A \to B} = \int_0^1 d\lambda \left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle_\lambda \approx \sum_i (\lambda_{i+1} - \lambda_i) \left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle_{\lambda_m}, \qquad (7)$$

where $A \to B$ refers to the alchemical transformation between compounds A and B. The canonical average $\langle \cdot \rangle_\lambda$ is performed over the phase space generated by the Hamiltonian $\mathcal{H}(\lambda)$, and $\lambda_m = (\lambda_i + \lambda_{i+1})/2$. For the LJ and PC derivatives CHARMM's PERT module using soft-core potentials for the LJ interactions[39,40] is used. No long-range corrections were applied to the LJ-interactions, as no noticeable change was found when increasing the nonbonded cutoff beyond $r_c = 12$ Å.

The LJ and electrostatic interactions are turned on separately. First, the LJ interactions with soft-core potentials are fully grown, followed by the electrostatics in the presence of the full van der Waals interactions, thereby avoiding the need for soft-core electrostatic potentials. The change in free energy due to MTP electrostatics with coupling λ_m was computed by first performing a simulation where all MTP energies (see Eq. 4), forces, and torques were linearly scaled by λ_m. In a post-processing step the energies with the original Hamiltonian (unscaled, $\lambda = 1$) are extracted and averaged over the solute-solute and solute-solvent energies (i.e., solvent-solvent interactions are not affected by λ_m) in such a way that its derivative with respect to λ yields the original energy (unscaled $\lambda = 1$).

Using a thermodynamic cycle, the hydration free energy is computed according to $\Delta G_{\mathrm{hyd}} = \Delta G_{\mathrm{sol}} - \Delta G_{\mathrm{vac}}$, where ΔG_{sol} and ΔG_{vac} correspond to the free energy of insertion of the compound in a box of water and vacuum, respectively. For the simulations in water the solute was placed in a box of ≈ 500 solvent molecules.

The grid of $\lambda-$points is chosen in different ways. The most accurate, automatic procedure starts from 20 evenly-spaced λ windows between 0 and 1. For further refinement, windows at the two ends of the $\lambda-$interval (typically $\lambda \in [0, 0.1]$ and $\lambda \in [0.9, 1]$) are further partitioned to retain accuracy. However, introduction of additional partitions is inconvenient in the present context as it requires to run an a priori unknown number of simulations in a sequential manner. Hence, further different strategies were explored for this step. The best performance was found for grid spacing $\Delta\lambda = 0.025$ for $\lambda \in [0.0, 0.1]$; $\Delta\lambda = 0.100$ for $\lambda \in [0.1, 0.9]$ and $\Delta\lambda = 0.025$ for $\lambda \in [0.9, 1.0]$. Such a procedure allows to submit all $\lambda-$windows at once which considerably speeds up turnover times for individual fitting cycles. However, for accuracy checks the interface also allows automatic subdivision of the windows for particularly relevant parametrization problems.

Database of Compounds

Fitting force fields for condensed-phase simulations requires reference data for adjusting the parameters, as described above. In the present approach, the atomic multipoles are fit to best reproduce the electrostatic potential from electronic structure calculations whereas adjustment of van der Waals parameters requires solution-phase data. For this, a database containing experimental values from the literature has been built. The current version of the database includes mass, density, enthalpy of vaporization and the hydration free energy (where available) as reference data. Mass, density, and ΔH_{vap} are those from PubChem[41], and solvation free energies were taken from the FreeSolv[42,43] database built by Mobley et al., which contains values collected from the literature.

The database is searchable by name, chemical formula or `SMILES`[44,45] and uses the `SQL` language.[46] It was decided to provide, as an embedded feature within the Fitting Wizard, the access to a database of chemical compounds. The database was built according to the follow-

ing procedure: (i) the version v0.31 of the database was downloaded, containing values for ΔG_{hyd} for 643 compounds, together with their PubChem ID, `SMILES` notation, IUPAC name and a DOI literature reference. (ii) a `MySQL` database was created using the database content. (iii) the PubChem ID was used for automatically retrieving (using the provided Application Programming Interface (API)[47]) the previously mentioned properties (m, ρ, ΔH_{vap}, ΔG_{hyd}). However, missing values or inconsistencies may remain for some of the compounds even after data curation: thus the database is editable, and then provided as a starting set the user can use and improve. See the SI section I and Figure S1 for further details concerning the database.

Validation and Results

For validating the Fitting Wizard several problems are considered. First, the parametrization of N-methylacetamide (NMA) is reconsidered and extended as it serves as a model for peptides and proteins. Second, the parametrization of substituted benzenes is presented as a case where - particularly for the case of halogen-substituted benzenes - MTPs have been found to be essential for an accurate description of solvent properties.[2,48]

N-Methyl-Acetamide

As the central building block for peptides and proteins, NMA is a meaningful test system. Experimental data is available for all three observables considered. Starting from an optimized MTP model, the LJ parameters are adjusted to best reproduce the experimentally measured ρ, ΔH_{vap}, and ΔG_{hyd}. The influence of scaling the LJ parameters is summarized in Table 1 where results for $\ell \in [0.9; 1.1]$ are presented, meaning that LJ parameters were changed by up to 10 % around their reference CGenFF-values. In order to determine the

best-performing model, a simple weighted score $S = \sum_{i=1}^{3} w_i(\mathrm{Obs_i} - \mathrm{Calc_i})^2$ with $w_\rho = 1$, $w_{\Delta H} = 3$ and $w_{\Delta G} = 5$ is introduced to differently weight the three observables. Such a weighting puts more emphasis on hydration free energies but alternative choices are possible for particular purposes and applications. The model with $\ell = 0.95$ yields the lowest score ($S = 0.1$) and is therefore the preferred one. Both, ΔH_{vap} and ΔG_{hyd} are reproduced to within less than 1 % compared to the reference data whereas the density differs by 6 %. Obvious extensions involve separate scaling factors for σ and ϵ which, however, further increases computational demands. Nevertheless, other models yield competitive scores well below $S = 1.0$. It should be noted that the experimental ΔH_{vap} used in force field parametrizations has been studied recently and it was found that $\Delta H_{\mathrm{vap}} = 13.0 \pm 0.1$ kcal/mol at 410 K is the preferred value.[49] As ΔH_{vap} increases with decreasing temperature, the value used in the present work ($\Delta H_{\mathrm{vap}} = 14.2$ at 300 K) should be qualitatively correct. However, refinement of this based on the detailed study in Ref.[49] may be desirable.

Table 1: Dependence of ρ (g/cm^3), ΔH_{vap} and ΔG_{hyd} (both in kcal/mol) when scaling the Lennard-Jones parameters. In bold face is shown the value of ℓ minimizing the score S.

Scaling ℓ	ρ	ΔH_{vap}	ΔG_{hyd}	Score S
0.9	1.13	14.24	-9.82	0.4
0.925	1.08	13.95	-9.89	0.4
0.95	**1**	**14.11**	**-9.99**	**0.1**
0.975	0.99	13.84	-10.22	0.5
1	0.95	13.82	-9.88	0.6
1.025	0.92	13.68	-9.06	6.0
1.05	0.88	13.57	-8.75	10.0
1.075	0.84	13.29	-8.38	16.9
1.1	0.81	13.47	-8.07	21.8
Expt.	0.94[41,50]	14.2[41,51]	-10.08[52]	

Performance of a Pre-defined $\lambda-$Grid: As mentioned in the Methods section (see "Thermodynamic Integration"), automated refinement of the $\lambda-$grid on either side of the interval $\lambda \in [0,1]$ is computationally inconvenient as each subdivision can only be made once the updated hydration free energy is available. Ideally, one would work with a pre-defined grid of

$\lambda-$values which allows to submit all necessary simulations at the same time. This improves turnover times, and the total time for an entire optimization (a few hours for a molecule such as NMA) can therefore be estimated *a priori*. The choice of this subdivision is flexibly handled in the fitting wizard. Here, it is merely illustrated that such a pre-defined grid can yield good-quality parametrizations but the subdivision is likely to depend on the particular molecule or class of molecules considered. Three possibilities (I) to (III) were explored in the following.

(I): [$\lambda \in [0, 0.1]$ with $\Delta\lambda = 0.010$; $\lambda \in [0.1, 0.9]$ with $\Delta\lambda = 0.100$; $\lambda \in [0.9, 1.0]$ with $\Delta\lambda = 0.025$].

(II): [$\lambda \in [0, 0.1]$ with $\Delta\lambda = 0.020$; $\lambda \in [0.1, 0.9]$ with $\Delta\lambda = 0.100$; $\lambda \in [0.9, 1.0]$ with $\Delta\lambda = 0.020$].

(III): [$\lambda \in [0, 0.1]$ with $\Delta\lambda = 0.025$; $\lambda \in [0.1, 0.9]$ with $\Delta\lambda = 0.100$; $\lambda \in [0.9, 1.0]$ with $\Delta\lambda = 0.025$].

The results for MTP/LJ optimizations for trans-NMA with different subdivisions of the $\lambda-$windows are summarized in Table 2. It is found that the hydration free energy changes by about 5 % depending on the subdivision of the $\lambda-$interval. On the other hand it is possible to find a subdivision (here case (I)) which provides an accurate estimate and is computationally efficient.

ΔG_{hyd} *for cis- and trans-NMA from a polarizable Drude Model:* In previous and also more recent computational studies it was observed that calculated ΔG_{hyd} values differ for the cis- and trans-isomers for NMA.[55,56] Experimentally, the direct determination of $\Delta G_{\mathrm{hyd}}^{\mathrm{cis}}$ is diffi-cult due to the low population of this isomer (< 2 %) in solution[53] although it is generally

Table 2: Comparison of ΔG_{hyd} for different λ subdivision (sets (I) to (III)), or heuristically decided by CHARMM's PERT module (based on the fluctuation of the average). "Automated" is the default mode. Predefined $\lambda-$windows speed up the process through a pre-determined number of simulations.

Method	ΔG_{hyd}
Expt.[52–54]	−10.08
Simulation[6]	−9.90
Automated division	−9.99
Case (I)	−9.61
Case (II)	−9.38
Case (III)	−9.29

believed that the differential hydration free energy $\Delta\Delta G_{\mathrm{hyd}} = \Delta G_{\mathrm{hyd}}^{\mathrm{trans}} - \Delta G_{\mathrm{hyd}}^{\mathrm{cis}} \approx 0.$[57,58] Hence, experimental values for $\Delta G_{\mathrm{hyd}}^{\mathrm{cis}}$ are indirect and also have been put into question for different reasons.[59] As a comparison with a recently published generic and polarizable force field, hydration free energies were also determined for *cis-* and *trans*-NMA using the Drude force field.[6] As recommended, simulations are carried out with the SWM4-DP[5] water model instead of TIP3P, the (default) automatic $\lambda-$division procedure is used for TI, and all other simulation parameters are identical to those in the MTP-simulations. For the two isomers $\Delta G_{\mathrm{hyd}}^{\mathrm{cis}} = -8.67$ kcal/mol and $\Delta G_{\mathrm{hyd}}^{\mathrm{trans}} = -9.81$ kcal/mol were found. The value for $\Delta G_{\mathrm{hyd}}^{\mathrm{trans}}$ agrees to within 0.09 kcal/mol with the reference value[6] which validates the present protocol. Despite using a polarizable model, $\Delta\Delta G_{\mathrm{hyd}} = 1.1$ kcal/mol between the two isomers, which differs from the assumed value of close to zero from experiment. Compared to this, the present non-polarizable simulations yield $\Delta\Delta G_{\mathrm{hyd}} = 1.8$ kcal/mol. As the Drude-simulations do not employ multipoles and the present MTP-simulations are non-polarizable it is possible that combining the two will yield satisfactory agreement with experiment. As another comparison, a recent parametrization study based on electron density partitioning found $\Delta\Delta G_{\mathrm{hyd}} = -1.0$ kcal/mol with the *cis*-isomer more stable than *trans*-NMA.[56]

Halogenated and Substituted Benzenes

Next, a validation study was performed for halogenated and substituted benzenes. They constitute important building blocks in medicinal chemistry and pharmaceutically active substances.[60–63] Also, halogenated amino acid side chains have recently found to be useful modifications in protein biochemistry, such as in insulins.[64] Besides the accuracy of such a parametrization it is also of interest to test the transferability of the final parameters. This is important in situations when the chemical environment of a group changes and the accuracy of the original parametrization should be retained.

Halogenated Phenols: As a first example, 4-BrPhOH is considered. Table 3 reports the calculated hydration free energy depending on the parametrization and level of optimization used for 4-BrPhOH. The calculated ΔG_{hyd} with PCs and LJ parameters transferred from CGenFF (i.e. unoptimized PC and LJ) overestimates the solvation energy by 4.22 kcal/mol. A considerable improvement of the calculated ΔG_{hyd} is obtained by including MTP electrostatics whereby ΔG_{hyd} drops from –10.07 kcal/mol (CGenFF parameters) to –6.37 kcal/mol (optimized MTP and CGenFF LJ parameters) that differs from the experimental value by only 0.52 kcal/mol. This can be explained by the fact that a simple unoptimized PC electrostatic model cannot describe the large electronic reorganization around the Bromobenzene ring when -OH is introduced in the *para* position. Moreover, when using previously optimized LJ parameters for Bromobenzene (PhBr) (Bereau et al.[2]) instead of standard-CGenFF parameters, the error in ΔG_{hyd} further reduces to 0.29 kcal/mol. Finally, with a slight optimization of the "-OH" group LJ parameters (scaling of σ and ϵ, see above) the calculated ΔG_{hyd} reproduces the experimental value with a difference of 0.04 kcal/mol, which falls within the statistical error typically found on computed values (around 0.05 kcal/mol).

Table 3: Hydration free energies calculated for 4-BrPhOH depending on the Electrostatic and LJ parameter treatment used. $|\Delta\Delta G_{\text{hyd}}|$ represent the absolute error relative to the experimental value (-5.85 kcal/mol)[65].
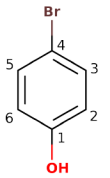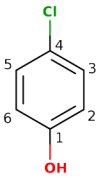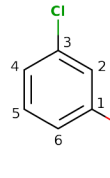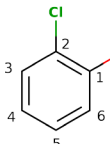
Treatment	ΔG_{hyd}	$|\Delta\Delta G_{\text{hyd}}|$
CGenFF parameters (unoptimized PC and LJ)	-10.07	4.22
Optimized MTP / CGenFF LJ	-6.37	0.52
Optimized MTP / LJ transferred from Bereau et *al.*[2]	-5.56	0.29
Optimized MTP / LJ transferred from Bereau et *al.*[2] for C, H, Br and optimization of (ϵ, σ) for -OH	-5.89	0.04

Transferability: One essential aspect in modern force field development and practical applications is the transferability of parametrizations for a chemical building block (e.g. an amino acid side chain) between two different chemical environments which can considerably speed up parametrization tasks and is also conceptually appealing. To assess transferability within the given fitting methodology the hydration free energy of different parametrizations was computed for Br- and ClPhOH. Here, the differential solvation free energy $\Delta\Delta G_{\text{hyd}}$ in transferring LJ parameters (ϵ and σ) for common atom types (aromatic C, H, Cl, and Br) obtained from previous parametrizations[2] of PhBr and PhCl to 4-BrPhOH and 4-ClPhOH, and from the current parametrization of 4-BrPhOH's polar "-OH" group (see Table 3) to 2,3,4-ClPhOH, is considered. The effect of reoptimizing the LJ parameters on the "-OH" group in positions (2-, 3- and 4-) is also evaluated. For all molecules considered (2-, 3-, 4-ClPhOH and 4-BrPhOH), MTP electrostatics was first fitted individually and not transferred since the impact of the "-OH" group insertion on the electron distribution varies depending on its position (2-, 3- or 4-) and the type of the halogen present (Cl or Br).

Table 4 reports calculated hydration free energies for 2-, 3-, 4-ClPhOH and 4-BrPhOH with LJ parameters transferred for aromatic C, H, Cl and Br from previous parametrizations of PhBr and PhCl[2], and with LJ parameters for the "-OH" group a) taken from CGenFF or b) optimized for 4-BrPhOH and transferred to 2,3,4-ClPhOH and compares them to experimentally determined values[65,66]. For the transferred parametrizations, the difference be-

tween computed and experimentally determined ΔG_{hyd} is 0.3 kcal/mol and 0.8 kcal/mol for 4-BrPhOH and 2-ClPhOH, respectively. For 3- and 4-ClPhOH they are below 0.4 kcal/mol. Hence, results with transferred parameters are well within 1 kcal/mol which points towards a good degree of transferability. Furthermore, improved hydration free energies after reoptimizing the -OH LJ parameters (ϵ and σ) are also reported for 4-BrPhOH and 2-ClPhOH in Table 4. They decrease to 0.04 kcal/mol and 0.34 kcal/mol for 4-BrPhOH and 2-ClPhOH, respectively.

Table 4: Hydration free energies calculated depending on the radical's position (here -OH) relative to the halogen and the LJ parameters used. With $|\Delta\Delta G_{\text{hyd}}| = |\Delta G_{\text{hyd}}^{\text{Exp}} - \Delta G_{\text{hyd}}^{\text{Calc}}|$. All MTPs optimized individually.

Molecule		$\Delta G_{\text{hyd}}^{\text{Exp 65,66}}$	$\Delta G_{\text{hyd}}^{\text{Calc}}$	$|\Delta\Delta G_{\text{hyd}}|$	Transferred	Optimized
	4-BrPhOH	−5.85	−5.56	0.29	C, H, Br Bereau et *al.* -OH CGenFF	
			−5.89	0.04	C, H, Br Bereau et *al.*	-OH (ϵ, σ)
	4-ClPhOH	−7.03	−6.65	0.38	C, H, Cl Bereau et *al.* -OH LJ optimized for 4-BrPhOH	
	3-ClPhOH	−6.62	−6.34	0.28	C, H, Cl Bereau et *al.* -OH LJ optimized for 4-BrPhOH	
	2-ClPhOH	−4.55	−5.32	0.77	C, H, Cl Bereau et *al.* -OH LJ optimized for 4-BrPhOH	
			−4.89	0.34	C, H, Cl Bereau et *al.*	-Cl (ϵ) -OH (ϵ, σ)

As an illustration of the effect of different LJ parameters, the water structure around 2-ClPhOH is considered. For this, the radial distribution function $g(r)$ of water around

the solute is determined. As an example for a recent application, it has been shown for fluoro-acetonitrile solvated in water that the combination of optical spectroscopy and atomistic simulations is able to detect incipient halogen bond formation.[67] Figure 2 reports the radial distribution function between 2-ClPhOH(O) and water(O) from 2 ns of NPT simulations using two different sets of -OH LJ parameters. The first set ($\Delta G_{\mathrm{hyd}} = -5.32$ kcal/mol) uses the LJ parameters transferred from 4-BrPhOH to which they were optimized (Table 4 2nd row). The second set ($\Delta G_{\mathrm{hyd}} = -4.89$ kcal/mol; Table 4 last row) uses the -OH LJ parameters (σ and ϵ) optimized specifically for 2-ClPhOH with respect to ρ, ΔH_{vap} and ΔG_{hyd}, starting from the parameters of the first set. The first set was optimized for an "-OH" group in position 4- (opposite to the halogen atom) whereas the second set was optimized for an "-OH" group in position 2- (adjacent to the halogen). The O-O_W pair distribution function obtained with both LJ parameter sets (Figure 2, black line and red line) peaks at ~ 2.8 Å. However, the amplitude of the peak is smaller for the second set (Figure 2, red line) and the first minimum is less pronounced. The reduced amplitude of $g(r)$ also decreases the occupation number $N(r_s) \propto \int_0^{r_s} g(r)r^2 dr$ of water molecules within a distance r_s around the "-OH" group, which also reduces the hydration free energy by 0.4 kcal/mol.
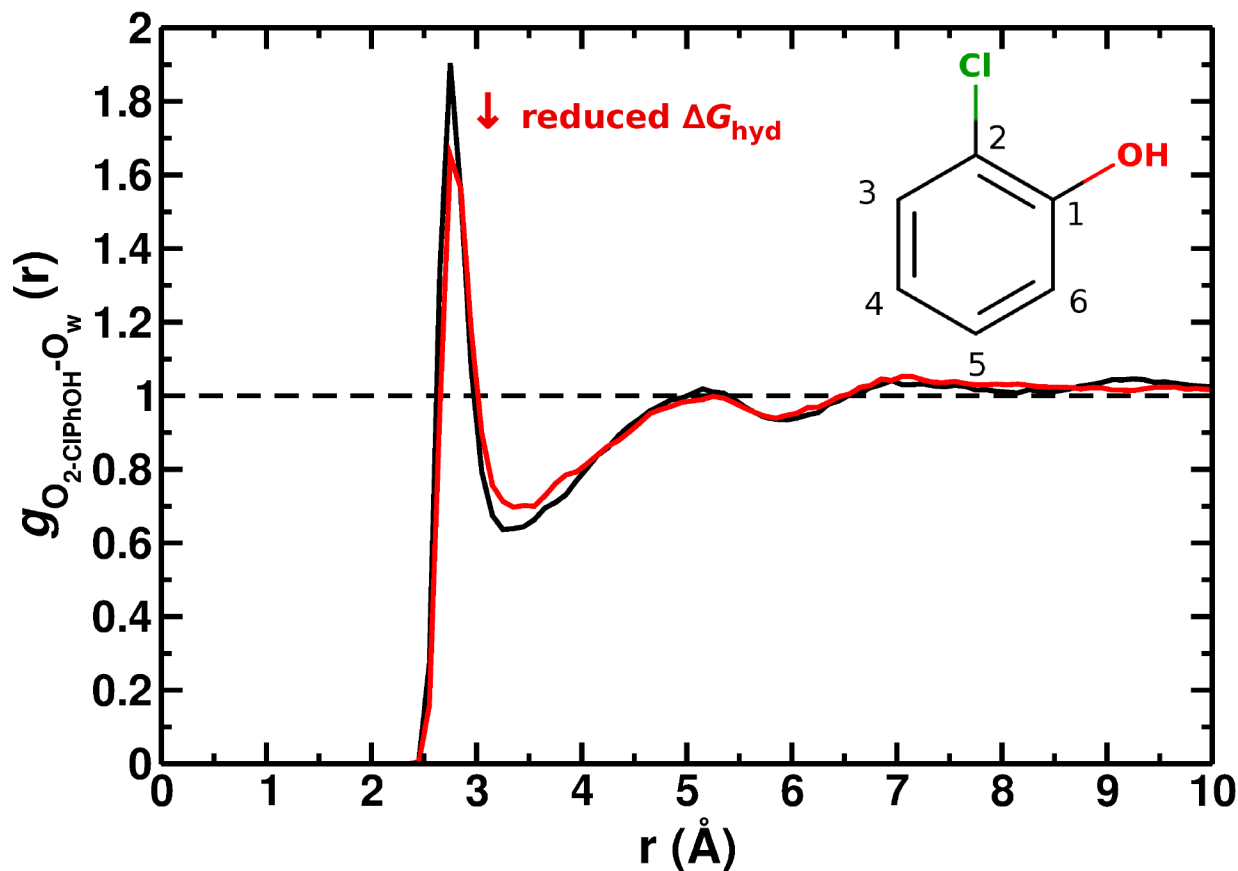
Figure 2: Radial distribution function $g(r)$ for (2-ClPhOH)O–O(water). The black and red traces represent the distribution functions before (first set) and after (second set) optimizing the -OH LJ parameters, respectively. The inset represents the chemical structure of 2-ClPhOH.

Degrees of Parameter Optimization

To further illustrate the effect of LJ reparametrization, Table 5 reports the three thermodynamic observables (ρ, ΔH_{vap}, ΔG_{hyd}) for different optimization levels for NMA and 4-ClPhOH. While ρ varies little throughout the range studied, ΔH_{vap} and ΔG_{hyd} strongly change.

The results in Table 5 establish that depending on the system studied (NMA or 4-ClPhOH), an optimized PC/LJ model ($S = 0.3$) can perform very well compared to an optimized MTP/LJ parametrization ($S = 0.1$). This is the case for NMA. Contrary to that, the halo-

Table 5: The computed (ρ, ΔH_{vap}, ΔG_{hyd})-values for force fields of different optimization levels for NMA and 4-ClPhOH. The score is used to differentiate between different levels of optimization. Units are g· cm^{-3} for ρ and kcal/mol for ΔH_{vap} and ΔG_{hyd}.

		ρ	ΔH_{vap}	ΔG_{hyd}	Score S
NMA					
	CGenFF (PC and LJ)	0.98	15.09	−11.03	6.9
	opt PC / CGenFF LJ	0.99	14.49	−10.11	0.3
	opt PC / opt LJ	0.99	14.49	−10.11	0.3
	opt MTP / CGenFF LJ	0.95	13.82	−9.88	0.6
	opt MTP / opt LJ	0.95	14.11	−9.99	0.1
	Exp.	0.94[41,50]	14.20[41,51]	−10.08[52]	
4-ClPhOH				-	
	CGenFF (PC and LJ)	1.28	15.74	−5.47	72.9
	opt PC / CGenFF LJ	1.27	10.78	−5.44	13.2
	opt PC / opt LJ	1.25	11.85	−5.61	11.2
	opt MTP / CGenFF LJ	1.27	12.86	−5.91	14.2
	opt MTP / opt LJ	1.25	11.46	−7.14	0.2
	Exp.	1.22[41,68]	11.24[41,69]	−7.03[66]	

genated system 4-ClPhOH evidently requires optimized MTP electrostatics and optimized

LJ parameters. It is also important to note that LJ parameters can be transferred from a

previous optimization of a similar compound as in Table 4 (second row), where the latter

were transferred from previous optimizations of PhCl and 4-BrPhOH to 4-ClPhOH and yield

a ΔG_{hyd} of –6.65 kcal/mol that only differs by 0.38 kcal/mol from the experimental value.

The score for the plain CGenFF parametrization reduces by a factor of 6 upon optimization

of the PC model but essentially remains unchanged in the next few refinements. Only when

both, MTP and LJ parameters, are optimized the score improves by almost two orders of

magnitude and excellent agreement with experiment is obtained. This highlights that not

all chemical building blocks may need the same level of parameter optimization and for some

systems good and computationally inexpensive PC-based parametrizations can be obtained.

A Broader Parametrization Study

Additional halogenated and substituted benzenes were parametrized along the same protocol and all results for ρ, ΔH_{vap} and ΔG_{hyd} are summarized and discussed in the following. Table 6 and Figure 3 compares the free energy of hydration ($\Delta G_{\text{hyd}}^{\text{Calc}}$) as calculated using the FW and compares them to experimental data ($\Delta G_{\text{hyd}}^{\text{Exp}}$).

The agreement between computed and observed ΔG_{hyd} is excellent. Over a range of 8 kcal/mol the RMSE is 0.36 kcal/mol and $R^2 = 0.99$, see Figure 3. As a comparison, in a study of the solvation free energies of amino acid side chains the RMSE for ΔG_{hyd} using TIP3P water and the OPLS-AA force field was 0.79 kcal/mol with an $R^2 = 0.93$ which changed to 0.51 kcal/mol and $R^2 = 0.94$ upon modification of the LJ parameters of the TIP3P water model.[70] In a more recent study focusing on 40 small organic molecules and charges from atoms-in-molecules electron density partitioning, using environment-specific charges and LJ parameters from quantum chemical calculations, the mean unsigned errors relative to experiment are 0.014 g/cm^3 for the density ρ, 0.65 kcal/mol for the heat of vaporization ΔH_{vap} and 1.03 kcal/mol over a range of 12 kcal/mol for ΔG_{hyd}.[56] In yet another, broader study of 239 molecules, the mean unsigned error for ΔG_{hyd} was 1.93 kcal/mol (CHARMM), 1.17 kcal/mol (GAFF) and 0.73 kcal/mol (OPLS2.1).[71]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 6: ΔH_{vap}, ΔG_{hyd} (kcal/mol) and ρ (g/cm^3) as calculated using the FW (Calc.) with optimized MPT and LJ parameters compared to experimental references (Exp.). The absolute deviation is also reported (|Dev.|). Experimental values of ρ and ΔH_{vap} were taken from Pubchem[41], and values of ΔG_{hyd} from the FreeSolv database[42,43]. See Table S1 from the SI for an extended version.

	ρ			ΔH_{vap}			ΔG_{hyd}			Score S
	Exp.[41]	Calc.	|Dev.|	Exp.[41]	Calc.	|Dev.|	Exp.[42,43]	Calc.	|Dev.|	
Benzene	0.88	0.9	0.02	7.89	7.88	0.01	−0.86	−0.89	0.03	0.01
Fluorobenzene	1.02	1.05	0.03	8.26	8.6	0.34	−0.80	−0.75	0.05	0.36
Chlorobenzene	1.11	1.14	0.03	9.97	10.13	0.16	−1.12	−1.11	0.01	0.08
Bromobenzene	1.5	1.47	0.03	10.65	11.98	1.33	−1.46	−1.40	0.06	5.33
Iodobenzene	1.83	1.84	0.01	11.85	12.43	0.58	−1.83	−1.97	0.14	1.11
1H-Pyrrole	0.97	0.99	0.02	10.78	10.87	0.09	−4.78	−3.74	1.04	5.43
6-chloropyridin-3-ol	1.39	1.36	0.03	14.81	15.36	0.55	−6.73	−6.32	0.41	1.75
6-chloropyridin-3-amine	1.33	1.29	0.04	12.71	12.44	0.27	−5.60	−5.47	0.13	0.30
4-chlorophenol	1.22	1.25	0.03	11.24	10.46	0.78	−7.03	−7.14	0.11	1.89
4-chloroaniline	1.17	1.19	0.02	11.2	10.51	0.69	−5.90	−6.01	0.11	1.49
4-bromophenol	1.84	1.83	0.01	14.04	14.1	0.06	−5.85	−5.89	0.04	0.02
2-chloropyridine	1.2	1.21	0.01	10.18	9.93	0.25	−4.39	−4.57	0.18	0.35
4-fluorophenol	1.31	1.32	0.01	10.43	10.77	0.34	−6.19	−5.66	0.53	1.75
4-fluoroaniline	1.17	1.18	0.01	10.16	9.43	0.73	−5.06	−5.28	0.22	1.84
4-fluoro-N-methylaniline	1.04	1.08	0.04	9.98	10.06	0.08	−4.26	−4.88	0.62	1.94
N-methylacetamide	0.94	1	0.06	14.2	14.11	0.09	−10.08	−9.99	0.09	0.07
Average Deviation			0.03			0.40			0.24	

For ΔH_{vap} the RMSE (estimated for the same family of compounds than for ΔG_{hyd}) is 0.53 kcal/mol with an $R^2 = 0.97$, see Figure 4. This compares with 0.65 kcal/mol from a recent study on a different set of small molecules.[56] For the pure liquid density (see Figure S2 from the SI) the current study yields an RMSE of 0.02 g/cm^3 with an $R^2 = 0.99$, compared with an RMSE of 0.01 g/cm^3 of the same recent parametrization work.[56] Hence, for a range of compounds the fitting environment presented here yields comparable if not superior performance based on a user-friendly interface.

For an extended version of Table 6, including also compounds for which one or more of the experimental references are missing, see the SI Section II Table S1.
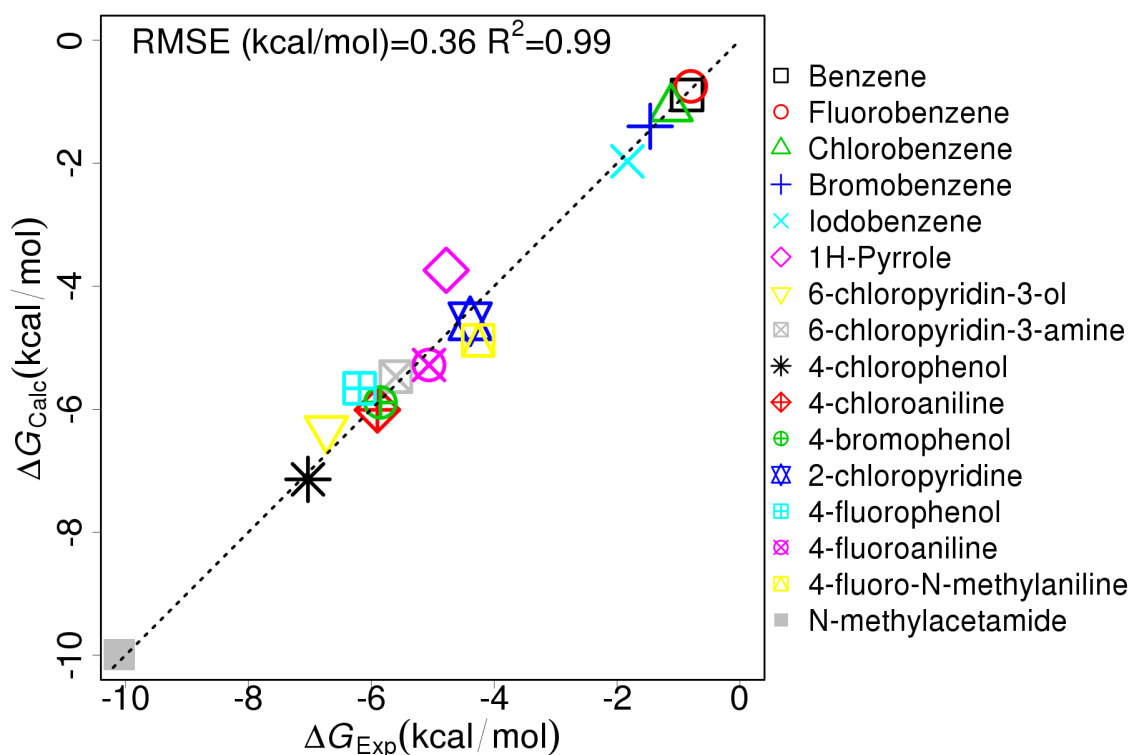
Figure 3: Correlation between experimental and computed solvation free energies ΔG_{hyd} (kcal/mol, respectively, x−axis and y−axis) for a range of compounds of interest. Computed values obtained after optimization of the LJ parameters.
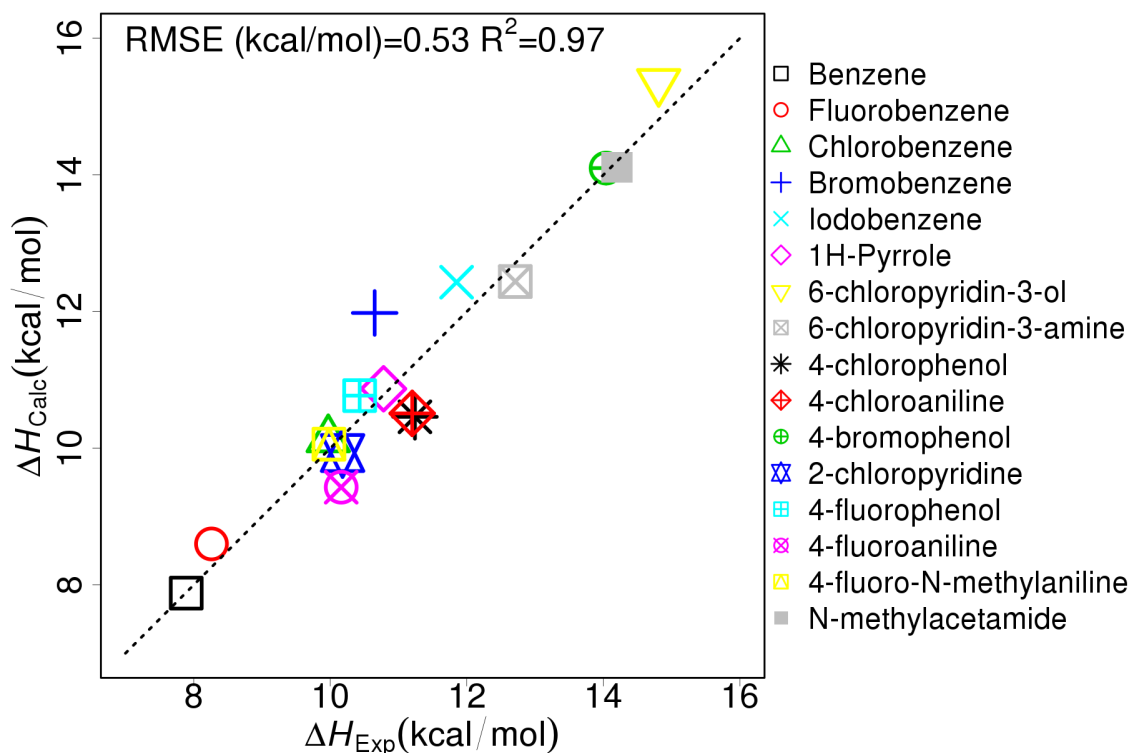
Figure 4: Correlation between experimental and computed enthalpy of vaporization ΔH_{vap} (kcal/mol, respectively, $x-$axis and $y-$axis) for a range of compounds of interest. Both, MTP and LJ parameters were optimized.

Outlook and Perspectives

The present work introduces a graphics-based, versatile, and extensible fitting environment for PC- and MTP-based force fields for condensed-phase simulations. It is demonstrated that accurate parametrizations for solution-phase simulations can be obtained and that medium-scale (several 10 molecules) parametrization tasks can be routinely undertaken as a typical parametrization takes a few hours for a molecule the size of NMA. Within the chemical space covered, the transferability of parametrizations yields results well within chemical accuracy.

The fitting environment can be easily adapted to different and higher levels of theory for the reference data from electronic structure calculations. Also, extension to other molecular dynamics codes (AMBER, GROMACS, TINKER) is possible because of the modular architecture of the software provided that multipolar interactions can be computed. For

molecules exhibiting two or more linked ring systems (e.g. biphenyl) it will be important to consider refitting dihedral parameters because of multipole-multipole interactions between atoms on different ring systems.

Currently, thermodynamic properties ($\rho, \Delta H_{\mathrm{vap}}, \Delta G_{\mathrm{hyd}}$) are used to improve the force field. This can be easily extended to additional interesting (and experimentally accessible) quantities such as diffusion coefficients D, or heat capacities C_p. Also, infrared and NMR spectroscopic data may be of interest in the future.[72–75]

A valuable extension will be the computation of derivatives $\frac{d}{dp}\langle A \rangle_p$ of observables A with respect to the LJ-parameters p from suitable ensemble averages. This has recently been done for the parametrization of the iAMOEBA force field for water.[76] It will be of interest to assess whether a grid-based search as proposed here or a gradient-based approach to improve parameter values converges more rapidly in concrete applications.

Furthermore, it was found in recent work that averaging over a number of conformations can yield meaningful parametrizations of conformationally dependent multipoles.[77] Including such effects should further improve transferability of the parametrizations. A final asset is the storage and retrieval of particular parametrizations for validated simulation and parametrization conditions, in particular for chemically and pharmaceutically important molecular fragments. If the transferability of the parametrizations can be ascertained, this will allow simple assembly of larger molecules from well-parametrized building blocks (molecules) and considerably speed up future parametrizations.

In summary, the present work describes a user-friendly, graphics-based interface for the parametrization of multipolar force fields for quantitative atomistic simulations of small molecular building blocks.

Author Information

Funding

The authors declare no competing financial interest.

Software availability

This software is available free of charge, and under the 3-clause BSD license, from the following github repository:

https://github.com/MMunibas/FittingWizard

Archived releases can be downloaded from

https://github.com/MMunibas/FittingWizard/releases

and documentation and instructions are available from

https://github.com/MMunibas/FittingWizard/wiki

Supporting Information Available

The Supporting Information is available free of charge on the ACS Publications website at

DOI: 10.1021/acs.jcim.xxxxxxx

- Details concerning the database design, construction and curation.

- Supplementary correlation plot for density ρ (similar to Figs 3 and 4)

- Extended version of Table 6 including compounds for which experimental data is incomplete.

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. *Adv. Prot. Chem.* **2003**, *66*, 27–85.

(2) Bereau, T.; Kramer, C.; Meuwly, M. Leveraging Symmetries of Static Atomic Multipole Electrostatics in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 5450–5459.

(3) Simmonett, A. C.; Pickard, F. C., IV; Schaefer, H. F., III; Brooks, B. R. An Efficient Algorithm for Multipole Energies and Derivatives Based on Spherical Harmonics and Extensions to Particle Mesh Ewald. *J. Chem. Phys.* **2014**, *140*, 184101.

(4) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio Jr, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B.* **2010**, *114*, 2549–2564.

(5) Lamoureux, G.; MacKerell, A. D., Jr.; Roux, B. A Simple Polarizable Model of Water Based on Classical Drude Oscillators. *J. Chem. Phys.* **2003**, *119*, 5185–5197.

(6) Lin, B.; Lopes, P. E. M.; Roux, B.; MacKerell, A. D., Jr. Kirkwood-Buff Analysis of Aqueous N-methylacetamide and Acetamide Solutions Modeled by the CHARMM Additive and Drude Polarizable Force Fields. *J. Chem. Phys.* **2013**, *139*, 084509.

(7) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(8) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the

OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(9) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prod-hom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B.* **1998**, *102*, 3586–3616.

(10) MacKerell, A. D. Empirical Force Fields for Biological Macromolecules: Overview and Issues. *J. Comput. Chem.* **2004**, *25*, 1584–1604.

(11) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D., Jr. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Comput. Inf. Model.* **2012**, *52*, 3155–3168.

(12) Yesselman, J. D.; Price, D. J.; Knight, J. L.; Brooks, C. L., III MATCH: An Atom-Typing Toolset for Molecular Mechanics Force Fields. *J. Comput. Chem.* **2012**, *33*, 189–202.

(13) Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *J. Chem. Theory Comput.* **2011**, *7*, 4026–4037.

(14) Zoete, V.; Cuendet, M. A.; Grosdidier, A.; Michielin, O. SwissParam: A Fast Force Field Generation Tool for Small Organic Molecules. *J. Comput. Chem.* **2011**, *32*, 2359–2368.

(15) Halgren, T. A. Merck molecular force field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.

(16) Halgren, T. A. Merck molecular force field. II. MMFF94 van der Waals and Electrostatic Parameters for Intermolecular Interactions. *J. Comput. Chem.* **1996**, *17*, 520–552.

(17) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

(18) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(19) Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. Rapid Parameterization of Small Molecules Using the Force Field Toolkit. *J. Comput. Chem.* **2013**, *34*, 2757–2770.

(20) Vanommeslaeghe, K.; MacKerell, A. D., Jr. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Comput. Inf. Model.* **2012**, *52*, 3144–3154.

(21) Huang, L.; Roux, B. Automated Force Field Parameterization for Nonpolarizable and Polarizable Atomic Models Based on Ab Initio Target Data. *J. Chem. Theory Comput.* **2013**, *9*, 3543–3556.

(22) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Jr., J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.;

Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, Revision B.01,. 2010.

(23) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(24) Head-Gordan, M.; Pople, J. A.; Frisch, M. MP2 Energy Evaluation by Direct Methods. *Chem. Phys. Lett.* **1988**, *153*, 503.

(25) Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(26) Woon, D. E.; Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. III. The Atoms Aluminum Through Argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.

(27) Wilson, A. K.; Woon, D. E.; Peterson, K. A.; Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. IX. The Atoms Gallium Through Krypton. *J. Chem. Phys.* **1999**, *110*, 7667–7676.

(28) Stone, A. J. Distributed Multipole Analysis: Stability for Large Basis Sets. *J. Chem. Theory Comput.* **2005**, *1*, 1128–1132.

(29) Kramer, C.; Bereau, T.; Spinn, A.; Liedl, K. R.; Gedeck, P.; Meuwly, M. Deriving Static

Atomic Multipoles from the Electrostatic Potential. *J. Comput. Inf. Model.* **2013**, *53*, 3410–3417.

(30) Kramer, C.; Gedeck, P.; Meuwly, M. Atomic Multipoles: Electrostatic Potential Fit, Local Reference Axis Systems and Conformational Dependence. *J. Comput. Chem.* **2012**, *33*, 1673–1688.

(31) Stone, A. J. The Description of Bimolecular Potentials, Forces and Torques: The S and V Function Expansions. *Mol. Phys.* **1978**, *36*, 241–256.

(32) Price, S. L.; Stone, A. J.; Alderton, M. Explicit Formulae for the Electrostatic Energy, Forces and Torques Between a Pair of Molecules of Arbitrary Symmetry. *Mol. Phys.* **1984**, *52*, 987–1001.

(33) Koch, U.; Popelier, P. L. A.; Stone, A. J. Conformational Dependence of Atomic Multipole Moments. *Chem. Phys. Lett.* **1995**, *238*, 253–260.

(34) Stone, A. J. *The Theory of Intermolecular Forces*; Clarendon Press Oxford: Oxford, UK, 1996; Vol. 32.

(35) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(36) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(37) van Gunsteren, W.; Berendsen, H. Algorithms for Macromolecular Dynamics and Constraint Dynamics. *Mol. Phys.* **1977**, *34*, 1311–1327.

(38) Wang, J.; Hou, T. Application of Molecular Dynamics Simulations in Molecular Property Prediction. 1. Density and Heat of Vaporization. *J. Chem. Theory Comput.* **2011**, *7*, 2151–2165.

(39) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-Shifted Scaling, a New Scaling Method for Lennard-Jones Interactions in Thermodynamic Integration. *J. Chem. Phys.* **1994**, *100*, 9025–9031.

(40) Boresch, S. The Role of Bonded Energy Terms in Free Energy Simulations - Insights from Analytical Results. *Mol. Sim.* **2002**, *28*, 13–37.

(41) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. PubChem Substance and Compound Databases. *Nucleic Acids Research* **2016**, *44*, D1202–D1213.

(42) Mobley, D. L. Experimental and Calculated Small Molecule Hydration Free Energies. 2013; `http://escholarship.org/uc/item/6sd403pz`, Accessed: 2016-05-18.

(43) Mobley, D. L.; Guthrie, J. P. FreeSolv: A Database of Experimental and Calculated Hydration Free Energies, with Input Files. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 711–720.

(44) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.

(45) Weininger, D.; Weininger, A.; Weininger, J. L. SMILES. 2. Algorithm for Generation of Unique SMILES Notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97–101.

(46) Chamberlin, D. D.; Boyce, R. F. SEQUEL: A Structured English Query Language. Proceedings of the 1974 ACM SIGFIDET (Now SIGMOD) Workshop on Data Description, Access and Control. New York, NY, USA, 1974; pp 249–264.

(47) PubChem REST API. `https://pubchem.ncbi.nlm.nih.gov/pug_rest/PUG_REST.html`, Accessed: 2016-05-18.

(48) Cazade, P.-A.; Bereau, T.; Meuwly, M. Computational Two-Dimensional Infrared Spectroscopy without Maps: N-Methylacetamide in Water. *J. Phys. Chem. B.* **2014**, *118*, 8135–8147.

(49) MacKerell, A. D., Jr.; Shim, J. H.; Anisimov, V. M. Re-evaluation of the Reported Experimental Values of the Heat of Vaporization of N-methylacetamide. *J. Chem. Theory Comput.* **2008**, *4*, 1307–1312.

(50) *CRC Handbook of Chemistry and Physics, 96th Edition*; Boca Raton, Florida: CRC press, 2015.

(51) Riddick, J. A.; Bunger, W. B.; Sakano, T. K. *Physical Properties and Methods of Purification*, 4th ed.; John Wiley & Sons: New York, NY, 1985; Vol. II. Organic Solvents.; p 660.

(52) Abraham, M. H.; Andonian-Haftvan, J.; Whiting, G. S.; Leo, A.; Taft, R. S. Hydrogen Bonding. Part 34. The Factors that Influence the Solubility of Gases and Vapours in Water at 298 K, and a New Method for its Determination. *J. Chem. Soc., Perkin Trans. 2* **1994**, 1777–1791.

(53) Wolfenden, R. Interaction of the Peptide Bond with Solvent Water: a Vapor Phase Analysis. *Biochemistry* **1978**, *17*, 201–204.

(54) Radzicka, A.; Pedersen, L.; Wolfenden, R. Influences of solvent water on protein folding: free energies of solvation of cis and trans peptides are nearly identical. *Biochemistry* **1988**, *27*, 4538–4541.

(55) Jorgensen, W. L.; Gao, J. Cis trans energy difference for the peptide-bond in the gas-phase and in aqueous-solution. *J. Am. Chem. Soc.* **1988**, *110*, 4212–4216.

(56) Cole, D. J.; Vilseck, J. Z.; Tirado-Rives, J.; Payne, M. C.; Jorgensen, W. L. Biomolecular Force Field Parameterization via Atoms-in-Molecule Electron Density Partitioning. *J. Chem. Theory Comput.* **2016**, *12*, 2312–2323.

(57) Ding, Y.; Bernardo, D. N.; Krogh-Jespersen, K.; Levy, R. M. Solvation Free Energies of Small Amides and Amines from Molecular Dynamics/Free Energy Perturbation Simu-

lations Using Pairwise Additive and Many-Body Polarizable Potentials. *J. Phys. Chem.* **1995**, *99*, 11575–11583.

(58) Rick, S. W.; Berne, B. J. Dynamical Fluctuating Charge Force Fields: The Aqueous Solvation of Amides. *J. Am. Chem. Soc.* **1996**, *118*, 672–679.

(59) Spector, T. I.; Kollman, P. A. Investigation of the Anomalous Solvation Free Energies of Amides and Amines: FEP Calculations in Cyclohexane and PS-GVB Calculations on Amide-Water Complexes. *J. Phys. Chem. B.* **1998**, *102*, 4004–4010.

(60) Lu, Y.; Shi, T.; Wang, Y.; Yang, H.; Yan, X.; Luo, X.; Jiang, H.; Zhu, W. Halogen Bonding — A Novel Interaction for Rational Drug Design? *J. Med. Chem.* **2009**, *52*, 2854–2862.

(61) Wilcken, R.; Zimmermann, M. O.; Lange, A.; Joerger, A. C.; Boeckle, F. M. Principles and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **2013**, *56*, 1363–1388.

(62) Hernandes, M. Z.; Cavalcanti, S. M. T.; Moreira, D. R. M.; de Azevedo, J.; Filgueira, W.; Leite, A. C. L. Halogen Atoms in the Modern Medicinal Chemistry: Hints for the Drug Design. *Curr. Drug Targets* **2010**, *11*, 303–314.

(63) Müller, K.; Faeh, C.; Diederich, F. Fluorine in Pharmaceuticals: Looking Beyond Intuition. *Science* **2007**, *317*, 1881–1886.

(64) Pandyarajan, V.; Phillips, N. B.; Cox, G. P.; Yang, Y.; Whittaker, J.; Ismail-Beigi, F.; Weiss, M. A. Biophysical Optimization of a Therapeutic Protein by Nonstandard Mutagenesis: Studies of an iodo-insulin derivative. *J. Biol. Chem.* **2014**, *289*, 23367–23381.

(65) Parsons, G. H.; Rochester, C. H.; Wood, C. E. C. Effect of 4-substitution on the Thermodynamics of Hydration of Phenol and the Phenoxide Anion. *J. Chem. Soc. B* **1971**, 533–536.

(66) Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions. *J. Chem. Theory Comput.* **2006**, *2*, 128–139.

(67) Cazade, P.-A.; Tran, H.; Bereau, T.; Das, A. K.; Kläsi, F.; Hamm, P.; Meuwly, M. Solvation of Fluoro-Acetonitrile in Water by 2D-IR Spectroscopy: A Combined Experimental-Computational Study. *J. Chem. Phys.* **2015**, *142*.

(68) Williams, M. The Merck Index: an Encyclopedia of Chemicals, Drugs, and Biologicals. 14th Edition. *Drug Development Research* **2006**, *67*, 870–870.

(69) Yaws, C. L. *Chemical Properties Handbook : Physical, Thermodynamic, Environmental, Transport, Safety, and Health Related Properties for Organic and Inorganic Chemicals*; McGraw-Hill,: New-York, NY, 1999.

(70) Shirts, M. R.; Pande, V. S. Solvation Free Energies of Amino Acid Side Chain Analogs for Common Molecular Mechanics Water Models. *J. Chem. Phys.* **2005**, *122*, 134508.

(71) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703.

(72) Lee, M. W.; Meuwly, M. On the Role of Nonbonded Interactions in Vibrational Energy Relaxation of Cyanide in Water. *J. Phys. Chem. A.* **2011**, *115*, 5053–5061.

(73) Schmid, F. F.-F.; Meuwly, M. Direct Comparison of Experimental and Calculated NMR Scalar Coupling Constants for Force Field Validation and Adaptation. *J. Chem. Theory Comput.* **2008**, *4*, 1949–1958.

(74) Huang, J.; Meuwly, M. Explicit Hydrogen-Bond Potentials and Their Application to NMR Scalar Couplings in Proteins. *J. Chem. Theory Comput.* **2010**, *6*, 467–476.

(75) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34*, 2135–2145.

(76) Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic Improvement of a Classical Molecular Model of Water. *J. Phys. Chem. B.* **2013**, *117*, 9956–9972.

(77) Kramer, C.; Gedeck, P.; Meuwly, M. Atomic Multipoles: Electrostatic Potential Fit, Local Reference Axis Systems, and Conformational Dependence. *J. Comput. Chem.* **2012**, *33*, 1673–1688.

Graphical TOC Entry