Automatic GROMACS Topology Generation and Comparisons of Force Fields for Solvation Free Energy Calculations

Magnus Lundborg[†] and Erik Lindahl[*,†,‡]

Department of Theoretical Physics and Swedish e-Science Research Center, Royal Institute of Technology, Science for Life Laboratory, Box 1031, SE-171 21 Solna, Sweden, and Department of Biochemistry and Biophysics, Center for Biomembrane Research, Stockholm University, 106 91, Stockholm, Sweden

E-mail: erik@kth.se

Abstract

Free energy calculation has long been an important goal for molecular dynamics simulation and force field development, but historically it has been challenged both by limited performance, accuracy, and creation of topologies for arbitrary small molecules. This has made it difficult to systematically compare different sets of parameters to improve existing force fields, but in the last few years several authors have developed increasingly automated procedures to generate parameters for force fields such as Amber, CHARMM, and OPLS. Here, we present a new framework that enables fully automated generation of GROMACS topologies for any of these force fields and an automated setup for parallel adaptive optimization of high-throughput free energy calculation by adjusting lambda point placement on the fly. We have calculated solvation free energies of

[*]To whom correspondence should be addressed
[†]Royal Institute of Technology
[‡]Stockholm University

50 different small molecules using the GAFF, OPLS-AA and CGenFF force fields and four different water models, and by including the often neglected polarization costs we show that the common charge models are somewhat underpolarized.

1 Introduction

Free energy is of paramount importance in chemistry. Almost all the experimental properties traditionally interpreted e.g. in terms of concentration, reaction rates, stability, folding, complex formation, binding catalysis, or solubility can equally well be described with free energy concepts, in particular on the molecular level. If we could rapidly calculate free energies for arbitrary complex reactions (such as protein folding or an antibody binding an antigen) it would not only be possible to make much more accurate predictions of experimental results from simulations, but it would enable an entirely new level of computational molecular design.

While the most complex systems are still limited by computational performance, the calculation of solvation free energies (i.e., the change in Gibbs free energy upon transfer from gas phase to solvent) has matured rapidly. It is already used in pharmaceutical applications since only a small fraction of commercially available compounds have had their solvation free energy determined experimentally.[1,2] This makes computational predictions tractable, if they are proven to be reliable, and likely to pave the way to more complex applications. For a long time, the calculation of hydration free energies has been a critical performance test of biomolecular force fields used in molecular dynamics (MD) simulations.[3] There have also been a number of blind challenges to predict hydration free energies of provided compounds, with experimental data that is difficult to find, in order to assess the state of the art and to improve current methodology.[1,2,4,5]

A central concern for solvation free energy computations has been whether their accuracy is sufficient for practical use.[6] The biomolecular force fields currently used in molecular dynamics simulations were originally parameterized with amino acids and nucleic acids in

mind. Over a number of years they have been extended to cover generic organic molecules, but some parameters still need to be improved in order to give satisfying results. In particular, it is easy to fool ourselves and believe that the free energy accuracy we get for 20 amino acid residues (for which we have spent almost 40 years improving parameters) are somehow typical for the general force field parameter quality.

Likewise, the commonly used water models are good at reproducing properties of pure liquid water, but they are not quite as reliable for modeling hydration free energies.[7] Lately there have been efforts to amend this, by tweaking the water model parameters to improve the interaction energies without sacrificing the water properties.[7,8] It can be argued, though, that it would be better to use a good water model, such as TIP4P-Ew[9] or SPC/E[10] and modify the force field parameters to improve solvation free energies.[11] The polarization cost when using a fixed charge force field is also often overlooked; studies have suggested that the partial charges commonly used in force fields are somewhat underpolarized.[12,13] This means that the force fields should be re-parameterized using more accurate charges, followed by re-calibration of the van der Waals parameters.[14,15] With these improvements it might be possible to further improve the accuracy of free energy calculations of current fixed charge force fields, rather than switching to polarizable force fields that are both computationally expensive and difficult to parameterize.[13]

These advances have been made possible both by faster computers, and because methods for free energy calculations have improved to the level where the *precision* (but not necessarily accuracy) of calculated solvation free energies now rivals experimental measurements.[7,16] In particular for small systems, this finally makes it possible to separate the classical simulation challenges of sampling efficiency vs. parameter quality and systematically improve both of them.

Calculating the solvation free energy of a small molecule is an important first step to predict its free energy of binding to a protein, which in turn is of interest when studying its effects in a biological system. However, doing this with MD simulations (or Monte Carlo,

which is occasionally used as an alternative sampling technique[17]) requires molecular force field topologies describing the molecules to be studied, which is a particularly difficult hurdle in the early phase of a project when thousands of molecules need to be screened rapidly. For these applications, the question is not how accurately we in theory could parameterize a molecule with manual tuning (cf. the amino acids above), but how efficient automatic methods can be with only a couple of hours of computer time.

To facilitate these types of studies with the GROMACS molecular dynamics package[18,19], we have developed a new tool that enables automatic generation of topologies for generic small molecules: STaGE (Small molecule Topology GEnerator). The name is also meant to describe its usage as a large-scale staging/preprocessing tool ahead of the actual simulations. GROMACS comes with a number of widely used force fields, and an important goal for this development was to enable automatic topology generation for usage both with AMBER[20], OPLS[21,22], and CHARMM[23] force fields to facilitate comparisons. STaGE uses both internal and several external tools, but they have been selected with the criteria that they must be possible to install locally (no web-service-only components) and preferably free open source, or at least completely free for academic research. Our scientific aim is to significantly increase the deployment of free energy calculations by enabling critical assessment of their scientific merits, and avoid issues whether confidential compound information can be sent outside the organization, or whether specific programs justify high licensing costs.

Some functions of STaGE are specific to one or two (optional) external programs, but in general the external components are exchangable, and it is possible to choose e.g. alternative charge generation algorithms. In particular, this means it is possible to completely avoid the few tools that are not freely available even in a commercial setting. The input to STaGE can be almost any molecular file format, including SMILES. A flexible plug-in system makes it easy to add other force fields or modify the provided generation protocols.

In order to illustrate the usability of the program it has been employed to generate GROMACS topologies of an evaluation set of 50 small neutral molecules chosen from the

selection of 504 compounds used by Mobley *et al.*[24] We utilized four different explicit solvent models and three force fields in order to evaluate their performance. The polarization cost of transferring a molecule from vacuum to water was also taken into account to get a more correct free energy estimation.[12,13] This is something that is often overlooked or ignored when calculating hydration free energies with fixed charge models.

The purpose of the STaGE program is to quickly generate topologies of many molecules with as little intervention as possible and with a low error rate. Just as with other automatic, or semi-automatic, topology generation approaches there is no guarantee that the generated topologies are perfect. In order to achieve that they would have to be verified manually. STaGE is equally useful for generation topologies for binding free energy studies, since it can easily combine the topology file of the macromolecule with those of different ligands.

2 Application overview

It is important to point out that STaGE is meant to be used for quickly generating topologies of a large number of molecules. When generating a topology of a single molecule it is highly advisable to invest more time and manually inspect all parameters. However, this is simply not realistic for high-throughput projects using hundreds or thousands of compounds, so in order to create a level playing field that is representative of high-throughput usage we have not touched the automatically generated topologies here.

2.1 Installation

STaGE itself is written in Python and does not need any installation as such. There is, however, a CMake setup that makes it easier to download and compile (if required) external tools, although some of them require accepting license agreements and must be downloaded and installed manually. See Fig. 1 for more information about the external tools used by STaGE.

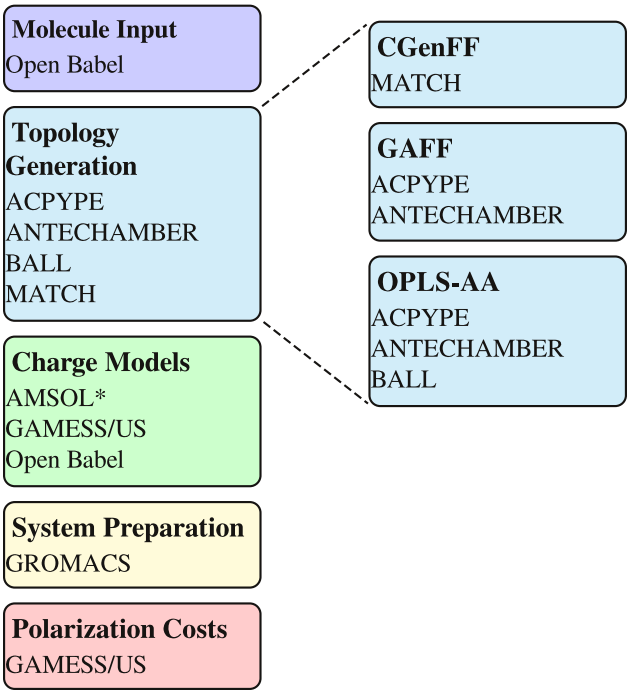


Figure 1: External applications used by STaGE. All applications (ACPYPE[25], AMSOL[26], ANTECHAMBER[27], BALL[28], GAMESS/US[29], GROMACS[18,19], MATCH[30], Open Babel[31]) are free for academic use and most of them are released under open source licenses. Only AMSOL[26] requires a license fee for commercial use, and this is an entirely optional component of STaGE. MATCH is a freely downloadable alternative to the CGenFF program[32,33] for generating parameters. It is possible that parameters assigned using MATCH are not completely correct, which means that erroneous parameters might not be due to CGenFF itself. The right column lists the external applications required for generating topologies of each currently supported force field. The applications in the Charge Models field can be used to assign charges based on specific alternative charge models - the default force field ones are always available. Most functions of STaGE only require a subset of the external applications.

2.2 Molecule input

Any molecular format that can be converted to the mol2 format by Open Babel[31] can be used as input. A specific pH can be specified to set the approximate protonation state accordingly, and when using mol2 as input it is possible to retain the partial charges. In this study, SMILES strings have consistently been used as input.

2.3 Topology generation

Topologies are created using plugins for each force field. STaGE comes with plugins for GAFF[27], OPLS-AA[22] and CGenFF (CHARMM General Force Field)[34], but it will be easy to add future plugins for other force fields as well. Fig. 1 illustrates the applications required for generating topologies for the different force fields.

By default, the GAFF partial charges are assigned according to AM1-BCC[35,36], as generated by ANTECHAMBER[27,37]. For OPLS-AA the default charges are based on the force field atom types and often not suited for assigning charges to molecules with combinations of several functional groups, since there are not enough OPLS-AA atom types to correctly describe all possible combinations of functional groups. This can result in a non-integer net charge, in which case the user is alerted - the easiest solution for this is simply to use one of the alternative partial charge models available in STaGE. The CGenFF partial charges are generated by MATCH using a bond charge increment (BCI) approach.[30]

MD simulations of collinear atoms are typically ill-defined since the force direction from angles at, or near, 180° fluctuates heavily. In GROMACS this problem can be circumvented by turning atoms into massless virtual interaction sites[38], with other properties unmodified, and creating new sites that will act as mass centers, without electrostatic or van der Waals properties. Only one new mass center for each group of collinear atoms is created by STaGE. At least one of the atoms in either end of the linear sequence is retained as a normal atom. If both atoms at the respective end of the collinear sequence have bonds to other atoms they are kept unmodified. The total mass and the moment of inertia is preserved by spreading

the mass of the atoms that are turned into virtual sites to the adjacent mass center and/or the retained end atoms. The GAFF force field usually circumvents the problem with linear angles by setting the angles to slightly below 180°, such as 179.4°. These angles, close to 180°, can still sometimes lead to unstable systems, so we recommend the virtual interaction site approach regardless of timestep.

2.3.1 Alternative Charge Models

While it is possible to use a specific charge model for all force fields, it is important to keep in mind that molecular force fields are parameterized using a specific method for applying partial charges. By definition, all the different charge models available in STaGE cannot correspond to the optimal charges for a specific force field. On the other hand, some of these charge models may reproduce actual charge distributions (dipole moments and electronic surface potentials) better than others, which can be a good reason to use one of them. Many of them have also been successfully used when calculating solvation and binding free energies.[39,40]

AM1-BCC is the default charge model when using GAFF and it is based on Austin Model 1 (AM1) charges[41] with an applied bond charge correction to reproduce the HF/6-31G* electrostatic potential.[35,36] The charges are assigned using ANTECHAMBER.[27,37]

CM1A is a class IV charge model based on AM1[41] wave functions, parameterized to reproduce experimental properties.[42] AMSOL[26] is used for the calculations.

CM3A is similar to CM1A, but developed using a larger training set and more robust[43] and the charges are also assigned using AMSOL[26].

SM5.4/AM1 is the aqueous solvation model SM5 with charges derived from AM1 wave functions.[44] These charges are polarized, as opposed to CM1A and CM3A charges. The calculations are performed using AMSOL[26].

MMFF94 are the charges used in the force field with the same name[45] and assigned using Open Babel[31].

EEM (the Electronegativity Equalization Method) is a quick method to calculate charges similar to B3LYP/6-31G*[46,47]. The charges are calculated using Open Babel.[31]

B3LYP/PCM are charges reproducing the electrostatic potential from quantum mechanics chemistry (QM) using the B3LYP[48] functional method, a polarizable water model (c-PCM)[49–52] and a cc-pV(T+d)Z basis set.[53–56]. The QM calculations is peformed using GAMESS/US[29]. RESP (restrained electrostatic potential)[57] charges are applied using gmstoresp.sh (by Sarnoff Corporation, Princeton, NJ, USA), which in turn uses the respgen and resp programs in the ANTECHAMBER[27,37] program suite. It is important to keep in mind that these QM based charges can take a long time to calculate.

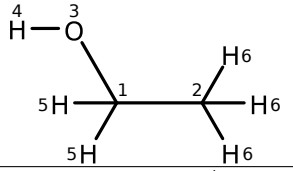
Tables 1 and 2 illustrate the different charge models applied to a few example molecules.

As mentioned above, it is also possible to retain previosly calculated atom charges, for instance when using a mol2 file as input. Charges can be multiplied by a constant factor in order to polarize the molecular charges in case the charge model does not take polarization into account.[39,58,59]

2.4 Solvation

If requested, a rhombic dodecahedron (the periodic unit cell most similar to a sphere, which minimizes the number of water molecules required) solvent box will be generated using the GROMACS editconf and genbox commands with the default minimum distance from the molecule to the edge of the solvent box set to 1.1 nm. In addition to the standard solvent models contained in the GROMACS installation, TIP3P-MOD[8] and TIP3P-M25[7] are also available in STaGE. If the system net charge is not zero it will automatically have ions added to make it neutral, unless the user explicitly asks for charged systems in this case. In this

Table 1: Atomic partial charges of ethanol using the available charge models and the two force field-specific alternatives for CGenFF and OPLS-AA. The atoms are numbered as in the figure. Equivalent hydrogens share index, but the charges are not indentical for all models. GAFF does not have a charge model of its own, but uses AM1-BCC by default. The difference in partial charge on the hydroxyl group (atoms 3 and 4) can make a large difference in hydration free energy.



Charge model	Atoms					
	1	2	3	4	5	6
CGenFF	0.050	-0.270	-0.650	0.420	0.090	0.090
OPLS-AA	0.145	-0.180	-0.683	0.418	0.060	0.060
AM1-BCC	0.126	-0.136	-0.600	0.396	0.043	0.042
CM1A	0.000	-0.254	-0.510	0.352	0.063, 0.103	0.073, 0.085, 0.088
CM3A	0.010	-0.227	-0.493	0.340	0.055, 0.096	0.064, 0.076, 0.079
SM5.4/AM1	-0.003	-0.252	-0.561	0.392	0.078, 0.085	0.077, 0.091, 0.093
MMFF94	0.280	0.000	-0.680	0.400	0.000	0.000
EEM	-0.016	-0.430	-0.582	0.276	0.122, 0.134	0.157, 0.164, 0.175
B3LYP/PCM	0.411	-0.270	-0.711	0.419	-0.021	0.065

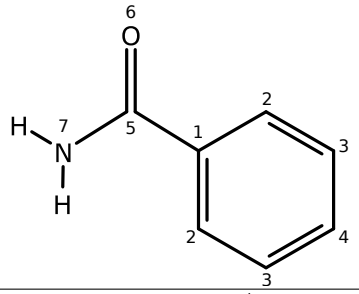
context it should be emphasized that the solvation process only refers to the generation of the solvent box around the molecule - it does not alter the atomic charges of the solute.

The water model can be a suprisingly difficult choice; by default we recommend the the TIP4P-Ew[9] and SPC/E[10] water models since they have been parameterized with the polarization cost of water taken into account, and they reproduce water properties well[11], but as evident from the results below this does not automatically mean they provide the most accurate results in all cases.

2.5 Polarization costs

Along with STaGE there is also a Python script for calculating the free energy cost of transferring a molecule from vacuum to a solvent when using a force field with fixed partial charges.[12,13] The calculations are performed as described by Swope *et al.*[13] Only the dipolar component of the polarization cost is calculated. The user can either provide the output of a

Table 2: Atomic partial charges of benzamide using the available charge models and the two force field-specific alternatives for CGenFF and OPLS-AA. The atoms are numbered as in the figure. Only the partial charges of the heavy atoms are shown in the table and for space reasons nonpolar hydrogens are not shown (their specific charges have relatively little effect). Equivalent atoms share index, but the charges are not identical for all models. GAFF does not have a charge model of its own, but uses AM1-BCC by default. The differences in partial charge for the carbonyl oxygen (index 6) and amine nitrogen (index 7) have significant impact on hydration free energy.



Charge model	Atoms						
	1	2	3	4	5	6	7
CGenFF	-0.020	-0.115	-0.115	-0.115	0.530	-0.510	-0.680
OPLS-AA	-0.115	-0.115	-0.115	-0.115	0.615	-0.500	-0.760
AM1-BCC	-0.142	-0.091	-0.139	-0.109	0.671	-0.610	-0.674
CM1A	-0.116	-0.076, -0.103	-0.138, -0.139	-0.108	0.590	-0.400	-1.132
CM3A	-0.118	-0.068, -0.095	-0.129, -0.131	-0.100	0.512	-0.483	-0.861
SM5.4/AM1	-0.143	-0.105, -0.109	-0.149, -0.150	-0.118	0.559	-0.519	-0.990
MMFF94	0.086	-0.150	-0.150	-0.150	0.544	-0.570	-0.800
EEM	-0.057	-0.084, -0.097	-0.101, -0.102	-0.100	0.555	-0.522	-0.873
B3LYP/PCM	-0.088	-0.108	-0.118	-0.117	0.707	-0.636	-0.870

GAMESS/US[29] calculation, with dipole polarizability and dipole moment, or supply a mol2 file to start a structure optimization (B3LYP with a cc-pV(T+d)Z basis set[48,53–56]), followed by calculations to generate the dipole moment and polarizability (B3LYP calculations with an aug-cc-pV(T+d)Z basis set[48,56]). The dipole moment of the molecule in solvent is calculated from the partial charges in a GROMACS topology file and the coordinates from the optimized structure in gas phase from the previous GAMESS/US calculations. The molecule center point for the dipole moment calculations is take from the GAMESS/US output of the gas phase dipole calculations. The polarization cost depends on the dipole polarizability and the difference in molecular dipole moment between gas phase and solvent phase as[13]

$$W_{\mathrm{pol}}^{\mathrm{D}} = \frac{1}{2}(\mu - \mu^0)^{\mathrm{T}}(\alpha^{-1})^{\mathrm{T}}(\mu - \mu^0), \tag{1}$$

where μ and μ^0 are the dipole moments in solvent and gas phase, respectively, and α is the dipole–dipole polarizability tensor. The superscript T indicates that the expression should be transposed. The included STaGE script makes it straightforward to account for this polarizability for all solvation free energy calculations, either before or after the actual MD simulations. It is trivial to compare the polarization costs of different charge models since the time consuming QM calculations do not have to be re-executed.

3 Methods

3.1 System preparation

50 molecules were selected from the test set used by Mobley *et al.*[6] to obtain a good coverage of different functional groups and low to high solvation free energies. The small molecule topologies and the solvated systems were generated from SMILES, using STaGE to obtain GAFF (General Amber Force Field)[27], OPLS-AA[22] and CGenFF (CHARMM General Force Field)[34] topologies. Unless otherwise stated, the suggested charge model was used for each

force field, i.e. the MATCH bond charge increment method for CGenFF, AM1-BCC version 1 for GAFF[35,36] and atom type-based partial charges for OPLS-AA. Rhombic dodecahedron solvent boxes were generated with a minimum distance of 1.1 nm between the small molecule and the nearest edge of the box, as illustrated in Figure 2.

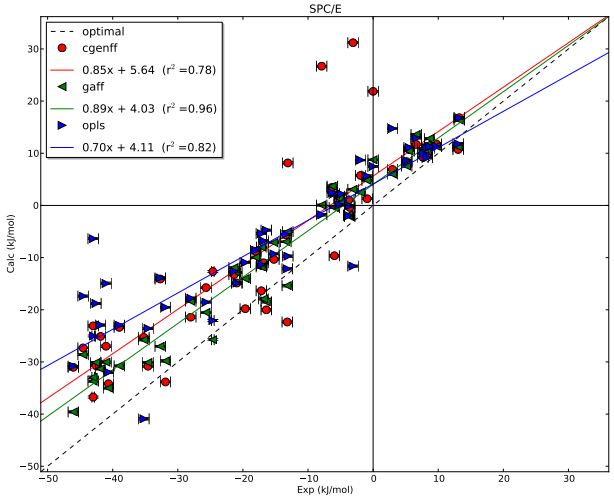


Figure 2: 4-methyl-1h-imidazole in a rhombic dodecahedron water box. 461 water molecules are included to keep the distance between the solute molecule and the nearest box edge above 1.1 nm.

The evaluation set was prepared using four different water models, viz. TIP3P[60], TIP3P-M25[7], TIP4P-Ew[9] and SPC/E[10]. The total number of systems to simulate was 600 (50 molecules, three force fields and four water models). In addition, the B3LYP/PCM charge model was used in combination with GAFF in SPC/E water in order to evaluate the effect of using a more polarized charge model.

3.2 Simulation parameters

The simulations were performed using GROMACS[18,19] version 4.6.3. The simulation protocol started with steepest-descent energy minimizations, first 1500 steps with flexible bonds, followed by 1500 additional steps with bonds constrained using the P-LINCS[61] algorithm, which was also used in all subsequent stages. The minimizations and simulations were run

using smooth Particle-Mesh Ewald (SPME) electrostatics.[62] The temperature during the simulations was 298 K, coupled using a velocity rescaling thermostat[63] and the pressure (when running NPT) was 1 bar, controlled using a Parrinello-Rahman barostat.[64] Equilibration was performed in three stages, the first stage in the NVT ensemble and the subsequent stages, as well as the actual production phase, in the isothermal-isobaric ensemble (NPT). During the first two equilibration stages the atoms of the solute were restrained. The simulation time step was 2.0 fs and the simulation length was 50 ps in each equilibration stage. A cut-off set to 1.0 nm was used for van der Waals interactions and the same radius was used for the short-range PME component. This is shorter than what is recommended for CGenFF, which could influence the results somewhat, but it was decided to use the same settings for all force fields - it is also a very common choice for simulations where performance matters. There should not be any systematic effects from the van der Waals cutoff distance since long-range dispersion corrections were applied for energy and pressure.

The solvation free energy calculations were performed using GROMACS and the Copernicus[65] parallel adaptive simulation toolkit. The lambda point distribution is optimized fully automatically in Copernicus, by starting a number of shorter trial simulation, then calculating the sampling overlap between points based on the provisional lambda point distribution, and finally adjusting the location and spacing of lambda points. This is followed by automatic execution of the production simulations on all hardware clients available to the Copernicus server, after which the server uses the Bennett Acceptance Ratio (BAR) method[66] to calculate the change in free energy upon turning off interaction with the environment using lambda points. Coulomb and van der Waals interactions were decoupled independently. A brief summary of the procedure is given in Figure 3.

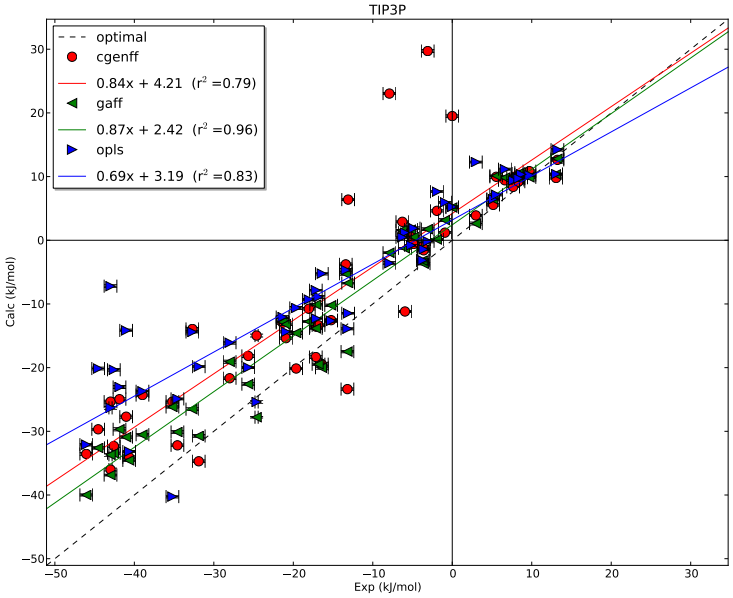


Figure 3: Iterative procedure for calculating $\Delta G_{solvation}$. When optimizing lambda point distributions the target for standard deviation per sample in each lambda interval is 1 $k_B T$. The provisional lambda point distribution is retained if the number of lambda points changed or if any lambda interval changed by more than 20%, otherwise the lambda distribution is not changed from what was previously used. ΔG and the estimated ΔG error are calculated using the Bennett Acceptance Ratio (BAR) (using the g_bar GROMACS tool).

4 Results and Discussion

When generating the OPLS-AA topologies the SMARTS matching of atoms of 1-methyl-imidazole did not correctly assign some atom types, resulting in a non-zero net charge (-0.306) of the molecule. The molecules 3-acetylpyridine, 3-methyl-1h-indole, 4-acetylpyridine and 4-methyl-1h-imidazole seemed to have their atom types set correctly, but their net charges were still not correct (-0.040, -0.057, 0.050 and 0.080 respectively). These molecules are still retained in the OPLS-AA simulations, but it should be kept in mind that the statistics might be unfair because of this. The experimental reference hydration free energy values were obtained from Rizzo *et al.*[67], except for 3-methyl-1h-indole[16], 4-methyl-1h-imidazole[16] and cyanobenzene (benzonitrile)[68].

4.1 SPC/E

A plot of the calculated vs. experimental hydration free energies is presented in Figure 4 and more detailed results are available in Table 3. The calculated results from CGenFF had a root-mean-square error (RMSE) of 7.94 kJ/mol (mean error 4.83 kJ/mol) compared to the experimental data, whereas GAFF had an RMSE of 5.95 kJ/mol (mean error 4.69 kJ/mol) and OPLS-AA had 8.92 (mean error 5.45 kJ/mol). Notably, when including the polarization costs the RMSE increased to 11.61, 6.89 and 11.84 kJ/mol. The hydration free energies are overestimated in SPC/E for all the studied force fields. The OPLS-AA outliers are mainly the molecules with low hydration free energies, four of which have too high calculated ΔG and one too low. The four obvious CGenFF outliers all have an overestimated ΔG. It seems like the chloro and bromo compounds get too high hydration free energy, whereas fluoro and iodo compounds perform better in CGenFF. GAFF has a few outliers from the linear correlation, but those are closer to the experimental ΔG, which makes it difficult to draw any conclusions from them.

The results from using the more polarized charge model B3LYP/PCM are presented in

Table 4. In this case only GAFF was used and the RMSE was 6.30 kJ/mol (mean error -0.08 kJ/mol). Including polarization costs reduced the RMSE to 5.15 kJ/mol as expected from a properly polarized charge model. The overall agreement with experimental data is much better using this charge model. The two main outliers are 2-ethoxyethanol and especially trimethylamine.

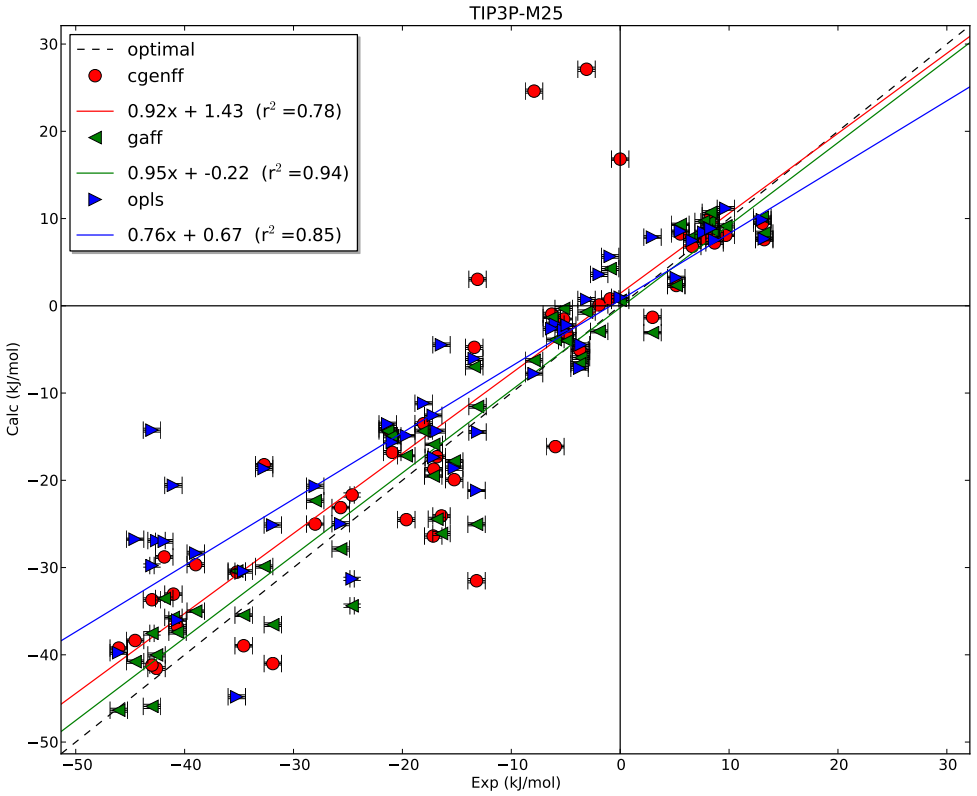


Figure 4: Calculated solvation free energies of 50 compounds in SPC/E water using three different force fields. Polarization costs are applied to the calculated values. The GAFF results agree best with experimental data for this set. The CGenFF fit also has a fairly good slope, but there are some clear outliers negatively affecting the predictability for an individual compound.

4.2 TIP3P

The results of the calculations using the TIP3P water model are illustrated in Figure 5 and Table 5. The calculated results from CGenFF had an RMSE of 7.03 kJ/mol (mean error

Table 3: Solvation free energies in SPC/E water using the CGenFF, GAFF and OPLS-AA force fields. All values in kJ/mol.

Molecule	**Exp.**	**CGenFF**				**GAFF**				**OPLS-AA**			
	ΔG	ΔG	UE	ΔG_{corr}	UE_{corr}	ΔG	UE	ΔG_{corr}	UE_{corr}	ΔG	UE	ΔG_{corr}	UE_{corr}
1-chlorohexane	0.00	16.83	16.83	21.87	21.87	8.63	8.63	8.75	8.75	6.46	6.46	7.47	7.47
1-methyl-imidazole	−35.21	−31.41	3.80	−25.30	9.91	−25.92	9.29	−25.75	9.46	−57.65	22.44	−40.91	5.70
1,3-dichloropropane	−7.91	19.53	27.44	26.70	34.61	0.03	7.94	0.07	7.98	−1.83	6.08	−1.79	6.12
2-ethoxyethanol	−28.01	−24.38	3.63	−21.41	6.60	−18.48	9.53	−18.36	9.65	−19.95	8.06	−17.79	10.22
2-iodopropane	−1.93	2.26	4.19	5.75	7.68	0.18	2.11	2.54	4.47	−0.31	1.62	8.70	10.63
2-methylpropane	9.71	11.69	1.98	11.77	2.06	11.28	1.57	11.35	1.64	11.25	1.54	11.33	1.62
2-nitropropane	−13.10	−0.38	12.72	8.15	21.25	−5.67	7.43	−5.02	8.08	−10.07	3.03	−9.72	3.38
2,2,2-trifluoroethanol	−18.05	−11.69	6.36	−8.81	9.24	−14.19	3.86	−9.96	8.09	−15.49	2.56	−8.59	9.46
3-acetylpyridine	−34.58	−31.97	2.61	−30.82	3.76	−30.29	4.29	−30.12	4.46	−25.04	9.54	−23.58	11.00
3-methyl-1h-indole	−24.62	−18.72	5.90	−12.67	11.95	−25.87	1.25	−25.74	1.12	−22.41	2.21	−22.05	2.57
4-acetylpyridine	−31.90	−35.62	3.72	−33.83	1.93	−30.95	0.95	−29.81	2.09	−23.14	8.76	−19.53	12.37
4-cyanophenol	−42.58	−36.66	5.92	−30.71	11.87	−32.12	10.46	−30.12	12.46	−30.91	11.67	−18.78	23.80
4-methyl-1h-imidazole	−43.00	−36.79	6.21	−36.73	6.27	−33.32	9.68	−32.95	10.05	−27.81	15.19	−25.02	17.98
4-nitroaniline	−43.00	−31.24	11.76	−23.09	19.91	−35.75	7.25	−33.69	9.31	−28.73	14.27	−6.38	36.62
4-nitrophenol	−44.55	−32.29	12.26	−27.31	17.24	−31.80	12.75	−28.60	15.95	−29.73	14.82	−17.38	27.17
acenaphthene	−13.19	−24.41	11.22	−22.34	9.15	−15.70	2.51	−15.37	2.18	−12.19	1.00	−12.15	1.04
benzaldehyde	−16.83	−12.33	4.50	−11.00	5.83	−19.37	2.54	−17.92	1.09	−12.12	4.71	−6.79	10.04
benzamide	−46.05	−33.05	13.00	−31.03	15.02	−42.51	3.54	−39.58	6.47	−36.96	9.09	−30.82	15.23
benzene	−3.60	−0.38	3.22	−0.38	3.22	−2.12	1.48	−2.12	1.48	−0.08	3.52	−0.08	3.52
bromoethane	−3.09	−1.66	1.43	31.19	34.28	3.07	6.16	3.08	6.17	−17.38	14.29	−11.63	8.54
butane	8.67	10.93	2.26	10.93	2.26	12.81	4.14	12.81	4.14	11.31	2.64	11.31	2.64
cyanobenzene	−17.17	−17.50	0.33	−16.36	0.81	−11.83	5.34	−11.78	5.39	−8.93	8.24	−5.41	11.76
cyclohexane	5.15	8.39	3.24	8.39	3.24	7.55	2.40	7.55	2.40	8.60	3.45	8.60	3.45
decan-1-ol	−15.24	−10.82	4.42	−10.34	4.90	−7.32	7.92	−7.11	8.13	−9.86	5.38	−9.29	5.95
di-n-propylether	−4.86	−1.54	3.32	0.57	5.43	−0.55	4.31	0.60	5.46	1.38	6.24	2.21	7.07
ethanamide	−40.65	−35.41	5.24	−34.18	6.47	−37.60	3.05	−35.01	5.64	−33.73	6.92	−31.98	8.67
ethane	7.66	9.22	1.56	9.22	1.56	10.89	3.23	10.89	3.23	9.95	2.29	9.95	2.29
ethanol	−20.93	−18.37	2.56	−14.87	6.06	−14.08	6.85	−12.85	8.08	−18.94	1.99	−14.88	6.05
fluoromethane	−0.92	1.07	1.99	1.29	2.21	4.79	5.71	4.81	5.73	3.06	3.98	5.64	6.56
hex-1-ene	6.62	9.25	2.63	11.70	5.08	13.25	6.63	13.60	6.98	11.74	5.12	12.95	6.33
methane	8.33	9.75	1.42	9.75	1.42	10.39	2.06	10.39	2.06	9.16	0.83	9.16	0.83
methanethiol	−5.19	0.38	5.57	1.75	6.94	−1.03	4.16	1.31	6.50	−1.23	3.96	0.13	5.32
methanol	−21.35	−18.88	2.47	−13.27	8.08	−13.56	7.79	−11.92	9.43	−17.94	3.41	−12.61	8.74
methyl benzoate	−16.41	−22.73	6.32	−20.01	3.60	−20.29	3.88	−18.49	2.08	−9.40	7.01	−4.74	11.67
methyl ethyl sulfide	−6.28	3.18	9.46	3.49	9.77	3.06	9.34	3.70	9.98	0.63	6.91	2.47	8.75
n-acetylpyrrolidine	−41.03	−27.31	13.72	−26.98	14.05	−31.88	9.15	−30.04	10.99	−17.84	23.19	−14.96	26.07
n-butylacetamide	−38.98	−28.78	10.20	−23.38	15.60	−32.39	6.59	−30.75	8.23	−23.31	15.67	−22.96	16.02
n-decane	13.23	16.81	3.58	16.82	3.59	16.37	3.14	16.37	3.14	16.76	3.53	16.76	3.53
n-methylacetamide	−41.87	−25.15	16.72	−25.10	16.77	−32.50	9.37	−31.32	10.55	−23.06	18.81	−22.93	18.94
n,n-dimethylformamide	−32.70	−21.96	10.74	−14.10	18.60	−29.76	2.94	−27.01	5.69	−16.46	16.24	−13.90	18.80
oct-1-yne	2.97	6.38	3.41	6.89	3.92	5.86	2.89	6.02	3.05	10.84	7.87	14.75	11.78
octan-1-ol	−17.12	−13.05	4.07	−11.53	5.59	−9.26	7.86	−7.99	9.13	−13.09	4.03	−11.38	5.74
p-cresol	−25.67	−16.50	9.17	−15.76	9.91	−21.32	4.35	−20.51	5.16	−21.74	3.93	−18.54	7.13
propane	8.21	9.76	1.55	9.80	1.59	11.54	3.33	11.58	3.37	10.90	2.69	10.94	2.73
propene	5.53	6.74	1.21	10.69	5.16	10.00	4.47	10.44	4.91	9.12	3.59	11.10	5.57
pyridine	−19.64	−20.08	0.44	−19.80	0.16	−14.51	5.13	−14.04	5.60	−11.02	8.62	−10.95	8.69
tetrafluoromethane	13.06	10.72	2.34	10.72	2.34	11.24	1.82	11.24	1.82	11.81	1.25	11.81	1.25
thiophene	−5.95	−9.75	3.80	−9.64	3.69	−0.57	5.38	−0.41	5.54	−0.33	5.62	2.26	8.21
toluene	−3.73	0.67	4.40	1.08	4.81	−1.91	1.82	−1.82	1.91	−2.05	1.68	−1.95	1.78
trimethylamine	−13.40	−12.50	0.90	−5.64	7.76	−13.23	0.17	−6.95	6.45	−11.00	2.40	−5.55	7.85
			5.95		8.70		5.17		6.03		6.97		9.29

Table 4: Solvation free energies in SPC/E water using the GAFF force field and the B3LYP/PCM charge model. All values in kJ/mol.

Molecule	**Exp.**	**GAFF**			
	ΔG	ΔG	UE	ΔG_{corr}	UE_{corr}
1-chlorohexane	0.00	2.81	2.81	3.74	3.74
1-methyl-imidazole	−35.21	−41.39	6.18	−36.49	1.28
1,3-dichloropropane	−7.91	−3.55	4.36	−1.04	6.87
2-ethoxyethanol	−28.01	−18.13	9.88	−16.83	11.18
2-iodopropane	−1.93	−0.52	1.41	2.04	3.97
2-methylpropane	9.71	11.71	2.00	11.73	2.02
2-nitropropane	−13.10	−21.28	8.18	−19.06	5.96
2,2,2-trifluoroethanol	−18.05	−15.47	2.58	−11.35	6.70
3-acetylpyridine	−34.58	−32.77	1.81	−32.51	2.07
3-methyl-1h-indole	−24.62	−22.84	1.78	−21.93	2.69
4-acetylpyridine	−31.90	−32.62	0.72	−30.73	1.17
4-cyanophenol	−42.58	−48.97	6.39	−44.54	1.96
4-methyl-1h-imidazole	−43.00	−45.32	2.32	−41.50	1.50
4-nitroaniline	−43.00	−62.27	19.27	−47.06	4.06
4-nitrophenol	−44.55	−50.74	6.19	−44.76	0.21
acenaphthene	−13.19	−6.57	6.62	−6.26	6.93
benzaldehyde	−16.83	−20.32	3.49	−17.25	0.42
benzamide	−46.05	−50.57	4.52	−45.61	0.44
benzene	−3.60	−1.31	2.29	−1.31	2.29
bromoethane	−3.09	−0.99	2.10	1.49	4.58
butane	8.67	12.10	3.43	12.10	3.43
cyanobenzene	−17.17	−25.92	8.75	−22.42	5.25
cyclohexane	5.15	8.33	3.18	8.33	3.18
decan-1-ol	−15.24	−11.80	3.44	−11.04	4.20
di-n-propylether	−4.86	2.64	7.50	3.68	8.54
ethanamide	−40.65	−52.79	12.14	−45.72	5.07
ethane	7.66	10.86	3.20	10.86	3.20
ethanol	−20.93	−20.52	0.41	−18.01	2.92
fluoromethane	−0.92	1.61	2.53	4.03	4.95
hex-1-ene	6.62	10.26	3.64	10.32	3.70
methane	8.33	10.63	2.30	10.63	2.30
methanethiol	−5.19	−1.84	3.35	1.87	7.06
methanol	−21.35	−19.04	2.31	−15.64	5.71
methyl benzoate	−16.41	−20.64	4.23	−19.75	3.34
methyl ethyl sulfide	−6.28	−0.01	6.27	1.91	8.19
n-acetylpyrrolidine	−41.03	−51.67	10.64	−44.64	3.61
n-butylacetamide	−38.98	−44.35	5.37	−39.53	0.55
n-decane	13.23	17.42	4.19	17.43	4.20
n-methylacetamide	−41.87	−51.42	9.55	−43.83	1.96
n,n-dimethylformamide	−32.70	−43.92	11.22	−35.39	2.69
oct-1-yne	2.97	−2.52	5.49	−1.75	4.72
octan-1-ol	−17.12	−17.86	0.74	−15.00	2.12
p-cresol	−25.67	−17.46	8.21	−16.83	8.84
propane	8.21	11.92	3.71	11.92	3.71
propene	5.53	7.40	1.87	7.51	1.98
pyridine	−19.64	−17.32	2.32	−15.50	4.14
tetrafluoromethane	13.06	11.88	1.18	11.88	1.18
thiophene	−5.95	−2.99	2.96	−2.90	3.05
toluene	−3.73	0.98	4.71	1.02	4.75
trimethylamine	−13.40	2.04	15.44	3.09	16.49
			4.98		4.19

3.59 kJ/mol), whereas GAFF had an RMSE of 4.93 kJ/mol (mean error 3.29 kJ/mol) and OPLS-AA had 8.07 kJ/mol (mean error 4.70 kJ/mol). When including the polarization costs the RMSE increased to 10.65, 5.88 and 11.09 kJ/mol respectively. Compared to SPC/E it is clear that TIP3P performs better, with overall results closer to the experimental values. The outliers of the respective force fields are the same as when using SPC/E.

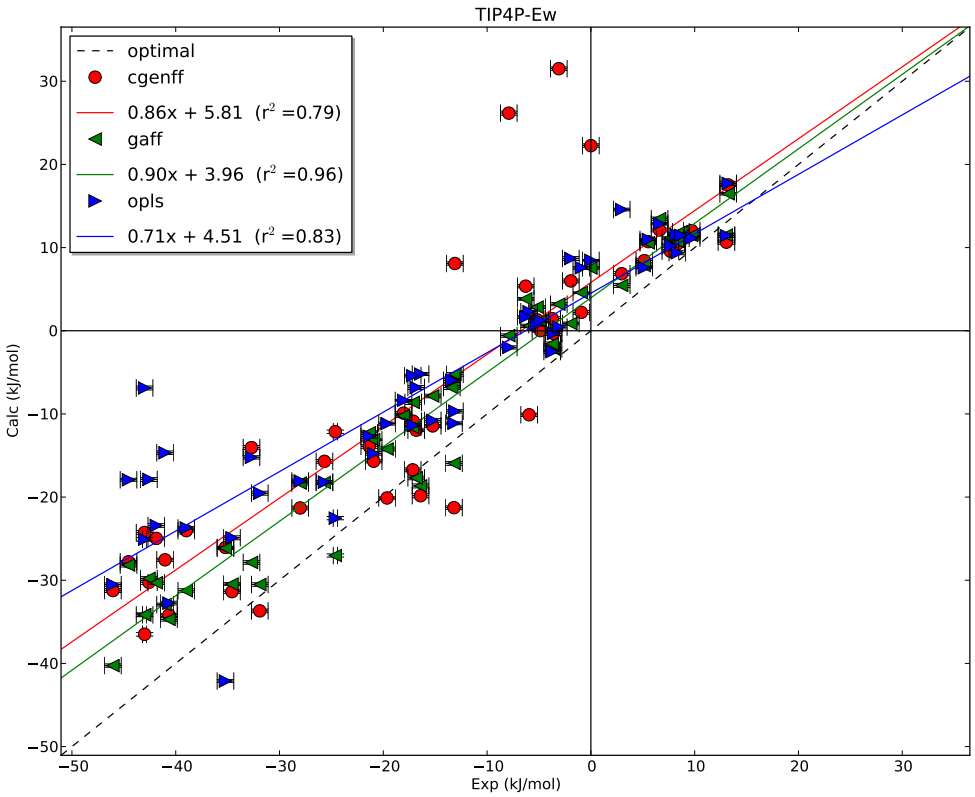


Figure 5: Calculated solvation free energies of 50 compounds in TIP3P water using three different force fields. Polarization costs are applied to the calculated values. All force fields produce lower hydration free energies in TIP3P water than SPC/E (Figure 4). GAFF performs better than the two other force fields, but the very hydrophilic compounds still have too high solvation free energies, which is a trend visible in the other two force fields as well.

4.3 TIP3P-M25

Results of the calculations using the TIP3P-M25 water model are presented in Figure 6 and Table 6. The calculated results from CGenFF had an RMSE of 6.79 kJ/mol (mean error

Table 5: Solvation free energies in TIP3P water using the CGenFF, GAFF and OPLS-AA force fields. All values in kJ/mol.

Molecule	**Exp.**	**CGenFF**				**GAFF**				**OPLS-AA**			
	ΔG	ΔG	UE	ΔG_{corr}	UE_{corr}	ΔG	UE	ΔG_{corr}	UE_{corr}	ΔG	UE	ΔG_{corr}	UE_{corr}
1-chlorohexane	0.00	14.46	14.46	19.50	19.50	5.08	5.08	5.14	5.14	4.23	4.23	5.24	5.24
1-methyl-imidazole	−35.21	−31.45	3.76	−25.34	9.87	−26.34	8.87	−26.17	9.04	−56.99	21.78	−40.25	5.04
1,3-dichloropropane	−7.91	15.87	23.78	23.04	30.95	−2.00	5.91	−1.96	5.95	−3.63	4.28	−3.59	4.32
2-ethoxyethanol	−28.01	−24.63	3.38	−21.66	6.35	−19.20	8.81	−19.08	8.93	−18.24	9.77	−16.08	11.93
2-iodopropane	−1.93	1.13	3.06	4.62	6.55	−2.24	0.31	0.12	2.05	−1.38	0.55	7.63	9.56
2-methylpropane	9.71	10.77	1.06	10.85	1.14	9.82	0.11	9.89	0.18	10.16	0.45	10.24	0.53
2-nitropropane	−13.10	−2.14	10.96	6.39	19.49	−7.37	5.73	−6.72	6.38	−11.81	1.29	−11.46	1.64
2,2,2-trifluoroethanol	−18.05	−13.64	4.41	−10.76	7.29	−16.99	1.06	−12.76	5.29	−16.21	1.84	−9.31	8.74
3-acetylpyridine	−34.58	−33.35	1.23	−32.20	2.38	−30.28	4.30	−30.11	4.47	−26.37	8.21	−24.91	9.67
3-methyl-1h-indole	−24.62	−21.01	3.61	−14.96	9.66	−27.91	3.29	−27.78	3.16	−25.75	1.13	−25.39	0.77
4-acetylpyridine	−31.90	−36.49	4.59	−34.70	2.80	−31.87	0.03	−30.73	1.17	−23.46	8.44	−19.85	12.05
4-cyanophenol	−42.58	−38.26	4.32	−32.31	10.27	−35.44	7.14	−33.44	9.14	−32.46	10.12	−20.33	22.25
4-methyl-1h-imidazole	−43.00	−36.06	6.94	−36.00	7.00	−34.06	8.94	−33.69	9.31	−29.11	13.89	−26.32	16.68
4-nitroaniline	−43.00	−33.49	9.51	−25.34	17.66	−38.93	4.07	−36.87	6.13	−29.56	13.44	−7.21	35.79
4-nitrophenol	−44.55	−34.66	9.89	−29.68	14.87	−35.80	8.75	−32.60	11.95	−32.49	12.06	−20.14	24.41
acenaphthene	−13.19	−25.44	12.25	−23.37	10.18	−17.81	4.62	−17.48	4.29	−13.94	0.75	−13.90	0.71
benzaldehyde	−16.83	−14.68	2.15	−13.35	3.48	−20.92	4.09	−19.47	2.64	−14.22	2.61	−8.89	7.94
benzamide	−46.05	−35.57	10.48	−33.55	12.50	−42.92	3.13	−39.99	6.06	−38.23	7.82	−32.09	13.96
benzene	−3.60	−1.58	2.02	−1.58	2.02	−3.67	0.07	−3.67	0.07	−1.42	2.18	−1.42	2.18
bromoethane	−3.09	−3.14	0.05	29.71	32.80	1.73	4.82	1.74	4.83	−1.28	1.81	−0.16	2.93
butane	8.67	9.82	1.15	9.82	1.15	10.61	1.94	10.61	1.94	10.35	1.68	10.35	1.68
cyanobenzene	−17.17	−19.48	2.31	−18.34	1.17	−13.88	3.29	−13.83	3.34	−11.36	5.81	−7.84	9.33
cyclohexane	5.15	5.55	0.40	5.55	0.40	6.44	1.29	6.44	1.29	6.45	1.30	6.45	1.30
decan-1-ol	−15.24	−13.02	2.22	−12.54	2.70	−10.60	4.64	−10.22	5.02	−13.22	2.02	−12.65	2.59
di-n-propylether	−4.86	−2.69	2.17	−0.58	4.28	−0.71	4.15	0.51	5.37	1.06	5.92	1.89	6.75
ethanamide	−40.65	−35.31	5.34	−34.08	6.57	−37.15	3.50	−34.56	6.09	−34.93	5.72	−33.18	7.47
ethane	7.66	8.38	0.72	8.38	0.72	10.04	2.38	10.04	2.38	9.39	1.73	9.39	1.73
ethanol	−20.93	−18.87	2.06	−15.37	5.56	−14.42	6.51	−13.19	7.74	−18.43	2.50	−14.37	6.56
fluoromethane	−0.92	0.98	1.90	1.20	2.12	3.15	4.07	3.17	4.09	3.38	4.30	5.96	6.88
hex-1-ene	6.62	6.97	0.35	9.42	2.80	9.48	2.86	9.83	3.21	9.95	3.33	11.16	4.54
methane	8.33	9.20	0.87	9.20	0.87	10.48	2.15	10.48	2.15	9.70	1.37	9.70	1.37
methanethiol	−5.19	−0.67	4.52	0.70	5.89	−0.66	4.53	1.68	6.87	−2.15	3.04	−0.79	4.40
methanol	−21.35	−18.57	2.78	−12.96	8.39	−14.47	6.88	−12.83	8.52	−17.39	3.96	−12.06	9.29
methyl benzoate	−16.41	−22.17	5.76	−19.45	3.04	−21.82	5.41	−20.02	3.61	−9.53	6.88	−5.23	11.18
methyl ethyl sulfide	−6.28	2.59	8.87	2.90	9.18	0.94	7.22	1.58	7.86	−1.38	4.90	0.46	6.74
n-acetylpyrrolidine	−41.03	−28.03	13.00	−27.70	13.33	−32.75	8.28	−30.91	10.12	−17.04	23.99	−14.16	26.87
n-butylacetamide	−38.98	−29.71	9.27	−24.31	14.67	−32.97	6.01	−30.53	8.45	−24.06	14.92	−23.71	15.27
n-decane	13.23	12.62	0.61	12.63	0.60	12.80	0.43	12.80	0.43	14.22	0.99	14.22	0.99
n-methylacetamide	−41.87	−25.00	16.87	−24.95	16.92	−30.86	11.01	−29.68	12.19	−23.17	18.70	−23.04	18.83
n,n-dimethylformamide	−32.70	−21.78	10.92	−13.92	18.78	−29.27	3.43	−26.52	6.18	−16.96	15.74	−14.40	18.30
oct-1-yne	2.97	3.39	0.42	3.90	0.93	2.47	0.50	2.63	0.34	8.37	5.40	12.28	9.31
octan-1-ol	−17.12	−15.09	2.03	−13.57	3.55	−11.46	5.66	−10.11	7.01	−14.00	3.12	−12.29	4.83
p-cresol	−25.67	−18.90	6.77	−18.16	7.51	−23.37	2.30	−22.60	3.07	−23.17	2.50	−19.97	5.70
propane	8.21	9.43	1.22	9.47	1.26	9.84	1.63	9.88	1.67	9.72	1.51	9.76	1.55
propene	5.53	6.01	0.48	9.96	4.43	9.69	4.16	10.13	4.60	5.16	0.37	7.14	1.61
pyridine	−19.64	−20.42	0.78	−20.14	0.50	−15.03	4.61	−14.56	5.08	−10.70	8.94	−10.63	9.01
tetrafluoromethane	13.06	9.78	3.28	9.78	3.28	10.35	2.71	10.35	2.71	10.40	2.66	10.40	2.66
thiophene	−5.95	−11.29	5.34	−11.18	5.23	−1.42	4.53	−1.26	4.69	−1.40	4.55	1.19	7.14
toluene	−3.73	−1.41	2.32	−1.00	2.73	−3.29	0.44	−3.20	0.53	−3.24	0.49	−3.14	0.59
trimethylamine	−13.40	−10.64	2.76	−3.78	9.62	−11.54	1.86	−5.26	8.14	−10.13	3.27	−4.68	8.72
			5.13		7.86		4.20		5.14		5.84		8.41

-0.37 kJ/mol), whereas GAFF had an RMSE of 4.36 kJ/mol (mean error -0.40 kJ/mol) and OPLS-AA had 6.41 kJ/mol (mean error 1.14 kJ/mol). When including the polarization costs the RMSE for CGenFF increased to 9.30 kJ/mol, GAFF to 4.54 kJ/mol and OPLS-AA to 8.39 kJ/mol. TIP3P-M25 outperforms TIP3P at predicting hydration free energies. This is not surprising since it was developed to improve exactly that, but at the expense of general water properties.

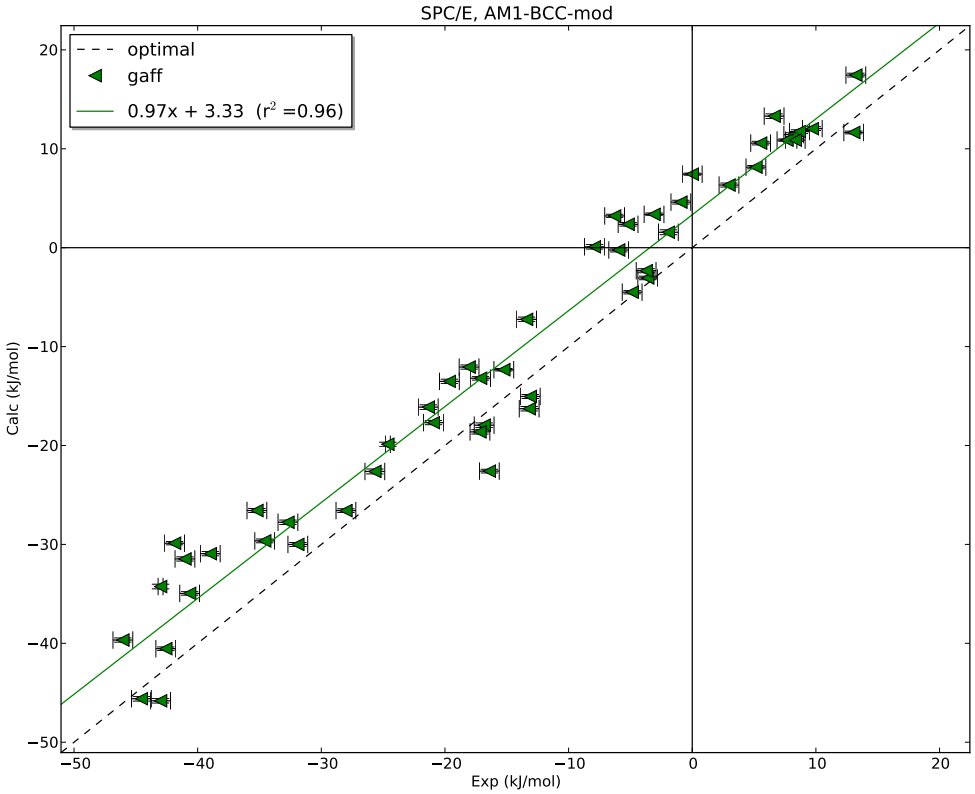


Figure 6: Calculated solvation free energies of 50 compounds in TIP3P-M25 water using three different force fields. Polarization costs are applied to the calculated values. The modified Lennard-Jones parameters of the TIP3P-M25 force field have improved the results for all force fields. The fitted lines of both the GAFF and CGenFF force fields have a slope close to unity, but the accuracy of GAFF is clearly better here.

Table 6: Solvation free energies in TIP3P-M25 water using the CGenFF, GAFF and OPLS-AA force fields. All values are in kJ/mol.

Molecule	**Exp.**	**CGenFF**				**GAFF**				**OPLS-AA**			
	ΔG	ΔG	UE	ΔG_{corr}	UE_{corr}	ΔG	UE	ΔG_{corr}	UE_{corr}	ΔG	UE	ΔG_{corr}	UE_{corr}
1-chlorohexane	0.00	11.77	11.77	16.81	16.81	0.56	0.56	0.62	0.62	−0.04	0.04	0.97	0.97
1-methyl-imidazole	−35.21	−36.64	1.43	−30.53	4.68	−30.56	4.65	−30.39	4.82	−61.55	26.34	−44.81	9.60
1,3-dichloropropane	−7.91	17.44	25.35	24.61	32.52	−6.29	1.62	−6.25	1.66	−7.81	0.10	−7.77	0.14
2-ethoxyethanol	−28.01	−27.98	0.03	−25.01	3.00	−22.46	5.55	−22.34	5.67	−22.83	5.18	−20.67	7.34
2-iodopropane	−1.93	−3.39	1.46	0.10	2.03	−5.29	3.36	−2.93	1.00	−5.41	3.48	3.60	5.53
2-methylpropane	9.71	8.00	1.71	8.08	1.63	9.09	0.62	9.16	0.55	11.09	1.38	11.17	1.46
2-nitropropane	−13.10	−5.49	7.61	3.04	16.14	−12.19	0.91	−11.54	1.56	−14.81	1.71	−14.46	1.36
2,2,2-trifluoroethanol	−18.05	−16.39	1.66	−13.51	4.54	−18.59	0.54	−14.36	3.69	−18.07	0.02	−11.17	6.88
3-acetylpyridine	−34.58	−40.11	5.53	−38.96	4.38	−35.63	1.05	−35.46	0.88	−31.88	2.70	−30.42	4.16
3-methyl-1h-indole	−24.62	−27.73	3.11	−21.68	2.94	−34.51	9.89	−34.38	9.76	−31.65	7.03	−31.29	6.67
4-acetylpyridine	−31.90	−42.79	10.89	−41.00	9.10	−37.69	5.79	−36.55	4.65	−28.74	3.16	−25.13	6.77
4-cyanophenol	−42.58	−47.51	4.93	−41.56	1.02	−42.01	0.57	−40.01	2.57	−39.03	3.55	−26.90	15.68
4-methyl-1h-imidazole	−43.00	−41.23	1.77	−41.17	1.83	−37.91	5.09	−37.54	5.46	−32.56	10.44	−29.77	13.23
4-nitroaniline	−43.00	−41.84	1.16	−33.69	9.31	−47.97	4.97	−45.91	2.91	−36.60	6.40	−14.25	28.75
4-nitrophenol	−44.55	−43.34	1.21	−38.36	6.19	−43.98	0.57	−40.78	3.77	−39.10	5.45	−26.75	17.80
acenaphthene	−13.19	−33.58	20.39	−31.51	18.32	−25.37	12.18	−25.04	11.85	−21.20	8.01	−21.16	7.97
benzaldehyde	−16.83	−18.62	1.79	−17.29	0.46	−25.91	9.08	−24.46	7.63	−19.69	2.86	−14.36	2.47
benzamide	−46.05	−41.25	4.80	−39.23	6.82	−49.28	3.23	−46.35	0.30	−45.89	0.16	−39.75	6.30
benzene	−3.60	−5.22	1.62	−5.22	1.62	−5.94	2.34	−5.94	2.34	−4.46	0.86	−4.46	0.86
bromoethane	−3.09	−5.74	2.65	27.11	30.20	−0.73	2.36	−0.72	2.37	−0.39	2.70	0.73	3.82
butane	8.67	7.21	1.46	7.21	1.46	8.51	0.16	8.51	0.16	7.56	1.11	7.56	1.11
cyanobenzene	−17.17	−27.53	10.36	−26.39	9.22	−19.56	2.39	−19.51	2.34	−16.08	1.09	−12.56	4.61
cyclohexane	5.15	2.33	2.82	2.33	2.82	2.33	2.82	2.33	2.82	3.23	1.92	3.23	1.92
decan-1-ol	−15.24	−20.40	5.16	−19.92	4.68	−18.27	3.03	−17.79	2.55	−19.18	3.94	−18.61	3.37
di-n-propylether	−4.86	−5.37	0.51	−3.26	1.60	−4.27	0.59	−4.00	0.86	−3.04	1.82	−2.21	2.65
ethanamide	−40.65	−38.01	2.64	−36.78	3.87	−39.99	0.66	−37.40	3.25	−37.77	2.88	−36.02	4.63
ethane	7.66	7.72	0.06	7.72	0.06	9.67	2.01	9.67	2.01	8.43	0.77	8.43	0.77
ethanol	−20.93	−20.30	0.63	−16.80	4.13	−16.13	4.80	−14.90	6.03	−19.70	1.23	−15.64	5.29
fluoromethane	−0.92	0.58	1.50	0.80	1.72	4.25	5.17	4.27	5.19	3.08	4.00	5.66	6.58
hex-1-ene	6.62	4.35	2.27	6.80	0.18	7.58	0.96	7.93	1.31	6.27	0.35	7.48	0.86
methane	8.33	10.31	1.98	10.31	1.98	10.71	2.38	10.71	2.38	8.86	0.53	8.86	0.53
methanethiol	−5.19	−2.89	2.30	−1.52	3.67	−2.64	2.55	−0.30	4.89	−4.15	1.04	−2.79	2.40
methanol	−21.35	−19.72	1.63	−14.11	7.24	−15.98	5.37	−14.34	7.01	−18.86	2.49	−13.53	7.82
methyl benzoate	−16.41	−26.80	10.39	−24.08	7.67	−27.90	11.49	−26.10	9.69	−8.76	7.65	−4.46	11.95
methyl ethyl sulfide	−6.28	−1.26	5.02	−0.95	5.33	−1.99	4.29	−1.35	4.93	−4.45	1.83	−2.61	3.67
n-acetylpyrrolidine	−41.03	−33.37	7.66	−33.04	7.99	−37.55	3.48	−35.71	5.32	−23.46	17.57	−20.58	20.45
n-butylacetamide	−38.98	−35.07	3.91	−29.67	9.31	−36.64	2.34	−35.00	3.98	−28.69	10.29	−28.34	10.64
n-decane	13.23	7.58	5.65	7.59	5.64	8.34	4.89	8.34	4.89	7.68	5.55	7.68	5.55
n-methylacetamide	−41.87	−28.83	13.04	−28.78	13.09	−34.73	7.14	−33.55	8.32	−27.13	14.74	−27.00	14.87
n,n-dimethylformamide	−32.70	−26.07	6.63	−18.21	14.49	−32.63	0.07	−29.88	2.82	−21.21	11.49	−18.65	14.05
oct-1-yne	2.97	−1.83	4.80	−1.32	4.29	−3.23	6.20	−3.07	6.04	3.96	0.99	7.87	4.90
octan-1-ol	−17.12	−20.24	3.12	−18.72	1.60	−16.89	0.23	−15.88	1.24	−19.08	1.96	−17.37	0.25
p-cresol	−25.67	−23.86	1.81	−23.12	2.55	−28.68	3.01	−27.87	2.20	−28.17	2.50	−24.97	0.70
propane	8.21	8.05	0.16	8.09	0.12	9.59	1.38	9.63	1.42	8.92	0.71	8.96	0.75
propene	5.53	4.28	1.25	8.23	2.70	8.90	3.37	9.34	3.81	6.58	1.05	8.56	3.03
pyridine	−19.64	−24.79	5.15	−24.51	4.87	−17.66	1.98	−17.19	2.45	−14.95	4.69	−14.88	4.76
tetrafluoromethane	13.06	9.46	3.60	9.46	3.60	10.16	2.90	10.16	2.90	9.86	3.20	9.86	3.20
thiophene	−5.95	−16.24	10.29	−16.13	10.18	−4.03	1.92	−3.87	2.08	−4.69	1.26	−2.10	3.85
toluene	−3.73	−5.44	1.71	−5.03	1.30	−6.74	3.01	−6.65	2.92	−7.25	3.52	−7.15	3.42
trimethylamine	−13.40	−11.62	1.78	−4.76	8.64	−13.27	0.13	−6.99	6.41	−11.50	1.90	−6.05	7.35
			4.60		6.39		3.31		3.73		4.11		6.07

4.4 TIP4P-Ew

The results of the calculations using the TIP4P-Ew water model are presented in Figure 7 and Table 7. The calculated results from CGenFF had an RMSE of 7.86 kJ/mol (mean error 4.81 kJ/mol), whereas GAFF had an RMSE of 5.85 kJ/mol (mean error 4.53 kJ/mol) and OPLS-AA had 8.63 kJ/mol (mean error 5.67 kJ/mol). When including the polarization costs the RMSE increased to 11.53, 6.76 and 11.68 kJ/mol respectively. There are no large differences between the results from TIP4P-Ew and SPC/E. In general this water model seems to have a minor advantage, but more molecules would need to be studied to draw any conclusions.

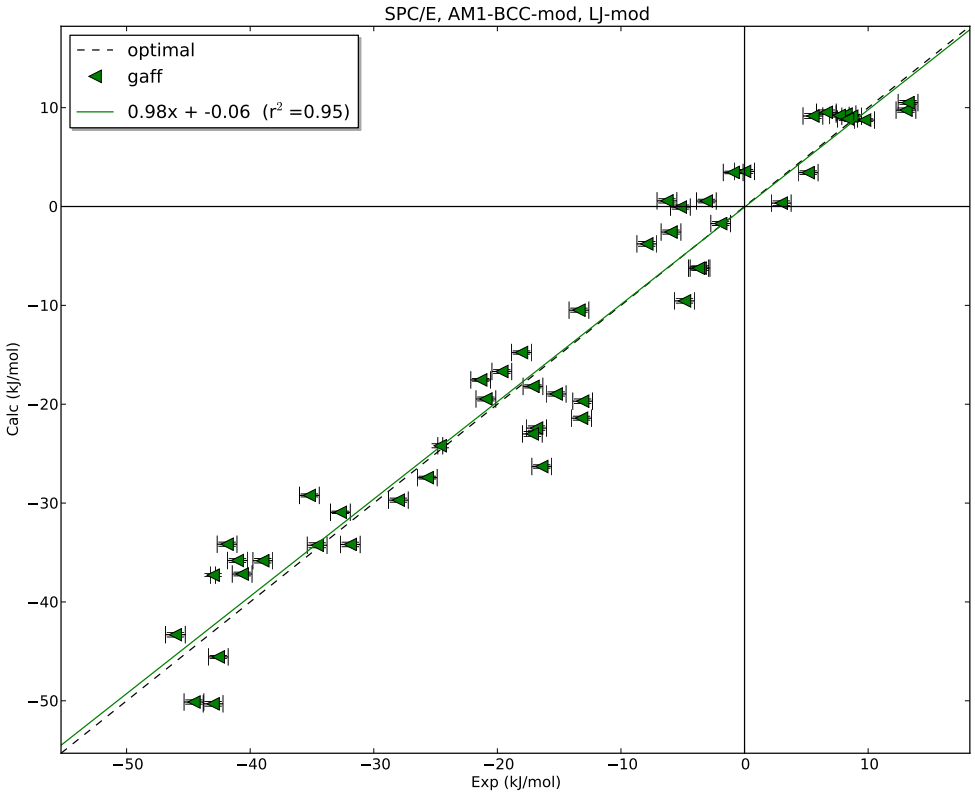


Figure 7: Calculated solvation free energies of 50 compounds in TIP4P-Ew water using three different force fields. Polarization costs are applied to the calculated values. The results are similar to SPC/E (Figure 4), but it seems like the results are slightly improved in TIP4P-Ew.

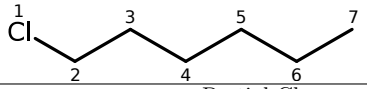
Table 7: Solvation free energies in TIP4P-Ew water using the CGenFF, GAFF and OPLS-AA force fields. All values are in kJ/mol.

Molecule	**Exp.**	**CGenFF**				**GAFF**				**OPLS-AA**			
	ΔG	ΔG	UE	ΔG_{corr}	UE_{corr}	ΔG	UE	ΔG_{corr}	UE_{corr}	ΔG	UE	ΔG_{corr}	UE_{corr}
1-chlorohexane	0.00	17.22	17.22	22.26	22.26	7.45	7.45	7.55	7.55	7.46	7.46	8.47	8.47
1-methyl-imidazole	−35.21	−32.17	3.04	−26.06	9.15	−26.31	8.90	−26.14	9.07	−58.86	23.65	−42.12	6.91
1,3-dichloropropane	−7.91	19.00	26.91	26.17	34.08	−0.62	7.29	−0.58	7.33	−2.04	5.87	−2.00	5.91
2-ethoxyethanol	−28.01	−24.27	3.74	−21.30	6.71	−18.44	9.57	−18.32	9.69	−20.22	7.79	−18.06	9.95
2-iodopropane	−1.93	2.50	4.43	5.99	7.92	−1.49	0.44	0.87	2.80	−0.32	1.61	8.69	10.62
2-methylpropane	9.71	11.97	2.26	12.05	2.34	11.40	1.69	11.47	1.76	11.06	1.35	11.14	1.43
2-nitropropane	−13.10	−0.44	12.66	8.09	21.19	−5.89	7.21	−5.24	7.86	−10.02	3.08	−9.67	3.43
2,2,2-trifluoroethanol	−18.05	−12.79	5.26	−9.91	8.14	−14.45	3.60	−10.22	7.83	−15.27	2.78	−8.37	9.68
3-acetylpyridine	−34.58	−32.52	2.06	−31.37	3.21	−30.64	3.94	−30.47	4.11	−26.35	8.23	−24.89	9.69
3-methyl-1h-indole	−24.62	−18.19	6.43	−12.14	12.48	−27.15	2.53	−27.02	2.40	−22.89	1.73	−22.53	2.09
4-acetylpyridine	−31.90	−35.47	3.57	−33.68	1.78	−31.66	0.24	−30.52	1.38	−23.15	8.75	−19.54	12.36
4-cyanophenol	−42.58	−36.20	6.38	−30.25	12.33	−31.77	10.81	−29.77	12.81	−30.00	12.58	−17.87	24.71
4-methyl-1h-imidazole	−43.00	−36.55	6.45	−36.49	6.51	−34.52	8.48	−34.15	8.85	−27.87	15.13	−25.08	17.92
4-nitroaniline	−43.00	−32.41	10.59	−24.26	18.74	−36.22	6.78	−34.16	8.84	−29.23	13.77	−6.88	36.12
4-nitrophenol	−44.55	−32.76	11.79	−27.78	16.77	−31.37	13.18	−28.17	16.38	−30.27	14.28	−17.92	26.63
acenaphthene	−13.19	−23.34	10.15	−21.27	8.08	−16.27	3.08	−15.94	2.75	−11.17	2.02	−11.13	2.06
benzaldehyde	−16.83	−13.28	3.55	−11.95	4.88	−19.15	2.32	−17.70	0.87	−12.15	4.68	−6.82	10.01
benzamide	−46.05	−33.26	12.79	−31.24	14.81	−43.19	2.86	−40.26	5.79	−36.66	9.39	−30.52	15.53
benzene	−3.60	−0.33	3.27	−0.33	3.27	−2.19	1.41	−2.19	1.41	−0.31	3.29	−0.31	3.29
bromoethane	−3.09	−1.34	1.75	31.51	34.60	3.18	6.27	3.19	6.28	−0.68	2.41	0.44	3.53
butane	8.67	11.11	2.44	11.11	2.44	11.90	3.23	11.90	3.23	11.46	2.79	11.46	2.79
cyanobenzene	−17.17	−17.87	0.70	−16.73	0.44	−11.67	5.50	−11.62	5.55	−8.92	8.25	−5.40	11.77
cyclohexane	5.15	8.44	3.29	8.44	3.29	7.84	2.69	7.84	2.69	7.58	2.43	7.58	2.43
decan-1-ol	−15.24	−11.92	3.32	−11.44	3.80	−8.31	6.93	−7.85	7.39	−11.35	3.89	−10.78	4.46
di-n-propylether	−4.86	−2.09	2.77	0.02	4.88	0.91	5.77	1.18	6.04	0.41	5.27	1.24	6.10
ethanamide	−40.65	−35.40	5.25	−34.17	6.48	−37.29	3.36	−34.70	5.95	−34.50	6.15	−32.75	7.90
ethane	7.66	9.57	1.91	9.57	1.91	11.15	3.49	11.15	3.49	10.21	2.55	10.21	2.55
ethanol	−20.93	−19.19	1.74	−15.69	5.24	−14.34	6.59	−13.11	7.82	−18.81	2.12	−14.75	6.18
fluoromethane	−0.92	1.99	2.91	2.21	3.13	4.57	5.49	4.59	5.51	4.99	5.91	7.57	8.49
hex-1-ene	6.62	9.67	3.05	12.12	5.50	13.12	6.50	13.47	6.85	11.66	5.04	12.87	6.25
methane	8.33	10.45	2.12	10.45	2.12	10.76	2.43	10.76	2.43	9.38	1.05	9.38	1.05
methanethiol	−5.19	−0.06	5.13	1.31	6.50	0.51	5.70	2.85	8.04	−0.61	4.58	0.75	5.94
methanol	−21.35	−19.51	1.84	−13.90	7.45	−13.91	7.44	−12.27	9.08	−18.02	3.33	−12.69	8.66
methyl benzoate	−16.41	−22.55	6.14	−19.83	3.42	−20.46	4.05	−18.66	2.25	−9.50	6.91	−5.20	11.21
methyl ethyl sulfide	−6.28	5.05	11.33	5.36	11.64	3.19	9.47	3.83	10.11	−0.18	6.10	1.66	7.94
n-acetylpyrrolidine	−41.03	−27.87	13.16	−27.54	13.49	−34.77	6.26	−32.93	8.10	−17.55	23.48	−14.67	26.36
n-butylacetamide	−38.98	−29.44	9.54	−24.04	14.94	−32.90	6.08	−31.26	7.72	−24.06	14.92	−23.71	15.27
n-decane	13.23	17.53	4.30	17.54	4.31	16.47	3.24	16.47	3.24	17.72	4.49	17.72	4.49
n-methylacetamide	−41.87	−25.02	16.85	−24.97	16.90	−31.48	10.39	−30.30	11.57	−23.55	18.32	−23.42	18.45
n,n-dimethylformamide	−32.70	−21.90	10.80	−14.04	18.66	−30.61	2.09	−27.86	4.84	−17.79	14.91	−15.23	17.47
oct-1-yne	2.97	6.35	3.38	6.86	3.89	5.29	2.32	5.45	2.48	10.67	7.70	14.58	11.61
octan-1-ol	−17.12	−12.39	4.73	−10.87	6.25	−9.80	7.32	−8.62	8.50	−13.08	4.04	−11.37	5.75
p-cresol	−25.67	−16.44	9.23	−15.70	9.97	−19.05	6.62	−18.24	7.43	−21.40	4.27	−18.20	7.47
propane	8.21	10.45	2.24	10.49	2.28	11.00	2.79	11.04	2.83	11.52	3.31	11.56	3.35
propene	5.53	6.76	1.23	10.71	5.18	10.14	4.61	10.58	5.05	9.00	3.47	10.98	5.45
pyridine	−19.64	−20.39	0.75	−20.11	0.47	−14.69	4.95	−14.22	5.42	−11.24	8.40	−11.17	8.47
tetrafluoromethane	13.06	10.65	2.41	10.65	2.41	11.51	1.55	11.51	1.55	11.44	1.62	11.44	1.62
thiophene	−5.95	−10.21	4.26	−10.10	4.15	0.43	6.38	0.59	6.54	−0.31	5.64	2.28	8.23
toluene	−3.73	1.05	4.78	1.46	5.19	−1.67	2.06	−1.58	2.15	−2.60	1.13	−2.50	1.23
trimethylamine	−13.40	−12.84	0.56	−5.98	7.42	−13.11	0.29	−6.83	6.57	−11.43	1.97	−5.98	7.42
			5.93		8.66		5.07		5.92		6.68		9.13

4.5 Outliers

In order to improve topology generation and force fields it is important to understand why certain molecules have a large error in the calculated hydration free energy, especially if only one force field suffers from it. We have limited this analysis to the SPC/E water model and summarize some of the molecules with relatively high mean unsigned error in Tables 8 through 16. Note that the CGenFF parameters were assigned using MATCH[30] and that there might be differences to parameters assigned using the CGenFF program[32,33].

Table 8: Parameters of 1-chlorohexane in three different force fields, and the B3LYP/PCM charge model used with GAFF, taking polarization costs into account. Only heavy atoms are presented. The solvation free energy of 1-chlorohexane in SPC/E has a large error in CGenFF ($\Delta\Delta$G=21.87 kJ/mol), whereas it is lower in GAFF ($\Delta\Delta$G=8.75 kJ/mol), OPLS-AA ($\Delta\Delta$G=7.47 kJ/mol) and GAFF with B3LYP/PCM ($\Delta\Delta$G=3.74 kJ/mol). The main difference between CGenFF and the other force fields is the van der Waals ϵ of the chlorine, but the low partial charge of that atom might also contribute to the high hydration free energy.

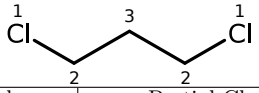


Force field	Partial Charge						
	1	2	3	4	5	6	7
CGenFF	**−0.04**	−0.24	−0.18	−0.18	−0.18	−0.18	−0.27
GAFF	−0.20	0.04	−0.09	−0.08	−0.08	−0.08	−0.09
- B3LYP/PCM	−0.24	−0.18	0.18	0.02	0.00	0.22	−0.24
OPLS-AA	−0.20	−0.01	−0.12	−0.12	−0.12	−0.12	−0.18
	van der Waals σ (nm)						
	1	2	3	4	5	6	7
CGenFF	0.34	0.36	0.36	0.36	0.36	0.36	0.36
GAFF	0.35	0.34	0.34	0.34	0.34	0.34	0.34
OPLS-AA	0.34	0.35	0.35	0.35	0.35	0.35	0.35
	van der Waals ϵ (kJ/mol)						
	1	2	3	4	5	6	7
CGenFF	**0.13**	0.23	0.23	0.23	0.23	0.23	0.23
GAFF	1.11	0.46	0.46	0.46	0.46	0.46	0.46
OPLS-AA	1.26	0.28	0.28	0.28	0.28	0.28	0.28

4.6 Force Field Parameter Modifications

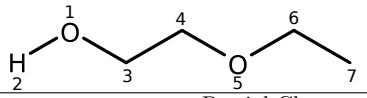
Automated free energy calculations for entire sets of compounds make it easier to apply systematic changes in order to improve the parameterization, sometimes with quite modest means. As an example of such an attempt we focused on the data obtained for the SPC/E

Table 9: Parameters of 1,3-dichloropropane in three different force fields, and the B3LYP/PCM charge model used with GAFF, taking polarization costs into account. Only heavy atoms are presented. The solvation free energy of 1,3-dichloropropane in SPC/E has a large error in CGenFF ($\Delta\Delta$G=34.61 kJ/mol), whereas it is lower in GAFF ($\Delta\Delta$G=7.98 kJ/mol), GAFF with B3LYP/PCM ($\Delta\Delta$G=6.87 kJ/mol) and OPLS-AA ($\Delta\Delta$G=6.12 kJ/mol). The problem with chlorine parameters of CGenFF as identified for 1-chlorohexane (Table 8) applies to this molecule too.



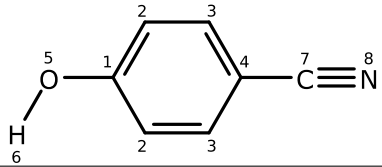
Force field	Partial Charge		
	1	2	3
CGenFF	**−0.04**	−0.24	−0.18
GAFF	−0.20	0.03	−0.10
- B3LYP/PCM	−0.21	−0.21	0.07
OPLS-AA	−0.20	−0.01	−0.12
	van der Waals σ (nm)		
	1	2	3
CGenFF	0.34	0.36	0.36
GAFF	0.35	0.34	0.34
OPLS-AA	0.34	0.35	0.35
	van der Waals ϵ (kJ/mol)		
	1	2	3
CGenFF	**0.13**	0.23	0.23
GAFF	1.11	0.46	0.46
OPLS-AA	1.26	0.28	0.28

Table 10: Parameters of 2-ethoxyethanol in three different force fields, and the B3LYP/PCM charge model used with GAFF, taking polarization costs into account. Only heavy atoms are presented. The solvation free energy of 2-ethoxyethanol in SPC/E has a large error in GAFF with B3LYP/PCM ($\Delta\Delta$G=11.18 kJ/mol), OPLS-AA ($\Delta\Delta$G=10.22 kJ/mol) and GAFF ($\Delta\Delta$G=9.65 kJ/mol), whereas it is lower in CGenFF ($\Delta\Delta$G=6.60 kJ/mol). Especially B3LYP/PCM, but also AM1-BCC (GAFF) and OPLS-AA partial charges of atoms 3 and 6 appear to be too high.



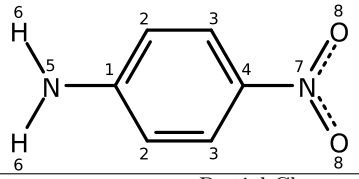
Force field	Partial Charge						
	1	2	3	4	5	6	7
CGenFF	−0.65	0.42	0.05	−0.01	−0.34	−0.01	−0.27
GAFF	−0.60	0.41	0.13	0.09	−0.44	0.13	−0.10
- B3LYP/PCM	−0.64	0.39	**0.28**	**−0.11**	−0.37	**0.25**	−0.26
OPLS-AA	−0.68	0.42	0.15	0.14	−0.40	0.14	−0.18
	van der Waals σ (nm)						
	1	2	3	4	5	6	7
CGenFF	0.31	0.04	0.36	0.36	0.29	0.36	0.37
GAFF	0.31	0.00	0.34	0.34	0.30	0.34	0.34
OPLS-AA	0.31	0.00	0.35	0.35	0.29	0.35	0.35
	van der Waals ϵ (kJ/mol)						
	1	2	3	4	5	6	7
CGenFF	0.88	0.19	0.23	0.23	0.42	0.23	0.33
GAFF	0.88	0.00	0.46	0.46	0.71	0.46	0.46
OPLS-AA	0.71	0.00	0.28	0.28	0.59	0.28	0.28

Table 11: Parameters of 4-cyanophenol in three different force fields, and also the B3LYP/PCM charge model used with GAFF, taking polarization costs into account. Only heavy atoms are presented. The solvation free energy of 4-cyanophenol in SPC/E has a large error in OPLS-AA ($\Delta\Delta$G=23.80 kJ/mol), GAFF ($\Delta\Delta$G=12.46 kJ/mol) and CGenFF ($\Delta\Delta$G=11.87 kJ/mol), whereas it is lower in GAFF with B3LYP/PCM ($\Delta\Delta$G=-1.96 kJ/mol). The charge on the nitrogen in the cyano group (atom 8) is more negative in B3LYP/PCM and the charge difference between the carbon and nitrogen in the cyanogroup (atoms 7 and 8) is smaller in OPLS-AA compared to the other charge models. In B3LYP/PCM atom 1 has a higher charge than in the other charge models.



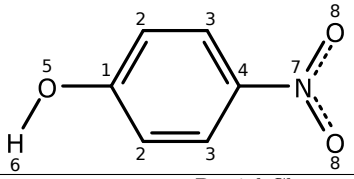
Force field	Partial Charge							
	1	2	3	4	5	6	7	8
CGenFF	0.11	−0.12	−0.12	0.10	−0.53	0.42	0.36	−0.46
GAFF	0.15	−0.19	−0.05	−0.06	−0.49	0.43	0.24	−0.37
- B3LYP/PCM	**0.39**	−0.26	−0.12	−0.04	−0.53	0.41	0.39	−0.56
OPLS-AA	0.15	−0.12	−0.12	0.04	−0.59	0.44	**0.40**	**−0.43**
	van der Waals σ (nm)							
	1	2	3	4	5	6	7	8
CGenFF	0.36	0.36	0.36	0.36	0.31	0.04	0.31	0.33
GAFF	0.34	0.34	0.34	0.34	0.31	0.00	0.34	0.33
OPLS-AA	0.36	0.36	0.36	0.36	0.31	0.00	0.37	0.32
	van der Waals ϵ (kJ/mol)							
	1	2	3	4	5	6	7	8
CGenFF	0.29	0.29	0.29	0.29	0.80	0.19	0.84	**2.51**
GAFF	0.36	0.36	0.36	0.36	0.88	0.00	0.88	0.71
OPLS-AA	0.29	0.29	0.29	0.29	0.71	0.00	0.64	0.71

Table 12: Parameters of 4-nitroaniline in three different force fields, and also the B3LYP/PCM charge model used with GAFF, taking polarization costs into account. Only heavy atoms are presented. The solvation free energy of 4-nitroaniline in SPC/E has a large error in OPLS-AA ($\Delta\Delta$G=36.62 kJ/mol) and CGenFF ($\Delta\Delta$G=19.91 kJ/mol), whereas it is lower in GAFF ($\Delta\Delta$G=9.31 kJ/mol) and GAFF with B3LYP/PCM ($\Delta\Delta$G=-4.06 kJ/mol). OPLS-AA and CGenFF might have slightly too low van der Waals ϵ for the nitrogen and oxygen atoms (atoms 7 and 8) in the nitro group, respectively. The oxygen atoms in the nitro group (atom 8) in B3LYP/PCM have a more negative charge and the charge difference between the nitrogen and oxygen atoms in the nitro group is more pronounced in B3LYP/PCM. Similarly to 4-cyanophenol (Table 11), atom 1 has a higher positive charge in B3LYP/PCM than in the other charge models. In OPLS-AA the charge on the amino nitrogen seems to be too negative.



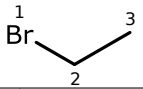
Force field	Partial Charge							
	1	2	3	4	5	6	7	8
CGenFF	0.07	−0.12	−0.18	0.32	−0.83	0.38	0.40	−0.34
GAFF	0.22	−0.22	−0.02	−0.24	−0.85	0.42	0.32	−0.22
- B3LYP/PCM	**0.41**	−0.24	−0.18	0.05	−0.79	0.39	0.69	**−0.49**
OPLS-AA	0.18	−0.12	−0.12	0.09	**−0.90**	0.36	0.65	−0.37
	van der Waals σ (nm)							
	1	2	3	4	5	6	7	8
CGenFF	0.36	0.36	0.36	0.36	0.33	0.04	0.33	0.30
GAFF	0.34	0.34	0.34	0.34	0.33	0.11	0.33	0.30
OPLS-AA	0.36	0.36	0.36	0.36	0.33	0.00	0.33	0.30
	van der Waals ϵ (kJ/mol)							
	1	2	3	4	5	6	7	8
CGenFF	0.29	0.29	0.29	0.29	0.84	0.19	0.84	**0.50**
GAFF	0.36	0.36	0.36	0.36	0.71	0.07	0.71	0.88
OPLS-AA	0.29	0.29	0.29	0.29	0.71	0.00	**0.50**	0.71

Table 13: Parameters of 4-nitrophenol in three different force fields, and also the B3LYP/PCM charge model used with GAFF, taking polarization costs into account. Only heavy atoms are presented. The solvation free energy of 4-nitrophenol in SPC/E has a large error in OPLS-AA ($\Delta\Delta$G=27.17 kJ/mol), CGenFF ($\Delta\Delta$G=17.24 kJ/mol) and GAFF ($\Delta\Delta$G=15.95 kJ/mol), whereas it is lower in GAFF with B3LYP/PCM ($\Delta\Delta$G=-0.21 kJ/mol). The probable culprits in the parameters in this case are already covered by 4-cyanophenol (Table 11) and 4-nitroaniline (Table 12).



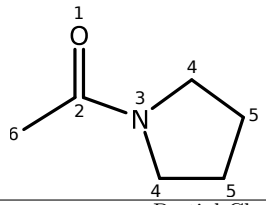
Force field	Partial Charge							
	1	2	3	4	5	6	7	8
CGenFF	0.11	−0.12	−0.18	**0.32**	−0.53	0.42	**0.40**	−0.34
GAFF	0.17	−0.20	−0.03	**−0.22**	−0.48	0.43	**0.32**	−0.21
- B3LYP/PCM	**0.37**	−0.23	−0.16	0.03	−0.51	0.42	0.72	−0.47
OPLS-AA	0.15	−0.12	−0.12	0.09	−0.59	0.44	0.65	−0.37
	van der Waals σ (nm)							
	1	2	3	4	5	6	7	8
CGenFF	0.36	0.36	0.36	0.36	0.31	0.04	0.33	0.30
GAFF	0.34	0.34	0.34	0.34	0.31	0.00	0.33	0.30
OPLS-AA	0.36	0.36	0.36	0.36	0.31	0.00	0.33	0.30
	van der Waals ϵ (kJ/mol)							
	1	2	3	4	5	6	7	8
CGenFF	0.29	0.29	0.29	0.29	0.80	0.19	0.84	**0.50**
GAFF	0.36	0.36	0.36	0.36	0.88	0.00	0.71	0.88
OPLS-AA	0.29	0.29	0.29	0.29	0.71	0.00	**0.50**	0.71

Table 14: Parameters of bromoethane in three different force fields, and also the B3LYP/PCM charge model used with GAFF, taking polarization costs into account. Only heavy atoms are presented. The solvation free energy of bromoethane in SPC/E has a large error in CGenFF ($\Delta\Delta$G=34.28 kJ/mol), whereas it is lower in OPLS-AA ($\Delta\Delta$G=-8.54 kJ/mol), GAFF ($\Delta\Delta$G=6.17 kJ/mol) and GAFF with B3LYP/PCM ($\Delta\Delta$G=4.58 kJ/mol). In CGenFF the bromine atom appears too have too high partial charge and atom 3 seems to have too low partial charge. This causes a drastic increase in the dipole moment, which is the cause of the bad agreement with experimental data. The too low hydration free energy in OPLS-AA might be explained by the high van der Waals ϵ of the bromine atom.



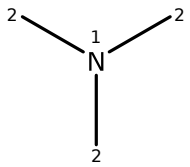
Force field	Partial Charge		
	1	2	3
CGenFF	**−0.10**	0.07	**−0.52**
GAFF	−0.18	−0.01	−0.10
- B3LYP/PCM	−0.25	−0.09	−0.05
OPLS-AA	−0.22	−0.01	−0.18
	van der Waals σ (nm)		
	1	2	3
CGenFF	0.37	0.36	0.37
GAFF	0.36	0.34	0.34
OPLS-AA	0.35	0.35	0.35
	van der Waals ϵ (kJ/mol)		
	1	2	3
CGenFF	1.76	0.23	0.33
GAFF	1.76	0.46	0.46
OPLS-AA	**1.97**	0.28	0.28

Table 15: Parameters of n-acetylpyrrolidine in three different force fields, and also the B3LYP/PCM charge model used with GAFF, taking polarization costs into account. Only heavy atoms are presented. The solvation free energy of n-acetylpyrrolidine in SPC/E has a large error in OPLS-AA ($\Delta\Delta$G=26.07 kJ/mol), CGenFF ($\Delta\Delta$G=14.05 kJ/mol) and GAFF ($\Delta\Delta$G=10.99 kJ/mol), whereas it is lower in GAFF with B3LYP/PCM ($\Delta\Delta$G=-3.61 kJ/mol). The most important difference between the charge models seems to be the larger magnitude in partial charge on atoms 1 and 6 in B3LYP/PCM.



Force field	Partial Charge					
	1	2	3	4	5	6
CGenFF	−0.51	0.51	−0.36	0.00	−0.18	−0.27
GAFF	−0.61	0.66	−0.47	0.09	−0.09	−0.17
- B3LYP/PCM	**−0.67**	0.56	−0.12	−0.02	−0.03	**−0.36**
OPLS-AA	−0.50	0.50	−0.14	−0.05	−0.12	−0.18
	van der Waals σ (nm)					
	1	2	3	4	5	6
CGenFF	0.30	0.36	0.33	0.36	0.36	0.37
GAFF	0.30	0.34	0.33	0.34	0.34	0.34
OPLS-AA	0.30	0.38	0.33	0.35	0.35	0.35
	van der Waals ϵ (kJ/mol)					
	1	2	3	4	5	6
CGenFF	0.50	0.46	0.84	0.25	0.25	0.33
GAFF	0.88	0.36	0.71	0.46	0.46	0.46
OPLS-AA	0.88	0.44	0.71	0.28	0.28	0.28

Table 16: Parameters of trimethylamine in three different force fields, and also the B3LYP/PCM charge model used with GAFF, taking polarization costs into account. Only heavy atoms are presented. The solvation free energy of trimethylamine in SPC/E has a large error in GAFF with B3LYP/PCM ($\Delta\Delta$G=16.49 kJ/mol), whereas it is slightly lower in OPLS-AA ($\Delta\Delta$G=7.85 kJ/mol), CGenFF ($\Delta\Delta$G=7.76 kJ/mol) and GAFF ($\Delta\Delta$G=6.45 kJ/mol). It is apparent that the partial charge on the nitrogen (atom 1) is not negative enough in B3LYP/PCM.



Force field	Partial Charge	
	1	2
CGenFF	−0.63	−0.06
GAFF	−0.74	0.16
- B3LYP/PCM	**−0.23**	**−0.22**
OPLS-AA	−0.63	0.03
	van der Waals σ (nm)	
	1	2
CGenFF	0.36	0.35
GAFF	0.33	0.34
OPLS-AA	0.33	0.35
	van der Waals ϵ (kJ/mol)	
	1	2
CGenFF	0.15	0.29
GAFF	0.71	0.46
OPLS-AA	0.71	0.28

water model above. Since GAFF was the force field that agreed best with experimental data that was used as the starting point. The B3LYP/PCM charge model outperformed AM1-BCC, but it is not as computationally efficient, so the bond charge corrections for three functional groups were modified (AM1-BCC-mod) to more closely resemble the B3LYP/PCM charges. See Table 17 for more details on the changes. Using these modified charges the RMSE dropped from 6.89 to 5.27 kJ/mol and the average error from 5.70 to 3.79 kJ/mol (see Figure 8).

Table 17: Modifications to bond charge corrections used in AM1-BCC. Atom 1, atom 2 and bond order correspond to the BCC atom types and bond orders[36]. The BCC column lists the correction used in AM1-BCC and BCC-mod the modified parameters (used in the AM1-BCC-mod charge model).

Atom 1	Atom 2	Bond order	Examples	BCC	BCC-mod
11	31	1	Alcohol, ether	0.0718	0.1218
15	16	1	Cyanobenzene	0.0040	−0.0200
16	23	1	Nitrobenzene	−0.0452	0.0552
31	91	1	Alcohol	−0.2010	−0.2210
15	25	3	Cyano	0.3258	0.4300
23	31	9	Nitro	−0.1500	0.0300

The Lennard-Jones interactions were scaled in a fashion similar to how TIP3P-MOD was developed.[8] At first, approximately the same factors were used as for TIP3P-MOD, but applied to all non-water atoms instead of water, i.e., f_σ=0.99 and f_ϵ=1.25. However, this made the solvation free energies too negative, leading to a new iteration with f_σ=0.993 and f_ϵ=1.17. With the combination of the modified bond charge corrections and the LJ parameters the RMSE was reduced to 4.15 kJ/mol and the average error 0.17 kJ/mol (see Figure 9). While there is certainly still room for improvement, this brings the accuracy of calculated free energies of solvation below a single kcal/mol even for fully automated parameterization of randomly selected small compounds - a considerably harder challenge than amino acids.

The simulation time (per molecule) for acquiring the reported hydration free energies in SPC/E water was typically in the range of 3 to 20 core-hours, depending on the input

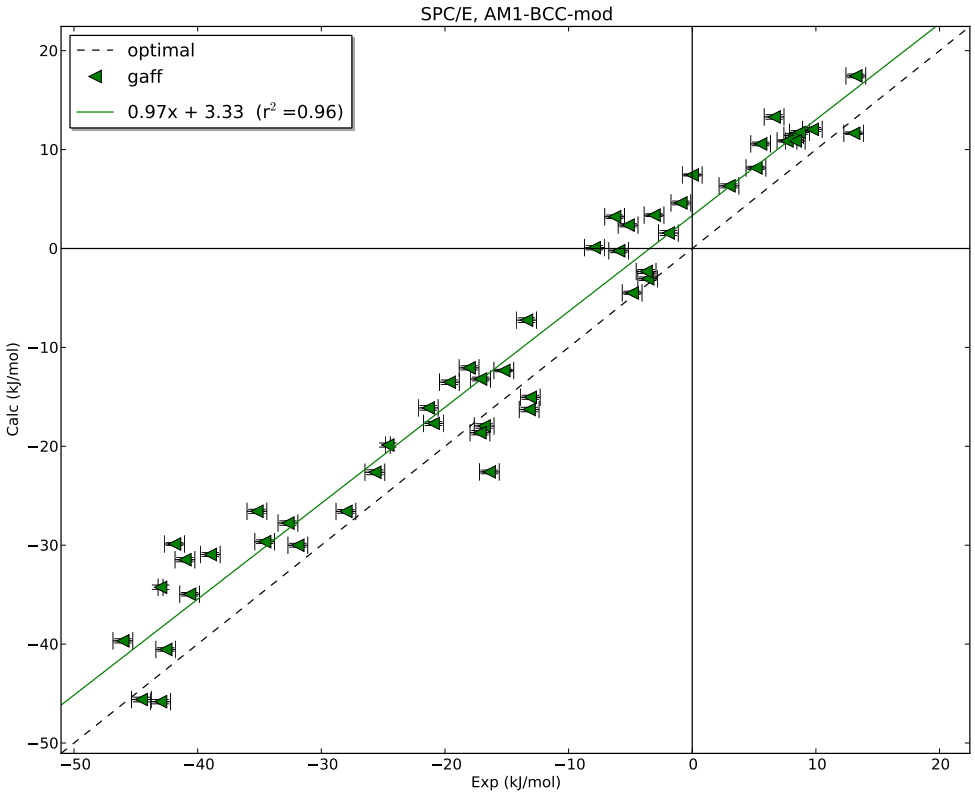


Figure 8: Calculated solvation free energies of 50 compounds in SPC/E water using the GAFF force field and the modified AM1-BCC charge model. Polarization costs are applied to the calculated values. As can be seen from the plot the slope is good, but the calculated hydration free energies are systematically too high.

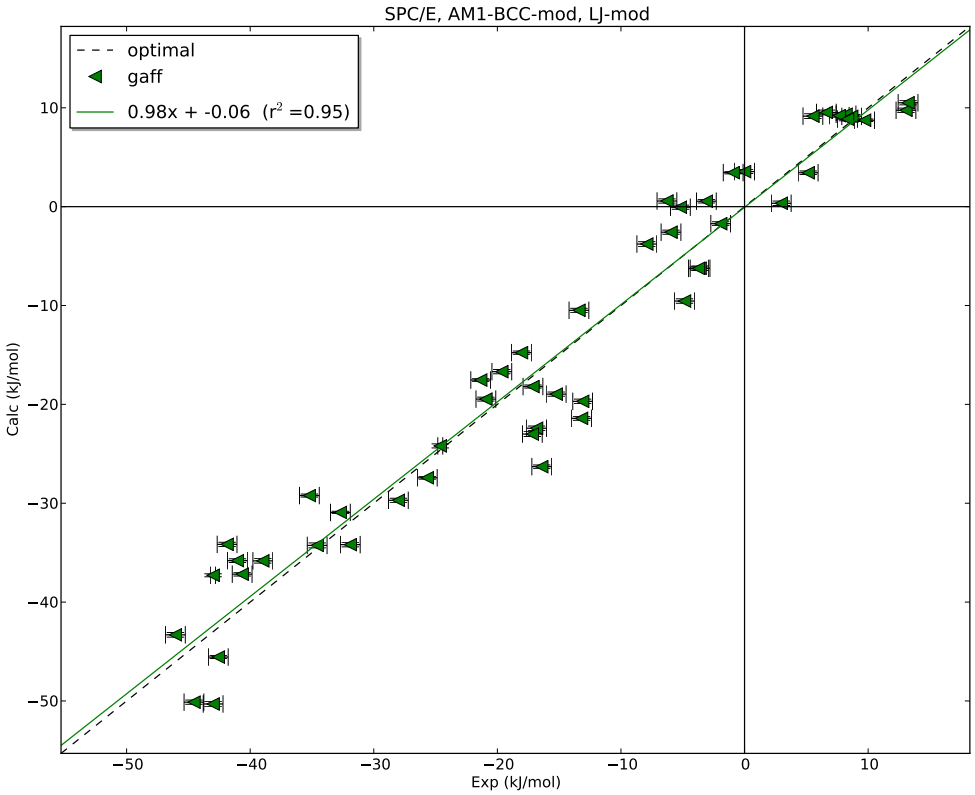


Figure 9: Calculated solvation free energies of 50 compounds in SPC/E water using the GAFF force field with modified Lennard-Jones interactions and the modified AM1-BCC charge model. Polarization costs are applied to the calculated values. With the scaled Lennard-Jones parameters the agreement with experimental values gets very good.

molecule. One molecule used as much as 64 core-hours. Since the simulations run in parallel the wallclock time is often less than a single hour per compound, which makes it straightforward to use e.g. cloud resources rather than supercomputers for this type of calculations. After this study it was noted that the simulation times in TIP4P-Ew was generally no slower than SPC/E, which could make that water model a better choice.

5 Conclusions

STaGE can be used to generate GROMACS topologies for multiple force fields using common molecular file formats as input. It can generate partial charges using a number of different charge models and also provides basic functionality for scaling or adjusting force field parameters, if required. There is no automatic parameter calibration, but it would be easy to implement a scheme to improve e.g. solvation free energies by modifying the van der Waals parameters, in approaches similar to those used by Nerenberg *et al.*[14] and Cerutti *et al.*[15] (the latter work first calibrated the partial charges). The generated system can be solvated and/or combined with previously generated macromolecular topologies. Most operations done by STaGE depend on external tools, all of which are freely available for academic research and all important progroms are also free for commercial use.

While STaGE will continue to evolve as a program (in particular with new functionality and force fields), it is fully ready for production use and an important addition to the GROMACS free energy calculation pipeline - small molecule topologies no longer require deep expertise in force field atom type selection, experience of quantum mechanics chemistry (QM) for partial charge calculation, or manual topology assembly in a text editor. Similarly, the fully automated optimization of free energy calculations, execution of dozens of independent simulations and BAR analysis made possible with Copernicus (e.g. in the cloud) means free energy calculations are more accessible than ever.

At the time of writing (spring 2014) the spot price for one core hour at a major cloud

vendor was approximately $0.01, meaning the total cost for calculating the hydration free energies of these 50 compounds for one of the figures would be approximately $10, based on the estimated calculation time per compound. This is an interesting alternative to maintaining hardware, and it emphasizes that free energy MD simulations do not necessarily require major hardware investments.

This should be useful for many applications, but one of the most important aspects is that it enables systematic critical assessment and comparisons both of force fields and methods to perform free energy calculations. There are huge efforts behind all modern force fields, and it is remarkable how much they have improved the last two decades, but they only way to further improve free energies is to find discrepancies and shortcomings.

The solvation free energy calculations of the 50 compounds included in this study show that all tested force fields reproduce the experimental results fairly well, but there is certainly room for improvements, with a mean unsigned error under 1.5 kcal/mol in almost all cases (except for the combination of OPLS-AA with SPC/E or TIP4P-Ew). Unfortunately, the force fields give worse results with the more correct water models, SPC/E and TIP4P-Ew, for which they have not been parameterized. When accounting for polarization costs it is clear that the charge models recommended for use with the three force fields employed in this study are underpolarized, since the errors increase when correctly applying the polarization cost. A slightly better agreement with experimental values can be achieved by using QM-based partial charges (B3LYP/PCM). This indicates the force fields might profit from being reparameterized taking polarization costs into account.

By modifying the AM1-BCC bond charge corrections for a handful of groups to better resemble B3LYP/PCM charges, and slight modifications of the GAFF LJ parameters, it was possible to achieve a clear improvement of the solvation free energies for the present test set - the final setup has a RMSE of less than a kcal/mol for a diverse set of arbitrary compounds with both topology generation and free energy calculations being fully automated.

This might be a starting point for re-parameterizing force fields to properly take polar-

ization costs into account. It is important to keep in mind that the modifications herein have just been a proof of concept that small changes can make a large difference for the solvation free energies. The parameters need to be verified for other properties and further modifications for other functional groups would certainly be good, but that will be covered in a future publication.

Importantly, we do not suggest using one force field over any other based on this limited study. Many things need to be taken into account when selecting a force field, for instance whether the small molecule should be used as part of a larger system that has already been simulated with one of the force fields. Ultimately, STaGE leaves the force field decision to the user, and we hope it will lead to more direct comparisions even for complex systems. STaGE is open source and freely available from the GROMACS website (http://www.gromacs.org).

Acknowledgement

Contract grant sponsor: EU FP7 project CRESTA; Contract grant number: 287703. Contract grant sponsor: Swedish Foundation for International Cooperation in Research and Higher Education; Contract grant number: IG2011-2072. Contract grant sponsor: Swedish e-Science Research Center; Contract grant sponsor: Swedish research council; Contract grant number: 2010-491, 2013-5901. Computational resources were provided by the Swedish National Infrastructure for Computing (2013/26-24).

References

(1) Guthrie, J. P. *J. Phys. Chem. B* **2009**, *113*, 4501–4507.

(2) Geballe, M. T.; Guthrie, J. P. *J. Comput. Aided Mol. Des.* **2012**, *26*, 489–496.

(3) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. *J. Chem. Theory Comput.* **2009**, *5*, 350–358.

(4) Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. *J. Med. Chem.* **2008**, *51*, 769–779.

(5) Geballe, M. T.; Skillman, A. G.; Nicholls, A.; Guthrie, J. P.; Taylor, P. J. *J. Comput. Aided Mol. Des.* **2010**, *24*, 259–279.

(6) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Dill, K. A. *J. Phys. Chem. B* **2009**, *113*, 4533–4537.

(7) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508–134513.

(8) Sun, Y.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1164–1169.

(9) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665–9678.

(10) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(11) Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2006**, *110*, 17616–17626.

(12) Swope, W. C.; Horn, H. W.; Rice, J. E. *J. Phys. Chem. B* **2010**, *114*, 8621–8630.

(13) Swope, W. C.; Horn, H. W.; Rice, J. E. *J. Phys. Chem. B* **2010**, *114*, 8631–8645.

(14) Nerenberg, P. S.; Jo, B.; So, C.; Tripathy, A.; Head-Gordon, T. *J. Phys. Chem. B* **2012**, *116*, 4524–4534.

(15) Cerutti, D. S.; Rice, J. E.; Swope, W. C.; Case, D. A. *J. Phys. Chem. B* **2013**, *117*, 2328–2338.

(16) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740–5761.

(17) Jorgensen, W. L.; Tirado-Rives, J. *Perspect. Drug Discovery Des.* **1995**, *3*, 123–138.

(18) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Spoel, D. v. d.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845–854.

(19) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(20) 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. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(21) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.

(22) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(23) MacKerell, J., A D; Banavali, N.; Foloppe, N. *Biopolymers* **2000**, *56*, 257–265.

(24) Mobley, D. L.; Dill, K. A.; Chodera, J. D. *J. Phys. Chem. B* **2008**, *112*, 938–946.

(25) Silva, A. W. S. d.; Vranken, W. F. *BMC Research Notes* **2012**, *5*, 367.

(26) Hawkins, G. D.; Giesen, D. J.; Lynch, G. C.; Chambers, C. C.; Rossi, I.; Storer, J. W.; Li, J.; Thompson, J. D.; Winget, P.; Lynch, B. J.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. AMSOL-version 7.1.

(27) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(28) Hildebrandt, A.; Dehof, A. K.; Rurainski, A.; Bertsch, A.; Schumann, M.; Toussaint, N. C.; Moll, A.; Stöckel, D.; Nickels, S.; Mueller, S. C.; Lenhof, H.-P.; Kohlbacher, O. *BMC Bioinformatics* **2010**, *11*, 531.

(29) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.

(30) Yesselman, J. D.; Price, D. J.; Knight, J. L.; Brooks, r., Charles L *J. Comput. Chem.* **2012**, *33*, 189–202.

(31) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *J. Cheminf.* **2011**, *3*, 33.

(32) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. *J. Chem. Inf. Model.* **2012**, *52*, 3155–3168.

(33) Vanommeslaeghe, K.; MacKerell, A. D. *J. Chem. Inf. Model.* **2012**, *52*, 3144–3154.

(34) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. *J. Comput. Chem.* **2010**, *31*, 671–690.

(35) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132–146.

(36) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(37) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

(38) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. *J. Comput. Chem.* **1999**, *20*, 786–798.

(39) Udier–Blagović, M.; Morales De Tirado, P.; Pearlman, S. A.; Jorgensen, W. L. *J. Comput. Chem.* **2004**, *25*, 1322–1332.

(40) Wallin, G.; Nervall, M.; Carlsson, J.; Åqvist, J. *J. Chem. Theory Comput.* **2009**, *5*, 380–395.

(41) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

(42) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Aided Mol. Des.* **1995**, *9*, 87–110.

(43) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **2003**, *24*, 1291–1304.

(44) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385–16398.

(45) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490–519.

(46) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Van Alsenoy, C.; Tollenaere, J. P. *J. Phys. Chem. A* **2002**, *106*, 7895–7901.

(47) Bultinck, P.; Langenaeker, W.; Carbó-Dorca, R.; Tollenaere, J. P. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 422–428.

(48) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(49) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.

(50) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. *J. Chem. Phys.* **2002**, *117*, 43–54.

(51) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669–681.

(52) Wang, Y.; Li, H. *J. Chem. Phys.* **2009**, *131*, 206101–206102.

(53) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(54) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(55) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1993**, *98*, 1358–1371.

(56) Dunning, T. H.; Peterson, K. A.; Wilson, A. K. *J. Chem. Phys.* **2001**, *114*, 9244–9253.

(57) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620–9631.

(58) Carlson, H. A.; Nguyen, T. B.; Orozco, M.; Jorgensen, W. L. *J. Comput. Chem.* **1993**, *14*, 1240–1249.

(59) Kaminski, G. A.; Jorgensen, W. L. *J. Phys. Chem. B* **1998**, *102*, 1787–1796.

(60) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(61) Hess, B. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.

(62) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(63) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

(64) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182.

(65) Pronk, S.; Larsson, P.; Pouya, I.; Bowman, G.; Haque, I.; Beauchamp, K.; Hess, B.; Pande, V.; Kasson, P.; Lindahl, E. Copernicus: A new paradigm for parallel adaptive molecular dynamics. International Conference for High Performance Computing, Networking, Storage and Analysis (SC), 2011. 2011; pp 1–10, Article 60.

(66) Bennett, C. H. *J. Comput. Phys.* **1975**, *19*, 267–279.

(67) Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. *J. Chem. Theory Comput.* **2006**, *2*, 128–139.

(68) Abraham, M. H.; Andonian-Haftvan, J.; Whiting, G. S.; Leo, A.; Taft, R. S. *J. Chem. Soc., Perkin Trans. 2* **1994**, 1777–1791.