Published in final edited form as:

J Phys Chem B. 2017 April 20; 121(15): 3864–3870. doi:10.1021/acs.jpcb.7b00272.

1.14*CM1A-LBCC: Localized Bond Charge Corrected CM1A Charges for Condensed-Phase Simulations

Leela S. Dodda, Jonah Z. Vilseck, Julian Tirado-Rives, and William L. Jorgensen*

Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107

Abstract

The quality of the 1.14*CM1A and 1.20*CM5 charge models was evaluated for calculations of free energies of hydration. For a set of 426 neutral molecules, 1.14*CM1A and 1.20*CM5 yield MADs of 1.26 and 1.20 kcal/mol, respectively. The 1.14*CM1A charges, which can be readily obtained for large systems, exhibit large deviations only for a subset of functional groups. The results for these cases were systematically improved using Localized Bond Charge Corrections (LBCC) by which off-setting adjustments are made to the partial charges for atoms in specified bond types. Only 19 LBCCs were needed to yield 1.14*CM1A-LBCC charges that reduce the errors for the 426 G_{hyd} values to only 0.61 kcal/mol. The modified charge method was also tested in computation of heats of vaporization and densities for pure organic liquids, yielding average errors of 1.40 kcal/mol and 0.024 g/cm³, similar to those for 1.14*CM1A.

Graphical Abstract

*Corresponding Author: william.jorgensen@yale.edu.

Notes

The authors declare no competing financial interests.

SUPPORTING INFORMATION

The complete results and additional analyses for computed and experimental free energies of hydration, heats of vaporization, and pure liquid densities are provided. This information is available free of charge *via* the Internet at <http://pubs.acs.org>.

INTRODUCTION

Electrostatic interactions play a key role in force-field based molecular simulations.¹ For example, calculated condensed-phase properties such as free energies of hydration (G_{hyd}) and heats of vaporization (H_{vap}) are very sensitive to the electrostatic model. They are equally critical in representing correctly the intermolecular interactions in host-guest complexes, *e. g.* calculating free energies of binding (G_{bind}) in Computer-Aided Drug Design (CADD).² Not surprisingly, much effort has been devoted to their representation in force fields.^{3–5} In spite of recent advances made in the development of polarizable force fields,^{6,7} fixed point charge models are still indispensable because of their low computational cost and reasonable accuracy.^{8–12}

Since most computer based-lead optimization campaigns cover large amounts of chemical space, general methods are needed to obtain atomic charges for arbitrary molecules. The commonly used alternative is to use partial charges derived from a quantum mechanical (QM) calculation on the molecule of interest such as in the CM1A,¹³ CM3P,¹⁴ CM5,¹⁵ AM1-BCC,^{16,17} and CGenFF¹⁸ approaches, which are used in conjunction with the OPLS-AA,³ Amber-GAFF,¹⁹ and CHARMM¹⁸ force fields. Point charge models such as CM1A, CM3P and AM1-BCC are derived from semiempirical QM methods, AM1²⁰ and PM3,^{21,22} but they also differ in terms of the empirical corrections applied after obtaining charges from Mulliken^{23,24} or Löwdin²⁵ population analyses. In contrast, the more recent CM5¹⁵ charge model is based on density functional theory (DFT)²⁶ and uses a Hirshfeld²⁷ population analysis. Since these charge models are parameterized to reproduce gas-phase properties such as dipole moments and electrostatic potentials, the charges need to be increased in magnitude for use in condensed-phase simulations to help take account of polarization effects from the surrounding media. Thus, a scaling factor is commonly employed for neutral molecules, *e. g.* 1.14 for CM1A^{3,28} and 1.20 for CM5.^{29,30} These scale factors are obtained by minimizing errors of condensed-phase properties such as free energies of hydration, densities and heats of vaporization. The AM1-BCC model uses bond-based incremental corrections, while the CGenFF fits gas-phase interaction energies between the compound of interest and water. Since different fragments of any given molecule have different capabilities to become polarized by interactions with their environment, a single scale factor is an over-simplification.

An alternative to the scale factor approach is to make specific corrections to the atomic charges for individual bonds or groups. In order to address this problem, we follow a methodology akin to that of the AM1-BCC model. For example, the AM1-BCC method uses 309 unique bond-charge corrections (BCC), which are parameterized to fit AM1 Mulliken charges to HF/6-31G* electrostatic potentials.^{16,17} In its current implementation, AM1-BCC does well for non-polar molecules such as saturated and aromatic hydrocarbons; however, significant errors persist for some polar compounds such as pyridines, alkyl amines, alkyl and aryl halides, sulfides, and nitriles. The problems likely arise from inadequacies of the HF/6-31G* electrostatic potentials and point-charge model for representing condensed-phase electrostatics. A higher-level QM treatment such as in the CM5 charge model might be expected to help, but the CM5 charges still suffer from the inability of a single, fixed scale factor to account for both internal electron delocalization and external polarization

effects. Thus, in the spirit of combining CMx and AM1-BCC approaches, we decided to explore if it might be possible to make a small number of BCC corrections to correct the key problems with CM1A charges. CM1A was chosen as the basis since the underlying AM1 calculations have low-computational cost and can easily be applied to large collections of molecules and/or to very large molecules such as proteins.

In the initial part of this study, the performance of the 1.14*CM1A and 1.20*CM5 charge models with the OPLS-AA force field for a dataset of 426 free energies of hydration³¹ was thoroughly examined. The methodology to parameterize the LBCC corrections is discussed, and calculations with the resultant 1.14*CM1A-LBCC charges are found to reduce the error in G_{hyd} by half without affecting the performance for pure liquid properties. Comparisons of results with both experimental data and the results from other charge models such as 1.14*CM1A, 1.20*CM5 and AM1-BCC/GAFF are provided.

COMPUTATIONAL DETAILS

Charge Models

All force-field energies were calculated using stretching, bending, torsional and Lennard-Jones parameters from the OPLS-AA force field augmented with the appropriate charge models for the Coulomb interactions. All the required files were prepared using the BOSS¹ program and its included utilities starting from either mol files or SMILES strings. For ease of use, these utilities have been integrated into a publicly accessible web server, LigParGen,³² which was used in this work to generate the CM1A and CM1A-LBCC charges. The needed CM5 charges were obtained from the recently developed BOSS-Gaussian interface³³ which calls the Gaussian program³⁴ to do the quantum mechanical calculations and Hirschfeld analyses, followed by the CM5PAC program³⁵ to calculate the charges. All the Monte Carlo (MC) simulations used to compute the pure liquid properties and free energies of hydration were performed with the BOSS program allowing all degrees of freedom of the organic molecules to be sampled.

Free Energies of Hydration

All simulations were performed in the isothermal-isobaric ensemble at 25 °C and 1 atm in a periodic cube, 35 Å on an edge, containing 1439 TIP4P water molecules³⁶ to handle even the largest molecules in the data set without edge effects. 426 molecules from the FreeSolv database were treated.³¹ All solvent-solvent and solute-solvent intermolecular interactions were truncated at a distance of 10 Å and quadratically smoothed over the last 0.5 Å. Combined Metropolis MC³⁷ and Free Energy Perturbation³⁸ annihilations followed previously reported procedures.^{29,30,39} In brief, the Zwanzig equation³⁸ was used to calculate free energy differences using a combination of double wide (DWS)⁴⁰ and simple overlap sampling (SOS) schemes.⁴¹ Molecules were annihilated to “null” or non interacting states in two steps in each medium. First the electrostatic interactions of the solute were removed by perturbing all partial atomic charges to zero over 15-SOS windows. In a separate calculation, all solute van der Waals interactions were annihilated in 18-SOS windows by perturbing Lennard-Jones parameters to zero and shrinking the molecule. Long range Lennard-Jones corrections were added to the aqueous simulation *post facto*, and the λ

schedule used for performing the annihilation of solute interactions is given in the SI. For solvent phase annihilations, 60 million (M) configurations of equilibration and 110 M configurations of averaging were used in each window, and 8M/8M configurations for gas phase. The computed uncertainties for free energies of hydration were less than 0.20 kcal/mol based on multiple independent runs.

Pure Liquid Simulations

Pure liquid simulations were performed in the isothermal-isobaric ensemble at 1 atm and at different temperatures where experimental data were available as detailed in the Supplementary Information (SI). Metropolis Monte Carlo simulations were performed for a box of 267 molecules with periodic boundary conditions as previously reported.³⁰ All inter and intra-molecular degrees of freedom were sampled. The cutoff distances were chosen based on molecular size; 11 Å for molecules with less than 3 heavy atoms, 13 Å for molecules with 3–5 heavy atoms, and 15 Å for molecules with more than 5 heavy atoms. Intermolecular interactions were truncated at these cutoff distances and quadratically smoothed over the last 0.5 Å. In addition, a long-range van der Waals energy correction was included for interactions neglected beyond the cutoff. Each simulation included 20 million (M) Steps of equilibration and 20 M steps of averaging to ensure convergence. The needed gas-phase simulations were performed at the same temperatures as the liquid simulations with a single molecule for 4 M steps of equilibration and 4 M steps of averaging. Computed uncertainties were less than 0.05 kcal/mol for heats of vaporization and 0.002 g/cm³ for densities and were computed by the batch means method.

GB/SA Free Energies of Hydration

The free energies of hydration using the GB/SA implicit solvent model were also calculated for molecules using the 1.14*CM1A-LBCC charges and compared to the results of the 1.14*CM1A and 1.20*CM5 charge models and those with explicit solvent. The motivation was to elucidate if local changes were needed or a more global measure such as a different scaling factor was required. An exhaustive conformational search with GB/SA invoked was done for each of the 426 molecules, and the minimum energy conformation was used for calculating the GB/SA free energies of hydration. These calculations were also done with the BOSS program, which includes an implementation of the GB/SA model of Still *et al.*^{42,43}

RESULTS

Free Energies of Hydration with 1.14*CM1A and 1.20*CM5 Charges

Free energies of hydration (G_{hyd}) of the 426 molecules were computed in combination with OPLS-AA force field. All values of G_{hyd} obtained in this study are tabulated in the SI; and a graphical comparison between computed and experimental values is shown in Figure 1.

Overall, the results from 1.14*CM1A compare favorably with those from the more computationally expensive 1.20*CM5 charges. The Mean Absolute Deviations (MAD) with 1.14*CM1A and 1.20*CM5 are similar at 1.26 and 1.21 kcal/mol for the 426-molecule

dataset. For a subset of 42 molecules, a MAD of 0.94 kcal/mol had been reported previously for the 1.20*CM5 model.³⁰ The 1.14*CM1A results show larger deviations for specific functional groups such as nitro, alcohols, thiols, sulfides, sulfoxides, aldehydes, ketones, primary amines, aromatic amines and molecules with aromatic cores such as benzene, pyridine and other extended aromatic systems (see SI). On the other hand, the results for alkanes, alkenes and alkynes, and for functional groups such as esters, secondary and tertiary amines, amides, ethers fare well with a MAD of only 0.5 kcal/mol. The localization of the larger errors to a few functionalities means that the performance of the 1.14*CM1A charge model can be improved if deviations for these errant groups are minimized. The 1.20*CM5 model, on the other hand, has a more uniform distribution of errors which would require many corrections. In addition, it makes computed G_{hyd} values too positive with a Mean Signed Deviation (MSD; experiment - calculated) of -1.14 kcal/mol; however, use of a larger scale factor with CM5 would lead to poorer results for liquid densities and heats of vaporization.^{29,30} The other advantage of correcting the CM1A rather than CM5 charges is the relative simplicity of the underlying AM1 calculations for the former.

Localized Bond Charge Corrections

The 1.14*CM1A charges for problematic functional groups were corrected systematically using localized bond charge corrections (LBCC), without affecting groups that are already doing well. As an illustration, the methodology used to obtain the single LBCC correction for aliphatic alcohols is indicated in Figure 2. For each problematic functional group, a representative molecule was chosen, such as ethanol for aliphatic alcohols. In this representative molecule, a small amount of charge is redistributed between the atoms in the bond central of the functional group, for example, the C-O bond in ethanol, which has OPLS atom types CT-OH. The arrow indicates the localized movement of charge in the bond due to the LBCC. The amount of charge is designated the localized bond charge correction, B_{X-Y} , where X and Y are the bonded neighbors. The signed numerical value should be added to the first atom in the pair, and subtracted from the second.

The value for B_{X-Y} needed to minimize the error in G_{hyd} was obtained by evaluating the change in free energy of hydration for different values of the correction using the requisite FEP calculations in water and the gas phase. These changes were then added to the absolute G_{hyd} of the uncorrected molecule to obtain the final results shown in Figure 2. In the case of ethanol, the value of 0.1 e minimized the difference between the computed and experimental G_{hyd} and was chosen to be the LBCC for the CT-OH bond.

The transferability of the resulting LBCC for the CT-OH bond ($B_{CT-OH} = 0.1 e$) was examined by calculating the G_{hyd} for other aliphatic alcohols as shown in Figure 3. As can be seen, the results are significantly improved with a reduction of the MAD from 1.95 to 0.43 kcal/mol for a set of 32 aliphatic alcohols. The small errors obtained are within the numerical uncertainties of the calculations and experiments and did not warrant following an alternate approach based on fitting the correction to a large number of training molecules.²⁸

The same procedure was utilized to obtain LBCC corrections for other errant functional groups. Of the 102 different bonds present in the dataset of 426 molecules, 19 bonds unique to the problematic functional groups were selected. 18 molecules were used to evaluate the

LBCC values, since aniline was sufficient to address both the CA-NP and the NP-HN bonds. Some corrections were evaluated sequentially to avoid interdependencies, *e.g.*, the aromatic core CA-HA was needed in order to calculate the CA-NC LBCC for pyridine and the various CA-X corrections for aromatic functional groups. Table 1 contains the summary of the representative molecules, corresponding bonds and derived LBCC parameters. It is important to note that the C-O LBCC is applied to all carbonyl compounds except amides and esters.

VALIDATION

Having determined the LBCC corrections, testing was carried out for computing G_{hyd} in TIP4P water and GB/SA implicit solvent, and H_{vap} and density of pure liquids. G_{hyd} was calculated in explicit and implicit water for the 426 molecules collected from the FreeSolv database;³¹ full results are provided in the SI. H_{vap} and density were calculated for a subset of molecules at various temperatures, as summarized in the SI. More than 300 experimental observations obtained from references 47 and 48 were used to benchmark the performance of the charge models. The comparisons to the experimental data are plotted in Figure 4, and a summary of the performance of the different charge models is provided in Table 2.

Free energies of hydration in TIP4P water

1.14*CM1A-LBCC provides the best results of the four charge models considered, with a MAD of 0.61 kcal/mol. Unlike the other three charge models, it also has a mean signed error of zero. A comparison of the results in Figures 1 and 4a shows that the use of the LBCCs with the 1.14*CM1A charges greatly improves the MAD from 1.26 to 0.61 kcal/mol, while also improving the correlation coefficient from 0.81 to 0.94. It is clear from Table 3 that, as designed, the application of LBCCs reduces significantly the large systematic errors with CM1A for nitro compounds, anilines, phenols, aliphatic alcohols, nitriles, and sulfur compounds such as thiols, sulfides and sulfoxides, thereby greatly reducing the overall MAD. Some of the most interesting cases are discussed below.

The free energies of hydration calculated with 1.14*CM1A-LBCC charges do well for compounds containing benzene and pyridine cores, with MADs of 0.65 and 0.59 kcal/mol, respectively. The corresponding 1.14*CM1A and 1.20*CM5 charge models yield too exoergic G_{hyd} , whereas AM1-BCC does well for benzene but not pyridine containing compounds. For sulfur containing compounds, 1.14*CM1A-LBCC performs by far the best, with MADs of 0.39 and 0.31 kcal/mol for sulfides and thiols. The 1.14*CM1A and 1.20*CM5 charge models perform poorly for sulfur containing compounds with overall MADs of >2 kcal/mol, while AM1-BCC charges fare somewhat better at 1.35 kcal/mol. For nitriles, the 1.14*CM1A-LBCC model modifies the 1.14*CM1A charges with B_{CZA-NZ} and B_{CZT-NZ} along with core LBCCs like B_{HA-CA} and B_{CA-NC}, reducing the MAD in aromatic and aliphatic nitriles to 0.22 and 0.24 kcal/mol, respectively.

In general, AM1 based charge models such as 1.14*CM1A and AM1-BCC charge models perform poorly for nitro compounds, as discussed before.^{14,27} On the other hand, 1.20*CM5 does well with a MAD below 1.0 kcal/mol. The use of a single B_{NO-ON} LBCC improved the performance of the 1.14*CM1A charge model for aliphatic nitro compounds, while for

aromatic nitro compounds an additional B_{CA-NO} LBCC was also applied. Their use decreased the overall MAD for nitro compounds from 4.18 to 0.69 kcal/mol.

Considering Figure 3a, it could be assumed that alcohols need scale factors larger than 1.14 and 1.20 for CM1A and CM5 charges, respectively. They were indeed modeled poorly by all existing charge models. The B_{CT-OH} and B_{CA-OH} corrections in 1.14*CM1A-LBCC were, however, sufficient to reduce the MAD from 1.82 to 0.45 kcal/mol for aliphatic alcohols and from 1.22 to 0.61 kcal/mol for aromatic alcohols. Even so, problems still persist for 2-nitrophenol as it shows a deviation of 2.9 kcal/mol even after LBCC; without the correction the deviation is much worse at 4.7 kcal/mol. An intramolecular hydrogen bond effect may not be properly accounted for in the AM1 calculations, since the experimental G_{hyd} values for 4-, 3-, and 2-nitrophenols (−10.65, −9.63, −4.58 kcal/mol) do not correlate well with the experimental dipole moments (5.05, 3.91, 3.13)⁴⁴ or pK_a values (7.14, 8.35, 7.23).⁴⁵ The 1.20*CM5 charges reproduce the trend in relative values better, but underestimate the absolute values by 1.2–2.4 kcal/mol. Primary amines such as 1° aliphatic amines and anilines are modeled poorly by 1.14*CM1A charges, but are well represented by the 1.20*CM5 and AM1-BCC models. The use of the B_{CT-NP} , B_{CA-NP} and B_{H-NP} LBCCs for aliphatic and aromatic primary amines reduce the error from 2.39 to 0.41 kcal/mol. Importantly, with LBCCs the absolute deviations for nitro anilines are reduced from 5.6 to 0.78 kcal/mol. Similarly, all charge models do reasonably well for carbonyl compounds such as aldehydes and ketones, except for 1.14*CM1A. The single $B_{C=O}$ LBCC used for carbonyl compounds corrects the problem, reducing the MAD from 1.28 to 0.49 kcal/mol.

In addition, the present results with 1.14*CM1A-LBCC charges were compared with those using Schrödinger's OPLS2/CM1A-BCC charges,⁴⁶ for the 193 molecules common to both studies. The results are provided graphically in SI Figure S1. For these cases, 1.14*CM1A-LBCC performs significantly better with an MAD of 0.55 kcal/mol versus 1.05 kcal/mol for OPLS2/CM1A-BCC. Overall, among general QM-based charge alternatives, the 1.14*CM1A-LBCC charge model in conjunction with the OPLS-AA force field gives by far the best performance for computing free energies of hydration to our knowledge.

Free Energies of Hydration Using GB/SA

Improvement in calculated GB/SA free energies of hydration was not as significant as with explicit TIP4P hydration. This is due to the fact that small charge differences can strongly affect solute – explicit water interactions, especially hydrogen bonds, while they are expected to have modest effects for a continuum solvent model. Overall, 1.20*CM5 charges perform best in GB/SA calculations with an MAD of 1.07 kcal/mol, while use of 1.14*CM1A-LBCC charges does provide a small improvement (1.23 kcal/mol) over uncorrected 1.14*CM1A charges (1.31 kcal/mol).

Heats of Vaporization

The average errors from the four tested charge models are all 1.2 – 1.4 kcal/mol in Table 2. This is significantly worse than the average error of 0.3 kcal/mol for ca. 100 liquids with OPLS-AA, for which the parameterization has emphasized this property and liquid densities.^{3,30} The 1.14*CM1A-based charge models on average over-estimate heats of vaporization

by ca. 1 kcal/mol as shown by the MSDs, although the values for the MADs are essentially unchanged. This is expected since the charges that are appropriate for computing free energies of hydration are likely over-polarized for the lower dielectric environment in pure organic liquids.³⁰ Consistently, the scale factor of 1.20 for CM5 charges was derived as a compromise for fitting G_{hyd} as well as H_{vap} , and it gives a smaller average over-estimate (0.5 kcal/mol) and a lower overall error (1.20 kcal/mol). However, the use of the LBCCs does lead to a better correlation coefficient, slope and intercept in Table 2 than the other charge models. This suggests that the LBCC changes correct some serious issues with the CM1A charges and reduce large deviations.

Nitro compounds, in particular, are poorly modeled with 1.14*CM1A and AM1-BCC charges, leading to large errors in both heats of vaporization and liquid densities. The results from 1.14*CM1A and AM1-BCC give errors of 4.89 and 3.69 kcal/mol for nitrobenzene, and MADs of 4.39 and 3.78 for aliphatic nitro compounds, respectively. These errors are significantly reduced with the LBCC corrections, leading to much improved performance from the 1.14*CM1A-LBCC charge model, with MADs of 1.99 and 1.60 kcal/mol for nitrobenzene and aliphatic nitro compounds. For sulfur-containing compounds, 1.20*CM5 charges perform less well than average with an MAD of 1.29 kcal/mol, while 1.14*CM1A charges do better with an MAD of 1.03 kcal/mol that is further improved by the LBCCs to 0.66 kcal/mol. The 1.14*CM1A-LBCC and 1.14*CM1A charges also do well for nitriles with MADs of 0.86 and 0.38 kcal/mol, respectively. 1.20*CM5 charges overestimate H_{vap} in this case with an MAD of 1.26 kcal/mol.

Overall for alcohols, 1.14*CM1A-LBCC gives the best results with a MAD of 0.76 kcal/mol compared to 1.14*CM1A, 1.20*CM5 and AM1-BCC/GAFF. 1.20*CM5 underestimates the heats of vaporization, whereas AM1-BCC/GAFF overestimates them. A recent study reported an interesting variant of AM1-BCC charges that reduces some of the over-estimations by reparameterization to reproduce the dielectric constants for pure liquids (GAFF-DC).⁸

Pure Liquid Densities

Calculations with the OPLS-AA force field in combination with the 1.14*CM1A, 1.14*CM1A-LBCC and 1.20*CM5 charge model give very good results for pure liquid densities. The MADs are all similar near 0.02 g/cm³. This is expected as the Lennard-Jones (LJ) parameters, which strongly influence liquid densities, are adopted from OPLS-AA. The change of 0.0006 g/cm³ in MAD when the LBCC corrections are added to the 1.14*CM1A charges is well within the simulation uncertainty of 0.002 g/cm³. Simulations with both the 1.14*CM1A and 1.14*CM1A-LBCC charge models slightly overestimate densities with MSDs of -0.060 and -0.018 g/cm³, mirroring their behavior for heats of vaporization. The errors are greater for the AM1-BCC/GAFF model, with a MAD of 0.039 g/cm³. Curiously, though simulations with the AM1-BCC/GAFF charge model overestimate the heats of vaporization, they underestimate the densities, suggesting that the Lennard-Jones parameters might benefit from some adjustments.

CONCLUSION

In this study, the performance of 1.14*CM1A and 1.20*CM5 charge models in combination with the OPLS-AA force field was evaluated for free energies of hydration. For a set of 426 molecules, 1.14*CM1A and 1.20*CM5 yield MADs of 1.26 and 1.20 kcal/mol, respectively. The deviations with the 1.14*CM1A were noted to arise primarily from a subset of functional groups, while performing well for others. Thus, the 1.14*CM1A charges for the problematic groups were addressed by the addition of localized bond charge corrections, LBCCs. It was found that use of only 19 LBCC was able to lower the MAD for the 426 free energies of hydration to 0.61 kcal/mol. Further improvements are expected to be limited by the accuracy of the experimental data³⁰ and the simplicity of the atom-centered point charge model. The 1.14*CM1A-LBCC charges in conjunction with the OPLS-AA force field were also tested for calculating heats of vaporization (H_{vap}) and densities of ca. 150 pure organic liquids. The densities are in excellent accord with experiment with an MAD of only 0.024 g/cm³, while the heats of vaporization with an MAD of 1.40 kcal/mol are significantly less accurate than from OPLS-AA. Of course, the later force field is limited in its coverage since it does not include partial charge assignments for most of the vast range of molecules that can be treated with the 1.14*CM1A-LBCC approach. The new charge model features the addition of only a small number of parameters and yields substantial improvements over the existing alternatives for use in computing free energies of hydration, which are a cornerstone for biomolecular modeling. The few parameters needed to implement the method are given in this paper, and 1.14*CM1A-LBCC charges for most molecules can be obtained at no cost through the LigParGen web service.³²

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Gratitude is expressed to the National Institutes of Health (GM32136) for support. This work was supported in part by the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Center. Gratitude is expressed to Israel Cabeza de Vaca and Michael Robertson for helpful discussions.

References

1. Jorgensen WL, Tirado-Rives J. Molecular Modeling of Organic and Biomolecular Systems Using BOSS and MCPRO. *J Comput Chem.* 2005; 26:1689–1700. [PubMed: 16200637]
2. Jorgensen WL. Efficient Drug Lead Discovery and Optimization. *Acc Chem Res.* 2009; 42:724–733. [PubMed: 19317443]
3. Jorgensen WL, Tirado-Rives J. Potential Energy Functions for Atomic-Level Simulations of Water and Organic and Biomolecular Systems. *Proc Nat Acad Sci USA.* 2005; 102:6665–6670. [PubMed: 15870211]
4. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14sb: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99sb. *J Chem Theory Comput.* 2015; 11:3696–3713. [PubMed: 26574453]
5. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD Jr. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *J Chem Theory Comput.* 2012; 8:3257–3273. [PubMed: 23341755]

6. Ponder JW, Wu C, Ren P, Pande VS, Chodera JD, Schnieders MJ, Haque I, Mobley DL, Lambrecht DS, DiStasio RA, Head-Gordon M, Clark GNI, Johnson ME, Head-Gordon TJ. Current Status of the AMOEBA Polarizable Force Field. *J Phys Chem B*. 2010; 114:2549–2564. [PubMed: 20136072]
7. Lopes PEM, Huang J, Shim J, Luo Y, Li H, Roux B, MacKerell AD Jr. Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. *J Chem Theory Comput*. 2013; 11:3696–3713.
8. Fennell CJ, Wymer KL, Mobley DL. A Fixed-Charge Model for Alcohol Polarization in the Condensed Phase, and Its Role in Small Molecule Hydration. *J Phys Chem B*. 2014; 118:6438–6446. [PubMed: 24702668]
9. Cerutti DS, Rice JE, Swope WC, Case DA. Derivation of Fixed Partial Charges for Amino Acids Accommodating a Specific Water Model and Implicit Polarization. *J Phys Chem B*. 2013; 117:2328–2338. [PubMed: 23379664]
10. Mobley DL, Dumont E, Chodera JD, Dill KA. Comparison of Charge Models for Fixed-Charge Force Fields: Small Molecule Free energies of hydration in Explicit Solvent. *J Phys Chem B*. 2011; 115:1329.
11. Mobley DL, Bayly CI, Cooper MD, Shirts MR, Dill KA. Small Molecule Free energies of hydration in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *J Chem Theory Comput*. 2009; 5:350–358. [PubMed: 20150953]
12. Swope WC, Horn HW, Rice JE. Accounting for Polarization Cost When Using Fixed Charge Force Fields. II. Method and Application for Computing Effect of Polarization Cost on Free Energy of Hydration. *J Phys Chem B*. 2010; 114:8621–8630. [PubMed: 20540503]
13. Storer J, Giesen D, Cramer C, Truhlar D. Class IV charge models: A new semiempirical approach in quantum chemistry. *J Comput-Aided Mol Des*. 1995; 9:87–110. [PubMed: 7751872]
14. Thompson JD, Cramer CJ, Truhlar DG. Parameterization of charge model 3 for AM1, PM3, BLYP, and B3LYP. *J Comput Chem*. 2003; 24:1291–1304. [PubMed: 12827670]
15. Marenich AV, Jerome SV, Cramer CJ, Truhlar DG. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J Chem Theory Comput*. 2012; 8:527–541. [PubMed: 26596602]
16. Jakalian A, Bush BL, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J Comput Chem*. 2000; 21:132–146.
17. Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and Validation. *J Comput Chem*. 2002; 23:1623–1641. [PubMed: 12395429]
18. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem*. 2010; 31:671–690. [PubMed: 19575467]
19. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem*. 2004; 25:1157–1174. [PubMed: 15116359]
20. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J Am Chem Soc*. 1985; 107:3902–3909.
21. Stewart JJP. Optimization of parameters for semiempirical methods I. Method. *J Comput Chem*. 1989; 10:209–220.
22. Stewart JJP. Optimization of parameters for semiempirical methods II. Applications. *J Comput Chem*. 1989; 10:221–264.
23. Mulliken RS. Electronic Structures of Molecules X. Aldehydes, Ketones and Related Molecules. *J Chem Phys*. 1935; 3:564–573.
24. Mulliken RS. Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I. *J Chem Phys*. 1955; 23:1833–1840.
25. Löwdin PO. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *J Chem Phys*. 1950; 18:365–375.

26. Zhao Y, Truhlar DG. Density Functionals with Broad Applicability in Chemistry. *Acc Chem Res.* 2008; 41:157–167. [PubMed: 18186612]
27. Hirshfeld FL. Bonded-atom fragments for describing molecular charge densities. *Theor Chim Acta.* 1977; 44:129–138.
28. Udier-Blagovic M, Morales De Tirado P, Pearlman SA, Jorgensen WL. Accuracy of free energies of hydration using CM1 and CM3 atomic charges. *J Comput Chem.* 2004; 25:1322–1332. [PubMed: 15185325]
29. Vilseck JZ, Tirado-Rives J, Jorgensen WL. Evaluation of CM5 Charges for Condensed-Phase Modeling. *J Chem Theory Comput.* 2014; 10:2802–2812. [PubMed: 25061445]
30. Dodda LS, Vilseck JZ, Cutrona KJ, Jorgensen WL. Evaluation of CM5 Charges for Nonaqueous Condensed-Phase Modeling. *J Chem Theory Comput.* 2015; 11:4273–4282. [PubMed: 26575922]
31. Mobley DL, Guthrie JP. FreeSolv: a database of experimental and calculated hydration free energies, with input files. *J Comput-Aided Mol Des.* 2014; 28:711–720. [PubMed: 24928188]
32. Dodda, LS., de Vaca, IC., Tirado-Rives, J., Jorgensen, WL. The LigParGen server is accessible at the url. LigParGen: OPLS/CM1A Parameter Generator for Organic Ligands. <http://jorgensenresearch.com/ligpargen>
33. Vilseck JZ, Kostal J, Tirado-Rives J, Jorgensen WL. Application of a BOSS—Gaussian interface for QM/MM simulations of Henry and methyl transfer reactions. *J Comput Chem.* 2015; 36:2064–2074. [PubMed: 26311531]
34. Frisch, MJ., Trucks, GW., Schlegel, HB., Scuseria, GE., Robb, MA., Cheeseman, JR., Scalmani, G., Barone, V., Mennucci, B., Petersson, GA., Nakatsuji, H., Caricato, M., Li, X., Hratchian, HP., Izmaylov, AF., Bloino, J., Zheng, G., Sonnenberg, JL., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, JA., Jr, Peralta, JE., Ogliaro, F., Bearpark, M., Heyd, JJ., Brothers, E., Kudin, KN., Staroverov, VN., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, JC., Iyengar, SS., Tomasi, J., Cossi, M., Rega, N., Millam, JM., Klene, M., Knox, JE., Cross, JB., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, RE., Yazyev, O., Austin, AJ., Cammi, R., Pomelli, C., Ochterski, JW., Martin, RL., Morokuma, K., Zakrzewski, VG., Voth, GA., Salvador, P., Dannenberg, JJ., Dapprich, S., Daniels, AD., Farkas, Ö., Foresman, JB., Ortiz, JV., Cioslowski, J., Fox, DJ. *Gaussian 09, revision A.02.* Gaussian, Inc; Wallingford, CT: 2009.
35. Marenich, AV., Cramer, CJ., Truhlar, DG. *CM5PAC*, version 2011. University of Minnesota; Minneapolis, MN: 2011.
36. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys.* 1983; 79:926–935.
37. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. *J Chem Phys.* 1953; 21:1087–1092.
38. Zwanzig RW. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J Chem Phys.* 1954; 22:1420–1426.
39. Vilseck JZ, Tirado-Rives J, Jorgensen WL. Determination of partial molar volumes from free energy perturbation theory. *Phys Chem Chem Phys.* 2015; 17:8407–8415. [PubMed: 25589343]
40. Jorgensen WL, Ravimohan C. Monte Carlo simulation of differences in free energies of hydration. *J Chem Phys.* 1985; 83:3050–3056.
41. Jorgensen WL, Thomas LL. Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria. *J Chem Theory Comput.* 2008; 4:869–876. [PubMed: 19936324]
42. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc.* 1990; 112:6127.
43. Jorgensen WL, Ulmschneider JP, Tirado-Rives J. Free Energies of Hydration from a Generalized Born Model and an All-Atom Force Field. *J Phys Chem B.* 2004; 108:16264–16270.
44. Catalán J, Macías A. Dipole Moment, Isomerization, and intramolecular hydrogen bond in some benzene derivatives. *J Mol Str.* 1977; 38:209–220.
45. Ehlerova J, Trevani L, Sedlbauer J, Tremaine P. Spectrophotometric Determination of the Ionization Constants of Aqueous nitrophenols at Temperatures up to 225°C. *J Solution Chem.* 2008:857–874.

46. Shivakumar D, Harder E, Damm W, Friesner RA, Sherman W. Improving the Prediction of Absolute Solvation Free Energies Using the Next Generation OPLS Force Field. *J Chem Theory Comput.* 2012; 8:2553–2558. [PubMed: 26592101]
47. Caleman C, van Maaren PJ, Hong M, Hub JS, Costa LT, van der Spoel D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J Chem Theory Comput.* 2012; 8:61–74. [PubMed: 22241968]
48. Lide, DR. *CRC Handbook of Chemistry and Physics.* 90. CRC Press; Cleveland, OH: 2009.

Figure 1.
Free energies of hydration (G_{hyd}) calculated with 1.14*CM1A and 1.20*CM5 charges compared to the experiment.

Figure 2. LBCC methodology demonstrated for ethanol (left). FEP-based parameterization strategy for obtaining LBCC parameters for aliphatic alcohols (right). Q_X is the partial charge on atom X.

Figure 3. Comparison of experimental and computed G_{hyd} using 1.14*CM1A charges without and with LBCC ($B_{CT-OH} = 0.1 e^-$) correction. G_{hyd} are calculated for 32 aliphatic alcohols

Figure 4. Comparison between experimental and calculated values from simulations using 1.14*CM1A-LBCC charges for (a) G_{hyd} in TIP4P water, (b) G_{hyd} with the GB/SA model, (c) H_{vap} and (d) liquid densities. Best-fit lines are shown in red.

Table 1

Localized Bond Charge Corrections and Representative Molecules

Molecule	Bond	LBCC ^a (e ⁻)
<i>Aromatic</i>		
Benzene	HA-CA	-0.01
Pyridine	CA-NC	0.07
Nitrobenzene	CA-NO	-0.08
Aniline	CA-NP	0.06
Aniline	H-NP	-0.05
Fluorobenzene	CA-F	0.13
Chlorobenzene	CA-Br	0.19
Benzonitrile	CZA-NZ	0.09
Phenol	CA-OH	0.22
<i>Aliphatic</i>		
Dimethyl sulfoxide	OY-SZ	0.06
Ethanol	CT-OH	0.10
Acetone	C-O	0.05
Dimethyl sulfide	CT-S	0.08
Ethanethiol	CT-SH	0.175
Nitroethane	NO-ON	-0.18
Ethylchloride	CT-Cl	0.10
Ethylbromide	CT-Br	0.08
Ethylamine	CT-NP	0.04
Acetonitrile	CZT-NZ	0.03

^aThe value is added to the 1.14*CM1A charge of the first atom in the designated bond and subtracted from the second.

Table 2

Summary of Computed Results for Liquid-State Properties^a

Models	MAD	MSD	R ²	Slope	Intercept	N
<i>G_{hyd}</i> (kcal/mol)						
1.14*CM1A	1.26	-0.28	0.81	0.76	-0.87	426
1.20*CM5	1.21	-1.14	0.91	0.98	-1.17	426
AM1-BCC/GAFF ^b	0.94	-0.71	0.91	0.95	-0.80	421
<i>GB/SA G_{hyd}</i>						
1.14*CM1A	1.31	0.41	0.81	0.71	-0.49	426
1.20*CM5	1.07	-0.49	0.85	0.83	-0.87	426
<i>H_{vap}</i> (kcal/mol)						
1.14*CM1A	1.40	-0.96	0.71	0.69	2.43	155
1.20*CM5	1.20	-0.50	0.73	0.79	1.65	155
AM1-BCC/GAFF ^c	1.31	-0.97	0.77	0.70	2.30	147
<i>1000*p</i> (g/cm ³)						
1.14*CM1A	23.79	-5.98	0.99	0.95	39	153
1.20*CM5	21.50	-0.61	0.99	0.95	53	153
AM1-BCC/GAFF ^c	39.27	11.58	0.95	1.23	-212	147

^aMAD and MSD are the mean absolute and signed deviations to the experimental data; N is the number of data points.^bFrom ref. 31.^cFrom ref. 47.

Table 3

Average Errors for Computed Free Energies of Hydration (kcal/mol)

TYPE	N	1.14*CM1A	1.14*CM1A-LBCC	1.20*CM5	AM1-BCC/GAFF
aromatic	145	1.43	0.65	1.37	0.92
nitro	15	4.18	0.69	0.75	1.23
sulfoxides	1	3.24	0.25	1.58	1.79
sulfides	7	2.23	0.31	2.44	1.56
thiols	4	2.22	0.39	2.33	0.86
aliphatic alcohols	35	1.82	0.45	2.13	1.47
phenols	32	1.22	0.61	2.21	1.36
aromatic nitriles	5	2.43	0.22	1.41	2.08
aliphatic nitriles	4	0.55	0.24	0.76	1.27
primary amines	24	2.39	0.41	1.25	0.67
carbonyl ^a	50	1.28	0.49	0.9	0.48
chloroalkanes	27	1.26	0.53	1.54	1.29
bromoalkanes	16	1.22	0.38	0.77	1.59
bromoarenes	3	0.71	0.05	1.57	1.39

^aExcluding amides and esters.