

Improving the Point-Charge Description of Hydrogen Bonds by Adaptive Force Matching

Omololu Akin-Ojo and Feng Wang*

Department of Chemistry, Boston University, Boston, Massachusetts 02215

Received: October 21, 2008; Revised Manuscript Received: December 04, 2008

The MP2_f water force field, recently developed using the adaptive force matching method, is improved significantly after the introduction of a short-range repulsion term. This term is designed to address the deficiency of point-charge models at short charge–charge separations. It may also provide a more realistic description of hydrogen bonds by capturing exchange–repulsion interactions. When compared with MP2_f, the new MP2_f-hb force field predicts properties of liquid water in much better agreement with experiments. The calculated site–site radial distribution functions are in good agreement with X-ray diffraction data corroborating the argument that point-charge models with only Lennard-Jones repulsion underestimate the nearest neighbor O–O distance. The incorporation of the new short-range repulsion term is argued to be important for any force field model that uses point charges to model electrostatics.

The development of a simple force field that can accurately describe the intermolecular interactions in the condensed phase is an important topic that has drawn a great deal of attention. Recently, the adaptive force matching (AFM)¹ method has been developed that is capable of parametrizing simple force fields to the quality of condensed phase QM/MM calculations. The AFM method was tested by generating force fields for water.¹ The resulting force fields were found to reasonably reproduce the QM forces in the condensed phase QM/MM calculations. The properties of water calculated using the BLYP_f, the B3LYP_f, and the MP2_f force fields were in good agreement with experimental values. It was suggested that the quality of these force fields is restricted mostly by the functional forms used in the force matching (FM). Indeed, although a relatively faithful representation of the QM forces was achieved, the scatter plots of the various forces showed a systematic error in that the AFM forces were slightly weaker than the QM reference forces. This underestimation of the QM reference forces by the AFM force fields is evident from Figure 6 of ref 1. The systematic error is very pronounced in the plots of the net forces and torques. Least squares fitting of the net forces produced a slope of 0.7, which is far from the ideal value of 1. In order to understand the origin of this systematic error, the configurations corresponding to the most serious underestimations were examined. We inspected a total of six configurations. In every inspected configuration, the molecule that is associated with the force underestimation participates in a very short hydrogen bond (Figure 1). The net forces on the water molecules forming the short hydrogen bonds are repulsive and the repulsion is underestimated by the AFM force fields developed in ref 1. It is very likely that the simple point-charge based model being parametrized is having trouble reproducing hydrogen bond interactions.

It is not surprising that a point-charge approximation of the underlying QM charge distribution leads to inaccurate forces at short separations. When a short hydrogen bond is formed,

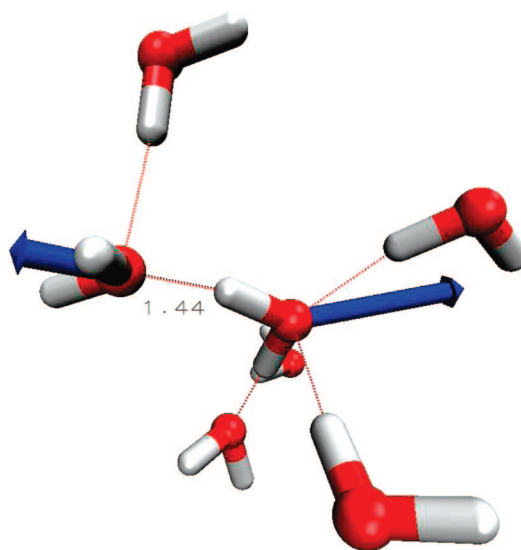


Figure 1. One representative configuration for which the MP2_f model underestimates the intermolecular forces. The blue arrows are the total forces on the molecules.

the point-charge model overestimates the attraction between the positive charge on the hydrogen atom and the negative charge on the oxygen atom. Thus, when combined with Lennard-Jones (L-J) repulsion between the two molecules, the overall repulsive interaction is underestimated. If this is indeed the correct explanation for the underestimation of the repulsion, adding a short-range intermolecular OH repulsion term will correct the systematic error and improve the quality of the fit. In this letter, the importance of such an OH repulsion term will be tested. The potential will be parametrized following the AFM procedure reported previously.¹ The AFM method samples the configuration space using molecular mechanics and parametrizes an MM force field to reproduce QM forces from QM/MM calculations. Starting from a guess force field the AFM is comprised of three steps.

* To whom correspondence should be addressed.

The Molecular Dynamics (MD) Step. MD is carried out to sample the configuration space using the guess force field.

The QM/MM Step. After equilibrium has been reached for a number of configurations sampled in the MD step, QM/MM calculations are performed treating part of the system as the QM region and the remaining as the MM region.

The FM Step. After the QM forces are obtained in the QM/MM step, the FM technique is applied to reparametrize a new force field that minimizes the difference between the AFM and QM forces.

After the FM step, the new reparametrized force field is used as the guess force field and AFM restarts from the MD step. One iteration over the three steps is referred to as a “generation.” In a typical AFM calculation, multiple generations of FM are done until convergence is achieved. After convergence is reached, the QM-quality force field from the FM step has been used for the MD sampling and for representing the MM part in the QM/MM calculations. Thus, not only would the configuration space sampled in the MD step be of QM quality, the quality of the QM/MM calculation is also ensured. For a more detailed description of the AFM method see ref 1. In this work, the ab initio MP2 method will be used as the QM method in the QM/MM calculations.

The objective function being optimized in the FM step is

$$\chi^2 = \sum_{\alpha} w_{\alpha} (f_{\alpha}^{\text{fit}} - f_{\alpha}^{\text{ai}})^2 + \sum_i w_i (F_i^{\text{fit}} - F_i^{\text{ai}})^2 + \sum_i u_i (\tau_i^{\text{fit}} - \tau_i^{\text{ai}})^2 \quad (1)$$

where f_{α} is the atomic force on the α th atom, F_i is the net force, and τ_i the net torque of the i th molecule about its centroid. The superscripts “fit” and “ai” refer to quantities obtained from the model or the ab initio calculations, respectively.

A 7-water QM region was used as done previously.¹ The weights for the net molecular force w_i and molecular torque u_i were chosen to be $w_i = s_i |F_i^{\text{ai}}|^2 / \sum_i s_i |F_i^{\text{ai}}|^2$ and $u_i = s_i |\tau_i^{\text{ai}}|^2 / \sum_i s_i |\tau_i^{\text{ai}}|^2$. This choice better ensures larger net forces and torques are well reproduced. The factors s_i were determined using a distance based cutoff. If a QM molecule does not have an MM atom within 2.1 Å from any of its atom, the QM molecule will have a weight $s_i = 1.0$; otherwise the forces and torques on the QM molecule will be excluded from the fit ($s_i = 0.0$). In the AFM parametrization of the MP2_f force field, point charges, Lennard-Jones, harmonic bond, and harmonic angle terms were included in the potential. In this work, a short-range intermolecular interaction was added between oxygen and hydrogen atoms. We experimented with a few different terms that are functions of the intermolecular O–H separation r_{OH} . The quality of the terms can be judged by the root-mean-square error (RMSE) of the fit. The shifted truncated $1/r_{\text{OH}}^4$ term provides the largest improvement in RMSE:

$$U_{\text{hb}}(r_{\text{OH}}) = \begin{cases} A_4/r_{\text{OH}}^4 - A_4/r_c^4 - 4A_4(r_{\text{OH}} - r_c)/r_c^5 & r \leq r_c \\ 0 & r > r_c \end{cases} \quad (2)$$

Both $U_{\text{hb}}(r_{\text{OH}})$ and its derivatives go to zero smoothly at the cutoff distance r_c . It is worth noting that although eq 2 was designed to capture the overestimation of the repulsion between oxygen and hydrogen atoms at short distance, it was not restricted to be repulsive. However, it indeed came out to be repulsive through AFM. Other terms in the potential may be important such as anharmonic bond stretching. We are working on the systematic survey of important functional forms with

TABLE 1: Parameters for the MP2_{f-hb} and MP2_f Force Fields

| | MP2 _{f-hb} | MP2 _f |
|---|---------------------|------------------|
| q_{O} [au] | − 0.938 | − 0.854 |
| q_{H} [au] | 0.469 | 0.427 |
| r_e (OH) [Å] | 0.960 | 0.960 |
| k_r [kcal mol ^{−1} Å ^{−2}] | 1092.21 | 1095.62 |
| θ_e (HOH) [degrees] | 109.19 | 109.35 |
| k_{θ} [kcal mol ^{−1} rad ^{−2}] | 86.28 | 85.00 |
| A_{OO} [10 ³ kcal mol ^{−1} Å ¹²] | 459.896 | 510.720 |
| C_{OO} [kcal mol ^{−1} Å ⁶] | − 158.859 | 257.475 |
| A_4 [kcal mol ^{−1} Å ⁴] | 31.799 | |
| r_c [Å] | 2.29 | |

AFM. As the focus of this letter is to establish the importance of the short-range OH repulsion term, we decided not to include other additional terms in the fit. We would like to note, however, that when anharmonic OH stretching potentials are included, the OH covalent bond is more likely to sample larger OH separations when another water molecule is nearby, thus providing an even shorter intermolecular hydrogen bond. When such potentials are included, the importance of the explicit intermolecular O–H repulsion term will be more pronounced.

All the parameters in the force field are linear parameters that can be obtained by solving the generalized linear least-squares problem using singular value decomposition (SVD) with the only exception being r_c . As the optimization of r_c for each generation resulted in only negligible improvement in the quality of the fits, a fixed value of $r_c = 2.1$ Å was used for each generation of AFM. The QM forces from our original MP2_f parametrization were refitted with the additional $U_{\text{hb}}(r_{\text{OH}})$ term. The resulting force field was used as the initial guess for the AFM iterations. A total of 10 AFM generations were run. A global fit was performed using all generations. In the global fit, r_c was optimized nonlinearly in an iterative manner using $r_c = 2.1$ Å as the initial guess. Starting with this r_c , SVD was used to determine the linear parameters. After the linear parameters were determined, a one-dimensional nonlinear optimization was performed to obtain a new r_c . These steps were repeated until the objective function was minimized with respect to all parameters. The final optimized r_c is 2.29 Å. The final force field parameters from the fit are reported in Table 1. The MP2_f parameters are also reported in this table for comparison. Since the force field with the intermolecular OH repulsion term is designed to better capture short-range hydrogen bond interactions, the force field will be referred to as MP2_{f-hb}.

MD simulations were performed at 298.15 K at the experimental density using a cubic box containing 216 water molecules as done previously.¹ Figure 2 compares the site–site radial distribution functions (RDFs) obtained using the MP2_{f-hb} and MP2_f force fields and the RDFs published by Soper in 2000 using the empirical potential structure refinement (EPSR) procedure.² Normal-mode path-integral MD (PIMD) simulations were performed to model the quantum nuclear effect. The first peak of the MP2_{f-hb} O–O RDF is slightly shifted to a larger r value when compared with the 2000 EPSR RDF. As pointed out recently by Soper in 2007,³ the EPSR procedure, when combined with the point-charge model and an O–O L-J repulsion term will necessarily underestimate the O–O separation.² Not surprisingly, as published in our previous work, the MP2_f model with only point charges and O–O L-J repulsion also underestimated the O–O separation as did the 2000 EPSR study. After a soft-core repulsive potential was added,³ the position of the first peak in EPSR was shifted from 2.73 to 2.76 Å. However, Soper commented that even with the soft-core

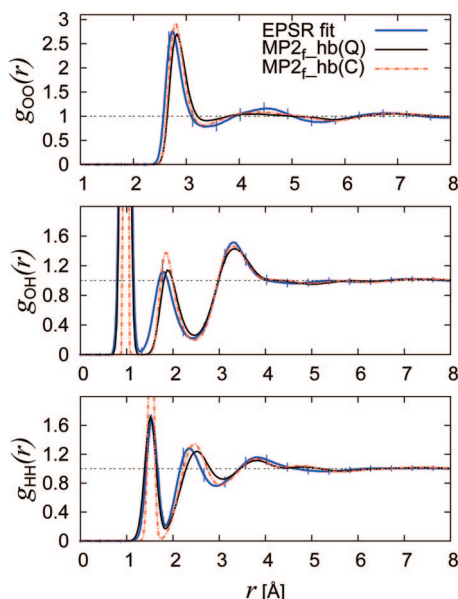


Figure 2. A comparison of site–site RDFs predicted by the MP2_f-hb model and the EPSR fit reported by Soper in 2000.² Soper argued that the EPSR fit underestimates the location of the first O–O peak.³ The solid blue lines with error bars are the experimental data, the solid black lines are the quantum results (Q) while the dashed-dotted red lines are the classical results (C).

potential, the nearest-neighbor O–O distance is underestimated. Our AFM classical simulation predicted a nearest-neighbor O–O separation of 2.79 Å. This separation increased to 2.81 Å after PIMD was used to model the quantum nuclear effects. The 2.81 Å estimate agrees well with the X-ray data set of Head-Gordon and co-workers.^{3,4}

The diffusion coefficient D_s was determined to be 2.3×10^{-5} cm²/s from classical MD simulations with MP2_f-hb. A centroid molecular dynamics simulation^{5–7} estimated the diffusion constant to be 3.3×10^{-5} cm²/s. This differs from the experimental value⁸ of 2.3×10^{-5} cm²/s, and we acknowledge a finite size effect would increase D_s by about 0.4×10^{-5} cm²/s. This overestimation of the diffusion constant by our model is most likely due to the implicit treatment of polarization. Polarization effects will cause extra two-body stabilization and, thus, make the water–water boundary stickier. The heat of vaporization ΔH_{vap} from classical and quantum simulations were 10.64 and 8.57 kcal/mol, respectively. The classical ΔH_{vap} was obtained following the same procedure as described in ref 1. The quantum ΔH_{vap} was determined from the difference in energies between an MP2_f-hb water in liquid and an MP2 water in the gas phase.⁹ The virial estimator¹⁰ was used to determine the PIMD internal energy. When compared with the experimental estimate of 10.52 kcal/mol,¹¹ our model underestimates ΔH_{vap} . We believe that this is due to the repulsive van der Waals $1/r^6$ term in our force field. Although the van der Waals term was argued to be unimportant for accurate fitting of the intermolecular forces in liquid water,¹ its contribution to the heat of vaporization may not be negligible. We are working on addressing this problem in AFM. The constant volume heat capacity C_v was calculated using finite difference. A value of $C_v = 17.58$ cal/(mol·K) was obtained with PIMD which is in very good agreement with the experimental value¹² of 17.77 cal/(mol·K). A classical calculation with quantum correction following the procedure we used previously¹ produced a C_v of 19.23 cal/(mol·K).

The average dipole moment of the water molecule from a classical simulation with MP2_f-hb is 2.74 D. This is larger than

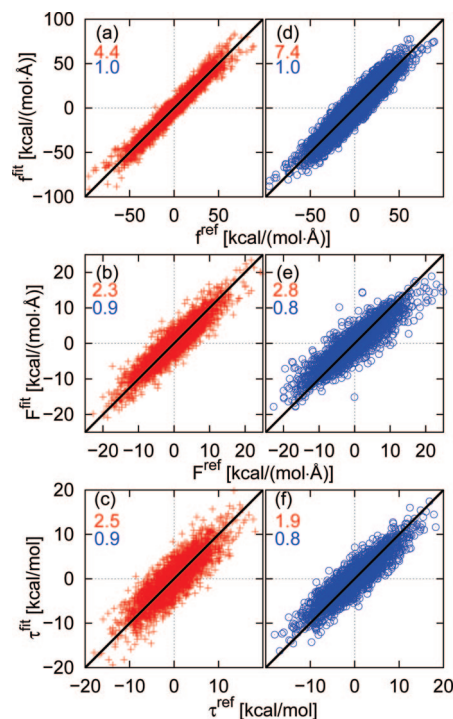


Figure 3. Scatter plots of atomic forces (a,d) molecular forces (b,e) and molecular torques (c,f) for the MP2_f-hb model (left) and the TTM2.1-F model (right). In each panel, the RMSE of the fit is reported in red and the slope in blue. The units for the forces are kcal/(mol·Å) and the torques are in kcal/mol. The reference forces were calculated using the ab initio MP2 method in QM/MM. Perfect agreement is achieved when all the data points lie on the diagonal line. When the TTM2.1-F forces were calculated, the entire box was treated using the TTM2.1-F force field.

2.47 D obtained from MP2_f and compares more favorably with experimental estimates^{13,14} that are scattered around 2.6 to 2.9 D with error bars as large as 0.6 D. This is not surprising as AFM underestimated the magnitude of the point charges in the parametrization of the MP2_f force field in order to avoid too strong an attraction at short-range. Once the explicit repulsion term is included in the force field, the AFM method does a better job at predicting the charges thus improving the description of the electrostatic interaction. The incorporation of the intermolecular OH repulsion term significantly improved the quality of the fit reducing the RMSE of the net forces to 2.3 from a value of 4.7 kcal/(mol·Å) achieved by MP2_f. Our $U_{\text{hb}}(r_{\text{OH}})$ repulsion can be considered as an explicit hydrogen bond term.^{15–17} One possible explanation for the significant improvement in the fit is that this term addresses the breakdown of the point-charge model at short charge–charge distances. One could also argue that exchange-repulsion, which becomes stronger at shorter distances, is important for an accurate description of the hydrogen bond.

In order to investigate the importance of exchange-repulsion, scatter plots were produced comparing both the MP2_f-hb potential and the TTM2.1-F¹⁸ potential with the reference condensed phase QM forces (Figure 3). The MP2_f-hb model performs better than TTM2.1-F producing not only smaller RMSE but also improved slope in the scatter plot. TTM2.1-F is a polarizable model that uses smeared charges and a long list of terms to faithfully reproduce the MP2 electrostatics under a broad range of conditions. The fact that TTM2.1-F underestimates the slope may indicate that exchange-repulsion is indeed important when the hydrogen bond separation is short.

It is worth commenting that while TTM2.1-F predicts accurate intermolecular forces and torques, the quality of the atomic forces predicted by TTM2.1-F is less than stellar when compared with MP2. This could be due to the fact that TTM2.1-F determines the intramolecular potential energy surface from gas phase monomer calculations. AFM opens the possibility of fitting intramolecular potential energy surfaces directly in the condensed phase.

The accurate modeling of hydrogen bonds plays a crucial role in many problems. Simple point-charge models have been used extensively to model hydrogen bond interactions, which apparently are inadequate as seen from this study. Another scenario where a simple point-charge model is likely to fail is the solvation of charged ions as the electrostatic interactions can get quite short-ranged in these systems. We expect a similar repulsion term will be important for all the cases where short-range point-charge interactions play a crucial role. For systems other than water, the shifted truncated $1/r^4$ term in particular may or may not be the best choice; however, any choice of a short-range repulsion term can be judged by examining the RMSE in AFM. We acknowledge smeared-charge models have been used extensively to address the deficiency of point-charge models at short-range. However, the repulsive term in our MP2_f_hb force field is much simpler than smeared-charge models. We would like to note that the requirement for a short-range repulsion to supplement a point-charge model for water follows naturally after inspecting configurations where AFM with only point-charge electrostatics produced significant errors. The inclusion of a simple repulsion term significantly reduced the error of the fit proving its importance. Without AFM, it is hard to imagine this term could play such an important role in a water force field. We believe AFM is an important method in understanding the basic interactions between atoms and will become more popular in the near future as a powerful tool for force field development.

Acknowledgment. This work was supported by ACS PRF 47001-AC6 and NSF CAREER award CHE0748628. The computer resource for this study was provided by the National Center for Supercomputing Applications under Grant MRAC TG-CHE070060 and by the Boston University's Center for Scientific Computing and Visualization.

References and Notes

- (1) Akin-Ojo, O.; Song, Y.; Wang, F. *J. Chem. Phys.* **2008**, *129*, 064108.
- (2) Soper, A. K. *Chem. Phys.* **2000**, *258*, 121.
- (3) Soper, A. K. *J. Phys.: Condens. Matter* **2007**, *19*, 335206.
- (4) Hura, G.; Russo, D.; Glaeser, R. M.; Head-Gordon, T.; Krack, M.; Parrinello, M. *Phys. Chem. Chem. Phys.* **2003**, *5*, 1981.
- (5) Feynman, R. P.; Kleinert, H. *Phys. Rev. A* **1986**, *34*, 5080.
- (6) Ramirez, R.; López-Ciudad, T. *Phys. Rev. Lett.* **1999**, *83*, 4456.
- (7) Cao, J.; Voth, G. A. *J. Chem. Phys.* **1994**, *100*, 5093.
- (8) Krynicki, K.; Green, C. D.; Sawyer, D. W. *Faraday Discuss. Chem. Soc.* **1978**, *66*, 199.
- (9) AFM was done on a single water molecule in vacuum using MP2/ aug-cc-pvdz. This gave $k_r = 1124.60$ kcal/(mol·Å²), $r_e = 0.9690$ Å, $k_\theta = 85.14$ kcal/(mol·rad²), and $\theta_e = 103.533^\circ$.
- (10) Herman, M. F.; Bruskin, E. J.; Berne, B. J. *J. Chem. Phys.* **1982**, *76*, 5150.
- (11) 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.
- (12) Wagner, W.; Pruss, A. *J. Phys. Chem. Ref. Data* **2002**, *31*, 387.
- (13) Clough, S. A.; Beers, Y.; Klein, G. P.; Rothman, L. S. *J. Chem. Phys.* **1973**, *59*, 2254.
- (14) Badyal, Y. S.; Saboungi, M. L.; Price, D. L.; Shastri, S. D.; Haeflner, D. R.; Soper, A. K. *J. Chem. Phys.* **2000**, *112*, 9206.
- (15) Singh, R. D.; Ferro, D. R. *J. Phys. Chem.* **1974**, *78*, 970.
- (16) Fabiola, F.; Bertram, R.; Korostelev, A.; Chapman, M. S. *Protein Sci.* **2002**, *11*, 1415.
- (17) Kortemme, T.; Morozov, A. V.; Baker, D. *J. Mol. Biol.* **2003**, *326*, 1239.
- (18) Fanourgakis, G. S.; Xantheas, S. S. *J. Phys. Chem. A* **2006**, *110*, 4100.

JP809324X