Lennard−Jones Parameters for B3LYP/CHARMM27 QM/MM Modeling of Nucleic Acid Bases

Ulla Pentikäinen,*[†,‡] Katherine E. Shaw,[†] Kittusamy Senthilkumar,[†,§] Christopher J. Woods,[†] and Adrian J. Mulholland*[†]

Centre for Computational Chemistry, School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, United Kingdom, and Department of Biological and Environmental Science and NanoScience Center, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland

Abstract: Combined quantum mechanics/molecular mechanics (QM/MM) methods allow computations on chemical events in large molecular systems. Here, we have tested the suitability of the standard CHARMM27 forcefield Lennard−Jones van der Waals (vdW) parameters for the treatment of nucleic acid bases in QM/MM calculations at the B3LYP/6−311+G(d,p)-CHARMM27 level. Alternative parameters were also tested by comparing the QM/MM hydrogen bond lengths and interaction energies with full QM [B3LYP/6−311+G(d,p)] results. The optimization of vdW parameters for nucleic acid bases is challenging because of the likelihood of multiple hydrogen bonds between the nucleic acid base and a water molecule. Two sets of optimized atomic vdW parameters for polar hydrogen, carbonyl carbon, and aromatic nitrogen atoms for nucleic acid bases are reported: base-dependent and base-independent. The results indicate that, for QM/MM investigations of nucleic acids, the standard forcefield vdW parameters may not be appropriate for atoms treated by QM. QM/MM interaction energies calculated with standard CHARMM27 parameters are found to be too large, by around 3 kcal/mol. This is because of overestimation of electrostatic interactions. Interaction energies closer to the full QM results are found using the optimized vdW parameters developed here. The optimized vdW parameters [developed by reference to B3LYP/6−311+G(d,p) results] were also tested at the B3LYP/6−31G(d) QM/MM level and were found to be transferable to the lower level. The optimized parameters also model the interaction energies of charged nucleic acid bases and deprotonation energies reasonably well.

Introduction

Computational studies of chemical reactions in condensed phases ideally require a method that describes electronic changes in the region of interest. Current quantum mechanical (QM)−molecular electronic structure approaches can provide such descriptions, but the relatively high cost of these

methods limits the size of the systems that can be treated.[1] However, when for example enzyme-catalyzed reactions are studied, inclusion of the surrounding enzyme and water molecules can be crucial for the reliable treatment of the reaction energetics.[2] This, however, increases the size of the system significantly, making calculations too computationally expensive for pure QM methods. To investigate the effects of the environment on chemical events, an implicit or explicit representation of the environment is needed. For biological systems, combined QM and molecular mechanical (MM) methods are increasingly popular and important.[3−9] QM/ MM methods enable computations on complex chemical events in large systems, by dividing the system into a

---

* E-mail: ulla.m.pentikainen@jyu.fi; Adrian.Mulholland@bris.ac. uk.

† University of Bristol.

‡ University of Jyväskylä.

§ Present address: Department of Physics, Bharathiar University, Coimbatore 641 046 India.

B3LYP/CHARMM27 QM/MM Modeling of Nucleic Acid Bases

J. Chem. Theory Comput., Vol. 5, No. 2, 2009 **397**

quantum region and molecular mechanics region. QM/MM methods are well suited for studying enzyme-catalyzed reactions, which take place in a solvent and biomolecular (e.g., protein, nucleic acids, carbohydrates, or lipids) environment, involving thousands of atoms.[3,6,7]

The total Hamiltonian for the molecular system under consideration in the QM/MM framework can be written as:

$$H = H_{QM} + H_{QM/MM} + H_{MM} \tag{1}$$

where H_{QM} and H_{MM} are the normal QM and MM Hamiltonians that correspond to the atoms in the QM and MM regions, respectively. In eq 1 the QM/MM coupling term, $H_{QM/MM}$, typically contains terms for the electrostatic, van der Waals (vdW), and bonded interactions (eq 2).[10−12]

$$H_{QM/MM} = H_{QM/MM(vdW)} + H_{QM/MM(elec)} + H_{QM/MM(bonded)} \tag{2}$$

The $H_{QM/MM(bonded)}$ term is required only where the partitioning into the QM and MM regions breaks covalent bonds. For such partitioning, the molecular mechanical bonding term is usually retained for interactions between covalently bonded QM and MM atoms, at the QM/MM boundary. The valency of the QM region is satisfied with the addition of link atoms[10−14] or by frozen orbital[6,15] or generalized hybrid orbital[17−21] approaches.

Several different methods have been used to describe the electrostatic interactions between the QM and MM regions, of which a so-called 'electrostatic embedding' scheme is the most common. In this model, interactions with the MM atomic point charges are included in the one-electron Hamiltonian of the QM region.[14,16] This model directly allows for the electronic polarization of the QM region by the MM environment. This is likely to be important to include in QM/MM studies of biological macromolecules because of their polar nature.

The vdW interaction between the QM and MM atoms in QM/MM calculations is typically included through a Lennard−Jones 12−6 potential,[11] as in standard biomolecular MM forcefields:[22,23]

$$V_{vdW}^{QM/MM} = \sum_A \sum_B 4\varepsilon_{AB} \left[\left(\frac{\sigma_{AB}}{R_{AB}} \right)^{12} - \left(\frac{\sigma_{AB}}{R_{AB}} \right)^6 \right] \tag{3}$$

where A and B are indices representing the QM and MM atoms, respectively, R_{AB} is the distance between the QM and MM atoms, and ε_{AB} and σ_{AB} are calculated from vdW parameters for each atom in the forcefield, using standard combination rules. In the widely used CHARMM27 forcefield, the Lorentz−Berthelot combination rules are used.[23] This nonelectrostatic interaction term represents dispersion attractions, that fall off as R^{-6}, and also prevents molecular collapse at short distances between the QM and MM atoms (the R^{-12} term is used for computational convenience). The vdW interaction term is written in a slightly different form in CHARMM:

$$V_{vdW}^{QM/MM} = \sum_A \sum_B \varepsilon_{AB} \left[\left(\frac{R_{min}^{A,B}}{R_{AB}} \right)^{12} - 2 \left(\frac{R_{min}^{A,B}}{R_{AB}} \right)^6 \right] \tag{4}$$

where

$$R_{min} = 2^{\frac{1}{6}} \sigma \tag{5}$$

and

$$\varepsilon_{AB} = \sqrt{\varepsilon_A \varepsilon_B} \tag{6}$$

Hence in this paper we optimize $R_{min}/2$ and ε.

Typical MM forcefields used in calculations on biomolecules use vdW parameters that have been optimized to describe bonded and nonbonded interactions or to reproduce experimental thermodynamic data for small molecules.[24] Derivation of vdW parameters can be a time-consuming and laborious process. It would be convenient to be able to use existing MM vdW parameters in QM/MM modeling. It is, however, possible that this could lead to significant errors in QM/MM calculations, which employ a different theoretical basis. In addition, the fact that electronic polarization (of the QM region by the MM region) is included in the QM/MM calculations but is modeled only indirectly, in an average way, by MM, could lead to optimal MM and QM/MM parameters being quite different. The suitability of vdW parameters for normal forcefield calculations for use in QM/MM calculations for different complexes has been studied earlier by several groups.[25−31] A systematic study of Lennard−Jones parameters and QM/MM hydrogen bonging energies was reported by Gao and Xia in 1992.[25] They calculated hydrogen bonding energies and geometries of 53 water complexes, covering functional groups on amino acids and nucleotide bases, using Monte Carlo AM1/TIP3P simulations. Comparison of AM1/TIP3P and ab initio 6−31G(d) results showed that adjustment of the OPLS Lennard−Jones parameters for H, C, N, and O in the QM region was necessary to get the best agreement between QM/MM and pure QM hydrogen bonding energies. This, however, resulted in a reduction in the accuracy of the geometries. The necessity to optimize vdW parameters for atoms in the QM region was also observed later, in a study of 45 organic small molecule−water complexes, using the AM1/TIP3P model.[26] Refinement of vdW parameters has also been found to be necessary when ab initio QM/MM methods are used.[27] When optimized vdW parameters were used, a good agreement between ab initio 3−21G and the MM OPLS -TIP3P potential and ab initio 6−31G(d) results was observed, for over 80 hydrogen bonded complexes of organic compounds with water. In the parametrization process it was also found that it was necessary to use different Lennard−Jones parameters for oxygen and nitrogen atoms in ionic and neutral molecules.[27] However, for hydrogen and carbon atoms, the same parameters were found to be suitable for both ionic and neutral complexes.[27] In addition to hydrogen bonded complexes, refinement of vdW parameters has also found to be necessary when chemical reactions in the condensed phase are studied with QM/MM methods (AM1/CHARMM).[28] More recently, Riccardi et al.[29] compared their optimized vdW parameters for the SCC-DFTB/CHARMM method with results for vdW parameters selected from the CHARMM22 forcefield.[23] While the different parameter sets gave clear differences in results for gas-phase clusters and solvent structure around the solutes, it was observed that thermodynamic quantities (activation free

energies and reduction potentials) in the condensed phase were not very sensitive to the vdW parameters used in the QM/MM calculations. In contrast, Luque and co-workers, who studied the hydrogen bonded complexes between various functional groups and a water molecule at B3LYP, AM1, and PM3 levels, observed that the vdW parameters given in normal forcefields are not transferable from MM to QM/MM calculations.[30] In addition, they pointed out that vdW parameters are sensitive to the QM/MM formalism and parametrization details and warned against the direct transferability of vdW parameters between different QM/MM methods.[30] However, in several studies optimization of vdW parameters has been observed to be necessary, for example Lennard−Jones parameters in the B3LYP/6−31G(d)/AMBER potential were found to accurately reproduce B3LYP/6−31G(d) hydrogen bond energies and geometries when amino acid−water complexes are studied.[31]

In the current work, we have studied the geometries and interaction energies of nucleic acid base−water complexes at a full QM level [B3LYP/6−311+G(d,p), using the popular B3LYP hybrid density functional technique] and an equivalent QM/MM [B3LYP/6−311+G(d,p)-CHARMM27] level. We have evaluated the suitability of the standard CHARMM27 forcefield vdW parameters for nucleic acid bases[23,24,32,33] for the QM/MM calculations at the B3LYP/6−311+G(d,p) and B3LYP/6−31G(d) levels. With these relatively high levels of QM theory, it is not yet feasible to calculate thermodynamic properties by simulations. Previous studies of the suitability of normal (MM) forcefield vdW parameters for QM/MM calculations have generally concentrated on complexes containing other small molecules and not nucleic acid bases[26−28,31] except for example the early works of Gao et al., where the AM1/TIP3P model was used to study hydrogen bonding energies and geometries of nucleic acid base−water complexes.[25,34] In those studies, good agreement between ab initio Hartree−Fock 6−31G(d) and AM1/TIP3P interaction energies was observed. The interactions distances from AM1/TIP3P calculations were also observed to be in reasonable agreement with the ab initio results. Here we apply significantly higher levels of theory than these early studies. Reliable models of the reactions and energetics of nucleic acids are crucial because of their vital roles in all living systems. For example, the reactions of catalytic RNA molecules, i.e. ribozymes,[35−38] interstrand cross-linking of DNA bases[39,40] (which is believed to be responsible for the biological activity of a number of antitumor agents[41]), stacking of nucleic acid bases in nucleic acid structures,[42,43] drug binding, and also reactivity[44−48] have been studied. Here, we report optimized vdW parameters for polar hydrogen, carbonyl carbon, and aromatic nitrogen atoms of nucleic acid bases by using a set of hydrogen bonded complexes at the B3LYP/6−311+G(d,p) QM and the B3LYP/6−311+G(d,p)-CHARMM27 QM/MM levels. In addition, the transferability of the optimized parameters to the lower B3LYP/6−31G(d)-CHARMM27 level, which is commonly used in QM/MM calculations,[7,49] is also tested. The suitability of the optimized nucleic acid base parameters was also tested on complexes other than those used in the parameter optimization. Also the suitability of the standard

CHARMM27[23,24,32,33] and optimized vdW parameters was tested for modeling chemical changes for simple reactions of nucleic acid bases.

Computational Details

QM calculations were performed using the Jaguar 6.0[50] and Gaussian03[51] programs. Full QM optimizations of nucleic acid base (Figure 1A)−water hydrogen bonded complexes shown in Figures 1B−E, 2, and 3 were performed at the B3LYP/6−311+G(d,p) level of theory. Work by Jorgensen and co-workers has shown that the B3LYP/6−311+G(d,p) level of theory (compared to electron correlated ab initio methods) treats hydrogen-bonded complexes well.[52] The counterpoise correction scheme of Boys and Bernardi[53] was used as an estimate to correct for basis set superposition error (BSSE) in the calculated QM interaction energies.

In the QM/MM calculations, the QM calculations were performed using Jaguar 5.0[50] at the B3LYP/6−311+G(d,p) or B3LYP/6−31G(d) level. The Tinker molecular mechanics program[54] with the CHARMM27 all-atom forcefield[23,24,32,33] was used to evaluate the MM terms. The results from the QM and the MM calculations were combined using the QM/MM interface program QoMMMa.[55] The QoMMMa program optimizes the QM geometry within the MM environment, and the positions of the MM atoms are fully relaxed at each QM step using the Tinker program.[54] In the nucleic acid−water complexes, the nucleic acid base was always defined as the QM region and the water molecule as the MM region. In nucleic acid base pair calculations, one base was defined as the QM region and the other as the MM region. MM calculations were performed with the CHARMM program.[56,57]

Parameter Optimization

The vdW parameters of polar hydrogen, carbonyl carbon, and aromatic nitrogen atoms in four nucleic acid bases (Figure 1A) were optimized. The Lennard−Jones parameters (the minimum in the vdW interaction energy surface $R_{min}/2$ and the well depth, ε) for the QM atoms were adjusted systematically to obtain the best agreement for hydrogen bond energies and geometries between the QM/MM and the full QM calculations. At the beginning of the optimization, an increment of 0.1 Å for $R_{min}/2$ and 0.01 kcal/mol for ε were used. As we got closer to hydrogen bond energies and geometries obtained from the full QM calculations, smaller increments were used. Both $R_{min}/2$ and ε were adjusted simultaneously. The best parameters were determined by eye, and if it was difficult to get parameters which gave both optimal hydrogen bond energies and geometries, then parameters were chosen to give optimal hydrogen bond energy rather than geometry. Two different types of parameters, base-dependent and base-independent, have been developed here. By base-dependent parameters, we mean that unique parameters for each atom type were developed for each nucleic acid base. Base-independent parameters are, in contrast, the same for each nucleic acid base, i.e., dependent only on atom type. Parameters of atoms in the MM region, (i.e., the atomic charges and Lennard−Jones parameters in
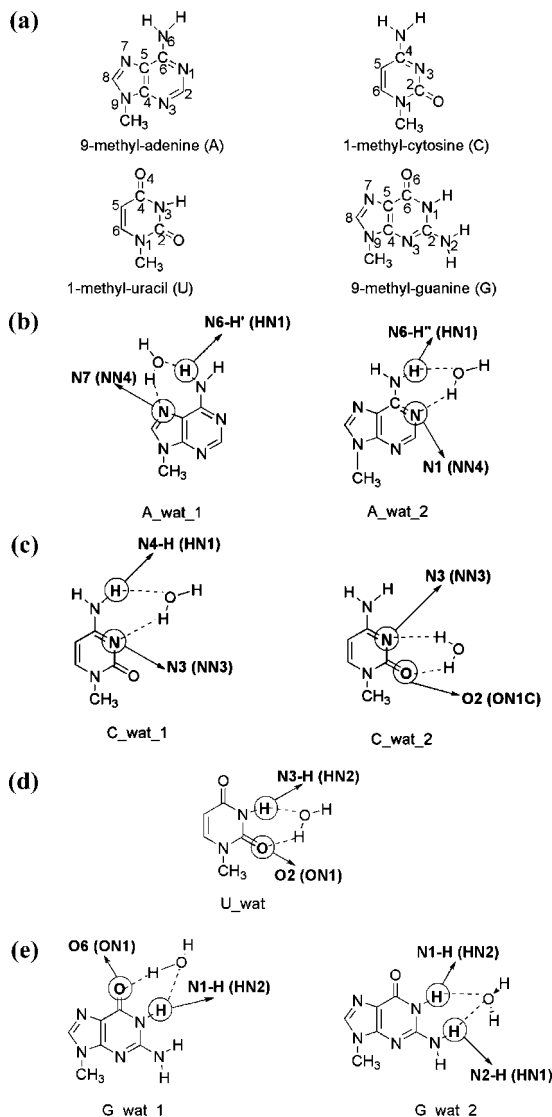
Figure 1. A. The four nucleic acid bases used for parameter optimization. The structures also show atom numbering. B. The two hydrogen bonded complexes of adenine are shown here. The bold circled atoms are the two QM nucleic acid base atoms that form hydrogen bonds (see Figure 1A for atom numbering), and it is for these atoms that parameters are optimized. The CHARMM27 atom types for these hydrogen bonding atoms are given in parentheses. C. The two hydrogen bonded complexes of cytosine are shown here. The bold circled atoms are the two QM nucleic acid base atoms that form hydrogen bonds (see Figure 1A for atom numbering), and it is for these atoms that parameters are optimized. The CHARMM27 atom types for these hydrogen bonding atoms are given in parentheses. D. The hydrogen bonded complex of uracil is shown here (used as a model for thymine). The bold circled atoms are the two QM nucleic acid base atoms that form hydrogen bonds (see Figure 1A for atom numbering), and it is for these atoms that parameters are optimized. The CHARMM27 atom types for these hydrogen bonding atoms are given in parentheses. E. The two hydrogen bonded complexes of guanine are shown here. The bold circled atoms are the two QM nucleic acid base atoms that form hydrogen bonds (see Figure 1A for atom numbering), and it is for these atoms that parameters are optimized. The CHARMM27 atom types for these hydrogen bonding atoms are given in parentheses.
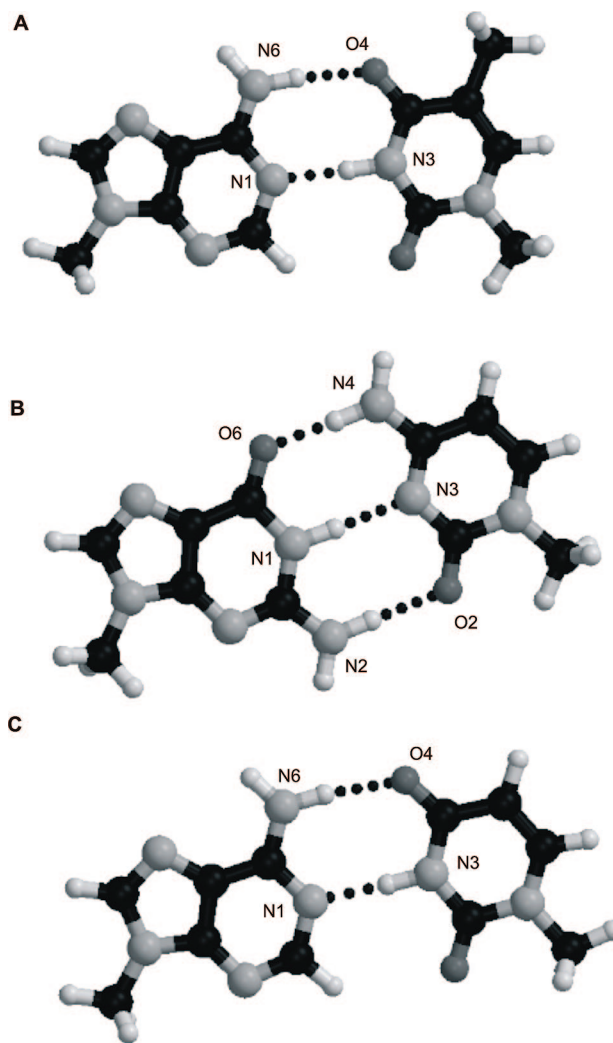

Figure 2. Structures of Watson Crick base pairs used in vdW parameter testing. A. Adenosine-thymine (AT) base pair. B. Guanosine-cytidine (GC) base pair. C. Adenosine-uracil (AU) base pair. Coloring of atoms: carbon atoms are in black, hydrogen atoms are in white, nitrogen atoms are in light gray, and oxygen atoms are in dark gray. The atom numbers of the atoms involved in the hydrogen bonding are shown on the structures (for atom numbering, see Figure 1A).

the CHARMM version of the TIP3P water model (TIPS3P)[58] (which is modified to include vdW parameters for the water hydrogen atoms[12,59]) were kept unchanged, to ensure transferability and consistency with the CHARMM27 force-field. Geometries and interaction energies were also calculated using the optimized (base-dependent and independent) parameters at the lower B3LYP/6−31G(d) QM/MM level.

Results and Discussion

Complexes Used in vdW Parameter Optimization. To model nucleosides, nucleic acid bases with a methyl group in place of the 1′ carbon of the ribose ring were used (Figure 1A). The structures of the nucleic acid base−water complexes used in vdW parameter optimization are shown in Figures 1B−E. Uracil (U, Figure 1E) was used as a model for both uracil and thymine because uracil (found in RNA) is smaller than thymine (found in DNA). All the complexes
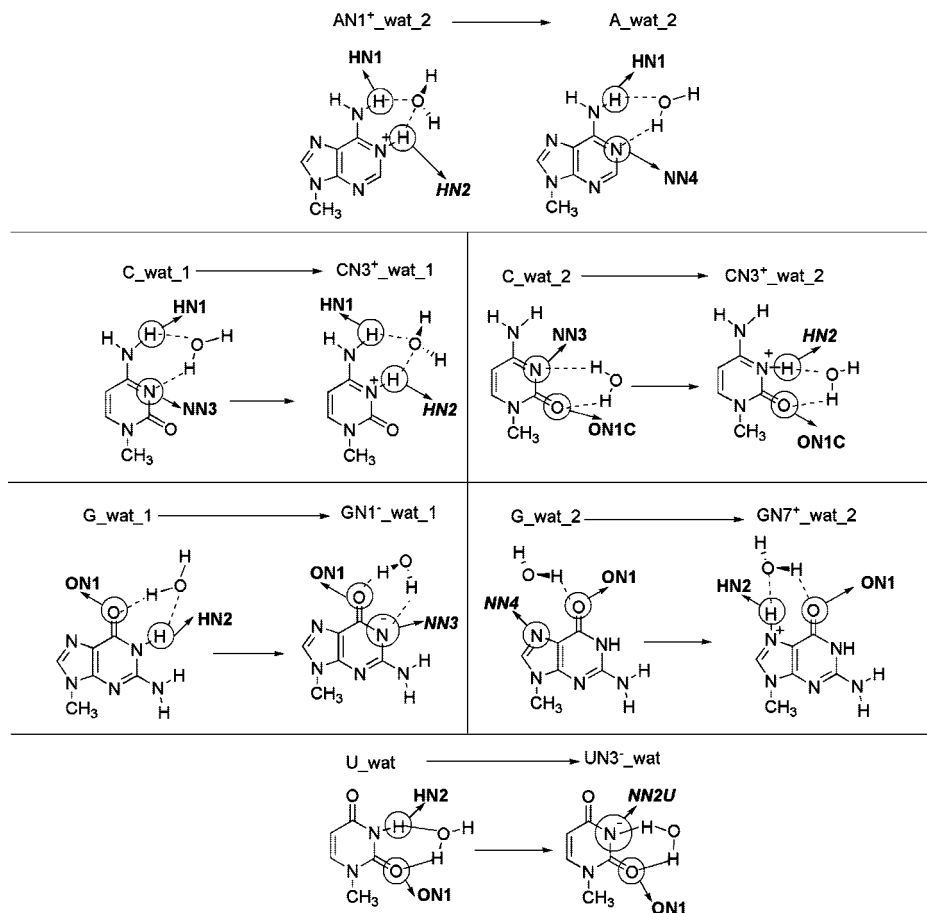
Figure 3. Protonation and deprotonation reactions studied with the newly developed vdW parameters. CHARMM27 atom types for nucleic acid base atoms involved in the reaction are shown (circled atoms). For atom types shown in italics, standard CHARMM27 parameters were used, as no base-dependent parameters were available. For atom numbering, see Figure 1A.

were constructed so that they contained two hydrogen bonds between the water molecule and the nucleic acid base, so that either the oxygen atom of the water molecule interacts with either the hydrogen atom of the exocyclic or endocyclic amino group, or the hydrogen atom of the water molecule interacts with either the carbonyl oxygen atom or the ring nitrogen atom of the nucleic acid base. For adenosine (A, Figure 1B), cytosine (C, Figure 1C), and guanosine (G, Figure 1D), two different water complexes were constructed (denoted A_wat_1, A_wat_2, etc.). In this way, vdW parameters for all different nitrogen [adenosine and guanosine N7 (CHARMM atom type NN4), adenosine N1 (NN4), cytosine N3 (NN3), see Figures 1B−D], hydrogen [adenosine N6-H (HN1), cytosine N4-H (HN1), uracil N3-H (HN2), guanosine N2-H (HN2), see Figures 1B−E], and oxygen [uracil O2 (ON1) and guanosine O6 (ON1)); see Figures 1E and 1D] atom types found in nucleic acid bases could be studied.

Comparison of QM and MM Geometries and Interaction Energies with QM/MM Results Using Standard CHARMM27 vdW Parameters. The hydrogen bond distances for the nucleic acid base−water complexes obtained from the QM [B3LYP/6−311+G(d,p)] and QM/MM [B3LYP/6−311G+(d,p)-CHARMM27] optimizations are listed in Table 1. Generally, the hydrogen bond distances are shorter (by 0.02−0.22 Å, measured between the heavy atom and the hydrogen atom) in the QM/MM optimized

structures than in the full QM optimized structures, consistent with earlier studies.[29,30,34] However, in some cases, the opposite situation is observed: in the A_wat_2 complex (between N6-H″ and water), in the C_wat_2 complex (between N3 and water), and in the U_wat complex (between O2 and water), the hydrogen bonds are clearly longer (by 0.3−0.5 Å) in the QM/MM optimized complexes than in the QM optimized complexes (Table 1; see Figures 1B−E for details of complexes). In the A_wat_1 complex, the hydrogen bond between N7 and water is also slightly longer (by 0.03 Å) in the QM/MM optimized complex than in the QM optimized complex (Table 1). The largest deviation between the QM and QM/MM optimized structures is seen for the C_wat_2 complex, where the distance between the nitrogen N3 atom and the hydrogen of the water molecule is 3.4 Å in the QM/MM optimized structure but only 2.9 Å in the QM optimized structure. However, this interaction is not a strong hydrogen bond, being too long. A similar observation was earlier made by Freindorf et al., who observed that the largest discrepancies in hydrogen bond distances optimized at QM [B3LYP/6−31+G(d)] and QM/MM [B3LYP/6−31+G(d)-AMBER] levels were observed when the A···H distance is longer than 2.5 Å, i.e., having more long-range electrostatic interaction character than hydrogen bond character.[31] In many (but not all) cases, the conformations of NH2 groups are planar but slightly pyramidal in some QM (and QM/MM) calculations. As the

Table 1. Hydrogen Bond Distances (H-acceptor distances, Å) Calculated at the B3LYP/6-311+G(d,p) QM and B3LYP/6-311+G(d,p)-QM/MM Levels Using Either CHARMM27, Base-Dependent Parameters (Table 3), or Base-Independent Parameters (Table 4) for Complexes Shown in Figures 1B–E[a]

complex[b]	H-bond[c]	QM	QM/MM (CHARMM)[d]	QM/MM (OPT-BI)[e]	QM/MM (OPT-BD)[f]	MM (CHARMM)[d]
A_wat_1	N7−Hwat	1.89	1.92 (−0.03)	1.95 (−0.07)	1.95 (−0.07)	1.89 (0.00)
A_wat_1	N6-H′−Owat	1.95	1.80 (0.16)	2.00 (−0.05)	2.00 (−0.05)	1.88 (0.07)
A_wat_2	N6-H″−Owat	2.05	2.38 (−0.33)	2.39 (−0.34)	2.39 (−0.34)	2.01 (0.04)
A_wat_2	N1−Hwat	1.92	1.93 (−0.01)	1.93 (−0.01)	1.93 (−0.01)	1.90 (0.02)
C_wat_1	N4-H−Owat	2.01	1.93 (0.08)	2.11 (−0.10)	1.99 (0.01)	2.01 (0.00)
C_wat_1	N3−Hwat	1.92	1.90 (0.02)	1.91 (0.01)	1.94 (−0.02)	1.95 (−0.3)
C_wat_2	N3−Hwat	2.90	3.40 (−0.50)	3.53 (−0.63)	3.36 (−0.46)	1.94 (0.96)
C_wat_2	O2−Hwat	1.95	1.79 (0.16)	1.85 (0.10)	2.12 (−0.17)	2.39 (−0.44)
G_wat_1	O6−Hwat	1.86	1.80 (0.06)	1.87 (−0.01)	1.91 (−0.05)	1.87 (−0.01)
G_wat_1	N3-H−Owat	1.95	1.93 (0.01)	2.08 (−0.13)	2.19 (−0.25)	1.99 (−0.04)
G_wat_2	N1-H−Owat	2.10	1.93 (0.16)	2.07 (0.02)	2.17 (−0.07)	2.08 (0.02)
G_wat_2	N2-H−Owat	2.21	2.12 (0.09)	2.14 (0.08)	2.22 (−0.01)	2.00 (0.21)
U_wat	N3-H−Owat	2.00	1.79 (0.22)	1.98 (0.02)	1.97 (0.03)	1.99 (0.01)
U_wat	O2−Hwat	1.97	2.25 (−0.28)	2.23 (−0.27)	2.21 (−0.24)	1.94 (0.02)
STDEV[g]			0.21	0.22	0.15	0.31
MSE[h]			0.01	−0.01	0.12	−0.04
MUE[i]			0.15	0.13	0.13	0.15

[a] Numbers in parentheses are the deviation from the fully QM calculation {Δ[QM − (QM/MM)]}. The hydrogen bond distances are calculated between hydrogen atoms and hydrogen bond acceptors. [b] See Figures 1B–E for names of complexes. [c] See Figure 1A for labeling. [d] Standard CHARMM27 vdW parameters were used. [e] Optimized base-independent parameters were used. [f] Optimized base-dependent parameters were used. [g] Standard deviation of the error from the full QM calculated value. [h] Mean signed error (MSE) from the full QM calculated value, MSE = $1/n \sum_{i=1}^{n} (f_i - y_i)$, f_i = calculated value, y_i = QM or experimental value. [i] Mean unsigned error (MUE) from the full QM calculated value, MUE = $1/n \sum_{i=1}^{n} |f_i - y_i|$.

Table 2. Interaction Energies (kcal/mol) Calculated at the B3LYP/6-311+G(d,p) QM and B3LYP/6-311+G(d,p)-QM/MM Levels Using Either CHARMM27, Base-Dependent Parameters (Table 3) or Base-Independent Parameters (Table 4) for Complexes Shown in Figures 1B–E[a]

complex[b]	QM	QM/MM (CHARMM)[c]	QM/MM (OPT-BI)[d]	QM/MM (OPT-BD)[e]	MM (CHARMM)[c]
A_wat_1	−9.9	−13.3 (3.4)	−11.9 (2.0)	−12.6 (2.7)	−10.4 (0.5)
A_wat_2	−8.9	−10.0 (1.1)	−10.2 (1.2)	−10.3 (1.4)	−9.5 (0.6)
C_wat_1	−10.5	−13.8 (3.2)	−12.9 (2.4)	−12.7 (2.3)	−11.1 (0.6)
C_wat_2	−7.0	−9.1 (2.1)	−8.7 (1.7)	−7.4 (0.4)	−10.4 (3.4)
G_wat_1	−10.8	−14.1 (3.3)	−12.8 (2.0)	−11.3 (0.6)	−11.5 (0.7)
G_wat_2	−7.7	−10.8 (3.1)	−10.6 (2.9)	−9.7 (2.1)	−10.9 (3.2)
U_wat	−7.5	−10.4 (2.9)	−9.3 (1.8)	−9.3 (1.8)	−9.6 (1.1)
STDEV[f]		0.8	0.5	0.9	1.3
MSE[g]		−2.7	−2.0	−1.6	−1.4
MUE[h]		2.7	2.0	1.6	1.4

[a] Numbers in parentheses are the deviation from the fully QM calculation {Δ[QM − (QM/MM)]}. [b] See Figure 1B–E for names of complexes. [c] Standard CHARMM27 vdW parameters were used. [d] Optimized base-independent parameters were used. [e] Optimized base-dependent parameters were used. [f] Standard deviation of the error from the full QM calculated value. [g] Mean signed error of from the full QM calculated value, see notes in Table 1. [h] Mean unsigned error from the full QM calculated value, see notes in Table 1.

QM data are the target for optimization, and the pyramidality is not affected by the vdW parameters (and has little effect on hydrogen bond strength), this is not an important factor here.

The pure MM (CHARMM27) optimized hydrogen bond distances are very close to the full QM optimized results (Table 1). Only in the C_wat_2 complex is there a large difference between the MM and QM structures: the N3−Hwat hydrogen bond distance is 0.96 Å shorter in the MM optimized structure than in the QM optimized structure, and the O2−Hwat hydrogen bond is 0.44 Å longer in the MM optimized structure than in the QM optimized structure (Table 1, see Figures 1C for details of complex).

The hydrogen bond angles, θ(D−H···A), calculated at the full QM and QM/MM levels, are usually within ∼5° of each other, although, in the U_wat complex and in the C_wat_2 complex (hydrogen bond from water to O2), a larger deviation (∼12°) is seen (see Supporting Information, Table 12). Comparison of the MM optimized structures with

the QM optimized structures shows a somewhat larger deviation in the hydrogen bond angles than seen for the comparison of the QM/MM results with the pure QM results (see Supporting Information, Table 12).

The QM/MM interaction energies are overestimated in all cases when using standard CHARMM27 vdW parameters,[23,24,32,33] compared to those calculated at the pure QM level (Table 2). In most cases, the interaction energies are overestimated by around 3 kcal/mol (2.9 to 3.4 kcal/mol), but in the A_wat_2 complex, the difference is only 1.1 kcal/mol and in C_wat_2 it is 2.1 kcal/mol. It is interesting that in A_wat_2, the interaction energy is overestimated in the QM/MM calculations compared to the QM calculations, even though both hydrogen bond distances are longer in the QM/MM optimized structure than in the QM optimized structure. The complexes contain examples of QM groups as hydrogen bond acceptors and donors. There does not seem to be any clear difference in the deviations in hydrogen bond energies depending on whether the donor is QM or MM.

402 *J. Chem. Theory Comput., Vol. 5, No. 2, 2009*

Pentikäinen et al.

Table 3. Standard CHARMM27 [23; 27; 28] and Optimized Base-Dependent vdW Parameters ($R_{min}/2$ and ε) for Nitrogen Polar, Hydrogen, and Carbonyl Oxygen Atoms for Nucleic Acid Bases

atom type[a]	CHARMM		OPT	
	$R_{min}/2$ (Å)	ε (kcal/mol)	$R_{min}/2$ (Å)	ε (kcal/mol^{-1})
	Uracil/Thymine			
HN2	0.2245	−0.046	0.7	−0.046
ON1	1.7	0.120	1.7	0.120
	Cytosine			
NN3	1.85	−0.200	1.95	−0.200
HN1	0.2245	−0.046	0.4245	−0.0460
ON1C	1.7	−0.120	1.9	−0.120
	Guanine			
ON1	1.7	−0.120	1.8	−0.086
HN1	0.2245	−0.046	0.8245	−0.086
HN2	0.2245	−0.046	0.8245	−0.086
	Adenine			
NN4	1.85	−0.200	1.9	−0.200
HN1	0.2245	−0.046	0.7	−0.046

[a] See Figures 1B−E for atom types.

The (CHARMM27) MM interaction energies are closer than the QM/MM interaction energies to the QM interaction energies (Table 2). In five complexes out of seven, the MM interaction energies are within 0.5−1.1 kcal/mol of the full QM interaction energies. This again demonstrates that the CHARMM27 parameters provide a good description of nucleic acid interactions in MM calculations. In two complexes, namely C_wat_2 and G_wat_2, the deviation from the QM interaction energies is larger, 3.4 and 3.2 kcal/mol, respectively.

Optimized Base-Dependent vdW Parameters. As described above, both hydrogen bond distances (Table 1) and interaction energies (Table 2) calculated at the QM/MM level with standard CHARMM27 Lennard−Jones parameters differed significantly from the full QM results. We therefore tested the effects of varying the vdW parameters for the nucleic acid bases treated with QM in the QM/MM calculations, in an attempt to improve the agreement, particularly for interaction energies. The vdW parameters of nitrogen (CHARMM atom types NN3 and NN4), polar hydrogen (atom types HN1 and HN2), and carbonyl oxygen (atom types ON1 and ON1C) atoms of the nucleic acid bases were changed systematically. The vdW parameters of the water molecule forming the MM part of the system were always kept unchanged (CHARMM TIPS3P[58]), for consistency with the CHARMM27 forcefield,[23,24,32,33] and because this represents likely interactions in typical QM/MM studies of nucleic acid bases (i.e., nucleic acid base treated by QM and water with MM).

To get the best agreement between the hydrogen bond distances and interaction energies calculated at the QM and QM/MM levels, base-dependent vdW parameters (i.e., different atomic parameters for different nucleic acid bases), were developed. These base-dependent parameters are listed in Table 3. Each nucleic acid base−water complex contained two hydrogen bonds between water and the nucleic acid base, which makes vdW parameter optimization challenging. The adenosine−water complexes, A_wat_1 and A_wat_2 (Figure 1B), were especially difficult. This was due to the fact that

when the original CHARMM27 vdW parameters were used, the hydrogen bond distance between HN1 and water in the complex A_wat_1 is shorter in the QM/MM optimized structure than in the QM optimized structure, whereas in the complex A_wat_2 the situation is reversed (Table 1). As the error in the interaction energy between the original QM and the QM/MM calculations is larger for the A_wat_1 complex than the A_wat_2 complex (Table 2), the hydrogen bond distance between HN1 and water in A_wat_1 complex was optimized rather than that in the A_wat_2 complex. A similar situation is also seen with C_wat_1 and C_wat_2. When the standard CHARMM27 vdW parameters were used, in the C_wat_1 complex the distance between NN3 and water is shorter in the QM/MM optimized structure than in the QM optimized structure, while in the C_wat_2 complex the situation is the opposite (Table 1).

As can be seen from Table 3, where the optimized base-dependent parameters are listed, the largest change was made for the $R_{min}/2$ value of hydrogen atoms. The original CHARMM27[23,24,32,33] value of 0.2245 Å is increased to 0.4245 Å on cytosine HN1, 0.7 Å on uracil/thymine HN2 and adenine HN1, and 0.8245 Å on guanine HN1 and HN2. The ε values for guanine HN1 and HN2 were also changed from −0.046 to −0.086 kcal/mol. VdW parameters for nitrogen and oxygen atoms were changed only slightly: the $R_{min}/2$ value for cytosine nitrogen (NN3) was changed from 1.85 Å to 1.95 Å, for cytosine oxygen (ON1C) from 1.7 Å to 1.9 Å, guanine oxygen (ON1) from 1.7 Å to 1.8 Å, and adenine nitrogen (NN4) from 1.85 Å to 1.9 Å. In addition, the ε value of the guanine oxygen was made slightly less negative, changed from −0.12 to −0.086 kcal/mol. Otherwise the change of the ε value did not have any meaningful effect, either on the hydrogen bond lengths or the interaction energies, and so they are left at the original CHARMM27 values.

The hydrogen bond distances obtained with the optimized base-dependent vdW parameters, as well as the differences in hydrogen bond lengths between the QM optimized and QM/MM optimized structures using both standard CHARMM27 parameters and optimized base-dependent parameters, are shown in Table 1. In six cases, the hydrogen bond distances obtained were improved, compared to those obtained with the standard CHARMM27 vdW parameters. In five cases, the optimization of vdW parameters did not have any substantial effect and in three cases the hydrogen bond distances, compared to the full QM results, were worse than when the original parameters were used. This illustrates the problem of optimization of parameters for systems having two or more hydrogen bonds.

The interaction energies calculated using the optimized base-dependent parameters are generally overestimated by 0.4−2.7 kcal/mol (Table 2), compared to those calculated at the full QM level. This is a significant improvement, by 0.7 to 2.8 kcal/mol, compared to the interaction energies calculated using the standard CHARMM27 parameters (Table 2). The A_wat_2 complex was the only one where the difference in the interaction energies between the QM and the QM/MM calculations became worse with optimized parameters. This was because the hydrogen bond distance

B3LYP/CHARMM27 QM/MM Modeling of Nucleic Acid Bases

J. Chem. Theory Comput., Vol. 5, No. 2, 2009 **403**

Table 4. Standard CHARMM27[23; 27; 28] and Optimized Base-Independent vdW Parameters for Nitrogen, Polar Hydrogen, and Carbonyl Oxygen in Nucleic Acid Bases

	CHARMM		OPT	
atom type[a]	$R_{min}/2$ (Å)	ε (kcal/mol)	$R_{min}/2$ (Å)	ε (kcal/mol)
HN1	0.2245	−0.046	0.7	−0.046
HN2	0.2245	−0.046	0.7	−0.046
NN3	1.85	−0.2	1.9	−0.2
NN4	1.85	−0.2	1.9	−0.2
ON1C	1.7	−0.12	1.8	−0.12
ON1	1.7	−0.12	1.8	−0.12

[a] See Figures 1B−E for atom types.

between HN1−water was longer in the QM/MM optimized A_wat_2 structure and was shorter in the A_wat_1 complex, as explained above. However, in the A_wat_2 complex, the differences between the calculated QM and QM/MM interaction energies are very small, regardless of which parameters are used (Table 2).

The hydrogen bond angles (D−H···A) with the optimized base-dependent vdW parameters changed only slightly from those obtained with the standard CHARMM27 vdW parameters (see Supporting Information, Table 12). In some cases the difference between the hydrogen bond angles, calculated at the full QM level and at the QM/MM level, were decreased and in some cases increased with the optimized base-dependent vdW parameters. Given the relative lack of sensitivity of hydrogen bond angles to the parameters, hydrogen bond angles were not considered specifically during the vdW parameter optimization process.

Optimized Base-Independent vdW Parameters. A typical approach for the development of MM forcefield parameters is that the MM parameters, optimized for small molecules or fragments, are then used in large compounds, in different molecular environments. Accordingly, the same vdW parameters are used, for example, for carbonyl oxygen atoms in all nucleic acid bases. Therefore, starting from the base-dependent parameters listed in Table 3, base-independent parameters were developed (shown in Table 4). In this set of parameters, $R_{min}/2 = 0.7$ Å and $\varepsilon = −0.046$ kcal/mol were used for polar hydrogen atoms. For nitrogen and oxygen atoms, the optimized $R_{min}/2$ and ε parameters are 1.85 Å and −0.2 kcal/mol, and 1.7 Å and −0.12 kcal/mol, respectively (Table 4).

The hydrogen bond distances and interaction energies calculated using the base-independent parameters are listed in Tables 1 and 2, respectively. Hydrogen bond angles are reported in Table 12, Supporting Information. With the base-independent parameters, the deviation of the QM/MM calculated interaction energies and hydrogen bond distances from the QM values is slightly larger than that seen with base-dependent parameters, as is to be expected. Hydrogen bond distances found with the base-independent parameters differ by 0.01 to 0.6 Å (Table 1) from the QM values, while with the base-dependent parameters (Table 4), the deviation is 0.01−0.5 Å (Table 1). The interaction energies deviate by 1.2−2.9 kcal/mol (Table 2) from those calculated at the full QM level, whereas with base-dependent parameters, the deviation is 0.4−2.7 kcal/mol (Table 2). However, with the base-independent parameters, the interaction energies are still

significantly closer to the full QM interaction energies than when standard CHARMM27 parameters are used, as clearly seen from Table 2. Interestingly, the base-independent parameters provide the best agreement with hydrogen bond angles between the QM and QM/MM calculated structures (see Supporting Information Table 11).

Transferability of the Optimized Parameters to the B3LYP/6−31G(d) Level. QM/MM calculations are often performed at the B3LYP/6−31G(d) level. This is partly because this level of theory gives reasonable descriptions of many (e.g., biomolecular) systems,[6,7,31,49,55] while being computationally less demanding than, for exaxample, the B3LYP/6−311+G(d,p) level, but also because diffuse functions can cause QM/MM calculations to be unstable.[60] The transferability of the optimized vdW parameters [both base-dependent (Table 3) and independent (Table 4)] to the widely used B3LYP/6−31G(d) QM/MM level was tested. The results show that the vdW parameters optimized at the B3LYP/6−311+G(d,p) level can be transferred to the lower level. Comparison of the interaction energies calculated at the B3LYP/6−31G(d) QM level and B3LYP/6−31G(d)/ CHARMM27 QM/MM levels shows that with both base-dependent and base-independent parameters, very similar values are obtained (Table 5). The deviation between the QM and QM/MM interaction energies is only 0.1−1.4 kcal/mol with base-dependent parameters (Table 3) and 0.1−3.3 kcal/mol with base-independent parameters (Table 4). Accordingly, especially with base-dependent parameters, the agreement between the QM and QM/MM interaction energies is very good. Table 5 also clearly shows that both sets of optimized parameters give relatively consistent QM/MM interaction energies that are closer to the QM results than when standard CHARMM27 parameters are used. Comparison of the hydrogen bond distances also shows that the QM and QM/MM geometries are very similar, as seen from Table 6. However, the differences between the hydrogen bond distances calculated at QM and QM/MM levels are larger when the 6−31G(d) basis is used instead of the 6−311+G(d,p) basis. In addition, although better agreement with the QM interaction energies is achieved with the optimized parameters than with the standard CHARMM27 parameters, no significant improvement is observed in QM/MM geometries with the optimized parameters when the 6−31G(d) basis is used (Table 6 for hydrogen bond distances and Supporting Information, Table 13 for hydrogen bond angles).

Nucleic Acid Base Pairs. The standard Watson−Crick base pairs [adenosine-uracil (AU), guanosine-cytosine (GC), and adenosine-thymine (AT); Figure 2] were used to test the transferability of the newly optimized parameters to different structures. For nucleic acid base pairs, it is only possible to easily use the base-independent parameters (Table 4), because in the standard Tinker format[54] (which QoM-MMa[55] uses to evaluate the MM terms) the same atom types are found in different bases (see Figures 1B−E for details of atom types in different nucleic acid bases). Calculations were performed for each base pair in both possible QM/ MM combinations, i.e., first with one base treated QM and the second by MM, then separately with the other base QM (and the first MM). Hydrogen bond distances (measured

Table 5. Interaction Energies (kcal/mol) Calculated at the B3LYP/6-31G(d) QM and B3LYP/6-31G(d)-QM/MM Levels, Using Base-Dependent Parameters (Table *3*), Base-Independent Parameters (Table *4*), or the Original CHARMM27 Parameters for the Complexes Shown in Figures *1*B−E[a]

complex[b]	QM	QM/MM (OPT-BI)[c]	QM/MM (OPT-BD)[d]	QM/MM (CHARMM)[e]
A_wat_1	−11.1	−11.6 (0.5)	−11.6 (0.5)	−13.4 (2.3)
A_wat_2	−9.6	−9.7 (0.1)	−9.7 (0.1)	−9.8 (0.2)
C_wat_1	−11.1	−12.1 (0.9)	−11.8 (0.7)	−12.9 (1.9)
C_wat_2	−6.9	−10.2 (3.3)	−6.7 (−0.3)	−8.6 (1.7)
G_wat_1	−11.4	−11.9 (0.5)	−11.9 (0.5)	−13.5 (2.1)
G_wat_2	−11.9	−9.8 (−2.1)	−9.3 (−2.6)	−10.5 (−1.4)
U_wat	−8.3	−8.3 (0.0)	−9.7 (1.4)	−9.4 (1.1)
STDEV[g]		1.6	1.3	1.3
MSE[h]		−0.5	−0.0	−1.1
MUE[h]		1.1	0.9	1.5

[a] Numbers in parentheses are the deviation from the fully QM calculation {Δ[QM − (QM/MM)]}. [b] See Figures 1B−E for names of complexes. [c] Optimized base-independent parameters were used. [d] Optimized base-dependent parameters were used. [e] Original CHARMM parameters were used. [f] Standard deviation of the error from the full QM calculated value. [g] Mean signed error from the full QM calculated value, see notes in Table 1. [h] Mean unsigned error from the full QM calculated value, see notes in Table 1.

Table 6. Hydrogen Bond Distances (Å) Calculated at the B3LYP/6-31G(d) QM and B3LYP/6-31G(d)-QM/MM Levels, Using Either Base-Dependent Parameters (Table *3*), Base-Independent Parameters (Table *4*), or the Original CHARMM27 Parameters, For the Complexes Shown in Figures *1*B−E[a]

complex[b]	H-bond[c]	QM	QM/MM (OPT-BI)[d]	QM/MM (OPT-BD)[e]	QM/MM (CHARMM)[f]
A_wat_1	N7−Hwat	1.90	1.87 (0.03)	1.87 (0.03)	1.85 (0.05)
A_wat_1	N6-H"−Owat	1.89	2.01 (−0.12)	2.01 (−0.12)	1.80 (0.09)
A_wat_2	N6-H″−Owat	1.96	2.42 (−0.46)	2.42 (−0.46)	2.39 (−0.43)
A_wat_2	N1−Hwat	1.95	1.85 (0.10)	1.85 (0.10)	1.90 (0.05)
C_wat1	N4-H−Owat	1.92	2.13 (−0.21)	2.06 (−0.14)	1.98 (−0.06)
C_wat1	N3−Hwat	1.94	1.90 (0.03)	1.92 (0.02)	1.88 (0.06)
C_wat2	N3−Hwat	2.18	1.95 (0.22)	3.37 (−1.20)	3.46 (−1.28)
C_wat2	O2−Hwat	2.24	2.70 (−0.46)	2.12 (0.12)	1.74 (0.50)
G_wat_1	O6−Hwat	1.86	1.85 (0.01)	1.78 (0.07)	1.76 (0.10)
G_wat_1	N3-H−Owat	1.88	2.12 (−0.23)	2.18 (−0.30)	1.96 (−0.08)
G_wat_2	N1-H−Owat	1.90	2.13 (−0.23)	2.22 (−0.32)	2.05 (−0.15)
G_wat_2	N2-H−Owat	2.52	2.13 (0.39)	2.18 (0.34)	1.96 (0.56)
U_wat	N3-H−Owat	1.93	1.02 (0.90)	2.14 (−0.22)	1.81 (0.12)
U_wat	O2−Hwat	1.94	2.23 (−0.29)	1.83 (0.11)	2.24 (−0.30)
STDEV[g]			0.36	0.37	0.44
MSE[h]			0.02	0.14	0.06
MUE[g]			0.27	0.25	0.27

[a] Numbers in parentheses are the deviation from the fully QM calculation {Δ[QM − (QM/MM)]}. The hydrogen bond distances are calculated between hydrogen atoms and hydrogen bond acceptor. [b] See Figures 1B−E for names of complexes. [c] See Figure 1A for labeling. [d] Optimized base-independent parameters were used. [e] Optimized base-dependent parameters were used. [f] Standard CHARMM parameters were used. [g] Standard deviation of the error from the full QM calculated value. [h] Mean signed error from the full QM calculated value, see notes in Table 1. [g] Mean unsigned error from the full QM calculated value, see notes in Table 1.

between heavy atoms) between the base pairs AU and CG obtained from QM, QM/MM, and MM (CHARMM27) calculations are very similar to those observed in crystal structures[61] (Table 7). For the AT base pair, no crystallographic structure is currently available. The hydrogen bond distances in base pair structures optimized using pure QM [at the B3LYP/6−311+G(d,p) level of theory] deviate from experimental values[61] by 0.01−0.11 Å (Table 7). Similar differences from experimentally determined hydrogen bond distances are also observed in the pure MM optimized structures (Table 7). In the QM/MM optimized structures, hydrogen bond distances seem to deviate more from experimental values, compared to the pure QM and pure MM optimized structures. This is not affected by the vdW parameters or the level of QM/MM theory used (Table 7). The AT base pair geometry calculated at the fully QM level is very similar to that of the fully QM optimized AU base pair structure. The deviations of the MM and QM/MM optimized geometries from the fully QM optimized structures are very similar to those seen with AU and CG base pairs.

Comparison of the interaction energies (Table 8) shows that at the QM/MM level, our optimized base-independent parameters give better agreement with the full QM interaction energies than the standard CHARMM27 vdW parameters.[23,24,32,33] With the B3LYP/6−311+G(d,p) level of theory and optimized base-independent parameters, the QM/MM interaction energies in five complexes out of six are within 0.8−1.8 kcal/mol of the QM interaction energies (Table 8). The largest deviation is seen for the AT base pair, when thymine is treated with QM. In this case, the interaction energy is overestimated by 4.4 kcal/mol. With the lower [B3LYP/6−311G(d)] level of QM/MM theory, for the same base pair, some overestimation is also observed but to a smaller extent (only 2.3 kcal/mol). When standard CHARMM27 parameters[23,24,32,33] are used in QM/MM calculations, the deviation of QM/MM interaction energies from the QM interaction energies is 2.4−6.3 kcal/mol (Table 8). The MM interaction energies (with standard CHARMM27 parameters) are very close (within 1−2.5 kcal/mol) to the QM interaction

Table 7. Hydrogen Bond Distances (Å) for Nucleic Acid Base Pairs Shown in Figure 2[a]

complex[b]	H-bond[c]	exp.[d]	QM[e]	MM (CHARMM)[f]	QM/MM[OPT-BI, 6−311+G(d,p)][g]	QM/MM[OPT-BI, 6−31G(d)][h]	QM/MM[CHARMM, 6−311+G(d,p)][i]
AU*	N1···H−N3	2.82	2.88 (−0.06)	2.87 (−0.05)	3.04 (−0.22)	3.05 (−0.23)	2.84 (−0.3)
AU*	N6···O4	2.95	2.94(0.01)	2.89 (0.06)	3.01 (−0.06)	3.00 (−0.06)	2.79 (0.16)
A*U	N1···H−N3	2.82	2.88 (−0.06)	2.87 (−0.05)	2.98 (−0.16)	2.99 (−0.17)	2.84 (−0.02)
A*U	N6···O4	2.95	2.94(0.01)	2.89 (0.06)	3.08 (−0.13)	3.10 (−0.15)	2.84 (0.11)
CG*	N4-H···O6	2.91	2.80 (0.11)	2.83 (0.08)	3.04 (−0.13)	3.07 (−0.16)	2.82 (0.09)
CG*	N3···H−N1	2.95	2.96 (−0.01)	2.92 (0.03)	3.10(−0.15)	3.09 (−0.14)	2.85(0.10)
CG*	O2···H−N2	2.86	2.91 (−0.06)	2.84 (0.02)	3.03 (−0.17)	3.03 (−0.17)	2.75 (0.11)
C*G	N4-H···O6	2.91	2.80(0.11)	2.83 (0.08)	2.98 (−0.07)	3.02 (−0.11)	2.74 (0.17)
C*G	N3···H−N1	2.95	2.96 (−0.01)	2.92 (0.03)	3.10 (−0.15)	3.07 (−0.13)	2.89 (0.06)
C*G	O2···H−N2	2.86	2.91 (−0.06)	2.84 (0.02)	3.07(−0.12)	3.03 (−0.08)	2.85(0.10)
AT*	N1···H−N3	−	2.89	2.87 (0.01)	2.98 (−0.10)	3.04 (−0.15)	2.82 (0.07)
AT*	N6-H···O4	−	2.94	2.90 (0.04)	3.09 (−0.15)	3.08 (−0.14)	2.77 (0.17)
A*T	N1···H−N3	−	2.89	2.87 (0.01)	2.99 (−0.10)	3.13 (−0.24)	2.84 (0.04)
A*T	N6-H···O4	−	2.94	2.90 (0.04)	3.09 (−0.14)	3.02 (−0.08)	2.85 (0.09)
STDEV - EXP[j]			0.06	0.05	0.05	0.09	0.14
STDEV -QM[k]				0.02	0.03	0.07	0.06
MSE -EXP[l]			0.00	−0.03	0.14	0.14	−0.06
MSE- QM[m]				−0.03	0.12	0.16	−0.09
MUE -EXP[n]			0.05	0.05	0.14	0.14	0.12
MUE - QM[o]				0.03	0.12	0.16	0.09

[a] Numbers in parentheses are the deviation from experimental hydrogen bond distances or in case of the AT base pair the fully QM calculation {Δ[QM − (QM/MM)]}. The hydrogen bond distances are calculated between hydrogen bond acceptors and donors (i.e., heavy atoms). [b] See Figure 2 for names of complexes.* denotes the base treated with QM. [c] See Figure 2 for labeling. [d] Taken from ref.[61] [e] Fully QM. The B3LYP/6−311+G(d,p) level of theory was used. [f] Fully MM. [g] Optimized base-independent parameters (Table 4)and the. [h] Optimized base-independent parameters (Table 4). [i] Standard CHARMM27 vdW parameters (Table 4. [j] Standard deviation of the error from the experimental value. [k] Standard deviation of the error from the full QM calculated value. [l] Mean signed error from the experimental value, see notes in Table 1. [m] Mean unsigned error from the full QM calculated value, see notes in Table 1. [n] Mean unsigned error from the experimental value, see notes in Table 1. [o] Mean unsigned error from the full QM calculated value, see notes in Table 1.

Table 8. Interaction Energies (kcal/mol) Calculated for Nucleic Acid Base Pairs Shown in Figure 2[a]

complex[b]	QM[c]	MM(CHARMM)[d]	QM/MM [OPT-BI, 6−311+G(d,p)][e]	QM/MM [OPT-BI, 6−311G(d)][f]	QM/MM [CHARMM, 6−311+G(d,p)][g]
A*U	−12.1	−13.2 (1.1)	−13.8 (1.7)	−12.5 (0.4)	−16.9 (4.8)
AU*	−12.1	−13.2 (1.1)	− 13.8 (1.7)	−12.5 (0.4)	−17.0 (4.9)
A*T	−11.8	−12.8 (1.0)	−13.3 (1.5)	−11.0 (−0.8)	−14.2 (2.4)
AT*	−11.8	−12.8 (1.0)	−16.2 (4.4)	−14.1 (2.3)	−18.1 (6.3)
C*G	−23.7	−26.4 (2.5)	−24.7 (0.8)	−22.5 (−1.4)	−29.06 (5.2)
CG*	−23.7	−26.4 (2.5)	−25.7 (1.8)	−24.6 (0.7)	−30.2 (6.3)
STDEV [h]		0.8	1.2	1.3	1.4
MSE [i]		−1.5	−2.0	−0.3	−5.0
MUE [j]		1.5	2.0	1	5.0

[a] Numbers in parentheses are the deviation from the fully QM calculation {Δ[QM − (QM/MM)]}. [b] See Figure 2 for names of complexes. * denotes the base which is treated with QM. [c] Full QM, the B3LYP/6−311+G(d,p) level of theory was used. [d] Full MM, standard CHARMM27 vdW parameters were used. [e] Optimized base-independent parameters (Table 4)and the B3LYP/6−311+G(d,p) level of theory were used. [f] Optimized base-independent parameters (Table 4 and the B3LYP/6−311G(d) level of theory were used. [g] Standard CHARMM27 vdW parameters (Table 4)and the B3LYP/6−311+G(d,p) level of theory were used. [h] Standard deviation of the error from the full QM calculated value. [i] Mean signed error from the full QM calculated value, see Table 1. [j] Mean unsigned error from the full QM calculated value, see Table 1.

energies, providing an illustration of their known good quality for (MM) calculations on nucleic acid complexes.

Suitability of the Optimized Parameters for Modeling Chemical Reactions. When chemical reactions are studied by QM/MM methods, the atoms directly involved in the reaction must be treated quantum mechanically, while the surrounding environment can be treated with MM. vdW parameters optimized, for example, for the initial (reactant) state may not be suitable for the product or transition state. The fact that the same Lennard−Jones parameters are typically used for the QM atoms throughout a chemical reaction is a limitation of most current QM/MM methods. Therefore, we tested the suitability of our optimized vdW parameters for investigating changes associated with chemical reaction, using a test set consisting of protonated or

deprotonated nucleic acid bases (Figure 3). The most physiologically important protonation and deprotonation sites were selected, i.e., protonation of adenosine at N2, cytosine at N3, and guanosine at N7 and deprotonation of uracil at N1 and guanosine at N1 positions. Interaction and deprotonation energies were calculated at the [B3LYP/6−311+G(d,p)] QM/MM level using standard CHARMM27,[23,24,32,33] base-dependent (Table 3), and base-independent (Table 4) parameters. The interaction and deprotonation energies obtained were compared to those from fully QM calculations [B3LYP/6−311+G(d,p); Table 9]. Standard vdW parameters were not available for some of the atoms in the complexes shown in Figure 3 for the interaction and deprotonation energy calculations. This is because these calculations involved modified nucleic acid base chemical structures (e.g., proto-

Table 9. Interaction Energies (kcal/mol) Calculated at the B3LYP/6-311+G(d,p) QM and B3LYP/6-311+G(d,p)-QM/MM Levels Using Either CHARMM27, Base-Independent Parameters (Table 4), or Base-Dependent Parameters (Table 3) for Neutral, Protonated and Deprotonated Nucleic Acid Bases (Figure 3)[a]

complex[b]	QM	QM/MM(CHARMM)[c]	QM/MM (OPT-BI)[d]	QM/MM (OPT-BD)[e]
A−wat	−8.9	−13.3 (4.4)	−10.2 (1.3)	−10.3 (1.4)
AN1$^+$−wat	−16.1	−18.4 (2.3)	−17.0 (−1.4)	−18.4 (2.3)
U−wat	−7.5	−10.4 (2.9)	−9.3 (1.8)	−9.3 (1.8)
UN3$^-$−wat	−16.0	−19.7 (3.7)	−19.3 (3.3)	−19.7 (3.7)
G−wat1	−10.8	−10.7 (−0.1)	−12.8 (2.0)	−11.3 (0.5)
GN1$^-$−wat	−15.9	−20.0 (−4.1)	−19.4 (3.5)	−19.7 (3.8)
G−wat2	−5.4	−9.0 (−3.6)	−6.6 (0.6)	−6.7 (1.3)
GN7$^+$−wat	−16.3	−19.4 (3.1)	−16.4 (0.1)	−16.0 (−0.3)
C−wat1	−10.5	−13.8 (3.3)	−12.9 (2.4)	−12.9 (2.4)
CN3$^+$−wat1	−16.5	−19.9 (3.4)	−18.5 (2.0)	−19.0 (2.5)
C−wat2	−7.0	−9.1 (2.1)	−8.7 (1.7)	−7.4 (0.4)
CN3$^+$−wat2	−14.6	−18.5 (3.9)	−15.6 (1.0)	−16.7 (2.1)
STDEV [f]		2.9	1.4	1.3
MSE [g]		−1.8	−1.5	−1.8
MUE [h]		3.1	1.8	1.9

[a] Numbers in parentheses are the deviation from the fully QM calculation {Δ[QM − (QM/MM)]}. [b] See Figure 3 for names of complexes. [c] Standard CHARMM27 vdW parameters were used. [d] Optimized base-independent parameters (Table 4) were used. [e] Optimized base-dependent parameters (Table 3)were used. [f] Standard deviation of the error from the full QM calculated value. [g] Mean signed error from the full QM calculated value, see notes in Table 1. [h] Mean unsigned error from the full QM calculated value, see notes in Table 1.

Table 10. Deprotonation Energies [(energy of the deprotonated complex) − (energy of the protonated complex) in kcal/mol] Calculated at the B3LYP/6-311+G(d,p) Level for the Reactions Shown in Figure 3[a]

reaction[b]	QM	QM/MM (CHARMM27)[c]	QM/MM (OPT-BI)[d]	QM/MM (OPT-BD)[e]
AN1$^+$ → A	−242.6	−240.6(−2.0)	−242.3 (−0.2)	−243.7 (1.1)
1: CN3$^+$ → C	−245.3	−246.0(0.2)	−244.9 (−0.4)	−245.4(0.1)
2: CN3$^+$ → C	−246.9	−248.8 (1.9)	−246.2 (−0.7)	−246.3 (0.4)
U → UN3$^-$	−346.8	−346.0 (−0.8)	−345.2 (−1.6)	−344.8 (−1.9)
G → GN1$^-$	−341.9	−337.3 (−4.6)	−340.1 (−1.8)	−338.3 (−3.6)
GN7$^+$ → G	−251.3	−252.3 (1.0)	−249.8(−1.5)	−249.2 (−2.1)
STDEV [f]		2.3	0.7	1.8
MSE [g]		0.7	1.0	1.0
MUE[h]		1.9	1.0	1.5

[a] Numbers in parentheses are the deviation from the fully QM calculation {Δ[QM − (QM/MM)]}. [b] See Figure 3 for names of reactions. [c] Standard CHARMM27 vdW parameters were used. [d] Optimized base-independent parameters (Table 4)were used. [e] Optimized base-dependent parameters (Table 3) were used. [f] Standard deviation of the error from the full QM calculated value. [g] Mean signed error from the full QM calculated value, see notes in Table 1. [h] Mean unsigned error from the full QM calculated value, see notes in Table 1.

nated/deprotonated), not included in the CHARMM27 parameter set (nor in our optimized set). Base-independent vdW parameters were available for all atom types except NN2U (in the UN3$^-$_wat complex, Figure 3). For the UN3$^-$_wat complex, standard CHARMM27 vdW parameters were used in all calculations. In addition, for the following atom types, the base-dependent vdW parameters (Table 3) were not available (see Figure 3 for details of the complexes): HN2 (AN1$^+$_wat, CN3$^+$_wat1 and CN3$^+$_wat2), NN3 (GN1$^-$_wat), and NN4 (G_wat2). For these atom types in the complexes mentioned, standard CHARMM27 vdW parameters were used in the base-dependent calculations.

The QM/MM interaction energies for the charged complexes are overestimated, compared to the QM interaction energies, in almost all cases (Table 9). The optimized vdW parameters [both base-dependent (Table 3) and base-independent (Table 4)] in the QM/MM calculations give generally better agreement with the QM interaction energies than the standard CHARMM27 parameters[23,24,32,33] (Table 9). Only in one case is a larger error observed with our optimized parameters than with standard CHARMM27 parameters (G_wat_1, Table 9). With the optimized parameters, the interaction energies differ from the QM results by 0.1−3.8 kcal/mol, while with the standard CHARMM27

parameters[23,24,32,33] all except G_wat_1 vary by 2.1−4.4. kcal/mol. G_wat_1 is the only complex where the standard CHARMM27 parameters[23,24,32,33] give very good agreement with the QM interaction energies (an error of only 0.1 kcal/mol, Table 9). With our optimized parameters, the largest deviations between the QM and QM/MM results are seen with deprotonated complexes, i.e., UN3$^-$_wat and GN1$^-$_wat_1 complexes, regardless of which parameter set is used (Figure 3 and Table 9).

Comparison of deprotonation energies for nucleic acid bases in complex with one water molecule (see Figure 3 for details of reactions) shows that deprotonation energies calculated at the full QM [B3LYP/6−311+G(d,p)] level and at the QM/MM [B3LYP/6−311+G(d,p)] level are generally surprisingly similar, within 0.1−4.6 kcal/mol (Table 10). In 8 reactions out of 18, the QM/MM deprotonation energies are within 1.0 kcal/mol of the QM results, and in 7 reactions out of 18, they are within 1−2 kcal/mol. Only in three cases do the QM/MM deprotonation energies deviate by more than 2.0 kcal/mol from the QM deprotonation energies (Table 10). Comparison of the deprotonation energies obtained with different vdW parameters (i.e., the standard CHARMM27,[23,24,32,33] base-dependent and independent) shows that the optimized base-independent parameters seem to be little better than the

Table 11. Electrostatic Contribution to the Interaction Energy (kcal/mol) Calculated at the B3LYP/6−31G(d) QM/MM and CHARMM27 MM Levels for Some Sample Base−Water Complexes

complex[a]	structure	hydrogen bonds[b]		QM/MM energy[c]	MM energy[d]
G_wat_1	MM[e]	MM	QM	−16.23	−13.50
	QM [f]	MM	QM	−15.41	−12.55
	QM/MM (CHARMM27[g])	MM	QM	−16.63	−14.02
	QM/MM (OPT_BI) [h]	MM	−	−14.06	−12.14
	QM/MM (OPT-BD)[i]	MM	−	−14.32	−12.33
G_wat_2	MM [e]	−	QM	−13.06	−12.38
	QM[f]	−	−	−10.75	−10.44
	QM/MM (CHARMM27)[g]	QM	-	−12.87	−12.88
	QM/MM (OPT_BI)[h]	−	−	−11.08	−12.14
	QM/MM (OPT-BD)[i]	−	−	−10.22	−12.33
A_wat_1	MM[e]	MM	QM	−15.96	−13.15
	QM [f]	MM	QM	−15.02	−12.29
	QM/MM (CHARMM27)[g]	MM	QM	−17.90	−14.58
	QM/MM (OPT_BI)[h]	MM	QM	−15.00	−12.37
	QM/MM (OPT-BD)[i]	MM	QM	−15.00	−12.37
C_wat_2	MM[e]	MM	−	−15.39	−13.23
	QM[f]	−	MM	−10.79	−10.03
	QM/MM (CHARMM27)[g]	−	MM	−10.36	−10.04
	QM/MM (OPT_BI)[h]	−	MM	−11.35	−10.59
	QM/MM (OPT-BD)[i]	−	−	−7.83	−7.96

[a] See Figures 1B−E for names of complexes. [b] Characterization of hydrogen bonds in the system. QM denotes a hydrogen bond with a QM hydrogen, MM a hydrogen bond with a MM hydrogen, and − no hydrogen bond present (a hydrogen bond was deemed to be present at acceptor to H distances ≤2 Å). The distances are shown in Table 1. [c] Electrostatic contribution to the interaction energy, calculated using QM/MM. [d] Electrostatic contribution to the interaction energy, calculated using pure MM. [e] Pure MM calculation using CHARMM27. [f] Pure QM calculation, using Gaussian. [g] Standard CHARMM27 parameters were used. [h] Optimized base-independent parameters (Table 4) were used. [i] Optimized base-dependent parameters (Table 3) were used.

optimized base-dependent or the standard CHARMM27 parameters.[23,24,32,33] This may be due, at least partly, to the fact that in the base-independent calculations, optimized vdW parameters were available for all atoms (except NN2U in the UN3⁻_wat complex (as mentioned previously standard CHARMM27 parameters were used)). However, in the base-dependent calculations more optimized parameters were unavailable, and hence standard CHARMM27 vdW parameters[23,24,32,33] had to be used (see above). Therefore, the base-independent parameters developed here (Table 4) would appear to be a better choice than standard CHARMM27 vdW parameters for QM/MM modeling of such reactions in nucleic acid bases.

Electrostatic Contribution to the Calculated QM/MM Interaction Energies. While the parametrization here has been generally successful, the improvements have not been uniform. When the original parameters were used, we found that typically QM/MM hydrogen bond interaction energies were too strong and hydrogen bond lengths were too short (Tables 1 and 2).[29,30,34] We attempted to address this problem by optimizing the vdW parameters, however, as can be seen from Tables 3 and 4, only the parameters of the hydrogens were altered significantly. This has been seen previously.[27,31] To further investigate why hydrogen bond strengths are overestimated when the original parameters are used, we have calculated the QM/MM and MM electrostatic contributions to the interaction energy for the QM optimized, MM optimized, and QM/MM optimized geometries. The results are shown in Table 11. The QM/MM electrostatic interactions in the QM optimized and MM optimized structures are consistently stronger than the corresponding MM electrostatic interactions. This indicates that the electrostatic interactions are the source of the error. Given that the electrostatic interactions are overestimated, it is not

surprising that the QM/MM optimized geometries have shorter hydrogen bond lengths. Our optimizations increased the vdW radii of QM (hydrogen) atoms, and, as expected, this lengthens the hydrogen bonds, resulting in a reduction in the QM/MM electrostatic interaction energy for these complexes. The results vary for the different complexes. In three cases hydrogen bonds are broken, resulting in either significant improvements (G_wat_2) or in overcompensation (G_wat_1 and C_wat_2). For A_wat_1, there are significant reductions in the electrostatic contribution that cannot be readily explained. For the base-independent results for C_wat_2, the optimizations result in the electrostatics becoming even more favorable. It is apparent that our optimization of the vdW parameters does not consistently improve the results, suggesting that more refined QM/MM models (e.g., with different treatment of QM/MM electrostatic interactions[62]) may be required for some applications.

The optimization of the vdW parameters compensates for the overestimated QM/MM electrostatic interaction by increasing the vdW radii of the QM hydrogen atoms involved in hydrogen bonds (large increases of the hydrogen $R_{min}/2$ values are seen). While this has been reasonably successful, it has not addressed the root cause of the problem. Optimizing the vdW parameters alone provides an improved model, but ideally treatment of QM/MM electrostatic interactions should also be considered and improved. Further study of the balance of electrostatic interactions in QM/MM systems is required.

Conclusions

QM/MM interaction energies and geometries for the hydrogen bonded systems are fairly sensitive to the values of the van der Waals parameters used to describe the nonbonded

dispersion interactions between the QM and MM atoms. In general, the use of the standard CHARMM27 parameters for the QM atoms in QM/MM calculations on nucleic acid base−water complexes gives shorter hydrogen bond lengths than those obtained from full QM calculations. QM/MM calculations with the standard CHARMM27 parameters overestimate the interaction energies by 1.1−3.4 kcal/mol, compared to the equivalent QM results (using B3LYP hybrid density functional theory). Adjusting the vdW parameters on the QM atoms gives improved results for interaction energies and in most cases also for hydrogen bond lengths. As expected, better agreement between the QM/MM and QM geometries and interaction energies is obtained with base-dependent parameters rather than with base-independent parameters. Using vdW parameters optimized at the B3LYP/6−311+G(d,p) level in QM/MM calculations at the B3LYP/6−31G(d) level reproduces the full QM [B3LYP/6−31G(d)] interaction energies (very well) and also geometries (reasonably well). Thus, vdW parameters can be transferred from the higher level to the lower level, when the same DFT model is used, and so should be useful in QM/MM calculations applying the popular B3LYP/6−31G(d) level of QM theory. Transferability to other levels of QM/MM treatment is not guaranteed and should be tested, particularly when different types of QM/MM approach (e.g., applying ab initio molecular orbital methods) are used. For example, Hartree−Fock level ab initio QM/MM methods tend to show more overestimation of intermolecular interaction energies.[63] Luque et al. have reported that vdW parameters are sensitive to the QM/MM formalism and cannot be transferred between different QM/MM methods.[30] Further work is required to analyze QM/MM interactions. In other work, we have examined QM/MM modeling of biomolecular hydrogen bonds. Examination of changes in total electron density and natural bond orbital atomic charges, due to hydrogen bond formation, in selected complexes showed that charge leakage from the QM atoms to MM atomic point charges close to the QM/MM boundary is not a serious problem, at least with limited basis sets.[64] Clearly, current QM/MM methods can provide good descriptions of many biomolecular systems.[65] Further analysis and development of more sophisticated treatment of QM/MM interactions (in particular of electrostatic interactions) should remain a central goal of research in this area, however. This is particularly true now that it is possible to model biomolecules (e.g., enzyme reactions) with QM/MM techniques employing correlated ab initio methods that are capable of high accuracy[66,67] and to carry out extensive high-level QM/MM free energy calculations.[68] QM/MM free energy calculations may provide a more general approach to optimizing parameters for QM/MM interactions.[69,70]

In summary, the results here show that, when appropriate parameters are used, hydrogen bonding interactions of nucleic acids can be modeled well with QM/MM methods. Obtaining parameters which reliably reproduce both interaction energies and geometries (hydrogen bonds and angles) can be difficult, especially if there are multiple hydrogen bonds in a system, as is the case for the interactions of nucleic acid bases with water. The parameters presented here give better agreement with full QM calculations, especially for interaction energies. They have also successfully been used for complexes not used in the parameter optimization. Therefore, these parameters should be useful for QM/MM investigations of nucleic acid structure employing the B3LYP hybrid density functional QM method. It should, however, be noted that these parameters have been developed to represent hydrogen bonding interactions of nucleic acid bases, not base-stacking interactions. The modifications made here seem unlikely to affect the treatment of stacking significantly, but this should be checked in applications where base-stacking is important. The parameters developed here (especially the base-independent parameters) also appear to perform reasonably well for modeling chemical changes, based on results for deprotonation reactions. However, parameters specifically optimized for a given reaction may be preferable, for enhanced accuracy.[71] For the chemical changes modeled here, the optimized parameters performed well, but it should be remembered that in current QM/MM implementations a compromise must usually be made between representing QM/MM interactions of the substrate and product (and for example the transition state), as the same van der Waals parameters are typically used for all stages of a reaction. It may be that a single set of (MM) parameters can represent the interactions of all species in the reaction satisfactorily, but this cannot be guaranteed in advance. Accordingly, we believe that the best approach for QM/MM reaction modeling, in general, is to test the results for a given reaction and to optimize the parameters for that specific application as required. The parameters developed here should be a good basis for the optimization of reaction-specific Lennard−Jones parameters for QM atoms in QM/MM (CHARMM27) studies of reactions involving nucleic acids.

Supporting Information Available: Hydrogen bond angles for nucleic acid base and water complexes, shown in Figures 1B−E, calculated with pure QM, pure MM and combined QM/MM methods, using CHARMM27 and optimized parameters. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Gogonea, V.; Suarez, D.; van der Vaart, A.; Merz, K. M. *Curr. Opin. Struct. Biol.* **2001**, *11*, 217.

(2) Warshel, A. In *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley & Sons: New York, 1991.

(3) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.

(4) Warshel, A.; Karplus, M. *J. Am. Chem. Soc.* **1972**, *94*, 5612.

B3LYP/CHARMM27 QM/MM Modeling of Nucleic Acid Bases

J. Chem. Theory Comput., Vol. 5, No. 2, 2009 **409**

(5) Gao, J.; Thompson, M. *Combined quantum mechanical and molecular mechanical methods.* ACS Symposium Series 712; American Chemical Society: Washington, D.C., 1998; p 712.

(6) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **2000**, *21*, 1442.

(7) Mulholland, A. J. *Drug Discovery Today* **2005**, *10*, 1393.

(8) Monard, G.; Merz, K. M. *Acc. Chem. Res.* **1999**, *32*, 904.

(9) Gao, J. Methods and Applications of Combined Quantum Mechanical and Molecular Mechanical Potentials. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Donald, B. B., Eds.; VCH Publishers Inc.: New York, 1996; Vol. 7, pp 119−185.

(10) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720.

(11) Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 1959.

(12) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.

(13) Singh, U.; Kollman, P. *J. Comput. Chem.* **1986**, *7*, 718.

(14) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580.

(15) Thery, V.; Rinaldi, D.; Rivail, J.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269.

(16) Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290.

(17) Pu, J.; Gao, J.; Truhlar, D. *J. Phys. Chem. A* **2004**, *108*, 5454.

(18) Pu, J.; Gao, J. L.; Truhlar, D. *ChemPhysChem* **2005**, *6*, 1853.

(19) Garcia-Viloca, M.; Gao, J. *Theor. Chem. Acc.* **2004**, *111*, 280.

(20) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714.

(21) Pu, J. Z.; Gao, J. L.; Truhlar, D. *J. Phys. Chem. A* **2004**, *108*, 632.

(22) Mulholland, A. J. In *Chemical Modelling: Applications And Theory; RSC Specialist Periodical Reports*; Hinchliffe, A., Ed.; The Royal Society of Chemistry: London, 2006; Chapter 2, pp 23−68.

(23) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

(24) MacKerell, A. D. *Annu. Rep. Comput. Chem.* **2005**, *1*, 91.

(25) Gao, J.; Xia, X. *Science* **1992**, *258*, 631.

(26) Gao, J. Computation of Intermolecular Interactions with the Combined Quantum Mechanical and Classical Approach In *Modeling the Hydrogen Bond*; Smith, D. A., Eds.; ACS Symposium Series 569, 1994, Chapter 2, pp 8−21.

(27) Freindorf, M.; Gao, J. *J. Comput. Chem.* **1996**, *17*, 386.

(28) Bash, P. A.; Ho, L. L.; MacKerell, A. D., Jr.; Levine, D., Jr.; Hallsrom, P. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 3698.

(29) Riccardi, D.; Li, G.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 6467.

(30) Luque, F. J.; Reuter, N.; Cartier, A.; Ruiz-Lopez, M. F. *J. Phys. Chem. A* **2000**, *104*, 10923.

(31) Freindorf, M.; Shao, Y.; Furlani, T. R.; Kong, J. *J. Comput. Chem.* **2005**, *26*, 1270.

(32) Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86.

(33) MacKerell, A. D., Jr.; Banvali, N. K. *J. Comput. Chem.* **2000**, *21*, 105.

(34) Gao, J. *Biophys. Chem.* **1994**, *51*, 253.

(35) Zhou, D. M.; Taira, K. *Chem. Rev.* **1998**, *98*, 991.

(36) Takagi, Y.; Warashina, M.; Stec, W. J.; Yoshinari, K.; Taira, K. *Nucleic Acids Res.* **2001**, *29*, 1815.

(37) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. *Cell* **1983**, *35*, 849.

(38) Cech, T. R.; Zaug, A. J.; Grabowski, P. J. *Cell* **1981**, *29*, 487.

(39) Noll, D.; Mason, T. M.; Miller, P. *Chem. Rev.* **2006**, *106*, 277.

(40) Elcock, A. H.; Lyne, P. D.; Mulholland, A. J.; Nandra, A.; Richards, W. G. *J. Am. Chem. Soc.* **1995**, *117*, 4706.

(41) Hopkins, P. B.; Millard, J. T.; Woo, J.; Weidner, M. F.; Kirchner, J. J.; Sigurdsson, S. T.; Raucher, S. *Tetrahedron* **1991**, *47*, 2475.

(42) Nordberg, J. L.; Nilsson, L. *Biopolymers* **1996**, *39*, 765.

(43) Hobza, P.; Sponer, J. *Chem. Rev.* **1999**, *99*, 3247.

(44) Li, Y.; Breaker, R. R. *J. Am. Chem. Soc.* **1999**, *121*, 5364.

(45) Kaukinen, U.; Venäläinen, T.; Lönnberg, H.; Peräkylä, M. *Org. Biomol. Chem.* **2003**, *1*, 2439.

(46) Kaukinen, U.; Lönnberg, H.; Peräkylä, M. *Org. Biomol. Chem.* **2004**, *2*, 66.

(47) Bibillo, A. M.; Figlerowicz, M.; Kierzik, R. *Nucleic Acids Res.* **1999**, *27*, 3931.

(48) Kierzek, R. *Nucleic Acids Res.* **1992**, *20*, 5079.

(49) Claeyssens, F.; Ranaghan, K. E.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J. *Chem. Commun.* **2005**, 5086.

(50) Jaguar 6.0, Schrödinger Inc., Portland, OR, 2002.

(51) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant,J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.;Kitao,O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev,O.; Austin,A. J.; Cammi,R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth,G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.;Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.;Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.;Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez,C.; Pople, J. A. Gaussian 03, Revision C.02, Gaussian, Inc.,Wallingford, CT, 2004.

(52) Rablen, P. R.; Lockman, J. W.; Jorgensen, W. L. *J. Phys. Chem. A* **1998**, *102*, 3782.

(53) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(54) Ponder, J. W. Tinker - Software Tools for Molecular Design. St. Louis, MO,2003 (http://dasher.wustl.edu/tinker/); accessed 10/17/08.

(55) Harvey, J. N. *J. Faraday Discuss.* **2004**, *127*, 165.

(56) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(57) MacKerell, A. D., Jr.; Brooks, B.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus M. In *The Encyclopedia of Computational Chemistry*; John Wiley & Sons: Chichester, 1998; pp 271−277.

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

(59) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902.

(60) Mulholland, A. J.; Lyne, P. D.; Karplus, M. *J. Am. Chem. Soc.* **2000**, *122*, 534.

(61) Saenger, W. In *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

(62) Biswas, P. K.; Gogone, V. *J. Chem. Phys.* **2005**, *123*, 1.

(63) Ridder, L.; Harvey, J. N.; Rietjens, I. M. C. M.; Mulholland, A. J. *J. Phys. Chem. B.,* **2003**, *107*, 2118.

(64) Senthilkumar, K.; Mujika, J. I.; Ranaghan, K. E.; Manby, F. R.; Mulholland, A. J.; Harvey, J. N. *J. R. Soc. Interface* **2008**, *5*, S207.

(65) van der Kamp, M. W.; Shaw, K. E.; Woods, C. J.; Mulholland, A. J. *J. R. Soc. Interface* **2008**, *5*, S173.

(66) Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schutz, M.; Thiel, S.; Thiel, W.; Werner, H. J. *Angew. Chem.* **2006**, *45*, 6856.

(67) Mulholland, A. J. *Chem. Cent. J.* **2007**, *1*, 19.

(68) Woods, C. J.; Manby, F. R.; Mulholland, A. J. *J. Chem. Phys.* **2008**, *128*, 014109.

(69) Cummins, P. L.; Gready, J. E. *J. Comput. Chem.* **1997**, *18*, 1496.

(70) Ridder, L.; Rietjens, I. M. C. M.; Vervoort, J.; Mulholland, A. J. *J. Am. Chem. Soc.* **2002**, *124*, 9926.