JCTC Journal of Chemical Theory and Computation

Improved Hydrogen Bonding at the NDDO-Type Semiempirical Quantum Mechanical/Molecular Mechanical Interface

Qiantao Wang and Richard A. Bryce*

School of Pharmacy and Pharmaceutical Sciences, University of Manchester, Oxford Road, Manchester M13 9PT, U.K.

Received May 25, 2009

Abstract: A semiempirical quantum mechanical (QM)/molecular mechanical (MM) potential with reformulated QM core-MM charge interactions is introduced, specifically to more accurately model hydrogen bonding at the QM/MM interface. Application of this potential using the PM3 Hamiltonian shows improved prediction of geometry and interaction energy for hydrogen bonded small molecule complexes typical of biomolecular interactions, without significantly impacting the modeling of other interaction types. Using this potential, more quantitative prediction of interaction energies is also found at a protein−ligand interface.

1. Introduction

Biological structure and function are dictated in major part by the influence of electrostatic interactions, in particular hydrogen bonds. These interactions typically dominate in substrate recognition, solvent rearrangements on binding and in enzyme reaction mechanisms. An ability to faithfully describe hydrogen bonding is fundamental to accurate biomolecular modeling methods. Ideally, it is desirable to model such complex biological systems at a quantum chemical level. A computationally efficient way to achieve this is *via* quantum mechanical/molecular mechanical (QM/MM) methods, where a quantum region is coupled to a solvent and biomolecular environment.[1] Since their first introduction,[2,3] QM/MM approaches have evolved in their power and popularity alongside improvements in the underlying QM and MM approaches.

QM/MM methods have also been employed to calculate thermodynamic properties, such as free energy reaction profiles, for comparison with experiment.[1] Here, however, in order to account for the large number of conformational states required in statistical mechanical calculations, the use of *ab initio* and density functional methods remains highly expensive. Correspondingly, there has been considerable interest in the development of fast and accurate semiempirical QM methods. Recent approaches include OMx,[4,5] PM3-D,[6] and PDDG/MNDO and PDDG/PM3[7] methods. Some approaches[8−11] seek to address the functional form of the NDDO[12] core−core interaction, contributing to the substantial improvements observed in the chemical accuracy of these methods.

In parallel with developments in QM and MM methodology, considerable effort has been invested into refining the description of the QM/MM interface, which can be modeled via mechanical or electronic embedding.[13] For the latter, within an *ab initio* QM/MM framework, the MM partial point charges enter the Hamiltonian of the QM region, forming electronic and nuclear interactions, while nonelectrostatic QM-MM interactions are modeled using a van der Waals potential. However, the atoms of a NDDO-based semiempirical QM region comprise valence electrons and an atomic core; this QM core subsumes the core electrons and nucleus of the atom into an *s* orbital term scaled by an effective nuclear charge, Z_a. Following the early work of Field et al.,[3] most current semiempirical QM/MM methods analogously treat MM point charges as scaled *s* orbital cores. The resulting QM/MM core−core interaction is of the same form as the QM/QM core−core expression that required correction as described above.

In this work, we therefore revisit the model of the semiempirical QM/MM interface, specifically exploring the effect of reformulating core−core interactions to more accurately model hydrogen bonding interactions in a biological context. In order to evaluate the ability of the method to model biologically important noncovalent interactions, we use the benchmark S22 data set,[14] a representative set of 22 bimolecular complexes with interaction energies determined at the level of CCSD(T) in conjunction with extrapolation methods to estimate the complete basis set (CBS) limit. We also apply our modified NDDO-based semiempirical QM/MM method to modeling noncovalent interactions at a protein−ligand interface.

2. Methods

Here we use the PM3[15] Hamiltonian in conjunction with the AMBER force field.[16] In the commonly adopted approach of Field et al.,[3] the energetic contribution arising from interaction of the core of a given QM atom *a* with MM atom *m* is given by

* Corresponding author phone: (0)161-275-8345; fax: (0)161-275-2481; e-mail: R.A.Bryce@manchester.ac.uk.

10.1021/ct9002674 CCC: $40.75
Published on Web 07/31/2009

$$E_{QM/MM}^{core} = Z_a q_m (s_a s_a, s_m s_m)(1 + e^{-\alpha_a R_{am}} + e^{-\alpha_m R_{am}}) \quad (1)$$

where Z_a is the effective charge of QM core a, q_m is the partial charge on MM atom m, s_a is an s orbital on the QM atom, s_m is a notional s orbital on the MM atom, and R_{am} is the QM-MM interatomic separation. α_m and α_a are parameters, as are $\rho_{a,0}$ and $\rho_{m,0}$ upon which the two-center two-electron integrals in eq 1 rely. After Field et al., α_m and $\rho_{m,0}$ are typically taken as 5.0 $Å^{-1}$ and 0.0 au, respectively.[3] The scaling of the integrals used in modeling core−core repulsions was originally introduced in the context of a purely semiempirical NDDO model and was intended to reflect increased screening by core electrons as two nuclei approach each other. Clearly, the physical premise alters in a QM/MM context, as now the expression contributes to the interaction of a MM partial point charge with a QM core, alongside a Lennard-Jones potential between QM and MM atoms. Our approach is to introduce a straightforward, general modification to this interaction term. Alteration of the final right-hand term of eq 1 leads to the following QM/MM core−core energy contribution by a given QM/MM atom pair

$$E_{QM/MM}^{core} = Z_a q_m (s_a s_a, s_m s_m)\left[1 + \frac{|q_m|}{q_m} \cdot (-e^{-f_1^a \cdot R_{am}} + e^{-f_2^a \cdot R_{am}})\right] \quad (2)$$

where $\rho_{m,0}$ is taken as 0.0 au as before, and f_1^a and f_2^a are exponential scale factors which depend on QM atom a. Optimal values were heuristically determined for f_1^a and f_2^a (Table 1) based on structures and energetics of complexes **1**−**3** of the S22 data set (Table 2).

A consequence of this new expression eq 2, relative to eq 1, is that the interaction of a QM core with a negatively charged (typically heavy) MM atom is enhanced; conversely, this QM/MM interaction is reduced for interaction of a QM core with a positively charged (typically light) MM atom. Use of eq 2 should therefore increase the strength of hydrogen bonding predicted by the semiempirical QM/MM method. All calculations were performed using a modified version of the Amber 9 software package.[16] For the MM molecules of S22 complexes, nonelectrostatic parameters for QM and MM atoms were obtained from the general AMBER force field (GAFF) for organic molecules.[17] No significant benefit was obtained by variation of van der Waals parameters for QM atoms. Atom-centered partial point charges for these molecules were derived using the AM1-BCC method.[18] For HIV-1 protease, parameters were adopted from the Cornell et al. force field for proteins.[19]

3. Results and Discussion

The S22 set[14] contains a range of binary complexes representative of biomolecular interactions (Table 2). Complexes **1**−**7** are bound together principally by hydrogen bonding; complexes **8**−**15** model dispersion-dominant interactions; and **16**−**22** are complexes with both types of interaction present. Starting from the *ab initio* geometries from Hobza et al. (at the MP2/cc-pVTZ level or higher[14]), interaction energies and optimal structures of the complexes were calculated at the PM3/MM level of theory. This led to a mean unsigned error in geometry across the 22 molecules of 0.22 Å (Table 3). The associated mean unsigned error (MUE) in energy relative to CCSD(T) energies at the complete basis set limit was 2.1 kcal/mol (Table 2). What is striking is the poor agreement for the hydrogen bonding complexes **1**−**7**, with mean unsigned errors in geometry and

Table 1. f_1^a and f_2^a Parameters for QM Atom *a* ($Å^{-1}$)

	f_1^a	f_2^a
H	2.2	2.7
H*	3.4	3.6
C	3.4	3.9
N	2.9	3.4
O	3.6	3.6

* When both QM and MM atoms are H.

Table 2. Interaction Energies of S22 Complexes (kcal/mol)[a] Calculated via the PM3/MM* Model (Using Eq *2*), Compared with PM3/MM and Reference Values[14]

		QM region	ref 14	PM3/MM	PM3/MM*
	Hydrogen Bonded Complexes (7)				
1	ammonia dimer (C_{2h})	*ammonia*	−3.17	−2.17	−3.18
2	water dimer (C_s)	*donor*	−5.02	−3.51	−5.25
		acceptor	−5.02	−3.92	−5.22
3	formic acid dimer (C_{2h})	*formic acid*	−18.61	−10.14	−14.31
4	formamide dimer (C_{2h})	*formamide*	−15.96	−8.51	−10.86
5	uracil dimer (C_{2h})	*uracil*	−20.65	−13.56	−16.61
6	2-pyridoxine/ 2-aminopyridine (C_1)	*2-py*	−16.71	−9.61	−13.23
		2-am	−16.71	−9.82	−13.60
7	adenine/thymine WC (C_1)	*adenine*	−16.37	−10.11	−14.36
		thymine	−16.37	−9.18	−12.59
MUE				**5.41**	**2.63**
MSE				**5.41**	**2.54**
	Complexes with Predominant Dispersion Contribution (8)				
8	methane dimer (D_{3d})	*methane*	−0.53	−0.49	−0.51
9	ethene dimer (D_{2d})	*ethene*	−1.51	−0.79	−0.88
10	benzene/methane (C_3)	*benzene*	−1.50	−0.97	−0.99
		methane	−1.50	−1.06	−1.15
11	benzene dimer (C_{2h})	*benzene*	−2.73	−2.56	−2.67
12	pyrazine dimer (C_s)	*pyrazine*	−4.42	−5.33	−5.64
13	uracil dimer (C_2)	*uracil*	−10.12	−9.31	−10.21
14	indole/benzene (C_1)	*indole*	−5.22	−5.57	−6.01
		benzene	−5.22	−3.99	−5.10
15	adenine/thymine stack (C_1)	*adenine*	−12.23	−12.18	−13.46
		thymine	−12.23	−12.87	−14.36
MUE				**0.54**	**0.65**
MSE				**0.19**	**−0.34**
	Mixed Complexes (7)				
16	ethene/ethine (C_{2v})	*ethene*	−1.53	−0.57	−0.62
		ethane	−1.53	−0.67	−0.73
17	benzene/water (C_s)	*benzene*	−3.28	−2.18	−3.57
		water	−3.28	−3.45	−3.79
18	benzene/ammonia (C_s)	*benzene*	−2.35	−1.59	−2.59
		ammonia	−2.35	−2.87	−3.12
19	benzene/HCN (C_s)	*benzene*	−4.46	−2.51	−2.69
		HCN	−4.46	−3.17	−3.30
20	benzene dimer (C_{2v})	*vertical*	−2.74	−1.98	−2.10
		horizontal	−2.74	−1.78	−1.90
21	indole/benzene T-shape (C_1)	*indole*	−5.73	−5.55	−6.00
		benzene	−5.73	−4.10	−5.27
22	phenol dimer (C_1)	*donor*	−7.05	−5.15	−6.17
		acceptor	−7.05	−5.29	−6.74
MUE				**1.06**	**0.70**
MSE				**0.96**	**0.41**
	Overall				
MUE				**2.14**	**1.24**
MSE				**1.99**	**0.78**
r^2				**0.86**	**0.94**

[a] MUE (mean unsigned error); MSE (mean signed error); correlation with reference values, r^2.

Table 3. Interaction Distances of S22 Complexes (in Å)[a] Calculated via the PM3/MM* Model (Using Eq *2*), Compared with PM3/MM and Reference Values[14]

		QM region	ref 14	PM3/MM	PM3/MM*
	Hydrogen Bonded Complexes (7)				
1	ammonia dimer (C_{2h})	*ammonia*	2.504	2.221	2.112
			2.504	3.530	2.179
2	water dimer (C_s)	*donor*	1.952	2.008	1.791
		acceptor	1.952	1.973	1.770
3	formic acid dimer (C_{2h})	*formic acid*	1.670	1.851	1.637
			1.670	1.891	1.739
4	formamide dimer (C_{2h})	*formamide*	1.841	1.987	1.874
			1.841	2.105	1.886
5	uracil dimer (C_{2h})	*uracil*	1.775	1.882	1.792
			1.775	1.941	1.822
6	2-pyridoxine/ 2-aminopyridine (C_1)	*2-py*	1.859	2.055	1.954
			1.874	1.978	1.856
		2-am	1.859	2.031	1.882
			1.874	2.037	1.869
7	adenine/thymine WC (C_1)	*adenine*	1.819	1.978	1.833
			1.929	2.026	1.886
		thymine	1.819	2.014	1.891
			1.929	1.989	1.841
MUE				**0.201**	**0.092**
MSE				**0.169**	**−0.046**
	Complexes with Predominant Dispersion Contribution (8)				
8	methane dimer (D_{3d})	*methane*	3.718	3.639	3.621
9	ethene dimer (D_{2d})	*ethene*	3.718	3.899	3.842
10	benzene/methane (C_3)	*benzene*	3.716	3.907	3.895
		methane	3.716	3.874	3.838
11	benzene dimer (C_{2h})	*benzene*	3.765	3.823	3.806
12	pyrazine dimer (C_s)	*pyrazine*	3.479	3.444	3.410
13	uracil dimer (C_2)	*uracil*	3.166	3.440	3.356
14	indole/benzene (C_1)	*indole*	3.498	4.500	4.535
		benzene	3.498	4.117	4.577
15	adenine/thymine stack (C_1)	*adenine*	3.172	3.403	3.360
		thymine	3.172	3.344	3.275
MUE				**0.273**	**0.294**
MSE				**0.252**	**0.263**
	Mixed Complexes (7)				
16	ethene/ethine (C_{2v})	*ethene*	2.752	3.082	3.021
		ethane	2.752	2.899	2.859
17	benzene/water (C_s)	*benzene*	3.435	3.309	3.010
		water	3.435	3.063	3.012
18	benzene/ammonia (C_s)	*benzene*	3.592	3.463	3.166
		ammonia	3.592	3.170	3.121
19	benzene/HCN (C_s)	*benzene*	3.387	3.497	3.472
		HCN	3.387	3.592	3.575
20	benzene dimer (C_{2v})	*vertical*	3.513	3.728	3.700
		horizontal	3.513	3.769	3.728
21	indole/benzene T-shape (C_1)	*indole*	3.238	3.062	3.020
		benzene	3.238	3.228	3.050
22	phenol dimer (C_1)	*donor*	1.937	2.093	1.906
			4.921	4.705	4.904
		acceptor	1.937	2.024	1.828
			4.921	5.138	5.090
MUE				**0.198**	**0.220**
MSE				**0.017**	**−0.068**
	Overall				
MUE				**0.218**	**0.187**
MSE				**0.135**	**0.022**

[a] MUE (mean unsigned error); MSE (mean signed error). The definition of interaction distances follows ref 6.

binding energy of 0.20 Å and 5.4 kcal/mol, respectively. The corresponding PM3/MM hydrogen bond distances are systematically too long, with a mean signed error (MSE) in intermolecular distance of 0.17 Å (Table 3). Single-point PM3/MM energy calculations at the high level *ab initio* QM geometries do not significantly improve the systematically underestimated interaction energies (Supporting Information).

It is instructive to consider the case of water dimer. The binding energy of water dimer at the CCSD(T)/CBS level is −5.0 kcal/mol (Table 2). At the QM/MM level, where PM3 water is the proton donor, the binding energy is considerably underestimated, at −3.5 kcal/mol (Table 2); we note the same level of underbinding (−3.5 kcal/mol) is exhibited by the entirely PM3 water dimer.[20] Reversing the QM/MM model, such that PM3 water is the acceptor, the interaction is stronger by 0.4 kcal/mol. Applying the modified potential which uses eq 2, denoted here as PM3/MM*, interaction energies of −5.2 kcal/mol are obtained, regardless of whether QM water is donor or acceptor (Table 1). This is in good agreement with the CCSD(T)/CBS value. The improved hydrogen bond energy arises from enhanced $H_{QM}\cdots O_{MM}$ interactions which, due to choice of f_1^a and f_2^a (Table 1), are pronounced relative to $O_{QM}\cdots O_{MM}$ interactions. As observed above, the modified potential also addresses to some extent the variation in predicted hydrogen bond strength, depending on which water is treated as QM. Larger exponentials for f_1^a and f_2^a of 3.6 and 3.6 $Å^{-1}$ respectively when the acceptor oxygen is treated as QM (*i.e.* a $O_{QM}\cdots H_{MM}$ hydrogen bond) ensure a smaller contribution relative to the converse situation of a QM hydrogen donor, where f_1^a and f_2^a are 2.2 and 2.7 $Å^{-1}$, respectively. Thus, this approach reduces the imbalance in the QM/MM hydrogen bond in water dimer, compensating for underpolarization of the QM oxygen. In terms of geometry of the water dimer complex, where QM water is the proton donor, the O−H···O angle improves from 197.3° at PM3/MM to 187.6° at PM3/MM*. This is closer to the angle observed in the CCSD(T)/cc-pVQZ geometry of 172.8°. For QM water as the acceptor, a smaller improvement is seen in the O−H···O angle, from 188.5° to 181.7°. We note that the improved binding energy for water dimer does correspond to O···H distances of 1.79 and 1.77 Å for QM donor and acceptor, respectively (Table 3); these are about 0.15 Å shorter than the CCSD(T)/cc-pVQZ values and 0.20 Å shorter than PM3 values. However, by comparison, very short PM3/MM hydrogen bond distances of 1.62 and 1.66 Å are required to obtain interaction energies of −5.2 and −5.1 kcal/mol for the QM water as hydrogen bond donor and acceptor respectively (achieved through optimization of van der Waals' parameters of the QM atoms).

For ammonia dimer, there exists a sensitive balance of H···H and N···H noncovalent interactions which determine the optimal C_{2h} structure, a tilted geometry with N−H···N angles of 121.9° and H···N distances of 2.50 Å at the CCSD(T)/cc-pVQZ level (Figure 1). The symmetry of $(NH_3)_2$ is lost by optimization *via* PM3/MM, such that the N−H···N angles are 65.4° and 175.3° and the H···N distances are 2.22 and 3.53 Å (Table 3, Figure 1). Through use of the PM3/MM* potential, the tilted C_{2h} symmetry of the $(NH_3)_2$ complex is approximately recovered with H···N distances of 2.11 and 2.18 Å (Table 3, Figure 1), despite the inherently asymmetric treatment of the waters via QM and MM models. In this improved orientation, the binding energy of ammonia dimer correspondingly improves from −2.2 kcal/mol at PM3/MM to the high level *ab initio* value of −3.2 kcal/mol at the PM3/MM* level (Table 2).

The remaining hydrogen bonding complexes **3**−**7** observe geometries in closer agreement with high level *ab initio* values. For example, formamide dimer has distances of 1.87 and 1.89 Å at the PM3/MM* level (Table 3), in better agreement with

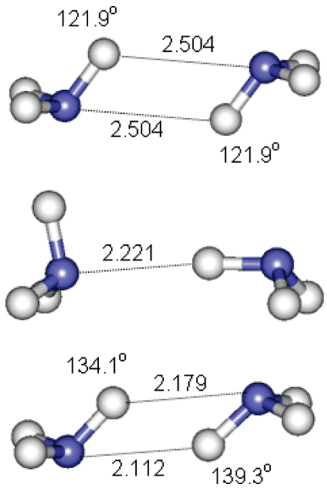

Figure 1. Minimum energy structure of ammonia dimer at the (a) CCSD(T)/aug-cc-pVQZ, (b) PM3/MM, and (c) PM3/MM* levels of theory. Distances in Å.

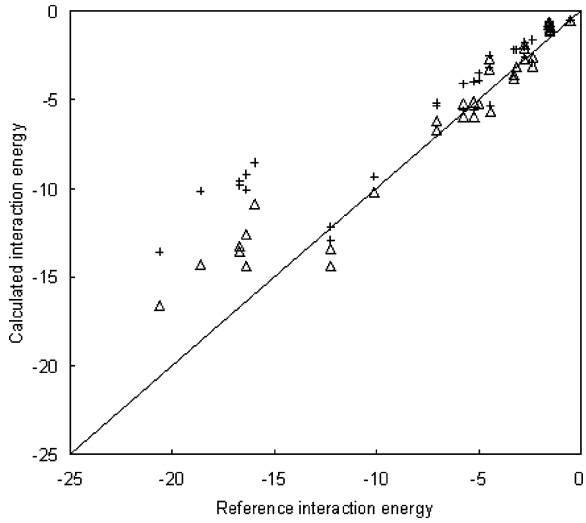

Figure 2. Interaction energies (kcal/mol) of S22 set calculated by CCSD(T)/complete basis set (reference), PM3/MM (+) and PM3/MM* (Δ) levels of theory.

the two CCSD(T)/aug-cc-pVTZ distances of 1.84 Å than PM3/MM values of 1.99 and 2.11 Å. Indeed, for hydrogen bonded systems **1**−**7**, the MUE in geometry improves from 0.20 Å to 0.09 Å, and the MSE reduces from 0.17 Å to −0.05 Å, showing a small underestimation (Table 3). This compares with a more modest overall improvement in geometry, from a MUE of 0.22 Å for the 22 molecules at PM3/MM to 0.19 Å at PM3/MM* (Table 3). Alongside improved geometries, the overall MUE in interaction energy for the S22 set reduces from 2.1 kcal/mol at PM3/MM to 1.2 kcal/mol at PM3/MM* (Table 2); this is illustrated by an improved correlation r^2 with *ab initio* QM binding energies, from 0.86 at PM3/MM to 0.94 at PM3/MM* (Figure 2). For the hydrogen bonded complexes **1**−**7**, the improvement is particularly apparent, with a reduction in MUE from 5.4 kcal/mol at the PM3/MM level to 2.6 kcal/mol *via* PM3/MM* (Table 2).

As expected, less improvement in modeling structures and energetics is observed for dispersion-dominant and mixed interaction sets. However, the errors in energy at the PM3/MM level are already rather smaller relative to hydrogen bonded

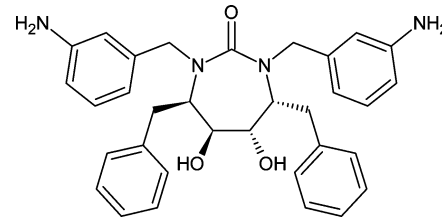

Figure 3. HIV-1 protease inhibitor, mozenavir.

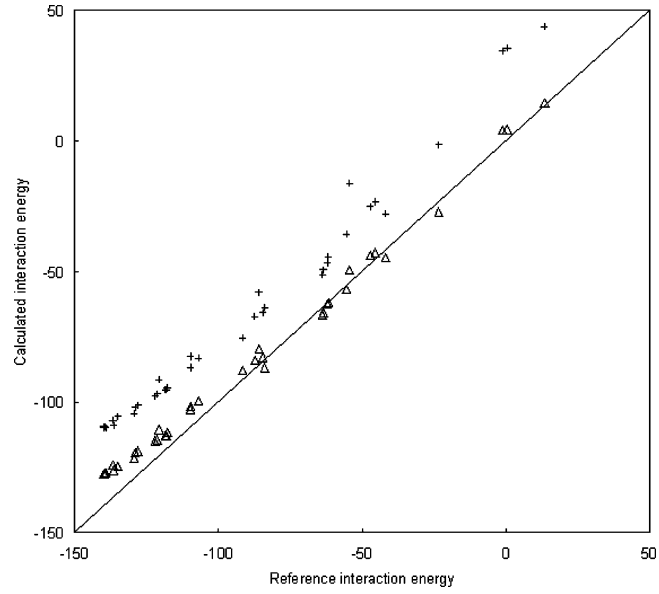

Figure 4. Comparison of PM3/MM (+) and PM3/MM* (Δ) models with HF/6-31G* (reference) QM/MM model in predicting total electrostatic interaction energies (kcal/mol) of native and decoy ligand poses with HIV-1 protease.

Table 4. Mean Unsigned Error (MUE), Mean Signed Error (MSE), Range of Error (Δ*E*), and Correlation r^2 for Electrostatic Interaction Energies of Native and Decoy Ligand/HIV-1 Protease Complexes Computed at the PM3/MM and PM3/MM* Levels of Theory, Relative to Reference *Ab Initio* QM/MM Calculations

method	MUE	MSE	Δ*E*	r^2
PM3/MM	25.20	25.20	25.39	0.85
PM3/MM*	6.29	5.53	15.71	0.96

complexes **1**−**7**. For complexes of mixed interaction type **16**−**22**, a modest decrease in mean unsigned error of binding energy from 1.1 kcal/mol for PM3/MM to 0.7 kcal/mol for PM3/MM* is observed (Table 2). However, no improvement in performance is found for dispersion dominant complexes **8**−**15**, where a small increase in MUE by 0.1 kcal/mol is observed. The mean unsigned errors in geometry for mixed and dispersion-dominant complexes are marginally poorer using the modified QM/MM potential, by 0.02 Å in both cases (Table 3). The subtle dispersive interactions of these complexes are treated in the QM/MM models rather simplistically *via* the QM/MM van der Waals' potential. The largest error in geometry of any bimolecular complex in the S22 set, for both PM3/MM and PM3/MM* models, is found for the indole/benzene stacked dimer (complex **14**). Here, the stacked conformation of the MP2/aug-cc-pVTZ geometry is distorted into a proto-T-shape conformation at the PM3/MM and PM3/MM* levels of theory (complex

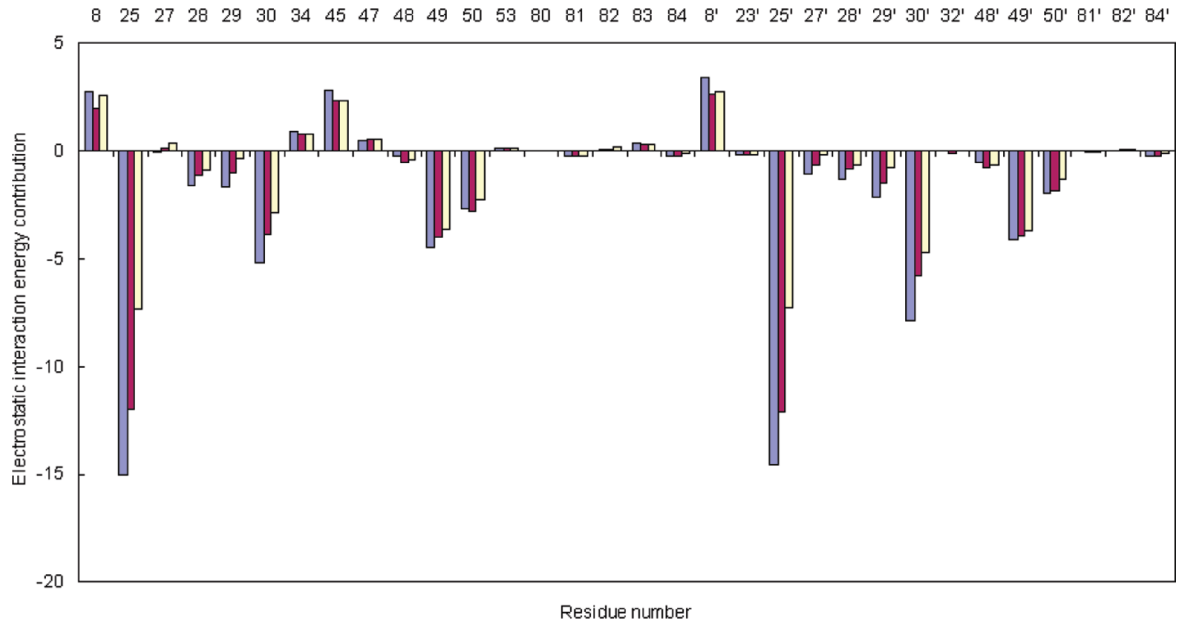

Figure 5. Electrostatic interaction energy contribution (kcal/mol) of selected residues to binding of mozenavir in its native pose, using PM3/MM (white), PM3/MM* (dark gray), and HF/6-31G*:MM (light gray).

21), leading to intercentroid distances of around 1 Å larger than the *ab initio* QM values (Table 3 - note that interaction distances for complexes **14** and **21** are defined differently). This reflects for this complex an imbalance of polar and dispersive interactions in both the semiempirical QM/MM models.

Having adopted a reformulated QM core-MM charge term and applied the resulting PM3/MM* model to the S22 set, we now consider the case of a protein−ligand interaction, specifically that of HIV-1 protease with mozenavir (Figure 3).

The native and 39 non-native structures of the complex have been studied previously by us, using a QM/MM approach, with the QM region (the ligand) described both at the HF/6-31G* and PM3 levels of theory.[21] In accord with available experimental evidence, the catalytic aspartyl dyad was modeled in its neutral form. The ligand conformations adopted in the non-native structures span a root-mean-square deviation in atom positions from 0.15 to 6.86 Å from the native pose. It was found that PM3/MM calculations underestimated the affinity of protein−ligand interactions relative to the *ab initio* QM/MM model quite significantly (Figure 4), with a MSE and MUE of 25.2 kcal/mol relative to HF for the 40 structures.[21]

Application of the modified PM3/MM* potential to the set of mozenavir/HIV-1 protease structures reduces the MSE in interaction energy to 5.5 kcal/mol and MUE to 6.3 kcal/mol; the range in error is also 10 kcal/mol smaller (Table 4). The correlation, r^2, between calculated semiempirical and *ab initio* QM/MM electrostatic interaction energies improves from 0.85 to 0.96.

To obtain insight into contributions of individual amino acid residues to binding mozenavir in its native pose, MM charges of individual key amino acids were deleted and the electrostatic interaction energy difference obtained (Figure 5). Generally more quantitative results are found for PM3/MM* relative to the PM3/MM model. For example, at the *ab initio* QM/MM level, the catalytic residues Asp25 and Asp25′ are predicted to contribute −15.0 and −14.6 kcal/mol, respectively, to the binding of mozenavir (Figure 5). Whereas the PM3/MM model finds contributions of −7.4 and −7.3 kcal/mol for Asp25 and Asp25′, the PM3/MM* method obtains closer agreement with the HF model, with respective estimates of −12.0 and −12.1 kcal/mol. Similar improvement is found for other amino acid contributions, with an overall reduction in MUE from 1.0 to 0.5 kcal/mol. We note that PM3/MM* calculations are in reasonable agreement with HF/6-31G*/MM interaction energies for protein−ligand complexes as well as much higher level CCSD(T)/CBS calculations for the S22 small molecule complexes. Although the reasons for this are not entirely clear, agreement may in part stem from the consistency of the AMBER point charge model with the HF/6-31G* level of theory and from a fortuitous balance struck by the HF/6-31G* model.

4. Conclusions

In this Letter, we have introduced a semiempirical QM/MM potential with reformulated QM core-MM charge interactions. Application of the QM/MM potential shows improved prediction of geometry and interaction energy for hydrogen bonded small molecule complexes typical of biomolecular interactions. This is without significant deterioration in the modeling of dispersive interactions. Further refinement in the approach can be envisaged - for example, here we introduced a specific f_1^a and f_2^a set for H_{QM}-H_{MM} interactions (Table 1). This essentially bond-specific approach could be generalized to other QM-MM atom type pairs, although, in doing so, one introduces more parameters and may reduce the generality of the approach. The modification to the core−core expression we have proposed here in eq 2 is readily applicable to existing semiempirical QM/MM codes. While we have explored here improved handling of hydrogen bonding through this QM/MM core−core expression, we note that an interesting alternative approach could be to recapture the original intent of the core−core expression as used in purely QM approaches, to model only short-range repulsion between atoms (dealt with in current QM/MM formulations principally via the r^{-12} component of the QM/MM Lennard-Jones

potential). Combined with other developments, such as improved QM-QM dispersion (PM3-D, OMx),[5,6] the capability to model metals (*e.g.* PM6[22]), and focused parametrizations (*e.g.* PM3CARB-1 for modeling carbohydrates[23]), semiempirical QM/MM methods remain a powerful cost-effective approach to solving a range of important biochemical and biophysical problems.

Acknowledgment. We thank Jonathan McNamara for useful discussions.

Supporting Information Available: Single point interaction energies of S22 set calculated by PM3 QM/MM models at high level *ab initio* geometries. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198.

(2) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227.

(3) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700.

(4) Kolb, M.; Thiel, W. Beyond the MNDO Model - Methodical Considerations and Numerical Results. *J. Comput. Chem.* **1993**, *14*, 775.

(5) Weber, W.; Thiel, W. Orthogonalization Corrections for Semiempirical Methods. *Theor. Chem. Acc.* **2000**, *103*, 495.

(6) McNamara, J. P.; Hillier, I. H. Semi-Empirical Molecular Orbital Methods Including Dispersion Corrections for the Accurate Prediction of the Full Range of Intermolecular Interactions in Biomolecules. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2362.

(7) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. PDDG/PM3 and PDDG/MNDO: Improved Semiempirical Methods. *J. Comput. Chem.* **2002**, *23*, 1601.

(8) Voityuk, A. A.; Rosch, N. AM1/d Parameters for Molybdenum. *J. Phys. Chem. A* **2000**, *104*, 4089.

(9) Winget, P.; Horn, A. H. C.; Selcuki, C.; Martin, B.; Clark, T. AM1*Parameters for Phosphorus, Sulfur and Chlorine. *J. Mol. Model.* **2003**, *9*, 408.

(10) Sharma, R.; McNamara, J. P.; Raju, R. K.; Vincent, M. A.; Hillier, I. H.; Morgado, C. A. The Interaction of Carbohydrates and Amino Acids With Aromatic Systems Studied by Density Functional and Semi-Empirical Molecular Orbital Calculations With Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2767.

(11) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods IV:Extension of MNDO, AM1, and PM3 to More Main Group Elements. *J. Mol. Model.* **2004**, *10*, 155.

(12) Pople, J. A.; Santry, D. P.; Segal, G. A. Approximate Self-Consistent Molecular Orbital Theory. I. Invariant Procedures. *J. Chem. Phys.* **1965**, *43*, S129.

(13) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. Combining Quantum Mechanics Methods With Molecular Mechanics Methods in ONIOM. *J. Chem. Theory Comput.* **2006**, *2*, 815.

(14) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. Benchmark Database of Accurate (MP2 and CCSD(T) Complete Basis Set Limit) Interaction Energies of Small Model Complexes, DNA Base Pairs, and Amino Acid Pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.

(15) Stewart, J. J. P. Optimisation of Parameters for Semi-Empirical Methods. 2. Applications. *J. Comput. Chem.* **1989**, *10*, 221.

(16) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668.

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

(18) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623.

(19) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(20) Bernal-Uruchurtu, M. I.; Martins-Costa, M. T. C.; Millot, C.; Ruiz-Lopez, M. F. Improving Description of Hydrogen Bonds at the Semiempirical Level: Water-Water Interactions As Test Case. *J. Comput. Chem.* **2000**, *21*, 572.

(21) Fong, P.; McNamara, J. P.; Hillier, I. H.; Bryce, R. A. Assessment of QM/MM Scoring Functions for Molecular Docking to HIV-1 Protease. *J. Chem. Inf. Model.* **2009**, *49*, 913.

(22) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007**, *13*, 1173.

(23) McNamara, J. P.; Muslim, A. M.; Abdel-Aal, H.; Wang, H.; Mohr, M.; Hillier, I. H.; Bryce, R. A. *Chem. Phys. Lett.* **2004**, *394*, 429.

CT9002674