# JCTC Journal of Chemical Theory and Computation

# Performance Evaluation of the Three-Layer ONIOM Method: Case Study for a Zwitterionic Peptide

Keiji Morokuma,*,† Qingfang Wang,† and Thom Vreven‡

*Cherry L. Emerson Center for Scientific Computation and Department of Chemistry, Emory University, Atlanta, Georgia 30322, and Gaussian, Inc., 340 Quinnipiac Street, Building 40, Wallingford, Connecticut 06492*

**Abstract:** The performance of the three-layer ONIOM method was systematically investigated by comparing the optimized geometries and calculated deprotonation energy of a zwitterionic peptide molecule, $NH_3^+-CH^nBu-CO-NH-CH_2-CO-NH-CH^nBu-COO^-$, using all possible combinations of B3LYP/6-31G* as the high-level quantum (HQ), AM1 as the low-level quantum (LQ), and Amber as the MM method. Results show that the three-layer ONIOM(HQ:LQ:MM) method, which includes a medium-level quantum method in the middle system to take into account the electronic effects of the middle layer and to keep the problematic QM-MM boundary away from the action region, is more reliable and more stable than the QM:MM:MM or generic QM/MM method and is the best compromise between accuracy and computational cost.

## I. Introduction

As theoretical/computational chemistry has gained a major role in studies of chemical problems in the last few decades, it became a challenge for theoreticians to accurately treat large molecular systems such as encountered in biochemistry. No single theoretical method is able to provide both the accuracy and acceptable computational cost that are required for the investigation of such chemical processes. Accurate ab initio quantum mechanics (QM) methods either scale nonlinearly with the size of the system, or have large pre-factors that prevent them to be applied to large systems. Low cost molecular mechanics (MM) methods are widely used and scale linearly but have the obvious weakness of the poor description of bond breaking and forming processes. Meanwhile, in many biological systems the actual reaction only localizes in a relatively small region.

A breakthrough came with the realization that it is not necessary to use a single computational method, which resulted in the development and application of a variety of hybrid methods. These methods utilize the localization feature of large reaction systems, and most combine the merit of accuracy of a QM method with the low cost of MM methods.[1−3] Our ONIOM method is a versatile and popular hybrid method.[4−15] ONIOM divides the system into several onion-like layers, treating the active center with the highest level ab initio QM method, while outer layers can be treated with less expensive methods, such as low-level ab initio QM, semiempirical QM, or MM methods.

Figure 1 illustrates the basic concept of the multilayered ONIOM method, taking three-layer ONIOM as an example. The three-layer ONIOM(high:medium:low) energy is an approximation to the energy at the high level for the real system, $E_{real}^{high}$, referred to as the target, and is given by

$$E_{real}^{ONIOM} = E_{model}^{high} + E_{middle}^{medium} + E_{real}^{low} - E_{model}^{medium} - E_{middle}^{low} \quad (1)$$

where *high*, *medium*, and *low* refer to the high-, medium-, and low-level theoretical methods, respectively, while *model*, *middle*, and *real* refer to the model, middle, and real systems, respectively. The middle system is a part cut from the real system, and the model system is a part cut from the middle system. The model and middle systems are mended by *link atoms* to satisfy the valencies if covalent bonds are cut. With the term "energy", we typically refer to a relative energy, such as the binding energy or the barrier height.

The accuracy of the ONIOM method is defined as the error of ONIOM relative to the target calculation (the high-level

---

\* Corresponding author e-mail: morokuma@emory.edu.

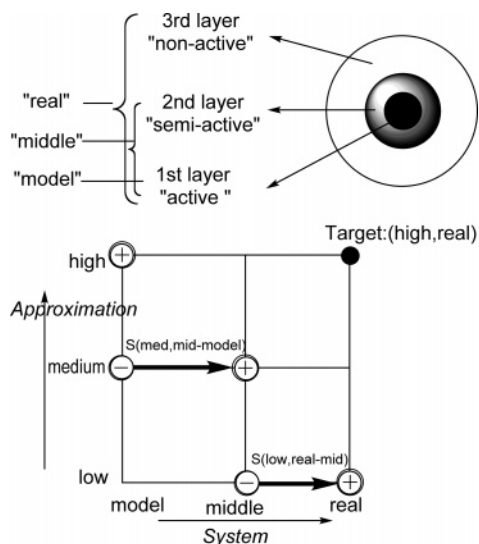† Emory University.

‡ Gaussian, Inc.

**Figure 1.** The layer partition and energy extrapolation scheme of the three-layer ONIOM method.

result for the real system):

$$\text{Err}_{\text{real}}^{\text{ONIOM}} = E_{\text{real}}^{\text{ONIOM}} - E_{\text{real}}^{\text{high}} \qquad (2)$$

$\text{Err}_{\text{real}}^{\text{ONIOM}}$ depends critically on the partitioning of the system into middle and model systems and the reliability of medium- and low-level methods used in ONIOM. The accuracy of any ONIOM combination can be tested using the $S$-values,[11] which are defined as

$$S_{\text{real}-\text{middle}}^{\text{low}} = E_{\text{real}}^{\text{low}} - E_{\text{middle}}^{\text{low}}$$

$$S_{\text{middle}-\text{model}}^{\text{medium}} = E_{\text{middle}}^{\text{medium}} - E_{\text{model}}^{\text{medium}}$$

$$S_{\text{real}-\text{model}}^{\text{high}} = E_{\text{real}}^{\text{high}} - E_{\text{model}}^{\text{high}} \qquad (3)$$

With these definitions of the $S$-values, the ONIOM energy can be written as

$$E_{\text{real}}^{\text{ONIOM}} = E_{\text{model}}^{\text{high}} + S_{\text{middle}-\text{model}}^{\text{medium}} + S_{\text{real}-\text{middle}}^{\text{low}} \qquad (4)$$

Therefore

$$\text{Err}_{\text{real}}^{\text{ONIOM}} = S_{\text{real}-\text{model}}^{\text{high}} - S_{\text{middle}-\text{model}}^{\text{medium}} - S_{\text{real}-\text{middle}}^{\text{low}}$$

$$= [S_{\text{middle}-\text{model}}^{\text{high}} - S_{\text{middle}-\text{model}}^{\text{medium}}] +$$

$$[S_{\text{real}-\text{middle}}^{\text{high}} - S_{\text{real}-\text{middle}}^{\text{low}}] \qquad (5)$$

$$= \Delta S_{\text{middle}-\text{model}}^{\text{high}-\text{medium}} + \Delta S_{\text{real}-\text{middle}}^{\text{high}-\text{low}}$$

Thus, the error of the three-layer ONIOM approximation is zero if both $S_{\text{middle}-\text{model}}^{\text{high}} = S_{\text{middle}-\text{model}}^{\text{medium}}$ and $S_{\text{real}-\text{middle}}^{\text{high}} = S_{\text{real}-\text{middle}}^{\text{low}}$, namely, if the energy difference between the middle and the model at the medium level reproduces that of the high level and at the same time the energy difference between the real and the middle systems at the low level reproduces that of the high level. The $S$-value is therefore a useful tool in the calibration of hybrid methods. Of course one can benchmark hybrid methods also by calculating the error simply from the target result and the hybrid method

result, but the $S$-value test has some advantages. First, it allows for a systematic and clear investigation of different method combinations and partitions, as we will show in the current paper. Second, for the three-layer ONIOM, it allows for investigation of the performance of the low and medium levels of theory individually. From eq 5 it is clear that the error in three-layer ONIOM can be zero when $\Delta S_{\text{middle}-\text{model}}^{\text{high}-\text{medium}}$ is the negative of $\Delta S_{\text{real}-\text{middle}}^{\text{high}-\text{low}}$, but neither is zero. Such cancellation of errors is undesirable and may be missed with a simple calibration based on only the target and hybrid method results.

Below we will systematically apply the three-layer, two-layer, and single-level (non-ONIOM, conventional) calculations for a sample test molecule. We just mention standard (one-layer) and two-layer calculations can be considered to be a special case of three-layer ONIOM in which two or more levels happen to be equal. For instance, ONIOM(high:low:low) is equivalent to the two-layer ONIOM(high:low) method using the real and the model system, and the same type of $S$-value analysis can be performed.

Two different approaches exist for the treatment of electrostatic interaction between the QM layers and the MM layer.[10,16] In the default ONIOM scheme, the electrostatic interaction is included in the MM calculations, which is referred to as *mechanical embedding* (ONIOM-ME). The energy expression for the three-layer method with MM as the low-level, is then

$$E_{\text{real}}^{\text{ONIOM}-\text{ME}} = E_{\text{model}}^{\text{high}} + E_{\text{middle}}^{\text{medium}} + E_{\text{real}}^{\text{MM}} - E_{\text{model}}^{\text{medium}} - E_{\text{middle}}^{\text{MM}} \qquad (6)$$

Here $E_{\text{real}}^{\text{MM}}$ includes the molecular mechanics electrostatic interaction in the full (real) system ($\text{ES}_{\text{real}}$), and $E_{\text{middle}}^{\text{MM}}$ includes the molecular mechanics electrostatic interaction for the middle system ($\text{ES}_{\text{middle}}$). Terms not involving the interactions between the QM parts and the MM part mostly cancel out between these two ES terms. $\text{ES}_{\text{real}}$ and $\text{ES}_{\text{middle}}$ are given as

$$\text{ES}_{\text{system}} = \sum_{J<K}^{\text{system}} \frac{s_{JK} q_J q_K}{r_{JK}} \qquad (7)$$

where $q_J$ is the partial charge on the atom $J$, $r_{JK}$ is the distance between atoms $J$ and $K$, and summation $J$ and $K$ are for all the atoms in the *system* (real or middle). Here $s_{JK}$ is the scale factor of electrostatic interaction introduced in most force field definitions (independent of the ONIOM or QM/MM applications). For instance in the Amber force field, it is 0, 1/1.2, and 1 for $JK$ atom pairs separated by one and two bonds (first and second neighbors), three bonds (third neighbors), and more than three bonds (fourth and further neighbors), respectively. The same scale factor is usually used for the generic QM/MM-ME or ONIOM-ME schemes.

In the electronic embedding ONIOM-EE scheme,[10] one includes the electrostatic interaction between the QM charges and the MM part as additional scaled electrostatic Hamiltonians $H'$ of the QM systems

Performance of the Three-Layer ONIOM Method

*J. Chem. Theory Comput., Vol. 2, No. 5, 2006* **1319**

$$H'_{system} = \sum_i \sum_J^{systemMM} \frac{\sigma_J q_J}{r_{iJ}} \qquad (8)$$

where system refers to model and middle systems, and $i$ refers to the electrons in that system. There is a fundamental difference between the scaling factors in the two eqs 7 and 8. In eq 7 individual interactions are scaled based on the number of covalent bonds between the atom pairs. This cannot be converted to a QM Hamiltonian. Therefore, in the screened Hamiltonian, eq 8, individual MM atomic charges are scaled by factor $\sigma_J$. In the present paper, we use $\sigma_J = 0$ for the first and second classical neighbor of the quantum link atom host (atom replaced by the link atom in the QM model system) and $\sigma_J = 1$ for the third and further classical neighbor. In addition to including the electrostatic interaction energy between the layers (which ONIOM-ME also does), electronic embedding also polarizes the wave function of the QM region by the MM atomic charges. It is therefore supposed to be better in describing a very polar system. However, the EE scheme is more expensive than the ME scheme, because a self-consistent charge iteration is required. Although the basic ideas of the QM/MM-EE and ONIOM-EE are the same, there are some subtle differences in cancellation of terms, and a detailed discussion on the comparison of ONIOM(QM:MM) and generic QM/MM methods has recently been published.[10]

Regardless of whether the ME or EE scheme is used, the arbitrary screening of an otherwise strong short-range electrostatic interaction, possible contributions of link atoms, and other coupling terms (for example the cancellation problem, discussed in detail in ref 10) between the QM and the MM regions is one of the weakest points of either the generic QM/MM or the ONIOM(QM:MM) scheme and affects their performance depending on the scaling and coupling schemes adopted. The most important advantage of the three-layer ONIOM(QM:QM:MM) method is that the inexpensive low-level QM layer pushes the problematic QM-MM boundary far away from the high-level QM region, and the arbitrariness in electrostatic scaling and coupling does not affect very much the final results for the events (or properties) that take place in the model region. Furthermore, often the ME and EE schemes would give similar results, since the polarization by long-distance MM charges does not significantly change the wave function in the vicinity of the action, in which case one would be able to avoid the somewhat more expensive EE scheme.

Currently two- and three-layer ONIOM methods are implemented in the Gaussian program.[17] The purpose of this paper is to present a systematic case study of the performance of the three-layer ONIOM(HQ:LQ:MM), where HQ is the high-level QM method and LQ is the low-level QM method, in comparison with single-level methods and various two-layer ONIOM combinations. We chose the proton dissociation energy of the zwitterionic form of peptide $NH_3^+-CHR_1-CO-NH-CH_2-CO-NH-CHR_2-COO^-$ to form $NH_2-CHR_1-CO-NH-CH_2-CO-NH-CHR_2-COO^-$ as our test system. The intramolecular charge separation makes this a good candidate to evaluate the performance of various ONIOM method combinations for very polar systems.

## II. Computational Methods

Peptides can exist either in the neutral form or the zwitterionic form (protonated at N-terminus and deprotonated at C-terminus). Usually for small peptides, the zwitterionic form is stabilized by solvent molecules or crystal force, while in the gas phase the neutral form is favored.[18] Our calculation at the B3LYP/6-31G* level for $NH_3^+-CHR_1-CO-NH-CH_2-CO-NH-CHR_2-COO^-$ in the gas phase shows that when both $R_1$ and $R_2$ are $n$-butyl ($^nBu$) groups, the zwitterionic form is a local minimum structure, and we use this as our case study system. The deprotonation energy of the zwitterionic form $NH_3^+-CH^nBu-CO-NH-CH_2-CO-NH-CH^nBu-COO^-$ at the $H_3N^+$ position, defined as $\Delta E = E(NH_2\cdots COO^-) + E(H^+) - E(NH_3^+\cdots COO^-)$, was calculated by one-layer, two-layer, and three-layer ONIOM methods using all possible combinations of B3LYP/6-31G* as high-level QM (called HQ throughout the paper), AM1 as low-level QM (LQ), and Amber as the MM method. The geometry of the zwitterionic form $NH_3^+\cdots COO^-$ was fully optimized by the respective ONIOM methods, and the geometry of $NH_2\cdots COO^-$ was assumed unchanged from that of $NH_3^+\cdots COO^-$ optimized geometry, removing the proton at H4, the proton closest to the Bu group on $NH_3^+$. All the calculations were performed using the Gaussian 03 package.[17]

Three partition schemes, depicted in Figure 2, were adopted. Partition C adopts a minimal model system consisting of only the reaction site $H_3N^+L$ (H: link H atom) as well as the minimal middle system consisting of the reaction site plus the next residue $H_3N^+CHL-CO-NHL$. This partition is just examined as to how small one can make the models and is not likely to be adopted in practical calculations. In addition, QM:MM:MM calculations with this partition will suffer from the aforementioned cancellation problem. Partition B, on the other hand, adopts the reaction site plus the next residue $H_3N^+CHL-CO-NHL$ as the model system for high-level calculation, and the middle system $NH_3^+-CH^nBu-CO-NH-CH_2-CO-NHL$ contains one more residue. Partition A adopts the same model system, $H_3N^+CHL-CO-NHL$ as in partition B, but uses a middle system $(H_3N^+)CHL-CO-NH-CH_2-CO-NH-CHL-(COO^-)$ that contains all the ionic components of the system. Even when the real system is much larger as in the case of polypeptides, these model and middle systems would be reasonable choices for practical ONIOM applications. Most of the calculations were performed with the mechanical embedding (ME). Electronic embedding (EE) calculations were performed only for the DFT:MM combinations. We did not perform the EE calculations involving the semiempirical AM1 method, since the one-electron integral corresponding to the QM-MM electrostatic Hamiltonian, eq 8, has not been implemented.

## III. Results and Discussions

**1. Optimized Geometry in Partition B.** At first we want to compare the optimized geometries of the present zwitterionic peptide, $NH_3^+-CH^nBu-CO-NH-CH_2-CO-NH-CH^nBu-COO^-$, using different ONIOM combinations in partition B. The fully optimized geometries are given in the Supporting Information, and Table 1 gives the root-mean-

## Partition A

model (dark) = $(H_3N^+)$CHL-CO-NHL , L is link hydrogen.
mid (dark+light) = $(H_3N^+)$CHL-CO-NH-$CH_2$-CO-NH-CHL(COO⁻)
real (all) = $(H_3N^+)$CHBu-CO-NH-$CH_2$-CO-NH-CHBu(COO⁻)
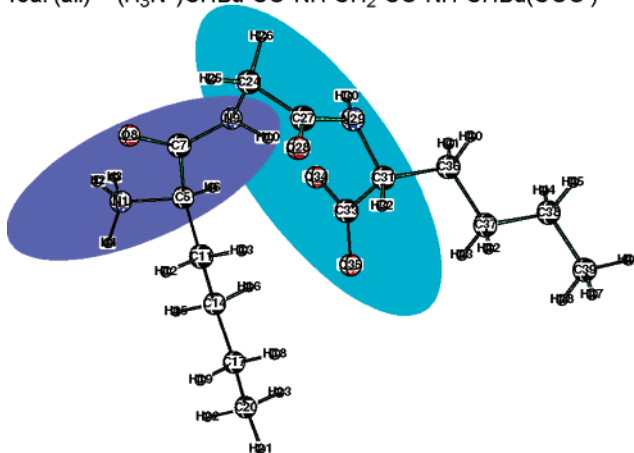
## Partition B

model (dark) = $(H_3N^+)$CHL-CO-NHL , L is link hydrogen.
mid (dark+light) = $(H_3N^+)$CHBu-CO-NH-$CH_2$-CO-NHL
real (all) = $(H_3N^+)$CHBu-CO-NH-$CH_2$-CO-NH-CHBu(COO⁻)
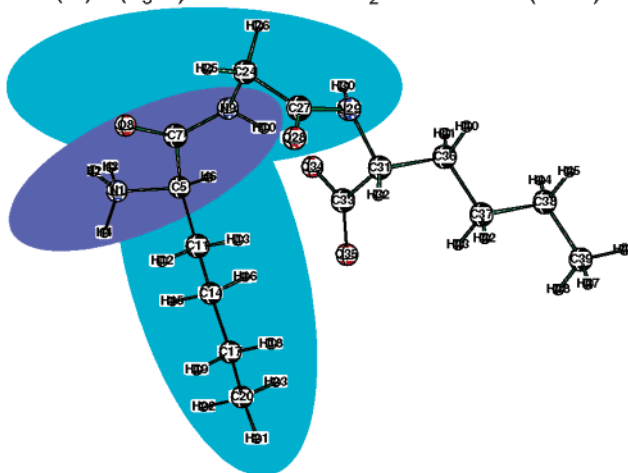
## Partition C

model (dark) = $(H_3N^+)$L , L is link hydrogen.
mid (dark+light) = $(H_3N^+)$CHL-CO-NHL
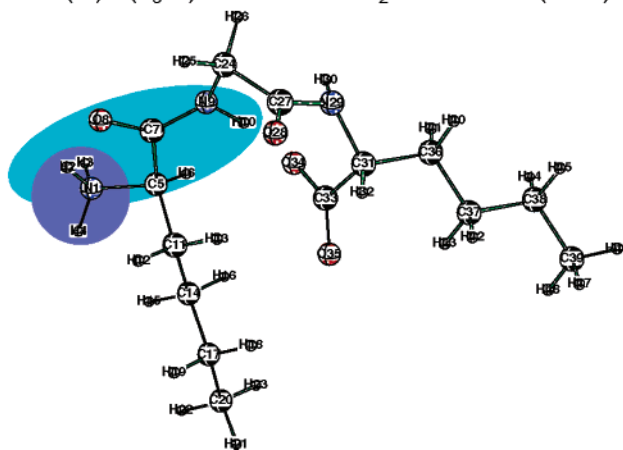real (all) = $(H_3N^+)$CHBu-CO-NH-$CH_2$-CO-NH-CHBu(COO⁻)

**Figure 2.** Three partition schemes used for the three-layer ONIOM calculations

square (RMS) deviations of all the bond distances, bond angles, and dihedral angles of the model system (excluding link hydrogens) and the real system, respectively, from those of the target calculation, HQ:HQ:HQ, which is nothing but the pure B3LYP/6-31G* calculation. One notices that both the pure Amber MM method, MM:MM:MM, and the pure semiempirical AM1 method, LQ:LQ:LQ, give very large deviations in geometry optimization of the entire peptide as well as those of the model system; among the two, Amber seems to do a little better than AM1. Among various ONIOM combinations, HQ:HQ:MM gives the smallest error. All the ONIOM combinations using HQ in the model system give rather small deviations in the model part of the geometry, which is expected but not necessarily automatic. An interesting finding is that this is true even for the geometry of the entire peptide, the real system; the correct geometry of the model system seems to dictate the errors in the rest of the system.

Among the methods using MM, the use of the Mulliken charge seems to give somewhat better geometries than the RESP charge; this is opposite to the trend in deprotonation energy to be discussed later. It is noted that when the MM is adopted already in the middle system, i.e., in HQ:MM:MM, the electronic embedding (EE) gives a substantially better geometry than the mechanical embedding (MM), in particular in the geometry of the model system. However, when the MM is used only in the real system, i.e., HQ:HQ:MM, the differences between EE and ME are negligible, because, as suggested earlier, the problematic boundary between QM and MM layers is now on the outside peripheral of the middle system and is located far away from the model system, and the polarization of the QM layer due to the MM charges becomes less important.

**2. Deprotonation Energy in Partition B.** Results of the deprotonation energies are shown in Table 2. The deprotonation energy of 322.09 kcal/mol at the pure B3LYP/6-31G* level is the target result which ONIOM approximations are trying to reproduce. At first we pay attention to the combinations without MM. The most expensive HQ:HQ:LQ combination, the two-layer HQ:LQ method with the large HQ region (up to the mid system), as expected gives the smallest error of only −4.4 kcal/mol or only 1.3% of the deprotonation energy. If one can afford a large HQ region (HQ:HQ) as well as LQ for the entire system, obviously this is an excellent approximation. The next level of approximation, HQ:LQ:LQ, gives a little larger error of −6.9 kcal/mol, with a smaller cost. The LQ:LQ:LQ or pure semiempirical AM1 calculation is not worth considering as this method is unable to describe the deprotonation reaction even qualitatively, with an absolute error of over 80 kcal/mol. These combinations are all quantum calculations and are likely to remain to be too expensive (see the rough estimated cost in Table 1) in the near future as tools for exploring potential energy surfaces of reactions of most very large (> thousands of atoms) biological systems.

In most real biomolecular calculations, one will have to use MM as the lowest level method for at least a part of the very large system. Of the methods that use MM, HQ:HQ:MM with the RESP[19] charges in the mechanical embedding

**Table 1.** RMS Errors of ONIOM Optimized Geometries (Relative to the Target Calculation HQ:HQ:HQ) of the $NH_3^+-C^nBuH-CO-NH-CH_2-CO-NH-CH^nBu-COO^-$ System with Partition B

| combination[a] | estimated cost[b] | atoms in the model system only | | | all atoms in the real system | | |
|---|---|---|---|---|---|---|---|
| | | 9 bond lengths (Å) | 13 bond angles (deg) | 10 dihedral angles (deg) | 47 bond lengths (Å) | 87 bond angles (deg) | 98 dihedral angles (deg) |
| HQ:HQ:HQ | 10000 | | | | | | |
| HQ:HQ:LQ | 200 | 0.009 | 1.34 | 9.04 | 0.014 | 1.27 | 6.32 |
| HQ:HQ:MM | 100 | 0.017 | 1.20 | 2.87 | 0.010 | 1.24 | 5.38 |
| | | (0.008) | (1.13) | (2.67) | (0.009) | (1.29) | (4.98) |
| HQ:HQ:MM (EE) | 500 | 0.017 | 1.19 | 2.85 | 0.010 | 1.24 | 5.38 |
| | | (0.009) | (1.17) | (2.77) | (0.010) | (1.31) | (5.08) |
| HQ:LQ:LQ | 110 | 0.012 | 1.62 | 14.27 | 0.022 | 1.70 | 8.04 |
| HQ:LQ:MM | 21 | 0.019 | 1.51 | 4.62 | 0.018 | 1.62 | 6.10 |
| | | (0.016) | (1.32) | (4.97) | (0.018) | (1.70) | (5.58) |
| HQ:MM:MM | 10 | 0.018 | 1.54 | 2.28 | 0.013 | 1.55 | 6.49 |
| | | (0.013) | (2.26) | (4.68) | (0.013) | (1.94) | (6.19) |
| HQ:MM:MM(EE) | 50 | 0.015 | 1.38 | 2.97 | 0.012 | 1.69 | 8.05 |
| | | (0.012) | (2.11) | (5.30) | (0.011) | (2.23) | (9.23) |
| LQ:LQ:LQ | 100 | 0.035 | 3.74 | 28.40 | 0.025 | 2.29 | 15.10 |
| LQ:LQ:MM | 11 | 0.033 | 3.77 | 23.87 | 0.022 | 2.22 | 11.62 |
| | | (0.027) | (3.44) | (19.64) | (0.021) | (3.42) | (10.26) |
| LQ:MM:MM | 2 | 0.032 | 3.87 | 17.41 | 0.018 | 2.20 | 9.33 |
| | | (0.028) | (4.12) | (16.59) | (0.017) | (3.27) | (9.31) |
| MM:MM:MM | 1 | 0.031 | 3.87 | 27.85 | 0.017 | 2.06 | 12.10 |
| | | (0.022) | (2.96) | (18.85) | (0.015) | (3.64) | (10.51) |

[a] HQ = B3LYP/6-31G*, LQ = AM1, MM = Amber. For the combinations involving the MM method, B3LYP/6-31G* RESP and Mulliken charges are adopted in the Amber calculation for values without and with parentheses, respectively, and mechanical embedding (ME) is used unless specified as electronic embedding (EE). [b] Very rough estimate of relative cost for a very large system, based on assumed cost: MM = $(10^{-3}, 10^{-2}, 1)$, LQ = $(1, 10, 10^2)$, HQ = $(10, 10^2, 10^4)$ for (model,mid,real) systems for ME, respectively. For EE the time for HQ and LQ calculations was multiplied by a factor of 5 to reflect the charge-iteration process.

**Table 2.** One-, Two-, and Three-Layered ONIOM Calculations for the Deprotonation Energy (in kcal/mol) of $NH_3^+-C^nBuH-CO-NH-CH_2-Co-NH-CH^nBu-COO^-$ System with Partition B Using the Optimized Geometries by the Respective Methods

| combination[a] | E(high/model) | E(med/mid) | E(med/model) | E(low/real) | E(low/mid) | E(ONIOM) | S(med/mid-model) | S(low/real-mid) |
|---|---|---|---|---|---|---|---|---|
| HQ:HQ:HQ | 241.70 | 245.51 | 241.70 | 322.09 | 245.51 | 322.09 | 3.81 | 76.59 |
| HQ:HQ:LQ | 244.04 | 248.18 | 244.04 | 244.96 | 175.48 | 317.66 | 4.15 | 69.48 |
| HQ:HQ:MM | 243.25 (238.86) | 243.11 (243.25) | 238.68 (238.86) | 107.13 (25.80) | 68.50 (−10.70) | 281.75 (279.75) | 4.43 (4.39) | 38.63 (36.51) |
| HQ:HQ:MM (EE) | | 244.47 (231.47) | | 69.56 (25.88) | 107.14 (−20.67) | 282.05 (278.02) | | 37.58 (46.55) |
| HQ:LQ:LQ | 243.61 | 174.55 | 172.43 | 244.03 | 174.55 | 315.21 | 2.11 | 69.48 |
| HQ:LQ:MM | 239.82 (239.29) | 170.29 (170.05) | 168.31 (168.08) | 104.93 (24.75) | 67.25 (−10.99) | 279.47 (277.00) | 1.98 (1.96) | 37.68 (35.74) |
| HQ:MM:MM | 238.50 (236.02) | 67.46 (−9.68) | 70.29 (28.53) | 106.68 (28.35) | 67.46 (−9.68) | 274.89 (235.83) | −2.83 (−38.21) | 39.22 (38.03) |
| HQ:MM:MM(EE) | 258.13 (242.20) | | 84.27 (38.44) | 109.22 (30.70) | | 283.09 (234.46) | | |
| LQ:LQ:LQ | 158.24 | 160.56 | 158.24 | 238.08 | 160.56 | 238.08 | 2.32 | 77.52 |
| LQ:LQ:MM | 158.69 (158.97) | 160.83 (161.21) | 158.69 (158.97) | 108.44 (27.02) | 66.72 (−12.34) | 202.55 (200.57) | 2.14 (2.24) | 41.72 (39.36) |
| LQ:MM:MM | 158.99 (158.76) | 68.03 (−10.33) | 69.33 (26.22) | 110.17 (30.85) | 68.03 (−10.33) | 199.82 (163.40) | −1.30 (−36.55) | 42.13 (41.19) |

[a] See footnote *a* of Table 1.

(ME) has an error in deprotonation energy of −40 kcal/mol, followed by HQ:LQ:MM of −43 kcal/mol and then HQ:MM:MM of −47 kcal/mol. The HQ:HQ:MM contains a large HQ region and is expensive. The three-layered HQ:LQ:MM method, which is inexpensive because the mid layer is calculated by the inexpensive LQ method, lost only 2.3 kcal/mol over the more expensive HQ:HQ:MM. The standard ONIOM(QM:MM) or generic QM/MM method corresponds to QM:MM:MM, which shows a larger error by 6.9 kcal/mol. The electronic embedding in QM:MM:MM reduces the error from −47 kcal/mol in ME to −38 kcal/mol but is expensive because it has to iterate QM calculations to

converge the polarized charges. Thus one can conclude clearly for this example that HQ:LQ:MM is an excellent approximation to the impractical HQ:HQ:MM method and is the method of choice which improves over QM:MM:MM or the standard QM/MM with very little additional cost of semiempirical calculation for the middle system. Again the combinations that use AM1 for the highest level, LQ:LQ:MM and LQ:MM:MM, are in error over 120 kcal/mol and are not worth considering.

One notes that the results of QM:MM:MM depend sensitively on the choice of the charges used in the Amber calculation. The use of the Mulliken charges, for instance,

**Table 3.** One-, Two-, and Three-Layered ONIOM Calculations for the Deprotonation Energy (in kcal/mol) of $NH_3^+-C^nBuH-CO-NH-CH_2-Co-NH-CH^nBu-COO^-$ System with Partition A

| combination[a] | E(high/ model) | E(med/ mid) | E(med/ model) | E(low/ real) | E(low/ mid) | E(ONIOM) | S(med/ mid-model) | S(low/ real-mid) |
|---|---|---|---|---|---|---|---|---|
| HQ:HQ:HQ | 241.70 | 318.06 | 241.70 | 322.09 | 318.06 | 322.09 | 76.36 | 4.04 |
| HQ:HQ:LQ | 245.04 | 318.67 | 245.04 | 248.06 | 242.11 | 324.62 | 73.63 | 5.95 |
| HQ:HQ:MM | 243.47 | 321.57 | 243.47 | 108.56 | 108.84 | 321.29 | 78.10 | −0.28 |
| HQ:HQ:MM (EE) | | 332.66 | | 108.53 | 118.89 | 322.30 | | −10.36 |
| HQ:LQ:LQ | 243.61 | 236.53 | 172.43 | 244.03 | 236.53 | 315.21 | 64.10 | 7.50 |
| HQ:LQ:MM | 237.81 | 235.91 | 166.04 | 110.17 | 110.43 | 307.41 | 69.86 | −0.27 |
| HQ:MM:MM | 238.50 | 106.72 | 70.29 | 106.68 | 106.72 | 274.89 | 36.43 | −0.04 |
| HQ:MM:MM(EE) | 258.13 | | 84.27 | 109.22 | | 283.09 | | |
| LQ:LQ:LQ | 158.24 | 232.51 | 158.24 | 238.08 | 232.51 | 238.08 | 74.27 | 5.57 |
| LQ:LQ:MM | 156.63 | 234.04 | 156.63 | 122.22 | 122.67 | 233.59 | 77.41 | −0.46 |
| LQ:MM:MM | 158.99 | 110.12 | 69.33 | 110.17 | 110.12 | 199.82 | 40.79 | 0.05 |

[a] See footnote *a* of Table 1.

in QM:MM:MM increases the error from −47 kcal/mol with RESP charges to −86 kcal/mol. This implies that the results will also depend sensitively on how to arbitrarily "scale" the near-border charges for the QM-MM interaction, because the problematic QM-MM boundary is very close to the reaction center. This is also reflected in the comparison of ME and EE schemes as well as RESP and Mulliken charges in the QM:QM:MM method. Here the differences among these choices are small because the problematic QM-MM boundary is far away from the action region, and the choice of EE vs ME or the choice of charges has small effects on the calculated deprotonation energies.

The performance of different methods can be evaluated more systematically by examining the *S*-values for the deprotonation energy. Looking at the *S*-value between the middle and model systems, *S*(med/mid-model) in Table 2, one sees that the target *S*-value is 3.8−4.4 kcal/mol for the HQ level, derived from HQ:HQ:HQ, HQ:HQ:LQ, and HQ:HQ:MM calculations. The *S*-value for LQ is in the range of 2.0−2.3 kcal/mol from a variety of combinations involving LQ for the middle and model system. This implies that the LQ in this middle system introduces an error of 2−4 kcal/mol in the deprotonation energy, suggesting that AM1 in the middle system is a good choice of the method. The *S*-value for MM is −2.8 to −1.3 kcal/mol with the RESP charges and −36 to −38 kcal/mol with the Mulliken charges. MM with RESP is not bad at least in this region but is very sensitive to the choice of charges.

Now we switch our attention to the *S*-value between the real and middle systems, *S*(low/real-mid). The target *S*-value for the deprotonation energy is 77.6 kcal/mol for the HQ level from the HQ:HQ:HQ calculation. The *S*-value for LQ is in the range of 69−78 kcal/mol from HQ:HQ:LQ to LQ:LQ:LQ. LQ is not bad for the real system but is too costly. The *S*-value for MM is 38−42 kcal mol/mol with RESP and 35−41 kcal/mol with Mulliken. Here we find a large source of error in deprotonation energy by using the Amber method for the real system. The present results clearly show that MM, even used as the low-level method in the outermost region of a large system, can introduce a substantial error in the energetics.

**3. Deprotonation Energy in Partition A.** Partition scheme A includes all the polar groups in the model and

middle systems. From RESP or Mulliken analysis, it appears that there is a charge transfer of about half an electron from the $CO_2$ to the $NH_3$ group. It is therefore likely that the small model in partition scheme B, with an integer charge, does not describe the deprotonation accurately, while AM1 is not able to recover the charge transfer quantitatively. Therefore partition A would be a more desired choice of the partition for a polar system. In this partition, all the combinations including large HQ calculations up to the middle system, HQ:HQ:LQ, HQ:HQ:MM, and HQ:HQ:MM(EE), reproduce the target results with an error of 2−3 kcal/mol, which is an excellent result. However, the large HQ calculation is relatively expensive. On the other hand, the combination with small HQ calculation, HQ:MM:MM-ME, gives a large error of 47 kcal/mol. Electronic embedding, HQ:MM:MM-EE, improves the results by 8 kcal/mol but still has a large error of 39 kcal/mol. The method with LQ as the highest method is again totally wrong, with an error of 90 kcal/mol or so. The three-layer ONIOM method, HQ:LQ:MM-ME, performs well, with the error of 15 kcal/mol. Considering the estimated cost of the method, this is a remarkable achievement. As discussed above, the three-layer HQ:LQ:MM-ME combination is substantially less expensive than the HQ:MM:MM-EE method and slightly more expensive than the HQ:MM:MM-ME method, but the error of the three-layer method is substantially smaller than either of the HQ:MM:MM-EE or -ME methods.

**4. Deprotonation Energy in Partition C.** We tested another partition scheme C, in which the small model consists of only $H_3N^+L$, and the middle system is equal to the model system in partition A. As discussed before, this is the minimal model and minimal middle system trying to push the ONIOM method to the extreme. Previously such minimal models worked reasonably well for C−H and C−C bond energies of organic molecules such as RR′R″C−H RR′R′C−CRR′R″,[20,21] and here we examine briefly whether such extreme models work for very polar complexes such as the present system.

The results are shown in Table 4. Comparing Tables 2 and 4, one can see clearly that the errors in deprotonation energy in partition C are significantly larger than in partition B. Only those combinations involving solely HQ and LQ (not MM) performed reasonably well. HQ:HQ:LQ here is

Performance of the Three-Layer ONIOM Method

*J. Chem. Theory Comput., Vol. 2, No. 5, 2006* **1323**

**Table 4.** One-, Two-, and Three-Layered ONIOM Calculations for the Deprotonation Energy (in kcal/mol) of $NH_3^+-C^nBuH-CO-NH-CH_2-CO-NH-CH^nBu-COO^-$ System with Partition C

| combination[a] | E(high/ model) | E(med/ mid) | E(med/ model) | E(low/ real) | E(low/ mid) | E(ONIOM) | S(med/ mid-model) | S(low/ real-mid) |
|---|---|---|---|---|---|---|---|---|
| HQ:HQ:HQ | 217.75 | 241.70 | 217.75 | 322.09 | 241.70 | 322.09 | 23.95 | 80.39 |
| HQ:HQ:LQ | 218.14 | 243.61 | 218.14 | 244.03 | 172.43 | 315.21 | 25.46 | 71.60 |
| HQ:HQ:MM | 217.62 | 236.02 | 217.62 | 28.35 | 28.53 | 235.83 | 18.40 | −0.18 |
| HQ:LQ:LQ | 218.16 | 158.57 | 159.13 | 238.30 | 158.57 | 297.32 | −0.56 | 79.73 |
| HQ:LQ:MM | 218.19 | 159.07 | 159.18 | 30.75 | 26.11 | 222.71 | −0.11 | 4.63 |
| HQ:MM:MM | 217.94 | 28.93 | 5.58 | 28.80 | 28.93 | 241.16 | 23.35 | −0.13 |
| LQ:LQ:LQ | 158.84 | 158.24 | 158.84 | 238.08 | 158.24 | 238.08 | −0.60 | 79.84 |
| LQ:LQ:MM | 158.95 | 158.76 | 158.95 | 30.85 | 26.22 | 163.40 | −0.18 | 4.64 |
| LQ:MM:MM | 158.89 | 28.97 | 5.64 | 28.84 | 28.97 | 182.10 | 23.33 | −0.13 |

[a] See footnote a of Table 1.

identical to HQ:LQ:LQ in partition B, and HQ:LQ:LQ here has an error of 22 kcal/mol, still noticeably better than any combinations involving MM. It appears that the MM method even at the outermost layer is not able to approximate the HQ method. The *S*-value between the real and middle systems, *S*(low/real-mid), is 80 kcal/mol in the target HQ method and 72−80 kcal/mol in the LQ method but is 0−5 kcal/mol in MM, introducing a huge error in the deprotonation energy. Obviously the minimal models are too small for the MM method to handle this very polar system.

## IV. Conclusions

We have systematically tested all possible three- and two-layer ONIOM combinations of high-level QM (HQ=B3LYP/ 6-31G*), low-level QM (LQ=AM1), and MM (Amber) for the deprotonation energy and structure of a test molecule, an ionic form of a peptide. Depending on the partition, we find the errors introduced in the ONIOM approximation, in comparison with the target HQ (or HQ:HQ:HQ) calculation, generally increase in the following order:

HQ:HQ:HQ (target) < HQ:HQ:LQ < HQ:LQ:LQ <
    HQ:HQ:MM < HQ:LQ:MM ≪ HQ:MM:MM ≪
        LQ:LQ:LQ < LQ:LQ:MM < LQ:MM:MM
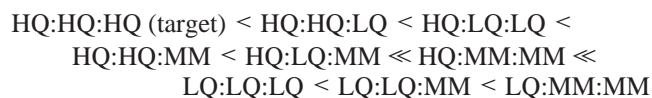
For realistic systems, the HQ calculation for the middle system and the LQ calculation for the real system can be expensive. The AM1 as the highest level (semiempirical QM/ MM) has too large an error to be useful, while also a QM-MM boundary close to the region of the action, in ONIOM-(QM:MM:MM), produces large errors.

We recommend the three-layer ONIOM(HQ:LQ:MM) method as the best tradeoff between accuracy and computational cost. It treats the innermost active center (small model) with a high-level quantum mechanical (HQ) method. The active center plus nearby environment (middle system) is handled with a low-level quantum mechanical (LQ) method, which provides a proper quantum mechanical description of the exchange as well as charge−charge interaction, can polarize the wave function of the active center, and allows charge-transfer between the active center and the environment. The real system is handled with a molecular mechanics (MM) method. The problematic boundary between the MM layer and the QM layer is sufficiently distant from the active center where the bond breaking and

forming takes place, and the intrinsically arbitrary choice of charge assignment and scaling does not affect the outcome of the calculations.

Despite the increasing computer power and further development of theoretical methods, it will not be to fully model very large molecular systems accurately with a single method. Therefore the use of the hybrid methods will remain essential. As we have shown in this case study, the ability of our ONIOM method to partition the system in more than two layers offers a valuable feature to the repertoire of hybrid methods in general.

**Supporting Information Available:** XYZ coordinates (in Å) for the optimized geometries. This material is available free of charge via the Internet at http://pubs.acs.org.

### References

(1) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.

(2) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718−730.

(3) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.

(4) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357−19363.

(5) Froese, R. D. J.; Morokuma, K. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Kollman, P. A., Clark, T., Schaefer, H. F., III, Gasteiger, J., Schreiner, P. R., Eds.; Wiley: Chichester, 1998; Vol. 2, pp 1244−1257.

(6) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct. (THEOCHEM)* **1999**, *461−462*, 1−21.

(7) Morokuma, K.; Musaev, D. G.; Vreven, T.; Basch, H.; Torrent, M.; Khoroshun, D. V. *IBM J. Res. Dev.* **2001**, *45*, 367−395.

(8) Vreven, T.; Morokuma, K. *Annual Reports in Computational Chemistry*; in press.

(9) Vreven, T.; Frisch, M. J.; Kudin, K. N.; Schlegel, H. B.; Morokuma, K. *Mol. Phys.* **2006**, *104*, 701−714.

(10) Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, J. A., Jr.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 815−826.

(11) Vreven, T.; Morokuma, K. *J. Comput. Chem.* **2000**, *21*, 1419−1432.

(12) Vreven, T.; Morokuma, K. *Theor. Chem. Acc.* **2003**, *109*, 125−132.

(13) Morokuma, K. *Philos. Trans. R. Soc. London, Ser. A* **2002**, *360*, 1149−1164.

(14) Morokuma, K. *Bull. Kor. Chem. Soc.* **2003**, *24*, 797−801.

(15) Svensson, M.; Humbel, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 3654−3661.

(16) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580−10594.

(17) 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.; 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 ed.*; Gaussian, Inc.: Wallingford, CT, 2004.

(18) Julian, R. R.; Jarrold, M. F. *J. Phys. Chem. A* **2004**, *108*, 10861−10864.

(19) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(20) Vreven, T.; Morokuma, K. *J. Chem. Phys.* **1999**, *111*, 8799−8803.

(21) Froese, R. D. J.; Morokuma, K. *J. Phys. Chem. A* **1999**, *103*, 4580−4586.