THE JOURNAL OF CHEMICAL PHYSICS **129**, 154106 (2008)

# A pseudobond parametrization for improved electrostatics in quantum mechanical/molecular mechanical simulations of enzymes

Jerry M. Parks, Hao Hu, Aron J. Cohen, and Weitao Yang[a)]
*Department of Chemistry, Duke University, 124 Science Drive, 5301 French Science Center, Durham, North Carolina 27708-0346, USA*

(Received 27 May 2008; accepted 11 September 2008; published online 17 October 2008)

The pseudobond method is used in quantum mechanical/molecular mechanical (QM/MM) simulations in which a covalent bond connects the quantum mechanical and classical subsystems. In this method, the molecular mechanical boundary atom is replaced by a special quantum mechanical atom with one free valence that forms a bond with the rest of the quantum mechanical subsystem. This boundary atom is modified through the use of a parametrized effective core potential and basis set. The pseudobond is designed to reproduce the properties of the covalent bond that it has replaced, while invoking as small a perturbation as possible on the system. Following the work of Zhang [J. Chem. Phys. **122**, 024114 (2005)], we have developed new pseudobond parameters for use in the simulation of enzymatic systems. Our parameters yield improved electrostatics and deprotonation energies, while at the same time maintaining accurate geometries. We provide parameters for $C_{ps}(sp^3)-C(sp^3)$, $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$, and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds, which allow the interface between the quantum mechanical and molecular mechanical subsystems to be constructed at either the $C_\alpha-C_\beta$ bond of a given amino acid residue or along the peptide backbone. In addition, we demonstrate the efficiency of our parametrization method by generating residue-specific pseudobond parameters for a single amino acid. Such an approach may enable higher accuracy than general purpose parameters for specific QM/MM applications. © *2008 American Institute of Physics*. [DOI: 10.1063/1.2994288]

## I. INTRODUCTION

One of the most challenging aspects of simulating complex biological systems such as enzymes is the proper description of bond breaking and forming. Although various levels of quantum mechanical (QM) methods provide accurate reaction energetics, their use is often limited by the available computational resources. In general, enzymatic systems are too large and computationally too intensive for *ab initio* or density functional theory (DFT) methods. Molecular mechanics (MM) methods, on the other hand, are efficient and computationally affordable for large systems, but are incapable of describing electronic degrees of freedom. As a result, macromolecular systems are often partitioned into a reactive QM subsystem and a classical MM subsystem that provide a reasonable environment in which reactions may take place. Combined QM/MM methods[1,2] thus enable the simulation of chemical reactions in large biological systems.

### A. Boundary problem in QM/MM simulations

The most pressing problem associated with QM/MM simulations then becomes an issue of properly interfacing the two subsystems together when they happen to be connected by one or more covalent bonds. Several methods have been introduced to resolve this issue including the hybrid orbital method, link atom approach, the generalized hybrid orbital method, the Frozen Orbital method, the local self-consistent field method, and the pseudobond method.[3–5] The practical application of many of these methods has been reviewed recently.[2] The pseudobond method, originally developed in our laboratory, provides a straightforward way to smoothly interface the QM and MM subsystems without introducing additional degrees of freedom into the system. This method has been applied successfully in numerous simulations of biomolecular systems (see for example, 6–10).

### B. Previous pseudobond formulation

In the pseudobond method, a single free valence atom with a specially parametrized basis set and effective core potential is used to satisfy the valency of the remainder of the QM subsystem. Thus, the pseudobond is designed to mimic the bond it replaces. The introduction of such a bond should not significantly perturb other nearby bonds, angles, or charges, nor should it adversely affect any other structural or electronic properties of interest.

Zhang[4] used a one-molecule training set consisting of ethane for the parametrization of $C_{ps}(sp^3)-C(sp^3)$ pseudobonds, and N-methylacetamide for the $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds. For each set of parameters, a geometry optimization was performed on the test set molecule in which a methyl group was replaced by the boundary carbon ($C_{ps}$). For the $C_{ps}(sp^3)-C(sp^3)$ parametrization, the criteria monitored in the fitting procedure included the C–C and C–H bond distances, C–C–H angle, Mulliken charges on C and H, and the

---

[a)] Electronic mail: weitao.yang@duke.edu.

bond dissociation energy (BDE) of the C–C bond. Analogous criteria were used for the $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds. The resulting preliminary parameters were then tested on sets of five to eight small molecules with various charges and peptide or nucleic acidlike functional groups. These tests included comparisons of selected bond distances, angles, Mulliken charges, and deprotonation energies between pair molecules in different charge states. The training set technically included these additional small molecules because the selection of parameters depended on their performance for all of the molecules. The optimizations of each parameter set were accomplished using a combined Monte Carlo/minimization approach.

### C. Scope of this work

Using the same form for the effective core potential (ECP) and basis set as Zhang,[4] we have developed new parameters for $C_{ps}(sp^3)-C(sp^3)$, $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$, and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds that are suitable for use in QM/MM simulations of enzyme systems. Our parameters yield improved electrostatic potential (ESP) charges relative to the previous parametrization for DFT and *ab initio* methods, while at the same time maintaining accurate geometries and deprotonation energies. The $C_{ps}(sp^3)-C(sp^3)$ parameters allow truncation at the $C_\alpha-C_\beta$ bond of any acceptable amino acid sidechain, whereas the $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ and $C_{ps}(sp^3)-N(sp^3)$ parameters allow QM and MM subsystems partitioning along the peptide backbone.

## II. METHODS

In this work, the parametrization efforts are focused on four criteria, namely, geometries, ESP fitted charges, deprotonation energies, and bond dissociation energies. ESP charges are commonly used to represent the QM subsystem during the evaluation of the electrostatic interactions between the two subsystems in QM/MM calculations. ESP charges more accurately reproduce the molecular ESP and as a result should yield better intermolecular interactions than Mulliken charges.

Note that the pseudobond method may be used with any standard quantum chemistry code capable of using effective core potentials, and no additional programming is required for its use. Also, the pseudobond method is completely independent of the MM force field. In fact, any standard force field may be used. Care must be taken to avoid the well known issues with spurious interactions between the QM subsystem and nearby MM point charges. Many approaches have been devised to deal with these issues and have been reviewed elsewhere,[2,11] so they will not be discussed further here.

All calculations were performed with Gaussian03 (Ref. 12) at the B3LYP/6-31G($d$) level of theory.[13,14] ESP charges were calculated using the Merz–Singh–Kollman scheme[15] with a total of ten layers of gridpoints and a gridpoint density of six. Note that no modification to the QM code is required because the ECP and basis set forms used in this work are standard in many QM packages. When the pseudobond method is used in a QM/MM simulation, the boundary atom ($C_{ps}$) is also covalently bonded to atoms in the MM subsystem. Bonds, angles, and torsions that include at least one atom in both the QM and MM subsystems are modeled using a MM force field.

### A. Pseudobond formulation

We have used the same STO-2G basis set and angular momentum-independent form for the effective core potentials as Zhang.[4] The basis set can be written as

$$\phi_s = g_s(\alpha_1,R) + d_1 g_s(\alpha_2,R)$$

$$\phi_p = g_p(\alpha_1,R) + d_2 g_p(\alpha_2,R),$$

where $g_s$ and $g_p$ are normalized $s$- and $p$-type Gaussians, $\alpha_1$ and $\alpha_2$ are the exponents, and $d_1$ and $d_2$ are the coefficients. The form for the effective core potential is

$$V^{\text{eff}}(r) = a\exp(-br^2)/r,$$

where $a$ is the coefficient and $b$ is the exponent.

In the previous pseudobond formulation, the parametrization criteria included Mulliken charges. Mulliken charges may have a more rigorous connection to the wave function or electron density than ESP charges, but tend to suffer from large basis set dependence effects and exaggerated charge separations. ESP charges, as the name implies, are inherently more accurate in reproducing the ESP associated with a given density. Because ESP charges are commonly used to represent the electrostatics of the QM subsystem as felt by the MM subsystem, it is desirable to develop pseudobonds that yield accurate ESP charges in QM/MM simulations.

### B. Training sets

The $C_{ps}(sp^3)-C(sp^3)$ training set used in this work consisted of 20 sidechain analogs. A sidechain analog is defined here as an amino acid sidechain in which the $C_\alpha$ is truncated as a methyl group. In the 20 molecule training set, Asp, Cys, Glu, Lys, Ser, Thr, and Tyr were included in both the neutral and ionized states, and His was modeled in the neutral $\delta$-protonated (Hsd) and $\epsilon$-protonated (Hse) states, as well as in the cationic state (Hsp) in which both $N_\delta$ and $N_\epsilon$ were protonated. Asn and Gln were modeled as neutral species, and Arg was only modeled in the cationic form.

For the parameter fitting, sidechain analogs incorporating pseudobonds were constructed by replacing the $C_\alpha$ methyl group of each standard sidechain analog with a boundary carbon $C_{ps}$. Figure 1 shows an example of the standard and $C_{ps}$-containing Hsp sidechain analog used in the parametrization and testing of the $C_{ps}(sp^3)-C(sp^3)$ pseudobond.

For the parametrizations of the $C_{ps}(sp^3)-C(sp^2$, carbonyl) and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds, two different training sets were considered. Initially, an 18-molecule subset of the 20 molecules used for the $C_{ps}(sp^3)-C(sp^3)$ pseudobonds was used, except the molecules were modeled as amino acid "dipeptides" rather than as sidechain analogs. Only 18 molecules were used because the deprotonated forms of the Cys and Ser dipeptides undergo spontaneous rearrangement reactions in the gas phase. As a second training set, we also used the much smaller five-molecule testing sets used by Zhang.[4]

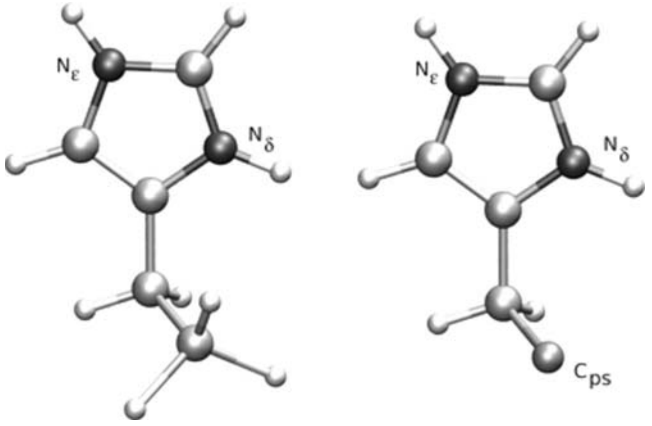

FIG. 1. Example of a sidechain analog (Hsp) used in the parametrization and testing of the $C_{ps}(sp^3)-C(sp^3)$ pseudobond.

It should be noted that the $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ and $C_{ps}(sp^3)-N(sp^3)$ parameter optimizations were performed independently of each other.

Instead of performing full geometry optimizations for each molecule in the training set every time a new set of parameters was generated during the parameter optimizations, we simply evaluated the gradient of each $C_{ps}$-containing molecule at the optimized geometry of the corresponding standard molecule from which it was generated. We also computed the ESP charges, deprotonation energies, and bond dissociation energy at these fixed geometries. This approach allowed more extensive and more chemically relevant training sets and partially circumvented problems associated with optimizations on rugged parameter landscapes. Deprotonation energies were computed for all ionizable residues. For His, both the Hsp→Hsd and Hsp→Hse deprotonation processes were considered.

## C. Test sets

For the $C_{ps}(sp^3)-C(sp^3)$ pseudobond, the test set was identical to the training set. During the testing, however, full geometry optimizations were performed and the rms errors were evaluated for all bond distances, angles, ESP charges (excluding $C_{ps}$), and deprotonation energies. For the $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds, the training sets consisted of the five-molecule test sets used by Zhang,[4] but testing was performed on the full 18-molecule dipeptide set described above. The tests were performed concurrently, i.e., the dipeptide systems were capped at both the N- and C-terminal methyl groups with the appropriate pseudobond. This is a realistic representation of how the pseudobonds would be used in a QM/MM simulation, although MM bonded and nonbonded terms would also be present in such a calculation. Again, the rms errors in all bond distances, angles, ESP charges (except $C_{ps}$ atoms), and deprotonation energies were evaluated.

## D. Objective function

The objective function used for the parameter fitting consisted of a weighted sum of four different terms,

$$E_{\text{total}} = \sum_{i}^{\text{prop}} \omega_i E_i(\lambda) \quad \lambda = (a,b,\alpha_1,\alpha_2,d_1,d_2), \tag{1}$$

where $E_i$ represents the error for one of the four properties of interest: the collective gradient norm, ESP charges, deprotonation energies, and bond dissociation energies. The collective gradient norm is the norm of the single vector consisting of all the individual geometric gradients of the pseudobond-containing molecules in the training set at their fixed, standard molecule geometries. The ESP charge error is the rms deviation between the ESP charges of all non-$C_{ps}$ atoms in the training set and the corresponding charges in the standard molecules. The bond dissociation energy error was computed as the rms deviation between the bond dissociation energies of the pseudobond-containing molecules and the corresponding standard molecules. The bond dissociation energy error of a given molecule is calculated by considering the following two equations:

$$\cdot CH_3 + \cdot R \rightarrow CH_3 - R \quad \Delta E_1$$

$$\cdot C_{ps} + \cdot R \rightarrow C_{ps} - R \qquad \Delta E_2,$$

where $\Delta E_1$ and $\Delta E_2$ correspond to the dissociation of the standard and pseudobond-containing molecules, respectively. However, $\cdot R$ appears in both equations, so it cancels out and the equations are simplified to

$$\cdot CH_3 \rightarrow CH_3 - R \quad \Delta E_1'$$

$$\cdot C_{ps} \rightarrow C_{ps} - R \qquad \Delta E_2'.$$

The error in the bond dissociation energy of a given molecule is then calculated as

$$\text{BDE} = \Delta E_1' - \Delta E_2'.$$

For the parameter optimizations, weights were chosen such that each property was treated approximately equally, although geometries and ESP charges were given greater emphasis. For all three parameter sets, a relative weight of 1.0 was assigned to the geometric gradient norm to ensure accurate molecular geometries. ESP charges were given a relative weight between 0.46 and 0.50. For deprotonation energies, relative weight factors of 0.018 to 0.020 were used in the parametrizations. Bond dissociation energies were weighted by a relatively small factor of between 0.0006 and 0.0008. Note that the units of each weight were the inverse of the corresponding property. It should also be mentioned that the results were fairly sensitive to the choice of the weights, and that some degree of manual adjustment was necessary to arrive at final parameter sets that provided balanced overall results for all molecular properties.

## E. Powell minimization

Because there is no analytic expression for the gradient of our objective function, we used the Powell method[16,17] to perform the parameter optimizations. The Powell method is a direction set method in which an approximate gradient is constructed, and bracketing and line searches are then used to locate minima. Iteration is performed over all variables

TABLE I. Optimized pseudobond parameters for QM/MM simulations of enzymes.

| | $C_{ps}(sp^3)-C(sp^3)$ | $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ | $C_{ps}(sp^3)-N(sp^3)$ |
|---|---|---|---|
| a | 16.49 | 23.1 | 18.4 |
| b | 7.75 | 10.8 | 9.5 |
| $\alpha_1$ | 0.9034 | 1.18 | 0.78 |
| $\alpha_2$ | 0.21310 | 0.29 | 0.17 |
| $d_1$ | 1.90904 | 4.99685 | 0.78895 |
| $d_2$ | 0.57864 | 0.79341 | 0.31400 |

until convergence is achieved. Additionally, a genetic algorithm[18] was used to search broad regions of the parameter space. However, we found that the Powell method was quite efficient for finding minima and actually gave superior results.

### F. Parametrization procedure

Initially, all molecules in each training set were fully optimized at the B3LYP/6-31G($d$) level of theory.[13,14] The training set molecules were constructed from the optimized standard molecules by replacing the appropriate methyl group by a fluorine atom ($C_{ps}$) with the effective core potential and STO-2G basis set. An initial guess for the parameters was chosen and the energy, geometric gradient, ESP charges, and bond dissociation energy were computed for each molecule, and the deprotonation energy was evaluated for all pairs of ionizable residues. The total weighted error was the computed according to the objective function and the process was iterated until convergence. Parameter optimizations were begun from a few different initial guesses to ensure that the best local minimum was found for each type of pseudobond.

## III. RESULTS AND DISCUSSION

The $C_{ps}(sp^3)-C(sp^3)$ pseudobond is designed for constructing the QM/MM interface between the $C_\alpha-C_\beta$ bond in peptides, whereas the $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds are designed for interfacing the QM and MM subsystems along the peptide backbone. Specifically, the $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ pseudobond connects the carbonyl carbon of a given QM residue to $C_\alpha$ of the adjacent MM residue. The $C_{ps}(sp^3)-N(sp^3)$ pseudobond connects the amide nitrogen atom of a given QM residue to $C_\alpha$ of the adjacent MM residue. As a result, the $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds are generally used together unless the QM residue is located at the C- or N-terminus of the peptide chain.

The optimized parameters for the $C_{ps}(sp^3)-C(sp^3)$, $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$, and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds are shown in Table I. It should be noted that our training set consisted of the sidechain analogs of residues that are likely to be designated as part of the QM subsystem in actual QM/MM simulations. Because these sidechain analogs were used in the training set, our test set was identical to the training set.

### A. $C_{ps}(sp^3)-C(sp^3)$ pseudobond performance

The performance of the $C_{ps}(sp^3)-C(sp^3)$ pseudobonds was tested on the 20-molecule set of sidechain analogs described in Sec. II, and the results are summarized in Table II. The rms error for all bonds in the test set was 0.009 Å. The maximum positive and negative deviations for all bonds were 0.068 ($C_\alpha-C_\beta$ of deprotonated Ser) and −0.032 Å ($C_\beta-O_\gamma$ of deprotonated Thr), respectively. For the angles, the rms error was 1.2° with a maximum positive and negative deviations of 4.189 ($C_\epsilon-N_\delta-H_\delta$ of cationic His) and −4.277° ($H_\delta-N_\delta-C_\gamma$ of cationic His), respectively. These somewhat large deviations are a result of the interaction of $H_\delta$ with $C_\alpha$, i.e., $C_{ps}$. However, in actual QM/MM simulations this interaction would be hindered because of adjacent MM van der Waals and bonded terms involving $C_\alpha$.

Significant improvements were made in the ESP charges relative to the previous parametrization. For the current $C_{ps}(sp^3)-C(sp^3)$ parametrization, the rms error in the ESP charges was 0.10. The maximum positive and negative deviations were 0.31 ($C_\gamma$ of Tyr) and −0.35 ($C_\beta$ of Glu). For the previous parametrization, the rms error in the ESP charges was 0.43 with maximum positive and negative deviations of 0.86 ($C_\gamma$ of Tyr) and −1.50 ($C_\beta$ of Tyr), respectively. These errors are quite severe and could produce spurious results in actual QM/MM simulations. As a clear way to visualize the improvements in the ESP charges, we have plotted the correlation between all non-$C_{ps}$ charges for the standard sidechain analogs and the corresponding molecules that include a $C_{ps}(sp^3)-C(sp^3)$ pseudobond in Fig. 2.

The rms error in the deprotonation energies for the 20 sidechain analogs with the $C_{ps}(sp^3)-C(sp^3)$ pseudobonds was 1.40 kcal mol$^{-1}$. The molecule with the largest positive error was Thr with a deprotonation energy that was overestimated by 1.93 kcal mol$^{-1}$ relative to the standard QM reference. Tyr had the largest negative error with a deprotonation energy that was underestimated by −2.52 kcal mol$^{-1}$.

TABLE II. RMS errors computed using new and previous pseudobond parameters. Bond errors are in Å, angles are in degrees, ESP charges are in units of $e_c$, and deprotonation energies are in kcal mol$^{-1}$. Values using parameters from Ref. 4 are shown in parentheses.

| | $C_{ps}(sp^3)-C(sp^3)$ | $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ and $C_{ps}(sp^3)-N(sp^3)$[a] |
|---|---|---|
| Bonds | 0.009 (0.006) | 0.026 (0.033) |
| Angles | 1.23 (0.91) | 0.67 (0.87) |
| ESP charges | 0.10 (0.44) | 0.11 (0.31) |
| Dep. energy | 1.4 (2.3) | 1.16 (1.19) |

[a] The $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ and $C_{ps}(sp^3)-N(sp^3)$ parameters were tested concurrently on the set of 18 dipeptide systems described in Sec. II.

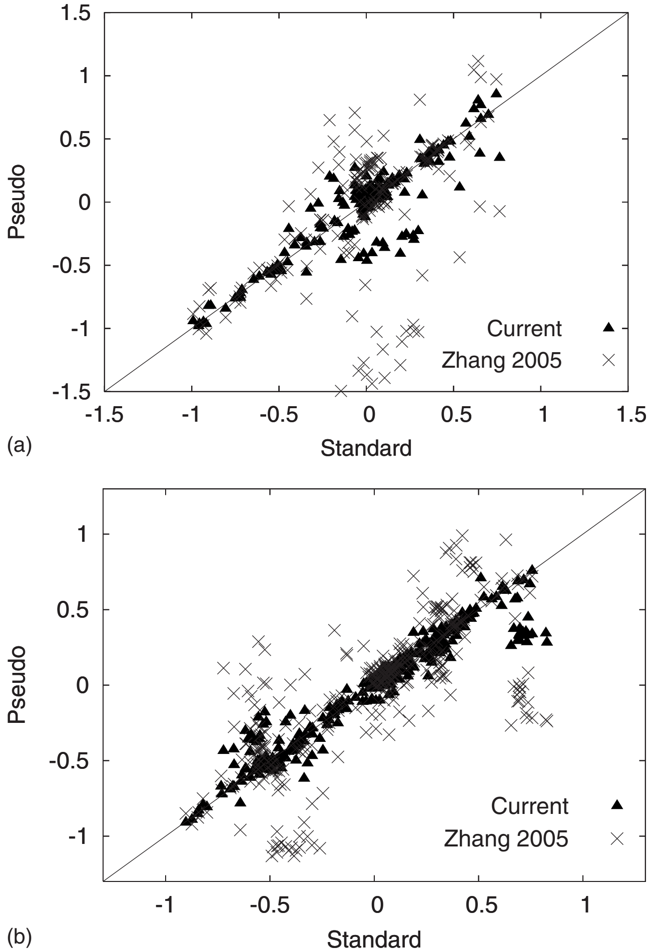

FIG. 2. Correlation of ESP charges in the standard molecule with molecules containing (a) $C_{ps}(sp^3)-C(sp^3)$ pseudobond and (b) $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds. All charges except those on $C_{ps}$ atoms are included.

As an additional test of our $C_{ps}(sp^3)-C(sp^3)$ pseudobond parameters, we computed CHELPG charges[19] for the set of 20 pseudobond-containing sidechain analogs and compared them with the corresponding standard molecules. The rms error over all non-$C_{ps}$ atoms was 0.10 (data not shown), demonstrating the transferability of the pseudobond parameters among the different ESP charge schemes.

### B. Pseudobond tests with various QM methods

In addition to the B3LYP method, we also tested the $C_{ps}(sp^3)-C(sp^3)$ pseudobond parameters with the HF, MP2, and BLYP methods. The results are shown in Table III. It is evident from the table that the parameters are transferable among various *ab initio* and DFT methods. The rms errors in the bonds and ESP charges across all methods are quite small and similar to the B3LYP results. For the MP2 results, the dominant errors arise from the interaction of $H_\delta$ with $C_{ps}$ in the protonated His sidechain analog. However, this interaction is somewhat artificial because it would be largely eliminated by MM terms in a QM/MM simulations as discussed previously. The relatively large errors in the Hartree–Fock deprotonation energies appear to be from systematic errors at this level of theory because all of the deprotonation energies except one are underestimated.

TABLE III. Performance of $C_{ps}(sp^3)-C(sp^3)$ pseudobonds with various QM methods. Bond errors are in Å, angles are in degrees, ESP charges are in units of $e_c$, and deprotonation energies are in kcal $mol^{-1}$.

| Method | Bonds | Angles | ESP charges | Dep. energy |
|---|---|---|---|---|
| B3LYP | 0.009 | 1.23 | 0.10 | 1.4 |
| BLYP | 0.011 | 1.32 | 0.10 | 1.2 |
| HF | 0.008 | 0.98 | 0.09 | 2.7 |
| MP2 | 0.007 | 1.59 | 0.09 | 1.5 |

### C. Dipeptide backbone test

As mentioned previously, the parametrizations for the $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ and $C_{ps}(sp^3)-N(sp^3)$ pseudobonds were performed independent of each other. However, these sets are designed to be used concurrently when the QM subsytem includes one or more sidechains and their associated backbone atoms. The exception is when the residue(s) to be included in the QM subsystem is at the N- or C-terminus. In this case, the $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ and $C_{ps}(sp^3)-N(sp^3)$ parameters may be used individually.

We tested the $C_{ps}(sp^3)-C(sp^2,\text{carbonyl})$ and $C_{ps}(sp^3)-N(sp^3)$ parameters concurrently on the 18-molecule test set of amino acid dipeptides. For these tests, both of the terminal methyl groups were replaced by pseudobonds, which is analogous to a QM/MM simulation with subsystem partitioning along the peptide backbone. The results of the dipeptide tests are shown in Table II. The rms error for all bonds was 0.026 Å with maximum positive and negative deviations of 0.046 and −0.11 Å, respectively. For the angles, the rms error was 0.67 Å with maximum positive and negative deviations of 2.39 and −2.88 degrees.

Again, the ESP charges showed the most significant improvement relative to the previous parametrization with an rms charge error over all non-$C_{ps}$ atoms of 0.11. The maximum positive and negative deviations compared to the corresponding charges in the standard molecules were 0.35 and −0.55, respectively. In the previous parametrization the rms error in the ESP charges was 0.31 with maximum positive and negative deviations of 0.84 and −1.07. For the deprotonation energies, the current parametrization yielded similar but slightly better results than in the previous parametrization with an rms error of 1.16 versus 1.19 kcal $mol^{-1}$.

### D. Residue-specific parametrization

Typically, general sets of pseudobond parameters designed to be transferable to many different systems are used in QM/MM simulations. An alternative approach is to perform parametrizations for a specific molecular systems of interest. For example, a specific set of pseudobond parameters can be obtained for a single amino acid residue that acts as a catalytic acid or base in an enzyme active site. The parametrization would include only two molecules in the training set, i.e., the protonated and deprotonated forms of the residue of interest. Greater accuracy is achievable relative to general purpose parameters because the error values are not averaged over a significantly larger set of molecules.

TABLE IV. Residue-specific $C_{ps}(sp^3)-C(sp^3)$ parameters for the Glu sidechain. The rms errors were 0.008 Å for bonds, 1.15 for angles, 0.088 for charges, and 0.18 kcal mol$^{-1}$ for deprotonation energies.

| | $C_{ps}(sp^3)-C(sp^3)$ |
|---|---|
| a | 18.8 |
| b | 8.0 |
| $\alpha_1$ | 0.89 |
| $\alpha_2$ | 0.21 |
| $d_1$ | 4.299 50 |
| $d_2$ | 0.474 99 |

To demonstrate the performance of residue-specific parameters, we have used our optimization method to generate pseudobonds for the Glu sidechain. As done previously, the objective function included four terms: the gradient norm, the rms errors in the ESP charges, the deprotonation energy, and the bond dissociation energies. However, this time the training set only consisted of Glu and protonated Glu. As a result, the optimization proceeded quickly. The weights were adjusted manually by hand in order to provide balanced error values for each of the tested properties. The parameters and error data for the residue-specific parametrization for the Glu sidechain are shown in Table IV.

### E. Other approaches

There are numerous viable options for describing covalent boundaries in QM/MM simulations. One approach similar in spirit to the pseudobond method is the quantum capping potential method of DiLabio *et al.*[20] The advantage of the quantum capping potential method is that it only requires a one-electron cap rather than the seven electrons of the boundary atom in the pseudobond method. In addition to requiring Pauli repulsion and spherical screening corrections, which complicate its implementation, the QCP method may also suffer from inaccurate charges and/or geometries. In the pseudobond method, fluorine is much more similar to carbon (and nitrogen) in terms of its electronic structure, which likely permits better reproduction of the desired bonding. Thus, although the pseudobond marginally increases the number of electrons and basis functions compared to the QCP method, its implementation is quite simple and its use appears justified based on our results.

## IV. CONCLUSIONS

The proper description of the QM ESP is of critical importance for obtaining reliable and accurate results in QM/MM simulations of enzyme systems. In this work, we have performed parametrizations for the QM/MM pseudobonds method that enables the interface between the QM and MM subsystems to be made at either the $C_\alpha-C_\beta$ bond of amino acid sidechains or at the $C_\alpha-C_{\mathrm{carbonyl}}$ and/or $C_\alpha-N_{\mathrm{amide}}$ bonds of the peptide backbone. The newly developed parameters yield accurate geometries and deprotonation energies that are comparable to the previous parametrization. At the same time, the ESP charges obtained with our parametrization offer a significant improvement compared to previous parametrizations. As a result, these parameters should enable QM/MM simulations of enzyme systems to achieve greater accuracy and reliability.

## ACKNOWLEDGMENTS

The authors thank the National Institutes of Health (NIH) (Grant No. R01-GM-061870) for financial support.

[1] A. Warshel and M.Levitt, J. Mol. Biol. **103**, 227 (1976).
[2] H. M. Senn and W. Thiel, Top. Curr. Chem. **268**, 173 (2007).
[3] Y. Zhang, T.-S. Lee, and W. Yang, J. Chem. Phys. **110**, 46 (1999).
[4] Y. Zhang, J. Chem. Phys. **122**, 024114 (2005).
[5] Y. Zhang, J. Chem. Phys. **116**, 43 (2006).
[6] H. Liu, Y. Zhang, and W. T. Yang, J. Am. Chem. Soc. **122**, 6560 (2000).
[7] G. A. Cisneros, H. Liu, Y. Zhang, and W. Yang, J. Am. Chem. Soc. **125**, 10384 (2003).
[8] T. H. Rod and U. Ryde, J. Chem. Theory Comput. **1**, 1240 (2005).
[9] J. Kästner, H. M. Senn, S. Thiel, N. Otte, and W. Thiel, J. Chem. Theory Comput. **2**, 452 (2006).
[10] H. Hu, Z. Lu, J. M. Parks, S. K. Burger, and W. Yang, J. Chem. Phys. **128**, 034105 (2008).
[11] H. Lin and D. G. Truhlar, Theor. Chem. Acc. **117**, 185 (2007).
[12] M. J. Frisch *et al.*, GAUSSIAN 03, Revision C.02, Gaussian, Inc., Wallingford, CT, 2004.
[13] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[14] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[15] U. C. Singh and P. A. Kollman, J. Comput. Chem. **5**, 129 (1984).
[16] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 77: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[17] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 90: The Art of Parallel Scientific Computing* (Cambridge University Press, Cambridge, 1996).
[18] G. Yang, L. E. Reinstein, S. Pai, Z. Xu, and D. L. Carroll, Med. Phys. **25**, 2308 (1998).
[19] C. M. Breneman and K. B. Wiberg, J. Comput. Chem. **11**, 361 (1990).
[20] G. A. DiLabio, M. M. Hurley, and P. A. Christiansen, J. Chem. Phys. **116**, 9578 (2002).