A Critical Evaluation of Different QM/MM Frontier Treatments with SCC-DFTB as the QM Method

P. H. König,[§] M. Hoffmann,[§] and Th. Frauenheim*

Theoretische Physik, Universität Paderborn, Warburger Strasse 100, 33098 Paderborn, Germany

Q. Cui*

Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, 1101 University Avenue, Madison, Wisconsin 53706

Received: December 18, 2004; In Final Form: February 18, 2005

The performance of different link atom based frontier treatments in QM/MM simulations was evaluated critically with SCC-DFTB as the QM method. In addition to the analysis of gas-phase molecules as in previous studies, an important element of the present work is that chemical reactions in realistic enzyme systems were also examined. The schemes tested include all options available in the program CHARMM for SCC-DFTB/MM simulation, which treat electrostatic interactions due to the MM atoms close to the QM/MM boundary in different ways. In addition, a new approach, the divided frontier charge (DIV), has been implemented in which the partial charge associated with the frontier MM atom ("link host") is evenly distributed to the other MM atoms in the same group. The performance of these schemes was evaluated based on properties including proton affinities, deprotonation energies, dipole moments, and energetics of proton transfer reactions. Similar to previous work, it was found that calculated proton affinities and deprotonation energies of alcohols, carbonic acids, amino acids, and model DNA bases are very sensitive to the link atom scheme; the commonly used single link atom approach often gives error on the order of 15 to 20 kcal/mol. Other schemes give better and, on average, mutually comparable results. For proton transfer reactions, encouragingly, both activation barriers and reaction energies are fairly insensitive (within a typical range of 2−4 kcal/mol) to the link atom scheme due to error cancellation, and this was observed for both gas-phase and enzyme systems. Therefore, the effect of using different link atom schemes in QM/MM simulations is rather small for chemical reactions that conserve the total charge. Although the current study used an approximate DFT method as the QM level, the observed trends are expected to be applicable to QM/MM methods with use of other QM approaches. This observation does not mean to encourage QM/MM simulations without careful benchmark in the study of specific systems, rather it emphasizes that other technical details, such as the treatment of long-range electrostatics, tend to play a more important role and need to be handled carefully.

1. Introduction

Quantum mechanical (QM) approaches can describe the breaking and formation of chemical bonds and hence are required for studying chemical reactivity. Many reactions of interest involve crucial contributions from the environment, which can be solvent or a macromolecule. Various approaches for including the effect of the environment on chemical reactivities have been suggested and implemented. For example, linear scaling techniques[1] are available in different theoretical frameworks and have been successfully applied to realistic biomolecular systems.[2] The potential of these full QM methods is enormous considering the speed gain of computational resources in recent years. Nevertheless, with commonly available facilities, it remains difficult to perform sufficient amount of statistical sampling for investigating chemical reactions in complex systems with QM methods alone even at the semiempirical[3,4] or approximate density functional theory (DFT) levels.[5]

An intuitive yet powerful alternative is to partition the system in a multiscale approach treating different parts of the system with different levels of theory. A popular choice is the hybrid quantum mechanical/molecular mechanical (QM/MM) approach,[6−8] in which the reactive part is described with QM and the remaining system by MM. The underlying assumption is that the changes in electronic structure are localized and that the response of the environment can be described with sufficient accuracy with use of a force field. Since the introduction of the QM/MM idea around the early 1980's,[9] numerous studies have been conducted examining both performance of different schemes as well as interesting applications to various solution and enzyme systems.[10−15] These previous studies have clearly demonstrated that careful QM/MM methods can provide useful insights into catalytic mechanisms in complex systems that are difficult to obtain otherwise. The current challenges lie in systematically improving the robustness of such methods for general chemical and biochemical systems, especially those involving open-shell species (e.g., radicals,[16,17] excited states,[18,19] and transition metals[20,21]) and complex structure/dynamics (e.g., molecular motors,[22−24] DNA-manipulating enzymes,[25−27] and membrane bound systems[28−30]).

In commonly used QM/MM schemes the Hamiltonian operator of the entire system, $\hat{H}$, is written as the sum of those for the QM partition, $\hat{H}_{QM}$, the MM partition, $\hat{H}_{MM}$ and the

[§] These authors contributed equally to this work.

10.1021/jp0442347 CCC: $30.25
Published on Web 04/08/2005

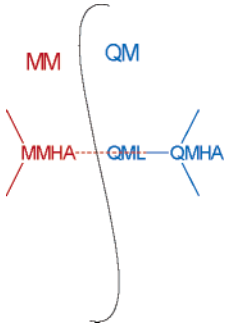

Figure 1. QM/MM partitioning across a covalent bond showing the MM host atom (MMHA), QM host atom (QMHA), and QM link atom (QML).

interaction between the two, $\hat{H}_{QM/MM}$,

$$\hat{H} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM} \quad (1)$$

The precise expression of $\hat{H}_{QM/MM}$ varies but generally has contributions from electrostatic, van der Waals, bonded interactions, and possibly additional constraints,[6]

$$\hat{H}_{QM/MM} = \hat{H}^{elec}_{QM/MM} + \hat{H}^{vdW}_{QM/MM} + \hat{H}^{bonded}_{QM/MM} + \hat{H}^{cons}_{QM/MM} \quad (2)$$

The bonded terms and constraints (e.g., fixed bond distance between boundary QM and MM atoms) are used for keeping the proper connectivities and geometries when cutting through covalent bonds at the QM/MM interface. The QM/MM van der Waals terms can be optimized to improve properties such as distribution of MM groups around the QM group.[31] The electrostatic component, which is missing in some early implementations for organometallic systems,[32] is crucial for polar environments such as water solution and biomolecules (for a recent example, see ref 33).

An issue that has been repeatedly raised concerns the treatment of the QM/MM boundary. The interaction between MM atoms and nearby QM atoms should be carefully treated to reliably describe the effect of the environment on chemical properties of the QM region. This is expected to be particularly important in cases where the QM/MM partition involves dividing the system across covalent bonds; a typical example involves partitioning catalytic side chains as QM and the remaining protein as MM. It is evident that special care has to be taken to avoid dangling bonds in the frontier QM atom(s) and to minimize artifactual effects on the electronic structure of the QM region.

Various approaches have been proposed to deal with capping of the QM frontier atoms. The most straightforward approach involves inserting a link atom, which is typically chosen as a hydrogen atom, between the QM host atom (QMHA, see Figure 1 and Table 1) and the MM host atom (MMHA). The link atom is treated at the QM level, and may be subject to an angular and distance constraint to lie along the bond between QMHA and MMHA at a fixed bond distance. The link atom typically interacts with MM atoms through electrostatic terms (however, see below) but not through van der Waals terms.

Instead of regular hydrogen atoms, hydrogen-like atoms or pseudohalogens have been used to terminate the QM region. In those approaches, the link atom coincides with the MMHA. The electronic nature of this atom is modified to mimic the behavior of the MM host atom or MM host group (MMHG). For example, in an approach described by Zhang et al.[34] for ab initio or DFT QM/MM, the boundary atom is described as a pseudohalogen with seven electrons, whose properties are adjusted to that of the atom it replaces by using pseudopotentials. Antes et al.[35] suggested an alternative for semiempirical QM/MM methods in which the integrals of the hydrogen-like boundary atom are re-parametrized to mimic the behavior of the host (e.g., methyl) group.

TABLE 1: Notations for Atoms and Groups Used in QM/MM Frontier Treatments

MM atom	MMA	generic atom in the molecular mechanics (MM) partition
QM atom	QMA	generic atom in the quantum mechanics (QM) partition
MM host atom	MMHA	atom on the MM side of a severed bond at the frontier (Figure 1)
MM host group	MMHG	charge neutral group including the MMHA
QM host atom	QMHA	atom on the QM side of a severed bond at the frontier (Figure 1)
QM link atom	QML	link atom introduced between the QMHA and MMHA to cap the QM region, treated with QM (Figure 1)
MM link atom	MML	link atom introduced between the MMHA and QMHA to balance dipole moment in the DLA approach, treated with MM (Figure 1)

Other approaches to the QM capping problem include the localized self-consistent field (LSCF)[36−39] and the generalized hybrid orbital method (GHO),[40] which use localized orbitals to describe the bonding at the QM/MM interface. Several studies with the LSCF approach in combination with DFT or HF[41−43] clearly demonstrated the capability and the pitfalls of the LSCF approach using small gas-phase models[41,42] as well as protein models.[42,43] Moreover, Reuter et al.[44] showed in their systematic analysis using gas-phase systems that the link atom approach and the LSCF method for AM1/CHARMM give comparable results. For small QM zones and MM atoms with large charges at the interface, the LSCF method gives worse results than the link atom approach; however, relatively large errors with both methods suggested that the QM/MM frontier should be chosen such that classical frontier atoms have small charges.

Within the link atom framework, it was argued that the treatment of QM/MM electrostatic interactions involving frontier MM atoms can influence the result substantially. Early implementation of the link atom approach sought to resolve the issue of spurious electrostatic interactions at the frontier by not taking into account interactions between the QM link atom and MM charges.[6,45] Multiple authors showed[44,46,47] that this treatment suffers from erroneous polarization and other artifacts, due to distortions in the QM charge distribution as the frontier QM atom and the QM link atom experience significantly different electric fields. These effects are evident in both the electronic structure (e.g., dipole moment) and the energetics of protonation/deprotonation of the QM region.

A popular scheme, which is the default in CHARMM, excludes electrostatic interactions between the MMHA and all QM atoms (including the link atom).[46,48−50] This has been termed as the "single link atom (SLA)" treatment in a recent overview by Das et al.[51] Noting that the SLA treatment leaves a fractional charge near the QM region, which can cause serious errors.[51] An alternative, which is termed the EXGR scheme in CHARMM and in this paper, is to exclude QM/MM electrostatic interactions involving the entire MM host group (MMHG, see Figure 2).[52] Sherwood et al. suggested yet another variation, referred to as the charge shift scheme (CHSH),[53,54] in which the charge of the MMHA is distributed to the neighboring atoms in the same group. Dipoles are placed on each of these atoms oriented in the direction of the bond to the MMHA to

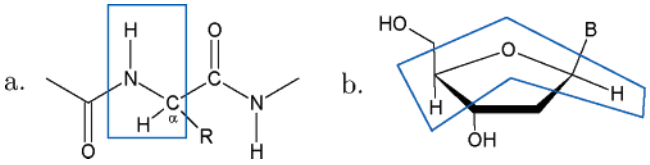

Figure 2. Typical MM host group (MMHG) for amino acids and proteins (a) and deoxyribose (b) in the framework of the CHARMM force field. R substitutes the side chains. B substitutes the purine or pyrimidine base.

compensate for the charge shift. The magnitude and position of these dipoles can be adjusted to obtain an accurate reproduction of the electrostatic potential in the QM region.

A number of authors suggested that no charges need to be exempt if the MM point charges are replaced by a distributed ("blurred") Gaussian charge density (DG scheme.)[38,51,55] Das et al.[51] further suggested that, to counter the dipole moment created by the QM link atom, another link atom should be added to the MMHA; this link atom is termed the molecular mechanics link atom (MML) and the scheme is referred to as the double link atom (DLA) scheme. Impressive accuracy for proton affinities and deprotonation energies were obtained when DLA and the Gaussian blurring approaches were combined.[51] A remaining issue, however, is that the blurring width may not be generally transferable; e.g., different blurring widths were used for optimal protonation and deprotonation energies,[51] which is problematic for studying proton-transfer reactions.

Despite those previous efforts, the quantitative performance of various QM/MM frontier treatments in realistic condensed phase simulations is not well established; previous studies have mainly used gas-phase molecules as benchmark systems, and many such molecules are aliphatic compounds where little polarization is expected from the MM part. Only a few studies have looked into the performance of different schemes for biologically relevant molecules.[13,44,47,55]

For the current work, we systematically evaluated the performance of different link atom approaches available in CHARMM. In addition to typical gas-phase test cases, we studied a number of polar molecules where polarization is important; a number of enzyme systems were also included. We focus on QM/MM calculations using an approximate DFT method (SCC-DFTB) as the QM level. The SCC-DFTB method has been used successfully in QM/MM simulations of a number of biological systems[56−63] and appears to be particularly promising for enzyme studies due to its speed and satisfactory accuracy for many types of reaction; an assessment of different link atom schemes in the SCC-DFTB/MM framework will further help establish a generally effective QM/MM simulation protocol. The insights gained from the current study can also shed light into similar issues concerning QM/MM frontier treatments involving other QM methods, such as ab initio or more sophisticated DFT.

In section 2, we describe the QM/MM formulation for SCC-DFTB/MM and clearly define the different link atom schemes investigated here. In section 3, we present and analyze results obtained with the various schemes; the test cases include a number of polar and nonpolar molecules in the gas phase (section 3.1) and proton-transfer reactions in both small molecules and two enzymes (section 3.2). Finally, we conclude in section 4 with a few remarks.

2. Methods

In this section, we first summarize the QM/MM Hamiltonian in the SCC-DFTB/MM framework. This is followed by the definition of different link atom schemes studied here. Finally, computational details for all test calculations are described.

2.1. The SCC-DFTB/MM Method. As described in eq 1 the potential energy of the QM/MM system is written as,

$$U^{\mathrm{tot}} = \langle \Psi | \hat{H}_{\mathrm{QM}} + \hat{H}^{\mathrm{elec}}_{\mathrm{QM/MM}} | \Psi \rangle + U^{\mathrm{vdW}}_{\mathrm{QM/MM}} + U^{\mathrm{bonded}}_{\mathrm{QM/MM}} + U_{\mathrm{MM}} + U^{\mathrm{cons}}_{\mathrm{QM/MM}} \quad (3)$$

where Ψ is the electronic wave function of the QM region. In this paper, the QM level is the SCC-DFTB method.[5] With different frontier treatment schemes, the operator describing QM/MM electrostatic interactions, $\hat{H}^{\mathrm{elec}}_{\mathrm{QM/MM}}$, takes different forms. In the original implementation,[64] $\hat{H}^{\mathrm{elec}}_{\mathrm{QM/MM}}$ has the form of Coulomb interaction between the MM point charge Q_{A} and the Mulliken charge Δq_{B} on the QM atom.[64]

$$\hat{H}^{\mathrm{elec,pointcharges}}_{\mathrm{QM/MM}} = \sum_{A \in \mathrm{MM}} \sum_{B \in \mathrm{QM}} \frac{Q_{\mathrm{A}} \Delta q_{\mathrm{B}}}{|\vec{r}_{\mathrm{A}} - \vec{r}_{\mathrm{B}}|} \quad (4)$$

To achieve a balanced treatment of electrostatics between MM atoms and QM atoms at different distance separation, a number of authors[38,51,55] suggested using blurred Gaussian charge distributions instead of point charges for the MM atoms in evaluating QM/MM interactions. The blurred charge distribution ρ for a MM atom with total partial charge Q_{A} located at $\vec{R}$ is given as

$$\rho(\vec{r}) = Q_{\mathrm{A}} \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^{3/2} \mathrm{e}^{-(\vec{r} - \vec{R})/2\sigma^2} \quad (5)$$

where σ is termed the blurring width. The corresponding contribution to the Hamiltonian operator similar to eq 4 is:

$$\hat{H}^{\mathrm{elec,blurred}}_{\mathrm{QM/MM}} = \sum_{A \in \mathrm{MM}} \sum_{B \in \mathrm{QM}} \frac{Q_{\mathrm{A}} \Delta q_{\mathrm{B}}}{|\vec{r}_{\mathrm{A}} - \vec{r}_{\mathrm{B}}|} \mathrm{erf}\left(\frac{|\vec{r}_{\mathrm{A}} - \vec{r}_{\mathrm{B}}|}{\sigma} \right) \quad (6)$$

Finally, in the charge shifting model (CHSH),[53,54] the QM atoms also interact with dipoles added to the MM atoms in the frontier region. The corresponding contribution to the Hamiltonian from a dipole $\vec{p}$ centered at $\vec{R}$ is:

$$\hat{H}^{\mathrm{elec,dipole}}_{\mathrm{QM/MM}} = \sum_{B \in \mathrm{QM}} \Delta q_{\mathrm{B}} \frac{\vec{p} \cdot (\vec{r}_{\mathrm{B}} - \vec{R})}{|\vec{r}_{\mathrm{B}} - \vec{R}|^3} \quad (7)$$

2.2. Different Link Atom Schemes. In all test calculations, the QM partition is capped by hydrogen link atoms. After the work was essentially finished, Gao et al.[65] implemented the GHO approach for SCC-DFTB/MM in CHARMM; therefore, the GHO results have not been compared to the link atom based approaches here. As described in section 1, different link atom schemes tested here differ in the way that electrostatic interactions at the QM/MM interface are treated:

- *Single link atom* (*SLA*):[46,48−50] MM partial charges are represented as point charges; charge on the MM host atom (MMHA) is excluded from QM/MM electrostatics.
- *Excluded group (EXGR)*:[52] MM partial charges are represented as point charges; charges on the entire MM host group (MMHG, see Figure 2) are excluded from QM/MM electrostatics.
- *Distributed Gaussian (DG)*:[38,51,55] MM partial charges are represented as Gaussian charge distributions centered on the

TABLE 2: Terms and Interactions Considered for Various QM/MM Schemes[a]

bonded terms		
	bond	MMA-X
	angle	X-MMA-X
	dihedral	X-MMA-X-X or X-X-MMA-X
	improper	MMA-X-X-X (first atom is central atom)
nonbonded terms		
QMA-MMA	electrostatics	interaction through eqs 4, 6, 7
	van der Waals	computed classically
QMA-QMA	electrostatics	not computed classically
	van der Waals	not computed
nonbonded terms involving link atoms		
QML	electrostatics	interaction through eqs 4, 6, 7
	van der Waals	invisible to both MMA and QMA
MML	electrostatics	classical interactions with all MMA, interaction with all QMA through eqs 4, 6, 7
	van der Waals	invisible to both MMA and QMA

[a] X is either MMA or QMA.

respective atoms; no exclusions for QM/MM electrostatic interactions are made. Identical blurring widths are used for all MM atoms.

- *Double link atom* (*DLA*):[51] MM partial charges are represented as Gaussian charge distributions centered on the respective atoms; no exclusions for QM/MM electrostatic interactions are made. In addition to the QM link atom, a MM link atom (MML) is added to the MMHA and placed along the bond between the MMHA and the QMHA. The MML bears a partial charge polarizing all QM atoms, and the magnitude of its charge is subtracted from the MMHA to maintain charge neutrality.
- *Divided (DIV) frontier charge*: MM partial charges are represented as point charges. The charge of the MMHA is deleted and distributed evenly over the remaining MMHG. Similar to CHSH, but no dipoles are added to the MMHG atoms.
- *Charge shift (CHSH)*:[53,54] Redistribution of charges as described for the DIV scheme. A dipole is added at the location of each of the remaining MMHG atoms to compensate for the charge shift from the MMHA. This dipole is exactly opposite to the dipole introduced by the amount of shifted partial charge from the MMHA to the respective atom multiplied by the distance between them.

Note that in almost all cases where partial charges are modified (SLA, EXGR, DIV, CHSH), such modifications are made *only* for the QM/MM interactions. This is crucial as otherwise the accuracy of the carefully parametrized MM/MM interaction by empirical force fields would be compromised. The only exception in which classical interactions are actually modified is the DLA approach. Furthermore, the sum of MM charges interacting with the QM atoms is conserved in all modifications except for SLA methods and EXGR when the overall charge of the MM host group is not neutral (no such cases are presented in this paper).

Although the link-atom based frontier treatments discussed above were used only in the CHARMM framework here, these schemes are applicable in the context of other MM force fields. For schemes that involve modifying charges on the MM host group (CHSH[53,54] and EXGR,[52] DIV), it is not necessary that the host group has an integer charge. In fact, the CHSH and EXGR schemes were initially developed to modify charges within a given number of bonds from the QM or MM host atom; the DIV scheme can be used in the same spirit.

In addition to electrostatics, other terms also have to be considered for QM/MM interactions. Table 2 summarizes terms included and discarded in QM/MM calculations as implemented in SCC-DFTB/CHARMM[64] with different link atom schemes.

2.3. Computational Details. All computations were carried out with the CHARMM27 force field[66] and the SCC-DFTB method.[5] The QM/MM interface implemented by Cui et al.[64] into the CHARMM program[67] was modified to include the DIV and CHSH approaches as well as the DG and DLA schemes described by Das et al.[51] No electrostatic cutoffs were used for the gas-phase models. No QM/MM restraints were used except for the DLA approach, in which the distance between the MM link atom (MML) and MMHA was restrained to be 1.1 Å and the MML was restrained to be along the bond connecting MMHA and QMHA, using a force constant of 4000 kcal/(rad^2·mol).

The group definitions in CHARMM, which are relevant in CHSH, DIV, and EXGR schemes, are as follows: for aliphatic compounds, each methyl and methylene group is charge neutral and constitutes a group in terms of the MMHG. For biomolecules tested here, the MMHG is schematically displayed in Figure 2 for amino acids and nucleotides.

For the gas-phase systems, the energy of all molecules was minimized with respect to their coordinates at the respective QM or QM/MM levels thus all comparisons were made with adiabatic energetics. When computing proton affinities and deprotonation energies, it should be noted that the energy of a proton in SCC-DFTB is not zero; a value of 141.8 kcal/mol is used as in standard parametrizations.[68,69,57]

To obtain the rotational profiles, an adiabatic mapping along the relevant dihedral angle was made between 0° (syn) and 180° (anti) in steps of 5°. At each step, the geometry was fully optimized while the relevant dihedral angle was harmonically restrained with a force constant of 2000 kcal/(rad^2·mol).

For proton transfer, adiabatic mappings were performed by using the anti-symmetric stretch coordinate involving the proton donor, the proton, and the proton acceptor for obtaining initial pathways, which were further refined by using the CPR algorithm[70] to locate the saddle points more precisely. The two enzyme systems studied here, triosephosphate isomerase (TIM) and methyl glyoxal synthase (MGS), were taken from previous SCC-DFTB/MM studies, using the stochastic boundary condition.[71] The systems contain approximately 4000 enzyme atoms and 1179 water molecules for TIM and approximately 7700 enzyme atoms and 384 water molecules for MGS. The QM region contains the substrate (DHAP) and side chains of two essential catalytic residues (His95 and Glu165 in TIM, His98 and Asp71 in MGS). For more details, refer to previous publications.[58,72−74] In the enzyme systems electrostatic force shifting with a cutoff of 13.0 Å was used for MM/MM interactions. No cutoffs were used for QM/MM interactions in any of the models.

Finally, because our focus here is on the influence of QM/MM frontier treatment, full SCC-DFTB results were used as the reference for all test calculations on small molecules. For the gas-phase models no cutoffs where used.

3. Results and Discussion

In the following we compare different link atom approaches using a number of model systems and model reactions. In section 3.1 we explore the performance of different frontier treatments for small gas-phase systems with different sizes of QM partition and MM partition. Specifically, we study proton affinities and deprotonation energies of alcohols and carbonic acids in sections 3.1.1 and 3.1.2. In section 3.1.3 we investigate how well energetics for different rotamers are reproduced using different

TABLE 3: Deprotonation Energies for QM Fragments Embedded in Various MM Environments (in kcal/mol)[a,b]

model		SCC-DFTB	SLA	EXGR[c]	DLA/1.0Å	DLA/1.7Å	DG/1.0Å	DG/1.7Å	CHSH
A	**CH_3-OH**	397.1							
B	CH_3-**CH_2-OH**	397.0	369.2	397.6	417.8	405.2	415.0	407.6	410.7
	dev		−27.8	0.6	20.8	8.2	18.0	10.6	13.7
C	CH_3-CH_2-**CH_2-OH**	395.5	379.0	402.0	414.8	405.2	412.4	407.7	410.2
	dev		−16.5	6.5	19.3	9.7	16.9	12.2	14.7
D	CH_3-CH_2-CH_2-**CH_2-OH**	394.9	378.3	401.4	412.7	404.1	411.9	407.2	409.7
	dev		−16.6	6.5	17.8	9.2	17.0	12.3	14.8
E	**CH_3-CH_2-OH**	397.0							
F	CH_3-**CH_2-CH_2-OH**	395.5	371.9	397.0	402.5	399.4	404.8	402.0	403.1
	dev		−23.6	1.5	7.0	3.9	9.3	6.5	7.6
G	CH_3-CH_2-**CH_2-CH_2-OH**	394.9	381.8	399.4	402.3	399.6	403.4	401.6	402.8
	dev		−13.1	4.5	7.4	4.7	8.5	6.7	7.9
H	CH_3-CH_2-CH_2-**CH_2-CH_2-OH**	394.6	381.6	399.1	400.6	398.9	403.2	401.4	402.6
	dev		−13.0	4.5	6.0	4.3	8.6	6.8	8.0
I	**CH_3-COOH**	366.4							
J	CH_3-**CH_2-COOH**	365.7	338.8	366.3	373.3	367.7	375.5	372.0	374.0
	dev		−26.9	0.6	7.6	2.0	9.8	6.3	8.3
K	CH_3-CH_2-**CH_2-COOH**	365.0	349.6	368.8	370.7	367.5	373.8	371.6	373.3
	dev		−15.4	3.8	5.7	2.5	8.8	6.6	8.3
L	CH_3-CH_2-CH_2-**CH_2-COOH**	364.7	349.3	368.6	370.8	368.2	373.6	371.4	372.6
	dev		−15.4	3.9	6.1	3.5	8.9	6.7	7.9
M	CH_3-CH_2-CH_2-CH_2-**CH_2-COOH**	364.5	349.9	368.3	370.8	368.1	373.6	371.2	372.4
	dev		−14.6	3.8	6.3	3.6	9.1	6.7	7.9

[a] The QM zone consists of the bold-faced portion of the molecule plus the link atom. Deviations are relative to the entire molecule computed with SCC-DFTB. [b] The column SCC-DFTB indicates a calculation encompassing the entire molecule. [c] For the molecules shown here, the DIV scheme is equivalent to the EXGR scheme and is hence not listed separately.

link atom schemes. To explore the situation in compounds with a more inhomogeneous electrostatic potential we examine protonation and deprotonation energies of amino acids and model DNA bases in sections 3.1.5 and 3.1.6. Finally, to make the transition from absolute energies to reactions, we study proton transfer in an adenosine-thymidine Watson−Crick base-pair in section 3.2.1 and in two realistic enzyme systems, TIM and MGS, in section 3.2.2.

3.1. Results for Gas-Phase Models. *3.1.1. Different Sizes of MM Region.* A major assumption in QM/MM applications is that the electronic structure of the QM region is localized in nature and perturbations due to the environment can be described well with a MM treatment for the environment. As a first test of the stability of QM/MM frontier treatments, we studied deprotonation energies for a series of aliphatic alcohols and carbonic acids (see Table 3); in both cases we kept the size of the QM partition constant but varied the size of the molecule (and therefore the size of the MM partition). For the alcohols, a methanol fragment was considered QM for models A−D and an ethanol fragment was taken to be QM for models E−H. For the carbonic acids (models I−M), the QM partition includes the ethanoic acid fragment (see Table 3).

The general trend is that the EXGR and DLA/1.7Å give better results than all other schemes. The widely used SLA approach, which is default in CHARMM, gives the largest error on the order of 20 kcal/mol! The accuracy of DLA and DG schemes depends rather sensitively on the blurring width, especially with a small QM region; changing it from 1.7 Å to 1.0 Å increases the error by the order of 10(2) kcal/mol when using the methanol (ethanol) QM fragment for the alcohols. As the size of the QM region increases from methanol (A−D) to ethanol (E−H), the error in all link atom schemes decreases, as expected; the sensitivity on the blurring width in DG and DLA results is also reduced significantly. One interesting observation is that for all three classes of molecules considered in Table 3, the deviation from full SCC-DFTB results varies very little as a function of the MM size, which shows that the error to a large extent is due to approximate treatment of frontier QM/MM interactions. The only exception found is for EXGR, which gives much smaller errors when the MM region contains merely a methyl group. With EXGR, the partial charges on the entire frontier methyl/methylene group are excluded from QM/MM interactions; as a result, A/E/I have very similar deprotonation energies as B/F/J, respectively, all of which agree very well with full SCC-DFTB calculations. As soon as the MM partition is larger than a single group such that there are MM charges interacting with the QM region, the deviations increase in magnitude to a range comparable to that of DLA/1.7Å.

Although we used SCC-DFTB to describe the QM region, the qualitative trends are very similar to those found in previous studies with different QM methods. For example, models C and F were also studied by Reuter et al.[44] using an LSCF approach and the AM1 Hamiltonian. The errors compared to full QM calculations are 6.6 and 0.6 kcal/mol, respectively, which are close to the values of 6.5 and 1.5 kcal/mol found here with EXGR. For model G, Das et al.[51] found errors of −14.5, 4.5, and 3.2 kcal/mol for SLA, EXGR, and DLA/1.7Å, respectively, using B3LYP/6-31G* as the QM method; these values are very similar to the SCC-DFTB/MM results, which are −13.1, 4.5, and 4.7 kcal/mol, respectively. These observations support the hypothesis that the general trends that we observe here can be generalized to QM/MM approaches with other QM methods.

We note that none of the QM/MM approaches reproduce the trend of decreasing deprotonation energies with increasing molecule sizes for test cases in Table 3. This is not surprising since the effects due to nonpolar MM groups are subtle and within the error bars of QM/MM calculations. For cases with more polar MM groups, QM/MM methods can indeed reproduce qualitative trends very well (see below).

3.1.2. Different Sizes of QM Region. An important choice in QM/MM simulations concerns the size of the QM region and therefore the separation between the QM/MM interface and the reactive site. Whether a small QM region is satisfactory depends critically on the accuracy of the MM force field and treatment of the QM/MM interface. To study the effect of the latter, we chose to study the deprotonation energies and proton affinities with different link atom methods and varying QM sizes, using pentanol and pentanoic acid as test cases (see Tables 4 and 5).

TABLE 4: Deprotonation Energies for Varying QM Zones (in kcal/mol)[a,b]

model		SCC-DFTB	SLA	EXGR[c]	DLA/1.0Å	DLA/1.7Å	DG/1.0Å	DG/1.7Å	CHSH
Q	CH_3-CH_2-CH_2-CH_2-**CH_2-OH**	397.1	378.1	401.2	414.3	404.7	411.8	407.1	409.6
	dev		−16.5	6.6	19.7	10.1	17.2	12.5	15.0
R	CH_3-CH_2-CH_2-**CH_2-CH_2-OH**	397.0	381.6	399.1	400.6	398.9	403.2	401.4	402.6
	dev		−13.0	4.5	6.0	4.3	8.6	7.0	8.0
S	CH_3-CH_2-**CH_2-CH_2-CH_2-OH**	395.5	383.9	397.2	401.0	397.6	400.8	399.1	400.0
	dev		−10.7	2.6	6.4	3.0	6.2	4.5	5.4
T	CH_3-**CH_2-CH_2-CH_2-CH_2-OH**	394.9	379.5	394.9	399.2	396.3	399.1	397.4	398.0
	dev.		−15.1	0.3	4.6	1.7	4.5	2.8	3.4
ref	**CH_3-CH_2-CH_2-CH_2-CH_2-OH**	394.6							
U	CH_3-CH_2-CH_2-**CH_2-COOH**	366.4	349.3	368.6	370.5	368.4	373.7	371.8	372.6
	dev		−15.4	3.9	5.8	3.7	9.0	7.1	7.9
V	CH_3-CH_2-**CH_2-CH_2-COOH**	365.7	353.1	367.2	371.7	367.9	371.4	369.5	370.6
	dev		−11.6	2.5	7.0	3.2	6.7	4.8	5.9
W	CH3-**CH_2-CH_2-CH_2-COOH**	365.0	348.5	364.9	369.7	366.5	369.6	367.6	368.3
	dev		−16.2	0.2	5.0	1.8	4.9	2.9	3.6
ref	**CH_3-CH_2-CH_2-CH_2-COOH**	364.7							

[a] The QM zone consists of the bold-faced portion of the molecule plus the link atom. Deviations are relative to the entire molecule computed with SCC-DFTB. [b] The column SCC-DFTB indicates a calculation including only the capped QM fragment. [c] For the molecules shown here, the DIV scheme is equivalent to the EXGR scheme and is hence not listed separately.

TABLE 5: Proton Affinities for Varying QM Zones (in kcal/mol)[a,b]

model		SCC-DFTB	SLA	EXGR[c]	DLA/1.0Å	DLA/1.7Å	DG/1.0Å	DG/1.7Å	CHSH
Q	CH_3-CH_2-CH_2-CH_2-**CH_2-OH**	−186.7	−169.9	−191.6	−198.8	−191.5	−198.1	−192.8	−195.0
	dev		23.6	1.9	−5.3	2.0	−4.6	0.7	−1.5
R	CH_3−CH_2-CH_2-**CH_2-CH_2-OH**	−191.7	−179.8	−195.7	−203.0	−196.7	−199.7	−198.6	−198.0
	dev		13.7	−2.2	−9.5	−3.2	−6.2	−5.1	−4.5
S	CH_3-CH_2-**CH_2-CH_2-CH_2-OH**	−194.6	−183.4	−196.1	−199.5	−196.4	−197.3	−197.7	−196.4
	dev		10.1	−2.6	−6.0	−2.9	−3.8	−4.2	−2.9
T	CH_3-**CH_2-CH_2-CH_2-CH_2-OH**	−195.3	−180.6	−195.4	−199.9	−196.7	−197.5	−197.7	−196.2
	dev		12.9	−1.9	−6.4	−3.2	−4.0	−4.2	−2.7
ref	**CH_3-CH_2-CH_2-CH_2-CH_2-OH**	−193.5							

[a] The QM zone consists of the bold-faced portion of the molecule plus the link atom. Deviations are relative to the entire molecule computed with SCC-DFTB. [b] The column SCC-DFTB indicates a calculation including only the capped QM fragment. [c] For the molecules shown here, the DIV scheme is equivalent to the EXGR scheme and is hence not listed separately.

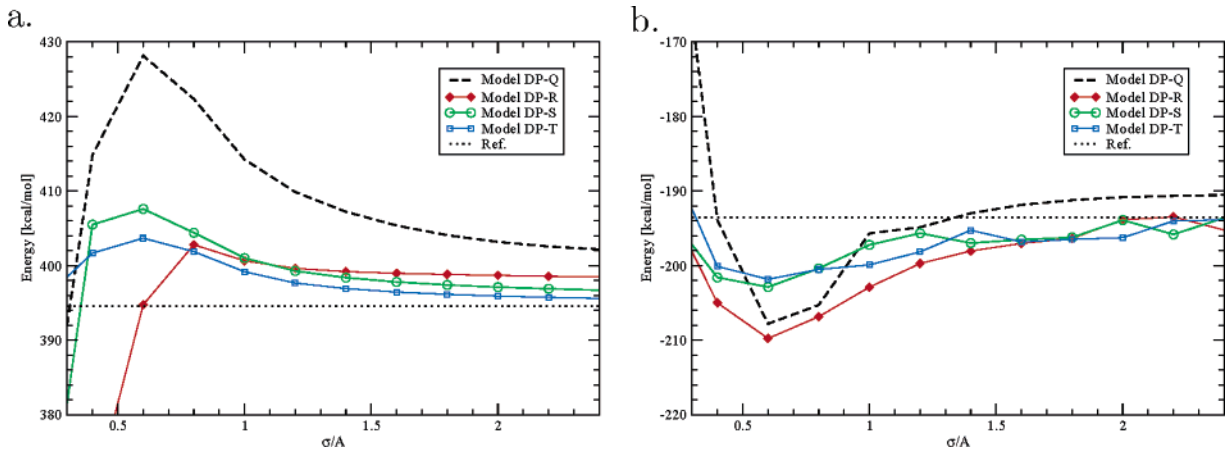

Figure 3. Deprotonation energies (a) and proton affinities (b) for the models introduced in Tables 4 and 5 for varying Gaussian blurring width σ, using the DLA scheme.

As shown in Table 4, the SLA approach gives the worst results for both pentanoic acid and pentanol with an average absolute deviation of ∼14 kcal/mol. Interestingly, the deviations do not decrease systematically with increasing size of the QM partition. For other link atom methods, by contrast, the error of QM/MM calculation does decrease monotonically with increasing size of the QM region. In the case of pentanol, for example, the error decreases from 6.6 (model Q) to 0.3 kcal/mol (model T) for EXGR, from 10.1 to 1.7 kcal/mol for DLA/1.7Å, from 12.5 to 2.8 kcal/mol for DG/1.7Å, and from 15.0 to 3.4 kcal/mol for CHSH. We note that the best link atom results are only slightly worse than those reported for GHO/SCC-DFTB,[65] which includes special parametrizations for the frontier atoms.

For the proton affinities (Table 5), similar trends are seen; e.g., the SLA approach shows a large mean error of 15 kcal/mol, while EXGR, DLA/1.7Å, DG/1.7Å and CHSH give much smaller deviations compared to full SCC-DFTB results. An interesting difference from deprotonation energies is that deviations in proton affinities are more or less constant with respect to the size of the QM region. This is likely due to the fact that the charge distribution in the cationic species involved in proton affinity calculations is highly localized in space, thus error in QM/MM calculations arises mainly from the frontier treatment and is independent from the size of the QM region. In deprotonation energy calculations, by contrast, anionic species are involved, which have a more delocalized charge distribution;

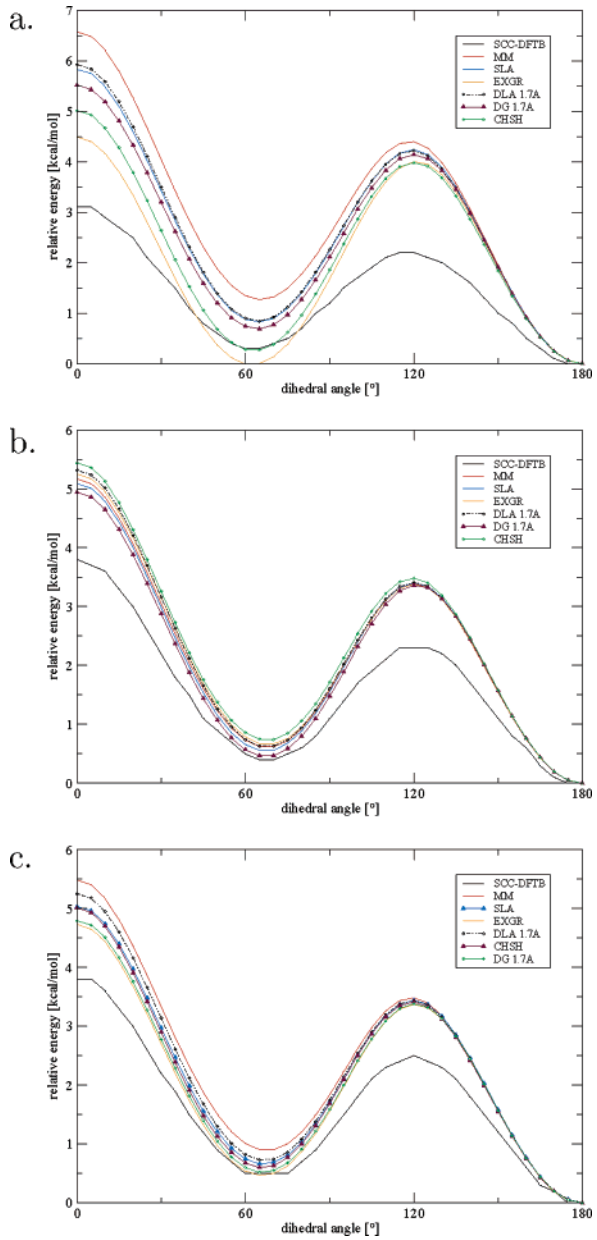

Figure 4. Rotational profile for (a) model Q of pentanol around the QM/MM frontier C-C bond, CH_3-CH_2-CH_2-CH_2-**CH_2-OH**, (b) model R of pentanol around the QM/MM frontier C-C bond, CH_3-CH_2-CH_2-**CH_2-CH_2-OH**, and (c) model S of pentanol around the QM/MM frontier C−C bond, CH_3-CH_2-**CH_2-CH_2-CH_2-OH** (QM zone in bold).

therefore, QM/MM results are more sensitive to the size of the QM region. A generally poorer description of deprotonation energies in comparison to proton affinities was also observed by Reuter et al.[44]

For methods based on Gaussian blurring, i.e., DG and DLA, the results depend on the blurring width, as found in previous work and the last subsection. Previous studies[51] found that different blurring widths are better suited for different properties; e.g., a blur width of 1.0 and 1.7 Å was found to give better deprotonation energy and proton affinity respectively with HF and B3LYP as the QM methods. The current calculations with SCC-DFTB as QM, in contrast, found that 1.7 Å gives better results for both quantities. This is confirmed by calculations with more extensive sampling of different blurring widths (Figure 3). For both deprotonation energies and proton affinities, the sensitivity toward the blurring width decreases sharply with increasing size of the QM region, and this effect is less

TABLE 6: Dipole Moments (magnitude) for Small Molecules in Debye[a]

	CH_3-**CH_3**	CH_3-**CH_2OH**	CH_3-CH_2-**CH_2OH**	CH_3-**CH_2-CH_2OH**
SCC-DFTB	0.00	1.35	1.32	1.32
SLA	0.50	1.40	1.27	1.40
EXGR	0.50	1.19	1.16	1.66
CHSH	1.08	1.42	1.20	2.10
DG/1.0 Å	1.59	1.14	1.32	2.56
DG/1.7 Å	0.90	1.14	1.18	1.96
DG/2.5 Å	0.62	1.16	1.17	1.73
DG/3.0 Å	0.56	1.18	1.19	1.68
DLA/1.0 Å	1.79	1.56	1.39	2.68
DLA/1.7 Å	0.56	1.15	1.18	1.68
DLA/2.5 Å	0.23	1.28	1.27	1.47
DLA/3.0 Å	0.14	1.33	1.31	1.41

[a] The QM fragment is indicated in bold.

pronounced for proton affinity than deprotonation energy. While deviations in deprotonation energies decrease with increasing blurring width, errors may increase for proton affinities when the blurring width is beyond a certain value. The proton affinities of model Q and R are well described by using a blurring width of 1.0 Å, and a blurring width of 1.7 Å seems to be a reasonable compromise for both proton affinity and deprotonation energy. We note that with a large blurring width the results should converge toward the EXGR values. Since EXGR may be problematic when there are important contributions from dipole and higher moments in the MMHG (see below), increasing the blur radius ad libidum is not generally robust.

3.1.3. Rotational Profiles. A useful probe for the performance of the QM/MM frontier treatment is the energy profile for the rotation around the bond at the QM/MM interface. As shown in eq 2, this is described mainly via a classical MM dihedral term, although the QM/MM frontier interaction may also contribute. Since the MM dihedral terms were parametrized to be consistent with the MM electrostatics (e.g., some force fields including CHARMM include no or scaled 1−4 interactions[75]), unbalanced QM/MM frontier interaction may significantly perturb the rotational profile of the frontier bond. The test system chosen in this context is pentanol with different QM/MM partitions that correspond to models Q−S in Table 4; the rotational profiles are shown in Figure 4a−c, where all energies are relative to the anti conformer (180°). In all cases full SCC-DFTB calculations yield too low rotational barriers compared to full MM results. This has also been observed in previous studies with semiempirical treatments.[76] We thus focus on how close can various QM/MM treatments reproduce full MM results as in previous work.[44]

Overall, except when the QM/MM frontier is close to the functional group (OH here), the deviations for different QM/MM approaches from the full MM rotational profile are small. For models R and S, the typical deviation of the various link atom schemes from full MM results is less than 0.5 and 1.3 kcal/mol, respectively. For model Q, in contrast, significant scattering for different link atom approaches is observed (Figure 4a). For example, while MM calculation predicts the gauche conformation to be 1.3 kcal/mol higher in energy than the anti conformer, the EXGR scheme finds the opposite trend. For the syn conformer, results of QM/MM calculations scatter to lower energies compared to full MM results. Here EXGR shows the largest deviation with an underestimation by 2.1 kcal/mol. The DLA/1.7Å is found to give results most similar to the full MM treatment.

The observed trend can be rationalized based on the charge distribution in the QM/MM interface region. In models R and S, atoms in the vicinity of the rotating bond have small partial

TABLE 7: Deprotonation Energies for Amino Acids (in kcal/mol)[a,b]

	SCC-DFTB	SLA	EXGR	DIV	DLA/1.0Å	DLA/1.7Å	DG/1.0Å	DG/1.7Å	CHSH
Ac-Lys-OCH_3	222.8	231.3	226.5	222.2	218.7	219.7	219.1	220.1	220.1
dev		8.5	3.7	−0.6	−4.1	−3.1	−3.7	−2.7	−2.7
Ac-Lys-OCH_3[c]	222.8	206.6	220.6		220.5	220.1	224.4	222.7	223.7
dev		−16.2	−2.2		−2.3	−2.7	1.6	−0.1	0.9
Ac-His$^{\delta}$-OCH_3	375.1	390.7	385.8	376.0	370.2	371.4	368.6	372.2	372.1
dev		15.6	10.7	0.9	−4.9	−3.7	−6.5	−2.9	−3.0
Ac-His$^{\epsilon}$-OCH_3	374.0	390.1	385.7	377.2	371.3	372.3	369.6	373.5	373.4
dev		16.1	11.7	3.2	−2.7	−1.7	−4.4	−0.5	−0.6
Ac-Tyr-OCH_3	356.4	368.7	365.2	358.4	353.5	354.5	352.4	355.3	355.3
dev		12.3	8.8	2.0	−2.9	−1.9	−4.0	−1.1	−1.1
rmsd		14.6	8.3	2.0	3.5	2.7	4.3	1.9	1.9

[a] Ac = acetyl protecting group, [b] The column SCC-DFTB indicates the property of the entire molecule calculated with SCC-DFTB. For the remaining QM/MM calculations the side chain including the link atom is treated with SCC-DFTB while the backbone is treated with MM. Deviations are relative to the entire molecule computed with SCC-DFTB. [c] QM/MM frontier between C^{β} and C^{γ}.

TABLE 8: Proton Affinities for Amino Acids (in kcal/mol)[a,b]

	SCC-DFTB	SLA	EXGR	DIV	DLA/1.0Å	DLA/1.7Å	DG/1.0Å	DG/1.7Å	CHSH
Ac-Asp-OCH_3	−356.7	−373.0	−368.5	−358.3		−350.4	−350.8	−353.2	−353.1
dev		−16.3	−11.8	−1.6		6.3	5.9	3.4	3.6
Ac-Glu-OCH_3	−359.7	−371.4	−367.6	−359.6	−351.7	−354.2	−352.5	−356.1	−355.9
dev		−11.7	−7.9	0.1	7.9	5.5	7.2	3.6	3.8
Ac-Glu-OCH_3[c]	−359.7	−340.5	−359.3		−363.2	−355.8	−363.9	−361.7	−363.5
dev		19.2	0.4		−3.5	3.9	−4.2	−2.0	−3.8
Ac-Tyr-OCH_3	−193.8	−200.7	−196.2	−191.3	−187.3	−187.8	−188.2	−188.9	−189.1
dev		−7.0	−2.4	2.5	6.5	6.0	5.6	4.9	4.7
Ac-His$^{\delta}$-OCH_3[d]	−255.1	−261.5	−252.7	−248.3	−243.3	−244.1	−243.2	−245.3	−245.0
dev		−6.4	2.4	6.8	11.8	11.0	11.9	9.8	10.1
Ac-GluH-OCH_3	−205.2	−213.4	−207.7	−202.7	−187.3	−188.4	−197.6	−199.3	−199.6
dev		−8.2	−2.5	2.5	17.9	16.8	7.6	5.9	5.6
rmsd		12.4	6.1	3.5	10.7	9.4	7.5	5.5	5.7

[a] Ac = acetyl protecting group, [b] The column SCC-DFTB indicates the property of the entire molecule calculated with SCC-DFTB. For the remaining QM/MM calculations the side chain including the link atom is treated with SCC-DFTB while the backbone is treated with MM. Deviations are relative to the entire molecule computed with SCC-DFTB. [c] QM/MM frontier between C^{β} and C^{γ}. [d] For the doubly protonated H is a hydrogen bridge formed between the proton attached to N^{δ} and the carbonyl oxygen. Fixing the backbone of the amino acid did not prevent this for all schemes. The deviations hence also reflect the imperfect description of a QM/MM hydrogen bridge.

charges and hence boundary QM/MM electrostatic interactions do not play an important role; as a result, the rotational profile is dominated by the classical MM dihedral term. In model Q, however, the rotating bond is close to the functional group (OH), which bears significant partial charges; therefore, frontier QM/MM electrostatic interactions may significantly modify the rotational profile, depending on the way such interactions are treated in different link atom schemes. This is most obvious for the EXGR scheme and model Q, where the gauche conformation is predicted to be more stable than the anti conformation. Still, this is the largest deviation for all cases shown here and the deviation with respect to full MM is smaller than 2 kcal/mol.

3.1.4. Dipole Moments. Charge distribution in molecules is of crucial importance for reactivity and interaction with the environment. A reliable QM/MM scheme should hence be able to reproduce the charge distribution correctly, and one way to analyze this is to compare dipole moments computed with QM/MM against full QM results. Systems chosen here include ethane as an unpolar test case and ethanol/propanol as polar test cases.

As seen in Table 6, *all* QM/MM schemes yield nonvanishing dipole moments for ethane. The DLA scheme, which is motivated by obtaining correct dipole moments in QM/MM calculations,[51] gives a large value (e.g., 1.0 Å) unless a large blurring width is used. For polar examples (ethanol and propanol) that involve a methanol QM region, the deviation from the full SCC-DFTB result is small relative to the absolute value of the dipole moment for all link atom schemes, and the maximum deviation is only about 0.2 D. With a larger QM region in propanol, somewhat unexpectedly, very large errors

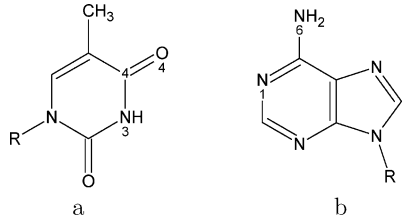

Figure 5. Structure and label of atoms of nucleotides studied here: (a) 2′-deoxythymidine and (b) 2′-deoxyadenosine. R = 2′-deoxyribose.

(on the order of 1 D!) are observed for CHSH and DG/DLA calculations with a small blurring width. As the blurring width increases, both DLA and DG results improve substantially and the DLA converges toward the correct value faster. The fact that errors in the dipole moment increase for some link atom schemes with a larger QM region while errors in deprotonation energies decrease for most cases (Table 3) is not difficult to understand. Deprotonation energy is a differential quantity, thus certain inaccuracies of the electronic structure due to QM/MM frontier treatment can be masked due to error cancellation; dipole moment, on the other hand, is a more direct and sensitive probe of the electron distribution in *one* molecule. Being a collective variable, however, an accurate dipole moment *alone* does not necessarily imply an accurate treatment; a good example is that SLA gives rather decent dipole moments for the molecules tested here (Table 6) but produces large errors for deprotonation energies (Table 3).

3.1.5. Amino Acids. When studying reactions in proteins, one obvious way to partition the system is to take the side chains of interest into the QM partition and the rest into the MM

TABLE 9: Deprotonation Energies for DNA Bases: 2′-Deoxyadenine (dA) and 2′-Deoxythymidine (in kcal/mol)[a]

	SCC-DFTB	SCC-DFTB (R = H)[b]	SLA	EXGR	DIV	DLA/1.0Å	DLA/1.7Å	DG/1.0Å	DG/1.7Å	CHSH
dA	363.4	369.8	382.1	367.4	374.1	365.0	366.2	367.2	367.9	374.6
dev			−18.7	−4.0	−10.7	−1.6	−2.8	−3.8	−4.5	11.2
dT	351.4	356.2	371.0	354.4	359.4	349.2	351.1	352.5	353.7	355.5
dev			−19.6	−3.0	−8.0	2.2	0.3	−1.1	−2.3	4.1

[a] The column SCC-DFTB indicates the property of the entire molecule calculated with SCC-DFTB. For the remaining QM/MM calculations the QM zone encompasses the purine or pyrimidine ring and the link atom; the sugar is treated with MM. Deviations are relative to the full molecule computed with SCC-DFTB. [b] Minimal model.

partition, cutting across the $C^{\alpha}-C^{\beta}$ bond. It is important to establish robust schemes for such a partitioning. In contrast to the aliphatic alcohols and carbonic acids discussed above, amino acids show a nonuniform charge distribution in the backbone, thus the QM/MM frontier scheme is expected to have an even greater impact on the result compared to the cases studied above. The benchmark chosen involves deprotonation energies (Tables 7) and proton affinities (Tables 8) for the polar and charged amino acid side chains. These properties are of interest because change in protonation state is a major mechanism for amino acid side chains to be involved in general acid−general base catalysis.[77] In all calculations, the N-terminus was capped with an acetyl residue and the C-terminus with a methyl-ester, to mimic QM/MM partition inside a protein.

Overall, all models other than SLA and EXGR give a rather similar performance. The SLA approach shows the largest deviations relative to full SCC-DFTB data with 16.2 kcal/mol for deprotonation and 19.2 kcal/mol for proton affinities. This is somewhat surprising (and alarming) as in many cases the protonation/deprotonation site is far from the QM/MM interface. For example, the deviation in the deprotonation energy for lysine is as high as 8.5 kcal/mol despite the large distance between the deprotonation site and the QM/MM interface. For tyrosine and histidine, significant deviations could be rationalized with unsatisfactory treatment for the polarization of the aromatic ring.

The EXGR results are mixed in quality. For deprotonation energies, the deviations for lysine are small and comparable with other methods; for the other amino acids, the deviations are about 10.6 kcal/mol and of similar magnitude as those for SLA. For the proton affinities, the deviations are generally lower (5.5 kcal/mol) and are dominated by the poor description of aspartate and glutamate.

The general difficulty associated with the EXGR approach is caused by the fact that the backbone N−H belongs to the same group as the $C^{\alpha}H^{\alpha}$ in the CHARMM force field, which means that a major dipole moment is excluded from QM/MM frontier treatment. This problem is alleviated to a certain extent by the DG, DLA, DIV, and CHSH schemes, which keep the N−H charges while still being careful about the charge neutrality at the interface (unlike SLA). Indeed, they give generally similar results, although the DIV approach appears to be most robust. For example, the DG and DLA calculations with a blurring width of 1.7 Å, have an absolute average deviation of 2.6 and 1.5 kcal/mol, respectively, for the deprotonation energies; the CHSH and DIV models give average errors of 1.7 kcal/mol. However, only the DIV treatment gives a comparably reliable treatment for the protonation affinities (1.7 kcal/mol), while other methods produce much larger errors: the DG/1.7Å, DLA/1.7Å, and CHSH methods have deviations of 4.0, 8.3, and 4.3 kcal/mol, respectively, and the DLA/1.0Å scheme even had the problem that the aspartate anion is not stable and dissociated between C^{α} and C^{β}.

One interesting question at this point is whether the quality of the description can be improved by choosing a different QM/MM partitioning scheme. For instance, both the lysine and the glutamate side chains are sufficiently long for shifting the frontier bond to $C^{\beta}-C^{\gamma}$, which would put the QM/MM interface closer to the protonation/deprotonation site but leave the electrostatics of the backbone unperturbed (note that EXGR and DIV become identical with this partition). As seen in Table 8, the proton affinities of glutamate improve significantly except for the SLA approach, which suffers from the decreased size of the QM zone; the error in CHSH did not improve either and became of different sign. The error for EXGR drops significantly to 0.4 kcal/mol, which is roughly the same as that for the DIV scheme in the original $C^{\alpha}-C^{\beta}$ partitioning. For the deprotonation energies (Table 7), improvements in lysine are substantial for some approaches. The EXGR result improves from 3.7 to −2.2 kcal/mol, and both the DLA and the DG schemes as well as the CHSH approach showed similar improvements. For glutamate, a similar decrease in error was seen in the study of Reuter et al. using AM1[44] as the QM method with different QM/MM partitioning schemes.

3.1.6. DNA Bases. To further explore the performance of various link atom schemes with different kinds of QM/MM interfaces, we chose to study the acidity and basicity of 2′-deoxyadenosine (dA) and 2′-deoxythymidine (dT), which are representative of DNA bases (Figure 5) with the phosphate groups omitted. We also studied the relative stability of different tautomers. Deprotonation refers to the removal of one proton at ^{3}N and ^{6}N in dT and dA, respectively, and protonation involves ^{4}O and ^{1}N, respectively. The tautomerization involves a proton exchange between the deprotonation and the protonation sites (e.g., between ^{3}N and ^{4}O for dT). The QM/MM partition occurs at the sugar-base bond, and the base was treated with QM (SCC-DFTB). To illustrate the effect of the sugar ring, QM results for the minimal models which are terminated with a hydrogen atom at R were also included (Figure 5).

Overall, the SLA scheme gives the largest error, as found above for other systems. The trends in the performance of the other link atom schemes, however, are somewhat different from those observed for amino acids, which highlights the importance of the charge distribution at the QM/MM interface. For the deprotonation energies (Table 9) and proton affinities (Table 10), EXGR, DLA, and DG give satisfactory results, while the DIV and CHSH schemes, which work well for amino acids (section 3.1.5), give much larger errors on the order of 10 kcal/mol in several cases. For methods using Gaussian blurring, no significant dependence on the blurring width is found.

The tautomerization energy (Table 11), as expected, depends much less on the link atom scheme presumably due to error cancellation. Here, the energetics are in excellent agreement with the full SCC-DFTB calculation with errors less than 2 kcal/mol, even with SLA. The only exception is for dA with the EXGR scheme, which gives a large deviation of 4.1 kcal/mol, which is even larger than that of the minimal model.

3.2. Results for Proton-Transfer Reactions. Results from the last subsection indicate that large errors in protonation and

TABLE 10: Proton Affinities for DNA Bases: 2′-Deoxyadenine (dA) and 2′-Deoxythymidine (in kcal/mol)[a]

	SCC-DFTB	SCC-DFTB (R = H)[b]	SLA	EXGR	DIV	DLA/1.0Å	DLA/1.7Å	DG/1.0Å	DG/1.7Å	CHSH
dA	237.2	235.5	256.3	241.4	244.5	236.9	238.3	239.9	241.0	238.0
dev			19.1	4.2	7.3	−0.3	1.1	2.7	3.8	0.8
dT	233.0	222.7	246.8	228.2	235.3	225.8	227.7	229.0	230.2	231.5
dev			13.8	−4.8	2.3	−7.2	−5.3	−4.0	−2.8	−1.5

[a] The column SCC-DFTB indicates the property of the entire molecule calculated with SCC-DFTB. For the remaining QM/MM calculations the QM zone encompasses the purine or pyrimidine ring and the link atom; the sugar is treated with MM. Deviations are relative to the full molecule computed with SCC-DFTB. [b] Minimal model.

TABLE 11: Energy Difference for the Tautomers for DNA Bases (in kcal/mol): 2′-Deoxyadenine (dA) and 2′-Deoxythymidine[a]

	SCC-DFTB	SCC-DFTB (R = H)[b]	SLA	EXGR	DIV	DLA/1.0Å	DLA/1.7Å	DG/1.0Å	DG/1.7Å	CHSH
dA	12.0	11.7	11.5	7.9	11.7	12.9	12.4	12.2	12.2	12.0
dev			0.5	4.1	0.3	0.9	0.4	0.2	0.2	0.0
dT	5.1	5.9	6.7	6.4	6.1	5.5	5.5	5.5	5.7	5.9
dev			1.6	1.3	1.0	0.4	0.4	0.4	0.6	0.8

[a] The column SCC-DFTB indicates the property of the entire molecule calculated with SCC-DFTB. For the remaining QM/MM calculations the QM zone encompasses the purine or pyrimidine ring and the link atom; the sugar is treated with MM. Deviations are relative to the full molecule computed with SCC-DFTB. [b] Minimal model.

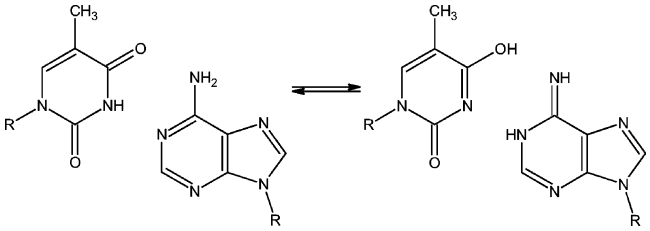

Figure 6. Tautomeric proton transfer in a 2′-deoxyadenine (dA)−2′-deoxythimidine (dT) Watson−Crick base pair.

TABLE 12: Activation Energies $\Delta E^{\ddagger}$ and Endothermicities ΔE (in kcal/mol) for the Tautomeric Proton Transfer in the dT-dA Base Pair[a]

	$\Delta E^{\ddagger}$ [kcal/mol]	ΔE [kcal/mol]
SCC-DFTB	15.8	8.0
SCC-DFTB (R = H)	14.9	8.5
SLA	15.2	8.2
EXGR	16.6	9.1
DIV	18.9	10.2
DLA/1.0 Å	15.0	9.0
DLA/1.7 Å	14.7	8.4
DG/1.0 Å	17.3	10.2
DG/1.7 Å	17.4	10.0
CHSH	17.5	9.9

[a] The column SCC-DFTB indicates the property of the entire system calculated with SCC-DFTB. For the remaining QM/MM calculations the QM zone encompasses the purine and pyrimidine ring and the link atoms; the sugars are treated with MM. [c] Minimal model.

deprotonation energies may arise if QM/MM interactions at the interface are not handled properly. The encouraging aspect, however, is that properties that conserve the total charge (e.g., rotational energy profile and tautomerization energy of DNA bases) show much less dependence on the link atom schemes due presumably to error cancellations. This observation motivated us to further explore the effect of QM/MM frontier treatment on *reaction energetics*, which are the quantity of primary interest in most QM/MM applications. We first examine results for proton transfer in a DNA base pair in the gas phase and then move on to more realistic enzyme systems.

3.2.1. Proton Transfer in a DNA Base Pair. Proton transfer (PT) was studied for an 2′-deoxyadenine (dA)−2′-deoxythimidine (dT) Watson−Crick base pair as shown in Figure 6. To simplify geometry optimizations, heavy atoms in the sugar moieties were kept fixed during the PT. The QM/MM partition

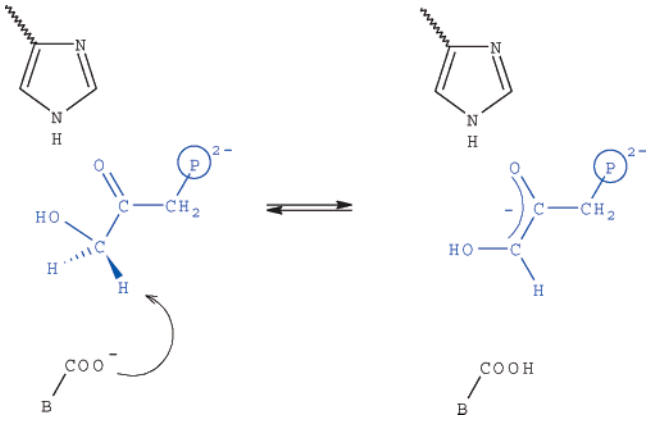

Figure 7. First proton-transfer step in TIM and MGS, in which the base B abstracting the proton from DHAP is Glu165 and Asp71 for TIM and MGS, respectively; the histidine residue is His95 and His98 in TIM and MGS, respectively.

TABLE 13: Activation Energies $\Delta E^{\ddagger}$ and Endothermicities ΔE (in kcal/mol) for the First Proton Transfer Step in TIM

	$\Delta E^{\ddagger}$ [kcal/mol]	ΔE [kcal/mol]
SLA	10.9	8.7
EXGR	12.3	11.0
DIV	10.1	8.8
DLA/1.0 Å	14.2	13.4
DLA/1.7 Å	14.3	13.5
DG/1.0 Å	12.0	10.8
DG/1.7 Å	11.6	10.4
CHSH	12.3	11.1

occurs at the sugar-base bond, and the base was treated with QM (SCC-DFTB). To illustrate the effect of the sugar atoms in PT, full QM results for a minimal model without the deoxyribose residue (R = H) are also included (Figure 6).

The PT is endothermic by about 8 kcal/mol, and has a barrier about 16 kcal/mol without the zero point energy correction (Table 12). Similar to the tautomerization energies, the PT energies (both activation barrier and endothermicity) show very minor dependence on the link atom schemes. Indeed, even the minimal models without the sugar part at all give very similar results; the differences for the barrier height ($\Delta E^{\ddagger}$) and endothermicity are 0.5 and 0.9 kcal/mol, respectively. In the QM/MM calculations, the maximum absolute deviation is found for the DIV scheme, which overestimates $\Delta E^{\ddagger}$ by 3.1 kcal/mol and the endothermicity by 2.2 kcal/mol. The absolute deviations of

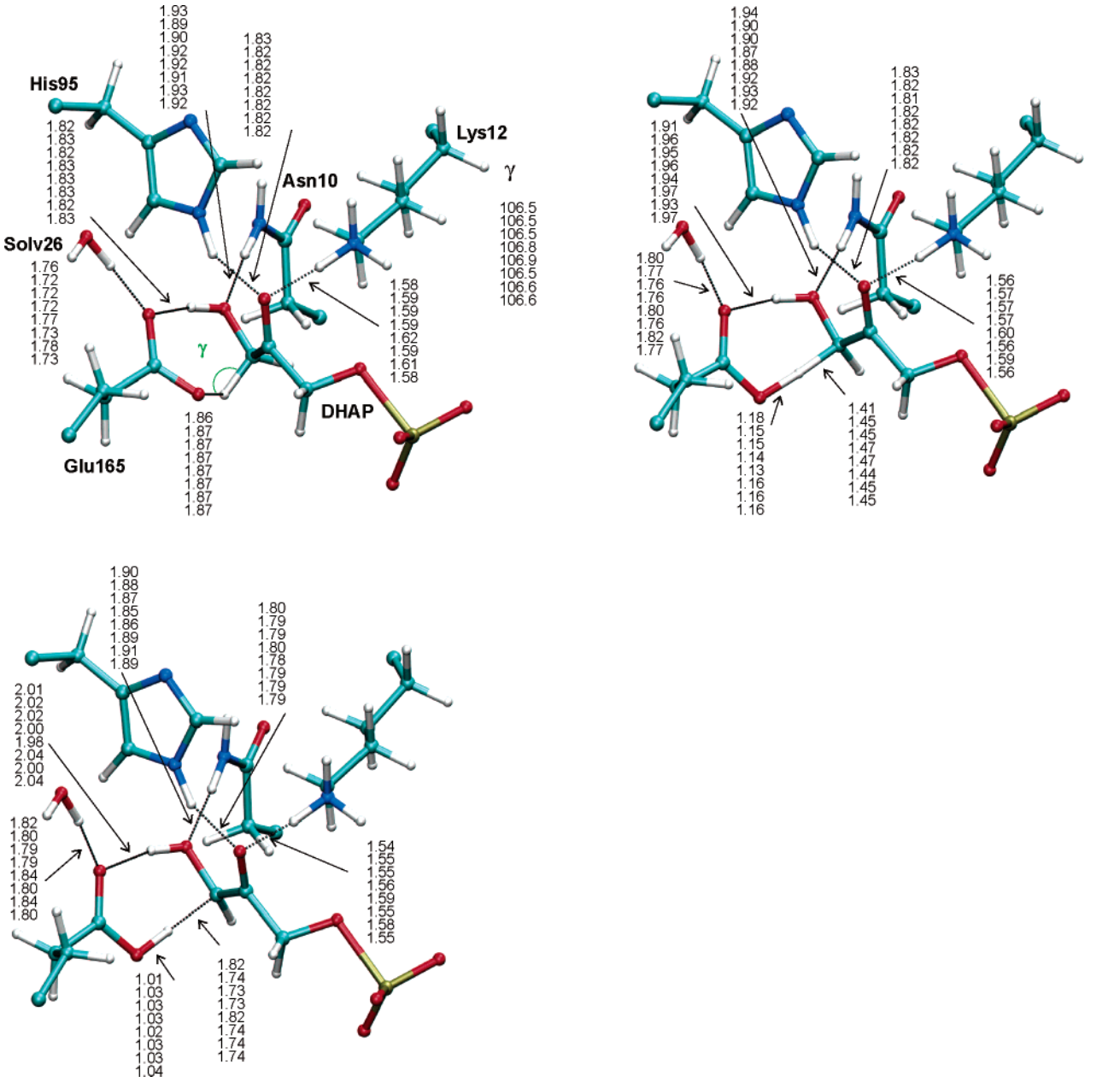

Figure 8. Structures and important geometrical parameters for the reactant (top left), transition state (top right), and product (below) for the first proton transfer in TIM. Distances are in Å, angles in deg. Numbers from *top to bottom* for SLA, EXGR, DIV, DLA/1.0Å, DLA/1.7Å, DG/1.0Å, DG/1.7Å, and CHSH. The figure was prepared with the VMD program.[110]

other methods are smaller in magnitude, and all schemes overestimate the energy difference while only the DLA and SLA schemes underestimate the barrier. For example, the DLA scheme underestimates the barrier by 1.1 kcal/mol.

Since the sugar ring does influence the deprotonation and protonation energetics rather significantly (Tables 9 and 10), the minimal dependence of the PT energetics on the QM/MM link atom scheme is interpreted as a simple consequence of error cancellation, similar to the case of tautomerization discussed above. According to the Hammond postulate, it is not surprising to see that errors in the barrier height also follow the trend in the reaction energy; this is very reassuring, however, for the application of QM/MM schemes to chemical reaction problems.

3.2.2. Proton Transfer in Enzymes: *TIM and MGS.* To illustrate the effect of link atom schemes for reactions in enzymes, the first PT step in triosephosphate isomerase (TIM) and methyl glyoxal synthase (MGS) was studied. These systems were chosen based on several considerations: Both systems have been well studied by previous experimental[78−87] and theoretical work[12,72−74,88−93] which provided firm grounds for benchmarking different link atom schemes. Although evolved independently, TIM and MGS have very similar active sites that bind to the same substrate, dihydroxyacetone phosphate (DHAP), and are expected to undergo the same first PT step[94,95] (Figure 7). The difference is that the general acid in this step is Glu165 for TIM and the shorter Asp71 for MGS; the electrostatic environment is also quite different in the two active sites, which was proposed to be crucial for their functional specificity.[12,58] Therefore, comparing the dependence of PT energetics to different QM/MM link atom schemes in these enzymes will clearly reflect the range of variations of QM/MM results in realistic biological systems.

As described in section 2.3, the QM region, treated with SCC-DFTB, included the substrate, the general base, and distal histidine (His95 in TIM and His98 in MGS). Considering the size of the system, instead of performing QM/MM calculations

TABLE 14: Activation Energies $\Delta E^{\ddagger}$ and Endothermicities ΔE (in kcal/mol) for the First Proton Transfer Step in MGS

	$\Delta E^{\ddagger}$ [kcal/mol]	ΔE [kcal/mol]
SLA	9.8	7.4
EXGR	8.8	7.0
DIV	11.7	10.2
DLA/1.0 Å	12.7	12.2
DLA/1.7 Å	12.8	11.9
DG/1.0 Å	11.8	10.5
DG/1.7 Å	11.9	10.8
CHSH	11.6	10.3

with a much larger QM region, we evaluated the performance of QM/MM frontier treatments by focusing on variations between results obtained from different link atom schemes.

As discussed in previous work, the first proton transfer in both TIM and MGS, despite being catalyzed by the enzymes, is endothermic by a substantial amount.[12,72−74] For TIM, the barrier heights range from 10.1 kcal/mol for DIV to 14.3 kcal/mol for DLA/1.7Å (Table 13). These schemes also give the minimum (8.8 kcal/mol) and maximum (13.5 kcal/mol) for the energy difference between the reactant and product. All other methods give energies within these ranges. In all cases the barrier is only slightly higher than the product.

These results are in qualitative agreement with previous work. For example, Cui et al. found a barrier of 17 kcal/mol and an energy difference of 9 kcal/mol using a specifically adjusted AM1 method.[74] More recently, using HF/3-21G as the QM method, the pseudo-bond frontier treatment,[34] and more extensive reaction path searches, Liu et al. found the barrier and endothermicity to be 21.3 and 19.4 kcal/mol, respectively.[96] Results from an approximate QM/MM free energy perturbation approach gave very similar results as well.[97]

As to critical geometrical parameters in the reactant, transition state, and product of the PT, all link atom schemes give very similar results (see Figure 8). For instance, the distances between the transferring proton and the donor/acceptor in the transition state are within 0.05 Å among all frontier treatments. Therefore, the trend that all link atom schemes yield comparable barriers and energy differences is also reflected in the geometries.

For MGS, which involves a shorter general base for the PT (Asp71), EXGR, and DLA/1.7Å schemes, Table 14 gives the minimum and maximum for the barrier height (8.8 kcal/mol, 12.8 kcal/mol); for the energy difference, EXGR and DLA/1.0Å give the minimum and maximum values (7.0 kcal/mol, 12.2 kcal/mol). In other words, the range of deviation in the QM/MM results for the barrier heights is approximately the same in MGS (4.0 kcal/mol) and in TIM (4.2 kcal/mol). Critical geometrical parameters are also very similar among all link atom schemes (see Supporting Information). Apparently, the shorter QM general base and more polar active site in MGS does not strongly increase the error bars for different link atom approaches. We note that in contrast to the PT in the DNA base pair discussed above, only one of the QM groups is directly influenced by the QM/MM interface; i.e., the entire substrate is treated with QM, and the QM/MM partition explicitly involves only the general base. Therefore, the degree of error cancellation for the PT between the substrate and the general base is expected to be different from that in DNA bases, and it is satisfying to see that relatively small differences between different QM/MM link atom schemes occur for both systems.

4. Concluding Discussions

Although numerous studies have demonstrated that QM/MM methods are powerful tools for studying reaction mechanisms[10−12,56] and related properties[98−101] in complex systems, many technical details such as the treatment of long-range electrostatics,[102−104] QM/MM van der Waals interactions,[31,105] and the extent of configuration samplings,[8,106] as well as the choice of MM partial charges,[42,107] *may* have a major impact on whether the results are meaningful. Precisely how important a contribution might be, which depends on the system under study and questions being addressed, needs to be established by careful and relevant benchmark calculations. In this study, we systematically evaluated the importance of the treatment of QM/MM interface (frontier), an issue that has been raised repeatedly in the QM/MM community. As emphasized in a recent study of QM/MM van der Waals interactions,[31] to evaluate whether this is the case in *typical* applications, one needs to investigate not only gas-phase systems as is often done but also realistic condensed phase systems and processes (e.g., reaction in enzymes). Accordingly, we systematically studied the effect of different link atom schemes in QM/MM calculations of proton affinity, deprotonation energy, and rotational barriers in a number of gas-phase systems of various polarity (alcohols, carbonic acids, amino acids, and model DNA bases), as well as proton transfer reactions in model DNA base-pairs and two enzymes. The breadth of the test cases was chosen to ensure a critical evaluation on the performance of QM/MM link atom schemes for realistic systems of biological interest. To focus on the effect of QM/MM frontier treatment, an approximate density functional theory (SCC-DFTB) was used throughout; in most cases, full SCC-DFTB results were taken as reference values for QM/MM results. However, we expect that the trends observed here concerning the effect of QM/MM frontier treatment apply to other QM/MM methods, which is supported by comparisons to previous studies[35,44,51] for several gas-phase systems with ab initio or DFT as the QM method.

For deprotonation energies and proton affinities, the current work found similar results as previous studies[51] in that the commonly used SLA approach, which (unfortunately) is the default option in CHARMM, gives large errors typically in the range of 15 to 20 kcal/mol;[51] more alarmingly, this error does not decrease systematically as the QM/MM interface moves away from the protonation/deprotonation site (i.e., with increasing QM size). Therefore, for absolute proton affinity and deprotonation energy calculations, such as pK_a evaluations,[63] the SLA approach should be avoided; we expect similar trends for other processes that involve significant changes in charge distribution, such as redox potential calculations.[59,62] In other QM/MM frontier treatments, by contrast, moving the boundary away from the reactive site generally improves proton affinities and deprotonation energies, as also pointed out by other authors.[42,44,107]

A major problem with SLA is that with the MM host atom (MMHA) excluded from the QM region, a net partial charge is created in the vicinity of the QM region, which causes spurious electrostatic interactions. This problem seems to be alleviated by most other schemes, which conserve the local charge neutrality at the QM/MM interface by either shifting the MMHA charge to the entire host group (CHSH, DIV) or excluding the entire host group (EXGR); alternatively, blurring MM point charges into Gaussian of finite width as in DG and DLA also works well. It is difficult to argue which among these schemes works consistently the best, since they all have successes and failures among the test cases examined here, which highlighted the importance of including benchmark systems of diverse chemical nature. For example, the EXGR approach works well when the MM host group does not bear any significant dipole moment but may cause serious error otherwise. A highly

relevant case involves partitioning amino acids (or proteins) across the $C^{\alpha}-C^{\beta}$ bond, where EXGR may cause significant error because a major dipole associated with the backbone N−H would be excluded from QM/MM electrostatic interactions (section 3.1.5). For amino acids with sufficiently long side chains (e.g., lysine and glutamate), it seems better to partition across $C^{\beta}-C^{\gamma}$, as suggested in another study.[44] The CHSH and DIV schemes (note that with aliphatic MM host groups, DIV is equivalent to EXGR due to the group definition in the CHARMM force field) gave encouraging results for amino acids, but substantial errors on the order of 10 kcal/mol in deprotonation energy were also observed in model DNA bases. On average, DIV and CHSH give similar results, although the DIV scheme tends to give better deprotonation energies. Methods that rely on Gaussian blurring (DLA and DG) generally give better results than SLA, but the accuracy depends quite sensitively on the blurring width although the dependence decreases as the QM/MM interface moves away from the reactive site (Table 4). For dipole moments, significant error may arise with small blurring widths (Table 6). With systematic tests on the blurring width, the DLA approach appears to be a promising choice for practical calculations of properties such as proton affinity and deprotonation energy.

The encouraging finding in the current work, although somewhat anticipated, is that reaction energies have a much less pronounced dependence on the link atom scheme. For the tautomerization energy of model DNA bases, proton transfers in model DNA bases pairs as well as in two enzymes, *all* link atom schemes gave very similar results, including the SLA approach. Apparently error cancellation is a major reason behind the reduced sensitivity toward QM/MM frontier treatment, because proton affinity and deprotonation energy depend sensitively on the link atom schemes. We note that although this is encouraging in general, there are still nonnegligible variations in the results of various link atom schemes; for proton transfers in MGS and TIM, for example, the range is on the order of 2−4 kcal/mol, and whether this is acceptable depends on the specific application.

Overall, we can conclude that while neither of the approaches examined here is a perfect solution to the frontier problems in QM/MM simulations, for many cases both DIV and the DLA scheme give reasonable results, especially for chemical reactions. Therefore, we agree with a previous study[13] that the effect of the frontier treatment on QM/MM study of chemical reactions is rather small, especially if the total charge is conserved during the reaction. This claim does not mean to encourage QM/MM simulations without careful benchmark in the study of specific systems, rather this investigation emphasizes that other technical details, especially the level of QM method and treatment of long-range electrostatics,[59,62,103,104,106] tend to play an even more important role and need to be handled carefully. For example, electrostatic interactions are known to be crucial to enzyme catalysis,[106] yet electrostatics are often treated in highly approximate manner in QM/MM simulations. Partial charges for charged residues on the surface of the enzyme are either quite arbitrarily set to zero or scaled down[108] according to Poisson−Boltzmann calculations based on a *single* conformation. Moreover, in stochastic boundary simulations,[71] atoms outside of the buffer region are often deleted and the reservior region is essentially vacuum; some studies keep the reservior atoms but the dielectric screening effect due to the bulk solvent is ignored. All these rather *ad hoc* approximations may work for the reaction path type of QM/MM analysis for very localized chemical reactions, but they are unlikely to be robust for QM/MM simulations with extensive conformational sampling. Indeed, inconsistent electrostatic treatments may generate unphysical behavior in QM/MM simulations. The immediate effects on the dynamics and energetics can be subtle, yet the influence on related properties can be of comparable importance as the problems discussed here. Such analyses will be presented soon.[102,109]

Acknowledgment. P.H.K. was supported by a stipend within the DFG Research Training Group GK-693 of the Paderborn Institute for Scientific Computation (PaSCo). M.H. was supported within the DFG research program 490. Some of the computations were carried out at the regional computing center in Cologne (RRZK). Q.C. is an Alfred P. Sloan Research Fellow. P.H.K. and M.H. acknowledge stimulating and helpful discussions with Drs. Marc Amkreutz and Christof Köhler.

Supporting Information Available: A figure containing critical geometrical parameters in the reactant, transition state, and product in the first proton transfer in methyl glyoxal synthase (MGS), which is similar to that for TIM in the main text. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Yang, W.; Lee, T.-S. *J. Chem. Phys.* **1995**, *103*, 5674.
(2) Liu, H.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Hermans, J.; Yang, W. *Proteins* **2001**, *44*, 484.
(3) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.
(4) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
(5) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260.
(6) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
(7) Aaqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523.
(8) Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298.
(9) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(10) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.
(11) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *32*, 425.
(12) Cui, Q.; Karplus, M. *Adv. Protein Chem.* **2003**, *66*, 315.
(13) Lennartz, C.; Schaefer, A.; Terstegen, F.; Thiel, W. *J. Phys. Chem. B* **2002**, *106*, 1758.
(14) Guallar, V.; Friesner, R. A. *J. Am. Chem. Soc.* **2004**, *126*, 8501.
(15) Cisneros, G. A.; Liu, H. Y.; Zhang, Y. K.; Yang, W. T. *J. Am. Chem. Soc.* **2003**, *125*, 10384.
(16) Himo, F.; Siegbahn, P. E. M. *Chem. Rev.* **2003**, *103*, 2421.
(17) Stubbe, J.; van der Donk, W. A. *Chem. Rev.* **1998**, *98*, 705.
(18) Wanko, M.; Hoffmann, M.; Strodel, P.; Koslowski, A.; Thiel, W.; Neese, F.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2005**, *109*, 3606.
(19) Hayashi, S.; Tajkhorshid, E.; Schulten, K. *Biophys. J.* **2003**, *85*, 1440.
(20) Holm, R. H.; Kennepohl, P.; Solomon, E. I. *Chem. Rev.* **1996**, *96*, 2239.
(21) Rees, D. C. *Annu. Rev. Biochem.* **2002**, *71*, 221.
(22) Bustamante, C.; Keller, D.; Oster, G. *Acc. Chem. Res.* **2001**, *34*, 412.
(23) Li, G.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 3342.
(24) Karplus, M.; Gao, Y. Q. *Curr. Opin. Struct. Biol.* **2004**, *14*, 250.
(25) Fromme, J. C.; Banerjee, A.; Verdine, G. L. *Curr. Opin. Struct. Biol.* **2004**, *14*, 43.
(26) Sancar, A.; Lindsey-Boltz, L. A.; Uensal-Kacmaz, K.; Linn, S. *Annu. Rev. Biochem.* **2004**, *73*, 39.
(27) Dinner, A. R.; Blackburn, G. M.; Karplus, M. *Nature* **2001**, *413*, 752.
(28) Michel, H.; Behr, J.; Harrenga, A.; Kannt, A. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 329.
(29) Wikstrom, M. *Curr. Opin. Struct. Biol.* **1998**, *8*, 480.
(30) Brzezinski, P. *Trends Biochem. Sci.* **2004**, *29*, 380.
(31) Riccardi, D.; Li, G.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 6467.
(32) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357.
(33) Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. *J. Mol. Biol.* **2003**, *327*, 549.
(34) Zhang, Y.; Lee, T.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46.
(35) Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290.

(36) Ferenczy, G. G.; Rivail, J. L.; Surjan, P. R.; Naray-Szabo, G. *J. Comput. Chem.* **1992**, *13*, 830.
(37) Thery, V.; Rinaldi, D.; Rivail, J. L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269.
(38) Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **1999**, *20*, 1468.
(39) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **2000**, *21*, 1442.
(40) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714.
(41) Assfeld, X.; Rivail, J. *Chem. Phys. Lett.* **1996**, *263*, 100.
(42) Ferre, N.; Assfeld, X.; Rivail, J. *J. Comput. Chem.* **2002**, *23*, 610.
(43) Assfeld, X.; Ferre, N.; Rivail, J. *Theor. Chem. Acc.* **2004**, *111*, 328.
(44) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720.
(45) Harrison, M. J.; Burton, N. A.; Hillier, I. H. *J. Am. Chem. Soc.* **1997**, *119*, 12285.
(46) Vasilyev, V. V. *THEOCHEM* **1994**, *110*, 129.
(47) Ferre, N.; Olivucci, M. *THEOCHEM* **2003**, *632*, 71.
(48) Waszkowycz, B.; Hillier, I. H.; Gensmantel, N.; Payling, D. W. *Perkin Trans. 2* **1991**, 225.
(49) Waszkowycz, B.; Hillier, I. H.; Gensmantel, N.; Payling, D. W. *Perkin Trans. 2* **1991**, 1819.
(50) Waszkowycz, B.; Hillier, I. H.; Gensmantel, N.; Payling, D. W. *Perkin 2* **1991**, 2025.
(51) Das, D.; Eurenius, K.; Billings, E.; Sherwood, P.; Chatfield, D.; Hodoscek, M.; Brooks, B. *J. Chem. Phys.* **2002**, *117*, 10534.
(52) Singh, U. C.; Kollmann, P. A. *J. Comput. Chem.* **1986**, *7*, 718.
(53) de Vries, A. H.; Sherwood, P.; Collins, S. J.; Rigby, A. M.; Rigutto, M.; Kramer, G. J. *J. Phys. Chem. B* **1999**, *103*, 6133.
(54) Sherwood, P.; de Vries, A. H.; Collins, S. J.; Greatbanks, S. P.; Burton, N. A.; Vincent, M. A.; Hillier, I. H. *Faraday* **1997**, *106*, 79.
(55) Amara, P.; Field, M. J. *Theor. Chem. Acc.* **2003**, *109*, 43.
(56) Elstner, M.; Frauenheim, T.; Suhai, S. *THEOCHEM* **2003**, *632*, 29.
(57) Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Comput. Chem.* **2003**, *24*, 565.
(58) Zhang, X.; Harrison, D. H. T.; Cui, Q. *J. Am. Chem. Soc.* **2002**, *124*, 14871.
(59) Formaneck, M. S.; Li, G.; Zhang, X.; Cui, Q. *J. Theor. Comput. Chem.* **2002**, *1*, 53.
(60) Cui, Q.; Elstner, M.; Karplus, M. *J. Phys. Chem. B* **2002**, *106*, 2721.
(61) Li, G.; Cui, Q. *J. Am. Chem. Soc.* **2003**, *125*, 15028.
(62) Li, G.; Zhang, X.; Cui, Q. *J. Phys. Chem. B* **2003**, *107*, 8643.
(63) Li, G.; Cui, Q. *J. Phys. Chem. B* **2003**, *107*, 14521.
(64) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569.
(65) Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 5454.
(66) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
(67) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. E.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(68) This energy term is not to be confused with the heat of formation of a proton. Rather, it is a consequence of the definition of the repulsive energy term in the SCC-DFTB formalism. A detailed explaination can be found in the supporting materials of ref 57.
(69) Elstner, M. Ph.D Thesis, Universität Paderborn, 1998.
(70) Fischer, S.; Karplus, M. *Chem. Phys. Lett.* **1992**, *194*, 252.
(71) Brooks, C. L. I.; Karplus, M. *J. Mol. Biol.* **1989**, *208*, 159.
(72) Cui, Q.; Karplus, M. *J. Am. Chem. Soc.* **2001**, *123*, 2284.
(73) Cui, Q.; Karplus, M. *J. Phys. Chem. B* **2002**, *106*, 1768.
(74) Cui, Q.; Karplus, M. *J. Am. Chem. Soc.* **2002**, *124*, 3093.
(75) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27.
(76) Stewart, J. J. P. *Rev. Comput. Chem.* **1990**, *1*, 45.
(77) Fersht, A. *Structure and Mechanism in Protein Science. Guide to Enzyme Catalysis and Protein Folding*; W. H. Freeman: New York, 1999.
(78) Davenport, R. C.; Bash, P. A.; Seaton, B. A.; Karplus, M.; Petsko, G. A.; Ringe, D. *Biochemistry* **1991**, *30*, 5821.
(79) Lolis, E.; Alber, T.; Davenport, R. C.; Rose, D.; Hartman, F. C.; Petsko, G. A. *Biochemistry* **1997**, *29*, 66609.
(80) Harris, T. K.; Abeygunawardana, C.; Mildvan, A. S. *Biochemistry* **1997**, *36*, 14661.
(81) Harris, T. K.; Cole, R. N.; Comer, F. I.; Mildvan, A. S. *Biochemistry* **1998**, *37*, 16828.
(82) Nickbarg, E. B.; Davenport, R. C.; Petsko, G. A.; Knowles, J. R. *Biochemistry* **1977**, *27*, 5948.
(83) Knowles, J. R.; Albery, W. J. *Acc. Chem. Res.* **1977**, *10*, 105.
(84) Albery, W. J.; Knowles, J. R. *Biochemistry* **1976**, *15*, 5627.
(85) Albery, W. J.; Knowles, J. R. *Biochemistry* **1976**, *15*, 5588.
(86) Maister, S. G.; Pett, C. P.; Albery, W. J.; Knowles, J. R. *Biochemistry* **1976**, *15*, 5607.
(87) Herlihy, J. M.; Maister, S. G.; Albery, W. J.; Knowles, J. R. *Biochemistry* **2000**, *15*, 5601.
(88) Perakyla, M.; Pakkanen, T. A. *Proteins* **1996**, *25*, 225.
(89) Alagona, G.; Desmeules, P.; Ghio, C.; Kollman, P. A. *J. Am. Chem. Soc.* **1984**, *106*, 3623.
(90) Alagona, G.; Ghio, C.; Kollman, P. A. *J. Mol. Biol.* **1986**, *191*, 23.
(91) Alagona, G.; Ghio, C.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 9855.
(92) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry* **1991**, *30*, 5826.
(93) Feierberg, I.; Luzhkov, V.; Aqvist, J. *J. Biol. Chem.* **2000**, *275*, 22657.
(94) Saadat, D.; Harrison, D. H. T. *Biochemistry* **2000**, *39*, 2950.
(95) Marks, G. T.; Harris, T. K.; Massiah, M. A.; Mildvan, A. S.; Harrison, D. H. T. *Biochemistry* **2001**, *40*, 6805.
(96) Liu, H.; Lu, Z.; Cisneros, G. A.; Yang, W. *J. Chem. Phys.* **2004**, *121*, 697.
(97) Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 3483.
(98) Cui, Q.; Karplus, M. *J. Chem. Phys.* **2000**, *112*, 1133.
(99) Cui, Q.; Karplus, M. *J. Phys. Chem. B* **2000**, *104*, 3721.
(100) Nonella, M.; Mathias, G.; Eichinger, M.; Tavan, P. *J. Phys. Chem. B* **2003**, *107*, 316.
(101) Klaehn, M.; Mathias, G.; Koetting, C.; Nonella, M.; Schlitter, J.; Gerwert, K.; Tavan, P. *J. Phys. Chem. A* **2004**, *108*, 6186.
(102) Schaefer, P.; Riccardi, D.; Cui, Q. *J. Chem. Phys.* **2005**, submitted for publication.
(103) Nam, K.; Gao, J.; York, D. M. *J. Chem. Theo. Comput.* **2005**, *1*, 2.
(104) Gregersen, B. A.; York, D. M. *J. Phys. Chem. B* **2005**, *109*, 536.
(105) Bash, P. A.; Ho, L. L.; MacKerell, A. D.; Levine, D.; Hallstrom, P. *Proc. Natl. Acad. Sci.* **1996**, *93*, 3698.
(106) Shurki, A.; Warshel, A. *Adv. Protein Chem.* **2003**, *66*, 249.
(107) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580.
(108) Simonson, T.; Archontis, G.; Karplus, M. *J. Phys. Chem. B* **1997**, *101*, 8349.
(109) Riccardi, D.; Schaefer, P.; Cui, Q. *J. Phys. Chem. B*, submitted for publication.
(110) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33.