Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches

Nathalie Reuter,[†,§] Annick Dejaegere,[*,‡,§] Bernard Maigret,[†] and Martin Karplus[*,§,||]

Laboratoire de chimie théorique, U.M.R. C.N.R.S. 7565, Faculté des Sciences, UHP, Nancy I, Vandoeuvre-lès-Nancy Cédex, France, Groupe de RMN, U.P.R. 9003, Ecole Supérieure de Biotechnologie de Strasbourg, Boulevard Sébastien Brant, pôle API, 67 400 Strasbourg-Illkirch, France, Laboratoire de chimie biophysique, Institut Le Bel, ULP, 67 000 Strasbourg, France, and Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138

Received: July 15, 1999; In Final Form: November 4, 1999

A major complication in hybrid QM/MM methods is the treatment of the frontier between the quantum part, describing the reactive region, and the classical part, describing the environment. Two approaches to this problem, the "link atom" method and the "local self-consistent field" (LSCF) formalism, are compared in this paper. For this purpose, the LSCF formalism has been introduced into the CHARMM program. A detailed description of the two approaches is presented. The results of semiempirical calculations of deprotonation enthalpies and proton affinities of propanol and a tripeptide with different treatments of the frontier bond are compared. Particular emphasis is placed on the effect of an external charge. It is shown that the choice of the QM/MM electronic interactions included in the frontier region is of considerable importance in determining the electron distribution of the QM region and the overall energy. The link atom and LSCF methods are generally of similar accuracy if care is taken in the choice of the frontier between the QM and MM regions. QM and QM/MM geometry optimizations of ethane and butane are also compared. The introduction of a link atom in the frontier bond is shown to lead to distortions of the internal coordinates unless the frontier bond is treated in a special way. A number of practical points concerning the choice of the frontier between the QM and MM regions are presented. It is not advisable to remove classical charges from the interactions with a subset of the quantum atoms, as this can introduce significant errors in the energy computations. The presence of a large charge on the classical atom involved in the QM/MM frontier also adversely influences the energy, especially with the LSCF method, and it is therefore advised to select classical frontier atoms with small charges. Charged atoms which are not directly bound to the QM frontier but which are in its proximity are also shown to be a source of errors, and it is advised to introduce warning messages in QM-MM codes when such a situation arises.

I. Introduction

Quantum mechanical approaches that can account for the making and breaking of chemical bonds have been widely used to probe the energetics and dynamics of chemical reactions. Most chemical reactions of interest take place in the condensed phase, where environmental effects can play a crucial role. Reactions in solution and in enzymes are important in this regard. The requirements of computer time limit the size of the systems that can be treated by quantum mechanical methods, even at the semiempirical level, so that the detailed analysis of complete enzymes or fluid systems of the required size is not yet possible. There is, therefore, a strong interest in developing methods that allow the study of the chemical reactivity of complex systems at a reasonable cost. An approach for large molecular systems is to partition them into different regions, which are modeled at different levels of approximation. Hybrid quantum mechanical/molecular mechanical (QM/MM) methods are a typical example of this type of approach,[1] and they are becoming increasingly popular for the study of chemical reactions in solution and in enzymes.[2,3] QM/MM methods combine a quantum mechanical treatment of the subset of atoms involved in the chemical reaction with a molecular mechanics description of the surroundings. The idea on which such a partitioning of the system is based is that the significant electron redistribution during the reaction is often limited to a small subset of the atoms, while the effects of the majority of the atoms in the system can be described adequately by a classical treatment of intermolecular interactions. Such a description fits in with chemical intuition, long standing ideas on the effect of solvent on chemical reactivity (e.g., cavity models[4]), and the focus on active site residues in the description of enzyme reaction mechanisms. In certain cases, where long-range charge redistribution is involved,[5] such a partitioning is not appropriate.

When using hybrid QM/MM methods for chemical reactions in solution, there is generally a natural separation between the reacting species, including possibly one or a few solvent molecules, which are treated quantum mechanically (QM), and the solvent, which can be described by molecular mechanics (MM) or continuum models. In enzymatic reactions, the partitioning into QM and the MM parts is not so straightforward and a separation must be defined between the active site amino acids for which a QM description is adopted and the other amino acids, which are described by molecular mechanics. This separation inevitably cuts covalent bonds and raises the question of how to best model the connection between the classical and quantum subsystems.

* Corresponding author. E-mail: annick@esbs.u-strasbg.fr; marci@brel.u-strasbg.fr.
† University of Nancy I.
‡ Ecole Supérieure de Biotechnologie de Strasbourg.
§ Institut Le Bel.
|| Harvard University.

10.1021/jp9924124 CCC: $19.00
Published on Web 02/03/2000

One commonly employed method is the link atom (LA) approach. The idea was introduced originally for purely quantum mechanical treatment in which a part of the molecule, presumed not to be important, was neglected and the unsaturated valences where satisfied by adding hydrogen atoms.[6,7] The first QM/MM formulation was introduced by Singh and Kollman.[8] A detailed description of the approach as implemented in the CHARMM[9] program has been presented by Field et al.[10] The link atom is included in the QM region, but its interaction with the MM atoms is adjustable and many schemes have been used. In the first description,[8] as well as in the original implementation of Field et al.,[10] interactions between links and MM atoms were not included, since the link atoms do not exist in the physical system under study. More recently, Hillier et al.[11] used the same scheme and, in addition, set the charges of the MM junction atom to zero. Eurenius et al.[12] did not include Coulombic interactions between the first MM group and all QM atoms, including link atoms. Vasilyev[13] et al. also used a link atom interacting with the MM charges but set the charges and van der Waals parameters of the MM atom of the frontier to zero. The added hydrogen atom along the frontier bond means that one must take care of the possible side effects induced in the electric field of this region. Removing the electrostatic interactions between the MM charges and the link atom, while the QM atom bonded to the link (QM frontier) feels the influence of these point charges, can lead to an unrealistic polarization of the bond between these two atoms. One must also make choices as to which classical force field terms should be maintained at the frontier. In some cases, QM and MM energy terms involving frontier and link atoms may be redundant.[12] In the original method implemented in CHARMM, all MM bonded terms are included if at least one MM atom is involved. Systems capped with methyl groups or pseudo-halogens[14] as links have also been described, but the hydrogen atom link remains the most common.

The local self-consistent field (LSCF) formalism is an alternative approach. It is based on the principle introduced by Warshel and Levitt[15] that a single hybrid sp^2 orbital with a single electron is included for each of the QM atoms at the junction. In the LSCF method, the QM/MM frontier bonds are described by strictly localized bond orbitals (SLBOs).[16,17] These localized bond orbitals, called frozen orbitals, are excluded from the SCF procedure and defined by their hybridization coefficients and their electron population. The two parameters are usually determined by quantum chemical calculations on small model systems and are assumed to be transferable to the system of interest. The method has been developed at the MNDO[18] semiempirical molecular orbital level (AM1 and PM3 methods) and is implemented in the GEOMOP[19] program. Since it avoids introducing extraneous hydrogen atoms in the system, the LSCF method is an attractive alternative to the link atom approach. However, no quantitative comparisons of the two methods have been made. A slightly different method, called generalized hybrid orbital method, also based on the principle of hybrid frontier orbitals, was proposed recently by Gao et al.[20]

Numerous QM/MM studies have been reported that use the LA method[11,13,21] and, more recently, the LSCF formalism.[22] The aim of the present work is to compare the two methods. For this purpose, we have connected the GEOMOP and CHARMM[9] programs so as to be able to use the LSCF formalism with the CHARMM[26] force field. Having both the SLBO and the link atom approaches available in the same program makes the comparison more straightforward. Moreover, the GEOMOP implementation was limited to the calculation of energy and forces on QM atoms under the influence of fixed MM charges, while the CHARMM implementation makes it possible to do minimization and dynamics of the quantum and classical regions simultaneously.

We first present a detailed description of the QM/MM Hamiltonian and the terms involved in the description of the frontier bond in the LSCF and LA methods. Calculations of properties for a series of small molecules using different treatments of the frontier bond between the quantum and the classical parts are then compared at the AM1/MNDO[18,23] semiempirical level. We examine the influence of an external charge in the neighborhood of the frontier bond by calculating deprotonation enthalpies and proton affinities of small polar systems. This is important for enzymatic reactions which often involve catalysis dominated by electrostatics.[2] It is shown that the effects of this polarizing field on the energy of the QM part depends on the description of the frontier bond and the choice of the QM/MM electronic interactions included in the boundary region. QM/MM results are compared to the results of full QM calculations using the same semiempirical Hamiltonian and are analyzed by studying the Mulliken charges of the QM atoms, which reflect the perturbations induced on the density matrix by an external charge.

Geometry optimizations and evaluation of rotational barriers for ethane and butane are also examined to analyze the structural perturbations introduced by the QM/MM description. The results obtained with the different QM/MM methods are compared with values computed by QM (AM1 and MP2/6-311G**) and MM methods, as well as to experimental values.

II. Methods

In this section, we describe the Hamiltonian used in hybrid quantum/classical methods. We then describe the local self-consistent field method, and the link atom method for treating the quantum/classical boundary. In the present description and the test calculations, the focus is on semiempirical quantum Hamiltonians.

II.1. The Hamiltonian. In the QM/MM formalism, the Hamiltonian is the sum of terms representing the QM region, the MM region, and the interaction between them; that is,

$$\hat{H}_{\text{eff}} = \hat{H}_{\text{QM/QM}} + \hat{H}_{\text{MM/MM}} + \hat{H}_{\text{QM/MM}} \tag{1a}$$

We consider each of these terms below. Given $\hat{H}_{\text{eff}}$, the energy of the system has the form

$$E = \langle\Phi|\hat{H}_{\text{eff}}|\Phi\rangle = \langle\Phi|\hat{H}_{\text{QM/QM}}|\Phi\rangle + \langle\Phi|\hat{H}_{\text{QM/MM}}|\Phi\rangle + E_{\text{MM/MM}} \tag{1b}$$

where Φ is the wave function describing the QM atoms.

II.1.a. $\hat{H}_{QM/QM}$. $\hat{H}_{\text{QM/QM}}$ is the Hamiltonian describing the quantum mechanical particles (nuclei and electrons) of the system and their interactions with each other. We outline here the Hartree−Fock (HF) LCAO-MO formulation at the semiempirical AM1 level.[18,23] Other quantum mechanical methods (such as DFT, ab initio LCAO-MO) can be used,[24] but semiempirical methods are studied here. They are most often employed in applications to large systems, such as the active sites of enzymes, because they allow calculations on a relatively large number of atoms (up to 100) with a reasonable computation time. The quantum term $H_{\text{QM/QM}}$ does not include any molecular mechanics interactions (i.e., valence force field or nonbonded terms) between QM atoms. The expression for $H_{\text{QM/QM}}$ is not discussed further here; a detailed description of the AM1 formulation is given in refs 18 and 23, see also ref 25.

II.1.b. $H_{MM/MM}$. $H_{MM/MM}$ is a molecular mechanics Hamiltonian which depends solely on the positions of the classical atoms. In the present study, the standard CHARMM22[26] all-atom force field is used.

II.1.c. $H_{QM/MM}$. The interaction Hamiltonian $H_{QM/MM}$ describes the interactions between the quantum and molecular mechanics atoms in the system. In the present formulation, $H_{QM/MM}$ is taken to have the form (in atomic units)[10]

$$\hat{H}_{QM/MM} = -\sum_{i,M}\frac{q_M}{r_{iM}} + \sum_{\alpha,M}\frac{Z_\alpha q_M}{R_{\alpha M}} + \sum_{\alpha,M}\left\{\frac{A_{\alpha M}}{R_{\alpha M}^{12}} - \frac{B_{\alpha M}}{R_{\alpha M}^{6}}\right\} + \hat{H}_{QM/MM}^{\text{int.coord.}} \quad (2)$$

where i and α represents the QM electrons and nuclei, respectively, and M corresponds to the MM atoms. The first term, the electrostatic interaction between an MM atom of charge q_M and electron i, is included in the SCF calculation, whereas the other terms do not depend on the electronic coordinates. The second term describes the Coulombic interactions between the classical MM charges and the quantum nuclei, and the third one the van der Waals interactions between the QM and MM atoms. The fourth term represents the bonded interaction terms introduced to connect the QM and MM systems; they are present only if the boundary between the QM and MM systems is within a molecule.

Electronic QM/MM Interaction Term. In MNDO-type semiempirical methods, including AM1, the electrostatic interactions between QM and MM regions are evaluated with a point-charge model, similar to that used for electron−electron repulsion integrals. The integrals are expanded in terms of multipole−multipole interactions.[25] For an sp^3 basis set, only monopoles, dipoles, and quadrupoles are included. Thus, the ss orbital charge distribution is represented by a single point charge at the nucleus and an sp_α charge distribution (where α is *x*, *y*, or *z*) by a dipole consisting of two equal charges located on opposite sides of the nucleus along the α-axis. Where the orbital is made up of two p functions, the charge distribution is represented by a quadrupole of four equal charges in a square configuration for $p_\alpha p_\beta$ and a linear quadrupole configuration for $p_\alpha p_\alpha$. The distances of the charges from the nucleus are related to the exponent of the orbitals. For more details, see ref 25.

The one electron integral between two atomic orbitals is then calculated as the sum of the interactions between the multipoles representing the charge distributions. Considering the point charge on the MM atom to reside in an s orbital for consistency, the one-electron integrals between the partial charge q_M on the MM atom and the basis function μ and ν on the QM atom have the form

$$I_{\mu\nu} = -q_M(\mu_Q\nu_Q|s_M s_M) \quad (3)$$

All of the terms contributing to $\langle\Phi|-\Sigma_{i,M}(q_M/r_{iM}) + \Sigma_{\alpha,M}(Z_\alpha q_M/R_{\alpha,M})|\Phi\rangle$ reduce to one-electron integrals of this type plus empirical terms described below.

Since the s orbitals are represented as point charges at the nucleus independent of their exponent, the interaction between the classical charges and the quantum nuclei is represented by the semiempirical core−core repulsion function, which takes the form

$$E_{QM/MM}^{\text{charge/core}} = \sum_{\alpha,M}\frac{Z_\alpha q_M}{R_{\alpha M}} = \sum_{Q,M} Z_Q q_M(s_Q s_Q|s_M s_M)(1 + f(R_{QM})) \quad (4)$$

where Z_Q is the core charge of the quantum atom Q, q_M the charge of the MM atom, and the function *f* depends on the interatomic distance R_{QM}. The s_Q and s_M are s orbitals associated with the nuclei and molecular mechanics atom, respectively. The core−core repulsion function $f(R_{QM})$ used in the QM/MM interaction has the same functional form as in semiempirical methods. For example, in MNDO the core−core interaction between atoms A and B is expressed as[18,25]

$$E^{\text{core/core}}(\text{MNDO}) = \sum_{A,B} Z_A Z_B(s_A s_A|s_B s_B)(1 + e^{-\alpha_A R_{AB}} + e^{-\alpha_B R_{AB}}) \quad (5)$$

In AM1 and PM3, an extra term is added to reduce the excessive core−core repulsion just outside bonding distances.[23] The core−core parameters used for the molecular mechanics atoms (i.e., α_M in MNDO) are different from those used for the quantum atoms (i.e., a nitrogen atom, for example, will have different core−core parameters if it belongs to the quantum or classical region). We used the parameters defined in the CHARMM QM/MM code and described in ref 10.

Molecular Mechanics van der Waals Interactions and Bonded Terms. The van der Waals interaction between quantum and molecular mechanics atoms is described by the classical force field. Nonbonded van der Waals interactions between QM and MM atoms are treated using the standard CHARMM22 force field convention.[9,26] It is sometimes necessary to adjust vdW parameters in the context of QM/MM interactions.[27] Nonbonded interactions in nonpolarizable force fields are represented as a sum of vdW and Coulombic terms, while the parameters (partial charges, vdW radii and well depth) are usually obtained by fitting to Hartree−Fock quantum chemical interaction energies on model systems. Hartree−Fock calculations include mutual polarization of the interacting partners, which is therefore implicitly introduced in the force field parameters. In hybrid QM/MM calculations with the CHARMM22 force field, the MM part is not polarized, but the quantum region is, so that some of the energy terms implicit in the force field are explicitly calculated. It is therefore reasonable to expect that in some cases an adjustment of the classical force field is necessary, but these adjustments are usually small and did not appear necessary for the particular examples considered here.

The classical interactions between QM and MM bonded atoms, the last term of eq 2, are part of the description of the frontier; they are discussed below.

II.2. The Frontier Bond in the Case of an Intramolecular QM/MM Boundary. In many practical applications (e.g., in enzymes), the boundary between the classical and quantum region cuts through covalent bonds. One has therefore to consider how best to model the frontier bonds that connect classical and quantum atoms. In what follows, we use the following conventions (see Figure 1): X is the last QM atom and Y is the first MM atom. The bond between X and Y (i.e., the frontier bond) is assumed to be single and both X and Y are atoms of the first row of the periodic table. Multiple bonds are more sensitive to polarization effects and are therefore not appropriate for QM/MM boundaries; as to elements in other rows of the periodic table, the principles are the same but localized orbitals descriptions would be more complex because of the involvement of higher atomic orbitals.

II.2.a. Local Self-Consistent Field (LSCF) Approach. The LSCF method represents the electronic density along the frontier bond by a frozen atomic orbital, which has a preset geometry and electronic population. This strictly localized bond orbital

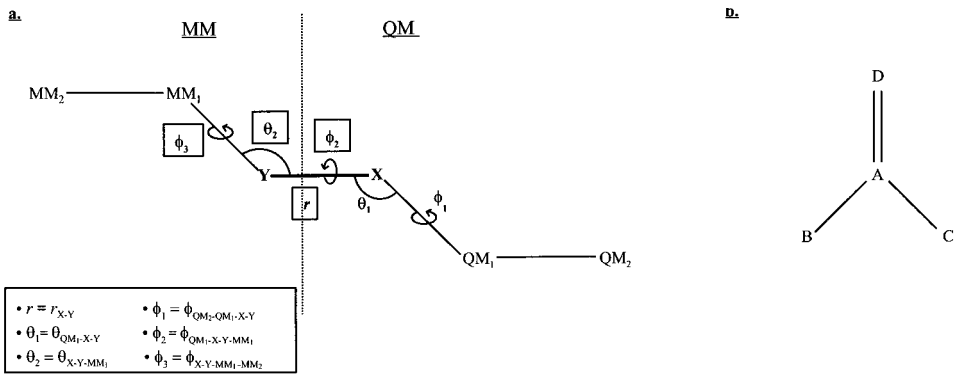

Figure 1. Schematic of the frontier bond, the frontier atoms, and the MM force field bonded terms considered at the QM−MM boundary.

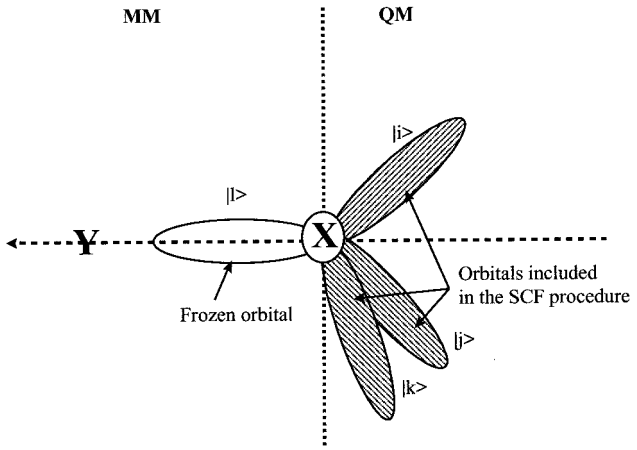

Figure 2. Partition between the QM and MM regions in the LSCF method.

is not included in the SCF procedure of the quantum fragment but acts as a frozen charge density, analogous to the charges on the MM atoms, which can polarize the QM portion of the system (Figure 2). Specifically, a set of four orthogonal hybrid orbitals appropriate for a first row atom is defined on the quantum atom X. The hybrid orbital, pointing toward the classical frontier atom Y, is expressed as

$$|1\rangle = a_{11}|s\rangle + a_{12}|x\rangle + a_{13}|y\rangle + a_{14}|z\rangle \tag{6}$$

where $|s\rangle$, $|x\rangle$, $|y\rangle$, and $|z\rangle$ are the s and p valence orbitals. The three hybrid orbitals $|i\rangle$, $|j\rangle$, $|k\rangle$ are included in the basis set of the quantum calculation on the quantum subsystem, whereas $|l\rangle$ is excluded from the self-consistent procedure. It is called the frozen orbital. Since we are using NDDO (neglect of differential diatomic overlap) type methods in the LCAO (linear combination of atomic orbitals) approximation, the hybrid orbitals are orthogonal to each other and to the other atomic orbitals of the quantum subsystem.

The transformation between the AO's $|s\rangle$, $|x\rangle$, $|y\rangle$, and $|z\rangle$ and the hybrid orbitals $|i\rangle$, $|j\rangle$, $|k\rangle$, and $|l\rangle$ can be expressed in the form

$$\begin{bmatrix}|i\rangle\\|j\rangle\\|k\rangle\\|l\rangle\end{bmatrix} = [H][R]\begin{bmatrix}|s\rangle\\|x\rangle\\|y\rangle\\|z\rangle\end{bmatrix} \tag{7}$$

First, a rotation [R] of the z axis bring it in coincidence with the $\overrightarrow{XY}$ direction, leading to a new set of orbitals $|s'\rangle$, $|x'\rangle$, $|y'\rangle$, and $|z'\rangle$, i.e.,

$$[R] = \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta\\ 0 & -\sin\phi & \cos\phi & 0\\ 0 & \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta\end{bmatrix} \tag{8}$$

The angles θ and ϕ are the spherical polar coordinates of Y in a basis centered on atom X.[28] A linear combination of the transformed orbital is then constructed to obtain the hybrid orbitals $|i\rangle$, $|j\rangle$, $|k\rangle$, and $|l\rangle$. With the z axis coincident with $\overrightarrow{XY}$, $|l\rangle$ can be expressed as a combination of $|s'\rangle$ and $|z'\rangle$:

$$|l\rangle = a_{11}|s'\rangle + a_{14}|z'\rangle \tag{9}$$

where a_{l4} is related to the s/p ratio of $|l\rangle$ by

$$a_{14} = (1 - a_{11}^2)^{1/2} \tag{10}$$

The orbitals $|j\rangle$ and $|k\rangle$ are identified, respectively, as $|x\rangle$ and $|y\rangle$. To satisfy the normalization and orthogonality conditions, $|i\rangle$ is expressed as

$$|i\rangle = a_{14}|s'\rangle - a_{11}|z'\rangle \tag{11}$$

These transformations lead to the matrix [H],

$$[H] = \begin{bmatrix}a_{14} & 0 & 0 & -a_{11}\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ a_{11} & 0 & 0 & a_{14}\end{bmatrix} \tag{12}$$

The frozen orbital is thus completely defined by its direction and its s/p ratio a_{l1}. In addition, the electronic density, P_{ll}, is specified. To determine the parameters P_{ll} and a_{l1}, a localized orbital computed from a model subsystem can be used. For example, for the C^{γ} atom of the $C^{\beta}-C^{\gamma}$ bond in the glutamic acid residue of the gly-glu-gly tripeptide, the values obtained for P_{ll} and a_{l1} are 0.98 and 0.43, respectively. P_{ll} and a_{l1} can also be treated as predefined parameters; for example, typically, for sp^3 hybridization, a_{l1} is expected to be around 0.5 and P_{ll} close to 1 for a nonpolarized bond.[16] Both P_{ll} and a_{l1} are entered as input parameters of the QM-MM calculation. The choice made for the electronic population P_{ll} influences the charge of the QM part, by introducing a partial electronic charge ($1 - P_{ll}$) at the frontier. If P_{ll} is different from one, fractional electronic charges are therefore introduced on the QM part.

II.2.b. Link Atom Method. The link atoms, generally hydrogens, are included in the QM system and treated by the SCF procedure. If there is one link hydrogen atom, as in most cases, one (hydrogen) nucleus and one electron are added to the QM system. The SCF calculations with the link atom thus include one more proton and two additional electrons, when compared

to SCF calculations performed with the LSCF method, since in the LSCF method, the electron from the localized orbital is excluded from the SCF. Unlike the LSCF approach, the QM system always has an integral charge. Since the link atoms are "artificial" (i.e., they are not part of the actual chemical system under study), there is considerable flexibility in the choice of their interactions; the appropriate criterion to use is to obtain the best approximation to the full quantum system. In the original version of the method,[8,10] the terms $I_{\mu\nu}$ and $E_{QM/MM}^{core/charge}$ (eqs 3 and 4) were set to zero when Q was a link atom so that there were no direct electrostatic interaction terms between the link and the MM charges. The link atom however interacted with the quantum subfragment, which was polarized by the MM charges so that it still experienced the influence of the MM atoms, albeit indirectly. Variants of the original method, where the link atoms interact with all or a subset of the MM atoms, have been used. For example, it has been advised to remove the interactions between the link atom and the classical atoms of the first electrostatic MM neutral group (MM frontier atom plus atoms bonded to it to maintain the neutrality of the system).[29] Hillier et al.[11] removed the charge of the first MM group and scaled the other MM charges to ensure the neutrality of the MM environment.

In what follows, we consider two limiting cases for the link atom interactions. In the first, no direct electrostatic interaction term is calculated between the atom and the MM charges, as originally proposed (this is called a "QQ" link), and in the second, the link atom interacts with all charges (the "HQ" link). HQ and QQ refer to the atom types used in the CHARMM program for the link atoms. When a link atom interacts with a subset of the classical charges, it is stated explicitly in the text.

II.2.c. Classical Bonded Terms. In treating the frontier, one must also decide which molecular mechanical force field terms to retain. Figure 1 summarizes the bonds, valence angles, and dihedral angles which involve both QM and MM atoms.

Most authors[10,30] introduce classical bonded force field terms when at least one MM atom is involved. Referring to Figure 1, classical energy terms would be computed for the X−Y bond (r), the valence angles θ_1, θ_2, and the dihedrals ϕ_1, ϕ_2, and ϕ_3. If present, improper dihedrals involving both QM and MM atoms would also be included in the potential energy function. Eurenius et al.[12] do not include all of these classical bonded terms to avoid duplicating interactions computed quantum mechanically. In their approach, classical bond terms between one QM atom and one MM atom are kept (to maintain bonds across the interface), and also angles and dihedral angles when at least one "central atom" is MM. For a valence angle A−B−C, there is one "central atom", B; for a dihedral between four atoms A−B−C−D, there are two "central atoms", B and C. Improper dihedral terms for A−B−C−D angles are included only when at least A and D are classical (see Figure 1b). For example, in a carbonyl group, the improper dihedral is defined by C−R[1]−R[2]−O where R[1] and R[2] are the first atoms of substituents of the carbonyl group. This improper dihedral is taken into account only if C and O are classical atoms. All the other bonded interactions are assumed to be described by the quantum Hamiltonian. Figure 1 summarizes the classical bonded terms that would be computed at a typical QM/MM interface, in general, and according to the Eurenius et al.[12] formulation (circled terms only).

When a link atom is present, additional terms involving the link atom and its interactions with the MM atoms must be considered. For clarity, Figure 1 does not show the presence of a link between atoms X and Y. When there is a link atom, one must define a valence angle between X, Y, and the link and constrain it to be equal to zero during the computations. If not, the interactions between the link and the MM atoms lead to a nonlinear X−link−Y bond. Constraints can also be introduced on the bond length between X and Y, and sometimes X and the link atom (see Results section for a detailed discussion). The link atom, whatever its type, is initially placed at 1 Å from the last QM atom X.

II.3. Computational Details. *II.3.a. Programs*. The calculations have been done with the CHARMM[9] program in which we implemented a new module to perform QM/MM calculations using the LSCF formalism. The GEOMOP program has been interfaced with CHARMM in a way that allows the user to select in the CHARMM input whether the QM/MM frontier will be described by a link atom or by a localized orbital.

If the link atom option is selected, the quantum calculations are carried out by MOPAC[31] (in CHARMM), whereas if the LSCF frontier orbital is selected, GEOMOP is called by the main routine of CHARMM and performs the quantum computations. Both programs have the same semiempirical parameters, the only difference being that MOPAC uses basis of Gaussians functions (STO6-G), whereas GEOMOP uses Slater functions. This is not a problem; it leads, for example, to differences of less than 0.003% in the AM1 energy of formation of a water molecule.

In practice, CHARMM calculates $E_{MM/MM}$ as well as the corresponding forces and calls GEOMOP to carry out quantum calculations ($E_{QM/MM}$ and $E_{QM/QM}$ and corresponding forces). $E_{QM/MM}$ is evaluated using the CHARMM22 charges and the MM atom coordinates which have been read by the main routine of CHARMM in the input files. The MM charges and coordinates are transmitted to GEOMOP at each call. The frontier atoms are automatically found and the construction of the frontier orbitals is done once, at the beginning of the simulation. The calls to GEOMOP are clear to the user during the simulation.

II.3.b. Classical Force Field. The force field used was CHARMM22[26] and the MM atoms are treated in a standard fashion.

II.3.c. Localization of Molecular Orbitals for the Determination of P_{ll} and a_{11}. The localization calculations are performed by the GEOMOS[32] program, based on the Edminston and Ruedenberg[33] criterion; that is, the molecular orbitals are calculated so as to maximize the interaction energy between electrons of the same orbital.

II.3.d. Computation of Forces. The forces related to the MM terms (bonds, angles, van der Waals, Coulombic) are computed in the usual way by the CHARMM[9] program. The QM and QM/MM forces are calculated by the quantum subprogram, MOPAC[31] in the case of the link atom method and GEOMOP[19] in the case of the LSCF formalism. In the latter method, the first Cartesian derivatives include also the derivatives of the matrix which transforms the atomic orbitals of the QM frontier atoms into the set of hybrid orbitals.[17] In both QM modules, the forces are calculated analytically; with MOPAC, they can also be calculated numerically. Specifically, the QM program calculates the forces between the QM atoms as well as the forces on the QM atoms due to the MM atoms. The forces exerted on the MM atoms by the QM atoms are simply the opposite of the forces exerted by the MM atoms on the QM atoms. These are calculated by the QM program and added to the purely classical forces calculated by the MM program.

II.3.e. Geometry Optimizations. The purely MM and also the QM/MM geometry optimizations were performed using suc-

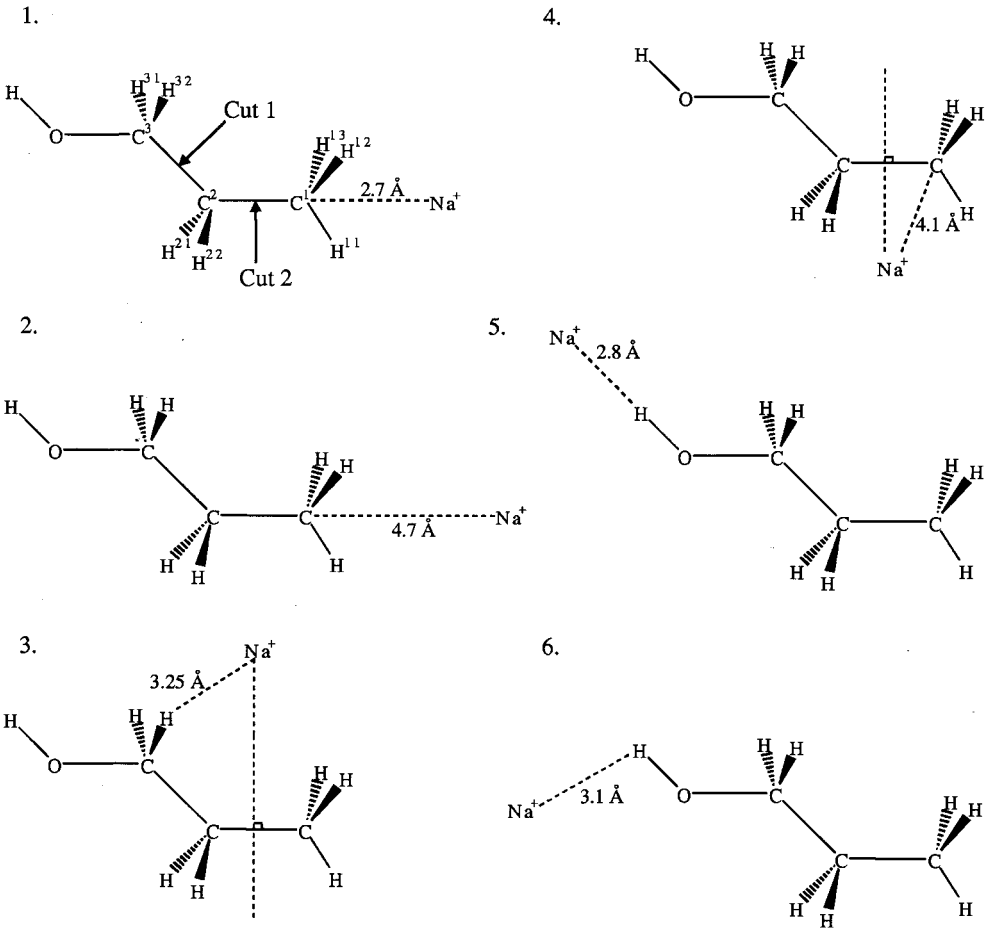

Figure 3. Propanol. The two positions used for the frontier bond are designed by arrows (Figure 3.1) and the positions of the sodium ion with respect to the propanol are indicated. The atoms are labeled in Figure 3.1. Full coordinates are given as Supporting Information.

cessively the steepest descent and the conjugate gradient algorithms in CHARMM, until the gradient is less than 10^{-4} kcal/(mol Å). The Cartesian second derivatives of the frontier orbitals are not implemented in GEOMOP, so it was not possible to use the Newton−Raphson type algorithm.

The rotational barriers of the butane molecule are also calculated. The staggered conformer is obtained by constraining the dihedral $H^{11}-C^1-C^2-H^{21}$ to 180° and minimizing in the presence of this constraint.

II.3.f. Ab Initio Calculations. To evaluate the accuracy of AM1 geometries (paragraph III.3) and the accuracy of AM1/MM geometries, ab initio computations on the ethane and butane molecules have been performed with GAUSSIAN94,[34] at the MP2 level with a 6-311G** basis set.

III. Results and Discussion

In this section, we present QM/MM data on several model systems to compare and test the LSCF and link atom boundary representations. The results are compared with full QM calculations performed at the same quantum level. We compute proton affinities and deprotonation enthalpies for propanol and a tripeptide gly-glu-gly in the presence of a sodium ion, which is represented as a classical charge. Different positions of the sodium ion around the QM/MM system are compared. The Mulliken charges obtained with the QM and QM/MM calculations are also compared since they are important for determining the interactions with the surrounding environment. The calculations test the performance of the various boundary descriptions in the presence of strong polarizing fields, such as could be encountered, for example, in QM/MM calculations of an enzyme; that is, charged side chains in the neighborhood of the substrate often play an important role in catalysis.[2] This type of effect has not been considered in other tests of QM/MM methods.[10] Finally, QM and QM/MM geometry optimizations of ethane and butane as well as the rotational barrier of butane are compared.

III.1. Propanol: Deprotonation Enthalpies (DE) and Proton Affinities (PA) in the Presence of an External Charge. Propanol was chosen as a test molecule because it is small enough to allow rapid quantum calculations on the entire molecule, yet it allows testing of two positions of the QM/MM boundary with respect to the hydroxyl group, which is the site of protonation and deprotonation. The division of the propanol molecule into QM and MM moieties is indicated in Figure 3; the arrows design the two possible positions for the frontier bond. The first is along the C^2-C^3 bond leading to a minimal QM region, C^3H_2OH. The second is along the C^1-C^2 bond, which yields a larger quantum ($C^2H_2C^3H_2OH$) and a smaller classical (C^1H_3) region. In this series of tests, the classical partial charges on the aliphatic carbons and hydrogen atoms are set to zero. The only classical charge in the system is, therefore, the sodium ion so as to allow a clear determination of its influence on the QM/MM results. The various positions of the sodium ion that were used in the DE and PA calculations are represented in Figure 3. The structures used for propanol (C_3H_7OH), propanolate ($C_3H_7O^-$), and the conjugated acid form ($C_3H_7OH_2^+$) are from full QM optimizations at the AM1 level (coordinates are given as Supporting Information). The structures are used without modification in all subsequent QM and QM/MM calculations in the presence of sodium.

TABLE 1: Comparison of Full QM (AM1) and QM-MM Values for the Proton Affinities and Deprotonation Enthalpies of Propanol[b]

	AM1	LSCF cut 1	LSCF cut 2	HQ cut 1	HQ cut 2	QQ cut 1	QQ cut 2
(a) Proton Affinities							
without sodium	172.4	165.0	167.6	167.7	171.6	167.7	171.6
Na (1)	114.8	112.1	114.1	114.1	116.1	120.0	123.1
Na (2)	129.9	124.5	126.9	127.0	130.0	131.2	134.6
Na (3)	98.0	88.0	92.3	91.4	96.7	96.7	101.4
Na (4)	116.9	112.6	113.5	114.5	116.7	120.5	122.1
Na (5)	87.1	76.6	80.3	79.8	85.6	84.4	89.4
Na (6)	90.2	80.1	83.4	83.3	88.8	87.8	95.2
RMS[a]	0	7.84	4.95	5.16	1.13	3.05	5.16
(b) Deprotonation Enthalpies							
without sodium	390.6	397.2	391.2	394.2	391.9	394.2	391.9
Na (1)	330.3	343.4	335.6	338.4	334.0	349.4	342.4
Na (2)	346.8	356.2	349.3	352.2	349.1	360.1	354.4
Na (3)	311.8	314.8	311.4	313.3	312.4	322.8	317.7
Na (4)	338.9	345.3	336.0	340.1	335.9	351.1	342.3
Na (5)	301.3	302.4	300.2	302.6	301.8	311.1	306.0
Na (6)	308.9	306.4	308.0	310.7	309.6	318.7	313.4
RMS[a]	0	8.23	2.61	4.82	2.01	13.84	7.66

[a] RMS refers to the root mean square difference between QM and QM-MM values (RMS $= \sqrt{(1/n)\Sigma_{i=1,n}(x_{QM/MM}-x_{QM})^2}$ where $x_{QM/MM}$ is the QM/MM value while x_{QM} is the corresponding value from QM. The sum extends over values Na (1) to Na (6). [b] The values are given in kcal/mol. The positions of the sodium are as described in Figure 3, i.e., Na (1) refers to Figure 3.1, Na (2) to Figure 3.2 etc. For each case, the position used for the sodium ion is the same in the neutral and protonated (PA) and neutral and anionic (DP) forms of propanol. See Figure 3 for definition of cut 1 and cut 2.

III.1.a. Energy Values. We computed the enthalpies of formation of $C_3H_7O^-$, C_3H_7OH, and $C_3H_7OH_2^+$ for each position of the sodium ion (cf. Figure 3). The *gas-phase* proton affinities are computed according to

$$PA = \Delta H_f(H^+) + \Delta H_f(B) - \Delta H_f(HB^+) \qquad (18a)$$

where $CH_3CH_2CH_2OH_2^+$ is the conjugated acid form (HB^+) and $CH_3CH_2CH_2OH$ is the base (B); that is, the reaction in the presence of Na^+ can be written

$$Na^+ + CH_3CH_2CH_2OH_2^+ = Na^+ + CH_3CH_2CH_2OH + H^+ \qquad (18b)$$

The deprotonation enthalpies are computed according to

$$DE = \Delta H_f(H^+) + \Delta H_f(B^-) - \Delta H_f(HB) \qquad (19)$$

where the acid form is $CH_3CH_2CH_2OH$ (HB) and $CH_3CH_2CH_2O^-$ is the conjugated base (B^-). The heat of formation of the proton, $\Delta H_f(H^+)$ is taken as the experimental value of 367.2 kcal mol^{-1}.[35]

The values of the proton affinity and deprotonation enthalpy obtained with the various hybrid QM/MM descriptions of propanol are compared to those obtained with a full QM computation on the propanol molecule; the Na^+ ion is always treated as an external MM charge. These results are presented in Table 1. It must be noted that semiempirical Hamiltonians are parametrized to reproduce experimental enthalpies of formation. Semiempirical energies therefore implicitly contain thermal contributions and can be properly compared to enthalpies. In the mixed semiempirical QM/MM calculations, we did not calculate thermal contributions for the classical part. These contributions are expected to cancel to a large extent in relative enthalpies, such as calculated here, and should not introduce any significant difference between QM/MM and purely QM results.

The AM1 values of the deprotonation enthalpy and proton affinity of propanol in the absence of the sodium charge are 390.6 and 172.4 kcal/mol, respectively. It can be seen from Table 1 that the contribution of the sodium ion to the energetics of the acid/base equilibrium is large, up to 90 kcal/mol on both DE and PA. In the presence of sodium, the DE varies between 301 and 347 kcal/mol and the PA between 87 and 130 kcal/mol. The positive sodium charge facilitates the deprotonation by stabilizing the anionic propanolate. The AM1 value for the enthalpy of formation of propanolate is −45 kcal/mol in the absence of sodium, while in the presence of sodium, the values are between −88 and −130 kcal/mol. The enthalpy of formation of propanol is changed by less than 5 kcal/mol in the presence of Na^+. The enthalpy of formation of the cationic CH_3CH_2-$CH_2OH_2^+$ is 128 kcal/mol in the absence of Na^+, and between 170 and 216 kcal/mol in the presence of the ion. Thus, the presence of Na^+ reduces the proton affinity of propanol.

Size of the QM Part. The results reported in Table 1 and in Figure 4 show that the position of the frontier bond significantly influences the quality of the QM/MM results. For HQ and LSCF, as expected, the agreement between QM and QM/MM is poorer when the frontier is positioned between the C^2 and C^3 carbons (i.e., cut 1, closer to the hydroxyl group). In the absence of sodium, there is no external classical charge and the HQ and QQ methods give the same results, but in the presence of sodium, the behavior of the QQ link is generally less reliable than the other methods and it is not systematically better when the QM fragment is larger. The QQ link is analyzed further in the next section.

With the smaller quantum region, the rms error between QM and QM/MM results, in the presence of sodium, is 8 kcal/mol for both PA and DE with the LSCF method, while it is 5 kcal/mol with HQ. Even though the rms errors are similar for PA and DE, Figure 4 show that the correlation between QM/MM and QM values is poorer for DE. This is most likely due to a smaller electronic delocalization on the aliphatic chain in the cationic $C_3H_7OH_2^+$ than in the anionic $C_3H_7O^-$ (see below). Overall, the LSCF method appears more sensitive to the position of the boundary than the HQ link method. The position of the frontier bond must thus be carefully chosen.

Comparison of the Three Methods. The results presented in Figure 4c and Table 1 clearly show that a link atom that does not interact with the outside charges (link atom "QQ") leads to significantly larger discrepancies between QM and QM/MM results: the correlation between the hybrid and full QM calculations can be poor and the rms difference between the two sets of values is 8 kcal/mol when the frontier bond is placed further from the hydroxyl group (i.e., between C^1 and C^2). Such problems with the original link atom method have been observed by others.[11,12,27] The problem is that the electric potential created by the classical charges (in this case, the sodium ion) is set to zero at the position of the link since the electrostatic QM/MM interaction terms (cf. eq 2) are equal to zero when i and α represent the electron and nucleus of a link atom, while the carbon atom directly bonded to the link experiences the MM potential. This creates an electric field along the carbon−link atom bond, and leads to distortions of the charge distribution (see below). Although the charge perturbation is generally localized on these two atoms (frontier carbon and link atom), the results presented in Table 1 and Figure 4c show that it also affects the relative energies of propanol and its conjugated base

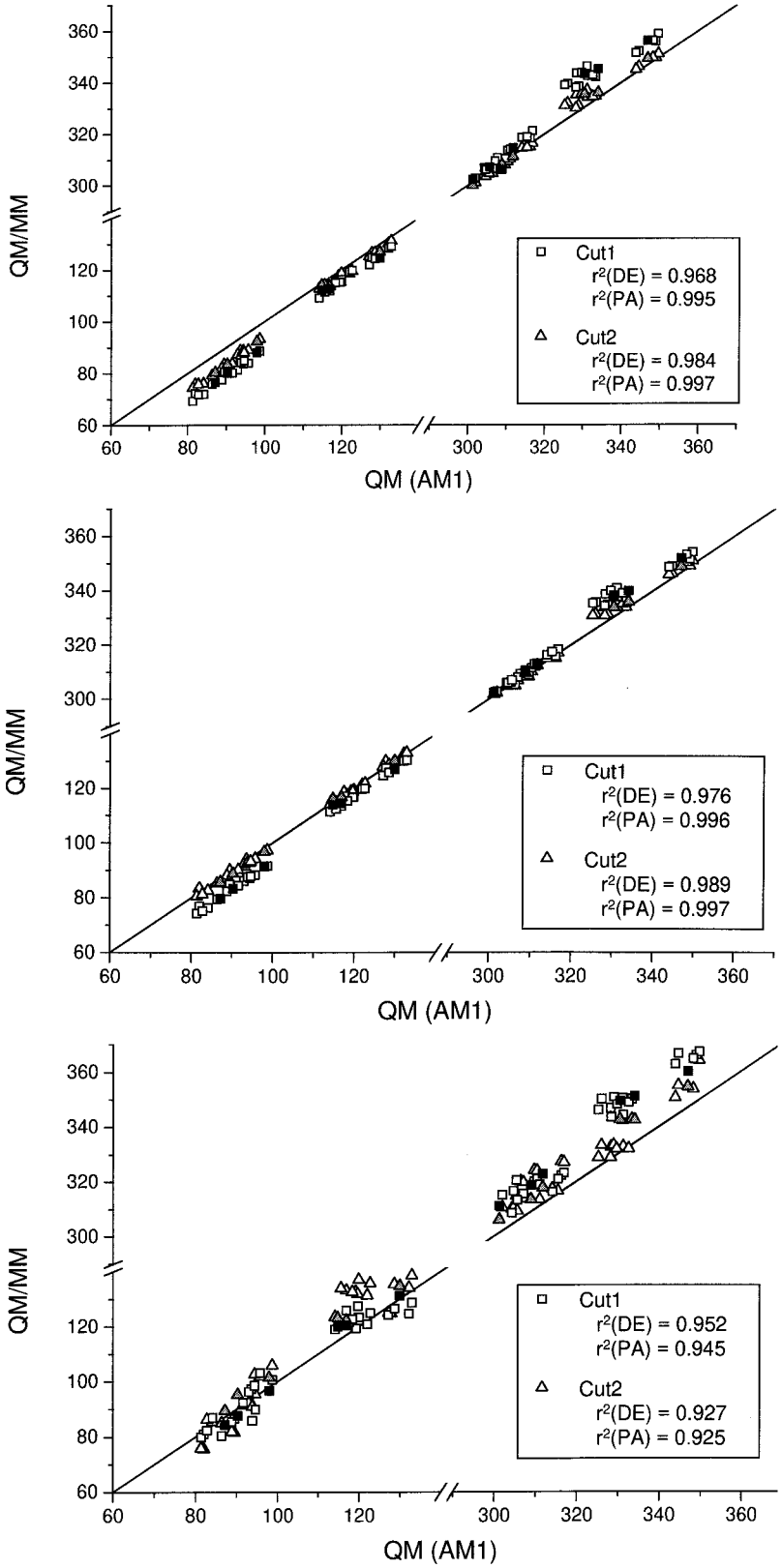

Figure 4. Comparison between QM/MM (AM1/CHARMM22) and full QM (AM1) deprotonation enthalpies and proton affinities of propanol (units, kcal/mol). LSCF values are displayed in part a, HQ link values in part b, and QQ link values in part c. PA values are between 80 and 140 kcal/mol and DP values between 300 and 360 kcal/mol. There are no data points between 140 and 290 kcal/mol, and these values are omitted from the *X* and *Y* axes. The $X = Y$ axis is represented as a solid line, the values obtained with cut 1 (cf. Figure 3.1) as squares and with cut 2 (cf. Figure 3.1) as triangles. Filled symbols indicate the data reported in Table 1, where the sodium ion is in the same position in the protonated and neutral (PA data) or neutral and anionic (DP data) systems. Open symbols correspond to data where the sodium ion is in a different position between the neutral and protonated, or neutral and anionic, forms of propanol; that is, these data correspond to a situation where the sodium ion would have moved during the course of the protonation/deprotonation reaction. By combining the six positions of the sodium from Figure 3, 36 values of PA and 36 values of DE are thus obtained for each boundary method. Generalized correlation coefficients (r^2) are given for each position of the frontier, each method, and each type of energy calculation (deprotonation enthalpy or proton affinity). Each correlation coefficient is thus calculated for 36 points of the figure.

TABLE 2: Comparison between QM (AM1) and QM-MM Mulliken Charges[b]

(a) Comparison between AM1, LSCF, HQ, and QQ Values of the Mulliken Charges for the Two Positions of the Boundary for the $C_3H_7O^-$ Anion with a Sodium Ion (Position 4)

	AM1	LSCF cut 1	LSCF cut 2	HQ cut 1	HQ cut 2	QQ cut 1	QQ cut 2
link atom (cut 2)					−0.01		0.31
C^2	−0.19		−0.31		−0.23		−0.54
H^{21}	0.03		0.04		0.02		0.01
H^{22}	0.03		0.04		0.02		0.01
link atom (cut 1)				−0.11		0.20	
C^3	0.09	−0.10	0.11	0.05	0.08	−0.21	0.10
H^{31}	−0.03	−0.03	−0.04	−0.06	−0.03	−0.07	−0.05
H^{32}	−0.03	−0.03	−0.04	−0.06	−0.03	−0.07	−0.05
O	−0.79	−0.85	−0.80	−0.82	−0.80	−0.85	−0.81

Comparison between AM1, LSCF, and HQ Mulliken Charges for Propanol, Propanolate, and the Acid Form Summed by Group (cf. Figure 3)

	C^1H_3			C^2H_2			C^3H_2			$OH/O^-/OH_2^+$		
	AM1	LSCF[a]	HQ[a]	AM1	LSCF	HQ	AM1	LSCF	HQ	AM1	LSCF	HQ
(b) In the Absence of Sodium												
propanol	0.02	0.00	0.00	0.02	0.01	0.04	0.10	0.11	0.09	−0.13	−0.12	−0.13
propanoate	−0.10	0.00	0.00	−0.07	−0.17	−0.15	−0.01	−0.01	−0.02	−0.82	−0.82	−0.82
oxonium ion	0.10	0.00	0.00	0.06	0.11	0.16	0.22	0.24	0.22	0.62	0.65	0.62
(c) With a Sodium Ion (Position 1)												
propanol	−0.11	0.00	0.00	0.09	−0.02	0.01	0.10	0.11	0.09	−0.10	−0.09	−0.10
propanoate	−0.21	0.00	0.00	−0.01	−0.21	−0.19	−0.01	−0.02	−0.03	−0.77	−0.77	−0.78
oxonium ion	−0.01	0.00	0.00	0.12	0.08	0.12	0.23	0.24	0.22	0.66	0.68	0.65
(d) With a Sodium Ion (Position 5)												
propanol	0.04	0.00	0.00	0.05	0.06	0.09	0.08	0.09	0.08	−0.17	−0.15	−0.17
propanoate	−0.07	0.00	0.00	−0.04	−0.12	−0.09	−0.03	−0.02	−0.04	−0.86	−0.86	−0.87
oxonium ion	0.14	0.00	0.00	0.09	0.16	0.22	0.20	0.23	0.21	0.57	0.62	0.58

[a] Classical charges are set to zero for C^1H_3 in the QM/MM computations. [b] For definition of cut 1 and cut 2 and atom labels, see Figure 3.1.

and acid. The QQ link atom thus gives poor results as soon as an external charge is in the neighborhood of the frontier bond.

Severe charge distortions around the boundary are not observed with HQ (i.e., a link atom that interacts with the MM charges) and the localized orbital of the LSCF method. These methods give smaller errors between QM/MM and full QM results (cf. Table 1 and parts a and b of Figure 4).

In the absence of sodium, the error in PA is 5 kcal/mol for LSCF and 1 kcal/mol for HQ. Similarly, in the presence of sodium, the rms error is 5 kcal/mol for LSCF and 1 kcal/mol for HQ. From Table 1, one can see that the largest errors in the LSCF values correspond to positions 3, 5, and 6 of the sodium ion (cf. Figure 3), where the ion is close to the hydroxyl group and affects most the energy of the $CH_3CH_2CH_2OH_2^+$ cation.

For the DE in the absence of Na^+, the error is 0.6 kcal/mol for LSCF and 1.3 kcal/mol for HQ, while in the presence of sodium the rms error is about 2 kcal/mol for the two methods. It can be seen from Table 1 that, as opposed to what is observed for the proton affinity, the results compare well with the full QM value when the sodium ion is close to the hydroxyl group, and the largest discrepancies are observed when the ion is close to the frontier bond, as in positions 1, 2, and 4 (cf. Figure 3). It is also clear from Figure 4 that there is a range of values around 330 kcal/mol where the correlation between QM/MM and full QM values is poor, and these points correspond to positions 1, 2, and 4 of the sodium ion. This is discussed further in relation to the resulting Mulliken charges (see below).

III.1.b. Mulliken Charges. As stated above, there are no partial charges on the MM aliphatic group in the QM/MM calculations on propanol. Since the only classical charge in the system is that of the sodium ion, its effect is clear.

We give a full Mulliken population analysis for position 4 of the sodium atom which is near the frontier bond and where the effects of the frontier on the charge distribution are particularly important. For other positions of the sodium, we give Mulliken charges summed by atom groups. In all cases, the QM/MM charges obtained with the three boundaries are compared to the full QM (AM1) values.

Position of the Frontier Bond. In Table 2a, we present the full Mulliken population analysis for $C_3H_7O^-$ in the presence of sodium (position 4). The Mulliken charges were computed for both positions of the boundary. From Table 2a, it is clear that in the case of the negative ion, for which the electronic distribution is more diffuse, placing the QM-MM boundary between the C^2 and C^3 carbons leads to a poor description of the charge distribution. In particular, the charge of the oxygen in the propanolate molecule is overestimated, which could lead to artifacts if, for example, this group was involved in hydrogen bonds to surrounding atoms. For propanol itself and for $C_3H_7OH_2^+$, the errors observed when using a small quantum fragment are not as large; the charges on the C^3 methylene group are not adequately described, but the charges at the site of protonation are within 0.01 e^- of the AM1 values (Tables S1 and S2 of the Supporting Information).

Comparison of the Three Methods: Problems with QQ. From the data presented in Table 2a for $C_3H_7O^-$, using the link atom (QQ) that does not interact with the sodium ion leads to a poor description of the charge distribution of the frontier bond; the link atom carries a large positive charge, compared to what is usually observed for hydrogen atoms, and the carbon bonded to the QQ link carries an unusually large negative charge, when compared to the carbon Mulliken charge obtained from both full QM (AM1) calculation and from the other QM/MM methods (similar results are obtained with propanol and $C_3H_7OH_2^+$ (Tables S1 and S2 of the Supporting Information). As mentioned above, the "hyperpolarization" of the link−C2 bond

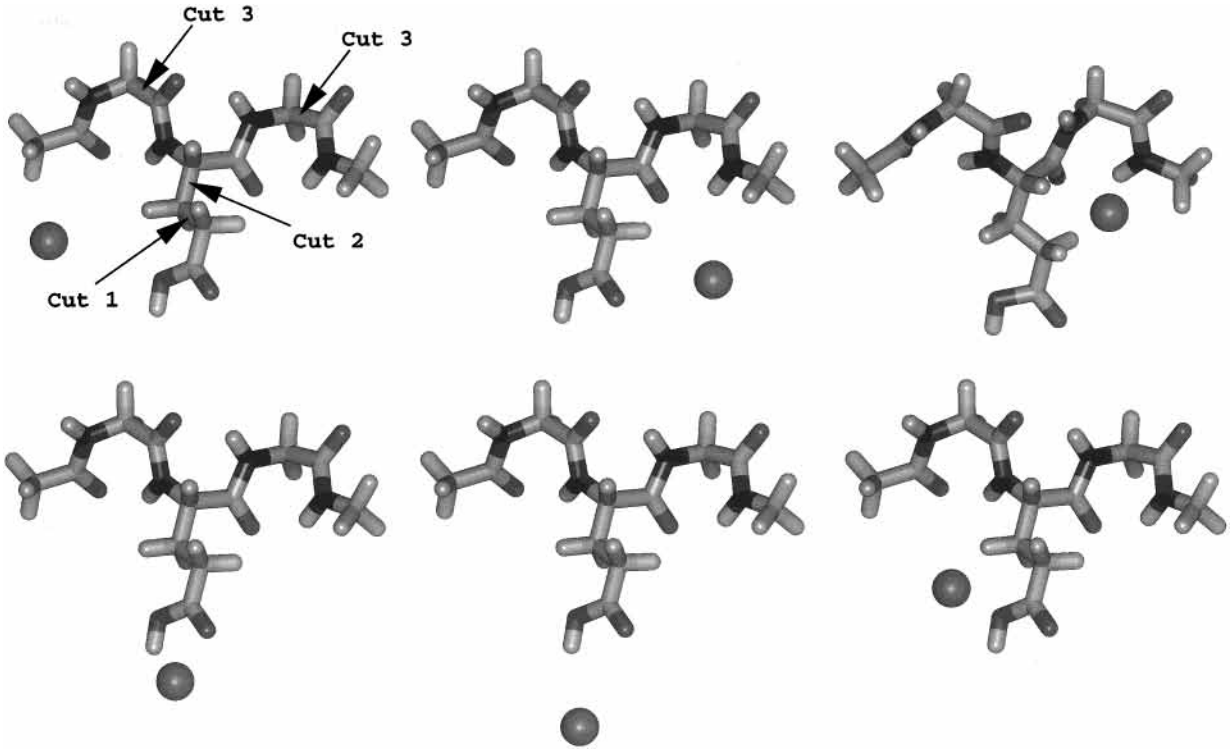

Figure 5. QM/MM division of the tripeptide Ace-Gly-Glu-Gly-Nme (GEG) and positions of the sodium ion. The three different positions of the frontier bond used for the calculations of deprotonation enthalpies are designated by arrows 1, 2, and 3. Cut 3 is on the backbone and therefore introduces two frontiers.

is due to the fact that the QQ link atom does not interact directly with the classical charges (cf. Methods section), which in turn generates a large electrostatic field along this bond. Since the frontier bond is distant from the hydroxyl group, this does not have a significant effect on the Mulliken charges of the hydroxyl group. It is however important enough to lead to values of deprotonation energies and proton affinities that do not compare well with the full QM values (see above).

HQ vs LSCF. In Table 2b−d, we present Mulliken charges summed by atom groups for propanol and the two ions, in the absence of sodium and with the sodium in positions 1 and 5, which are chosen as representative of the charge distributions. Position 1 (cf. Figure 3, with Na^+ close to the QM/MM frontier) yields large errors for DE, while position 5 (Na^+ close to the hydroxyl group) has large errors for PA. The charge data summed by group for positions 2, 3, 4, and 6 are given in the Supporting Information (Tables S3−S6).

From Table 2b−d, it is clear that in all cases the charge distribution on the hydroxyl group is well represented by the QM/MM methods. The difference between the full AM1 and QM/MM charge is at most 0.01 e^- for HQ and 0.01/0.02 e^- for LSCF. Only in the case of position 5 of the Na^+ (cf. Table 2c) is the charge on the OH_2^+ group 0.05 e^- larger for LSCF than for AM1. This is in accord with the energy results, where a larger discrepancy was observed for LSCF in PA when Na^+ is closer to the hydroxyl group. The charge distribution on the methylene group adjacent to the hydroxyl is also well represented, with differences of 0.01/0.02 e^- between AM1 and QM/MM results for both methods.

The largest differences between full AM1 and QM/MM charge distributions are observed for the methylene (C^2H_2) group involved in the QM/MM boundary. When the sodium is close to the boundary, the QM/MM propanolate charges for this group differ by 0.2 e^- from the full AM1 value. When the sodium is close to the hydroxyl, the $C_3H_7OH_2^+$ charge at this position differs by about 0.1 e^- from the AM1 value. It can be seen from Tables 2b−d that the error in the QM-MM C^2H_2 charge is more important since the AM1 charge on the C^1H_3 group gets larger. The QM-MM Mulliken charge on the C^2H_2 group is, in fact, closer to the sum of AM1 charges for the C^1H_3 and C^2H_2 groups.

In the absence of sodium, this C^1H_3 group is globally neutral in propanol, while it carries a small negative charge (−0.1 e^-) in $C_3H_7O^-$ and a small positive charge (+0.1 e^-) in $C_3H_7OH_2^+$. In the QM/MM calculations, the classical charge on this group is zero, as in the CHARMM22 force field. It can be seen from parts c and d of Table 2 that when the sodium ion is in position 1, the AM1 charge on this C^1H_3 group is more negative by about 0.1 e^-; while when the sodium ion is in position 5, the C^1H_3 group becomes less negative by about 0.05 e^- (except for propanol where the charge is unchanged) (cf. Table 2b). This electronic redistribution between the $HO-CH_2-CH_2$ moieties and the terminal CH_3 group cannot be reproduced by the QM/MM computations, where charge redistribution is limited to the $HO-CH_2-CH_2$ moiety.

The positions of sodium that yield larger errors in the energy computations are generally those where the charge redistribution to the C^1H_3 group is the more important. For example (cf. Table 1c), the error on DE is 4 (5) kcal/mol for HQ (LSCF) for position 1 of the sodium (using the larger quantum fragment).

It is important to note that the charge of the hydroxyl group is well represented by QM/MM calculations. So, even if the site of the electronic modifications (OH to O^- or OH_2^+) is well described, there can still be significant errors in the energy.

III.2. Tripeptide gly−glu−gly. We calculated deprotonation enthalpies (see section III.1.a.) of the peptide shown in Figure 5; the corresponding chemical reaction is

$$\text{ace-gly-glu-gly-nme} \rightarrow \text{ace-gly-glu}^-\text{-gly-nme} + H^+$$

TABLE 3: Deprotonation Enthalpies of the Tripeptide with and without External Ion: Comparison of Values Obtained with HQ and LSCF for the Three Positions of the Frontier[d]

		without Na^+	Na^+ (1)	Na^+ (2)	Na^+ (3)	Na^+ (4)	Na^+ (5)	Na^+ (6)	RMS
AM1		360.2	292.35	303.94	271.44	279.66	285.95	292.35	
HQ	cut 1	364.3	292.63	306.87	272.49	279.39	286.72	293.93	1.47
	cut 1[a]	362.3	290.57	304.84	270.53	277.37	284.75	291.97	1.39
	cut 2	362.9	291.73	305.32	271.68	278.62	285.39	292.82	0.81
	cut 2[a]	365.2	294.03	307.52	274.00	280.97	287.70	295.14	2.41
	cut 3	360.0	285.29	305.13	272.28	278.02	284.84	291.45	3.07
	cut 3[a]	363.0	285.25	307.78	274.81	281.15	287.87	294.49	3.81
LSCF	cut 1[c]	382.6							
	cut 1[a,c]	358.3							
	cut 1[b]	384.1	311.98	326.58	292.19	298.98	306.71	313.71	20.77
	cut 1[b,a]	363.4	291.43	306.03	271.45	278.10	286.18	293.13	1.18
	cut 2[c]	355.1							
	cut 2[a,c]	365.0							
	cut 2[b]	358.8	287.23	301.55	266.98	273.96	280.91	288.46	4.56
	cut 2[b,a]	366.4	294.85	309.17	274.65	281.68	288.52	296.09	3.38
	cut 3[c]	375.0							
	cut 3[a,c]	380.2							
	cut 3[b]	335.9	266.52	278.65	246.32	252.72	260.59	267.27	25.61
	cut 3[b,a]	361.8	292.54	304.56	272.30	278.81	286.43	293.13	0.67

[a] No electrostatic interaction between the MM frontier group and the QM atoms. [b] $P_{ll} = 1$ and $a_{l1} = 0.5$ (for LSCF calculations). [c] P_{ll} and a_{l1} from a localization procedure of the molecular orbitals of the whole peptide; the localization is performed on the neutral and anionic peptide, separately, so that two values of P_{ll} and a_{l1} are obtained for each position of the boundary (cf. Table 4). Different values of P_{ll} and a_{l1} were tested only in the absence of Na^+; see text for details. [d] See Figure 5 for atom labels and definition of cut 1, cut 2, and cut 3.

Charged amino acids are often found in enzyme active sites and the transfer of a proton from a substrate to an acid may play a role in catalysis.[21a] Thus, the question of the effect of the choice of the position of the frontier bond for this case is of interest. It can be on the side chain of the amino acids involved but it is useful to examine more distant positions. We consider the three positions shown in Figure 5.

III.2.a. Deprotonation Enthalpies. We first tested the influence of the peptide classical charges on the QM/MM results. This effect was not considered in the propanol example where all classical charges were set to zero except for the sodium ion. We calculated the deprotonation enthalpies of the tripeptide in the absence of sodium ion using the classical charges from the CHARMM22 force field. The calculations were performed for three different QM parts (see Figure 5), two positions of the frontier bond on the side chain of the glutamic acid and a larger QM part with two frontiers in the backbone. The full QM (AM1) and QM/MM results are reported in Table 3.

HQ Link Atom. It can be seen from Table 3 that the differences between the QM (AM1) and QM/MM values in the absence of sodium ion are all below 5 kcal/mol (1.4%) when using the HQ method. A comparison of AM1 and experimental values of deprotonation enthalpies for several carboxylic acids showed that AM1 values differ by up to 10 kcal/mol (3%) from experimental ones.[36] The errors introduced by the HQ QM/MM method are thus of comparable magnitude to those of the full AM1 calculations.

To test the influence of the classical charges on the QM/MM results, we used two alternative approaches. In the first, the charges of the MM group at the frontier interact with the quantum part (a group is defined as a small subset of bonded atoms which is globally neutral, e.g., $C^{\beta}H_2$). In the second case, these charges are removed from the QM/MM electrostatic interactions[12] (rather than being removed solely from the interactions with the link atom). When the charges on the first MM group interact with the quantum part, the maximum error on DE is 4.1 kcal/mol (1.1%) for cut 1. The error observed for cut 2 is 2.7 kcal/mol or 0.7%, and for cut 3 it is only 0.2 kcal/mol. The results are not significantly better when the charges from the first classical group are ignored; while the error decreases for cut 1, it actually increases for cut 2 and cut 3. Thus, the close proximity between the charge of the first MM group and the link hydrogen (0.5 Å) does not introduce significant errors in the QM/MM values of the deprotonation enthalpy.

TABLE 4: Values of a_{l1} and P_{ll} Obtained by the Localization Procedure and Used for the LSCF Calculations on the gly-glu-gly Tripeptide, Reported in Table 3[a]

	a_{l1}/P_{ll}			
	cut 1 (C^{γ})	cut 2 (C^{β})	cut 3 (C gly_1, C^{α} gly_2)	
charged GLU acid	0.43/1.04	0.51/1.07	0.69/1.08	0.46/0.92
neutral GLU acid	0.43/0.98	0.50/1.04	0.68/1.08	0.46/1.04

[a] See Figure 5 for atom labels and definition of Cut 1, Cut 2, and Cut 3.

LSCF Frontier. The LSCF results in the absence of sodium are reported in Table 3. The test cases used for the LSCF method are equivalent to those used for the HQ link method, but we also examine two choices for the electronic population (P_{ll}) and s/p ratio (a_{l1}) of the frontier orbital. A point often discussed in the LSCF method is the fact that the choice of the electronic population of the frozen orbital P_{ll} does not guarantee the neutrality of the QM part. Indeed, as described in the Method section, if P_{ll} is different from 1, there is a partial electronic charge $[(1 - P_{ll})e]$ at the frontier, and thus the total charge of the system can be fractional (positive if $P_{ll} < 1$ and negative if $P_{ll} > 1$). We tested the LSCF method with P_{ll} set to 1, as well as to a value determined by localization of the molecular orbitals from the full QM (AM1) system. To each P_{ll} value corresponds a value for a_{l1} (the s/p ratio); we use $a_{l1} = 0.5$ if $P_{ll} = 1$; otherwise, we use values obtained from the localization procedure (cf. Table 4).

It can be seen from Table 3 that when the charges of the first classical group are included in the QM/MM interactions, large errors (up to 25 kcal/mol or 7% of the DE value) occur, irrespective of the value chosen for P_{ll} and a_{l1}. These errors are most significant for cut 1 and cut 3 where the charges on the first MM atom are large (-0.18 for C^{β} and 0.51 for the C of the second glycine, respectively). For cut 2, the charge of the

classical atom at the boundary is small (0.07 e^- for C^α) and the error is below 5 kcal/mol, as with HQ.

When the electrostatic interactions between the quantum region and the first classical group are disregarded, the LSCF and HQ methods yield comparable results for cut 1 and cut 2, irrespective of the choice made for the electronic population of the frontier orbital. For cut 3, on the other hand, the LSCF values obtained with $P_{ll} = 1$ compare well to the HQ values, while using P_{ll} from a localization procedure yields large errors. The error for P_{ll} from the localization procedure can be traced to the relatively large net charges (+0.08 for glu^- and −0.04 for gluH) introduced at the QM/MM boundary with cut 3, which modify the relative energies of the neutral and anionic peptide. It has been argued that the necessity to predetermine P_{ll} and a_{l1} by localization calculations on model systems is a major drawback of the LSCF procedure.[20] The tests performed here show that it is not necessary to perform the localization procedure and that choosing $P_{ll} = 1$, which ensures the neutrality of the QM/MM frontier, yields better results.

III.2.b. Influence of an External Sodium Ion. From the data presented in Table 3, it is clear that, not surprisingly, the sodium ion has a strong influence on the energy values. The calculated value of deprotonation enthalpy of the gly-glu-gly tripeptide in the absence of the sodium ion is 360.2 kcal/mol (cf. Table 3), while in the presence of the ion, the values are between 265 and 313 kcal/mol. As for propanol, the sodium decreases the values of DE by stabilizing the negative form of the tripeptide.

The rms errors between full QM and QM/MM data obtained with HQ are 1.5, 0.8, 3.1 kcal/mol for cut 1, 2, and 3, respectively, when all the electrostatic interactions are taken into account, and the rms errors are 1.4, 2.4, and 3.8 kcal/mol when the electrostatic interactions between the QM part and the first MM group are removed. The smallest rms error is observed for cut 2, suggesting that placing the frontier between the C^α and C^β of the glutamic acid side chain is an appropriate choice, which also comes at a reasonable cost in computational time. If the quantum fragment is extended beyond the side chain, two frontier bonds have to be introduced. Although these frontier bonds are further from the site of deprotonation, the agreement between QM and QM/MM results is actually poorer with cut 3 than with cut 2. As was observed before, the results do not improve when the interactions with the first MM group are disregarded.

For all the calculations done with LSCF in the presence of sodium, we used a value of 1 for the P_{ll} and 0.5 for the s/p ratio. As was already observed in the absence of sodium ion, the error between QM and QM/MM values is large for cut 1 and cut 3 when the charges of the first MM group are taken into account. The rms error for cut 2, where the charge of the first MM atom is small (0.07), is approximately the same with and without this electrostatic interaction. It is clear from Figure 6 that the source of error with the LSCF method is the classical charges on the peptide itself, rather than the external sodium charge; there is a good correlation between the full QM and QM/MM results in the presence of sodium for the three positions of the boundary, but in the case of cut 1 and cut 3, the LSCF QM/MM values are systematically off. This systematic error is the same as the error found without sodium (20 kcal/mol for cut 1, 25 kcal/mol for cut 3) when the interactions between the first classical atom and the frontier are taken into account.

Thus, contrary to what was observed with the HQ link method, the LSCF method is very sensitive to the presence of a classical charge on the boundary MM atom. If the interactions with this MM charge are removed, the results are comparable

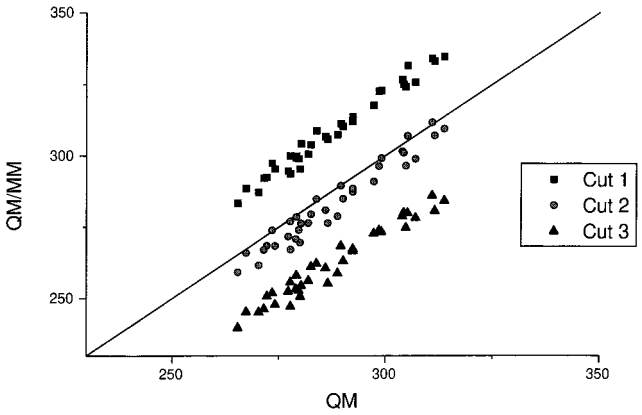

Figure 6. Comparison of QM (AM1) and QM-MM deprotonation enthalpies of the GEG peptide (kcal/mol) in the presence of sodium. Data obtained with the LSCF method and which include electrostatic interactions between the QM region and the first MM group (see text for details).

to those of the HQ method, and they improve when larger QM fragments are considered, as has been generally observed with the LSCF method.[19]

For both methods, the improvement achieved by placing the boundary on the main chain is not significant enough to warrant the added computational cost. In the case of the HQ method, the results do not systematically improve when the boundary is placed on the main chain rather than between the C^α and C^β.

III.3. Geometry Optimizations. The three different QM/MM methods, LSCF, the HQ link atom, and the QQ link, were used to obtain the geometries of ethane and butane and the rotational barriers of butane. The QM/MM geometries were compared with fully optimized structures obtained from ab initio MP2/6-311G**, semiempirical AM1, and molecular mechanics (CHARMM22 force field) calculations. These comparisons allow us to evaluate if the differences between full AM1 and QM/MM AM1 geometries are of the same magnitude or significantly larger than the differences between semiempirical, ab initio, and MM geometries. For the ethane molecule, one methyl group is computed by quantum mechanics while the other one is described by molecular mechanics (see Figure 7). The classical methyl group carries partial charges from the CHARMM22 force field (C, −0.27; H, +0.09). For the butane molecule, the frontier bond is placed between carbon atoms C^2 and C^3(see Figure 7). As for ethane, the classical groups carry partial charges from the CHARMM22 force field (C, −0.27; H, +0.09 for CH_3 and C, −0.18; H, +0.09 for CH_2)

III.3.a. Bond Lengths. The comparison of bond lengths is presented in Table 5. We first used the standard protocol where all QM/MM bonded terms including at least one MM atom and all electrostatic interactions between the quantum and classical region are considered (except for the QQ link, which does not interact with the classical charges).

It can be seen from Table 5 that all methods overestimate the frontier $C_{QM}-C_{MM}$ bond length. The results are particularly poor with the HQ link method, where the bond length is overestimated by 0.15 Å compared to the MP2 data. The LSCF method overestimates by 0.06 Å and QQ link by 0.03 Å. The other bond lengths (such as C−H) are in good agreement with the full QM results (data not shown).

It has been argued[12] that it is incorrect to introduce both a classical harmonic term for the frontier bond and an electrostatic interaction between the classical and quantum atoms of the frontier bond. We therefore redetermined the geometries using a modified protocol,[12] where the electrostatic QM/MM interaction between the quantum atoms and the first frontier classical

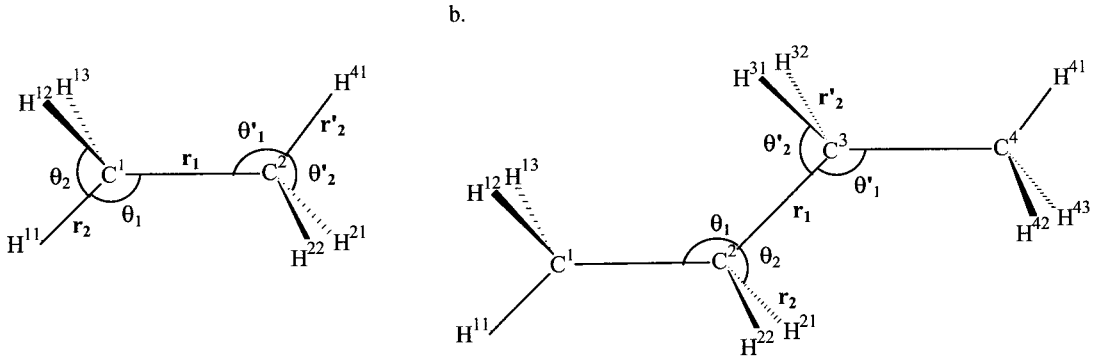

Figure 7. QM-MM partitioning and labeling of atoms used for the ethane and butane molecule.

TABLE 5: Carbon−Carbon Bond Length for the Frontier C−C Bond in Ethane and Butane (Å)

	ethane	butane
experiment	1.535	1.531
MP2/6-311G**	1.529	1.529
AM1	1.500	1.514
CHARMM22	1.533	1.539
LSCF	1.586	1.581
LSCF[a]	1.586	1.580
LSCF[b]	1.530	1.538
HQ	1.678	1.660
HQ[a]	1.694	1.670
HQ[b]	1.530	1.537
QQ	1.559	1.560
QQ[a]	1.561	1.560
QQ[b]	1.530	1.537

[a] Only angles suggested by Eurenius et al. [b] Only angles suggested by Eurenius et al. and no electrostatic interactions between MM frontier group and QM atoms.

group are removed. Another modification is that MM angle terms (valence and dihedrals) are included only if at least one central atom is MM.[12]

With this modified protocol, the QM/MM frontier C−C bond length is within 0.015 Å of the reference MP2 value. This improvement is brought about by the modification of the electrostatic interaction between the two frontier carbon atoms and between the link atom and the first classical atom, while the modification of the angle MM terms hardly changes the bond length values. We have however seen in the calculations performed on the tripeptide that removing the charges of the classical atom at the frontier generally yields poorer results for energies, as this frontier atom can interact with QM atoms further away from the QM/MM boundary. As an alternative to modifying the electrostatic interactions, it is possible to adopt a smaller equilibrium bond length in the MM force field for the $C_{QM}-C_{MM}$ bond. As the frontier bond is special, it is not inconsistent with the principle of a classical force field to parametrize it with a different value than a regular C−C bond in butane or ethane ($r_0 = 1.53$ Å). We made tests on butane, decreasing the r_0 values until the frontier bond length was in good agreement with full QM value. This leads to good results for LSCF ($r_0 = 1.46$ Å, $d_{C-C} = 1.53$ Å) and QQ ($r_0 = 1.49$ Å, $d_{C-C} = 1.53$ Å); for HQ, the equilibrium distance must be decreased more ($r_0 = 1.27$ Å) to obtain a distance of 1.53 Å. As the frontier bond differs according to the QM/MM partition, this reparametrization of the frontier bond should be done for each system.

III.3.b. Valence Angles. The QM/MM and MP2, AM1, and CHARMM22 valence angles are generally in good agreement (cf. Table 6). QM/MM methods tend to underestimate H−C−C angles and overestimate H−C−H angles, but all methods are

TABLE 6: Valence Angles at Carbon (Degrees) at the QM-MM Boundary[e]

	ethane[c]		butane[d]	
	θ_1/θ'_1	θ_2/θ'_2	θ_1/θ'_1	θ_2/θ'_2
MP2/6-311G**	111.1	107.8	112.8	109.0
AM1	110.7	108.2	111.6	109.4
CHARMM22	110.6	108.3	109.5	109.5
LSCF	106.5/109.9	112.3/109.0	110.0/114.1	106.7/108.1
LSCF[a]	106.8/110.3	112.0/108.6	107.1/114.1	106.1/108.1
LSCF[b]	107.9/110.8	111.0/108.1	106.7/114.3	106.4/108.7
HQ	109.7/109.1	109.2/109.8	113.6/114.1	108.2/107.2
HQ[a]	111.0/109.0	107. 9/110.0	112.3/114.0	109.3/107.0
HQ[b]	109.5/110.3	109.4/108.6	111.3/114.5	108.2/108.6
QQ	109.2/110.2	109.8/108.7	112.8/114.2	107.8/108.2
QQ[a]	108.9/110.2	110.0/108.8	111.2/114.2	107.5/108.3
QQ[b]	109.5/110.3	109.4/108.6	111.2/114.5	107.9/108.6

[a] Only angles suggested by Eurenius et al. [b] Only angles suggested by Eurenius et al. and no electrostatic interactions between MM frontier group and QM atoms. [c] For ethane, $\theta_1 = H_{QM}-C_{QM}-C_{MM}/\theta'_1 = C_{QM}-C_{MM}-H_{MM}$ and $\theta_2 = H_{QM}-C_{QM}-H_{QM}/\theta'_2 = H_{MM}-C_{MM}-H_{MM}$. [d] For butane: $\theta_1 = C_{QM}-C_{QM}-C_{MM}/\theta'_1 = C_{QM}-C_{MM}-C_{MM}$ and $\theta_2 = H_{QM}-C_{QM}-C_{MM}/\theta'_2 = C_{QM}-C_{MM}-H_{MM}$. [e] See Figure 7 for a description of the geometries.

within 2° of each other. An exception is LSCF, where differences of up to 4° are observed in valence angles that should be equal by symmetry. When using mixed QM/MM methods, an artificial asymmetry can indeed be introduced between angles with central MM atoms (such as $H_{MM}-C_{MM}-C_{QM}$) and angles with central QM atoms (such as $H_{QM}-C_{QM}-C_{MM}$), which are equivalent in full QM or full MM treatments. We observe this asymmetry to be significant only with the LSCF method, while for the HQ and QQ link the differences are small.

Using a modified protocol where the angle terms that do not involve central MM atoms are removed from the classical force field, does not lead to any improvement of the internal coordinates. As can be seen from Table 6, removing these angles terms generally increases the discrepancy with the reference values, and also increases the difference between the H−C−C angles with a central MM and with a central QM atom, as opposed to reducing this difference and preserving molecular symmetry. The error introduced by removing this valence term is especially important in the LSCF method.

III.3.c. Rotational Barrier of Butane. The main contributions to the rotational barriers of butane in the QM/MM computations arises from the classical energy terms such as the C−C−C−C dihedral angle term and the nonbonded van der Waals term. These classical terms bring the QM/MM rotational barrier closer to the MM force field results than to the full AM1 results, which are known to underestimate the barrier (cf. Table 7). Figure 8 represents the calculated QM/MM rotational barriers of butane compared with the CHARMM 22 values.

TABLE 7: Energies (kcal/mol) for Butane Conformers (cf Figure 8) Relative to the Trans Conformer ($\theta_{C-C-C-C} = 180°$)

	$\theta_{C-C-C-C} = 0°$	$\theta_{C-C-C-C} = 60°$	$\theta_{C-C-C-C} = 120°$
CHARMM22	5.25	0.93	3.48
AM1	3.28	0.87	1.52
QQ	4.97	0.71	3.45
QQ[a]	5.02	0.72	3.44
QQ[b]	4.84	0.62	3.42
HQ	4.43	0.56	3.34
HQ[a]	4.41	0.55	3.32
HQ[b]	4.83	0.62	3.42
HQ[c]	4.98	0.74	3.44
LSCF	5.37	0.82	3.52
LSCF[a]	5.79	0.98	3.57
LSCF[b]	5.39	0.78	3.47

[a] Only angles suggested by Eurenius et al. [b] No electrostatic interactions between the MM frontier group and QM atoms. [c] $r_0 = 1.27$.

It can be seen from Table 7 and Figure 8 that the value of the cis ($\theta = 0°$) barrier for HQ with the usual r_0 (1.53 Å) is 0.8 kcal/mol lower than the CHARMM value, while the LSCF and QQ values are within 0.25 kcal/mol (5%) of the CHARMM22 values. For $\theta = 60°$, LSCF is in good agreement with the CHARMM22 value (within 0.1 kcal/mol), while QQ gives a somewhat larger error (within 0.2 kcal/mol) and HQ (with $r_0 = 1.53$ A) is 0.4 kcal/mol lower. For $\theta = 120$ °, all methods are in good agreement with the CHARMM22 results. The barriers computed by the QQ and LSCF frontiers are thus in good agreement with the CHARMM22 value; while for the HQ link, the value of r_0 (1.27 Å) which gives a correct bond length (1.53 Å) should be used to obtain results that are in agreement with the others (cf. Table 7). If $r_0 = 1.53$ is used, the bond length is 1.60 Å (see above) and the results are much poorer (see Figure 8 and Table 7). Similarly, when the electrostatic interactions with the frontier classical group are removed, the bond length decreases and, as expected, the rotational barrier obtained with the HQ link improves (cf. Table 7). Using a protocol with modified angle terms, on the other hand, does not improve the rotational barrier, and even worsens it for the LSCF method.

Thus, for small molecules such as the one we have tested, the three methods yield satisfying structures, excepted for the length of the frontier bond. The problem can be alleviated by removing the electrostatic interactions between the quantum fragment and the frontier classical group or by reducing the equilibrium length in the MM energy term for the frontier bond. Other modifications of the frontier region, such as including only angles terms containing a central MM atom, do not lead to significantly improved geometries.

IV. Conclusion

Three descriptions of the frontier bond currently used in hybrid QM/MM methods have been compared. They are a hydrogen link atom that is invisible to the MM atoms (QQ link), a hydrogen link atom that interacts with the MM atoms (HQ link), and a description of the frontier that uses a fixed strictly localized orbital (LSCF method). For each method, we have examined the influence of the location of the frontier bond and the interaction terms in the frontier region on QM/MM energies and geometries. In particular, to emulate the large electrostatic interactions that can be present in the active site of enzymes, we computed deprotonation energies and proton affinities in the presence of an external MM charge.

The tests showed that using a link atom that does not interact with the MM charges (QQ link) can lead to large errors in energy computations and is therefore not recommended. More generally, it is not advisable to modify the electrostatic interactions between the MM atoms and a subset of the QM atoms, as this amounts to creating an artificial electric field on the QM region and leads to distortion of charge distributions. This, in turn, can adversely affect the energies.

The test performed on single point energies showed that using a link atom that interacts with the MM charges (HQ link) yields results in good agreement with full QM computations. When geometry optimizations are performed, the interactions between the link and the first classical atom at the frontier, which is situated at only about 0.5 Å from the link atom, leads to incorrect bond length for the frontier QM/MM bond, which in turn can

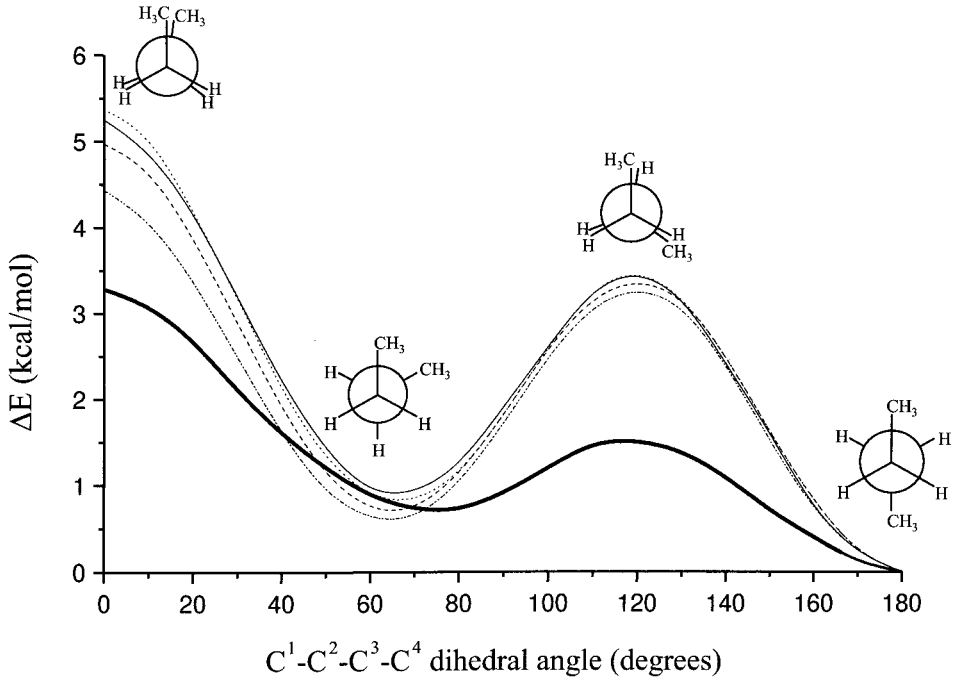

Figure 8. Comparison of QM-MM rotational barriers of butane (dotted line, LSCF; dashed line, QQ; dot−dashed line, HQ with long bond length 1.66 Å) with CHARMM22 barriers (continuous line) and full AM1 potential (bold continuous line).

affect the rotational barrier around the frontier bond. This difficulty can be solved by removing the interactions between the link and the classical group at the frontier, but then, to avoid the electric field problems mentioned above, it is necessary to prevent the frontier classical group from interacting with all QM atoms. This procedure is recommended only if the charge of the classical group at the frontier is small; otherwise important electrostatic interactions between the QM and MM regions are disregarded. A better option to correct the geometries is to introduce constraints that maintain (i) the angle between the link and the two frontier atoms equal to zero (i.e., the link atom is constrained to be along the bond) and (ii) the length of the frontier bond at its correct value; the latter can be achieved by simply changing the r_0 value for the bond term in the MM force field. The drawback of this procedure is that such constraints can introduce large forces at the frontier, but it is unlikely that this would introduce significant errors except for the specific vibrational frequencies.

The strictly localized orbital used in the LSCF method is an elegant way to terminate the quantum charge distribution without introducing extra atoms in the system. The problems associated with the geometry optimization of the frontier QM/MM bond with the HQ link method do not appear with the LSCF method. The energy computations have shown, however, that the LSCF method is generally less robust than the HQ link method. The computed energies are more sensitive to factors such as the size of the QM part, the values of the MM charges on the frontier MM atom, and the choice of the electronic population of the frozen orbital. Compared to the link atom method, the frozen localized orbital allows electron delocalization on a smaller number of molecular orbitals; i.e., the SCF calculations include two electrons more with the HQ link than with the LSCF method. This probably explains the fact that the LSCF method is more sensitive to the size of the quantum part.[37]

Large errors in energy computations have also been observed with LSCF when the charge of the classical frontier atom is large. The tests have shown that adopting a fixed value of one for the population of the frozen orbital P_{ll}, which avoids introducing fractional electronic charges at the frontier, is not only simpler (as it avoids localization procedures on model systems) but also yields better values of the energies.

We thus conclude that, except for the QQ link method, which should clearly be avoided, the HQ and LSCF methods give results of similar accuracy and neither one is systematically better than the other. If a small quantum fragment is used or if the classical atom at the frontier bears a large charge, the HQ method is preferable. If the frontier bond is sufficiently remote from the site of chemical modification, and the classical frontier atom bears a small charge, LSCF and HQ give equivalent energetic results, but LSCF is more straightforward since it is not necessary to introduce constraints at the frontier bond to maintain the proper geometry. I. Antes and W. Thiel also recently studied the description of frontier bonds in QM/MM methods.[38] They used different test cases, but their conclusions are in many respects similar to ours.

Finally, the tests have shown that when a charge is close to the frontier bond, errors can occur in relative energies even if the frontier is far enough from the site of the chemical modification so that the electronic distribution of the site is correctly modeled by the QM/MM method. By definition, QM/MM descriptions prevent charge delocalization across a frontier bond, which in turn can cause the errors in the relative energies when a charge is near the frontier. These errors are thus artifacts implicit to the QM/MM representation. They occur irrespective of the description used for the frontier bond.

From a practical point of view, it should be possible to introduce warnings in QM/MM codes that point to possible problems, as these errors in energies are associated with modifications of the Mulliken charge distribution at the frontier QM atom. It must also be noted that the present computations have been performed with a nonpolarizable MM region and that it is possible that some of the errors could be reduced if the MM part was allowed to be polarized by the large charges. It is necessary in using QM/MM methods in enzyme systems to carefully monitor MM partial charges in the active site to make sure that they do not introduce artifactual results. With this caution, we believe that QM/MM will continue to play a useful role in the study of systems that are too large for a full quantum treatment.

Acknowledgment. The authors thank Pr. J.-L. Rivail, Dr. C. Burgess, and Dr. S. Fischer for helpful discussions. N. Reuter acknowledges a doctoral fellowship from the French Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche and a short term grant from the Fondation pour la Recherche Medicale (FRM) France. N.R. also thanks Dr. P. Aplincourt for doing the ab initio calculations and Dr. A. Cartier for his support of this work. Part of this work was supported by a grant from the National Science Foundation (Harvard).

Supporting Information Available: Tables S1−S7 showing comparison between QM (AM1) and QM/MM Mulliken charges. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(2) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solution*; Wiley: New York, 1992.
(3) (a) Gao, J. In *Reviews in Computational Chemistry VII*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1996; p 119. (b) Field, M.; Gao, J. *Int. J. Quantum Chem.* **1995**, *60*, 1093. (c) Amara, P.; Field, M. *Encyclopedia of Computational Chemistry*; Wiley: New York, 1998.
(4) Cramer, C. J.; Truhlar, D. G. In *Reviews in Computational Chemistry VI*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1995; p 1.
(5) (a) Taubes, G. *Science* **1997**, *275*, 1420. (b) Dandliker, P. J.; Holmlin, R. E.; Barton, J. K. *Science* **1997**, *275*, 1465.
(6) (a) Christoffersen, R. E.; Maggiora, G. M. *Chem. Phys. Lett.* **1969**, *3*, 419. (b) Davis, T. D.; Maggiora, G. M.; Christoffersen, R. E. *J. Am. Chem. Soc.* **1974**, *96*, 7878.
(7) Allen, L. C. *Ann. N. Y. Acad. Sci.* **1981**, *367*, 383.
(8) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.
(9) Brooks, B. R.; Burccoleri, R. E.; Olafson, B. D.; States D. J.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(10) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
(11) Harrison, M. J.; Burton, N. A.; Hillier, I. H. *J. Am. Chem. Soc.* **1997**, *119*, 12285.
(12) Eurenius, K. P.; Chatfield, D. C.; Brooks, B. R.; Hodoscek, M. *Int. J. Quantum Chem.* **1996**, *60*, 1189.
(13) Vasilyev, V. V. *J. Mol. Struct. (THEOCHEM)* **1994**, *304*, 129.
(14) *HyperChem Users Manual*; Computational Chemistry, Hypercube; Hypercube Inc.: Waterloo, Ontario, Canada.
(15) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(16) Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, J. J. *J. Comput. Chem.* **1994**, *15*, 269.
(17) Monard, G.; Loos, M.; Théry, V.; Baka, K.; Rivail, J.-L. *Int. J. Quantum Chem.* **1996**, *58*, 153.
(18) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.
(19) Théry, V. Thesis, University of Nancy I, France, 1993.
(20) Gao, J.; Amara, P.; Alhambra, C.; Field, M. *J. Phys. Chem. A* **1998**, *102*, 4714.
(21) For example: (a) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Karplus, M. *Biochemistry* **1991**, *30*, 5826. (b) Lyne, P. D.; Mulholland, A. J.; Richards, W. G. *J. Am. Chem. Soc.* **1995**, *117*, 11345.

(c) Barnes, J. A.; Williams, I. H. *Biochem. Soc. Trans.* **1996**, *24*, 263. (d) Mulholland, A. J.; Richards, W. G. *Proteins: Structure, Function, and Genetics* **1997**, *27*, 9. (e) Cunningham, M. A.; Ho, L. L.; Nguyen, D. T.; Gillilan, R. E.; Bash, P. A. *Biochemistry* **1997**, *36*, 4800. (f) Waszkowycz, B.; Hillier, I. H.; Gensmantel, N.; Payling, D. W. *J. Chem. Soc., Perkin Trans.* **1991**, *2*, 225. (g) Chatfield, D. C.; Eurenius, K. P.; Brooks, B. R. *J. Mol. Struct. (THEOCHEM)* **1998**, *423*, 79. (h) Liu, H.; Müller-Plathe, F.; van Gunsteren, W. F. *J. Mol. Biol.* **1996**, *261*, 454.

(22) (a) Antonczak, S.; Monard, G.; Ruiz-López, M. F.; Rivail, J.-L. *J. Am. Chem. Soc.* **1998**, *120*, 8825. (b) Salvatella, L.; Mokrane, A.; Cartier, A.; Ruiz-López, M. F. *J. Org. Chem.* **1998**, *63*, 4664.

(23) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(24) Lyne, P. D.; Hodoscek, M.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 3462.

(25) Stewart, J. J. P. In *Reviews in Computational Chemistry I*; Lipkowitz, K. B., Boyds, D. B., Eds.; VCH: New York, 1990; p 45.

(26) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; 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.* **1998**, *102*, 3586.

(27) (a) Freindorf, M.; Gao, J. *J. Comput. Chem.* **1996**, *17*, 386. (b) Ho, L. L.; MacKerell, A. D., Jr.; Bash, P. A. *J. Phys. Chem.* **1996**, *100*, 4466. (c) Bash, P. A.; Ho, L. H.; MacKerell, A. D., Jr.; Levine, D.; Hallstrom, P. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 3698.

(28) Pople, J. A.; Beveridge, D. L. *Approximate Molecular Orbital Theory*; Series in Advanced Chemistry; McGraw-Hill: New York, 1970; p 25.

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

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

(31) Stewart, J. J. P. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1.

(32) Rinaldi, D.; Hoggan, P. E.; Cartier, A. *GEOMOS*, QCPE 584.

(33) Edminston, C.; Ruedenberg, K. *Rev. Mod. Phys.* **1963**, *35*, 457.

(34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez C.; Pople, J. A. *GAUSSIAN 94*, revision B.3; Gaussian, Inc.: Pittsburgh, PA, 1995.

(35) Stull, D. R. J. prophet, *JANAF Thermochemical Tables*, 1971; NSRDS-NBS37.

(36) Dewar, M. J. S.; Dieter, K. *J. Am. Chem. Soc.* **1986**, *108*, 8075.

(37) This importance of the number of active electrons was confirmed by tests recently performed using the GHO method on the Gly-Glu-Gly tripeptide (N. Reuter, Ph.D. Thesis, Nancy University, 1999). The GHO method, like the LSCF method, uses localized orbitals to describe the frontier bond. GHO differs from LSCF in that four atomic orbitals are placed on the frontier classical atom. Three of these orbitals (directed towards MM atoms) are frozen, while the atomic orbital directed towards the quantum frontier atom is included in the SCF calculations. As a result, the GHO method, although similar in spirit to the LSCF method, includes the same number of active electrons in the SCF calculations as a calculation using a hydrogen link atom at the frontier. In tests on the Gly-Glu-Gly tripeptide, it was found that the GHO method gives results closer to the HQ link method than to the LSCF method.

(38) Antes, I.; Thiel, W. *ACS Symp. Ser.* **1998**, *712*, 50 (Combined quantum mechanical and molecular mechanical methods).