Adjusted Connection Atoms for Combined Quantum Mechanical and Molecular Mechanical Methods

Iris Antes and Walter Thiel*,†

Organisch-chemisches Institut, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

Received: June 2, 1999

Connection atoms are proposed as an alternative to link atoms in semiempirical hybrid calculations that divide a system at a C−C single bond into a quantum mechanical (QM) and a molecular mechanical (MM) region. A connection atom interacts with the other QM atoms as a specially parametrized QM atom, and with the other MM atoms as a standard carbon MM atom. Detailed definitions of these interactions are given for three QM/MM coupling models (A mechanical embedding, B/C electronic embedding without/with MM polarization). Semiempirical connection atom parameters are derived for three standard methods (MNDO, AM1, PM3) such that the adjusted connection atoms closely reproduce the geometrical and electronic properties of methyl groups. The corresponding deviations are generally smaller than the intrinsic errors of these methods. QM/MM test calculations on proton affinities confirm the usefulness of the adjusted connection atoms, particularly in coupling model B. Connection atoms are conceptually superior to link atoms in that they do not introduce extra centers and thus lead to well-defined potential surfaces. In addition, they allow an improved semiempirical description of the QM/MM interactions.

1. Introduction

In recent years, advances in computer technology and theoretical methodology have significantly extended the size and complexity of systems that are accessible to quantum mechanical (QM) calculations. Normally, however, it is still not yet practical to study large-scale problems with thousands of atoms by pure QM methods. On the other hand, this is often not even necessary, because the processes that need to be described at the QM level usually occur in a fairly localized region (e.g., bond breaking and making in the active side of an enzyme, or electronic excitation within a chromophor). It may therefore be sufficient to treat only this relevant localized region by quantum mechanics and to include the influence of the surroundings at a simpler level, e.g., by molecular mechanics (MM). This is the idea behind the combined QM/MM potentials,[1−6] which have first been proposed by Warshel and Levitt in 1976[1] and have become quite popular during the past decade.[2−38] Hybrid QM/MM methods have been used to study solvent effects,[4,7−15] reaction mechanisms in enzymes,[14,16−23] electronic excitations in the condensed phase,[24−31] and topics in organometallic and inorganic chemistry, particularly catalysis.[32−38] These latter studies[32−38] have used ab initio and density functional QM methods, while most of the organic and biochemical QM/MM applications have employed semiempirical QM components.

A central issue in the QM/MM approach is how to partition the system into QM and MM regions and how to treat the QM/MM interactions for different QM/MM combinations.[2,5] Special problems arise when the QM/MM boundary disrupts covalent bonds, leading to one unsaturated valency per cut bond in the QM calculation. There are two major concepts for dealing with this situation.

In the approach of Rivail and co-workers, the electron density of this bond is precalculated, and the corresponding “frozen” orthogonal hybrid orbitals are then used in the actual QM/MM calculation.[39−41] These orbitals are assumed to be transferable and are not allowed to adapt during the QM/MM calculation. The generalized hydrid orbital (GHO) approach[42] is a refinement of this concept: the QM part is saturated by a pseudoatom with three frozen hybrid orbitals such that the electron density in the cut bond can readjust in the QM/MM calculation.

In the second and most widely used approach, each free valency is saturated by a so-called *link atom*,[2,3] normally a hydrogen atom (or a pseudohalogen atom in the case of the HYPERCHEM software[43]). The presence of additional “unphysical” link atoms may cause problems in the definition of a meaningful potential surface,[5] e.g., because of extra degrees of freedom due to link atoms or the need for link atom corrections to avoid double counting. Furthermore, there is some ambiguity on how to treat the QM/MM interactions involving the link atoms, and different link atom options have been considered in order to minimize any associated artifacts.[44]

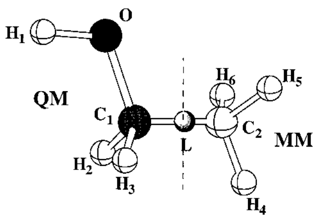
It would clearly be advantageous to saturate each free valency of the QM region by a pseudoatom that occupies the same position as the neighboring MM atom in the bond being cut and that mimics its electronic influence as closely as possible. The recently proposed pseudobond approach[45] addresses this goal at the ab initio and density functional QM level, by introducing a one-free-valence atom with seven valence electrons, a flourine basis set, and an effective core potential designed to mimic a C−C single bond. We have developed and implemented a conceptually similar approach[46] at the semiempirical QM level, which is outlined in this article (further details being available from ref 46).

In our approach we define an *adjusted connection atom* with a special parametrization that connects the QM and MM regions. It may be viewed as a QM atom with some additional classical bonded terms. Its interactions with the neighboring QM carbon atom (in the bond that is formally cut) are treated entirely by quantum mechanics, thereby removing any double counting and the associated corrections that have plagued the traditional link

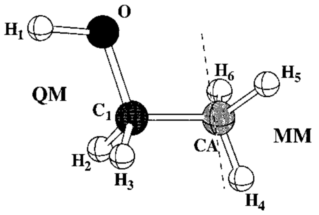
---

[†] Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, D-45470 Mülheim, Germany.

10.1021/jp991771w CCC: $18.00



link atom



adjusted connection atom

Figure 1. Definition of the QM and MM regions for ethanol. The QM/MM cut is made through the C_1-C_2 bond when using link atoms, or the C_2 atom (CA) in the connection atom approach.

atom approaches.[2,5] On the other hand, bonded interactions with MM atoms are handled classically, using carbon force field parameters for the connection atom. To be specific, we assume that carbon−carbon bonds are cut and that the connection atom is adjusted such that it mimics the properties of an attached methyl group. The parameters for the connection atom are determined by a semiempirical calibration designed to reproduce theoretical QM reference data (energies, geometries, dipole moments, and charge distributions) for suitably chosen methyl compounds. This ensures that the connection atom remains at approximately the same position as the carbon atom being replaced and that it has similar electronic properties.

2. Theoretical Background

We have previously defined a hierarchy of QM/MM coupling models[5] and investigated different treatments of link atoms.[5,44,46] In this section we present the corresponding definitions for connection atoms and relate them to those for link atoms, using an analogous notation as before.

Consider a molecule X−Y partitioned into an MM region X (atoms J) and a QM region Y (atoms A). The QM/MM division is made through a particular bond that is saturated by a link atom L in the established link atom scheme or through a particular connection atom (CA) in our current approach. Figure 1 illustrates these different QM/MM divisions for ethanol. When connection atoms are used, the total energy can generally be written as a sum of a pure MM term $E_{\mathrm{MM}}(\mathrm{X-CA})$, a pure QM term $E_{\mathrm{QM}}(\mathrm{Y-CA})$, and an interaction term $E_{\mathrm{QM/MM}}(\mathrm{X,Y})$.

$$E = E_{\mathrm{MM}}(\mathrm{X-CA}) + E_{\mathrm{QM}}(\mathrm{Y-CA}) + E_{\mathrm{QM/MM}}(\mathrm{X,Y}) \quad (1)$$

Comparing with the corresponding expression in the link atoms scheme, there is no correction E_{LINK},[5] which was previously needed to remove spurious link atom contributions from the total energy and thereby caused an artificial distinction between the total energy and the Born−Oppenheimer potential energy surface.[5] Because of the absense of such corrections, the connection atom treatment avoids this deficiency at the outset and provides a well-defined potential surface.

The simplest coupling model A[5] involves mechanical embedding of the QM region. The pure MM term $E^{\mathrm{A}}_{\mathrm{MM}}(\mathrm{X-CA})$ is evaluated according to the prescriptions of the chosen force field and includes all force field contributions that refer only to the atoms $\mathrm{J} \in \mathrm{X}$ plus any connection atoms CA (using standard carbon force field parameters for CA). The pure QM energy $E^{\mathrm{A}}_{\mathrm{QM}}(\mathrm{Y-CA})$ is obtained from a QM calculation for an isolated molecule Y−CA (using special connection atom parameters; see section 3). The interaction energy $E^{\mathrm{A}}_{\mathrm{QM/MM}}(\mathrm{X,Y})$ is determined solely from the force field.

$$E^{\mathrm{A}}_{\mathrm{tot}} = E^{\mathrm{A}}_{\mathrm{MM}}(\mathrm{X-CA}) + E^{\mathrm{A}}_{\mathrm{QM}}(\mathrm{Y-CA}) + E^{\mathrm{A}}_{\mathrm{QM/MM}}(\mathrm{X,Y}) \quad (2)$$

$$E^{\mathrm{A}}_{\mathrm{QM/MM}}(\mathrm{X,Y}) = E^{\mathrm{Bonded}}_{\mathrm{MM}}(\mathrm{X,Y}) + E^{\mathrm{vdW}}_{\mathrm{MM}}(\mathrm{X,Y}) + E^{\mathrm{coul}}_{\mathrm{MM}}(\mathrm{X,Y}) \quad (3)$$

Conceptually, the interactions of a connection atom with all MM atoms are included in $E^{\mathrm{A}}_{\mathrm{MM}}(\mathrm{X-CA})$, while those with all QM atoms are represented through $E^{\mathrm{A}}_{\mathrm{QM}}(\mathrm{Y-CA})$. Hence, in eq 3, the bonded terms $E^{\mathrm{Bonded}}_{\mathrm{MM}}(\mathrm{X,Y})$ only consist of bond angle and torsion angle contributions (involving the connection atom and at least one QM and one MM atom), while the nonbonded van der Waals terms $E^{\mathrm{vdW}}_{\mathrm{MM}}(\mathrm{X,Y})$ and Coulomb terms $E^{\mathrm{coul}}_{\mathrm{MM}}(\mathrm{X,Y})$ are composed of contributions from QM/MM atom pairs only (excluding connection atoms).

In model A of the link atom scheme,[5] the pure MM term $E^{\mathrm{A}}_{\mathrm{MM}}$ is handled exactly as described above, and $E^{\mathrm{A}}_{\mathrm{QM}}(\mathrm{Y-L})$ is also computed for an isolated molecule. However, there are more QM/MM interaction terms since the atom replaced by the link atom L (i.e., the equivalent of the connection atom) is treated as a regular MM atom. In the specific example of methanol (Figure 1), the link atom scheme includes QM/MM force field contributions for C_1-C_2, $O-C_1-C_2$, $H_2-C_1-C_2$, $H_3-C_1-C_2$, and $H_1-O-C_1-C_2$, as well as nonbonded force field interactions between the C_2 and the QM region, all of which are missing in the connection atom scheme, where they are represented by $E^{\mathrm{A}}_{\mathrm{QM}}(\mathrm{Y-CA})$.

Coupling model B[5] provides an electronic embedding of the QM region, by determining the QM wave function under the influence of the external charges:

$$\hat{H}_{\mathrm{el}}(\mathrm{Y-CA;X}) = \hat{H}_{\mathrm{el}}(\mathrm{Y-CA}) - \sum_{i}^{\mathrm{Y-CA}} \sum_{\mathrm{J} \notin \mathrm{CG}}^{\mathrm{X}} q_{\mathrm{J}}/r_{i\mathrm{J}} \quad (4)$$

$\hat{H}_{\mathrm{el}}(\mathrm{Y-CA;X})$ and $\hat{H}_{\mathrm{el}}(\mathrm{Y-CA})$ are the electronic Hamiltonians for Y−CA with and without the external field, respectively, which is generated by the MM atomic point charges q_{J} outside the connection group (consisting of the carbon connection atom and any covalently bound hydrogen atom: $\mathrm{CG} = \mathrm{CH_3}$ in $\mathrm{Y-CH_3}$, $\mathrm{CG} = \mathrm{CH_2}$ in $\mathrm{Y-CH_2-R}$, $\mathrm{CG} = \mathrm{CH}$ in $\mathrm{Y-CHRR'}$, and $\mathrm{CG} = \mathrm{C}$ otherwise). Excluding point charges from the connection group is motivated as follows: The QM parameters for the connection atoms are optimized such that they reproduce the properties of $\mathrm{Y-CH_3}$ (see section 3) and thus effectively absorb the influence of any internal methyl charges. Most force fields treat the $\mathrm{CH_3}$, $\mathrm{CH_2}$, and CH units as charge groups[47] with an overall net charge of zero, and it is thus reasonable to assume that a parametrized connection atom may represent each of those units (see section 4 for numerical validation). Excluding point charges from the connection group thus conserves the net charge of the interacting MM region and also avoids spurious short-range Coulomb effects. In option L2 of the link atom scheme,[44] all QM atoms and link atoms experience the influence of external charge according to eq 4 so that this option is the closest link atom analogue to the connection atom approach; in contrast, option L1 includes all MM charges (also from CG) but excludes any interactions between MM charges and link atoms.[5,44,46]

A QM calculation with the given Hamiltonian, eq 4, provides a pure QM energy $E^{\mathrm{B}}_{\mathrm{QM}}(\mathrm{Y{-}CA})$ that is lower than $E^{\mathrm{A}}_{\mathrm{QM}}(\mathrm{Y{-}CA})$ due to the polarization of the QM region induced by the MM point charges. It contains the Coulomb interaction energy between the QM and MM regions:

$$E^{\mathrm{coul}}_{\mathrm{QM}}(\mathrm{X,Y{-}CA}) = \sum_{\mu\nu}^{\mathrm{Y-CA}} \sum_{\mathrm{J}\notin \mathrm{CG}}^{\mathrm{X}} P_{\mu\nu} q_{\mathrm{J}} V^{\mathrm{J}}_{\mu\nu} + \sum_{\mathrm{A}}^{\mathrm{Y-CA}} \sum_{\mathrm{J}\notin \mathrm{CG}}^{\mathrm{X}} Z_{\mathrm{A}} q_{\mathrm{J}} V^{\mathrm{AJ}} \tag{5}$$

Here, $P_{\mu\nu}$ and Z_{A} are density matrix elements and nuclear charges, respectively. $V^{\mathrm{J}}_{\mu\nu}$ and V^{AJ} denote nuclear attraction integrals and Coulomb repulsion terms that describe the interaction between a unit charge at MM atom J and an electron or core A in the QM region, respectively. In eq 5, the QM summations include the connection atoms, while the MM summations exclude the charge in the connection groups. Having incorporated the Coulomb interactions from eq 5 into $E^{\mathrm{B}}_{\mathrm{QM}}(\mathrm{Y{-}CA})$, they must not appear in other energy terms to avoid double counting. This is ensured by the following definitions for model B:

$$E^{\mathrm{B}}_{\mathrm{tot}} = E^{\mathrm{B}}_{\mathrm{MM}}(\mathrm{X{-}CA}) + E^{\mathrm{B}}_{\mathrm{QM}}(\mathrm{Y{-}CA}) + E^{\mathrm{B}}_{\mathrm{QM/MM}}(\mathrm{X,Y}) \tag{6}$$

$$E^{\mathrm{B}}_{\mathrm{MM}}(\mathrm{X{-}CA}) = E^{\mathrm{A}}_{\mathrm{MM}}(\mathrm{X{-}CA}) - E^{\mathrm{coul}}_{\mathrm{MM}}(\mathrm{X,CA}) \tag{7}$$

$$E^{\mathrm{B}}_{\mathrm{QM/MM}}(\mathrm{X,Y}) = E^{\mathrm{A}}_{\mathrm{QM/MM}}(\mathrm{X,Y}) - E^{\mathrm{coul}}_{\mathrm{MM}}(\mathrm{X,Y}) \tag{8}$$

The explicit QM/MM interaction term, eq 8, thus contains only the bonded and the nonbonded van der Waals force field contributions from model A, see eq 3.

In model B of the link atom scheme 5, the atom replaced by the link atom is again treated as a regular MM atom, which inplies $E^{\mathrm{B}}_{\mathrm{MM}} = E^{\mathrm{A}}_{\mathrm{MM}}$. Concerning $E^{\mathrm{B}}_{\mathrm{QM/MM}}(\mathrm{X,Y})$, the definition is analogous to eq 8, and the distinction between the link atom and connection atom treatments is the same as for model A (see above).

In addition to electronic embedding, coupling model C (eq 5) accounts for MM polarization:

$$E^{\mathrm{C}}_{\mathrm{tot}} = E^{\mathrm{B}}_{\mathrm{tot}} - \frac{1}{2} \sum_{\mathrm{J}\notin \mathrm{UP}}^{\mathrm{X}} \sum_{\alpha} \mu^{\mathrm{J}}_{\alpha} \langle F^{\mathrm{J}}_{\alpha} \rangle \tag{9}$$

where $\mu^{\mathrm{J}}_{\alpha}$ and $\langle \mathrm{F}^{\mathrm{J}}_{\alpha} \rangle$ denote the components of the induced atomic dipole moment and the QM electric field, respectively. In the chosen classical treatment, short-range polarization is damped in an empirical manner,[48] and MM atoms at the QM/MM boundary are considered unpolarizable (UP) to avoid unreasonably large polarization effects. In the link atom scheme, the only unpolarizable atom is the one replaced by the link atom (e.g., C_2 in Figure 1), whereas all atoms bound covalently to the connection atom are unpolarizable in the present approach (e.g., H_4, H_5, H_6 in Figure 1). Conceptually, the parametrized connection atom should effectively absorb any polarization effects within the methyl group that it replaces, and therefore models B and C are equivalent in the connection atom treatment for methyl compounds $\mathrm{Y{-}CH_3}$.

This completes the definition of coupling models A−C when connection atoms are used. Compared with the link atom scheme, the QM/MM division has essentially been shifted from the QM/MM bond being cut to the connection atom (see Figure 1). The connection atom has a dual character. Generally speaking, it acts as an MM carbon atom in the MM region and as a specially parametrized QM atom in the QM region. By parametrizing the connection atom to mimic a methyl group, one can effectively extend the QM region compared with the link atom scheme. Therefore, fewer force field terms are required to describe the QM/MM interaction in general, and only more distant MM polarization effects need to be included in model C. Moreover, due to the absence of extra atoms, the potential surface is well-defined in the connection atom treatment.

3. Parametrization

The success of the present approach will largely depend on how well the connection atom can simulate the electronic and geometrical behavior of a methyl group (in the sense of a united pseudoatom). It would seem reasonable to use a pseudohalogen with seven valence electrons and an sp-basis for this purpose,[43,45] but in a semiempirical context an even simpler representation turns out to be feasible, namely a connection atom with one valence electron and a single 2s orbital.

Connection atom parameters have been determined for the standard semiempirical QM methods MNDO,[49] AM1,[50] and PM3.[51] Five parameters are needed in MNDO (orbital exponent ζ_s, one-center one-electron energy U_{ss}, one-center two-electron integral g_{ss}, resonance parameter β_s, repulsion term α). In the case of AM1 and PM3, the parameters for the core repulsion Gaussians are also required. Keeping the standard parameters for all other elements fixed, those for the connection atom were optimized by a nonlinear least-squares procedure.[46]

The reference data for the parametrization (heats of formation, geometries, dipole moments, net atomic charges) were generated by standard MNDO, AM1, and PM3 calculations, respectively. During the course of the parametrization, typically 20−30 methyl compounds were employed as reference molecules, including hydrocarbons, alcohols, ethers, carboxy acids, aldehydes, ketones, and amines; the final reference set contained 29 molecules (see Supporting Information). In the parametrization runs, the methyl groups were replaced by connection atoms whose parameters were adjusted to reproduce the reference data for these compounds. The optimized parameters are listed in Table 1.

The quality of these parameters may be judged by comparing the results for methyl compounds and the corresponding analogues with connection atoms. Such comparisons have been carried out for the 29 molecules of the final training set covering heats of formation, dipole moments, and net connection atom charges, as well as bond lengths, bond angles, and dihedral angles involving the connection atoms. A statistical evaluation of the available data[46] is given in Table 2. Obviously, heats of formation are well reproduced when connection atoms are used (see Tables S1−S3 of the Supporting Information). The mean absolute deviations of 1.3 (MNDO), 1.6 (AM1), and 0.9 (PM3) kcal/mol are considerably lower than the intrinsic errors relative to the experimental values,[49−52] and the maximum absolute deviations within the training set[46] are not excessive (MNDO 3.4, AM1 5.5, PM3 1.7 kcal/mol). Likewise, dipole moments are well reproduced, with mean absolute deviations of about 0.1 D. The net charges of the connection atoms are generally close to those of the methyl groups being replaced; they are uniformly somewhat too low in MNDO and somewhat too high in AM1 and PM3. Concerning the geometric variables, the substitution of a methyl group by a connection atom also leads to relatively minor changes, typically by less than 0.01 Å for bond lengths and 1−2° for bond angles and dihedral angles involving the connection atom (and significantly less for the remainder of the molecule). To be specific, the mean absolute deviations between the C−CA and the corresponding $\mathrm{C{-}CH_3}$ bond lengths amount to 0.007 (MNDO), 0.005 (AM1), and

TABLE 1: Parameters for the Connection Atom

parameter[a]	MNDO	AM1	PM3
ζ_s (au)	0.93258788	1.15644960	1.43824357
α (1/Å)	1.64501728	1.46212100	1.96719427
β_s (eV)	−12.67678925	−11.20857778	−9.94923778
U_{ss} (eV)	−13.37151483	−11.83958268	−11.59537348
g_{ss} (eV)	11.89194026	11.39327472	11.60459808
Core Repulsion Gaussians			
G1			
prefactor		−0.30401893	−4.25077548
exponent (1/Å^2)		5.07141717	6.90952417
width (Å)		1.62230340	1.51754243
G2			
prefactor		0.05480783	4.40160107
exponent (1/Å^2)		5.21442318	5.37194220
width (Å)		1.97508768	1.48927868
G3			
prefactor		−0.25449049	
exponent (1/Å^2)		5.10879333	
width (Å)		2.26466753	
G4			
prefactor		−0.11453638	
exponent (1/Å^2)		4.70889286	
width (Å)		2.92552463	

[a] Standard notation;[49−51] see text.

TABLE 2: Statistical Evaluation of the Deviations between the Results for Methyl Compounds and the Connection Atom Analogues[a]

method	property	N	mean error	mean abs error	std dev
MNDO	heats of formation (kcal/mol)	29	0.5010	1.3217	1.7115
	bond lengths (Å)	29	−0.0006	0.0069	0.0090
	bond angles (deg)	30	−0.3968	1.4867	1.8488
	dihedral angles (deg)	22	0.4486	1.1623	2.5525
	dipole moments (Debye)	29	0.0419	0.0807	0.0999
	atomic charges (e)	29	−0.0459	0.0459	0.0492
AM1	heats of formation (kcal/mol)	29	0.2293	1.5859	2.0747
	bond lengths (Å)	29	0.0009	0.0051	0.0067
	bond angles (deg)	30	−0.6165	1.3908	1.8151
	dihedral angles (deg)	22	0.2404	1.3173	2.4178
	dipole moments (Debye)	29	0.0494	0.1026	0.1236
	atomic charges (e)	29	0.0275	0.0275	0.0294
PM3	heats of formation (kcal/mol)	29	−0.2496	0.8828	1.0912
	bond lengths (Å)	29	−0.0006	0.0082	0.0096
	bond angles (deg)	30	0.0621	0.7023	1.0808
	dihedral angles (deg)	22	0.0191	0.3603	0.4902
	dipole moments (Debye)	29	0.0222	0.0677	0.0820
	atomic charges (e)	29	0.0342	0.0342	0.0360

[a] N comparisons for the final training set.[46]

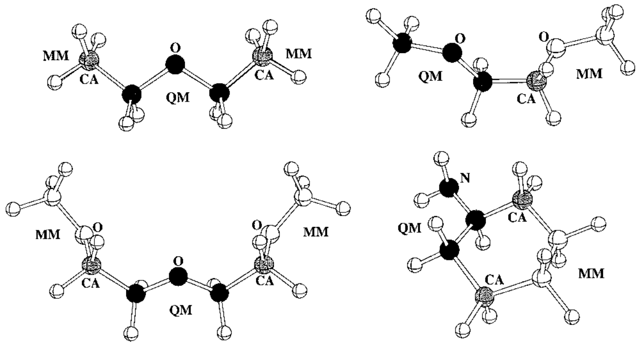


Figure 2. Definition of the QM and MM regions for diethyl ether, dimethoxyethane, di(methoxyethyl) ether, and cyclohexylamine.

0.008 Å (PM3) while the average absolute errors of the computed C−C bond lengths compared with experiment lie in the range 0.013−0.015 Å.

TABLE 3: C−C Bond Lengths (Å) at the QM/MM Boundary Calculated Quantum Mechanically (QM = AM1) and with the Coupled Potential (AM1/AMBER)[a] Using the Link Atom (L2) and the Adjusted Connection Atom (CA) Approach

QM	MM[b]	QM	L2	CA
$HOCH_2$	CH_3	1.512	1.529	1.522
$HOCH_2$	CH_2CH_3	1.519	1.533	1.523
$HOCH_2$	$CH_2CH_2CH_3$	1.519	1.532	1.523
$HOCH_2$	$CH(CH_3)_2$	1.527	1.539	1.526
$HOCH_2$	$CH(CH_2CH_3)_2$	1.528	1.545	1.534
$HOCH_2$	$C(CH_2CH_3)_3$	1.536	1.555	1.543
$HOCH_2$	CH_2CH_2OH	1.518	1.537	1.527
H_2NCH_2	CH_2CH_2OH	1.527	1.539	1.529
H_2NCH_2	$CH_2CH_2NH_2$	1.529	1.533	1.522
diethyl ether[c]		1.510	1.529	1.526
dimethoxyethane[c]		1.525	1.529	1.523
di(methoxyethyl) ether[c]		1.525	1.529	1.525
cyclohexylamine[c]		1.514/ 1.540	1.536/ 1.538	1.518/ 1.541

[a] Standard AMBER parameters: see refs 44 and 53. [b] In the CA approach, the first carbon atom is the connection atom. [c] QM/MM division; see Figure 2.

In an overall view, the results with the connection atoms are quite satisfactory. For the investigated properties, the substitution C−CH_3 → C−CA causes changes that are normally smaller than the intrinsic errors of the underlying semiempirical QM method. This is particularly important in the case of the geometrical variables, because the good agreement that has been found supports our basic assumption that the connection atom can indeed be constrained to coincide with a carbon atom at the QM/MM boundary (Figure 1). This holds for a variety of chemical environments since the training set[46] includes both C(sp^3)−CH_3 and C(sp^2)−CH_3 moieties, with diverse substituents attached to C(sp^3) (e.g., alkyl R, OH, OR, CH=O, COOH, NH_2, CN) and to C(sp^2) (e.g., =CH_2, =CHR, =O).

Finally, it should be emphasized that the parametrization of the connection atom occurs entirely at the QM level. The optimized parameters can therefore be used for QM/MM calculations in combination with any force field.

4. Validation

The preceding section has demonstrated that adjusted connection atoms can successfully simulate methyl groups in Y−CH_3 compounds. In practice, however, they will be employed to connect the QM region Y to a much larger MM region X, e.g., in systems such as Y−CH_2R, Y−CHR_2, or Y−CR_3 where the connection atoms serve to mimic CH_2, CH, or C, respectively. Their performance in this regard has been studied[46] through QM/MM calculations using MNDO and AM1 as QM components and the AMBER4.1 force field.[53] As in previous validation work,[5,44,46] we have chosen the proton affinities of simple organic molecules and the geometries of the corresponding unprotonated and protonated forms as our test cases. There are bound to be strong electrostatic effects upon protonation so that the QM/MM coupling model A (mechanical embedding) is not expected to be adequate. We have therefore only considered coupling models B and C (see section 2).

The selected set of test molecules Y−X consists of six aliphatic alcohols $HOCH_2$−R, three molecules with electronegative groups in the MM region ($HOCH_2$−CH_2CH_2OH, H_2NCH_2−CH_2CH_2OH, H_2NCH_2−$CH_2CH_2NH_2$), three acyclic ethers, and cyclohexylamine. Figure 2 shows the QM/MM division for the latter four molecules; in three of these cases, there are two QM/MM boundaries.

TABLE 4: Absolute[a] and Relative[b] Proton Affinities (kcal/mol) Calculated Quantum Mechanically (QM = AM1) and with the Coupled Potential (AM1/AMBER) According to Model B and Model C Using the Link Atom (L2) and the Adjusted Connection Atom (CA) Approach[c]

				L2		CA	
QM	MM	exp[d]	QM	model B	model C	model B	model C
$HOCH_3$		181.9	170.4				
$HOCH_2$	CH_3	6.4	6.9	0.5	6.3	3.4	3.4
$HOCH_2$	CH_2CH_3	8.9	7.1	3.4	9.4	6.2	8.5
$HOCH_2$	$CH_2CH_2CH_3$	9.2	7.5	2.9	10.0	5.9	8.7
$HOCH_2$	$CH(CH_3)_2$	9.3	8.0	7.2	15.8	9.3	14.8
$HOCH_2$	$CH(CH_2CH_3)_2$		12.4	8.4	19.3	10.2	18.6
$HOCH_2$	$C(CH_2CH_3)_3$		16.0	11.9	25.5	14.4	26.6
dimethyl ether		175.6	175.9				
diethyl ether		8.1	11.3	0.1	8.8	4.0	4.0
dimethoxyethane		12.8	11.0	7.7	14.1	10.1	11.9
di(methoxyethyl) ether		27.3	20.3	15.3	27.5	20.0	23.1
$HOCH_2$	CH_2CH_2OH		190.1	181.5	189.7	184.5	189.3
H_2NCH_2	CH_2CH_2OH	228.6	220.2	215.7	219.8	219.3	221.3
H_2NCH_2	$CH_2CH_2NH_2$	234.1	224.4	219.8	226.1	221.0	224.9
cyclohexylamine		221.2	216.9	217.9	224.2	219.8	222.0

[a] Values above 100 kcal/mol. [b] Values below 100 kcal/mol. [c] See footnotes a−c of Table 3. [d] From ref 54.

For the sake of brevity, we shall focus on AM1/AMBER. The results obtained with the use of adjusted connection atoms will be compared with the available experimental data, the pure AM1 values, and the results from the closest link atom analogue, i.e., option L2.[44,46] More extensive comparisons can be found in ref 46.

Table 3 lists the C−C bond lengths at the QM/MM boundary. The C−CA distances from the connection atom treatment show a mean absolute deviation from the pure AM1 values of only 0.0052 Å (essentially the same as for the methyl compounds $Y-CH_3$ in the training set; see Table 2). The variations in these C−C bond lengths for different molecules are relatively small, but the main trends in the pure AM1 results are reproduced by the CA approach (e.g., see cyclohexylamine). In the link atom scheme, the C−C distance at the QM/MM division is essentially determined by the force field. It is therefore not surprising that the L2 results exhibit larger deviations from the pure AM1 values (0.012 Å on average) and do not capture some of the AM1 trends (e.g., see cyclohexylamine). In addition, they are almost identical for AM1/AMBER and MNDO/AMBER whereas the CA results do indeed reflect the differences between the pure AM1 and MNDO distances.[46] Concerning bond angles and dihedral angles involving the boundary atoms, the CA and L2 approaches yield results of similar quality (deviations of 1−2° from pure AM1[46]), which is as expected since these angles are mostly affected by the force field.

In summary, these tests thus indicate that QM/MM calculations with adjusted connection atoms reproduce the QM geometries in the coupling region very well. The deviations are about as small as for the methyl compounds $Y-CH_3$ in the training set (section 3). This confirms the soundness of the connection atom concept.

Table 4 lists absolute and relative proton affinities for our test molecules Y−X. Generally speaking, the trends in the experimental and AM1 reference data are also found in the AM1/AMBER calculations. This can be seen, e.g., from the relative proton affinities for the alcohols and ethers, and from the absolute values for $HOCH_2-CH_2CH_2OH$, $H_2NCH_2-CH_2CH_2OH$, and $H_2NCH_2-CH_2CH_2NH_2$. In the special case $X = CH_3$ (ethanol, diethyl ether), option L2/B is known to be inadequate[44] while CA/B and CA/C are defined to be identical (see section 2); the resulting irregularities are not considered to be serious since realistic QM/MM applications will always involve MM regions larger than $X = CH_3$.

Closer inspection of Table 4 shows that the results for model C tend to be quite similar for the link atom (L2) and connection atom (CA) treatments. The corresponding trends are normally also apparent at the level of model B, but generally in a less pronounced manner. The differences between the results from models B and C are smaller in the CA than in the L2 approach, as theoretically expected since the CA treatment incorporates more of the QM/MM interactions at the level of model B (see section 2).

Comparison with pure AM1 results indicates that model C usually overestimates the variations in the QM proton affinities, particularly for the branched test molecules (both with L2/C and CA/C), while model B underestimates these variations in the link atom scheme (L2/B). The best overall performance is found for model B with connection atoms (CA/B; see Table 4). Even though one should be cautious about generalizations based on a relatively small number of tests, the present findings suggest that semiempirical QM/MM calculations with unpolarizable force fields should best be performed by using connection atoms and coupling model B (whenever mechanical embedding is not sufficient). This is quite fortunate for practical purposes since QM/MM calculations for model B are significantly less expensive than for model C.[5]

Concerning individual molecules in our test set, it seems noteworthy that the connection atom approach does not face any particular problems with cyclic molecules (cf. cyclohexylamine) or with strongly electronegative substituents in the MM region, even when these are covalently bound to the connection atom (cf. dimethoxyethane, di(methoxyethyl) ether). Likewise, it is possible to have two or more connection atoms as long as they are separated by at least two intervening QM atoms (see Figure 2). However, the use of geminal connection atoms that are bound to the same QM atom is not recommended since this leads to larger deviations from pure QM results.[46] This limitation seems plausible, however, since the spherical connection atoms cannot be expected to capture all of the QM interactions between neighboring nonspherical methyl groups.

5. Conclusions

Connection atoms have been introduced as an alternative to link atoms in semiempirical QM/MM calculations. They may be used to saturate free valencies in the QM region whenever the QM/MM boundary occurs at a C−C single bond. Connection atoms have a dual character because they interact with the other QM atoms as a specially parametrized QM atom, and with the other MM atoms as a standard carbon MM atom. Semiempirical connection atom parameters have been determined for MNDO, AM1, and PM3 by a calibration against theoretical reference data, such that the adjusted connection atoms closely reproduce the geometrical and electronic properties of methyl groups. The corresponding deviations are small enough to validate their use in QM/MM calculations. The available evidence suggests that connection atoms are superior to link atoms both conceptually and numerically.

Acknowledgment. This work was supported by the Schweizerischer Nationalfonds.

Supporting Information Available: Tables S1−S3 contain heats of formation from MNDO, AM1, and PM3 calculations, respectively, with and without the use of connection atoms, for all molecules of the training set. 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) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.

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

(4) Gao, J. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1995; Vol. 7, p 119.

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

(6) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.

(7) Gao, J.; Freindorf, M. *J. Phys. Chem. A* **1997**, *101*, 3182.

(8) Seghal, A.; Shao, L.; Gao, J. *J. Am. Chem. Soc.* **1995**, *117*, 11337.

(9) Furlani, T. R.; Gao, J. *J. Org. Chem.* **1996**, *61*, 5492.

(10) Tunon, I.; Martins-Costa, M. T. C.; Millot, C.; Ruiz-Lopez, M. F. *J. Phys. Chem.* **1997**, *106*, 3633.

(11) Stmad, M.; Martins-Costa, M. T. C.; Millot, C.; Tunon, I.; Ruiz-Lopez, M. F. *J. Chem. Phys.* **1997**, *106*, 3643.

(12) Barnes, J. A.; Williams, I. H. *J. Chem. Soc., Chem. Commun.* **1996**, 193.

(13) Gao, J.; Xia, X. *Science* **1992**, *258*, 631.

(14) Cummins, P. L.; Gready, J. E. *J. Comput. Chem.* **1997**, *18*, 1496.

(15) Freindorf, M.; Gao, J. *J. Comput. Chem.* **1996**, *17*, 386.

(16) Liu, H.; Müller-Plathe, F.; van Gunsteren, W. F. *J. Mol. Biol.* **1996**, *261*, 454.

(17) Merz, K.; Banci, L. *J. Am. Chem. Soc.* **1997**, *119*, 863.

(18) Lyne, P. D.; Mulholland, A. J.; Richards, W. G. *J. Am. Chem. Soc.* **1995**, *117*, 11345.

(19) Cunningham, M. A.; Ho, L. L.; Nguyen, D. T.; Gillian, R. E.; Bash, P. A. *Biochemistry* **1997**, *36*, 4800.

(20) Harrison, M. J.; Burton, N. A.; Hillier, I. H.; Gould, R. *J. Chem. Soc., Chem. Commun.* **1996**, 2769.

(21) Hart, J. C.; Burton, N. A.; Hillier, I. H.; Harrison, M. J. *J. Chem. Soc., Chem. Commun.* **1997**, 1431.

(22) Mulholland, A. J.; Richards, W. G. *Proteins* **1997**, *9*, 27.

(23) Moliner, V.; Turner, A.; Williams, I. H. *J. Chem. Soc., Chem. Commun.* **1997**, 1271.

(24) Thompson, M. A.; Schenter, G. K. *J. Chem. Phys.* **1995**, *99*, 6374.

(25) Thompson, M. A. *J. Chem. Phys.* **1996**, *100*, 14492.

(26) Gao, J. *J. Am. Chem. Soc.* **1994**, *116*, 9324.

(27) Gao, J.; Byun, K. *Theor. Chem. Acta* **1997**, *96*, 151.

(28) de Vries, A. H.; van Duijnen, P. T. *Int. J. Quantum Chem.* **1996**, *57*, 1067.

(29) Gao, J.; Li, N.; Freindorf, M. *J. Am. Chem. Soc.* **1996**, *118*, 4912.

(30) Gao, J.; Alhambra, C. *J. Am. Chem. Soc.* **1997**, *119*, 2962.

(31) Broo, A.; Pearl, G.; Zerner, M. C. *J. Phys. Chem. A* **1997**, *101*, 2478.

(32) Matsubara, T.; Maseras, F.; Koga, N.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 2573.

(33) Yoshida, T.; Koga, N.; Morokuma, K. *Organometallics* **1996**, *15*, 766.

(34) Deng, L.; Woo, T. K.; Cavallo, L.; Margl, P. M.; Ziegler, T. *J. Am. Chem. Soc.* **1997**, *119*, 6177.

(35) Woo, T. K.; Margl, P. M.; Blöchl, P. E.; Ziegler, T. *J. Phys. Chem. B* **1997**, *101*, 7877.

(36) Eichler, U.; Kölmel, C. M.; Sauer, J. *J. Comput. Chem.* **1997**, *18*, 463.

(37) Eichler, U.; Brändle, M.; Sauer, J. *J. Phys. Chem. B* **1997**, *101*, 1003.

(38) Brändle, M.; Sauer, J. *J. Mol. Catal. A* **1997**, *119*, 19.

(39) Thery, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. *J. Comput. Chem.* **1994**, *15*, 269.

(40) Monard, G.; Loos, M.; Thery, V.; Baka, K.; Rivail, J.-L. *Int. J. Quantum Chem.* **1996**, *58*, 153.

(41) Assfeld, X.; Rivail, J.-L. *Chem. Phys. Lett.* **1996**, *263*, 100.

(42) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714.

(43) Hyperchem, Inc. *Hyperchem Users Manual, Computational Chemistry*; Hypercube, Inc.: Ontario, Canada, 1994.

(44) Antes, I.; Thiel, W. In *Combined Quantum Mechanical and Molecular Mechanical Methods*; Gao, J., Thompson, M. A., Eds.; ACS Symposium Series Vol. 712; American Chemical Society: Washington, DC, 1998; p 50.

(45) Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46.

(46) Antes, I. Ph.D. Dissertation; Universität Zürich, Zürich, 1998.

(47) Smith, P. E.; van Gunsteren, W. F. In *Computer Simulations of Biomolecular Systems*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; ESCOM: Leiden 1993; Vol. 2, p 182.

(48) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341.

(49) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.

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

(51) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.

(52) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 221.

(53) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(54) Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Lias, S. G.; Bartmess, J. E.; Mallard, W. G. *J. Phys. Chem. Ref. Data* **1988**, *17*, Supplement 1.