Theor Chem Acc (2003) 109:43–52
DOI 10.1007/s00214-002-0413-3

Theoretical
Chemistry Accounts

Regular article

Evaluation of an ab initio quantum mechanical/molecular mechanical hybrid-potential link-atom method

Patricia Amara, Martin J. Field

Institut de Biologie Structurale – Jean-Pierre Ebel CEA/CNRS/UJF, 41, rue Jules Horowitz, 38027 Grenoble Cedex 1, France

Received: 27 September 2002 / Accepted: 21 October 2002 / Published online: 8 January 2003

Abstract. Hybrid potentials have become a common tool in the study of many condensed-phase processes and are the subject of much active research. An important aspect of the formulation of a hybrid potential concerns how to handle covalent bonds between atoms that are described with different potentials and, most notably, those at the interface of the quantum mechanical (QM) and molecular mechanical (MM) regions. Several methods have been proposed to deal with this problem, ranging from the simple link-atom method to more sophisticated hybrid-orbital techniques. Although it has been heavily criticized, the link-atom method has probably been the most widely used in applications, especially with hybrid potentials that use semiempirical QM methods. Our aim in this paper has been to evaluate the link-atom method for ab initio QM/MM hybrid potentials and to compare the results it gives with those of previously published studies. Given its simplicity and robustness, we find that the link-atom method can produce results of comparable accuracy to other methods as long as the charge distribution on the MM atoms at the interface is treated appropriately.

Keywords: Quantum mechanical/molecular mechanical hybrid potentials – Ab initio quantum chemistry – Link atoms

1 Introduction

Hybrid or combined quantum mechanical (QM)/molecular mechanical (MM) potentials are well adapted for the study of reactions and other processes in large systems. They work by treating the potential of a small region of the system – the region of interest – with a high-precision QM method and the rest of the system – the environment – with less-precise QM methods or empirical MM potentials. Probably the main impetus for the development of hybrid potentials was the desire to simulate reactions in enzymes but they have been applied in other areas as well, notably in solution and surface chemistry [1, 2, 3].

The principal problem in devising a hybrid potential is how to formulate the interactions between regions described with different potentials. For cases where there is one QM and one MM region, there will be non-bonding (electrostatic and, perhaps, Lennard-Jones) interactions between the QM and MM atoms. These are relatively straightforward to deal with. In many situations, however, it is necessary to split a molecule between QM and MM regions, which means that there will be covalent bonds between QM and MM atoms. These "dangling" bonds must be treated in some way because the presence of broken bonds and unpaired electrons at the boundary of the QM region dramatically changes the electronic structure of the QM system.

Although many different models have been developed to deal with covalent bonds at the interface, they can be broadly classified into two types. One class, which we denote as link-atom methods, introduces an extra atom – most often a hydrogen atom – along the broken QM–MM (single) bond at an appropriate distance from the QM atom. The link atom is treated quantum mechanically and serves to sate the unsatisfied valence of the QM atom. The bond itself between the QM and MM atoms is described molecular mechanically. The link-atom method is simple and because of this it has been the most widely used method, especially with semiempirical QM/MM hybrid potentials. It has, however, been criticized on a number of counts, not the least of which is the necessity of introducing extra, unphysical atoms into the system, one for each broken bond.

The second class of approaches are the hybrid-orbital methods, a version of which was used in the first hybrid potential study of an enzyme reaction mechanism by Warshel and Levitt [4]. These methods work by generating a set of hybrid orbitals on the QM atom which has the broken bond. Only orbitals pointing to other QM

Correspondence to: M. J. Field
e-mail: mjfield@ibs.fr

atoms are included in the QM calculation, whereas orbitals pointing to MM atoms are kept frozen. Hybrid-orbital approaches are elegant, because no unphysical atoms need be introduced, but their formulation and implementation are significantly more complex than for the link-atom methods.

Many groups have worked on these various classes of techniques both for semiempirical and ab initio QM/MM hybrid potentials. Thus, for example, hybrid-orbital methods have been developed by Rivail and coworkers (the local self-consistent-field (LSCF) method) [5, 6, 7, 8, 9], by Gao and coworkers (the generalized hybrid orbital method) [10, 11], by Friesner and coworkers [12, 13, 14] and by Kairys and Jensen [15]. Other groups have worked on link-atom methods, including Singh and Kollman (the junction dummy atom method) [16], our laboratory [17, 18], Karplus and coworkers [19, 20, 21], Eichinger et al. (the scaled-position link-atom method) [22], Eurenius et al. [23] and Nicoll et al. [24]. Other types of methods which are closely related to the link-atom approach are the adjusted connection atom model of Antes and Thiel [25], the pseudobond model of Zhang et al. [26], the integrated molecular orbital MM method of Maseras and Morokuma [27, 28], the harmonic cap method of Corchado and Truhlar [29] and the quantum capping potential of DiLabio et al. [30].

In a previous study from our laboratory, a version of the link-atom model developed for semiempirical QM/MM hybrid potentials was described [18]. The purpose of the present study is to see how this model must be modified for use with ab initio QM/MM hybrid potentials and to assess its accuracy in given situations against that of other published ab initio QM/MM models. One of our principal aims has been to seek a widely applicable, simple model that requires as little parameterization as possible.

The outline of the paper is as follows. Our version of the link-atom method is described in Sect. 2, the results of tests using the method are presented in Sect. 3 and Sect. 4 concludes.

2 The link-atom method

The link-atom method that we investigate in this paper is the same as the one described for the semiempirical QM/MM hybrid potential in Ref. [18], except for the way in which the electrostatic interactions at the interface of the QM and MM regions are handled. As full details of the method are given there, we shall only provide a brief outline of the general scheme before going on to detail the differences.

Consider the case of a system divided into two regions, one QM and one MM. The potential energy, E, of this system can be written as a sum of three terms:

$$E = E_{\mathrm{QM}} + E_{\mathrm{MM}} + E_{\mathrm{QM/MM}} \ , \qquad (1)$$

where E_{QM} and E_{MM} are the energies of the QM and MM regions, respectively, and the energy $E_{\mathrm{QM/MM}}$ contains the terms responsible for the interaction between the two regions. The latter will contain terms for the nonbonding (electrostatic and Lennard-Jones) interactions between QM and MM atoms as well as the terms that we are concerned with here and which are necessary for the treatment of the covalent bonds across the QM/MM interface. It is worth emphasizing that, in general, E_{QM} (and also E_{MM} for nonpairwise additive force fields) will not be equivalent to the energy of the isolated QM subsystem because the charges of the MM atoms will polarize the QM charge distribution and so give a modified energy.

A schematic of the link-atom method that we employ is shown in Fig. 1. For each broken bond between a QM atom, $\mathcal{Q}$, and an MM atom, $\mathcal{M}$, across the interface, we add a link atom, $\mathcal{L}$, along the covalent bond between $\mathcal{Q}$ and $\mathcal{M}$. We always use a hydrogen as the link atom, although other elements have also been tried [25, 26], and, for reasons previously explained [17], we partition the system so that only unconjugated single bonds are broken at the interface. In our programs, the two atoms $\mathcal{L}$ and $\mathcal{M}$ are not independent but together constitute what we call a boundary atom, $\mathcal{B}$.

The boundary atom, $\mathcal{B}$, is treated using both QM and MM potentials. $\mathcal{L}$ is the QM part of $\mathcal{B}$, and enters the QM calculation (as a hydrogen), whereas $\mathcal{M}$ is the MM part. Only the coordinates of $\mathcal{M}$ are ever stored and those of $\mathcal{L}$ are constructed, whenever they are needed, by placing $\mathcal{L}$ at an appropriate, fixed, distance from the atom $\mathcal{Q}$ along the $\mathcal{Q}$–$\mathcal{M}$ bond. This distance is a parameter which is chosen at the start of the calculation although we normally use values of 1.1 Å if $\mathcal{Q}$ is a carbon and of 1.0 Å if it is a nitrogen or an oxygen. As the coordinates of $\mathcal{L}$ are expressible as an analytic function of those of $\mathcal{Q}$ and $\mathcal{M}$ it is straightforward to transfer any forces on $\mathcal{L}$ to $\mathcal{Q}$ and $\mathcal{M}$ using the chain rule. This way of handling the link atom has the great advantage that there is no need for users to add extra atoms as link atoms or to employ constraints to keep the link atoms in reasonable configurations because this can be easily done automatically by the program. It should be noted that Eichinger et al. independently derived a similar approach [22].

As already mentioned, the atom $\mathcal{L}$ enters the QM calculation but it remains to specify exactly how $\mathcal{L}$ and $\mathcal{M}$ interact with the other atoms in the system. A complete specification of which interactions are calculated for all atoms is given in Table 1. The terms involved in these interactions are of three types – covalent MM terms, nonbonding MM Lennard-Jones interactions and electrostatic interactions. For the covalent terms, we follow the scheme of Eurenius et al. [23] and exclude all covalent MM interactions (bonds, angles, dihedrals, etc.) which involve only QM atoms or boundary atoms. The sole exceptions to this are bond terms between QM and boundary atoms (i.e. bonds of type $\mathcal{Q}$–$\mathcal{M}$) which must be kept. The Lennard-Jones interactions involving the atom $\mathcal{M}$ are dealt with as in a normal MM calculation and this atom retains its usual Lennard-Jones parameters. In contrast, atom $\mathcal{L}$ has no Lennard-Jones terms.

Up to now, the scheme we have presented is the same as the one published previously [18]. The differences reside in the electrostatic interactions. In our semiempirical hybrid potential, we normally

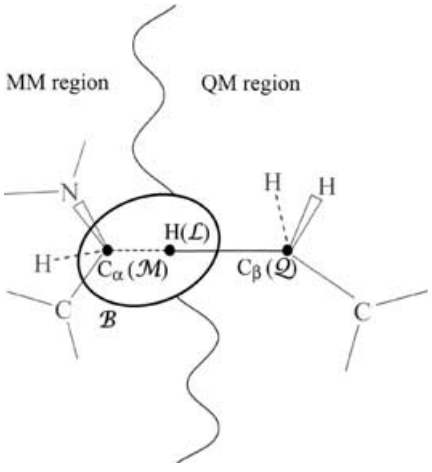

Fig. 1. The link-atom scheme for the case of an amino acid in a protein whose side chain is treated as quantum mechanical (*QM*) and the protein backbone as molecular mechanical (*MM*). The partitioning occurs across the C_β–C_α bond where C_β is QM ($\mathcal{Q}$) and C_α is MM ($\mathcal{M}$). The link atom is denoted $\mathcal{L}$, whereas $\mathcal{B}$ refers to the boundary atom which comprises both $\mathcal{L}$ and $\mathcal{M}$

Table 1. A summary of the different interaction terms calculated between quantum mechanical (*QM*) and molecular mechanical (*MM*) atoms in our hybrid potential. In this table, QM in the columns atom set 1 and atom set 2 refers to an atom in the QM region, other than those of type $\mathscr{Q}$ and $\mathscr{L}$, whereas MM refers to an atom in the MM region, other than those of type $\mathscr{M}$. The atom $\mathscr{L}$ has no MM interactions and only interacts electrostatically with other atoms in the system. Note that calculation of the QM/MM Lennard-Jones interactions includes nonbonding exclusions, whereas calculation of the QM/MM electrostatic interactions does not (meaning all interactions are determined)

Interactions	Atom set 1	Atom set 2
Self interactions		
MM/MM	MM + $\mathscr{M}$	MM + $\mathscr{M}$
QM/QM	QM + $\mathscr{Q}$ + $\mathscr{L}$	QM + $\mathscr{Q}$ + $\mathscr{L}$
Hybrid QM/MM interactions		
Bonds	$\mathscr{Q}$	$\mathscr{M}$
Angles, dihedrals and impropers	QM + $\mathscr{Q}$ + $\mathscr{M}$	MM
Lennard-Jones	QM + $\mathscr{Q}$	MM + $\mathscr{M}$
Electrostatics	QM + $\mathscr{Q}$ + $\mathscr{L}$	MM + $\mathscr{M}$

include all electrostatic interactions between QM atoms and the MM atoms' partial charges except for those between the MM atoms and the atom $\mathscr{L}$. This exclusion has caused some debate [20] but, in our tests, we have seen little advantage to adding these terms, especially as it becomes necessary to change the value of the covalent MM terms for the bonds of type $\mathscr{Q}$–$\mathscr{M}$ at the interface so as to maintain reasonable bond distances.

In contrast, however, we thought it worthwhile to investigate more consistent approaches for an ab initio hybrid-potential link-atom model. To this end we have developed a simple model in which all electrostatic interactions between QM (including link atoms $\mathscr{L}$) and MM atoms are calculated but in which the charge distributions of MM atoms adjacent to boundary atoms and the MM parts of the boundary atoms themselves (atoms $\mathscr{M}$) are represented by Gaussians instead of the usual δ functions. Thus, the charge distribution, ρ_{MM}, of an MM atom can be written as

$$\rho_{\mathrm{MM}}(\mathbf{r}) = q_{\mathrm{MM}} \exp\left[-\left(\frac{\mathbf{r}-\mathbf{r}_{\mathrm{MM}}}{\sigma_{\mathrm{MM}}}\right)^2\right] \Big/ \left(\sqrt{\pi}\sigma_{\mathrm{MM}}\right)^3 , \qquad (2)$$

where q_{MM}, σ_{MM} and $\mathbf{r}_{\mathrm{MM}}$ are the charge, width and center of the Gaussian distribution, respectively, and $\mathbf{r}$ is the position. The choice of a Gaussian form for the charge distribution is arbitrary, but expedient, and serves to soften the interactions between atoms which are only a short distance apart. In practice, it only proved necessary to use these smoothed charge distributions for the MM parts of the boundary atoms (atoms $\mathscr{M}$) and for the MM atoms directly bound to them. In what follows, we shall denote the Gaussian-width parameter used for the charge distribution of the atoms, $\mathscr{M}$, as σ_1 and that for the MM atoms bound to them as σ_2. Values of zero indicate that no smoothing has been performed and so the charge distribution is a δ function.

Before terminating the discussion of our method, we feel that it should be put in perspective with other work on ab initio hybrid potentials. The idea of using smeared charge distributions for the MM atoms has been suggested by others in the context of link-atom methods [22, 23, 24] but, to our knowledge, no comprehensive test of its effectiveness has, as yet, been presented in the literature. Other groups, who do not smooth the MM charge distribution at the interface, have resolved the problem of large QM/MM interactions in the link-atom region by simply zeroing the charges on the MM atoms [19, 23, 26] but this appears to us unsatisfactory as important electrostatic interactions could be missed.

The link-atom method described has been put into an ab initio hybrid-potential version of the DYNAMO program [31]. Implementation is straightforward and requires little modification of the parts of the program that deal with the calculation of the QM and MM energies. Derivatives of the energy – in particular the forces – are also easy to implement and may be determined analytically. For the tests in this paper, the QM calculations were performed at the Hartree–Fock (HF) level using either a 3-21G or a 6-31G* basis set [32] and the OPLS-AA force field was employed as the MM potential [33]. It is our normal strategy to use the terms from a force field unchanged in our hybrid potentials. The only exception to this is that, for certain QM/MM partitionings, it is necessary to modify the charges on the MM atoms at the QM/MM interface so as to maintain zero or integral charge in the MM region. In the tests presented here, this was not required and so all parameters were taken directly from the OPLS-AA force field. Minimizations were performed using a Broyden–Fletcher–Goldfarb–Shanno algorithm [34]. Although we report only HF results, the comparisons we obtain between the full QM and hybrid QM/MM results are very similar when we perform density functional theory calculations with the same basis sets and the BLYP or B3LYP functionals [35].

3 Results

In this section we present the results of calculations for various parameterizations of our link-atom method. A criticism that we have of other work done in this area, with one or two notable exceptions, is that tests of link-atom or hybrid-orbital methods are done using only a very small number of model systems. To avoid falling into the same trap, we have gathered together as comprehensive a series of tests as possible, including many that have been used by other workers in the field.

Full details of our calculations are given later but we found that six parameterizations of our link-atom method were useful, two models with Gaussian smoothing that were representative of all those that we tried, and four point-charge models that we employed for comparison purposes (even though some of them did not give very good results). For convenience, we summarize them here:

- $\mathrm{LA_g}$(4,0) A Gaussian smoothing model with (σ_1, σ_2) = (4.0 Å, 0.0 Å) (i.e. smoothing is only used for the atoms $\mathscr{M}$).
- $\mathrm{LA_g}$(4,3) A Gaussian smoothing model with (σ_1, σ_2) = (4.0 Å, 3.0 Å) (i.e. smoothing is used for the atoms $\mathscr{M}$ and their first neighbors).
- $\mathrm{LA_f}$ A point-charge model in which all QM/MM electrostatic interactions are calculated.
- LA As for model $\mathrm{LA_f}$ except that the electrostatic interactions that occur within each boundary atom $\mathscr{B}$ between the nuclei of atoms $\mathscr{L}$ and the MM charges of atoms $\mathscr{M}$ are omitted.
- $\mathrm{LA_0}$ As for model LA, except that the MM charges of the boundary atom and the MM hydrogens to which it is bound are set to zero.
- $\mathrm{LA_0'}$ As for model LA, except that the MM charges of the boundary atom and the MM hydrogens to which it is bound are excluded from the QM calculation. In other words, they interact with other MM charges but not with the QM atoms.

3.1 Ethane

As our simplest test, we started with ethane, which nicely illustrates the need for a proper treatment of the electrostatic interactions. In all cases, one methyl group

was treated as QM, at the HF/6-31G* level of theory, and the other methyl group as MM with the OPLS-AA force field. The geometries of ethane with each different link-atom model were optimized and the dipole and the vibrational frequencies of the resulting structures were determined.

Take first the results of calculations in which no charge smoothing is applied and the MM charges are point charges. For model LA_f, in which all electrostatic interactions between QM and MM atoms are calculated, the structure collapses. This is due to the very short range interaction between the link-atom (atom $\mathscr{L}$) nucleus and the MM charge on atom $\mathscr{M}$. If this interaction is excluded (to give model LA), the structure optimizes but the C–C bond is much too long at 1.670 Å and the dipole moment of the molecule has a value of 0.733 D. This situation is reminiscent of that observed for semi-empirical hybrid-potential link-atom models [20]. Zeroing the charge of the MM atoms (to give model LA_0) greatly improves the situation as the QM calculation is equivalent to that of a (slightly-distorted) methane molecule. The C–C bond now has a length of 1.530 Å, compared to the pure HF/6-31G* and OPLS-AA values of 1.527 and 1.531 Å, respectively. The dipole moment is also much smaller at 0.006 D.

The use of Gaussian charge distributions also gives good results. For ethane, it is only necessary to make the charge distribution of the MM methyl group carbon a Gaussian as the results are insensitive to the widths of the Gaussians on the MM hydrogens. Good values of the C–C bond length and the ethane dipole moment are found for σ_1 values ranging from 3.5 to 6.0 Å, but below this the C–C bond length becomes too short and the dipole moment is too large. Remaining aspects of the structure are also in good agreement with the pure QM or MM results. The angles are accurate to within 0.2–2.0° which is comparable to the results found by Reuter et al. [20] at the semiempirical level and slightly better than those of the LSCF method.

The vibrational frequencies of ethane in the range 800–1700 cm^{-1} are shown in Fig. 2. The experimental, hybrid-potential, pure QM and pure MM values are illustrated. Two levels of QM theory were used for the hybrid-potential and the pure QM calculations and the $LA_g(4,3)$ parameterization was employed with the hybrid potential. We show only unscaled frequencies as our principal interest is not the accuracy of the calculated results compared to experiment but how the hybrid-potential results compare to the pure QM or MM values. The main result is that the hybrid-potential results are intermediate between the pure QM and MM values (although for the most part nearer to the MM results), indicating that the link-atom method introduces no undue distortions in the values of the frequencies (particularly that of the C–C bond). Similar results were reported in Ref. [22] but, in contrast to that work, we do not adjust any of the hybrid potential's MM parameters.

3.2 Butane

Geometry optimization of the butane molecule with the hybrid-potential model, in which half the molecule is treated as QM and the other half as MM, gives similar results to that of ethane. Thus, the model LA gives a central C–C bond that is much too long (no matter what the QM method used) but this is redressed in the model LA_0 in which the charges on the MM CH_2 group adjacent to the broken bond are set to zero. Good structural results are also obtained with both Gaussian charge models.

Adiabatic energy profiles for rotation about butane's central C–C bond, calculated with pure QM, pure MM and hybrid-potential methods, are shown in Fig. 3, and the energies of certain structures, relative to the most stable structure (which has a C–C–C–C torsion, ϕ_{CCCC}, with a value of 180°) are listed in Table 2. The curves for the pure QM and pure MM results are both very similar.

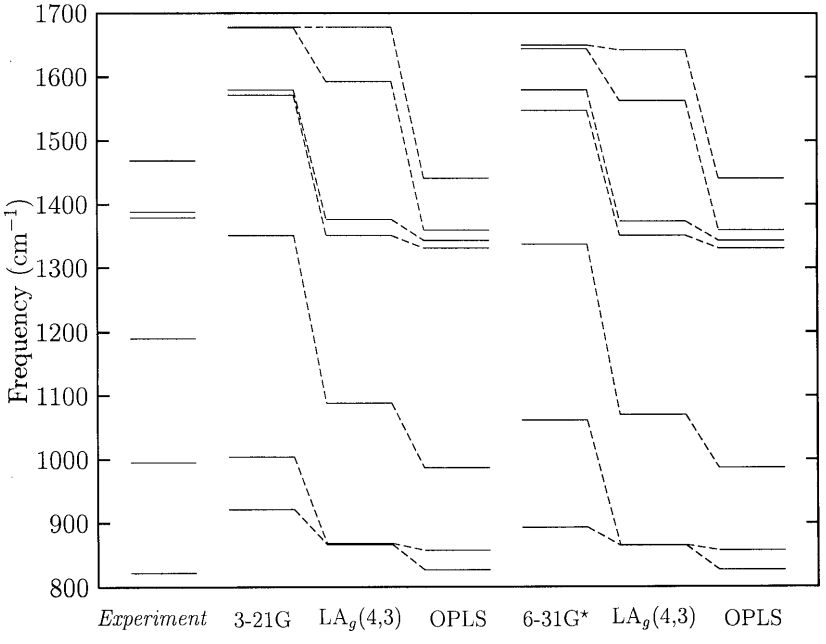

Fig. 2. The vibrational spectra of ethane in the range from 800 to 1700 cm^{-1}. The experimental results are shown together with values determined from full QM calculations (the HF/3-21G and HF/6-31G* methods), from the corresponding hybrid potential using the $LA_g(4,3)$ link-atom model and from full MM (OPLS-AA) calculations

The profile for the LA model is also in good agreement except near the secondary minimum ($\phi_{CCCC} \sim 60°$) where there is an error of about 2 kJ mol^{-1}. Zeroing the charges on the MM CH_2 group adjacent to the central C–C bond (via either model LA_0 or model LA_0') does not help matters as the whole curve moves away from the QM and MM results. A similar tendency is found for the $LA_g(4,0)$ model. In contrast, smoothing the charge distributions of the MM atoms one bond away from the central C–C bond by using the $LA_g(4,3)$ model produces the best results compared to the QM and MM curves.

3.3 Dipole moments

To test the ability of our models to reproduce dipole moments, we geometry-optimized a set of 12 molecules using each of our hybrid-potential models and compared their geometries and dipole moments to those obtained with full QM calculations. The calculations were performed with both the HF/3-21G and HF/6-31G* QM methods. The results are presented in Table 3.

The differences in the geometries, given by the root-mean-square (RMS) error in Table 3, are small and range from 0.04 to 0.14 Å. Except for the long bond at the QM/MM interface, the structures are well reproduced by the LA model with a total RMS deviation of 0.08 Å and 0.04 Å for the HF/3-21G and HF/6-31G* QM methods, respectively. The RMS error for the geometries is somewhat higher for the other models and goes up to 0.14 Åfor the $LA_g(4,3)$ model with the HF/3-21G QM method. This large value is caused by three molecules of the subset for which the RMS error ranges from 0.3 to 0.8 Å and which is due to differences in hydrogen positions. The RMS dipole error, which is about 1.08 D for the LA model, is reduced if the charges on MM atoms are zeroed (in the models LA_0 and LA_0') and is reduced even further if charges are smoothed. The lowest RMS dipole errors are about 0.6 D and are obtained with the $LA_g(4,0)$ and $LA_g(4,3)$ models with the HF/6-31G* QM method.

Table 3 presents results for cases in which the QM/MM partitioning leads to an ethyl group being treated as MM. As expected, partitioning the system further away from the functional group and putting only a methyl group in the QM region reduces the error significantly. Thus, the RMS dipole errors for the $LA_g(4,0)$ and $LA_g(4,3)$ link-atom models with the HF/6-31G* QM method fall to 0.20 and 0.31 D, respectively, whereas the RMS errors in the geometries become 0.05 and 0.01 Å.

Table 2. Energies of different conformations of butane taken from the torsional barriers of Fig. 3 (except for those of model LA_0). All energies are given relative to those of the structures with C–C–C–C dihedral angles (ϕ_{CCCC}) of 180 °. Angles are in degrees and energies are in kilojoules per mole. HF/6-31G* was the QM method used for the pure QM and hybrid-potential calculations

ϕ_{CCCC}	6-31G*	OPLS-AA	LA	LA_0	$LA_g(4,0)$	$LA_g(4,3)$
0	25.9	25.3	25.4	28.2	28.9	24.7
60	4.2	4.9	6.1	7.0	6.9	4.3
120	15.0	15.1	14.9	16.4	16.1	15.3

3.4 Proton affinities

The results of proton affinity calculations for a series of alcohols, carboxylic acids and amines are listed in Table 4. Both the HF/3-21G and the HF/6-31G* QM methods were used. The proton affinities were calculated as the differences in energies between the geometry-optimized protonated and unprotonated forms of the molecules. For each molecule, the link-atom calculations were performed with various partitionings so that the effect of increasing the size of the QM region could be assessed. The results with the model LA are not given because, although the values of the proton affinities are

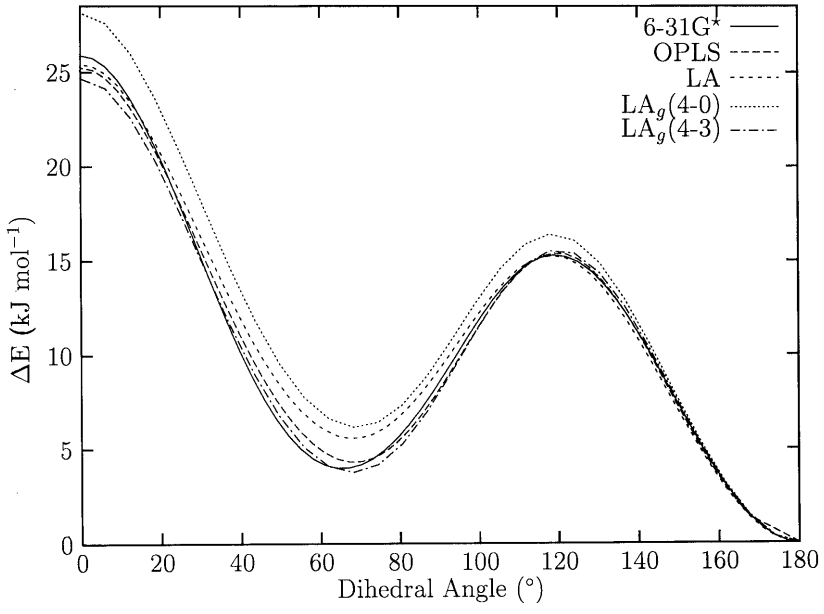

Fig. 3. The torsional barrier of butane about its central C–C bond calculated with full QM, full MM and a variety of link-atom method parameterizations. HF/6-31G* was the QM method used for the full QM and hybrid-potential calculations

Table 3. The dipole moments of a set of 12 test molecules calculated with different methods. Columns 2 (3-21G) and 8 (6-31G*) are the full QM results, whereas columns 3–7 and columns 9–13 are hybrid-potential results obtained with the HF/3-21G and HF/6-31G* QM methods, respectively. Parts of the molecule treated as MM in the hybrid-potential calculations are given in *italics*. The bottom of the table lists the root-mean-square (*RMS*) errors for the dipoles and the geometries of the link-atom results with respect to the full QM calculations. Figures in *bold* indicate the best values in each row. Dipoles are in debye and geometries are in angstrom

Molecule	3-21G	LA	LA_0	LA_0'	$LA_g(4,0)$	$LA_g(4,3)$	6-31G*	LA	LA_0	LA_0'	$LA_g(4,0)$	$LA_g(4,3)$
$CH_3CH_2(CH_2)_2NH$-	10.30	12.24	11.67	11.62	11.25	**11.23**	10.43	12.23	11.72	11.65	11.36	**11.33**
$CH_3CH_2(CH_2)_2NH_2$	1.33	1.59	1.27	1.48	**1.37**	1.44	1.43	1.71	1.60	**1.38**	**1.48**	1.55
$CH_3CH_2(CH_2)_2\ NH_3^+$	9.27	9.72	9.86	**9.79**	10.04	10.03	9.36	9.72	**9.87**	9.90	10.11	10.12
$CH_3CH_2CH_2COO$-	39.99	40.10	40.13	40.07	40.08	**40.04**	39.98	40.10	40.07	40.13	40.07	**40.04**
$CH_3CH_2CH_2COOH$	1.48	1.58	1.90	1.53	1.62	**1.51**	1.61	1.68	1.66	2.02	1.76	**1.65**
$CH_3CH_2CH_2NH$-	42.70	41.14	**40.97**	40.93	40.81	40.82	40.34	41.07	40.91	40.88	**40.78**	40.80
$CH_3CH_2CH_2NH_2$	1.32	1.40	1.55	1.35	1.36	**1.34**	1.40	1.47	1.45	1.64	1.48	**1.44**
$CH_3CH_2CH_2NH_3^+$	43.11	43.59	**43.61**	43.72	43.71	43.96	43.29	43.73	43.76	**43.65**	43.92	43.89
$CH_3CH_2CH_2O$-	6.33	8.92	8.48	8.24	7.98	**7.93**	6.43	8.72	8.22	8.24	7.95	**7.89**
$CH_3CH_2CH_2OH$	1.86	1.88	2.29	**1.93**	2.04	1.96	1.65	1.62	**1.66**	2.03	1.78	1.69
$CH_3CH_2CH_2OH_2^+$	39.96	40.43	40.47	**40.43**	40.54	40.49	40.88	41.45	40.09	40.06	**40.22**	40.19
$CH_3CH_2CH_3$	0.04	0.69	0.41	0.43	**0.22**	0.24	0.07	0.70	0.44	0.40	**0.22**	0.24
RMS error on dipole	0.00	1.08	0.95	0.89	**0.85**	**0.85**	0.00	0.92	0.73	0.74	0.64	**0.62**
RMS error on geometry	0.00	0.08	**0.06**	0.09	0.08	0.14	0.00	0.04	**0.06**	0.10	**0.06**	0.07

sometimes reasonable, the C–C bond distance between the QM and MM regions is always much too long.

The maximum differences between the link-atom and the full QM results are about 44 kJ mol^{-1} if one CH_2 group is treated as QM, 25 kJ mol^{-1} if there are two QM CH_2 groups and 13 kJ mol^{-1} if there are three QM CH_2 groups, in addition to the functional group. These results confirm those of previous studies [11, 17] which indicate that cutting too close to the functional group gives poorer energetics and also confirm the common received wisdom when performing hybrid-potential calculations that a molecule should be partitioned at least two bonds away from the chemical group of interest. Even though the energetics of the models with the smaller QM regions may be inaccurate, the structures are close to the full QM results. Thus, for example, the proton affinity for $CH_3(CH_2)_2CH_2OH$ is in error by 36.3 kJ mol^{-1} with the $LA_g(4,3)$ model and the HF/6-31G* QM method but this error drops to 3.9 kJ mol^{-1} if full QM calculations are done on the structures obtained with the link-atom method. This means that it could be appropriate in some circumstances to do hybrid-potential geometry optimizations with a small QM region followed by single-point hybrid-potential calculations with more QM atoms.

In their development of a link-atom model for an ab initio hybrid potential, Nicoll et al. [24] calculated proton affinities for various amino acids and presented the results for two of them, serine and histidine. The differences in the proton affinities that they obtained between their hybrid-potential and full QM HF/6-31G* results were 49.8 kJ mol^{-1} for $SerH_\gamma \rightarrow Ser^-$ and -21.0 kJ mol^{-1} for $HisH_\epsilon^+ \rightarrow His$. We repeated their calculations using our LA_0, $LA_g(4,0)$ and $LA_g(4,3)$ hybrid potentials and the same QM method. In these calculations, the side chain of the amino acid was in the QM region and both the protonated and unprotonated forms were fully optimized. Unlike Nicoll et al. no adjustment or reparameterization of the MM charges had to be done upon partitioning, owing to our use of the OPLS-AA force field. The importance of a good model for the charge distribution at the interface is confirmed by the bad performance of the LA_0 model, which gives energy differences of -160.4 and -92.4 kJ mol^{-1} for serine and histidine, respectively. Significant improvements are obtained with the $LA_g(4,0)$ model, which gives energy differences of -82.0 and -33.2 kJ mol^{-1}, respectively, and these reduce even further to 1.2 and -20.1 kJ mol^{-1} with the $LA_g(4,3)$ model. In spite of the good result for serine, it should be pointed out that the size of the QM region is inadequate and that, in actual applications, a partitioning further away from the serine oxygen should be employed.

In a slightly more recent paper, DiLabio et al. [30] also tested their quantum capping potential on the protonation energy of the N_ϵ of histidine (although with a terminating COOH group instead of a COH as in the previous example). The differences in the proton affinities between their hybrid potentials and the full QM HF/6-31G(d) results range from about 5 to 11 kJ mol^{-1}, depending on the capping model they use. In contrast, we obtain errors of 40.3 and 17.3 kJ mol^{-1} with our $LA_g(4,0)$ and $LA_g(4,3)$ link-atom models, respectively.

Finally in this section, we determined the effect of an external charge on proton affinities using tests similar to those performed by Reuter et al. [20], who studied the influence of a sodium ion on the proton affinity and the deprotonation enthalpy (DPE) of propanol with semi-empirical QM and hybrid-potential calculations. In all of their calculations, the sodium ion was treated as a point charge, whereas in the hybrid-potential calculations, the ethyl and CH_2OH groups were in the MM and QM regions, respectively, and all the charges of the MM atoms, apart from the sodium ion, were set to zero. The proton affinities and DPEs were calculated for propanol with six different positions of the sodium ion and the RMS differences in these values between the hybrid-potential and the full QM results were reported for a variety of link-atom and hybrid-orbital models. For the link-atom scheme in which electrostatic interactions were calculated between the link atom and the MM atoms, they obtained RMS deviations of about 22 and

Table 4. Proton affinities for several molecules calculated as the differences in energy between the geometry-optimized protonated and deprotonated forms. Column 1 shows the deprotonated form of the molecule with the part of the molecule treated as MM in the hybrid-potential calculations in *italics*. Columns 2 and 5 are the pure QM results at the HF/3-21G and HF/6-31G* levels, respectively. Columns 3–4 and 6–7 list the hybrid-potential results but the values given are the differences with respect to the full QM results. Energies are in kilojoules per mole

Molecule	3-21G	$LA_g(4,0)$	$LA_g(4,3)$	6-31G*	$LA_g(4,0)$	$LA_g(4,3)$
*CH_3*CH_2O-	−1752.2	16.8	−10.3	−1696.3	18.6	−7.7
*CH_3*CH_2CH_2O-	−1746.0	13.1	−1.9	−1691.7	13.9	−0.6
CH_3*CH_2*CH_2O-		−12.3	−20.5		−7.8	−9.6
*CH_3*$CH_2CH_2CH_2O$-	−1745.5	9.7	1.0	−1690.5	8.9	0.6
*CH_3CH_2*CH_2CH_2O-		0.2	−1.5		1.5	−0.1
*$CH_3CH_2CH_2$*CH_2O-		−9.6	−18.3		−6.4	−14.7
*CH_3*CH_2OH	−876.7	47.0	27.1	−817.6	51.0	29.3
*CH_3*CH_2CH_2OH	−882.5	22.9	9.5	−824.0	23.5	10.2
CH_3*CH_2*CH_2OH		36.5	33.5		40.3	36.5
*CH_3*$CH_2CH_2CH_2OH$	−884.5	12.1	4.8	−826.9	13.1	5.6
*CH_3CH_2*CH_2CH_2OH		14.8	13.5		16.3	14.9
*$CH_3CH_2CH_2$*CH_2OH		37.6	32.1		43.6	36.3
*CH_3*CH_2COO-	−1571.5	14.0	−2.7	−1534.0	17.0	1.1
*CH_3*CH_2CH_2COO-	−1568.9	10.0	−0.3	−1532.4	9.9	0.3
CH_3*CH_2*CH_2COO-		−1.7	−3.6		2.7	1.0
*CH_3*CH_2NH-	−1862.0	−18.7	−9.9	−1808.0	15.6	−8.1
*CH_3*CH_2CH_2NH-	−1842.8	0.1	−12.7	−1797.6	8.8	−4.2
CH_3*CH_2*CH_2NH-		−24.4	−25.9		−14.1	−14.7
*CH_3*$CH_2CH_2CH_2NH$-	−1857.0	7.9	0.4	−1803.6	7.4	0.0
*CH_3CH_2*CH_2CH_2NH-		−0.2	−1.8		0.8	−0.8
*$CH_3CH_2CH_2$*CH_2NH-		−7.7	−15.1		−5.1	−12.5
*CH_3*CH_2NH_2	−1004.8	41.7	21.2	−968.2	41.6	−20.5
*CH_3*$CH_2CH_2NH_2$	−1006.2	22.3	6.7	−972.6	25.2	10.2
CH_3*CH_2*CH_2NH_2		23.9	28.0		27.2	31.1
*CH_3*$CH_2CH_2CH_2NH_2$	−1012.5	11.4	4.3	−976.9	11.5	4.6
*CH_3CH_2*$CH_2CH_2NH_2$		13.1	11.8		13.7	12.6
*$CH_3CH_2CH_2$*CH_2NH_2		34.5	26.4		35.5	27.7

20 kJ mol^{-1} for the proton affinities and DPEs, respectively. For a hybrid-orbital method of LSCF type, they got values of about 33 and 34 kJ mol^{-1}.

We repeated their tests using 20 sodium ion positions which were generated randomly by choosing distances at a minimum (maximum) of 2.5 (5.0) Å from any atom. If all the charges on the MM atoms are zeroed (except for the sodium), the RMS deviations are about 15 and 16 kJ mol^{-1} for the proton affinities and DPEs, respectively, in rough agreement with the values of Reuter et al. For our LA_0 model, in which only the charges on the CH_2 group adjacent to the QM region are set to zero, the deviations are about 7 kJ mol^{-1} (proton affinities) and 27 kJ mol^{-1} (DPEs). The values for the models with Gaussian smoothing are a reasonable 29 and 5 kJ mol^{-1} for the $LA_g(4,0)$ model and 24 and 6 kJ mol^{-1} for the $LA_g(4,3)$ model.

3.5 Alanine dipeptide

The alanine dipeptide is composed of an alanyl residue with an acetyl group on the N-terminus and an *N*-methylamine group on the C-terminus. It was used as a model system by Philipp and Friesner [12] for the parameterization of their hybrid-orbital QM/MM model. Of the six different conformational minima that the dipeptide possesses, they were able to find four with their hybrid-potential model and they obtained good agreement for the relative energies of the structures.

For our calculations, we started with six structures that had been optimized with full QM calculations at the HF/6-31G** and HF/3-21G levels. We then tested two QM/MM partitionings, one in which the system was divided at the N–Cα bond of the alanine (14 QM atoms) and the second where it was divided at the Cα–carbonyl C bond of the alanine (eight QM atoms). In each case the first atom of the bond defines the limit of the QM region, whereas the second atom of the bond marks the start of the MM region. Philipp and Friesner only tested the first partitioning. All our hybrid-potential calculations were done with the HF/3-21G QM method. Results for the energies and the structures of the six conformers obtained with the various methods are presented in Tables 5 and 6.

Table 5 lists the values for the LA model even though the optimized bond length at the QM/MM interface is in poor agreement with the reference QM results. Thus, the length of the N–Cα bond optimizes to around 1.70 Å, compared to a reference value of around 1.45 Å, when

Table 5. Energies for different conformations of the alanine dipeptide relative to the C7eq form. Columns 2 and 3 refer to the full QM results with the HF/6-31G** and HF/3-21G QM methods, respectively. All the link-atom calculations were done with the HF/3-21G QM method. The superscript° indicates calculations in which the dipeptide is partitioned at the N–Cα bond, whereas an asterisk indicates partitioning at the Cα–C bond. The last column lists the full MM results. The values in *parentheses* after the energies are the RMS coordinate deviations for the heavy atoms with respect to the reference HF/6-31G** structures. Energies are in kilojoules per mole and geometries are in angstrom

Conf.	6-31G**	3-21G	LA°	$LA_g^\circ(4,0)$	$LA_g^\circ(4,3)$	LA*	$LA_g^*(4,0)$	$LA_g^*(4,3)$	OPLS-AA
C5	1.67	4.39 (0.14)	14.64 (0.15)	4.60 (0.09)	→C7eq	5.10 (0.15)	3.93 (0.10)	−2.80 (0.10)	5.40 (0.12)
C7eq	0.00	0.00 (0.10)	0.00 (0.20)	0.00 (0.08)	0.00 (0.23)	0.00 (0.07)	0.00 (0.08)	0.00 (0.05)	0.00 (0.08)
C7ax	11.82	11.97 (0.04)	8.62 (0.14)	13.35 (0.10)	10.50 (0.19)	10.42 (0.09)	11.55 (0.07)	18.54 (0.08)	10.63 (0.07)
β_2	10.80	15.40 (0.06)	→C7eq	→C7eq	→C7eq	→C7eq	→C7eq	→C7eq	→C7eq
α_L	19.93	25.19 (0.09)	→C7eq	33.18 (0.16)	→C7ax	40.38 (0.14)	→C7ax	42.93 (0.21)	→C7ax
α'	24.43	31.05 (0.08)	→C5	28.16 (0.22)	27.70 (0.21)	28.87 (0.20)	27.70 (0.19)	30.29 (0.19)	26.94 (0.21)

Table 6. Torsional ϕ and ψ angles for different conformations of the alanine dipeptide calculated with QM, MM and the best hybrid-potential models of Table 5. The angles are in degrees

	6-31G**		3-21G		OPLS-AA		$LA_g^\circ(4,0)$		$LA_g^*(4,0)$	
	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
C5	−157.9	160.3	−167.7	170.6	−143.7	161.5	−159.1	167.1	−147.3	160.0
C7eq	−85.8	78.5	−86.1	68.7	−80.3	67.8	−83.1	70.7	−82.5	70.5
C7ax	75.8	−56.5	76.8	−56.0	68.5	−55.1	70.6	−63.0	70.1	−56.3
β_2	−128.6	23.2	−131.4	28.9	→	C7eq	→	C7eq	→	C7eq
α_L	66.9	29.7	66.4	31.8	→	C7ax	56.7	48.2	→	C7ax
α'	−166.4	−40.1	−176.8	−43.8	−157.3	−62.5	−158.9	−65.9	−156.3	−59.9

the system is partitioned at this bond, whereas, for the second partitioning, the reverse behavior is observed, with the Cα–C having a distance of around 1.42 Å compared to the reference value of around 1.53 Å. The LA model finds only three out of six conformers for the N–Cα partitioning and the relative stability of C5 and C7ax are reversed compared to the full QM or MM OPLS-AA results. In contrast, the model does much better for the second partitioning, finding five out of six conformers, four of them with good energies.

Both the Gaussian link-atom models give good bond lengths for the QM–MM bond at the interface no matter which partitioning is used. Of the two models, $LA_g(4,0)$ does better overall. For the first partitioning it finds five conformers out of six, although the α_L and α' conformers do not have the correct relative energies, whereas it finds four out of six for the second partitioning. The $LA_g(4,3)$ model is less good. It finds only three conformers out of six for the first partitioning and, although it finds five different conformers for the second partitioning, the stability of the C5 and C7eq conformers is reversed. In general, both models reproduce well the geometries of the conformers they find. This can be seen for the $LA_g(4,0)$ model from Table 6. For the $LA_g(4,3)$ model, this can be verified by taking the optimized geometries and performing single-point energy calculations with the HF/6-31G** QM method. Thus, for example, the QM energies of the C5, C7eq and C7ax conformers obtained with the $LA_g^*(4,3)$ model are 5.72, 0.00 and 11.49 kJ mol^{-1}, respectively, in good agreement with the full QM results. Similar calculations on the α_L and α' structures, however, do not reverse their relative stabilities although the energy difference is reduced to 6 from 12 kJ mol^{-1}.

We also tested both partitionings of the dipeptide with the $LA_g(4,0)$ model and the HF/6-31G* QM method. Four conformers were found in each case and the energies were in error by between 0.04 and 2.50 kJ mol^{-1} per structure compared to the full HF 6-31G* results. These errors and those obtained with the HF/3-21G QM method are comparable to those found by Philipp and Friesner. It should be noted that their model, like some of our models and the OPLS-AA calculations, found neither the α_L nor the β_2 conformers.

3.6 Alanine tetrapeptide

The alanine tetrapeptide consists of three alanyl residues along with an N-terminal acetyl group and a C-terminal *N*-methylamine group. Extensive QM calculations have been performed on this system by Beachy et al. [36] and, like the dipeptide, it was used by Philipp and Friesner for testing their hybrid-potential model [12]. As for the dipeptide, we tested two partitions with our hybrid-potential model, one in which the system was cut at the central N–Cα bond (24 QM atoms) and the second at which the cut was at the central Cα–C bond (18 QM atoms). The HF/3-21G QM method was used for all our calculations.

The energetic and structural results we obtained on the ten lowest-energy conformers of the tetrapeptide are shown in Tables 7 and 8. As for the dipeptide, the LA model gives geometries about the QM/MM interface that are distorted but we nevertheless show the energies it gives in Table 7. The two Gaussian link-atom methods are

better than the LA model, with the $LA_g(4,3)$ and the $LA_g(4,0)$ models being best for the first and second partitionings, respectively. The values of the RMS energy differences that we obtain with our link-atom models are higher but compare favorably with those obtained by Philipp and Friesner (around 5.4 kJ mol^{-1}) and are, in any case, similar in magnitude for all our models (except for LA°) to those of the full QM HF/3-21G (6.55 kJ mol^{-1}) and OPLS-AA (7.07 kJ mol^{-1}) results. Torsional angles for the ten conformers are given in Table 8 for the best hybrid model compared to full ab initio and full MM calculations. From the RMS

Table 7. Energies for the ten lowest-energy conformations of the alanine tetrapeptide. Columns 2 and 3 refer to the full QM results with the LMP2 [36] and the HF/3-21G QM methods, respectively. All the link-atom calculations were done with the HF/3-21G QM method. The superscript ° indicates calculations in which the dipeptide is partitioned at the N–Cα bond, whereas an asterisk indicates partitioning at the Cα–C bond. The last column lists the full MM results. All energies (other than those of the LMP2 reference) are obtained by minimizing the RMS energy difference between the calculated and the LMP2 reference structures (see Ref. [12] for details). The last row gives the overall RMS energy difference for all structures. The values in *parentheses* after the energies are the RMS coordinate deviations for the heavy atoms with respect to the reference LMP2 structures. Energies are in kilojoules per mole and geometries are in angstrom

Conf.	LMP2 [35]	3-21G	LA°	$LA^{\circ}_g(4,0)$	$LA^{\circ}_g(4,3)$	LA*	$LA^{*}_g(4,0)$	$LA^{*}_g(4,3)$	OPLS-AA
1	11.34	15.38 (0.13)	25.40 (0.20)	10.38 (0.16)	→4	11.52 (0.33)	11.00 (0.25)	5.61 (0.23)	12.59 (0.27)
2	11.88	13.14 (0.22)	12.30 (0.22)	4.35 (0.26)	−0.13 (1.69)	11.28 (0.27)	7.03 (0.25)	3.10 (0.22)	11.17 (0.23)
3	0.00	−7.05 (0.20)	−14.35 (0.32)	−3.01 (0.22)	−6.02 (0.34)	−5.47 (0.24)	−9.46 (0.28)	−4.35 (0.28)	−5.61 (0.29)
4	17.28	21.31 (0.50)	16.78 (0.52)	8.74 (0.41)	14.23 (0.35)	15.90 (0.28)	16.02 (0.36)	13.47 (0.48)	15.90 (0.17)
5	16.23	21.79 (0.37)	→4	9.37 (0.93)	18.49 (0.80)	27.15 (0.32)	25.40 (0.30)	26.36 (0.42)	19.96 (0.21)
6	9.20	−6.57 (0.43)	7.03 (0.54)	9.75 (0.58)	8.87 (0.71)	17.03 (0.51)	11.25 (0.55)	11.72 (0.60)	−4.60 (0.73)
7	24.14	26.92 (0.13)	25.40 (0.34)	19.12 (0.25)	25.48 (0.31)	21.65 (0.30)	21.13 (0.30)	18.24 (0.27)	18.24 (0.25)
8	17.41	22.11 (0.23)	16.02 (0.58)	20.13 (0.48)	24.48 (0.64)	20.90 (0.41)	18.91 (0.41)	12.09 (0.45)	28.33 (0.42)
9	28.95	26.92 (0.08)	17.66 (0.16)	37.99 (0.12)	20.13 (0.16)	26.15 (0.13)	20.88 (0.12)	33.10 (0.11)	26.34 (0.12)
10	29.25	27.56 (0.13)	38.41 (0.16)	44.73 (0.76)	41.76 (0.16)	15.40 (1.08)	39.29 (0.38)	42.13 (0.81)	38.83 (0.73)
RMS	0.00	6.55	8.03	7.03	6.36	6.66	5.98	6.65	7.07

Table 8. Torsional angles of the alanine tetrapeptide. For each conformer, the first line has the full QM HF/3-21G values along with the HF/6-31G** values in *parentheses*, the second line has the OPLS-AA results and the third one the results with the $LA^{*}_g(4,0)$ model. The structures optimized at the HF/6-31G** level were taken from Ref. [36]. The angles are in degrees

Conf.	ϕ_1	ψ_1	ϕ_2	ψ_2	ϕ_3	ψ_3
1	−167.4 (−158.5)	170.9 (163.5)	−167.9 (−157.8)	171.0 (163.4)	−167.7 (−156.2)	170.3 (160.8)
	−143.9	164.5	−142.4	166.3	−141.6	163.8
	−144.6	165.1	−146.2	163.4	−157.0	160.2
2	−166.9 (−158.6)	170.7 (163.9)	−167.0 (−154.9)	169.2 (158.1)	−85.6 (−86.0)	70.0 (79.2)
	−142.6	162.2	−141.3	164.1	−81.0	70.2
	−144.5	164.5	−143.9	159.8	−86.3	78.9
3	−82.4 (−81.7)	81.3 (93.4)	75.8 (76.3)	−50.5 (−53.4)	−76.1 (−80.4)	84.2 (85.1)
	−78.9	74.0	71.8	−48.8	−73.5	76.0
	−90.5	59.8	67.5	−61.7	−66.3	126.2
4	−166.6 (−156.9)	170.6 (161.3)	−87.2 (−88.8)	68.8 (83.5)	−167.0 (−156.0)	167.0 (152.8)
	−143.0	162.1	−83.8	72.9	−136.2	160.2
	−140.8	163.3	−82.9	76.2	−160.5	146.5
5	−167.1 (−157.2)	176.6 (170.0)	−100.3 (−76.2)	13.5 (−19.6)	−166.2 (−153.8)	169.8 (160.8)
	−139.8	160.0	−73.4	−25.8	−142.6	157.5
	−138.5	160.3	−74.7	−27.3	−147.2	163.7
6	−90.4 (−89.0)	60.6 (67.3)	56.3 (63.0)	26.1 (24.3)	177.3 (−165.0)	162.0 (149.8)
	−89.5	62.1	68.4	−54.3	−89.0	142.9
	−87.7	57.5	57.3	22.7	−180.0	145.6
7	57.5 (56.0)	−165.8 (−158.5)	−86.0 (−93.0)	69.1 (63.8)	−164.8 (−163.3)	−55.1 (−50.0)
	57.3	−150.3	−87.5	57.3	−154.7	−57.0
	55.1	−150.2	−87.6	64.7	−166.8	−52.3
8	71.4 (72.8)	−69.9 (−70.5)	−56.8 (−58.1)	139.6 (134.7)	64.7 (62.0)	18.0 (25.7)
	69.1	−55.2	−63.6	98.2	64.0	34.3
	68.8	−51.8	−64.0	108.8	62.7	30.2
9	74.3 (75.7)	−55.6 (−59.5)	74.2 (76.1)	−54.7 (−55.3)	74.2 (75.5)	−56.8 (−53.0)
	67.8	−53.6	67.3	−52.0	67.7	−52.9
	66.7	−56.2	67.8	−53.7	76.2	−50.3
10	60.7 (62.5)	26.8 (29.0)	61.4 (65.1)	24.4 (20.6)	72.3 (73.8)	−56.4 (−51.5)
	60.6	−6.6	57.8	44.7	71.4	−50.8
	−64.3	87.9	55.0	49.1	70.4	−71.2

geometries in Table 7 and the torsional angles in Table 8, conformers 1–3, 7 and 9 are well reproduced by most of the models. The geometries of conformers 4–6, 8 and 10 are less accurate although the full QM HF/3-21G method also poorly reproduces conformers 4–6 and the OPLS-AA method does badly with conformers 6, 8 and 10.

4 Conclusions

In this paper we have tested a link-atom method for treating covalent bonds between atoms at the QM/MM interface of an ab initio QM/MM hybrid potential and find that it produces results that are not markedly inferior to those of more sophisticated techniques. The method itself is identical to that reported previously for use with semiempirical QM/MM potentials except that the electrostatic interactions between the QM and MM atoms at the interface are calculated differently. We emphasize that our tests were performed without modification of any of the parameters required by the separate QM or MM potentials and that the link-atom method itself introduces only three extra parameters (the bond length between the QM interface atom and the link atom and the widths of the Gaussian charge distributions for the MM atoms at the interface).

How to deal with covalent bonds at the QM/MM interface is only one part in the formulation of the interactions between the different regions of a hybrid potential. Whereas the current interaction models for semiempirical QM/MM potentials are probably satisfactory (at least, for the semiempirical QM methods that are now available), it seems to us important that further research on the "link-atom problem" in the ab initio case be supplemented with improvements in other aspects of the hybrid potential. Areas requiring attention include optimization of QM algorithms so that the size of the QM region can be increased, a better description of the charge distribution of the MM atoms (including polarization effects) and a more satisfactory model for the short-range and van der Waals interactions between the QM and MM atoms. Such work is ongoing in our laboratory and will be reported in due course.

Acknowledgements. The authors thank the Institut de Biologie Structurale – Jean-Pierre Ebel, the Commissariat à l'Energie Atomique and the Centre National de la Recherche Scientifique for support of this work.

References

1. Gao J (1995) In: Lipkowitz KB, Boyd DB (eds) Rev comput chem vol 7. VCH, New York, pp 119–185
2. Friesner RA, Beachy MD (1998) Curr Opin Struct Biol 8: 257–262
3. Amara P, Field MJ (1999) In: Leszczynski J (ed) Computational molecular biology. Elsevier, Amsterdam pp 1–33
4. Warshel A, Levitt M (1976) J Mol Biol 103: 227–249
5. Théry V, Rinaldi D, Rivail J-L, Maigret B, Ferenczy GG (1994) J Comput Chem 15: 269–282
6. Assfeld X, Rivail J-L (1996) Chem Phys Lett 263: 100–106
7. Gorb LG, Rivail J-L, Théry V, Rinaldi D (1996) Int J Quantum Chem 30: 1525–1536
8. Monard G, Loos M, Théry V, Baka K, Rivail J-L (1996) Int J Quantum Chem 58: 153–159
9. Ferré N, Assfeld X, Rivail J-L (2002) J Comput Chem 23: 610–624
10. Gao J, Amara P, Alhambra C, Field MJ (1998) J Phys Chem A 102: 4714–4721
11. Amara P, Field MJ, Alhambra C, Gao J (2000) Theor Chem Acc 104: 336–343
12. Philipp DM, Friesner RA (1999) J Comput Chem 20: 1468–1494
13. Murphy RB, Philipp DM, Friesner RA (2000) J Comput Chem 21: 1442–1457
14. Murphy RB, Philipp DM, Friesner RA (2000) Chem Phys Lett 321: 113–120
15. Kairys V, Jensen JH (2000) J Chem Phys A 104: 6656–6665
16. Singh UC, Kollman PA (1986) J Comput Chem 7: 718–730
17. Field MJ, Bash PA, Karplus M (1990) J Comput Chem 11: 700–733
18. Field MJ, Albe M, Bret C, Proust-De Martin F, Thomas A (2000) J Comput Chem 21: 1088–1100
19. Lyne PD, Hodoscek M, Karplus M (1999) J Phys Chem A 103: 3462–3471
20. Reuter N, Dejaegere A, Maigret B, Karplus M (2000) J Phys Chem A 104: 1720–1735
21. Cui Q, Elstner M, Kaxiras E, Frauenheim T, Karplus M (2001) J Phys Chem B 105: 569–585
22. Eichinger M, Tavan P, Hutter J, Parrinello M (1999) J Chem Phys 110: 10452–10467
23. Eurenius KP, Chatfield DC, Brooks BR, Hodoscek M (1996) Int J Quantum Chem A 60: 1189–1200
24. Nicoll RM, Hindle SA, MacKenzie G, Hillier IH, Burton NA (2001) Theor Chem Acc 106: 105–112
25. Antes I, Thiel W (1999) J Phys Chem A 103: 9290–9295
26. Zhang Y, Lee T-S, Yang W (1999) J Chem Phys 110: 46–54
27. Maseras F, Morokuma K (1995) J Comput Chem 16: 1170–1180
28. Maseras F (2000) Chem Commun 2000: 1821–1827
29. Corchado JC, Truhlar DG (1998) J Phys Chem A 102: 1895–1898
30. DiLabio GA, Hurley MM, Christiansen PA (2002) J Chem Phys 116: 9578–9584
31. Field MJ (1999) A Practical introduction to the simulation of molecular systems. Cambridge University Press, Cambridge
32. Hehre WJ, Radom L, Schleyer PvR, Pople JA (1986) Ab initio molecular orbital theory. Wiley, New York
33. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) J Am Chem Soc 118: 11225–11236
34. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1996) Numerical recipes in Fortran 77, 2nd edn. Cambridge University Press, Cambridge
35. Koch M, Holthausen MC (2000) A chemist's guide to density functional theory. Wiley-VCH, New York
36. Beachy MD, Chasman D, Murphy RB, Halgren TA, Friesner RA (1997) J Am Chem Soc 119: 5908–5920