

Frozen core orbitals as an alternative to specific frontier bond potential in hybrid Quantum Mechanics/Molecular Mechanics methods

Arianna Fornili ^a, Pierre-François Loos ^b, Maurizio Sironi ^{a,c,*}, Xavier Assfeld ^{b,*}

^a Dipartimento di Chimica Fisica ed Elettrochimica, Università degli Studi di Milano, Via Golgi 19, 20133 Milano, Italy

^b Equipe de Chimie et Biochimie Théoriques, UMR UHP-CNRS 7565, Faculté des Sciences et Techniques, Université Henri Poincaré, 54506 Vandoeuvre-lès-Nancy Cedex, France

^c Centre for Bio-molecular Interdisciplinary Studies and Industrial applications (CISI), Via F.lli Cervi, 93 Palazzo L.I.T.A. 20090 Segrate, Milano, Italy

Received 4 April 2006; in final form 19 June 2006
Available online 4 July 2006

Abstract

The use of Frozen Core Orbitals (FCOs) at the frontier atom is proposed as an alternative to the specific frontier bond potential in the Local-SCF/MM method. The calculations illustrate that frontier bond length is correctly recovered without adding any specific force-field potential. It is shown that the same accuracy is reached for calculations using FCOs arising either from an atomic (ROHF) or from a molecular calculation using Extremely Localized Molecular Orbitals (ELMOs). In addition FCOs do not modify the description of the remaining of the QM subsystem from both an energetic and/or geometric point of view.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

Quantum Mechanics/Molecular Mechanics (QM/MM) methods are now well established techniques for the study of complex systems and particularly of enzyme-catalyzed reactions. Although widely used, they cannot be considered as black boxes because the coupling of the two levels of theory presents some problems if the boundary between QM and MM regions is set across covalent bonds. Different ways of representing the frontier regions have been devised, based on the use of link-atoms [1–6], of boundary-atoms (examples are the pseudo-bond [7], the adjusted connection atom [8], the quantum capping potential (QCP) [9] and the generalized hybrid orbital (GHO) [10,11] approaches) or of frontier orbitals [12–15]. A recent mixed

method has also to be mentioned, which combines the link-atom approach and the distributed charge (and/or dipole) in a ‘GHO-like’ philosophy [16].

A critical issue common to all these methods is the accuracy in the description of the geometrical parameters across the boundary. In particular, special measures have to be adopted to avoid too short or too long frontier bonds. In link-atom methods modifications of either the bonded or non-bonded force-field parameters across the boundary [17,18] are used, while in boundary-atom approaches the semi-empirical [8,10] or pseudo-potential [7,9] parameters are adjusted to better reproduce the frontier geometry in model compounds.

In frontier-orbital methods, frontier bonds are described through strictly localized bond orbitals (SLBOs), which are kept frozen during the QM/MM calculation. Thus the frontier atom on the MM side of the boundary (‘Y’ in the following) has one of its electrons explicitly treated at the QM level. The combined effect of replacing the full QM atom with a one-electron Y atom carrying a +1 point charge and of freezing the frontier orbital generally yields too short frontier bonds. In the approach based on the Local SCF (LSCF) method [12], a special five-parameters force-field indepen-

* Corresponding authors. Address: Dipartimento di Chimica Fisica ed Elettrochimica, Università degli Studi di Milano, Via Golgi 19, 20133 Milano, Italy (M. Sironi), Equipe de Chimie et Biochimie Théoriques, UMR UHP-CNRS 7565, Faculté des Sciences et Techniques, Université Henri Poincaré, 54506 Vandoeuvre-lès-Nancy Cedex, France (X. Assfeld).

E-mail addresses: maurizio.sironi@unimi.it (M. Sironi), Xavier.assfeld@lctn.uhp-nancy.fr (X. Assfeld).

dent frontier bond potential has been developed for different frontier atom pairs, which allowed reducing the maximum deviation from full QM calculations to 0.05 Å [14].

In this Letter, we propose an alternative approach to the problem, by introducing in hybrid LSCF/MM calculations the explicit description of the core electrons of Y atoms through frozen core orbitals (FCOs). Analogously to frontier SLBOs, FCOs are determined on model systems and are kept unchanged during the LSCF calculation. Our preliminary calculations suggest that adding FCOs improves significantly the description of frontier bond lengths without the need for specific parameters. FCOs are here tested in conjunction with two different types of SLBOs, derived either from an a posteriori localization method [19] or by minimizing the energy under the constraint of strict localization, which yields the so-called Extremely Localized Molecular Orbitals (ELMOs) [20–25].

2. Methods

The LSCF method has been developed to optimize a wavefunction in the presence of frozen molecular orbitals, which is required when performing frontier orbital-based QM/MM calculations. It can be used at both the HF and post-HF level, as well as with DFT approaches. For further details see Ref. [12,14,26–28].

LSCF calculations were performed on test molecules (ethane, methylamine and methanol) first at the full QM level. SLBOs were calculated on each molecule at the RHF-optimized geometry. Localization was carried out either according to the a posteriori Weinstein–Pauncz (WP) criterion [19,29,30] followed by deletion of the delocalization tails [31] or using the tail-free ELMO wavefunction [20–24]. A comparison of the two strategies has been recently performed, suggesting that the second procedure yields orbitals that are generally more transferable [25]. FCOs were calculated both in a molecular environment at the ELMO level and on isolated atoms at the ROHF level. The localization scheme (i.e. the partitioning of the molecule into fragments) required by ELMO calculations was set up according to the molecular Lewis structure [23].

Subsequently, real QM/MM calculations were carried out on the two different protonation states of the lysine amino acid capped with an acetyl and an *N*-methyl group (see Fig. 1). The AMBER force-field [32] as implemented in GAUSSIAN98 [33] was used, together with atom charges determined according to the RESP protocol [34] on the RHF-optimized protonated form.

The 6-311G** basis set was used throughout this Letter (except for the RESP calculation, which employed the recommended 6-31G* basis set). All LSCF type calculations together with WP localizations were performed with a locally modified version of the GAUSSIAN98 package, while ELMO calculations were carried out with a local code [23] which has been recently interfaced with the GAMESS-UK package [35].

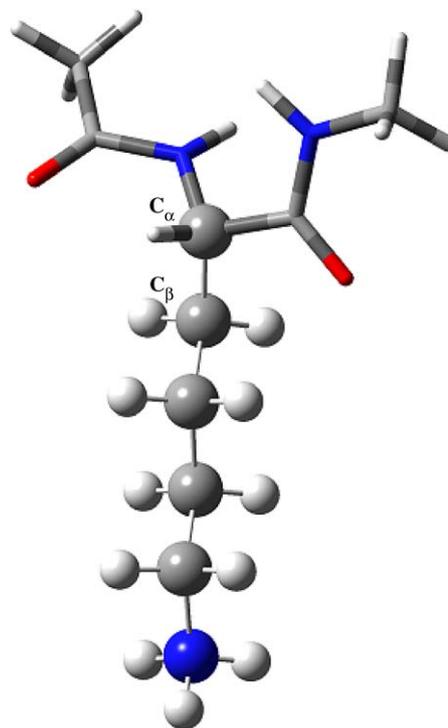


Fig. 1. Partitioning of the Ace-Lys-Nme dipeptide into QM (ball and sticks) and MM (thick sticks) regions.

3. Results and discussion

The effect of considering FCOs in QM/MM calculations was first investigated through pure LSCF calculations on the ethane molecule, where the frontier was set at the C–C bond. The hydrogen atoms of the classical region were removed to determine the influence of the QM wavefunction alone on the structure of the molecule and to devise a force-field independent model. The C_X – C_Y bond (where C_X is on the QM and C_Y on the MM side) and the C_Y core electrons were described through an SLBO and FCO, respectively, which were determined through an ELMO calculation on the complete ethane molecule and kept frozen during the LSCF iterations on the CH_3 –C fragment. In this way three of the C_Y electrons were explicitly included in the QM calculation and a +3 charge was assigned to the C_Y nucleus to keep electroneutrality. In the following, this type of calculation will be referred to as LSCF/+3. Calculations without FCOs were also performed for comparison and they will be referred to as LSCF/+1.

In Fig. 2 the energy profile obtained by rigid scan along the C_X – C_Y distance is reported for LSCF/+3 calculations, together with the LSCF/+1 and the full-RHF results obtained on the complete ethane molecule. It is evident that the LSCF/+3 equilibrium distance is much closer to the RHF value than the LSCF/+1. As previously mentioned, a specific frontier bond potential was developed to adjust the LSCF/+1 energy profile to the SCF one, which accounts for both the atomic orbital overlap varia-

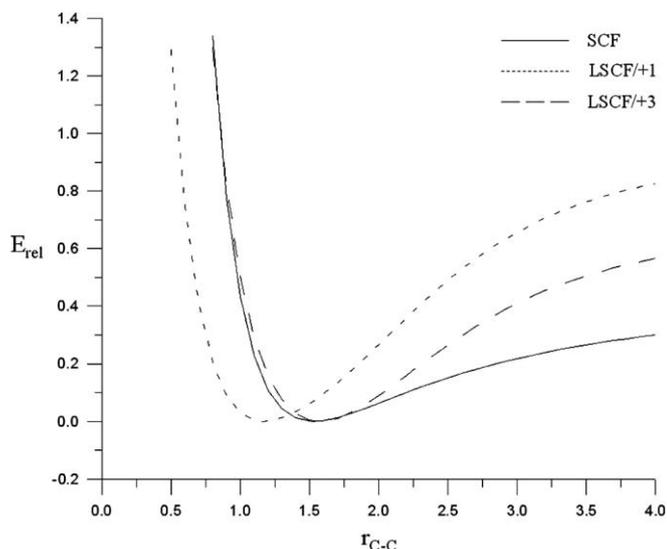


Fig. 2. Energy profile (in a.u.) along the C–C distance (in Å) calculated for ethane at the RHF and for the CH₃–C fragment at the LSCF/+1 and LSCF/+3 levels with the 6-311G** basis set.

tion and the nuclei-ion repulsion [14]. The use of FCOs in the present LSCF/+3 calculation can clearly model only the latter effect and one can see from Fig. 2 that it plays a significant role on the position of the minimum but also on the curvature at the minimum and on the width of the well. Since the frontier bond is not expected to break during QM/MM calculations, the behaviour at large atomic separation is irrelevant. One can conclude that FCOs are enough to obtain correct geometries and no more specific potential is needed.

To assess the dependence of the results from the way the FCO is determined, we tested the core orbital obtained from a ROHF calculation on the isolated carbon atom, but the resulting energy profile was indistinguishable from that obtained using an ELMO FCO, implying that the variation of the core orbital on going from atom to molecule is negligible, at least for our purposes¹.

Moreover, we checked that the introduction of the FCO does not perturb the description of other degrees

¹ Since the ground state of the isolated carbon atom has a non-null angular momentum, the ROHF core orbital obtained using polarization basis functions has not a perfect spherical symmetry (in fact, for cartesian d functions, one of the d_{XX} coefficients, with $X = x, y$ or z , is slightly different from the other two), so that the results obtained using it in a molecular environment depend on its orientation. Even if this dependence is small (in the present calculation reorienting the orbital by simply switching the unequal d_{XX} component with one of the other two in turn yields a variation of at most 0.2 kcal/mol in the minimum energy and of 0.0003 Å in the frontier bond length), using ELMO core orbitals allows to avoid this ambiguity. Another way to solve this problem could be calculating carbon core orbitals without d functions, i.e. using a 6-311G basis set. Also in this case, the variation from results obtained with ROHF/6-311G** core orbitals is negligible (at most 0.3 kcal/mol for the energy and 0.0003 Å for the bond length).

of freedom by determining the LSCF/+3 minimum geometry of the CH₃–C fragment. Results for selected distances and angles and for the overall RMS deviation from the reference RHF structure are reported in Table 1 as LSCF (ELMO)/+3, together with the full-RHF and LSCF (ELMO)/+1 values. One can see that the deviation from RHF results is generally lower for the LSCF(ELMO)/+3 calculations with respect to LSCF(ELMO)/+1 also for the internal coordinates involving only ‘full’ QM atoms.

At last, LSCF/+3 calculations with a SLBO determined according to the WP criterion were performed for comparison purposes. The LSCF(WP)/+3 results in Table 1 are of similar accuracy as the LSCF(ELMO)/+3 ones, even if the frontier bond distance is slightly better reproduced in the latter case.

Analogous calculations were performed on methylamine and methanol to test FCOs when the Y atom is a nitrogen or an oxygen². From Table 1 it is evident that also in these cases using FCOs greatly improves frontier bond lengths. For both molecules WP frontier distances are somewhat better than ELMO ones, but it is the opposite for other degrees of freedom, so that the overall RMS deviation from RHF is similar for the two kinds of SLBOs. Moreover, as for the carbon atom in the calculations on ethane, using a nitrogen or oxygen FCO obtained from ROHF calculations on the isolated atom does not influence significantly the results (data not shown).

Contrary to the ethane case, methyl hydrogens both in methylamine and methanol are not equivalent, as is visible from the different values of the angles H1–C–N(O) and H2–C–N(O) (see Fig. 3 for atom numbering) in Table 1. In LSCF calculations this asymmetry can be completely ascribed to the shape of the SLBO, as in our model the hydrogen atoms of the classical region do not influence the QM region. Moreover, the different values yielded by ELMO and WP SLBOs can be interpreted just on the basis of their different shape. In Fig. 3 the two SLBOs are compared for both molecules through two isosurfaces of the difference $|\varphi_{\text{ELMO}}(\mathbf{r})|^2 - |\varphi_{\text{WP}}(\mathbf{r})|^2$ plotted for the 0.001 (solid) and the –0.001 a.u. (mesh) values. It can be seen that the ELMO is larger than the WP SLBO in the region around the C–Y bond that is opposite to the lone pairs, which then explains why for methylamine we have $a_{\text{H1-C-N}}^{\text{ELMO}} > a_{\text{H1-C-N}}^{\text{WP}}$ and $a_{\text{H2-C-N}}^{\text{ELMO}} < a_{\text{H2-C-N}}^{\text{WP}}$. The same effect is responsible for the valence angles in methanol. These different shapes are expected from the variational nature of ELMOs, which implies that for them the repulsion and then the overlap between the bond and the lone pair orbitals is minimized.

² For these molecules the Lewis localization scheme used in ELMO calculations was modified so that the lone pairs of the heteroatom were delocalized over first-neighbor atoms. This prevented mixing between core and lone pairs, which could happen if the same partial basis set is used for their determination and which then would not yield a proper FCO.

Table 1

Selected distances (in Å) and angles (in degrees) calculated for CH₃-R (R = CH₃, NH₂, OH) using the RHF, LSCF(ELMO)/+1, LSCF(ELMO)/+3 and LSCF(WP)/+3 wavefunctions with the 6-311G** basis set

	RHF	LSCF(ELMO)/+1	LSCF(ELMO)/+3	LSCF(WP)/+3
Ethane				
<i>r</i> _{C-C}	1.527	1.162 (−0.365)	1.571 (0.044)	1.585 (0.057)
<i>r</i> _{C-H}	1.086	1.106 (0.019)	1.097 (0.011)	1.094 (0.008)
<i>a</i> _{H-C-C}	111.2	114.1 (2.9)	113.7 (2.5)	114.0 (2.8)
<i>a</i> _{H-C-H}	107.7	104.4 (−3.3)	104.9 (−2.8)	104.6 (−3.1)
RMSD		0.130	0.038	0.045
Methylamine				
<i>r</i> _{C-N}	1.454	1.060 (−0.394)	1.418 (−0.037)	1.439 (−0.015)
<i>r</i> _{C-H^a}	1.088	1.109 (0.021)	1.097 (0.009)	1.097 (0.009)
<i>a</i> _{H1-C-N}	114.6	120.8 (6.2)	119.6 (5.0)	116.2 (1.6)
<i>a</i> _{H2-C-N}	109.4	111.4 (2.0)	110.4 (1.0)	113.3 (3.9)
<i>a</i> _{H1-C-H2}	108.0	104.2 (−3.8)	105.2 (−2.7)	104.4 (−3.6)
<i>a</i> _{H2-C-H2'}	107.3	103.6 (−3.6)	105.0 (−2.3)	104.1 (−3.1)
RMSD		0.141	0.024	0.030
Methanol				
<i>r</i> _{C-O}	1.399	1.016 (−0.384)	1.312 (−0.088)	1.323 (−0.077)
<i>r</i> _{C-H^a}	1.085	1.117 (0.032)	1.101 (0.016)	1.100 (0.015)
<i>a</i> _{H1-C-O}	107.4	103.7 (−3.7)	105.0 (−2.4)	111.4 (4.1)
<i>a</i> _{H2-C-O}	112.0	119.5 (7.5)	116.9 (4.9)	115.8 (3.8)
<i>a</i> _{H1-C-H2}	108.3	103.8 (−4.6)	105.7 (−2.6)	104.1 (−4.2)
<i>a</i> _{H2-C-H2'}	108.7	104.4 (−4.4)	105.6 (−3.1)	104.3 (−4.4)
RMSD		0.140	0.041	0.033

Only unique hydrogen atoms are considered (see Fig. 3 for atom numbering). For LSCF calculations, deviations from the corresponding RHF values are reported in parentheses, together with the RMSD (in Å) of the structure of the CH₃-Y fragment.

^a Average over the methyl group.

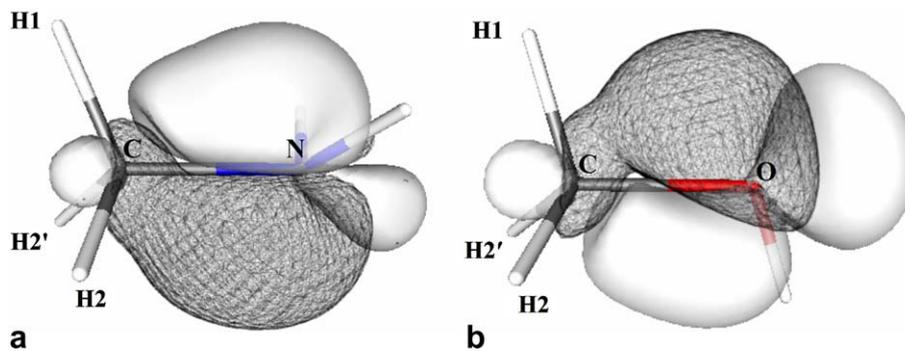


Fig. 3. Isosurfaces of the difference $|\varphi_{\text{ELMO}}(\mathbf{r})|^2 - |\varphi_{\text{WP}}(\mathbf{r})|^2$ plotted at the 0.001 (white solid surface) and -0.001 a.u. (black mesh surface) values for methylamine (a) and methanol (b).

To assess the performance of FCOs in a real QM/MM environment a further test was performed on the lysine amino acid. Full-RHF/6-311G** optimizations were carried out on both the protonated and deprotonated forms of the Ace-Lys-Nme dipeptide. Since the aim of this example is to compare different methods, we did not consider the full lysine conformational space, but only a conformer with a completely extended side chain. QM/MM optimizations were then performed setting the boundary at the C_α-C_β bond and describing the side chain at the QM level (see Fig. 1). Classical terms involving both QM and MM atoms were considered according to the scheme reported in Ref. [14]. Moreover, the φ and ψ dihedrals defining the orienta-

tion of the two amide groups were constrained to their RHF values³.

Also in this case calculations were performed both with and without the FCO, which was obtained on the ethane molecule together with the two types of SLBOs. As reported in Table 2, the frontier bond length deviations from RHF yielded by the three methods are similar to

³ Since the relative energies of the protonated and unprotonated forms are significantly influenced by the orientation of the two peptide bonds with respect to the amine group and for the current partitioning the φ and ψ angles are mainly determined by the force-field, this constraint was necessary to isolate the influence of FCOs on deprotonation energies.

Table 2

Energetic and geometric data calculated for the protonated (P) and neutral (N) form of lysine at the RHF, LSCF(ELMO)/+1, LSCF(ELMO)/+3 and LSCF(WP)/+3 level with the 6-311G** basis set

	RHF	LSCF(ELMO)/+1	LSCF(ELMO)/+3	LSCF(WP)/+3
$E_{\text{tot}}(\text{P})$	-665.536821	-212.301496	-213.426853	-213.442916
$E_{\text{tot}}(\text{N})$	-665.161689	-211.929661	-213.049365	-213.066150
ΔE	235.4	233.3 (-2.1)	236.9 (1.5)	236.4 (1.0)
$r_{\text{C}\alpha\text{-C}\beta}$	1.531	1.206 (-0.325)	1.576 (0.045)	1.591 (0.060)
$\text{RMSD}(r_{\text{QM-C}\beta})$	-	0.025	0.021	0.017
$\text{RMSD}(a_{\text{QM-C}\beta\text{-C}\alpha})$	-	3.7	3.4	3.5
$\text{RMSD}(a_{\text{C}\beta\text{-C}\alpha\text{-MM}})$	-	3.8	5.1	5.4
RMSD_{QM}	-	0.062	0.031	0.043

Total energies (including MM contributions for LSCF) are reported in a.u., while $\Delta E = E_{\text{tot}}(\text{N}) - E_{\text{tot}}(\text{P})$ is in kcal/mol. Distances and distance RMSD values are in Å, while RMSD over angles are in degrees. RMSD_{QM} (in Å) refers to the positions of QM atoms ($\text{C}\alpha$ excluded).

those of the $\text{CH}_3\text{-C}$ fragment, with LSCF(ELMO)/+3 giving the smallest value. Furthermore, distances and angles involving $\text{C}\beta$ and its QM neighbours ($\text{RMSD}(r_{\text{QM-C}\beta})$ and $\text{RMSD}(a_{\text{QM-C}\beta\text{-C}\alpha})$) are reproduced by the three methods with the same accuracy, which is comparable to the literature values [11,17]. However, the introduction of the FCO improves the overall deviation of the side chain from RHF (RMSD_{QM}).

At last, LSCF/+3 calculations allow reducing the deviation from the RHF deprotonation energy ΔE by 0.6–1.1 kcal/mol, with WP SLBO yielding the best value. All values are anyway within the accuracy that can be expected for QM/MM calculations of this kind [8,11,17,18].

4. Conclusions

In this Letter the introduction of frozen core orbitals (FCOs) is proposed as an alternative approach to solve the problem of the frontier bond length shortening in the LSCF/MM hybrid method. Our preliminary calculations on test molecules, representative of three of the most commonly found chemical bonds (namely the C–C, C–N and C–O single bonds), suggest that FCOs allow to recover a significant part of the repulsion between the frontier bond atoms, yielding an average error for the frontier bond length of 0.06 Å from the SCF value without the need of adding a potential with specific parameters. Moreover, further tests show that FCOs do not modify the description of the rest of the QM subsystem, both from the geometric and the energetic point of view. As a future improvement, the variation of the SLBO with the frontier bond length could be taken into account.

References

- [1] U.C. Singh, P.A. Kollman, *J. Comput. Chem.* 7 (1986) 718.
- [2] M.J. Field, P.A. Bash, M. Karplus, *J. Comput. Chem.* 11 (1990) 700.
- [3] S. Ranganathan, J.E. Gready, *J. Phys. Chem. B* 101 (1997) 5614.
- [4] P.L. Cummins, J.E. Gready, *J. Phys. Chem. B* 104 (2000) 4503.
- [5] D. Das, K.P. Eurenus, E.M. Billings, P. Sherwood, D.C. Chatfield, M. Hodoseck, B.R. Brooks, *J. Chem. Phys.* 117 (2002) 10534.
- [6] F. Maseras, K. Morokuma, *J. Comput. Chem.* 16 (1995) 1170.
- [7] Y. Zhang, T.-S. Lee, W. Yang, *J. Chem. Phys.* 110 (1999) 46.
- [8] I. Antes, W. Thiel, *J. Phys. Chem. A* 103 (1999) 9290.
- [9] G.A. DiLabio, M. M Hurley, P.A. Christiansen, *J. Chem. Phys.* 116 (2002) 9578.
- [10] J. Gao, P. Amara, C. Alhambra, M.J. Field, *J. Phys. Chem. A* 102 (1998) 4714.
- [11] J. Pu, J. Gao, D.G. Truhlar, *J. Phys. Chem. A* 108 (2004) 632.
- [12] X. Assfeld, J.-L. Rivail, *Chem. Phys. Lett.* 263 (1996) 100.
- [13] V. Kairys, J.H. Jensen, *J. Phys. Chem. A* 104 (2000) 6656.
- [14] N. Ferré, X. Assfeld, J.-L. Rivail, *J. Comput. Chem.* 23 (2002) 610.
- [15] R.B. Murphy, D.M. Philipp, R.A. Friesner, *J. Comput. Chem.* 21 (2000) 1442.
- [16] H. Lin, D.G. Truhlar, *J. Phys. Chem. A* 109 (2005) 3991.
- [17] N. Reuter, A. Dejaegere, B. Maigret, M. Karplus, *J. Phys. Chem. A* 104 (2000) 1720.
- [18] P. Amara, M.J. Field, *Theor. Chem. Acc.* 109 (2003) 43.
- [19] H. Weinstein, R. Pauncz, *Symp. Faraday Soc.* 2 (1968) 23.
- [20] H. Stoll, G. Wagenblast, H. Preuss, *Theoret. Chim. Acta* 57 (1980) 169.
- [21] M. Couty, C.A. Bayse, M.B. Hall, *Theor. Chem. Acc.* 97 (1997) 96.
- [22] M. Sironi, A. Famulari, *Theor. Chem. Acc.* 103 (1999) 417.
- [23] A. Fornili, M. Sironi, M. Raimondi, *J. Mol. Struct. (Theochem)* 632 (2003) 153.
- [24] Z. Szekeres, P.R. Surján, *Chem. Phys. Lett.* 369 (2003) 125.
- [25] A. Fornili, Y. Moreau, M. Sironi, X. Assfeld, *J. Comput. Chem.* 27 (2006) 515.
- [26] N. Ferré, X. Assfeld, *J. Chem. Phys.* 117 (2002) 4119.
- [27] Y. Moreau, P.-F. Loos, X. Assfeld, *Theor. Chem. Acc.* 112 (2004) 228.
- [28] N. Ferré, X. Assfeld, *J. Mol. Struct. (Theochem)* 632 (2003) 83.
- [29] J. Pipek, P.G. Mezey, *J. Chem. Phys.* 90 (1989) 4916.
- [30] H. Weinstein, R. Pauncz, *Adv. At. Mol. Phys.* 7 (1971) 97.
- [31] P.R. Surján, in: Z.B. Maksic (Ed.), *Theoretical Models of Chemical Bonding, Part 2, The Concept of the Chemical Bond*, Springer, Heidelberg, 1989.
- [32] W.D. Cornell et al., *J. Am. Chem. Soc.* 117 (1995) 5179.
- [33] M.J. Frisch et al., *GAUSSIAN 98 (Revision A.9)*, Gaussian Inc., Pittsburgh, PA, 1998.
- [34] C.I. Bayly, P. Cieplak, W.D. Cornell, P.A. Kollman, *J. Phys. Chem.* 97 (1993) 10269.
- [35] M.F. Guest et al., *Mol. Phys.* 103 (2005) 719.