Brill Academic Publishers
P.O. Box 9000, 2300 PA Leiden,
The Netherlands

Computing Letters (CoLe)
vol. 3, no. 2-4, 2007, pp. 473-486
ISSN 1574–0404

Removing extra frontier parameters in QM/MM methods : a tentative with the Local Self-Consistent Field approach

P.-F. Loos[1]

Equipe de Chimie et Biochimie Théoriques,
UMR 7565 CNRS-UHP, Institut Jean Barriol (FR CNRS 2843),
Faculté des Sciences et Techniques, Nancy-Université, B.P. 239,
54506 Vandoeuvre-lès-Nancy Cedex, France

A. Fornili
Biocrystallography Unit, DIBIT S. Raffaele Scientific Institute,
Via Olgettina, 58 I-20132, Milano, Italy

M. Sironi
Dipartimento di Chimica Fisica ed Elettrochimica,
Università degli Studi di Milano, Via Golgi 19, 20133 Milano, Italy

X. Assfeld[2]
Equipe de Chimie et Biochimie Théoriques,
UMR 7565 CNRS-UHP, Institut Jean Barriol (FR CNRS 2843),
Faculté des Sciences et Techniques, Nancy-Université, B.P. 239,
54506 Vandoeuvre-lès-Nancy Cedex, France

Received 30 May, 2007; accepted in revised form 13 June, 2007

Abstract: We present a new quantum mechanics/molecular mechanics (QM/MM) scheme based on the Local-SCF method that avoids the extra-parametrization of the QM/MM frontier by taking into account the core electrons of the hybrid frontier atoms by means of self-consistent core orbitals (SCCO). This study follows and extends our previous investigation based on frozen core orbitals (FCO) [Chem. Phys. Lett. **427** 236-240(2006)]. Test calculations on small organic compounds show that the most common atoms found in biochemical systems can be used as frontier atoms without adding any parameters, and that only relatively small geometric and energetic deviations compared to full-QM calculations are generated.

Keywords: QM/MM, Local Self-Consistent Field , Extremely Localized Molecular Orbitals, Core Orbitals, Parametrization

Mathematics Subject Classification: Here must be added the AMS-MOS or PACS Numbers

PACS: Here must be added the AMS-MOS or PACS Numbers

[1]Corresponding author. E-mail: Pierre-Francois.Loos@cbt.uhp-nancy.fr
[2]Corresponding author. E-mail: Xavier.Assfeld@cbt.uhp-nancy.fr

1 Introduction

Hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) methods are well known to be powerful tools for computational chemists to study chemical and biochemical processes, such as enzyme catalysis. The basic idea of QM/MM methods exploits the local nature of the electronic modifications that occurs during a chemical reaction or during an electronic excitation. According to this principle, QM methods can be used to treat a small localized region of the system where bond breaking/forming takes place, whereas the surroundings is treated with MM. When solvent-solute systems are considered, the QM:MM coupling is relatively straightforward because the two subsystems interact exclusively through non-bonded interactions. However, when one wants to study macromolecular systems, special care is needed when the boundary between QM and MM regions passes through a covalent bond. Several treatments of the QM/MM frontier have been designed to correctly handle this problem.

The most simple and widely used scheme, called the link-atom (LA), use a monovalent atom to cap the unsaturated QM atoms. The LA is usually a hydrogen atom[1, 2, 3, 4, 5, 6, 7, 8, 9, 10] or a parametrized atom, like in the pseudo-bond[11], the adjusted connection atom[12] or the capping potential[13] philosophies.

A second class of QM/MM methods use frozen orbitals to describe the boundary. For example, the Local Self-Consistent Field (LSCF) method, developed by Rivail, Assfeld and co-workers[14, 15, 16, 17, 18, 19], employs a localized orbital to represent the frontier bond. Using the same idea, Friesner and co-worker developed a "frozen orbital QM/MM interface methodology"[20, 21, 22]. An extension of the LSCF method, called generalized hybrid orbital[23, 24, 25, 26, 27] (GHO), use a set of four hybrid orbitals for each MM boundary atom, while the effective fragment potential (EFP) method define a buffer region consisting of frozen localized orbitals between the QM and MM parts[28].

To ensure reliable results and an accurate description of the frontier, all these methods introduce special treatments of the boundary parameters. In the LA scheme, special treatment of the MM boundary charges are used because of the over-polarization of the wave function due to the nearby point charges on the QM/MM frontier[1, 12, 29, 30, 31, 32, 10, 33]. These MM point charges might be zeroed, scaled, redistributed[33] or partially delocalized[10] to avoid the unphysical polarization. For the parametrized atom approach[11, 12, 13], the semi-empirical or pseudo-potential parameters are adjusted to reproduce the geometrical data of model systems. In the LSCF framework, a special five-parameters frontier bond potential[18] (FBP) has been developed to ensure a well-described bond length in the QM/MM frontier. Friesner and co-workers have parametrized a library of frozen density for the side chains of all amino acids[20, 21, 22], while in the GHO scheme, a minimal valence basis set and parametrized integrals are required for the boundary atoms[26].

These specific extra parameters, needed to achieve a correct description of the frontier, can be considered as side-effects of the initial idea applied to connect the MM part to the QM one. Some collateral damages result from these parametrizations. The parameters can be force field dependent, basis set dependent, etc, resulting in a poorly transferable method and thus, to a poorly applicable QM/MM treatment of any system. To improve the transferability and the applicability of QM/MM methods, it is desired to get rid of these extra parameters, even if this gain is accompagned by a slight loss of accuracy.

In a previous article[19], we have shown that the introduction of the explicit quantum description of the core electrons of the LSCF frontier atom, by means of a frozen core orbital (FCO), avoids the determination of the specific FBP. As a counterpart, the determination of the FCO was required.

In this article, we would like to show that the explicit description of the core electrons can be handled and that acceptable accuracy can be obtained without freezing any core orbitals. No FBP or FCO determinations are required and the core electrons of the MM boundary atom are treated

as any other electron of the remaining QM atoms. Hence, these core electrons are described by an orbital obtained self-consistently, that we call in the remaining of this letter a self-consistent core orbitals (SCCO). Our aim is to show that the LSCF/MM method combined with the SCCO scheme is a general theoretical method, free of additional parameters, that is highly transferable and widely applicable to the major part of chemical and biochemical systems.

2 Computational Details

The particularity of the LSCF method is that it optimizes a wave function in the presence of frozen molecular orbitals[17, 18, 34, 35, 36, 37]. The QM subsystem can be treated at both the spin-restricted or unrestricted Hartree-Fock (HF) and density-functional theory (DFT) level for single point calculations and geometry optimizations. Post-HF LSCF single-point calculations are also possible and extension to geometry optimizations is underway. In QM/MM calculations[38, 39, 40, 41, 19, 42], the connecting bond is described by a strictly localized bond orbital (SLBO). This SLBO is determined on a model compound that possesses the chemical bond of interest and applied to the system under investigation according to the transferability principle. The two atoms defining the connection between the QM and the MM subsystems are called X and Y for the atom on the QM side and on the MM side respectively. The Y atom can be called a quantum-classical atom (QCA) since, while on one side it possesses basis functions and a nuclear charge, on the other side it has a classical atomic point charge, van der Waals parameters, and force field parameters corresponding to bonding interactions with the MM part. The classical parameters are those of the force field used to describe the MM part without any fudging adjustment for QM/MM calculations. The QM/MM interactions are summarized in table 1. One shall not forget that, in addition to these non-electronic QM/MM interactions, the MM point charges interact with the electrons of the QM fragment giving rise to the polarization of the wave-function. This polarization is the most important component of QM/MM methods. If the X-Y bond is covalent then we consider that the X and Y atom contributes each by one electron to the SLBO. In that case, the Y atom bears a nuclear charge of +1. If the X-Y bond can be considered as dative then the nuclear charge of the Y is either 0 or +2 depending if it gives 0 or 2 electrons to the QM fragment. In a previous study[19], we proposed to add the core electron of the Y atom to the QM part. The nuclear charge of the Y atom is then equal to the number of core electron + the number of valence electrons participating to the QM system. In that previous study, the core electrons were described by means of FCO. In the present study, the core electrons are considered as any other electron of the QM fragment. Hence, they are described by self-consistent core orbitals (SCCO).

In order to compare the SCCO and the FCO scheme, we perform the same calculations at the same level of theory as the previous study[18]. First, benchmark LSCF calculations on small target molecules (ethane, methylamine and methanol) that possess the three most common bonds found in organic molecules (C-C, C-N and C-O bonds) are computed. The SLBO are determined at the full-QM RHF/6-311G** level of theory according to the *a posteriori* Weinstein-Pauncz (WP) criterion[43, 44] followed by the deletion of the orthogonalization-delocalization tails. Furthermore, FCOs are determined as Extremely Localized Molecular Orbitals (ELMOs)[45, 46, 47, 48, 49, 35] through calculations on the target molecules following a Lewis structure localization scheme (see Ref. [19] for more details). This point is motivated by the best transferability of the ELMO scheme[35]. Besides, it has been shown[19] that the way the FCO is determined (from a ROHF calculation on the isolated atom or an ELMO calculation on the molecular environment) leads to indistinguishable results.

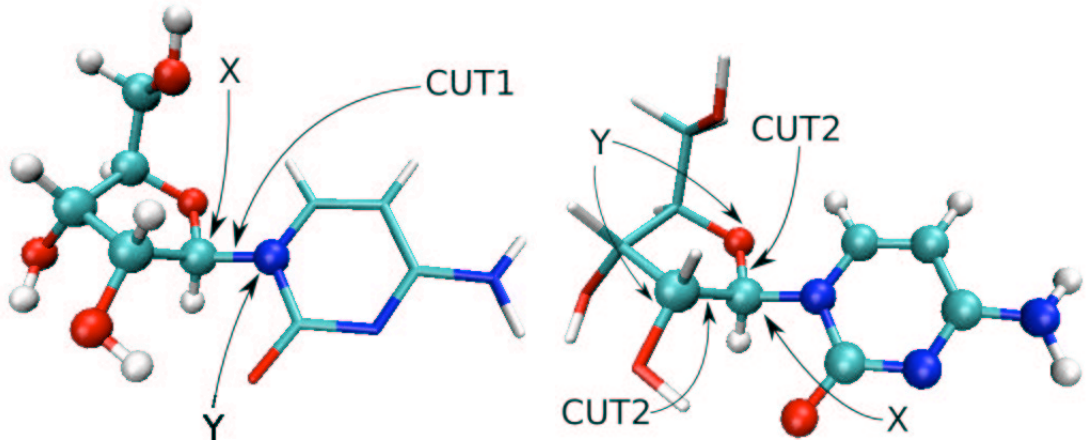
In addition to these model systems, we perform QM/MM calculations on the cytidine nucleoside to check the applicability of the new SCCO scheme to larger biomolecular systems (see fig. 1). Two QM/MM boundary locations have been considered, corresponding to the three cases studied.

Table 1: Details of the QM/MM interactions. The atom of the quantum part are labelled Q and the ones of the MM part C. Z(A) and q(A) represent respectively the nuclear charge and the classical atomic point charge of atom A. $P_{\mu\nu}$ is the $\mu\nu^{th}$ element of the density matrix.

Bonds	$Y\text{-}C_1$
Angles	$X\text{-}Y\text{-}C_1$, $Y\text{-}C_1\text{-}C_2$
Torsion	$Q_1\text{-}X\text{-}Y\text{-}C_1$, $X\text{-}Y\text{-}C_1\text{-}C_2$, $Y\text{-}C_1\text{-}C_2\text{-}C_3$
vdW[a]	$Q_i\text{-}Y_k$, $Q_i\text{-}C_j$, $Y_k\text{-}C_j$, $Y_k\text{-}Y_l$[b]
Electrostatics	
Z-Z	$Z(Q_i)\text{-}Z(Y_k)$, $Z(Y_k)\text{-}Z(Y_l)$[b]
Z-q	$Z(Q_i)\text{-}q(Y_k)$, $Z(Q_i)\text{-}q(C_j)$, $Z(Y_k)\text{-}q(C_j)$, $Z(Y_k)\text{-}q(Y_l)$[b]
q-q[a]	$q(Y_k)\text{-}q(C_j)$, $q(Y_k)\text{-}q(Y_l)$[b]
e-Z	$\sum_k \sum_{\mu\nu} P_{\mu\nu} \langle \mu \vert \frac{Z(Y_k)}{r_k} \vert \nu \rangle$
e-q	$\sum_k \sum_{\mu\nu} P_{\mu\nu} \langle \mu \vert \frac{q(Y_k)}{r_k} \vert \nu \rangle$, $\sum_j \sum_{\mu\nu} P_{\mu\nu} \langle \mu \vert \frac{q(C_j)}{r_j} \vert \nu \rangle$

[a]Standard 1-4 condition of classical force fields
[b]$k \neq l$

The Amber *ff99*[50, 51] force field is used and to ensure the charge neutrality of the MM part, the RESP[52, 53] protocol is applied to determine the classical charges.

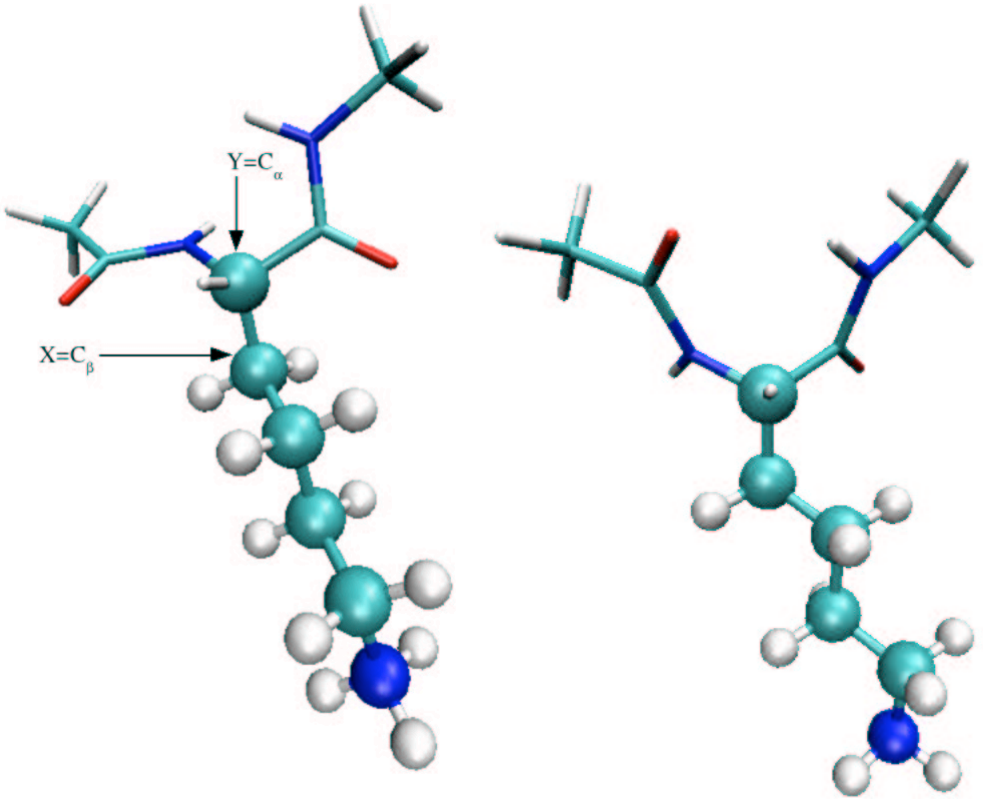


(a) Cut 1: C-N case (b) Cut 2: C-C and C-O cases

Figure 1: Partitioning of the cytidine nucleoside into QM (ball and sticks) and MM (thin sticks) regions. Two QM/MM boundary locations have been considered.

Next, the accuracy of the new scheme is checked against energetic values. We perform QM/MM calculations on the two protonation states of the lysine amino acid capped with an acetyl and a N-methylacetamide group at the N-terminus and C-terminus respectively (Figure 2). The QM/MM frontier is located between the C_α and C_β atoms. The Amber *ff99* force field and the RESP protocol is applied in the same manner as above.

All QM/MM calculations are performed with our modified version of the Gaussian 03[54] package linked to the Tinker software[55] for the MM calculations.



(a) Protonated form (b) Neutral form

Figure 2: Partitioning of the Ace-Lys-NMe dipeptide into QM (ball and sticks) and MM (thin sticks) regions. The QM/MM boundary is located between the C_α and C_β atoms.

3 Results and Discussion

3.1 Test Calculations

To study the intrinsic influence of the MM frontier atom on the geometric parameters of the QM/MM frontier, we performed LSCF calculations on the ethane, methylamine and methanol molecules. The frontiers were set, respectively, at the C-C, C-N and C-O bonds and, first, the MM hydrogen atoms are removed to provide a force field independent model. Because each Y atom participates to the SLBO with one electron, a +1 nuclear charge was set to the Y atom (we refer to these calculations as LSCF/+1 in the following as was proposed previously[19]). However, when FCO or SCCO are used, the core electrons must be taken into account, and consequently, a +3 nuclear charge is carried by the frontier atom to ensure the electroneutrality (LSCF/+3 calculations).

In table 2, we report the equilibrium distances and angles for the three molecules at the LSCF(WP)/+1, LSCF(WP)/+3 FCO and the LSCF(WP)/+3 SCCO level of theory (figure 3). One can see that the LSCF/+1 scheme gives reasonable deviation from RHF values for the angles ($RMSD_\angle$) but largely underestimate the C-Y bond lengths, which justifies the development of a FBP in the previous LSCF/MM methodology[18]. The introduction of the explicit description of the Y core electrons reduces the deviation between LSCF/+3 and RHF bond lengths without

affecting the other degrees of freedom. The SCCO scheme, in addition to avoid the FCO determination, gives bond lengths in better agreement with the RHF values compared to the FCO approach whatever the kind of chemical bond. More precisely, the SCCO scheme reduces the distance root mean square deviation ($RMSD_d$) from RHF results to 0.025, 0.006 and 0.032 Å for the ethane, methylamine and methanol molecules, respectively. Changing the localization criterion does not affect the just derived conclusion. For example, the C-C bond length of the ethane molecule in the LSCF/+3 SCCO framework is 1.557, 1.554, 1.559 and 1.562 Å for the Magnasco-Perico[56, 57], Boys-Foster[58], Pipek-Mezey[59] and Weinstein-Pauncz[43, 44] criterion respectively.

Table 2: Selected distances (in Å) and angles (in degrees) calculated for CH_3-R (R = CH_3, NH_2, OH) using the RHF, LSCF(WP)/+1, LSCF(WP)/+3 FCO and LSCF(WP)/+3 SCCO level of theory with the 6-311G** basis set. For LSCF calculations, deviations from the corresponding RHF values are reported in parentheses, together with the distance and angle RMSD ($RMSD_d$ and $RMSD_\angle$ (in Å and degrees) of the structure of the CH_3-Y fragment.

	RHF	LSCF(WP)/+1	LSCF(WP)/+3 FCO	LSCF(WP)/+3 SCCO
Ethane				
d(C-C)	1.527	1.118(-0.409)	1.585(0.057)	1.562(0.035)
d(C-H)	1.086	1.104(0.018)	1.094(0.008)	1.094(0.007)
∠(H-C-C)	111.2	114.0(2.8)	114.0(2.8)	114.6(3.4)
∠(H-C-H)	107.7	104.6(-3.1)	104.6(-3.1)	103.9(-3.8)
$RMSD_d$	—	0.290	0.041	0.025
$RMSD_\angle$	—	3.0	3.0	3.6
Methylamine				
d(C-N)	1.454	1.030(-0.424)	1.439(-0.015)	1.454(0.000)
d(C-H)[a]	1.088	1.110(0.022)	1.097(0.009)	1.096(0.008)
∠(H_1-C-N)	114.6	115.2(0.6)	116.2(1.6)	116.5(1.9)
∠(H_2-C-N)	109.4	114.6(5.2)	113.3(3.9)	113.4(4.0)
∠(H_1-C-H_2)	108.0	103.8(-4.2)	104.4(-3.6)	104.3(-3.7)
∠(H_2-C-H_2')	107.3	103.6(-3.7)	104.1(-3.1)	103.6(-3.7)
$RMSD_d$	—	0.300	0.030	0.006
$RMSD_\angle$	—	3.8	3.2	3.4
Methanol				
d(C-O)	1.399	0.947(-0.453)	1.323(-0.077)	1.356(-0.043)
d(C-H)[a]	1.085	1.116(0.031)	1.100(0.015)	1.098(0.013)
∠(H_1-C-O)	107.4	113.7(6.3)	111.4(4.1)	110.7(3.3)
∠(H_2-C-O)	112.0	116.1(4.1)	115.8(3.8)	115.6(3.6)
∠(H_1-C-H_2)	108.3	103.0(-5.3)	104.1(-4.2)	104.3(-4.0)
∠(H_2-C-H_2')	108.7	103.2(-5.5)	104.3(-4.4)	105.1(-3.6)
$RMSD_d$	—	0.321	0.033	0.032
$RMSD_\angle$	—	5.3	4.1	3.6

[a]Average over the methyl group

In figure 4, we report the energy profile of the ethane, methylamine and methanol molecules obtained by rigid scan along the C-Y bond (Y=C,N,O) at the RHF, FCO and SCCO levels of theory. As one can see, the LSCF/+3 schemes significantly improve the position of the minimum and exhibit similar energy profile for small atomic separation. This point is a great improvement

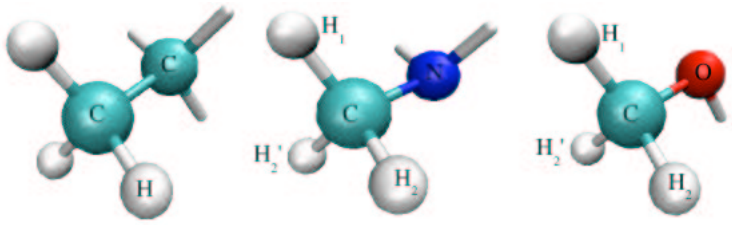


Figure 3: Atom labels and partitioning of the CH_3-R (R = CH_3, NH_2, OH) molecules into QM (ball and sticks) and MM (thin sticks).

compared to LSCF/+1 calculations. For large interatomic distance, the LSCF/+3 scheme present a different behaviour compared to full-RHF calculation, due to the neglect of the SLBO readjustment. This latter point is irrelevant for geometric considerations because the frontier bond is not expected to break during QM/MM calculations. From the energetic point of view, the readjustment of the SLBO had to be considered and preliminary results are encouraging[37].

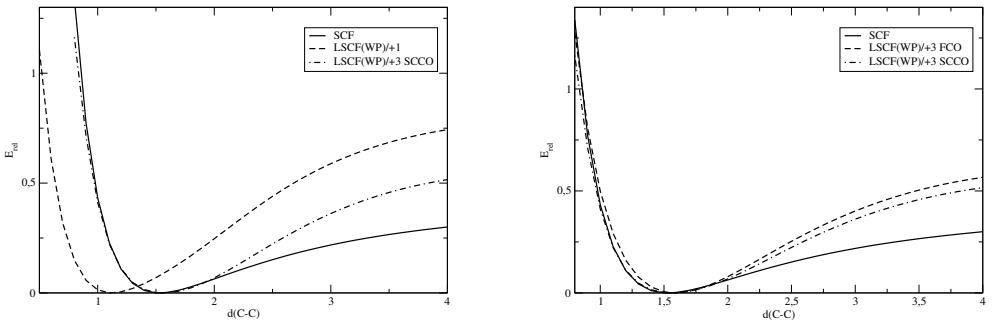
To determine the influence of the MM hydrogen atoms of the R fragment, we report the equilibrium distances and angles in the case of mechanical embedding (ME) (MM3 force field[60, 61, 62, 63, 64, 65, 66, 67]) only and with electronic embedding (EE) (OPLS-AA force field [68, 69, 70, 71, 72, 73, 74]. As shown in table 3, the inclusion of the MM surrounding does not influence the conclusions of the previous paragraph whatever the kind of embedding. However, we point out the influence of the classical point charge of the Y atom. For example, in the case of the methanol molecule, taking into account the MM charge of the Y atom induces a 0.034 Å decreasing of the C-O distance compared to ME only. As the native OPLS-AA atomic point charges do not preserve the neutrality of the MM part (expect for the ethane case due to symmetry considerations), we report the same calculations with atomic point charges obtained according to the RESP procedure. The results exhibit slight modifications of the equilibrium bond lengths and angles with respect to the change of the MM charges, which is expected for the C-Y distance that is sensitive to the charge of the Y atom. We conclude that the LSCF/+3 SCCO scheme allows to have a good geometrical description of the QM/MM boundary independently of the force field (ME or EE) and the set of classical atomic point charges.

3.2 Cytidine nucleoside

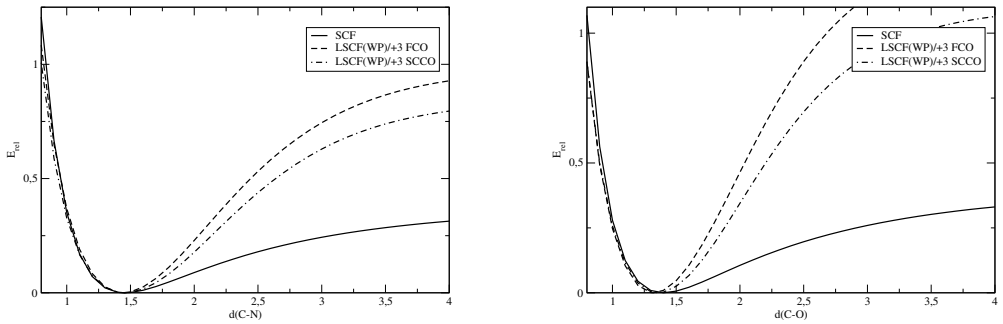
In order to check the reliability of the SCCO scheme on larger systems, we consider the cytidine molecule. The partitioning schemes between the QM and MM parts are depicted on fig. 1. For the first case (cut 1: fig. 1a), the QM/MM frontier has been set to the C-N bond connecting the ribose ring to the cytosine nucleobase, and the ribose ring which is considered at the QM level. The second case (cut 2: fig. 1b) corresponds to QM/MM calculations comprising two QM/MM frontiers located to the C-C and C-O bonds of the ribose moiety, and we treat the nucleobase at the QM level. For the three chemical bonds (C-C, C-N and C-O), the WP SLBO determined on the previous model systems (ethane, methylamine and methanol) are used.

The results for the cytidine nucleoside are summarized in table 4. For the two cases, the LSCF(WP)/+3 SCCO scheme gives reasonable equilibrium frontier bond lengths compared to the full-RHF calculations with a maximum deviation of 0.050 Å. The largest deviation is found for difficult case where the X atom is connected to several Y atoms (cut 2). One can see that the accuracy of the new scheme is similar to that of the LSCF/+1 FBP[18] and LSCF/+3 FCO[19] approach. This clearly shows the limitation of the point charge model for the representation of the frontier atom.

,



(a) C-C bond energy profile



(b) C-N bond energy profile

(c) C-O bond energy profile

Figure 4: Energy profile (in a.u.) along the C-C, C-N and C-O bond (in Å) for the ethane, methylamine and methanol molecules at the RHF level of theory and for the CH_3-R fragment at the LSCF(WP)/+3 FCO and SCCO level of theory with the 6-311G** basis set.

3.3 Ace-Lys-NMe Dipeptide

In this section, we examine the overall performance of the SCCO scheme on QM/MM systems. As a benchmark, the protonated and neutral forms of the Ace-Lys-NMe dipeptide are computed at the full-RHF/6-311G** level of theory for a conformer with a completely extended side chain, since proton affinities are very sensitive and very difficult cases[75]. We performed QM/MM calculations with full geometry optimization. The frontier is located between the C_α and C_β atoms (figure 2) and the side chain of the lysine residue is treated at the QM level. For the LSCF/+1 scheme, the C(sp3)-C(sp3) FBP[18] is applied to the boundary. The FCO and SLBO are determined on the ethane molecule thanks to ELMO calculations following the same methodology as previously (see above). The RMSD of the QM (MM resp.) part, RMSD_{QM} (RMSD_{MM} resp.), and the overall RMSD are computed following the method of Kabsch[76] implemented in the VMD software[77].

The results are gathered in table 5. The three theoretical schemes give reasonable frontier bond lengths and RMSD_{QM} values lower than 0.044 Å. The main deviations between the full-QM and the QM/MM structures come from the MM part ($\text{RMSD} \simeq \text{RMSD}_{MM}$). In the case of the protonated form of the lysine, although one can see that the C_α-C_β distance obtained with

Table 3: Classical point charges (in electrons) of the corresponding force field, selected distances (in Å) and angles (in degrees) calculated for CH_3-R (R = CH_3, NH_2, OH) using the LSCF(WP)/+3 SCCO level of theory with the 6-311G** basis set. Deviations from the corresponding RHF values are reported in parentheses, together with the distance and angle RMSD—$RMSD_d$ and $RMSD_\angle$—(in Å and degrees) of the structure of the CH_3-Y fragment.

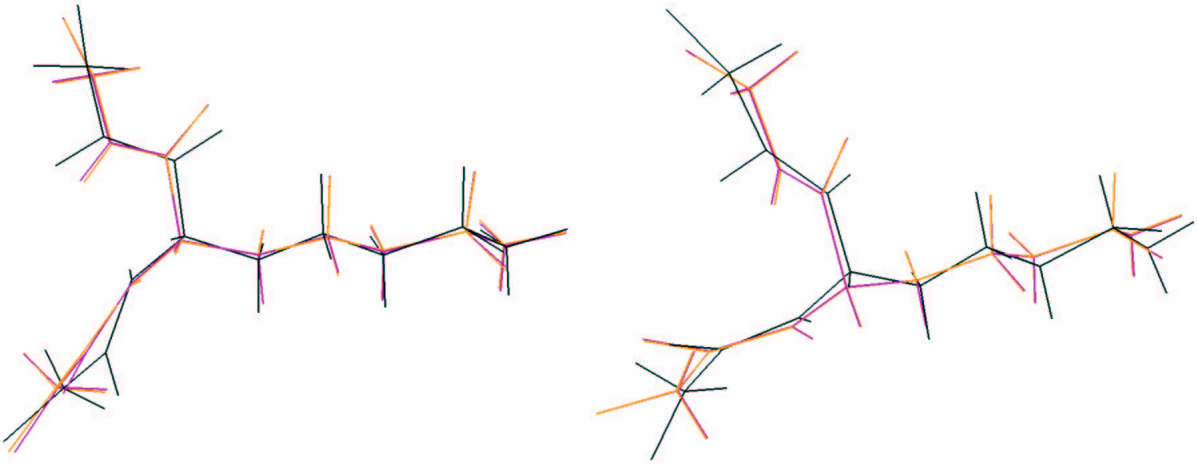
Force Field	MM3	OPLS-AA	OPLS-AA
Charge Set	—	OPLS-AA	RESP
Ethane			
q(C)	—	-0.1800	-0.0189
q(H)	—	0.0600	0.0063
d(C-C)	1.567(0.040)	1.589(0.042)	1.565(0.038)
d(C-H)	1.095(0.009)	1.095(0.009)	1.094(0.008)
∠(H-C-C)	114.8(3.6)	114.6(3.4)	114.6(3.4)
∠(H-C-H)	103.6(-4.1)	103.9(-3.8)	103.9(-3.8)
$RMSD_d$	0.029	0.030	0.027
$RMSD_\angle$	3.9	3.6	3.6
Methylamine			
q(N)	—	-0.9000	-0.4446
q(H)	—	0.3500	0.2223
d(C-N)	1.462(0.008)	1.465(0.011)	1.482(0.028)
d(C-H)[a]	1.097(0.009)	1.105(0.017)	1.100(0.012)
∠(H_1-C-N)	116.8(2.2)	117.1(2.5)	117.2(2.6)
∠(H_2-C-N)	113.8(4.4)	113.6(4.2)	113.4(4.0)
∠(H_1-C-H_2)	104.1(-3.9)	103.9(-4.1)	103.9(-4.1)
∠(H_2-C-H_2')	103.5(-3.8)	103.2(-4.1)	103.5(-3.8)
$RMSD_d$	0.009	0.014	0.022
$RMSD_\angle$	3.7	3.8	3.7
Methanol			
q(O)	—	-0.6830	-0.3177
q(H)	—	0.4180	0.3177
d(C-O)	1.358(-0.041)	1.325(-0.074)	1.354(-0.045)
d(C-H)[a]	1.099(0.014)	1.107(0.022)	1.101(0.016)
∠(H_1-C-O)	110.4(3.0)	111.4(4.0)	110.9(3.5)
∠(H_2-C-O)	116.0(4.0)	116.2(4.2)	116.0(4.0)
∠(H_1-C-H_2)	103.9(-4.4)	103.7(-4.6)	103.9(-4.4)
∠(H_2-C-H_2')	105.2(-3.5)	104.2(-4.5)	104.6(-4.1)
$RMSD_d$	0.031	0.055	0.034
$RMSD_\angle$	3.8	4.3	4.0

[a]Average over the methyl group

LSCF(ELMO)/+3 SCCO method is the less accurate one, all QM/MM schemes give similar RMSD values and structures (fig. 5). For the neutral form, the LSCF/+3 schemes do not permit to obtain RMSD values of the same accuracy as the LSCF/+1 framework. However, the FCO and SCCO schemes reduce the deviation from the RHF deprotonation energy ΔE by 1.4 and 1.0 kcal/mol respectively. We conclude that the description of the Y core electrons by means of FCO or SCCO does not significantly affect the geometry of both the QM and/or MM part compared to LSCF/+1 FBP calculations and reduces the energetic deviation from full-QM calculations. As one can see

Table 4: X-Y equilibrium distances of the cytidine nucleoside at the RHF and LSCF(WP)/+3 SCCO level of theory with the 6-311G** basis set. Two types of QM/MM boundary location have been considered and depicted on fig. 1. The Amber *ff99* force field is used for QM/MM calculations and the MM charges are determined according to the RESP procedure. Deviations from the corresponding RHF values are reported in parentheses.

	RHF	LSCF(WP)/+3 SCCO
Cut 1		
d(C-N)	1.473	1.488(0.015)
Cut 2		
d(C-O)	1.389	1.369(-0.020)
d(C-C)	1.538	1.588(0.050)



(a) Protonated form (b) Neutral form

Figure 5: Overlay of the protonated and neutral form of the Ace-Lys-NMe dipeptide structure at the full RHF (black), LSCF(ELMO)/+3 FCO (orange) and LSCF(ELMO)/+3 SCCO (purple) level of theory.

from the results gathered in table 5, taking into account the core electrons of the Y atom by means of a FCO or a SCCO leads to similar results. Besides, the SCCO scheme permits to avoid the FCO determination.

4 Conclusions

In this article, we have introduced the concept of self-consistent core orbitals (SCCO) as an alternative of our previously frozen core orbitals (FCO) approach based on the hybrid LSCF/MM method. The new theoretical scheme avoids the extra-parametrization of the QM/MM boundary and the FCO determination. We have shown that the explicit inclusion of the core electrons by means of a SCCO orbitals leads to results of comparable quality than the ones obtained with the FCO approach. Compared to full-QM calculations, the SCCO scheme induces small geometrical and/or energetic deviations on both model systems and larger biomolecules like polypeptides and nucleosides. Furthermore, we have shown that several types of chemical bonds with different

Table 5: Energetic and geometric data for the protonated (P) and neutral (N) form of lysine at the RHF, LSCF(ELMO)/+1 FBP, LSCF(ELMO)/+3 FCO and LSCF(ELMO)/+3 SCCO level of theory with the 6-311G** basis set. The Amber *ff99* force field is used for QM/MM calculations and the MM charges are determined according to the RESP procedure. Distances and distance RMSD values are in Å and the $\Delta E = E_{tot}(N) - E_{tot}(P)$ is in kcal/mol.

	RHF	LSCF(ELMO)/+1 FBP	LSCF(ELMO)/+3 FCO	SCCO
Protonated form				
d(C_α-C_β)	1.536	1.519(-0.017)	1.573(0.037)	1.589(0.053)
$RMSD_{QM}$	—	0.029	0.038	0.044
$RMSD_{MM}$	—	0.367	0.377	0.359
RMSD	—	0.374	0.393	0.362
Neutral form				
d(C_α-C_β)	1.526	1.520(-0.006)	1.573(0.047)	1.588(0.062)
$RMSD_{QM}$	—	0.028	0.035	0.037
$RMSD_{MM}$	—	0.571	1.154	1.140
RMSD	—	0.532	1.031	1.021
ΔE	235.4	239.0(3.6)	237.6(2.2)	238.0(2.6)

polarities (C-C, C-N, C-O) can be considered and correctly handled. This last point is a real improvement since it is not always possible to cut a C-C bond to divide the macromolecules.

According to these conclusions, the only user-dependent parameters for LSCF/MM calculations are the strictly localized bond orbitals (SLBO). As outlook, we are currently investigating and testing the reliability of this new theoretical scheme in the evaluation of the protonation affinities of some polypeptides for the case of the peptide bond QM/MM boundary location.

Acknowledgment

The authors are grateful to Nicolas Ferré for helpful and fruitful discussions. Two of the authors (X.A. and P.F.L.) are indebted to J. Preat, D. Jacquemin and E.A. Perpète (Namur) for their intense and friendly scientific support.

References

[1] U.C. Singh and P.A. Kollman, J. Comput. Chem. **7**, 718 (1986).

[2] M.J. Field, P.A. Bash and M. Karplus, J. Comput. Chem. **11**, 700 (1990).

[3] F. Maseras and K. Morokuma, J. Comput. Chem. **16**, 1170 (1995).

[4] M. Svensson, S. Humbel, D.J. Froese, T. Matsubara, S. Sieber and K. Morokuma, J. Phys. Chem. **100** 19357 (1996).

[5] S. Ranganathan and J. Gready, J. Phys. Chem. B **101** 5614 (1997).

[6] S. Dapprich, I. Komárino, K. S. Byun, K. Morokuma, and M. J. Frich, J. Mol. Struct. (Theochem) **461** 1 (1999).

[7] P. Cummins and J. Gready, J. Phys. Chem. B **104** 4503 (2000).

[8] T. Vreven, B. Mennucci, C.O. da Silva, K. Morokuma and J. Tomasi, J. Chem. Phys. **115** 62 (2001).

[9] T. Kerdcharoen and K. Morokuma, Chem. Phys. Lett. **355** 257 (2002).

[10] D. Das, K.P. Eurenius, E.M. Billings, P. Sherwood, D.C. Chatfield, M. Hodoseek and B.R. Brooks, J. Chem. Phys. **117** 10534 (2002).

[11] Y. Zhang, T.-S. Lee and W. Yang, J. Chem. Phys. **110** 46 (1999).

[12] I. Antes and W. Thiel, J. Phys. Chem. A **103** 9290 (1999).

[13] G.A. DiLabio, M.M. Hurley and P.A. Christiansen, J. Chem. Phys. **166** 9578 (2002).

[14] G.G. Ferenczy, J.-L. Rivail, P.R. Surján and G.Náray-Szabó, J. Comput. Chem. **13** 830 (1992).

[15] V. Théry, D. Rinaldi, J.-L. Rivail, B. Maigret and G. Ferenczy, J. Comput. Chem. **15** 269 (1994).

[16] G. Monard, M. Loos, V. Théry, K. Baka and J.-L. Rivail, Int. J. Quantum Chem. **58** 153 (1996).

[17] X. Assfeld and J.-L Rivail, Chem. Phys. Lett. **263** 100 (1996).

[18] N. Ferré, X. Assfeld and J.-L Rivail, J. Comput. Chem. **23** 610 (2002).

[19] A. Fornili, P.-F. Loos, M. Sironi and X. Assfeld, Chem. Phys. Lett. **427** 236 (2006).

[20] D.M. Philipp and R.A. Friesner, J. Comput. Chem. **20** 1468 (1999).

[21] R.B. Murphy, D.M. Philipp and R.A. Friesner, J. Comput. Chem. **21** 1442 (2000).

[22] R.B. Murphy, D.M. Philipp and R.A. Friesner, Chem. Phys. Lett. **321** 113 (2000).

[23] J. Gao, P. Amara, C. Alhambra and M.J. Field, J. Phys. Chem. A **102** 4714 (1998).

[24] P. Amara, M.J. Field, C. Alhambra and J. Gao, Theor. Chem. Acc. **104** 336 (2000).

[25] D.G. Truhlar, J. Gao, C. Alhambra, M. Garcia-Viloca, J. Corchado, M.L. Sánchez and J. Villá, Acc. Chem. Res. **35** 341 (2002).

[26] J. Pu, J. Gao and D.G. Truhlar, J. Phys. Chem. A **108** 632 (2004).

[27] J. Pu, J. Gao, and D.G. Truhlar, J. Phys. Chem. A **108** 5454 (2004).

[28] V. Kairys and J.H. Jensen, J. Phys. Chem. A **104** 6656 (2000).

[29] N. Reuter, A. Dejaegere, B. Maigret and M. Karplus, J. Phys. Chem. A **104** 1720 (2000).

[30] P. Amara and M. J. Field, Theor. Chem. Acc. **109** 43 (2003).

[31] D. Bakowies and W. Thiel J. Phys. Chem. **100** 10580 (1996).

[32] A. de Vries, P. Sherwood, S. Collins, A. Rigby, M. Rigutto and G. Kramer, J. Phys. Chem. B **103** 6133 (1999).

[33] H. Lin and D.G. Truhlar, J. Phys. Chem. A **109** 3991 (2005).

[34] N. Ferré and X. Assfeld, J. Chem. Phys. **117** 4119 (2002).

[35] A. Fornili, Y. Moreau, M. Sironi and X. Assfeld, J. Comput. Chem. **27** 515 (2006).

[36] P.-F. Loos and X. Assfeld, Int. J. Quantum Chem. **online** (2007)

[37] P.-F. Loos and X. Assfeld, J. Chem. Theory and Comput. **3** 1047 (2007).

[38] N. Ferré and X. Assfeld, J. Mol. Struct. (Theochem) **632** 632 (2003).

[39] X. Assfeld, N. Ferré and J.-L. Rivail, Theor. Chem. Acc. **111** 328 (2004).

[40] Y. Moreau, P.-F. Loos and X. Assfeld, Theor. Chem. Acc. **112** 228 (2004).

[41] Y. Moreau and X. Assfeld, *Lecture Series on Computer and Computational Sciences* **3** 1 (2005).

[42] P.-F. Loos, X. Assfeld and J.-L. Rivail, Theor. Chem. Acc. **online** (2007).

[43] H. Weinstein and R. Pauncz, Adv. Atomic. Molec. Phys. **7** 97 (1971).

[44] H. Weinstein and R. Pauncz, Symp. Faraday Soc. **2** 23 (1968).

[45] H. Stoll, G. Wagenblast and H. Preuss, Theoret. Chim. Acta **57** 169 (1980).

[46] M. Couty, C.A. Bayse and M.B. Hall, Theor. Chem. Acc **97** 96 (1997).

[47] M. Sironi and A. Famulari, Theor. Chem. Acc. **103** 417 (1999).

[48] A. Fornili, M. Sironi and M. Raimondi, J. Mol. Struct. (Theochem) **632** 157 (2003).

[49] Z. Szekeres and P.R. Surjan, Chem. Phys. Lett. **369** 125 (2003).

[50] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, Jr. K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell and P.A. Kollman, J. Am. Chem. Soc. **117** 5179 (1995).

[51] J. Wang, P. Cieplak and P.A. Kollman, J. Comput. Chem. **21** 1049 (2000).

[52] C.I. Bayly, W.D. Cornell and P.A. Kollman, J. Phys. Chem. **97** 10269 (1993).

[53] W.D. Cornell, P. Cieplak, C.I. Bayly and P.A. Kollman, **115** 9620 (1993).

[54] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, T. Vreven Jr., K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision B.05. Gaussian, Inc., Wallingford, CT, 2004.

[55] J.W. Ponder, Tinker, version 4.2, Washington University: St. Louis, MO, 2004.

[56] V. Magnasco and A. Perico, J. Chem. Phys. **47** 971 (1967).

[57] V. Magnasco and A. Perico, J. Chem. Phys. **48** 800 (1967).

[58] J.M. Foster and S.F. Boys, Rev. Mod. Phys. **32** 300 (1960).

[59] J. Pipek and P. Mezey, J. Chem. Phys. **90** 4916 (1989).

[60] N.L. Allinger, Y.H. Yuh and J.-H. Lii, J. Am. Chem. Soc. **111** 8551 (1989).

[61] J.-H. Lii and N.L. Allinger, J. Am. Chem. Soc. **111** 8567 (1989).

[62] J.-H. Lii and N.L. Allinger, J. Am. Chem. Soc. **111** 8576 (1989).

[63] N.L. Allinger, H.J. Geise, W. Pyckhout, L.A. Paquette and J.C. Gallucci, J. Am. Chem. Soc. **111** 1106 (1989).

[64] N.L. Allinger, F. Li and L. Yan, J. Comput. Chem. **11** 848 (1990).

[65] N.L. Allinger, F. Li, L. Yan, and J.C. Tai, J. Comput. Chem. **11** 868 (1990).

[66] J.-H. Lii and N.L. Allinger, J. Phys. Org. Chem. **7** 591 (1994).

[67] J.-H. Lii and N.L. Allinger, J. Comput. Chem. **19** 1001 (1998).

[68] W.L. Jorgensen, D.S. Maxwell and J. Tirado-Rives, J. Am. Chem. Soc. **117** 11225 (1996).

[69] D.S. Maxwell, J. Tirado-Rives, and W.L. Jorgensen, J. Comput. Chem. **16** 984 (1995).

[70] W.L. Jorgensen and N.A. McDonald, J. Mol. Struct. **424** 145 (1998).

[71] N.A. McDonald and W.L. Jorgensen, J. Phys. Chem. B **102** 8049 (1998).

[72] R.C. Rizzo and W.L. Jorgensen, J. Am. Chem. Soc. **121** 4827 (1999).

[73] M.L.P. Price, D. Ostrovsky and W.L. Jorgensen, J. Comput. Chem. **22** 1340 (2001).

[74] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives and W.L. Jorgensen, J. Phys. Chem. B **105** 6474 (2001).

[75] H. Lin and D.G. Truhlar, Theor. Chem. Acc. **117** 185 (2007).

[76] W. Kabsch, Acta Cryst. A **34** 827 (1978).

[77] W. Humphrey, A. Dalke and K. Schulten, J. Molec. Graphics, **14** 33 (1996).