ELSEVIER

Journal of Molecular Structure (Theochem) 632 (2003) 157–172

THEO CHEM

www.elsevier.com/locate/theochem

Determination of extremely localized molecular orbitals and their application to quantum mechanics/molecular mechanics methods and to the study of intramolecular hydrogen bonding

Arianna Fornili, Maurizio Sironi*, Mario Raimondi

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

Received 31 October 2002; revised 2 December 2002; accepted 9 December 2002

Abstract

The extremely localized molecular orbitals (ELMOs) are a set of molecular orbitals strictly localized only on a few atoms of a molecule. They are obtained in an a priori fashion through the direct application of the variation principle. Even if the theoretical aspects of their determination have been discussed already in the literature, stable and fast algorithms to obtain ELMOs are still not trivial and a comparison between different methods is reported. We furthermore investigate the applicability of ELMOs to quantum mechanics/molecular mechanics (QM/MM) methods which employ frozen localized orbitals to represent covalent bonds across the QM and the MM region. In addition it is shown that ELMOs can be used to describe species with intramolecular hydrogen bonds, where a correct elimination of the intramolecular basis set superposition error can be essential to perform accurate conformational studies.

Keywords: Extremely localized molecular orbitals; Ab initio; Quantum mechanics/molecular mechanics methods; Intramolecular basis set superposition error

1. Introduction

It is well known that the picture offered by the molecular orbitals (MOs) yielded by the self-consistent resolution of the canonical Hartree–Fock equations is in general much far from the classical representation of the structure of a molecule. Indeed, we often consider molecules as an assembly of functional groups whose properties are transferable from one molecule to another. Furthermore, in the rationalization of chemical structure and reactivity we heavily rely on *local* concepts, as when we draw lines or points to indicate lone pairs and electron pairs shared between atoms. MOs are instead spread out on the whole molecule and this prevents their transfer between molecules, even if differing for the addition of just a few atoms.

In the past years several efforts have been made to recover the concept of locality in theoretical calculations: a number of methods have indeed been developed to localize MOs by unitary transformations at the end of a traditional HF calculation. These

* Corresponding author. Tel.: +39-025031-4251; fax: +39-025031-4300.
E-mail address: maurizio.sironi@unimi.it (M. Sironi).

0166-1280/03/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0166-1280(03)00296-3

a posteriori approaches differ for the localization criterion, i.e. for the localization functional of the orbitals to be optimized: the Boys method [1] minimizes the spatial extension of the orbitals, while in the technique proposed by Edmiston and Ruedenberg [2] the self-repulsion energy is maximized. Also known is the Pipek and Mezey approach [3], based on the minimization of a functional correlated with the Mulliken population analysis that measures the number of atoms over which a given orbital extends. Even if localized molecular orbitals (LMOs) yielded by different a posteriori procedures have the same general main features, this variety of criteria gives to these methods some degree of arbitrariness. Furthermore, the LMOs show 'tails' beyond the localization region, i.e. they are not strictly localized. This drawback makes their transfer difficult: indeed, even if the coefficients associated to these tails are small, their effect on the energy is not negligible. At last, a posteriori methods need a complete traditional SCF calculation, so they cannot exploit the computational advantages that may arise from the local character of the electron distribution.

This is instead the purpose of the a priori approaches, which partition the molecule into subunits and try to calculate the global wavefunction or the electron density function as a sum of contributions coming from each fragment and from the interactions between them. Walker and Mezey proposed to build-up the electron density of a polypeptide superimposing the densities determined on smaller fragments, giving rise to the 'LEGO' approach [4]. They showed that the resulting electron density function is almost indistinguishable from that determined with a direct calculation on the whole molecule. Yang et al. [5–7] furthermore developed a 'divide and conquer' (D & C) approach, where the electron density corresponding to the solution of the Kohn–Sham equations is obtained through a proper definition of the electron densities of smaller parts of the target molecule. Due to the success obtained, the D & C strategy has been also implemented in the framework of the Hartree–Fock theory and semi-empirical approaches [8,9].

The methods based on the determination of 'extremely localized molecular orbitals' (ELMOs), known in the literature also as 'strictly localized molecular orbitals' (SLMOs) [10] and 'non-orthogonal localized molecular orbitals' (NOLMO) [11], are closely related to the group function method introduced by McWeeny [12], one of the first theoretical approaches to the decomposition of the total electronic wavefunction into functions describing subsets of electrons. As in the conventional 'linear combination of atomic orbitals' (LCAO) approximation, ELMOs are expanded in terms of atomic orbitals (AOs). However, for each ELMO it is possible to use only a part of the total basis, namely the functions centred on a given subset of atoms, whose choice can be defined on the basis of the chemical problem under investigation. Thus they are characterized by the absence of any tail, as the localization region is defined before the calculation. The orbitals obtained in this way are rigorously localized and this makes the ELMOs ideal candidates for transfer. It should also be noticed that the procedure for the determination of ELMOs yields strictly localized virtual orbitals as well. This could be an attractive feature, since attempts to localize virtual orbitals, useful to describe the correlation on a local basis, are not always successful, particularly with large basis sets [13].

In this work we will discuss two different applications of ELMOs at present under study in our laboratory, the first of which concerns the description of frontier regions in quantum mechanics/molecular mechanics (QM/MM) methods. The mixed QM/MM approaches are used for large systems which cannot be represented entirely at the quantum level, but have only a small part which requires to be modelled with high accuracy (e.g. the active site of an enzyme). The chief difficulty in these hybrid methods is the coupling between the QM and the MM level of theory, particularly when the QM and the MM regions are connected by a covalent bond. Several coupling scheme have been proposed [14–21]. In particular, there are methods (such as the LSCF of Rivail et al. [16,22,23]), which describe the frontier bonds by means of 'frozen' localized orbitals. These are generally obtained with the a posteriori localization techniques mentioned earlier, followed by 'tail deletion', i.e. the annihilation of the 'forbidden' coefficients. In the present work, instead, we will discuss the possibility to model the frontier bonds through the ELMOs and we will present a preliminary QM/MM calculation performed on a test molecule.

The second application of ELMOs that we are studying is the elimination of the basis set

superposition error (BSSE) in the calculation of intramolecular interactions. This problem is closely related to the *inter*molecular analogue, but it has been much less discussed in the literature. A few correction schemes of intramolecular BSSE have been suggested [24–26], which are based on modifications of the a posteriori counterpoise (CP) method introduced by Boys and Bernardi [27] for the intermolecular case. Here an a priori technique is proposed, that employs the ELMO wavefunction to describe molecular fragments constrained to be strictly localized on the appropriate centers. This approach can be considered as a generalization of the SCF for Molecular Interactions (SCF–MI) method, proposed by Gianinetti et al. [28,29] to correct intermolecular interaction energies.

The article is organized as follows: we will first discuss briefly the ELMO theory and the algorithms implemented for their determination. Then, some peculiar features of ELMOs (such as strict localization and transferability) will be exemplified through calculations on a test molecule. At last, the preliminary results of the ELMO-based approaches to the description of frontier bonds in hybrid methods and to the elimination of intramolecular BSSE will be presented.

2. Theory

Let us consider a system composed by one or more closed shell molecules. ELMOs are defined partitioning the N MOs (where $2N$ is the total number of electrons) into n subgroups or fragments and then assigning to each fragment j a 'partial' basis set $\{|\chi_p^j\rangle\}_{p=1}^{m_j}$, built-up with m_j functions of the total basis. The αth ELMO of the jth fragment can thus be written as

$$|\varphi_\alpha^j\rangle = \sum_{p=1}^{m_j} C_{p\alpha}^j |\chi_p^j\rangle.$$

The coefficients $C_{p\alpha}^j$ are determined minimizing the expectation value of the energy of the electronic ELMO wavefunction

$$|\psi\rangle = \hat{A}\left[\prod_{j=1}^{n}\prod_{\alpha=1}^{N_j}\varphi_\alpha^j\bar{\varphi}_\alpha^j\right]$$

where $\hat{A}$ is the antisymmetrizer and N_j the number of the doubly occupied ELMOs of the fragment j. Thus the ELMOs are determined with a straightforward use of the variation principle.

It must be noted that different fragments can share one or more AOs: therefore, if $\mathbf{\Phi}$ is the row vector built up with all the occupied ELMOs of the system, written as

$$\mathbf{\Phi} = [|\varphi_1^1\rangle\cdots|\varphi_{N_1}^1\rangle\cdots|\varphi_1^j\rangle\cdots|\varphi_{N_j}^j\rangle\cdots],$$

while $\mathbf{X}$ is the row vector constructed with the full atomic basis set $\{|\chi_q\rangle\}_{q=1}^{m}$ (ignoring the subdivision in fragments, so that only unique AOs are considered), written as $\mathbf{X} = [|\chi_1\rangle\cdots|\chi_m\rangle]$, the matrix $\mathbf{C}$ in $\mathbf{\Phi} = \mathbf{X}\cdot\mathbf{C}$ has the general structure shown in Fig. 1. It is worth pointing out that in general ELMOs cannot be orthogonalized without loosing the 'block' structure of $\mathbf{C}$.

As mentioned earlier, the SCF–MI equations were originally developed by Gianinetti et al. [28] to eliminate BSSE from the calculation of intermolecular interaction energies: in that case it was necessary to define only one fragment for each molecule of the system, so that each AO appeared in one subgroup only and the matrix $\mathbf{C}$ acquired a block diagonal structure. It was demonstrated that the minimum total energy is achieved by satisfying the following eigenvalue–eigenvector equations

$$\mathbf{F}'_j\mathbf{C}_j = \mathbf{S}'_j\mathbf{C}_j\mathbf{E}_j \tag{1a}$$

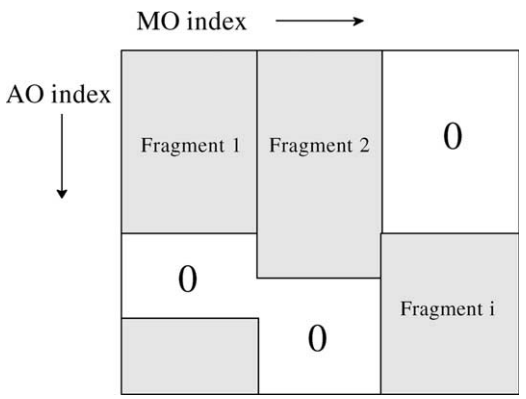

Fig. 1. Schematic representation of an ELMO coefficient matrix $\mathbf{C}$.

where

$$\mathbf{F}'_j = [(\mathbf{I} - \mathbf{SD} + db(\mathbf{SC}\tilde{\mathbf{S}}^{-1})\tilde{\mathbf{S}}^{-1}\mathbf{C}^{\mathrm{T}})\mathbf{F}(\mathbf{I} - \mathbf{DS} + \mathbf{C}\tilde{\mathbf{S}}^{-1}db(\tilde{\mathbf{S}}^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{S}))]_{jj}, \quad (1b)$$

$$\mathbf{S}'_j = [db(\mathbf{S}) - db(\mathbf{SDS}) + db(\mathbf{SC}\tilde{\mathbf{S}}^{-1})db(\tilde{\mathbf{S}}^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{S})]_{jj}, \quad (1c)$$

$\mathbf{C}_j = [\mathbf{C}]_{jj}$ and $\mathbf{E}_j$ is a diagonal matrix. In the preceding expressions $\mathbf{D}$ is the density matrix defined as $\mathbf{D} = \mathbf{C}\tilde{\mathbf{S}}^{-1}\mathbf{C}^{\mathrm{T}}$, $\mathbf{S}$ and $\tilde{\mathbf{S}}$ contain the overlap integrals between AOs and ELMOs, respectively, $\mathbf{F}$ is the Fock matrix and $\mathbf{I}$ is the identity matrix. Furthermore, the notation $[\mathbf{M}]_{jk}$or $\mathbf{M}_{jk}$ indicates the block of the matrix $\mathbf{M}$ with the row index running along the fragment j and the column index running along the fragment k, while $db(\mathbf{M})$ is the diagonal block matrix built-up with the n $\mathbf{M}_{jj}$ blocks.

The block diagonal shape of $\mathbf{C}$ can be recovered also in the general case if we 'expand' the AO row vector $\mathbf{X}$ into

$$\widehat{\mathbf{X}} = [|\chi_1^1\rangle \cdots |\chi_{m_1}^1\rangle \cdots |\chi_1^j\rangle \cdots |\chi_{m_j}^j\rangle \cdots],$$

where AOs belonging to more than one subgroup are repeated in each fragment, so that the dimension of the basis now is $\widehat{m} = \sum_{j=1}^{n} m_j$, which is in general greater than m. With this 'redundant' basis we thus have 'supermatrices' $\widehat{\mathbf{C}}$ and $\widehat{\mathbf{D}}$ defined as $\mathbf{\Phi} = \widehat{\mathbf{X}} \cdot \widehat{\mathbf{C}}$ and $\widehat{\mathbf{D}} = \widehat{\mathbf{C}}\,\tilde{\mathbf{S}}^{-1}\,\widehat{\mathbf{C}}^{\,T}$, respectively. Eqs. (1a)–(1c) are still valid if we replace matrices $\mathbf{F}$ and $\mathbf{S}$ with supermatrices $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{S}}$ with dimension $\widehat{m} \times \widehat{m}$, whose blocks ij are defined as

$$(\widehat{\mathbf{F}}_{ij})_{pq} = \langle \chi_p^i | \hat{F} | \chi_q^j \rangle \text{ and } (\widehat{\mathbf{S}}_{ij})_{pq} = \langle \chi_p^i | \chi_q^j \rangle,$$

with $p = 1, m_i$ and $q = 1, m_j$. Nevertheless, Sironi et al. [30] showed that $\mathbf{F}'_j$ and $\mathbf{S}'_j$ can be calculated as

$$\mathbf{F}'_j = \sum_{ik}^{n} [\widehat{\mathbf{I}}_{ii}\delta_{ij} - (\widehat{\mathbf{V}}_{ji}^{T} - \widehat{\mathbf{V}}_{jj}^{T}\tilde{\mathbf{S}}_{ji}^{-1})\widehat{\mathbf{C}}_i^{T}]\widehat{\mathbf{F}}_{ik}[\widehat{\mathbf{I}}_{kk}\delta_{kj} - \widehat{\mathbf{C}}_k(\widehat{\mathbf{V}}_{kj} - \tilde{\mathbf{S}}_{kj}^{-1}\widehat{\mathbf{V}}_{jj})] \quad (2a)$$

$$\mathbf{S}'_j = \widehat{\mathbf{S}}_{jj} - \sum_{i}^{n} \widehat{\mathbf{S}}_{ji}\widehat{\mathbf{C}}_i\widehat{\mathbf{V}}_{ij} + \widehat{\mathbf{V}}_{jj}^{T}\widehat{\mathbf{V}}_{jj}, \quad (2b)$$

where $\widehat{\mathbf{V}} = \tilde{\mathbf{S}}^{-1}\,\widehat{\mathbf{C}}^{\,T}\,\widehat{\mathbf{S}}$, so that its block ij is given by

$$\widehat{\mathbf{V}}_{ij} = \sum_{k}^{n} \tilde{\mathbf{S}}_{ik}^{-1}\,\widehat{\mathbf{C}}_k^{\,T}\,\widehat{\mathbf{S}}_{kj} \quad (2c)$$

Therefore, neither $\mathbf{F}'_j$ nor $\mathbf{S}'_j$ requires that the supermatrices $\widehat{\mathbf{M}}$ are actually used in the calculations. A drawback of the method based on the Eqs. (1a) and (2a)–(2c), which will be denoted as ELMO1, is that it involves a $\widehat{\mathbf{S}}$ matrix which is singular; the singularity is also transferred to the $\mathbf{S}'_j$ matrix, so that Eq. (1a) have to be solved with the canonical orthogonalization method [30,31]. Although the algorithm converges in most cases, it is sometimes affected by a certain instability, which makes convergence very difficult.

A related formalism for determining ELMOs proposed by Stoll et al. [32] explicitly uses the dual orbitals of the ELMOs, which are defined as

$$|\tilde{\varphi}_\alpha^i\rangle = \sum_{j}^{n}\sum_{\beta}^{Nj} (\tilde{S}_{ji}^{-1})_{\beta\alpha}|\varphi_\beta^j\rangle.$$

It was shown that ELMOs can be determined by solving the following equations [32]

$$\begin{cases} \langle \chi_q^j | \hat{\tilde{F}}^j | \varphi_\beta^j \rangle & = \varepsilon_\beta^j \langle \chi_q^j | \varphi_\beta^j \rangle \\ \langle \varphi_\alpha^j | \varphi_\beta^j \rangle & = \delta_{\alpha\beta} \end{cases}, \quad (3)$$

where

$$\hat{\tilde{F}}^j = (1 - \hat{\rho} + \hat{\rho}^{j+})\hat{F}(1 - \hat{\rho} + \hat{\rho}^j) \text{ and } \varepsilon_\beta^j = \langle \varphi_\beta^j | \hat{\tilde{F}}^j | \varphi_\beta^j \rangle.$$

The density and 'partial' density operators $\hat{\rho}$ and $\hat{\rho}^j$ are defined as

$$\hat{\rho} = \sum_{j}^{n}\sum_{\alpha}^{Nj} |\tilde{\varphi}_\alpha^j\rangle\langle\varphi_\alpha^j| \text{ and } \hat{\rho}^j = \sum_{\alpha}^{Nj} |\tilde{\varphi}_\alpha^j\rangle\langle\varphi_\alpha^j|.$$

In the following, we will refer to this approach as ELMO2. Both the sets of Eqs. (1a) and (3) have to be solved self-consistently. In general we observed that the ELMO1 algorithm, when converging, is much faster than ELMO2, which nevertheless proved to be more stable in some cases.

The convergence properties of these algorithms proved to be heavily affected also by the number of AOs shared by the fragments. For a given molecule,

there can be different ways of defining the ELMOs, i.e. it is possible to define several 'localization schemes'. In the most localized scheme, which generally can be built-up by inspection of the Lewis structure of the molecule, a fragment is defined for each bond, employing AOs centered only on the atoms involved in the bond, while the ELMOs which correspond to core electrons or lone pairs belonging to a given atom use only the basis functions located on it. Nevertheless it is possible to 'partially' delocalize ELMOs which are intended to describe bonds and lone pairs, e.g. allowing them to use also AOs centered on next neighbours: thus the number of AOs shared by the fragments increases. The localization scheme chosen generally depends on the particular application of the ELMOs and in this paper we will give some examples of different partitioning schemes. From our experience the convergence rate of both the ELMO1 and ELMO2 algorithms decreases as the delocalization of the partitioning scheme increases, probably as a consequence of a stronger coupling between equations relative to different fragments. In addition some partitionings schemes showed even a divergent behavior.

The convergence difficulties associated to the resolution of Eq. (3) were already illustrated in previous works [11,32]. Algorithms which exploit the information provided by exact first and approximate second derivatives of the energy with respect to the ELMO coefficients have been proposed [11,32,33]. These derivatives can be calculated with the following expressions, taken from Ref. [32]:

$$\frac{\partial E}{\partial C^j_{q\beta}} = 4\langle \chi^j_q | (1-\hat{\rho})\hat{F}|\tilde{\varphi}^j_\beta\rangle; \tag{4}$$

$$\begin{aligned}\frac{\partial^2 E}{\partial C^j_{p\beta}\partial C^k_{q\gamma}} &= 4(\tilde{\mathbf{S}}^{-1}_{kj})_{\gamma\beta}\langle\chi^j_p|(1-\hat{\rho})\hat{F}(1-\hat{\rho})|\chi^k_q\rangle \\ &\quad - 4\langle\chi^j_p|(1-\hat{\rho})|\chi^k_q\rangle\langle\tilde{\varphi}^k_\gamma|\hat{F}|\tilde{\varphi}^j_\beta\rangle \\ &\quad - 4\langle\chi^j_p|(1-\hat{\rho})\hat{F}|\tilde{\varphi}^k_\gamma\rangle\langle\chi^k_q|\tilde{\varphi}^j_\beta\rangle \\ &\quad - 4\langle\chi^k_q|(1-\hat{\rho})\hat{F}|\tilde{\varphi}^j_\beta\rangle\langle\chi^j_p|\tilde{\varphi}^k_\gamma\rangle \end{aligned} \tag{5}$$

where Eq. (5) was obtained from Eq. (4) neglecting the dependence of the Fock operator on the ELMO coefficients. A DIIS approach using these expressions was proposed [11]: nevertheless, even if DIIS techniques proved useful to overcome convergence difficulties, in certain cases we have seen that these algorithms can be very unstable. Hence we tried a conjugate gradient procedure, based on the Polak–Ribiere formula [34], which nevertheless was found to be still too slow near the minimum. Thus, we decided to implement an augmented Hessian Newton–Raphson algorithm, using the shift parameter suggested by Goldfeld et al. [35]. This method converges with much fewer iterations than both the ELMO2 and the conjugate gradient algorithms and it is free from the instability that affects the ELMO1 method.

In Fig. 2 and Table 1 we report a test calculation performed on benzene with the different methods considered, using a TZV basis set. The starting point was the same for each procedure and it was prepared performing a partial localization by means of the method of Pipek and Mezey [3] on Hartree–Fock canonical orbitals. Then some coefficients were annihilated according to a partitioning scheme where the ELMOs describing the C core electrons could use only the AOs centered on the corresponding atom, while the ELMOs describing the C–C single bonds could use only the AOs centered on the atoms involved in the bond. Each of the three ELMOs which were intended to represent the π electrons, was developed on AOs centered on three contiguous carbon atoms. One can see from Fig. 2 that the ELMO1 procedure was initially much faster than ELMO2, but it failed to converge. The conjugate gradient method (ELMO-CG) was slightly better than ELMO2, while the augmented Hessian Newton–Raphson (ELMO-NR) algorithm succeeded in reaching the convergence threshold in much fewer iterations (see also Table 1).

The ELMO-NR method was successful also for partitionings schemes which turned out to be particularly 'difficult' cases, where all the other algorithms failed to converge. However also this procedure has some drawbacks, i.e. the computational cost of each iteration and the storage of the Hessian matrix, which can become very large as the number of coefficients increases. To solve the first problem we have

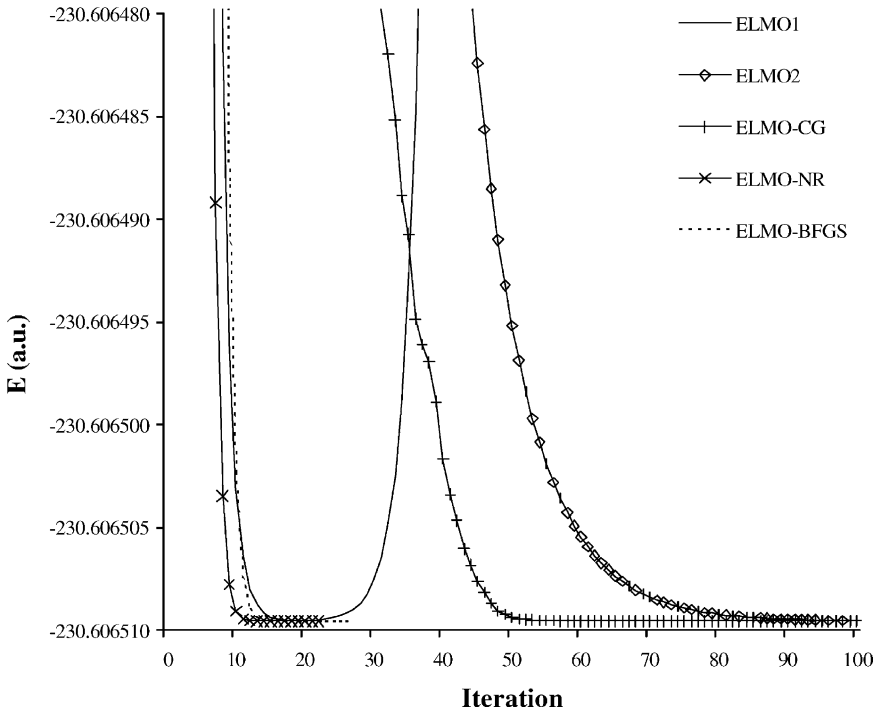

Fig. 2. ELMO total energy of benzene (in a.u.) vs iteration number, calculated with ELMO1, ELMO2, ELMO-CG, ELMO-NR and ELMO-BFGS algorithms. The basis set used is the TZV.

implemented a quasi-Newton procedure, where the approximate analytic Hessian given by Eq. (5) is calculated only in the first iteration and it is successively updated with a variable metric algorithm. Both the Davidon–Fletcher–Powell and Broyden–Fletcher–Goldfarb–Shanno updating formulae have been used [34]. The latter proved to be more efficient than the former, so that here we report only the BFGS results (see ELMO-BFGS in Fig. 2 and Table 1). This procedure converges with a number of iterations slightly larger than ELMO-NR, but, as each iteration is much cheaper, the overall computing time is largely reduced. Furthermore, the cost of the calculation, as well as the memory needed for the storage of the Hessian matrix, could be reduced using only a part of the Hessian, e.g. the blocks relative to a given fragment [33]. This aspect will be investigated in future work.

The development of a robust algorithm, which permits to obtain a tight convergence, was a necessary step to the subsequent implementation of the geometry optimization. The derivative of the electronic energy with respect to the generic coordinate a, $\frac{\partial E_{\text{elec}}}{\partial a}$, can be evaluated through the known expression [29,36]:

$$\frac{\partial E_{\text{elec}}}{\partial a} = 2\sum_{pq}^{m} D_{pq}\frac{\partial h_{pq}}{\partial a} + \sum_{pqrs}^{m}(2D_{pq}D_{rs} - D_{ps}D_{qr}) \times \frac{\partial(pq|rs)}{\partial a} - 2\sum_{pq}^{m} W_{pq}\frac{\partial S_{pq}}{\partial a},$$

where h_{pq} and $(pq|rs)$ are the one and two-electron AO integrals, respectively. W_{pq} is an element of

Table 1
Number of iterations needed by each ELMO algorithm to reach the convergence threshold (set to 5×10^{-7} a.u. on the maximum component of the gradient) in the calculations of Fig. 2

Algorithm	Number of iterations
ELMO2	192
ELMO-CG	121
ELMO-NR	22
ELMO-BFGS	27

The ELMO1 method is not reported as it failed to converge.

the Lagrangian matrix, that is defined as $W_{pq} = \sum_{\alpha}^{N} C'_{p\alpha} C'_{q\alpha} \varepsilon_{\alpha}$, where $C'_{p\alpha}$ are the components in the AO basis of the eigenvectors of the Fock operator represented in the basis of the occupied ELMOs $|\varphi_{\alpha}^{i}\rangle$ and ε_{α} are the corresponding eigenvalues. The algorithm has been interfaced with the PC-GAMESS version [37] of the GAMESS-US package [38]: it thus permits to carry out both conventional and direct 'single point' calculations and to perform geometry optimizations.

3. Calculations

3.1. Test calculations

Some test calculations have been made on the acetone molecule to compare some electronic properties (such as the electrostatic potential and the Mulliken populations) and the optimized geometry calculated with the conventional SCF and the ELMO wavefunction. All calculations have been performed with the 6-31G** basis set and for the ELMO calculations the most localized partitioning scheme (see Section 2) has been chosen. The structure used, unless otherwise stated, is the SCF optimized geometry shown in Fig. 3.

First we can note that the difference between ELMO and SCF energy values can be large: for acetone with the chosen basis set and localization scheme, the ELMO energy is ~40 kcal/mol higher than the SCF value. This result is to be ascribed to the reduction in the number of variational parameters caused by the a priori annihilation of some elements of the coefficient matrix. However, it is to be remarked that the ELMO wavefunction provides the lowest possible energy for a given partitioning scheme. Indeed, by performing an a posteriori localization on HF canonical orbitals and annihilating the 'tails', we obtain an energy value which is much above that yielded by the ELMO calculation. In our case the LMO wavefunction (localized according to the Pipek and Mezey criterion [3]) subjected to the 'tail deletion' produced an increment of ~80 kcal/mol with respect to the SCF value and thus of ~40 kcal/mol with respect to the ELMO. In Fig. 4 we report the LMO and ELMO describing one of the C–C bonds of acetone, to give a pictorial representation of the 'extreme' localization of ELMOs: on the atoms next to those involved in the bond, it is possible to see clearly the tails of the LMO, responsible for the significant contribution to the total energy. Furthermore, in Fig. 5(a) we compare the electrostatic potential calculated along the molecular plane of acetone with the SCF, the ELMO and the LMO subjected to the 'tail deletion' (*td* LMO) wavefunction: in spite of the large energy difference, both *td* LMO and ELMO contours have the same main features as the SCF wavefunction, which nevertheless is more closely approximated by the ELMO calculation.

Finally, we point out that the SCF energy can be recovered to a highly significant extent just allowing a partial delocalization of the ELMOs. If the ELMOs are partially 'relaxed' performing just a single SCF iteration, an energy value only 3 kcal/mol above the converged SCF value is obtained. This behavior

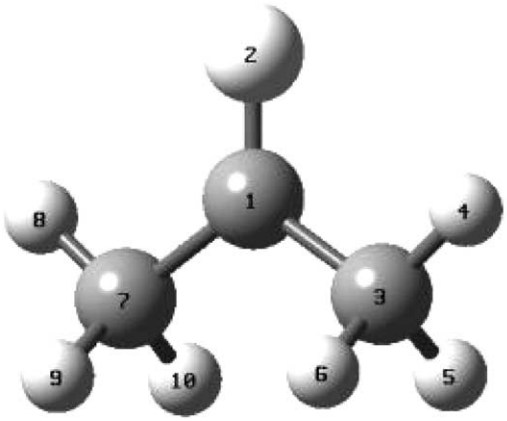

Fig. 3. Ball and stick representation of the structure of acetone optimized at the RHF/6-31G** level.

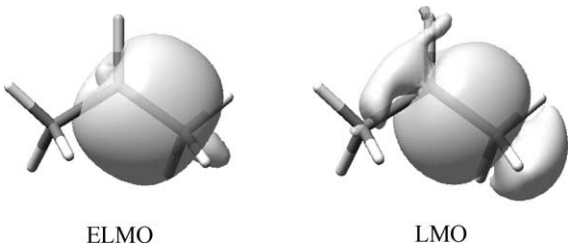

Fig. 4. Isosurfaces of the ELMO and LMO ($|\varphi(\mathbf{r})|^2 = 0.0009$ for each point $\mathbf{r}$ of the surface) describing one of the C–C bonds of acetone.

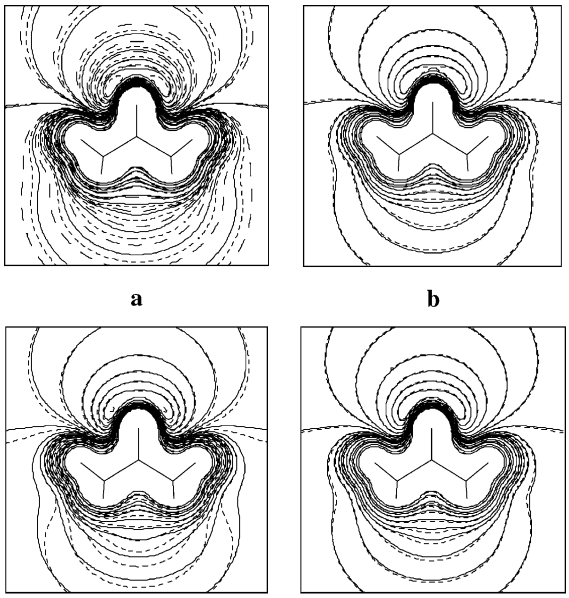

Fig. 5. Contour plots of electrostatic potential calculated along the molecular plane of the SCF optimized structure of acetone with the SCF (—) and (a) ELMO (- - -) and *td* LMO (– – –), (b) *rel* ELMO (- - -), (c) *t* ELMO (- - -) and (d) *rel-t* ELMO (- - -) wavefunctions.

has already been observed by Hierse et al. [39], who pointed out that ELMO 'relaxation' can increase significantly the accuracy of energy barriers evaluated with an ELMO calculation at DFT level. Electronic properties are improved as well, as one can see in Fig. 5(b), where electrostatic potentials generated by SCF and ELMO with partial relaxation (*rel* ELMO) wavefunctions are compared: it can be noted that the *rel* ELMO contours show deviations from the SCF electrostatic potential which are significantly smaller than the ELMO calculations (Fig. 5(a)). Moreover, in Table 2 we report the Mulliken population analysis performed with SCF, ELMO and *rel* ELMO wavefunctions: the partial 'relaxation' improves by one order of magnitude the overall deviation (σ) from the SCF values.

The complete absence of 'tails' makes ELMOs particularly suitable to be transferred from one molecule to another. We have thus performed on the test molecule a few preliminary calculations to study ELMOs transferability. In a recent article [40], it was shown that the electronic properties of some

Table 2
Mulliken populations (in electrons) relative to the SCF optimized structure of acetone (see Fig. 3) calculated from SCF, ELMO, ELMO with partial relaxation (*rel* ELMO), transferred ELMO (*t* ELMO) and transferred ELMO with partial relaxation (*rel-t* ELMO) wavefunctions

Atom	SCF	ELMO	*rel* ELMO	*t* ELMO	*rel-t* ELMO
C1	5.499	5.304	5.510	5.462	5.464
O2	8.516	8.559	8.497	8.500	8.537
C3	6.421	6.439	6.421	6.267	6.406
H4	0.841	0.855	0.837	0.916	0.850
H5	0.865	0.887	0.869	0.918	0.872
σ		0.101	0.011	0.092	0.022

Only 'unique' atoms are considered, since the conformation considered has C_{2v} symmetry. In the last row the root mean square deviation from SCF values (σ) is reported for each ELMO wavefunction.

ortho-substituted biphenyl molecules determined at the SCF level can be reproduced optimizing only the ELMOs associated to the substituent group, while keeping all the other ELMOs identical to those of biphenyl. In the present calculation ethane and formaldehyde were chosen to provide the building blocks of acetone: ELMOs calculated on ethane were used to describe the C–H and C–C bonds together with core electrons of methyl C atoms of acetone, while the carbonyl group was described by ELMOs calculated on formaldehyde. It can be seen from Table 2 that Mulliken populations calculated using transferred ELMOs without any further optimization (*t* ELMO) are better than expected, considering the crude approximation due to the use of the 'unpolarized' ELMO of ethane to describe the C–C bonds of acetone. Indeed, the chief deviations from SCF values stem from the populations of the methyl C and H atoms. The same trend is present in the electrostatic potential contours shown in Fig. 5(c), where the main departures from SCF contours are near the methyl groups. In Table 2 and Fig. 5(d) we report the results obtained by performing two standard SCF iterations on transferred ELMOs (named *rel-t* ELMO): it can be seen that this partial relaxation is sufficient to significantly improve both Mulliken populations and electrostatic potential contours. This behavior suggests that ELMOs describing small functional groups could be also employed to build-up good starting points for SCF calculations on large molecules.

Table 3
Geometric parameters of SCF and ELMO optimized structures of acetone, together with the percent deviation of the ELMO from the SCF values

	SCF	ELMO	Δ (%)
C1–O2	1.192	1.216	2.0
C1–C3	1.513	1.528	1.0
C3–H4	1.081	1.080	−0.2
C3–H5	1.086	1.084	−0.2
C3–C1–C7	116.70	119.18	2.1
C3–C1–O2	121.65	120.41	−1.0
C1–C3–H4	109.73	109.29	−0.4
C1–C3–H5	110.30	110.40	0.1

Units are Å for bond lengths and degrees for bond angles.

The possibility to perform geometry optimization with the ELMO wavefunction, which was mentioned in Section 2, turns out to be useful for both the generation of BSSE-free minimum structures and the optimization of systems described with a QM/MM method. Here we want to compare some geometric parameters of the ELMO and SCF optimized structures (see Table 3 and Fig. 3 for atom numbering). Both optimizations yielded the same conformation, in which the two methyl groups 'eclipse' the carbonyl group (see Fig. 3). It can be noted that the agreement between ELMO and SCF minimum geometries is good, with only the ELMO C–O bond length and C3–C1–C7 bond angle differing significantly from the SCF values. This behavior was already pointed out by other Authors, which showed that the observed variations were of the same order of magnitude as those which generally arise from a change of the basis set at the SCF level [11].

3.2. Description of the frontier regions in QM/MM methods

The QM/MM methods can be classified according to the strategy adopted to cope with the discontinuity created when two parts of a given system are represented at different levels of theory [41]. This problem is particularly complicated when the two regions strongly interact (e.g. through a covalent bond): in this case it cannot be solved at a satisfactory level in a simple and univocal way.

In the so-called 'link atom' methods the frontier bonds are 'cut' and the free valencies are saturated with 'capping' atoms, so that the resulting QM subsystem is well defined [14,15]. The link atoms are generally hydrogen atoms, but 'pseudoatoms' have also been used, to better reproduce the electron donating or electron withdrawing effect of the environment [42,43]. The obvious drawback of the link atom approach is that an extra non-physical atom is introduced in the original system. Some techniques try to eliminate the contribution of link atoms to the total energy adding molecular mechanical correction terms [17,44,45]. However, the definition of the potential surface is still difficult owing to the extra degrees of freedom relative to the link atoms [45].

Thus some approaches have been developed which 'merge' the classical frontier atom and the link atom in a single boundary atom with both QM and MM features. These methods ensure the correct behavior of the frontier bond using optimized semi-empirical parameters [18,20] or effective core potentials [19] for the boundary atom.

Another class of QM/MM techniques alternative to the link atom ones are those based on frozen localized frontier orbitals. They originate from an idea proposed by Warshel and Levitt [46], who employed hybrid orbitals to describe the QM/MM interface. Among these, we mention the local self-consistent field (LSCF) method, developed by Rivail and coworkers first at the semi-empirical [22] and then at the ab initio level [16,23], and the method of Friesner et al., characterized by the large effort in the parametrization of the interface interactions (both at the HF [21,47] and DFT [47,48] level).

In both these methods, the localized orbitals used to describe the frontier bonds are calculated on a model molecule (see later) by unitary transformation of the SCF canonical MOs and subsequent tail deletion: thus they are the *td* LMOs discussed in Section 3.1. Considering that the ELMOs reproduce better the SCF energy and other molecular properties (such as the electrostatic potential), we thought worth investigating their applicability to this kind of QM/MM methods.

Let us consider a system partitioned into a QM and a MM region, with the frontier atoms X (QM) and Y (MM) connected by a covalent bond (however the procedure can be readily generalized to more than one frontier bond). Correspondingly, the effective

hamiltonian of the total system can be written as

$$\hat{H}_{\text{eff}} = \hat{H}_{\text{QM}} + \hat{H}_{\text{MM}} + \hat{H}_{\text{QM/MM}},$$

where $\hat{H}_{\text{QM}}$ and $\hat{H}_{\text{MM}}$ concern only the QM and the MM region, respectively, while $\hat{H}_{\text{QM/MM}}$ comprises the interactions between the two parts. In the following we will describe the quantum mechanical part of the calculation (implemented at the Hartree–Fock level), postponing to future works the details on the calculation of the 'classical' contributions to $\hat{H}_{\text{MM}}$ and $\hat{H}_{\text{QM/MM}}$.

The electron density associated to the frontier bond is represented quantum mechanically through the localized orbital. Thus the total basis set used in the QM calculation comprises also the AOs centered on the MM frontier atom Y, which acquires a partial QM character. The quantum mechanical part of the QM/MM procedure differs from a traditional SCF calculation in the following aspects: (a) one of the MOs (the frontier ELMO determined in a previous step) is kept unchanged during the calculation; (b) the 'active' MOs are forbidden to use the AOs centered on the frontier atom Y, to prevent charge transfer to the MM subsystem [21]; (c) the electronic hamiltonian contains also the coulomb interactions between the QM electrons and the MM atoms, represented by point charges. The frontier ELMO in (a) is strictly localized on atoms X and Y and it is determined with an ELMO calculation on a model molecule built-up with the frontier atoms X and Y and as many atoms of the original system as needed to reproduce the chemical environment of the frontier bond [21]. The ELMO is then transferred to the system under study: during this operation one has to ensure that the orientation of the orbital matches that of the frontier bond and that its normalization is preserved. The points (a) and (b) can be achieved again by an ELMO calculation. Once the MOs of the QM subsystem and the frontier bond are defined, together with their different basis sets, the SCF procedure for the active QM subsystem is started, while the frontier orbital is kept frozen. This 'freezing' is due to the fact that the representation of the frontier region provided by the localized orbital should be as close as possible to that obtained with a QM calculation performed on the whole system. Indeed, the relaxation of the frontier orbital in a hybrid environment would produce artificial effects: it is to be considered that in the QM/MM calculation the atom Y has only a partial QM character (as only one of its electrons is explicitly represented) and the frontier bond is very close to the classical charges, so that the relaxed ELMO would be much different from that obtained with a full-QM calculation. The coulomb QM/MM interactions in (c) can be introduced using a modified 'core' operator

$$\hat{h}^*(1) = -\frac{\nabla_1^2}{2} - \sum_{\alpha=1}^{n_{\text{QM}}} \frac{Z_\alpha}{|\mathbf{R}_\alpha - \mathbf{r}_1|} - \sum_{p=1}^{n_{\text{MM}}} \frac{Q_{\text{p}}}{|\mathbf{R}_p - \mathbf{r}_1|},$$

where n_{QM} and n_{MM} are the total number of QM and MM atoms, respectively, while Z_α indicates the charge of the QM nucleus α and Q_{p} the charge of the MM atom p. The Q_{p} charges are taken from standard force fields, except for Q_{Y}, which is adjusted so that the total charge of the MM subsystem amounts to +1 (but other more 'refined' schemes can be used [21,23,49]).

As a first application of this procedure, we discuss here the following preliminary QM/MM calculations on the tripeptide GLY-HIS(ε)-GLY. The partitioning of the molecule into a QM and a MM region is shown in Fig. 6(a), where the QM and MM subsystems are represented with thick and thin sticks, respectively; black spheres indicate the QM and MM frontier atoms. In Fig. 6(b) the model molecule (represented with thick sticks) which was used for the calculation of the frontier orbital, is superimposed to the total system (represented with thin sticks); it is possible to

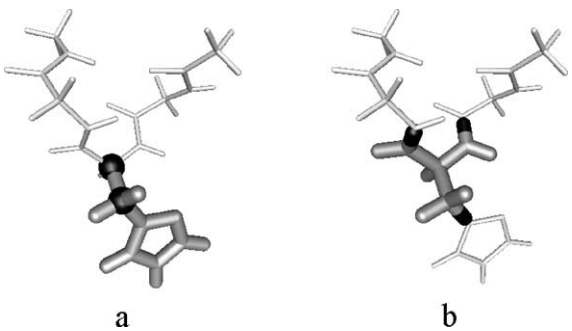

Fig. 6. (a) Representation of the partitioning of the GLY-HIS (ε)-GLY tripeptide into the MM region (thin sticks) and the QM region (thick sticks). Black spheres indicate the QM and MM frontier atoms. (b) Representation of the model molecule (thick sticks) superimposed to the tripeptide (thin sticks). Added hydrogen atoms (see text) are also visible in black.

Table 4
Mulliken populations (in electrons) of the GLY-HIS(ε)-GLY tripeptide calculated with SCF and QM/MM wavefunctions (see Fig. 7 for atom numbering)

Atom	SCF	QM/MM
C1	6.097	5.959
H2	0.928	0.967
H3	0.932	0.970
C4	5.953	5.948
C5	5.983	5.997
H6	0.920	0.926
N7	7.318	7.321
H8	0.763	0.769
C9	5.879	5.885
H10	0.916	0.922
N11	7.281	7.292

Root mean square deviation of QM/MM values from SCF is 0.05 electrons.

see also the hydrogen atoms (in black) which were added to fill the free valencies of the model molecule. The geometry of the tripeptide was optimized at the AM1 level [50]; owing to the preliminary nature of this study, the subsequent SCF and QM/MM calculations were performed with the STO-3G basis set. The 'classical' atom charges Q_p were taken from the AMBER force field [51]. Thanks to the relatively small dimensions of this molecule, it has been possible to perform also QM calculations on the entire system. In Table 4 we report the Mulliken populations of the QM subsystem calculated with both the SCF and the QM/MM wavefunction (see Fig. 7 for atom numbering). The general good agreement between the two sets of results is promising. As expected, the greater deviations of the QM/MM values from the SCF come from the atoms which are closer to the QM boundary.

3.3. Elimination of the intramolecular BSSE: an intramolecular hydrogen bond case

It is well known that the description of the intermolecular interaction at the traditional SCF level is affected by the basis set superposition error (BSSE). This error arises when two molecules approach each other, as one molecule can begin to use also the atomic functions centered on the other molecule to describe its wavefunction, yielding in this way an artificial lowering of the energy. One of

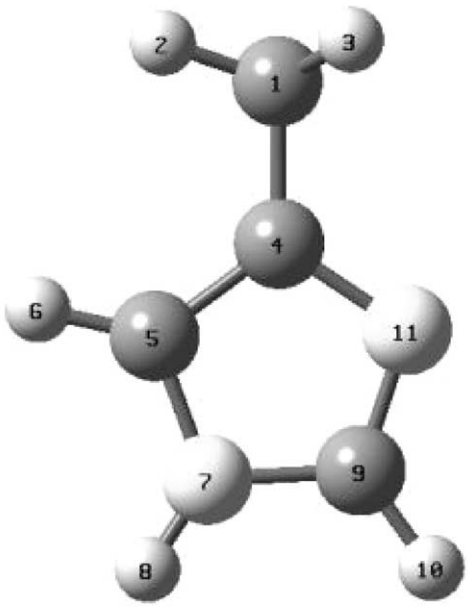

Fig. 7. Ball and stick representation of the QM region, with atom numbering.

the most used techniques to correct a posteriori this error is the CP method proposed by Boys and Bernardi [27]. However, there are drawbacks inherent to this approach [52–54]. The SCF–MI wavefunction [28,29], which avoids a priori this error, has been recently used for molecules ranging from water dimer [55] to DNA pairs [56].

Elimination of the intramolecular BSSE is a problem much less discussed. However, it is well known that the energy difference between conformations of a molecule depends on the basis set. This effect can be partially ascribed to the fact that energy calculations on different conformations are affected by a bias of the basis set, as the nuclei change position carrying their AOs, which can be better suited to describe other parts of the molecule [25]. The main difficulty in the elimination of this intramolecular BSSE is that it is not possible to uniquely determine and isolate the interacting fragments.

Previous attempts to eliminate intramolecular BSSE were all based on modifications of the CP method. Brickmann et al. [24], in their work on the relative stability of 1,2-ethanediol conformers, proposed two methods for the correction of BSSE: (1) in the additive CP (aCP) approach the greater flexibility

of the basis set in proximity of the OH groups in hydrogen-bonded conformers is compensated adding ghost atoms to the non H-bonded ones; (2) in the equivalence CP (eCP) method the intramolecular BSSE for each minimum conformation is evaluated by a standard CP calculation on a water dimer, with the relative orientation of the OH groups in the glycol conformer.

Jensen [25] suggested a new strategy and built-up a common basis for two different conformations of a given molecule superimposing the two structures and replacing the atoms of one conformation with ghost atoms. However, as stated by Jensen himself, this procedure may become computationally very demanding for large molecules. In addition, increasing the size of the basis set, problems of linear dependence can arise. Moreover, Senent et al. [26] pointed out that, if more than two conformations have to be compared, BSSE corrections evaluated with this method are additive (and thus consistent) only if the same basis set for each structure is used: this means that the basis for one conformation has to contain ghost atoms in place of the 'true' atoms of *all* the other structures. Another complication arises from geometry relaxation effects.

The use of the ELMO wavefunction to avoid intramolecular BSSE appears as a natural extension of the SCF–MI method developed for the study of intermolecular interactions. Indeed, strictly localized MOs allow the description of a given group of atoms in a molecule through a given subset of AOs, thereby preventing the use of 'forbidden' basis functions.

We present here our preliminary results on 1,2 ethanediol, already studied by Brickmann et al.[24]: we performed traditional SCF and ELMO calculations on ten conformations of the molecule (see Fig. 8), whose structures were found starting from the geometries reported by Brickmann et al. [24], employing various basis sets (i.e. 6-31G**, 6-311G** and 6-311++G(2d,2p)). The nomenclature of the conformers was taken from Ref. [24]: small letters relate to torsion around C–O bonds (*t* means a *trans* disposition, i.e. a dihedral angle H–O–C–C $\sim 180°$, while *g* and g' mean a *gauche* disposition, i.e. H–O–C–C $\sim 60°$ and $\sim -60°$, respectively), while capital letters describe the arrangement of the O–C–C–O torsion, with an analogous meaning.

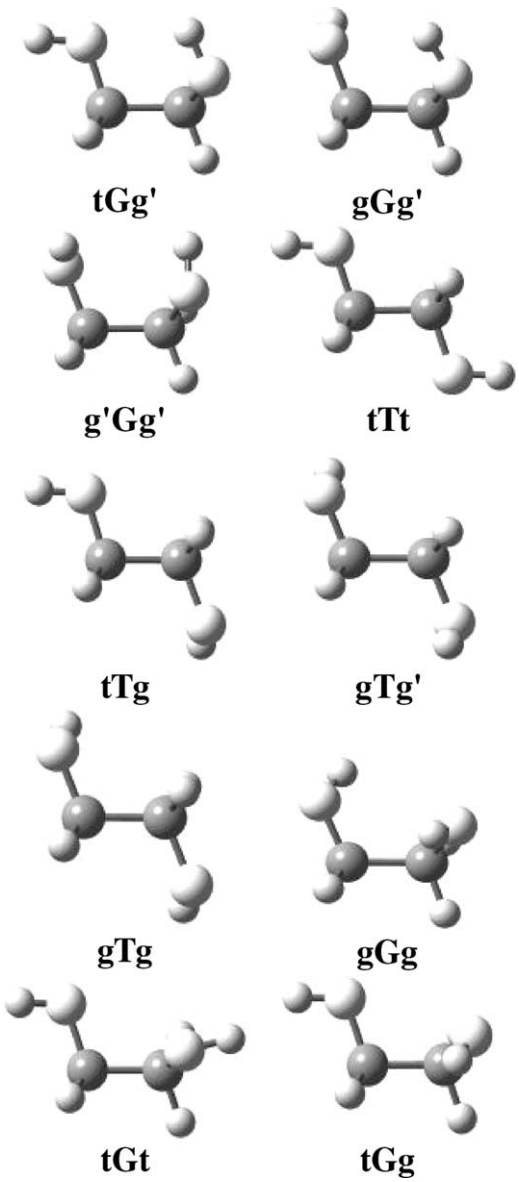

Fig. 8. Ball and stick representation of the ten conformations of 1,2-ethanediol (see text for nomenclature).

The SCF energies of the ten conformations relative to the most stable one (i.e. tGg′), are reported in Table 5 for each basis set. It can be noted that the first three conformers, due to the *G* torsion at the central bond and a particularly favourable interaction between the two OH groups (see Fig. 8), can be defined as hydrogen-bonded. The other three *G* conformations, which are not stabilized by an

Table 5
SCF and ELMO relative energies (in kcal/mol) of the ten conformations of 1,2-ethanediol, calculated with the 6-31G**, 6-311G** and 6-311++G(2d,2p) basis sets

	6-31G**			6-311G**			6-311++G(2d,2p)		
	SCF	ELMO	$BSSE_{rel}$	SCF	ELMO	$BSSE_{rel}$	SCF	ELMO	$BSSE_{rel}$
tGg′	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
gGg′	0.64	0.79	−0.15	0.72	1.01	−0.29	0.67	0.74	−0.07
g′Gg′	1.28	1.16	0.12	1.50	1.35	0.15	1.01	0.98	0.03
tTt	2.02	1.37	0.65	1.89	1.32	0.57	1.74	1.46	0.28
tTg	2.36	1.80	0.56	2.33	1.93	0.40	2.11	1.91	0.20
gTg′	2.43	2.01	0.42	2.49	2.26	0.23	2.30	2.18	0.12
gTg	2.80	2.45	0.35	2.86	2.72	0.14	2.60	2.52	0.08
gGg	3.36	2.79	0.57	3.29	2.94	0.35	2.98	2.88	0.10
tGt	3.66	3.18	0.48	3.50	3.04	0.46	2.90	2.94	−0.04
tGg	4.10	3.49	0.61	4.01	3.58	0.43	3.46	3.36	0.10

The geometry of each conformation has been minimized at SCF level for each basis set. Relative BSSE values (calculated as $BSSE_{rel}$ = SCF-ELMO) are also reported.

intramolecular hydrogen bond, prove to be less stable than *T* conformations mainly for steric reasons.

The ELMO calculations used a coefficient matrix corresponding to the localization scheme shown in Table 6 (see Fig. 9 for atom numbering): in this way no ELMO was allowed to use AOs belonging to the O3–H5 group *and* AOs belonging to the O4–H6 group, while ELMOs relative to C–H and C–C bonds, which do not directly participate in the hydrogen bond, were partially delocalized according to a 'next neighbours' criterion. Moreover, some delocalization of the oxygen lone pairs was allowed, to improve the description of polarization effects and recover a significant part of the SCF energy. Core electrons, which are not reported in Table 6, were strictly localized on their corresponding atoms.

The ELMO relative energies are shown in Table 5, together with the values of relative BSSE, calculated as difference between SCF and ELMO relative energies. It is worth noting that these BSSE values are internally consistent: if $BSSE_{rel}(x, y)$ is the BSSE of the species y relative to the x, $BSSE_{rel}(x, z)$-$BSSE_{rel}(x, y)$ can be taken as the BSSE of the species z relative to the y. Moreover, only one ELMO calculation is required for each conformer. It can be seen that the values of $BSSE_{rel}$ for H-bonded

Table 6
Localization scheme used in the ELMO calculations on 1,2-ethanediol

ELMO	AOs
C1–O3	C1, O3
O3–H5	O3, H5
C2–O4	C2, O4
O4–H6	O4, H6
C1–C2	C1, C2, H7, H8, H9, H10
C1–H7 and C1–H8	C1, H7, H8, C2, O3
C2–H9 and C2–H10	C2, H9, H10, C1, O4
O3 lone pairs	O3, C1, H5
O4 lone pairs	O4, C2, H6

ELMOs labelled in the first column with the corresponding bond or lone pair can use AOs centered on the atoms indicated in the second column.

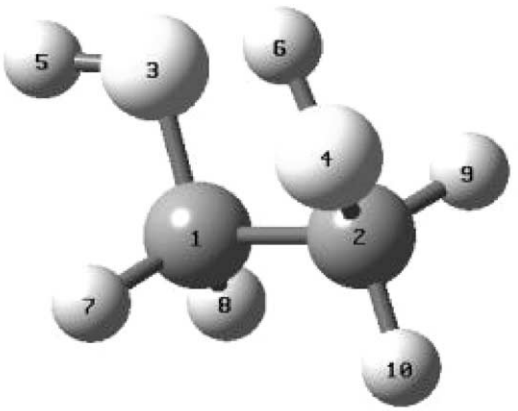

Fig. 9. Ball and stick representation of the tGg' conformation of 1,2-ethanediol with atom numbering.

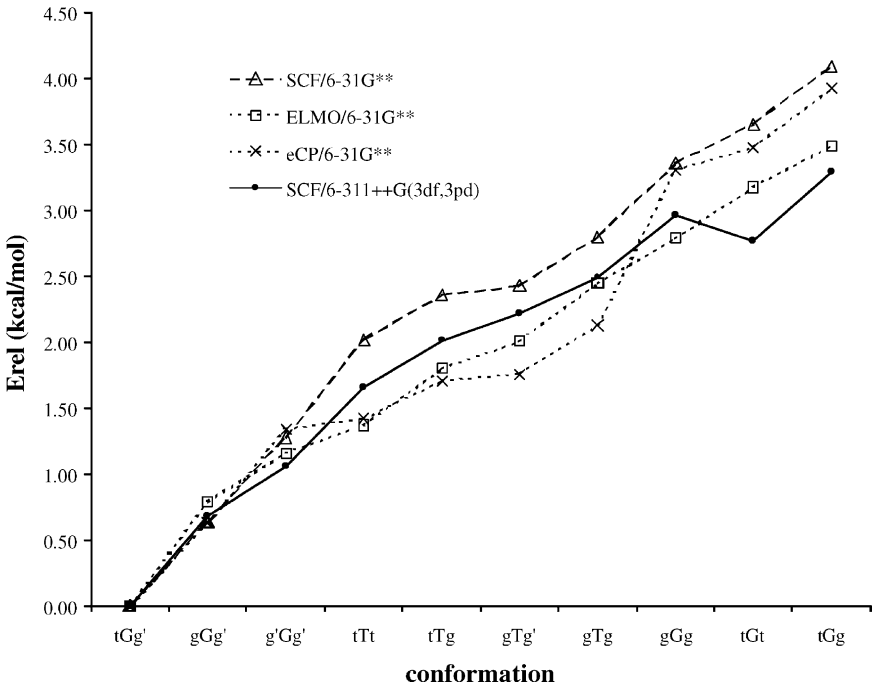

Fig. 10. SCF, ELMO and eCP relative energies (in kcal/mol) of the ten conformations of 1,2-ethanediol calculated with the 6-31G** and (only for SCF calculations) 6-311++G(3df, 3pd) basis sets on SCF/6-31G** optimized structures. The eCP values are taken from Ref. [24].

conformations are generally smaller than those for the non H-bonded ones: this can be explained considering that the reference conformation contains an intramolecular hydrogen bond. Furthermore, it can be observed that the values of $BSSE_{rel}$ calculated with the largest basis set decrease with respect to the smaller one, as expected.

To test the reliability of the ELMO approach to the elimination of the intramolecular BSSE, we compared both SCF and BSSE-corrected SCF calculations performed using the smallest of the basis sets considered (6-31G**), with SCF values obtained with a much larger basis set, namely the 6-311++G(3df,3pd) (see Fig. 10). In addition to the ELMO BSSE-corrected values, also the eCP relative energies previously calculated by Brickmann et al. [24] are reported. As SCF/6-311++G(3df,3pd) relative energies are supposed to be much less affected by BSSE than SCF/6-31G**, we can conclude that a good method for eliminating BSSE has to correct the values obtained with the small basis set so that they approach the results yielded by the larger. This seems to be the behavior of both the ELMO and eCP correction schemes, even if the ELMO approach in general proves to be more effective, particularly for the conformers with high relative energies. These observations suggest that the ELMO approach can be an alternative method to correct energy differences which are affected by intramolecular BSSE.

4. Conclusions

The ELMOs, contrary to the orbitals obtained from the a posteriori localization techniques (LMOs), are characterized by the complete absence of 'tails'. This feature allows to associate in a rigorous way the MOs to 'molecular fragments', whose size can be chosen a priori selecting the more appropriate localization pattern. So we can obtain ELMOs which describe single bonds, lone pairs or functional groups, to be used to assemble more complicated wavefunctions.

In this work we have presented the development of a robust algorithm for the determination of

the ELMOs; their intrinsic non-orthogonality makes convergence difficult, so that Newton–Raphson techniques were needed. Our code has been interfaced with the PC-GAMESS package [37,38] and it can be used to perform both single point (conventional and direct) energy calculations and geometry optimizations.

Furthermore, two possible applications of the ELMOs have been discussed. The first one concerns the description of the frontier regions in systems modelled with a mixed QM/MM method. As in the LSCF method of Rivail et al. [16], LMOs orbitals computed on a model molecule can be used to represent the frontier bonds between the QM and the MM part of the real molecule. Here we have shown how to use ELMOs in this context and we have presented our first calculations on the GLY-HIS(e)-GLY tripeptide. The ELMO QM/MM results relative to the Mulliken population analysis are in good agreement with the description of the system at the 'pure' QM level. The inclusion of the MM and QM/MM 'classical' contributions to the total energy and the implementation of energy first derivatives with respect to the atomic coordinates, as well as a more extensive testing of the method, are work in progress.

The second application of the ELMOs presented concerns the description of the intramolecular hydrogen bonding, which in a traditional SCF calculation is affected by the intramolecular BSSE. In this case the widely used CP correction [27] cannot be directly applied and some peculiar strategies have to be adopted [24–26]. The a priori ELMO-based method which is here suggested can be considered as a natural development of the SCF–MI approach to the study of intermolecular interactions [28]. We used the ELMO wavefunction in the study of the relative stability of ten conformations of 1,2 ethanediol and the results indicate that this approach can be one of the possible strategies to avoid intramolecular BSSE. It is worth pointing out that this method is not affected by the huge increment of the basis set which characterizes the CP-derived a posteriori approaches when the number of conformations to be compared increases. In addition it permits to include in a simple way the geometry relaxation effects.

Acknowledgements

We want to thank Dr Massimilano Meli for technical support.

References

[1] S.F. Boys, Rev. Mod. Phys. 32 (1960) 296.
[2] C. Edmiston, K. Ruedenberg, J. Chem. Phys. 43 (1965) S97.
[3] J. Pipek, P.G. Mezey, J. Chem. Phys. 90 (1989) 4916.
[4] P.D. Walker, P.G. Mezey, J. Am. Chem. Soc. 116 (1994) 12022.
[5] W. Yang, Phys. Rev. Lett. 66 (1991) 1438.
[6] W. Yang, J. Chem. Phys. 94 (1991) 1208.
[7] Q. Zhao, W. Yang, J. Chem. Phys. 102 (1995) 9598.
[8] S.L. Dixon, K.M. Merz Jr., J. Chem. Phys. 104 (1996) 6643.
[9] S.L. Dixon, K.M. Merz Jr., J. Chem. Phys. 107 (1997) 879.
[10] G. Náray-Szabó, Comput. Chem. 24 (2000) 287.
[11] G.F. Smits, C. Altona, Theor. Chim. Acta 67 (1985) 461.
[12] R. McWeeny, Methods of Molecular Quantum Mechanics, Academic Press, London, 1992.
[13] J. Rubio, A. Povill, J.P. Malrieu, P. Reinhardt, J. Chem. Phys. 107 (1997) 10044.
[14] U.C. Singh, P.A. Kollman, J. Comput. Chem. 7 (1986) 718.
[15] M.J. Field, P.A. Bash, M. Karplus, J. Comput. Chem. 11 (1990) 700.
[16] X. Assfeld, J.-L. Rivail, Chem. Phys. Lett. 263 (1996) 100.
[17] M. Svensson, S. Humbel, R.D.J. Froese, T. Matsubara, S. Sieber, K. Morokuma, J. Phys. Chem. 100 (1996) 19357.
[18] J. Gao, P. Amara, C. Alhambra, M.J. Field, J. Phys. Chem. A 102 (1998) 4714.
[19] Y. Zhang, T.-S. Lee, W. Yang, J. Chem. Phys. 110 (1999) 46.
[20] I. Antes, W. Thiel, J. Phys. Chem. A 103 (1999) 9290.
[21] D.M. Philipp, R.A. Friesner, J. Comput. Chem. 20 (1999) 1468.
[22] V. Théry, D. Rinaldi, J.-L. Rivail, B. Maigret, G.G. Ferenczy, J. Comput. Chem. 15 (1994) 269.
[23] N. Ferré, X. Assfeld, J.-L. Rivail, J. Comput. Chem. 23 (2002) 610.
[24] S. Reiling, J. Brickmann, M. Schlenkrich, P.A. Bopp, J. Comput. Chem. 17 (1996) 133.
[25] F. Jensen, Chem. Phys. Lett. 261 (1996) 633.
[26] M.L. Senent, S. Wilson, Int. J. Quantum Chem. 82 (2001) 282.
[27] S.F. Boys, F. Bernardi, Mol. Phys. 19 (1970) 553.
[28] E. Gianinetti, M. Raimondi, E. Tornaghi, Int. J. Quantum Chem. 60 (1996) 157.
[29] A. Famulari, E. Gianinetti, M. Raimondi, M. Sironi, Int. J. Quantum Chem. 69 (1998) 151.
[30] M. Sironi, A. Famulari, Theor. Chem. Acc. 103 (1999) 417.
[31] A. Szabo, N.S. Ostlund, Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory, McGraw-Hill, New York, 1989.
[32] H. Stoll, G. Wagenblast, H. Preuss, Theor. Chim. Acta 57 (1980) 169.

[33] M. Couty, C.A. Bayse, M.B. Hall, Theor. Chem. Acc. 97 (1997) 96.
[34] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes: the Art of Scientific Computing, Cambridge University Press, Cambridge, UK, 1986.
[35] S.M. Goldfeld, R.E. Quandt, H.F. Trotter, Econometrica 34 (1966) 541.
[36] Y. Yamaguchi, Y. Osamura, J.D. Goddard, H.F. Schaefer III, A New Dimension to Quantum Chemisty: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure Theory, Oxford University Press, Oxford, UK, 1994.
[37] A.A. Granovsky, www http://classic.chem.msu.su/gran/gamess/index.html
[38] M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.J. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Su, T.L. Windus, M. Dupuis, J.A. Montgomery, J. Comput. Chem. 14 (1993) 1347.
[39] W. Hierse, E.B. Stechel, Phys. Rev. B 54 (1996) 16515.
[40] M. Sironi, A. Famulari, M. Raimondi, S. Chiesa, J. Mol. Struct. (Theochem) 529 (2000) 47.
[41] J. Gao, in: K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 7, VCH Publishers, New York, 1996, Chapter 3.
[42] Hyperchem Inc., Hyperchem Users Manual, Computational Chemistry, Hypercube Inc., Ontario, Canada, 1994.
[43] P.L. Cummins, J.E. Gready, J. Phys. Chem. B 104 (2000) 4503.
[44] F. Maseras, K. Morokuma, J. Comput. Chem. 16 (1995) 1170.
[45] D. Bakowies, W. Thiel, J. Phys. Chem. 100 (1996) 10580.
[46] A. Warshel, M. Levitt, J. Mol. Biol. 103 (1976) 227.
[47] R.B. Murphy, D.M. Philipp, R.A. Friesner, J. Comput. Chem. 21 (2000) 1442.
[48] R.B. Murphy, D.M. Philipp, R.A. Friesner, Chem. Phys. Lett. 321 (2000) 113.
[49] R.J. Hall, S.A. Hindle, N.A. Burton, I.H. Hillier, J. Comput. Chem. 21 (2000) 1433.
[50] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, J. Am. Chem. Soc. 107 (1985) 3902.
[51] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, J. Am. Chem. Soc. 117 (1995) 5179.
[52] E.R. Davidson, S.J. Chakravorty, Chem. Phys. Lett. 217 (1994) 48.
[53] G. Karlstrom, A.J. Sadley, Theor. Chim. Acta 61 (1982) 1.
[54] F.B. van Duijneveldt, J.M.C.M. van Duijneveldt-van de Rijdt, J. H. van Lenthe, Chem. Rev. 94 (1994) 1873.
[55] A. Famulari, M. Raimondi, M. Sironi, E. Gianinetti, Chem. Phys. 232 (1998) 275.
[56] A. Famulari, F. Moroni, M. Sironi, E. Gianinetti, M. Raimondi, J. Mol. Struct. (Theochem) 529 (2000) 209.