Theor Chem Acc (2004) 112: 247–253
DOI 10.1007/s00214-004-0584-1

Theoretical
Chemistry Accounts

Regular article

Determination of extremely localized molecular orbitals in the framework of density functional theory

Emiliano Burresi[1], **Maurizio Sironi**[1,2]

[1] Dipartimento di Chimica Fisica ed Elettrochimica, Universita' di Milano, Via Golgi 19, 20133, Milan, Italy
[2] Centre for Biomolecular Interdisciplinary Studies and Industrial Applications CISI, Via Fratelli Cervi n. 93, Palazzo LITA, 20090, Segrate (MI), Italy

Received: 16 July 2003 / Accepted: 18 December 2003 / Published oline: 8 June 2004

Abstract. Extremely localized molecular orbitals are rigorously localized on only a preselected set of atoms and do not have any tails outside the localization region. The importance of these orbitals lies in their ability to be transferred from one molecule to another one. A new algorithm to determine extremely localized molecular orbitals in the framework of the density functional theory method is presented. This could also be a valuable tool in the quantum mechanics/molecular mechanics methodology where localized molecular orbitals are used to describe covalent bonds across the frontier region. The present approach is used to build up the electron density of thymopentin, a polypeptide constituted by five residues, starting from extremely localized molecular orbitals determined on a set of model molecules. The results obtained confirm good transferability properties for these orbitals.

Keywords: Extremely localized molecular orbitals – Density functional theory method – Transferability

Introduction

The methods of computational chemistry have been applied so far to a variety of small and large molecules. Even though a large body of results was obtained within the Hartree–Fock (HF) approach, the importance to take properly into account the effects of electronic correlation was widely demonstrated. To this aim, several methods have been developed and are routinely applied using the more diffuse packages of computational chemistry. Among them, the density functional theory (DFT) [1] has attracted and continues to attract the attention of theoreticians, as it represents a very powerful method to introduce the correlation while retaining quite limited computational requirements.

In the case of molecules of biological interest, however, such rigorous treatments would still be infeasible, owing to their large dimensions. This difficulty can be overcome by recognizing that the biological activity of such systems is sometimes exerted in only a small region of the molecule, the remaining part of it simply acting as a perturbing moiety. It is then possible to describe accurately, for example, at the quantum mechanics (QM) level, only the more interesting region of the molecule (e.g., the active site of a protein), and to treat the remaining part of the molecule at a lower level of theory, using, for example, molecular mechanics (MM). As the former region can be involved in the disruption of some bonds and in the formation of new ones, it is expected that a considerable amount of electronic reorganization can arise. This requires the use of methods which take into account the electronic correlation, and the DFT approach owing to its lower computational cost can be an obvious choice in the treatment of these large systems.

A large variety of such QM/MM hybrid methods have been developed [2, 3]. It should be observed that these methods are far from trivial when the classical and quantum regions of the system investigated are connected by one or more covalent bonds. Among the most promising methods which are not based on the use of nonphysical atoms [4], we recall to the work of Thery et al. [5], which has recently been generalized by Rivail's group [6, 7] and by Murphy et al.[8, 9].

These approaches describe the covalent bonds connecting the QM and the MM regions through localized molecular orbitals (LMOs), which can be transferred from model molecules after a traditional localization procedure [10, 11, 12]. An ab initio calculation is then carried out on the QM region in the presence of the frozen LMOs which describe the connections to the MM region. In the ab initio cal-

Proceedings of the 11th International Congress of Quantum Chemistry satellite meeting in honor of Jean-Louis Rivail

Correspondence to: Maurizio Sironi
e-mail: maurizio.sironi@unimi.it

culation the MOs of the quantum region and the LMOs are described using different basis sets that share some common atomic functions; the traditional ab initio algorithms were properly modified to account for this. Applications to large systems have been already carried out, so demonstrating the power offered by these methods [13].

Of course, these approaches rely on the concept of transferability, which is well established in the chemical view of a molecule, usually thought of as an assembly of different functional groups each of them possessing well-known properties.

We observe that LMOs are characterized by the presence of tails outside the localized region, which are necessary to preserve the orthogonality of the orbitals. These tails cannot be transferred from the model molecule to the target one, so they are annihilated by simply zeroing the corresponding coefficients. This procedure is therefore associated with a nonnegligible increase in the energy value.

In our laboratory we have recently investigated algorithms to determine extremely localized molecular orbitals (ELMOs), i.e., orbitals which are rigorously defined on only some molecular fragments and then characterized by a complete absence of tails. ELMOs or strictly localized molecular orbitals [14] or nonorthogonal localized molecular orbitals [15] were introduced a long time ago [16], but they are now the subject of renewed interest [17, 18, 19, 20]. The complete absence of tails is an appealing feature of the ELMOs, as they can be directly transferred without perturbation. They could represent a useful tool in both the QM/MM methodologies which use localized orbitals to describe the frontier regions and the methods which aim to assemble the electronic structure of a large molecule using molecular fragments defined on model molecules.

We have already tested the transferability of the ELMOs [21] and we showed that electronic properties of some ortho-substituted biphenyl molecules determined at the HF level are quite well reproduced by optimizing only the ELMOs defined on the substituent group, while keeping all the other ELMOs identical to those of the unsubstituted biphenyl molecule. These works suggest that the ELMOs provide good results owing to the fact that they can be transferred without deletion of tails.

In addition we have shown that if the transferred ELMOs are subjected to a relaxation, it is possible to obtain a close agreement with respect to the HF results. This step can be carried out by performing just a single self-consistent-field (SCF) iteration using the transferred ELMOs. Alternatively, we have recently devised a procedure based on a valence bond approach, denoted ELMO-VB [22], which permits us to take advantage of the extremely localized nature of the orbitals to obtain very compact wavefunctions, giving considerable improvement in the results.

The results obtained so far using the ELMOs at the restricted HF (RHF) level prompted us to extend this approach to the DFT framework. In this paper we present our first implementation of the method together with a preliminary test calculation. This approach could also be used to avoid a priori the basis set superposition error in the determination of inter-molecular and intramolecular interactions. In these cases it could be considered as a useful tool to introduce electronic correlation in the SCF for molecular interaction method proposed by Gianinetti et al. [23] for the intermolecular forces and in the ELMO method which has been already used to study the intramolecular hydrogen bond [18] at the RHF level.

Theory

In the framework of DFT [1], we can express the energy of a system as

$$E[\rho] = T_s[\rho] + \int v(r)\rho(r)\mathrm{d}r + E_{xc}[\rho] + \frac{1}{2}\iint \frac{\rho(r)\rho(r')}{|r-r'|}\mathrm{d}r\mathrm{d}r' + \sum_{A<B}\frac{Z_A Z_B}{R_{AB}}, \quad (1)$$

where $T_S[\rho]$ is the kinetic energy of a noninteracting electron gas in its ground state with electron density $\rho(r)$, $v(r)$ is the external potential, i.e., the electrostatic field generated by the nuclei, $E_{xc}[\rho]$ is the exchange–correlation energy, $\frac{1}{2}\iint \frac{\rho(r)\rho(r')}{|r-r'|}\mathrm{d}r\mathrm{d}r'$ is the classical Coulomb description of the repulsion of the electrons, and the last term is the nuclear repulsion energy.

The noninteracting electron gas in its ground state is usually described by a Slater determinant built up using N orthogonal Kohn–Sham orbitals (where $2N$ is the number of electrons), i.e.,

$$\rho(r) = 2\sum_{i}^{N}\phi_i^2(r). \quad (2)$$

The Kohn–Sham orbitals are determined by an iterative solution of the Kohn–Sham equations:

$$H_{KS}\phi_i = \left(-\frac{1}{2}\nabla^2 + v_{eff}(r)\right)\phi_i(r) = \varepsilon_i\phi_i(r), \quad (3)$$

where $v_{eff}(r) = v(r) + \int \frac{\rho(r')}{|r-r'|}\mathrm{d}r' + v_{xc}(r)$, with $v_{xc}(r) = \delta E_{xc}[\rho]/\delta\rho(r)$.

We are now faced with the problem to determine Kohn–Sham ELMOs, i.e., orbitals which are defined on a preselected subset of atoms. The partitioning choice depends on the successive use of the ELMOs, and it is usually based on chemical ideas. So it is possible to define ELMOs which describe single bonds, lone pairs, etc., or ELMOs which describe a larger set of atoms, for example, a residue in a polypeptide molecule. Hence, in general, we can consider dividing a molecule into n_f molecular fragments, each of them defined by its own basis set, $\left(\chi_\mu^i\right)_{\mu=1}^{m_i}$, where i denotes the generic fragment.

Each fragment can share some atomic basis functions with other fragments. On the basis of the chemical nature of the fragment, we will assign to each of them $2N_i$ electrons which will be described by N_i doubly occupied ELMOs φ_α^i defined with the usual linear combination of atomic orbitals (LCAOs) limited to the AOs of their own fragment, i.e.,

$$\varphi_\alpha^i = \sum_\mu^{m_i} c_{\mu\alpha}^i \chi_\mu^i \text{ or } \varphi_\alpha^i = \sum_\mu^{M} c_{\mu\alpha}^i \chi_\mu, \quad \text{where } c_{\mu\alpha}^i = 0 \text{ if } \chi_\mu \notin i\text{th fragment}, \tag{4}$$

where M is the total number of unique basis functions.Of course each fragment will also be characterized by m_i–N_i virtual ELMOs. In order to determine the ELMOs using the variation principle, we generalize a procedure described in Refs. [24, 25]. It should be noted that the set of all occupied ELMOs must be linearly independent as they enter in the Slater determinant describing the noninteracting electronic gas. As the fragments usually share some atomic functions, $M_{\text{tot}} = \sum_i^{n_f} m_i$ is, in general, greater than M so the set of virtual orbitals of all fragments can present a linear dependence, which must be correctly taken into account.

Let us indicate with $\mathbf{C}^{(0)}$ the matrix containing all the coefficients of the occupied and virtual ELMOs for each fragment; it will be a rectangular matrix of dimension $M \times M_{\text{tot}}$ with the following structure:

$$\mathbf{C}^{(0)} = \left| C_1^{\text{occ}} \dots C_i^{\text{occ}} \dots C_{n_f}^{\text{occ}} C_1^{\text{vir}} \dots C_i^{\text{vir}} \dots C_{n_f}^{\text{vir}} \right|, \tag{5}$$

where $C_i^{\text{occ}}/C_i^{\text{vir}}$ represent the occupied/virtual orbitals of the ith fragment, respectively.

At the beginning, the $\mathbf{C}^{(0)}$ matrix is built up using a proper guess, usually the localized Kohn–Sham orbitals obtained from the first iteration of a traditional DFT calculation, followed by tail deletion and proper renormalization. Owing to the extremely localized nature of the orbitals, as expressed by Eq. (4), the $\mathbf{C}^{(0)}$ matrix is sparse and its orbitals are not orthogonal, so it is not possible to express the electron density through Eq. (2).

The procedure that we have implemented optimizes cyclically the set of ELMOs of each fragment. Let us consider the steps involved to optimize the ELMOs of the ith fragment. Starting from the matrix $\mathbf{C}^{(0)}$, we assemble the matrix $\mathbf{C}^{(1)}$ just reordering the orbitals contained in $\mathbf{C}^{(0)}$:

$$\mathbf{C}^{(1)} = \left| C_1^{\text{occ}} \dots C_{i-1}^{\text{occ}} C_{i+1}^{\text{occ}} \dots C_{n_f}^{\text{occ}} C_i^{\text{occ}} C_i^{\text{vir}} C_1^{\text{vir}} \dots C_{i-1}^{\text{vir}} C_{i+1}^{\text{vir}} \dots C_{n_f}^{\text{vir}} \right| = \left| C_{(i)}^{\text{occ}} C_i^{\text{occ}} C_i^{\text{vir}} C_{(i)}^{\text{vir}} \right|,$$

where $C_{(i)}^{\text{occ}}/C_{(i)}^{\text{vir}}$ represent all the occupied/virtual orbitals except those of the ith fragment to be optimized.

This matrix is then subjected to a Gram–Schmidt orthogonalization, giving the matrix $\mathbf{C}^{(\text{GS})} = \left| \tilde{C}_{(i)}^{\text{occ}} \tilde{C}_i^{\text{occ}} \tilde{C}_i^{\text{vir}} \tilde{C}_{(i)}^{\text{vir}} \right|$. As is well known, the Gram–Schmidt orthogonalization keeps the first vector unaltered, orthogonalizes the second on the first one, then the third vector on the first one, and so on. If a vector is a linear combination of the previous orthogonalized vectors, it is annihilated by the procedure.

Owing to the adopted sequence of the orbitals in the $\mathbf{C}^{(1)}$ matrix, a $\mathbf{C}^{(\text{GS})}$ matrix is obtained where

1. The first columns, $\tilde{C}_{(i)}^{\text{occ}} \tilde{C}_i^{\text{occ}}$, contain a new set of orthogonal orbitals which are a mixture of only the occupied orbitals $C_{(i)}^{\text{occ}} C_i^{\text{occ}}$, so the associated wavefunction is not changed. They allow us to assemble the electron density using Eq. (2).

2. The central part, $\tilde{C}_i^{\text{occ}} \tilde{C}_i^{\text{vir}}$, contains the new set of Gram–Schmidt orthogonalized occupied and virtual orbitals of the fragment to be optimized; they define an adequate space to be used for a proper projection of the Kohn–Sham operator from the atomic basis to the molecular basis (see later).

Of course the matrix $\mathbf{C}^{(\text{GS})}$ is a square matrix of order M, so if $M_{\text{tot}} > M$ some virtual orbitals are annihilated from the orthogonalization step. In particular some virtual orbitals of the ith fragment can also be eliminated during this step. Let us indicate with m_i' the number of orbitals of the ith fragment that survive after the Gram–Schmidt step ($m_i' \le m_i$).

Using the occupied orbitals contained in the $\mathbf{C}^{(\text{GS})}$ matrix, $\tilde{C}_{(i)}^{\text{occ}} \tilde{C}_i^{\text{occ}}$, the density matrix and hence the Kohn–Sham operator on the atomic basis functions, $\mathbf{F}_{\text{KS}}^{\text{AO}}$, are obtained by using a traditional DFT package.

The Kohn–Sham matrix is then transformed on the MO basis, $\mathbf{F}_{\text{KS}}^{\text{MO}}$, using the matrix $\mathbf{C}^{(\text{GS})}$:

$$\mathbf{F}_{\text{KS}}^{\text{MO}} = \mathbf{C}^{(\text{GS})\dagger} \mathbf{F}_{\text{KS}}^{\text{AO}} \mathbf{C}^{(\text{GS})}.$$

The diagonal block of the $\mathbf{F}_{\text{KS}}^{\text{MO}}$ matrix, labelled by the orbitals $\tilde{C}_i^{\text{occ}} \tilde{C}_i^{\text{vir}}$ of the ith fragment, $\mathbf{F}_{\text{KS}}^{\text{MO}}(i)$, is selected. This is a square matrix of order m_i'. The optimized orbitals of the ith fragment are now computed by evaluating the eigenvectors $\mathbf{C}_i^{\text{MO}}$ of the $\mathbf{F}_{\text{KS}}^{\text{MO}}(i)$ matrix, which are then transformed on the atomic basis obtaining the rectangular $(M \times m_i')$ matrix $\mathbf{C}_i^{\text{AO}}$. The new orbitals are so obtained in the space defined by the orbitals $\tilde{C}_i^{\text{occ}} \tilde{C}_i^{\text{vir}}$ which contain contributions from the occupied orbitals of the other fragments owing to the Gram–Schmidt step. Hence the orbital coefficients of the matrix $\mathbf{C}_i^{\text{AO}}$ do not obey to the structure of the LCAOs indicated in Eq. (4).

Anyway it is possible to obtain again a matrix satisfying the required constraints, through an appropriate linear combination with the occupied orbitals of the other fragments, $C^{\mathrm{occ}}_{(i)}$, without energy change.

In order to get this result, the orbitals $\bar{\varphi}^i_\alpha$ defined by the $\mathbf{C}^{\mathrm{AO}}_i$ matrix, $\bar{\varphi}^i_\alpha = \sum_{\mu=1}^{M} \left[\mathbf{C}^{\mathrm{AO}}_i\right]_{\mu\alpha}\chi_\mu$, are subjected to the following transformation:

$$\overline{\varphi}^i_\alpha \Rightarrow \sum_{\mu=1}^{M}\left[\mathbf{C}^{\mathrm{AO}}_i\right]_{\mu\alpha}\chi_\mu + \sum_{\substack{j=1\\ j\neq i}}^{n_{\mathrm{f}}}\sum_{\beta}^{N_j}\lambda^\beta_j\varphi^j_\beta = \sum_{\mu=1}^{M}\left[\mathbf{C}^{\mathrm{AO}}_i\right]_{\mu\alpha}\chi_\mu + \sum_{\substack{j=1\\ j\neq i}}^{n_{\mathrm{f}}}\sum_{\beta}^{N_j}\lambda^\beta_j\sum_{\mu}^{M}c^j_{\mu\beta}\chi_\mu$$

$$= \sum_{\mu=1}^{M}\left(\left[\mathbf{C}^{\mathrm{AO}}_i\right]_{\mu\alpha} + \sum_{\substack{j=1\\ j\neq i}}^{n_{\mathrm{f}}}\sum_{\beta}^{N_j}\lambda^\beta_j c^j_{\mu\beta}\right)\chi_\mu,$$

where the coefficients λ^β_j are determined in such a way that

$$\left(\left[\mathbf{C}^{\mathrm{AO}}_i\right]_{\mu\alpha} + \sum_{\substack{j=1\\ j\neq i}}^{n_{\mathrm{f}}}\sum_{\beta}^{N_j}\lambda^\beta_j c^j_{\mu\beta}\right) = 0 \text{ if}$$

$\chi_\mu \notin$ ith fragment.

The number of coefficients λ^β_j to be determined is equal to the number of occupied ELMOs of the other fragments, while there is a different equation for each atomic function not belonging to the ith fragment. So these equations represent an overdetermined system of linear equations, which can be solved by means of the singular value decomposition technique [26]. It should be noted that the solution of this system has an exact solution, as the "contamination" is due to only the components of the occupied orbitals $C^{\mathrm{occ}}_{(i)}$. At the end of this procedure, the updated orbitals of the ith fragment satisfy exactly to the constraints of Eq. (4).

The number of orbitals obtained for the ith fragment is obviously equal to the dimension of the $\mathbf{F}^{\mathrm{MO}}_{\mathrm{KS}}(i)$ matrix, m_i'.

If $m_i' \leq m_i$, it is necessary to restore the original dimensionality of the space for the ith fragment. This can be performed by constructing its orthogonal complement, using either the atomic basis functions which define the fragment or the starting (i.e., contained in the $\mathbf{C}^{(0)}$ matrix) complete set of occupied and virtual orbitals of the same fragment. The optimization procedure is then cyclically repeated for all the n_{f} fragments until convergence is reached, for which 10–20 iterations are usually sufficient. At convergence the optimized occupied orbitals of all the fragments are orthogonalized, and the final electron density and energy are computed.

Of course the energy associated with the ELMO wavefunction is higher than that obtained with a traditional DFT calculation where the orbitals are completely delocalized, as the number of variational coefficients is largely reduced. The advantage offered by the proposed procedure relies on the possibility to have a set of orbitals that can now be easily transferred from one molecule to another. We also note that if one or more fragments must be kept fixed, as required in some QM/MM methods [6, 7, 8, 9], this procedure can be still adopted by simply excluding them by the optimization loop over the whole set of fragments.

In order to perform the test calculations described in this paper, we used the Gaussian 98 package [27] to obtain the Kohn–Sham matrix on the AO basis, giving as an input the appropriate density matrix.

Test calculations

The benzene and cyclobutadiene molecules

In order to test the convergence capabilities of the present approach, we first applied it to the benzene and cyclobutadiene molecules, which from our experience we already know to be critical systems [17].

Using the geometries optimized at the RHF level with the standard 6-31G basis set, we performed for both molecules two ELMO calculations differing in the choice of the molecular fragments.

In the first one we described the molecules using two molecular fragments: the first fragment, describing all the σ electrons, which is defined using only the σ atomic functions, and the second one, describing the π electrons, which is defined using only the π atomic functions. Obviously in this case the ELMO wavefunction reproduces a traditional DFT calculation.

In the second localization scheme we adopted the same definition for the σ molecular frame, while the π electrons were described using the π atomic functions of only two adjacent carbon atoms for each pair of electrons. For the benzene molecule, we realized a Kekulé structure in this way.

The program was always able to obtain convergence and the energy values are reported in Table 1. From the values reported, we can have an estimate of the reso-

Table 1. Energy values (atomic units). *ELMO* represents extremely localized molecular orbital

Benzene	
ELMO (delocalized π bonds)	−232.083118
ELMO (localized π bonds)	−231.945228
Cyclobutadiene	
ELMO (delocalized π bonds)	−154.563370
ELMO (localized π bonds)	−154.558942
Thymopentin	
ELMO-Transf	−2,512.453636
ELMO-Rel(1)	−2,512.781853
ELMO	−2,512.836704
Density functional theory	−2,513.015353

nance energy, which appears to be 86.5 kcal/mol for the benzene molecule, to be compared with just 2.8 kcal/mol for the cyclobutadiene molecule.

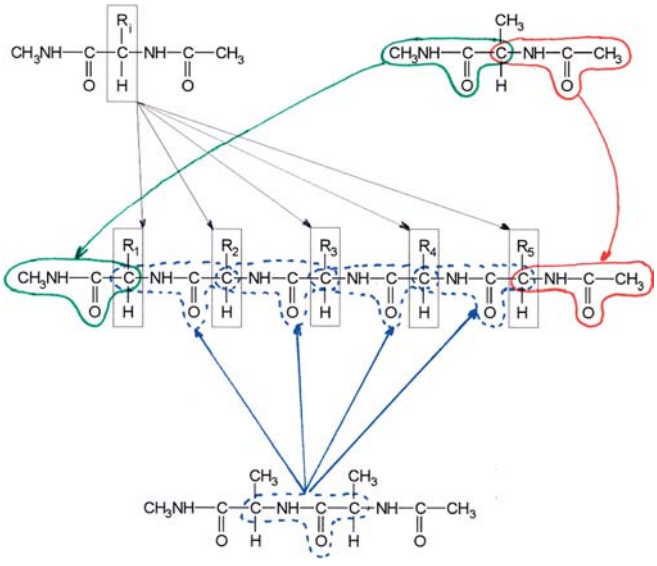

Fig. 1. The transferring scheme of the extremely localized molecular orbitals (*ELMOs*) for the thymopentin polypeptide

The thymopentin polypeptide

In order to investigate the transferability properties of the ELMOs, we looked at the possibility to build up an approximate DFT description of a little polypeptide using molecular fragments determined on the constituent amino acids. This approach closely resembles the LEGO method proposed by Walker and Mezey [28, 29], which builds up the electron density of a polypeptide by superimposing the electron densities computed for some molecular fragments.

As a test molecule we considered the thymopentin polypeptide, TP-5, which is constituted by the five residues of the active site of the thymopoietin hormone, isolated from thymus [30]. TP-5 presents immunoregulatory effects in animals as in humans, and clinical studies have also evidenced that TP-5 can restore the immunological responsiveness of patient affected by different diseases [31].

TP-5 is constituted by the following sequence of amino acids: Arg–Lys–Asp–Val–Tyr. Using the geometry the residues adopt in thymopoietin, as recovered from its crystal structure (Protein Data Bank entry 1H9E) [32], we performed a DFT calculation with the B-LYP functional [33, 34] and the standard 6-31G basis set, which will be considered as our reference. The molecular fragments to be used for building up the wavefunction of the target molecule were determined by selecting appropriate model molecules (Fig. 1). For all the fragments the corresponding Lewis structures were used to determine the number of doubly occupied ELMOs.

The amino acid alanine with its C- and N-termini protected by CH_3NH and $C(O)CH_3$ groups, respectively, was chosen to describe the molecular fragments which constitute the C- and N-termini of TP-5. In particular we performed an ELMO calculation using molecular fragments defined by the atomic functions of $CH_3NHC(O)C(\alpha)$, $C(\alpha)NHC(O)CH_3$ and $CH_3C(\alpha)H$ units. The ELMOs obtained for the first two units were then transferred to the TP-5 target molecule.

For each of the five amino acids (R_i = Arg, Lys, Asp, Val, Tyr) we carried out an ELMO calculation on the molecule $CH_3NHC(O)C(\alpha)HR_iNHC(O)CH_3$, which was partitioned into the following molecular fragments: $C(\alpha)HR_i$,$CH_3NHC(O)C(\alpha)$ and $C(\alpha)NHC(O)CH_3$. The first unit was then transferred for each of the five different amino acids to the TP-5 molecule. Finally, in order to describe the molecular fragment of the peptidic region, we performed an ELMO calculation on the dialanine dipeptide, $CH_3NHC(O)$-$C(\alpha)(H)(CH_3)NHC(O)C(\alpha)(H)(CH_3)NHC(O)CH_3$, using the following molecular fragments: $C(\alpha)NHC(O)C(\alpha)$, $CH_3NHC(O)C(\alpha)(H)(CH_3)$, $C(\alpha)(H)(CH_3)NHC(O)CH_3$, where the first one was transferred into the different regions of the TP-5 molecule to describe the corresponding four peptidic regions.

All the calculations on the fragments were carried out using the geometry that they adopt in the target polypeptide, in order to transfer the ELMOs without variations. In this way we assembled a set of ELMOs for the TP-5 molecule starting from smaller fragments, and the corresponding wavefunction will be denoted as ELMO-Transf.

The extremely localized nature of the ELMOs prevents, of course, a direct "charge transfer" between different fragments. Anyway the different residues are not electrically neutral as the ELMOs are not orthogonal and so a partial charge transfer is allowed through ELMOs of the nearest fragments. In order to allow a greater electronic rearrangement, the transferred ELMOs can be relaxed, as described later.

The wavefunction ELMO-Transf was compared with a conventional DFT calculation on the whole molecule.

The energy difference between the two wavefunctions (Table 1) amounts to 352.5 kcal/mol. This remarkable difference has already been discussed [21, 22] and it is, of course, due to the different number of the coefficients which define the two wavefunctions. For the present calculation, the DFT wavefunction is defined by 113,866 coefficients (variationally determined) while the ELMO-Transf wavefunction is set up by only 31,435 coefficients (determined on the molecular fragments).

Despite this large difference in the energy values, the agreement between the electron densities is considerably more favourable, as was evidenced by performing a Stone analysis [35] on the two wavefunctions. The standard deviations of the charges and of the higher moments of the electron density with respect to the DFT results are reported in Table 2. The comparison is quite satisfactory considering that the ELMOs were determined on molecular fragments and simply transferred without any change.

An obvious way to obtain better accuracy is to proceed in a relaxation of the transferred ELMOs by performing, for example, just one or more optimization cycles on the target molecule. The results obtained after just one optimization step, ELMO-Rel(1), and when the convergence on the target molecule is obtained (ELMO) are reported in Tables 1 and 2. It is encouraging to see that just one optimization cycle permits us to recover 86% of the energy with respect to the ELMO wavefunction with completely optimized molecular fragments. Also the accuracy of the charges and the higher moments is greatly increased.

The differences in the ELMO-Rel(1) charges with respect to the full DFT calculation are plotted in Fig. 2 using a colour scale. An inspection of the figure indicates, as expected, that the larger differences are principally located in the regions where the fragments overlap, i.e., where a greater reorganization of the electronic structure should arise. Anyway the overall charge distribution within the molecule has been largely reproduced.

This example clearly indicates that an approach based on an ELMO strategy can be a valuable tool in order to assemble the electron density of a large molecule using smaller molecular models, in the spirit of the LEGO approach [28, 29].

Table 2. Standard deviations of the atomic moments (atomic units) of the electron density according to the Stone analysis, with respect to those obtained with the density functional theory calculation

	ELMO-Transf	ELMO-Rel(1)	ELMO
Charges	0.111	0.074	0.070
Dipole *x*	0.106	0.049	0.059
Dipole *y*	0.100	0.038	0.039
Dipole *z*	0.107	0.042	0.043
Quadrupole *xx*	0.256	0.152	0.148
Quadrupole *yy*	0.268	0.150	0.130
Quadrupole *zz*	0.399	0.146	0.142
Quadrupole *xy*	0.111	0.063	0.050
Quadrupole *xz*	0.132	0.069	0.047
Quadrupole *yz*	0.094	0.037	0.034

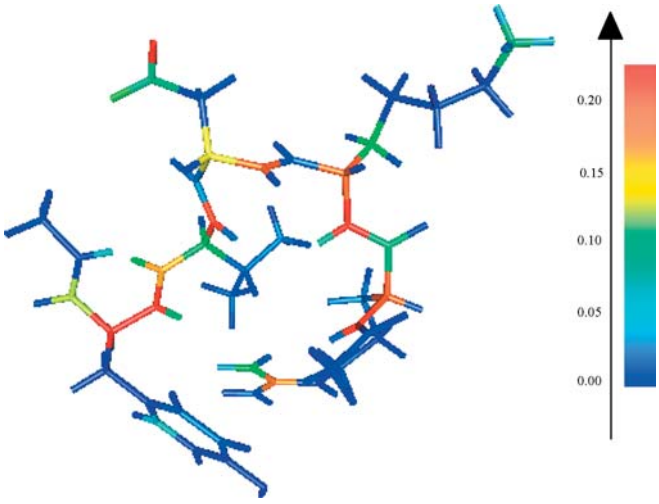

Fig. 2. Differences, plotted on a colour scale, in the charges between the wavefunction ELMO-Rel(1) (see text) and the traditional density functional theory calculation

Conclusions

We have presented an approach to determine ELMOs within the framework of the DFT method based on the solution of the Kohn–Sham equations.

The ELMOs are orbitals which are characterized by a priori elimination of the tails, so improving their transferability. The procedure devised optimizes the molecular fragments cyclically; hence, it is straightforward to keep one or more fragments frozen at their initial guess. The procedure could be used in the framework of the QM/MM methods based on localized orbitals, close to the LSCF strategy suggested by the Rivail's group [5, 6, 7]. This will be the subject of a future paper.

The approach developed was used to assemble a reasonable guess for the electron density of thymopentin, a polypeptide constituted by five amino acids, by transferring the ELMOs obtained on the constituent amino acids. The number of coefficients which define the ELMOs localized on the different molecular fragments of the system reported is less than one third with respect to the coefficients of the DFT calculation, and of course this ratio will decrease on increasing the dimension of the target molecule. Despite this, the electron density is quite well reproduced just using the transferred ELMOs, even if a partial relaxation appears to be important to permit mutual interactions. This relaxation was realized by a single iterative optimization step for each molecular fragment, as just one cycle was sufficient to recover a great percentage of energy with respect to an ELMO calculation on the target molecule with variationally determined molecular fragments. We are also studying the possibility to introduce relaxation using a different strategy [22], which takes into proper account the

localized nature of the orbitals, as already reported in previous studies [36, 37].

Acknowledgements. We would like to thank Dr. Arianna Fornili for helpful discussions.

References

1. Parr RG, Yang W (1989) Density functional theory of atoms and molecules. Oxford University Press, New York
2. Gao J (1996) In: Lipkowitz KB, Boyd DB (eds) Reviews in computational chemistry, vol 7. Wiley-VCH, Weinheim, chap 3
3. Monard G, Merz K (1999) J Acc Chem Res 32:904
4. Lyne P, Hodoscek M, Karplus M (1999) J Phys Chem A 103:3462
5. Thery V, Rinaldi D, Rivail J-L, Maigret B, Ferenczy GG (1994) J Comput Chem 15:269
6. Assfeld X, Rivail J-L (1996) Chem Phys Lett 263:100
7. Ferré N, Assfeld X, Rivail J-L (2002) J Comput Chem 23:610
8. Murphy RB, Philip DM, Friesner RA (2000) J Comput Chem 21:1442
9. Murphy RB, Philip DM, Friesner RA (2000) Chem Phys Lett 321:113
10. Boys SF (1960) Rev Mod Phys 32:296
11. Edmiston C, Ruedenberg K (1965) J Chem Phys 43:S97
12. Pipek J, Mezey PG (1989) J Chem Phys 90:4916
13. Reuter N, Loos M, Monard G, Cartier A, Maigret B, Rivail J-L (1997) Mol Eng 7:349
14. Naray-Szabo G (2000) Comput Chem 24:287
15. Smits GF, Altona C (1985) 67:461
16. Stoll H, Wagenblast G, Preuss H (1980) Theor Chim Acta 57:169
17. Sironi M, Famulari A (1999) Theor Chem Acc 103:417
18. Fornili A, Sironi M, Raimondi M (2003) J Mol Struct (THEOCHEM) 632:157
19. Szekeres Z, Surjan PR (2003) Chem Phys Lett 369:125
20. Couty M, Bayse CA, Hall MB (1997) Theor Chem Acc 97:96
21. Sironi M, Famulari A, Raimondi M, Chiesa S (2000) J Mol Struct (THEOCHEM) 529:47
22. Genoni A, Sironi M (2004) Theor Chem Acc
23. Gianinetti E, Raimondi M, Tornaghi E (1996) Int J Quantum Chem 60:157
24. Bagus PS, Hermann K, Bauschlicher CW (1984) 80:4378
25. Stevens WJ, Fink WH (1987) Chem Phys Lett 139:15
26. Press WH, Flannery BP, Teukolsky S, Vetterling WT (1990) Numerical recipes. Cambridge University Press, Cambridge
27. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA Jr, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA (1998) Gaussian 98, revision A.9. Gaussian, Pittsburgh, PA
28. Walker PD, Mezey PG (1993) J Am Chem Soc 115:12423
29. Walker PD, Mezey PG (1994) J Am Chem Soc 116:12022
30. Szewczuk Z, Siemion IZ, Wieczorek Z (1996) Mol Immunol 33: 903
31. Bodini G, Luisa A, Solfrini R, Marino M, Quadri A (1991) Curr Therap Res Clin Exp 50:189
32. Laguri C, Gilquin B, Wolff N, Romi-Lebrun R, Courchay K, Callebaut I, Worman HJ, Zinn-Justine S (2001) Struct 9:503
33. Becke AD (1988) Phys Rev A 38:3098
34. Lee C, Yang W, Parr RG (1988) Phys Rev B 37:785
35. Stone AJ (1981) Chem Phys Lett 83:233
36. Diner S, Malrieu JP, Claverie P (1969) Theor Chim Acta 13:1
37. Daudey JP, Diner S (1972) Int J Quantum Chem 6:575