



# Extremely localized molecular orbitals: theory and applications

Maurizio Sironi, Alessandro Genoni, Monica Civera, Stefano Pieraccini,  
Michela Ghitti

## ► To cite this version:

Maurizio Sironi, Alessandro Genoni, Monica Civera, Stefano Pieraccini, Michela Ghitti. Extremely localized molecular orbitals: theory and applications. Theoretical Chemistry Accounts: Theory, Computation, and Modeling, 2007, 117 (5-6), pp.685-698. 10.1007/s00214-006-0200-7 . hal-02196453

**HAL Id: hal-02196453**

**<https://hal.univ-lorraine.fr/hal-02196453>**

Submitted on 27 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Extremely localized molecular orbitals: theory and applications.

MAURIZIO SIRONI<sup>(A,B)\*</sup> ALESSANDRO GENONI<sup>(A)</sup> MONICA CIVERA<sup>(A,B)</sup> STEFANO  
PIERACCINI<sup>(A)</sup> MICHELA GHITTI<sup>(A)</sup>

*(A) Dipartimento di Chimica Fisica ed Elettrochimica – Università degli Studi di Milano, Via Golgi 19, 20133, Milano, Italy*

*(B) Centre for Bio-molecular Interdisciplinary Studies and Industrial Application, CISI, Via Fratelli Cervi 93, Palazzo LITA, 20090, Segrate (MI), Italy*

E-mail: [maurizio.sironi@unimi.it](mailto:maurizio.sironi@unimi.it)

Tel.: +390250314251

## Abstract

Orbitals which are extremely localized on molecular fragments represent a powerful tool for a number of purposes: just to cite a few examples, they allow to strongly reduce the complexity of calculations on large systems and they are easily transferable from one molecule to another one, providing a suitable and efficient way to build up the electronic structure of large molecules. Recently, we have developed efficient algorithms to determine Extremely Localized Molecular Orbitals (ELMOs) that will be reviewed in this paper.

As a rigorous localization is strictly connected to a reduction in the number of variational parameters, which reflects into an increased value of the associated energy with respect to the Hartree Fock value, we have developed a number of strategies to relax the wavefunction built up using transferred localized orbitals.

The extreme localization has also been exploited in connection with the ‘Divide & Conquer’ technique to determine large polypeptides electron densities assembled from orbitals computed on small model molecules.

Moreover, we will discuss the recent application of the ELMOs in the framework of the hybrid QM/MM methods to describe the frontier region.

We will also show that the ELMOs can be used to extract chemical interpretations from numerical results.

A variety of applications will be presented.

*Keywords: Extremely Localized Molecular Orbitals, transferability, VB approaches, Divide and Conquer, hyperconjugation.*

## Introduction

The two main challenges that computational chemists have to tackle are the quantitative prediction of molecular properties and the interpretation of the results.

A proper ‘insight’ of molecular electronic structure was offered by the Valence Bond methods (VB). Although they received a great attention in the early stages of quantum chemistry, the VB approaches were improved only by a small community of scientists [1,2]. On the contrary, the MO approach, which has the great advantage to efficiently compute molecular electronic properties in a cheaper way, has permitted the development of a large number of quantum chemistry packages, well known not only by the theoretical chemists, but also by almost all the chemical community. Nevertheless, such accuracy and easy of computation lead to the disappearance of some chemical concepts: for example, the wide delocalization of molecular orbitals is contrary to the chemist’s local vision. We learn chemistry using the concept of functional groups, we rationalize reactions drawing arrows which indicate the movements of electrons to break some bonds and to create new ones. We feel that a local vision of the molecules, like the one offered by the VB methods, would be more suitable to get a proper ‘understanding’ of the molecules.

Of course, these problems were immediately recognised by the MO community and a first solution of the problem was the possibility to transform the canonical delocalized Molecular Orbitals into a set of Localized Molecular Orbitals (LMOs) according to different localization strategies [3,4,5]. Even if the LMOs permit to recover to a great extent the chemical concepts of bonds, lone pairs, etc., they are not rigorously localized on a few atoms, but still preserve some tails on the whole molecule. The presence of such tails prevents a straightforward definition of chemical entities (i.e. functional groups) that could be transferred from one molecule to another one. Of course, it is possible to delete the LMOs tails to get transferable orbitals, but it is evident that this represents an arbitrary operation.

Moreover, it has been clearly shown that the small coefficients associated with the tails have a non negligible effect on the electronic energy.

A different choice is the use of molecular orbitals which are *a priori* localized only on a pre-selected set of atoms and which are variationally determined under this constraint. Methods based on this idea permit to obtain Extremely Localized Molecular Orbitals (ELMOs). Their use has been proposed by different research groups using different acronyms, such as ‘Strictly Localized Molecular Orbitals (SLMOs) [6], ‘Non-Orthogonal Localized Molecular Orbitals’ (NOLMOs) [7], and these approaches can be considered an extension of the group function method introduced by McWeeny [8] to decompose a wavefunction into functions describing subsets of electrons.

The ELMOs are directly transferable due to the complete absence of tails and can be properly defined in order to describe core electrons, lone pairs, functional groups, etc.. Of course their definition, i.e. their localization scheme, is arbitrary, but they are the best orbitals consistent with the chosen scheme and no further perturbation, like the tail deletion, is necessary to make them transferable from one molecule to another one. They could be considered as a tentative to introduce VB ideas in the framework of the MO theory. Like the VB orbitals, the ELMOs are not orthogonal and this introduces a mathematical complexity in the theory, which reflects into non trivial convergence problems in the algorithms devised to compute them. Of course, the ELMOs, unlike the VB orbitals, are doubly occupied and this permits to strongly bound the mathematical complexity with respect to the VB theory.

In this paper we will review the main results obtained by our group in the determination of the ELMOs and we will outline some of their applications.

Due to the absence of an available program for the ELMOs determination and due to the well known convergence difficulties [7,9], we have investigated different approaches to compute Extremely Localized Molecular Orbitals. In particular, we have used both a generalized pseudo Hartree-Fock approach [10] and a direct minimization of the energy based on the theoretical formulation proposed by Stoll [9,11]. The method has been recently extended to the density functional theory [12].

The ELMOs transferability has been previously analyzed on small systems [13] and the results obtained have encouraged us to devise a new method for the

construction of wavefunctions for polypeptides, using ELMOs determined on model molecules that suitably represent the side chains of all the aminoacids. In particular, a program to determine *ab initio* quality electron densities of polypeptides, DENPOL, has been written [14].

Even if ELMOs are not orthogonal, their extremely localized nature implies a fast decrease in the overlap values as the distance between the different molecular fragments increases. This important feature allowed to combine the “Divide and Conquer” (D&C) technique, a linear scaling strategy originally developed by Yang *et al.* in the framework of the Density Functional Theory [15,16], with the ELMO approach to efficiently compute the electron densities even for very large polypeptides [17].

Recently, the ELMOs transferability has also been exploited in the framework of the QM/MM methodology. As known, when the QM and the MM regions are connected by a bond, a suitable description of the frontier region is not trivial. In the LSCF approach proposed by Assfeld *et al.* [18,19,20] a frozen localized orbital describes the bond connecting the two regions. We have shown that the transferability of the ELMOs can be suitably used to describe this bond [21].

The extremely localized nature of the ELMOs is based on a reduction in the number of variational coefficients used to describe them, implying an increase in the energy with respect to the Hartree-Fock value. Anyway, just a few SCF iterations are usually sufficient to recover a large percentage of the energy difference [13]. Recently, we have also investigated the possibility to increase the accuracy of the ELMO wavefunction by means of a non orthogonal configuration interaction including excitations from the occupied ELMOs to a selected set of virtual ELMOs defining an ELMO-VB wavefunction [22,23]. Owing to the extreme localization of the virtual ELMOs, it is possible to greatly reduce the number of possible excitations introducing a criterion based on the proximity of the orbitals involved. The proposed method can be considered as a natural development of the Perturbative Configuration Interaction using Localized Orbitals (PCILO) [24,25] and of the studies carried out to reduce the CI expansion using LMOs [26].

The ELMO description also allows to gain a chemical interpretation of results that are otherwise of simple numerical nature. As example, we will show how this approach can be used to partition a torsional energy barrier into different chemical

contributions, following the ideas suggested by Mo et al. [27]. The results obtained at a correlated level, using the ELMOs determined in the framework of the Density Functional Theory, will be presented.

## Theory

### A. The determination of the ELMOs.

Let us consider a  $2N$  electrons closed shell molecule. We subdivide the molecule into  $n_f$  molecular fragments (lone pairs, bonds, functional groups, etc). Each of these subsystems is characterized by its own basis set,  $(\chi_\mu^i)_{\mu=1}^{m_i}$ ,  $m_i$  being the number of the atomic orbitals belonging to the  $i$ -th fragment.

The  $i$ -th fragment will be described by  $N_i^o$  ELMOs, whose number will depend on the chemical nature of the fragment itself (of course  $\sum_i^{n_f} N_i^o = N$ ). The ELMO wavefunction will be written as:

$$\Psi_{ELMO} = A(\Phi) \quad \text{with } \Phi = \prod_i^{n_f} \prod_\alpha^{N_i^o} \varphi_\alpha^i \alpha \varphi_\alpha^i \beta$$

where  $A$  is the antisymmetrizer operator and the  $\alpha$ -th ELMO of the  $i$ -th fragment will be described using only the atomic functions which belong to the same fragment:

$$\varphi_\alpha^i = \sum_{\mu=1}^{m_i} c_{\mu\alpha}^i \chi_\mu^i$$

In this way each ELMO is rigorously localized on the corresponding fragment. The coefficients  $c_{\mu\alpha}^i$  are determined through a minimization of the energy associated with the ELMO wavefunction.

The definition of the molecular fragments is arbitrary and it is established according to the application. A typical localization scheme can be obtained looking at the Lewis formula of the molecule and defining one ELMO for each bond or core/lone pair. Of course, molecular fragments can share atomic fragments. For example, if we adopt a ‘‘Lewis localization scheme’’, the atomic

functions centred on atoms involved in more than one bond, are common to different molecular fragments. This introduces a natural non orthogonality between the ELMOs of different fragments, which causes non trivial convergence problems. Different algorithms to determine ELMOs have been proposed both in the past and recently [9,10,11,28,29]. Here we briefly summarize two approaches on which our program [11] is based.

The first strategy is based on a generalization of the SCF-MI (Self Consistent Field for Molecular Interaction) theory developed by Gianinetti et al. [30]. In the SCF-MI approach, developed as a strategy to avoid *a priori* the Basis Set Superposition Error (BSSE), the wavefunction for a dimer is built up using two molecular fragments, one for each monomer. The difference between the ELMO and the SCF-MI wavefunction is that in the latter the fragments cannot share the same atomic orbitals.

It can be shown [10] that the optimal ELMOs must satisfy the following set of coupled equations:

$$\hat{F}_i \varphi_\alpha^i = \varepsilon_\alpha^i \varphi_\alpha^i \quad \left\{ \begin{array}{l} i = 1, \dots, n_f \\ \alpha = 1, \dots, m_i \end{array} \right\} \quad (1)$$

where  $\hat{F}_i$  is a proper hermitian operator for the  $i$ -th fragment built up using the occupied ELMOs of all the  $n_f$  molecular fragments. The exact expression of the operator [10] can be easily obtained generalizing the equations proposed by Gianinetti et al. [30]. It should be evidenced that each molecular fragment has its own set of occupied and virtual orbitals with the same localization scheme. Like in the SCF-MI method, orbitals belonging to different fragments are not orthogonal, while orbitals localized on the same fragment constitute an orthogonal set. We also note that as the fragments can share atomic orbitals, the total number of occupied and virtual ELMOs of all the fragments can overcome the basis-set dimension, giving rise to some ‘singularity’ problems which have been properly discussed [10] and which seriously afflict the convergence procedure in some cases.

When converging, the algorithm based on the resolution of Eq. (1) is quite rapid, but some difficulties in the convergence or some instabilities sometimes arise.

In these cases our program uses a second strategy based on a direct minimization of the energy functional, using a quasi-Newton procedure. The expressions of the

first and second derivatives of the energy with respect to the variational coefficients were obtained by Stoll using the dual orbitals formalism [9,11]. In the quasi-Newton procedure, an approximate analytic Hessian is evaluated at the first iteration and then it is updated with a variable metric algorithm in the following iterations. Also in this case, the extremely localized nature of the orbitals can be exploited to avoid the computation of the full Hessian matrix, which reduces to only diagonal blocks.

The ELMO program has been interfaced with the PC-GAMESS [31] version of the GAMESS\_US package [32] and with the GAMESS-UK package [33]. Both single point calculations (conventional and direct) and geometry optimizations can be carried out. Furthermore, the ELMO program has been recently extended to the DFT method [12].

#### **B. The relaxation of the ELMO wavefunction.**

The strict localization of the ELMOs is necessarily associated with an increase in the energy with respect to the RHF value, due to the reduction in the number of the variational coefficients. On the other hand, the ELMOs can be easily transferred from one molecule to another one in a straightforward way, without resorting to any tail deletion.

Once the wavefunction of the target molecule has been obtained by means of transferred ELMOs, it could be very useful to introduce a relaxation of the wavefunction in order to increase the accuracy of the results. We have shown that just carrying out one or two SCF cycles on the ELMO wavefunction can largely improve the results [13].

We have also investigated a different strategy based on the use of virtual ELMOs. Due to the Brillouin theorem, the energy of the ELMO wavefunction cannot be improved by a mixing of the occupied ELMOs of a fragment with its own virtual orbitals. Hence, a relaxation of the ELMO wavefunction can only be obtained by a mixing of the occupied ELMOs of a fragment with the virtual ELMOs of another one. In this way we introduce a delocalization of the orbitals which allows to improve the accuracy yielded by the ELMO description. It could be argued that the most significant contributions arise from permitting the ELMOs to delocalize towards the nearest fragments and that a complete delocalization on the whole molecule is not very important. This view is supported by the



‘nearsightedness of the electronic structure’ [34,35], a principle that can be considered as the physical basis of successful linear scaling strategies, such as the “Divide and Conquer” approaches [15,16,36,37].

Therefore, we have introduced the ELMO-VB method, which consists in improving the ELMO wavefunction by means of a configuration interaction calculation, using a list of single excitations from an occupied ELMO to the virtual ELMOs of the fragments located in its proximity [22].

Indicating with  $\Psi_{\alpha\mu}^{ir}$  the configuration characterised by the excitation  $\varphi_{\alpha}^i \rightarrow \varphi_{\mu}^r$  we can write:

$$\Psi_{ELMO-VB} = c_0 \Psi_{ELMO} + \sum_{ir}^{n_f} \sum_{\alpha}^{N_r^o} \sum_{\mu}^{N_r^v} c_{\alpha\mu}^{ir} \Psi_{\alpha\mu}^{ir} \quad (2)$$

where  $N_r^v = m_r - N_r^o$  is the number of virtual ELMOs of the  $r$ -th fragment.

The computation of the Hamiltonian matrix elements  $\langle \Psi_{\alpha\mu}^{ir} | \hat{H} | \Psi_{\beta\nu}^{js} \rangle$  should require the use of the Löwdin formula [38] due to the non orthogonality of the ELMOs belonging to different fragments. This fact, which would have reduced the dimensions of the systems to study, has prompted us to introduce the following transformation:

1. The occupied ELMOs are Löwdin orthogonalized in order to obtain a new set of equivalent occupied ELMOs. The Löwdin procedure was chosen in order to reduce the orbitals deformation; this permits to preserve the ELMOs localization scheme as much as possible.
2. Each virtual orbital is Gram-Schmidt orthogonalized with respect to the new set of occupied orbitals obtained at step 1.

After these transformations, the new occupied and virtual orbitals, which, from now on, will be denoted with the symbol  $\phi_{\alpha}^i$  instead of  $\varphi_{\alpha}^i$ , are characterized by the following properties:

- the occupied orbitals are now a set of orthogonal orbitals which give rise to the same energy with respect to the original non orthogonal set
- each virtual orbital is orthogonal to all the occupied orbitals belonging to any fragment;
- the virtual orbitals constitute a non orthogonal set

Let us now consider the following pair of excitations:  $\Psi_{\alpha\mu}^{ir}$  (i.e.  $\phi_\alpha^i \rightarrow \phi_\mu^r$ ) and  $\Psi_{\beta\nu}^{js}$  (i.e.  $\phi_\beta^j \rightarrow \phi_\nu^s$ ); these can be expressed in the following way:

$$\Psi_{\alpha\mu}^{ir} = A \left[ \Phi \begin{pmatrix} i j \\ \alpha \beta \end{pmatrix} \phi_\beta^j \alpha \phi_\beta^j \beta \phi_\alpha^i \phi_\mu^r \left( \frac{\alpha\beta - \beta\alpha}{\sqrt{2}} \right) \right]$$

$$\Psi_{\beta\nu}^{js} = A \left[ \Phi \begin{pmatrix} i j \\ \alpha \beta \end{pmatrix} \phi_\alpha^i \alpha \phi_\alpha^i \beta \phi_\beta^j \phi_\nu^s \left( \frac{\alpha\beta - \beta\alpha}{\sqrt{2}} \right) \right]$$

where  $\Phi \begin{pmatrix} i j \\ \alpha \beta \end{pmatrix}$  is the product of all the doubly occupied orbitals with the exception of the spin-orbital pairs  $\phi_\alpha^i \alpha$  and  $\phi_\beta^j \beta$ , namely

$$\Phi = \Phi \begin{pmatrix} i j \\ \alpha \beta \end{pmatrix} \phi_\alpha^i \alpha \phi_\alpha^i \beta \phi_\beta^j \alpha \phi_\beta^j \beta.$$

Hence, the couple of excitations  $\Psi_{\alpha\mu}^{ir}$  and  $\Psi_{\beta\nu}^{js}$  is characterized by a set of common  $N-4$  spin-orbitals,  $\Phi \begin{pmatrix} i j \\ \alpha \beta \end{pmatrix}$ , orthogonal to each other, and by 4 spin-orbitals whose spatial parts are not orthogonal between them but are orthogonal to the product  $\Phi \begin{pmatrix} i j \\ \alpha \beta \end{pmatrix}$ . These considerations have allowed us to derive an expression for the Hamiltonian matrix elements  $\langle \Psi_{\alpha\mu}^{ir} | \hat{H} | \Psi_{\beta\nu}^{js} \rangle$  which is much less computationally demanding than applying the Löwdin formula to the original set of orbitals. In fact, using the new set of ELMOs, the non orthogonal problem is substantially limited to just 4 orbitals. Moreover, we have shown that it is possible to efficiently evaluate the matrix elements through a ‘difference strategy’, consisting in computing some common quantities from which the elements of the Hamiltonian matrix can be evaluated by simple subtraction of some small terms [22].

As mentioned above, the total number of occupied and virtual orbitals usually exceeds the basis set dimension. This can give rise to a critical situation in the solution of the secular problem to obtain the expansion coefficients which appear in Eq. (2). In particular, we have found that it is quite common to find virtual orbitals which are linear combination of the occupied orbitals of the other fragments. In this case the transformation adopted at the step 2 eliminates the linear dependent virtual orbitals.

The number of excitations to be included in Eq. (2) can be greatly reduced taking into account the extreme localization of the orbitals. In fact, preliminary calculations have shown that the excitations from an occupied to a virtual orbital localized on a far fragment give a very small contribution, so they can be neglected. On the contrary, excitations from occupied to virtual orbitals belonging to nearby fragments are always significant. Despite this great saving in computational cost, there is necessarily a rapid growth in the number of excitations to be included in expansion (2), as the molecular size increases. For this reason we have considered the possibility to define a small set of optimized virtual ELMOs [23] that allow for the construction of very compact ELMO-VB wavefunctions and that take implicitly into account a very large amount of usual virtual ELMOs.

In particular, for each occupied ELMO  $\varphi_\alpha^i$ , we define a corresponding optimized virtual orbital  $\varphi_\alpha^{i(opt)}$  as the orbital which minimizes the energy of the following wavefunction:

$$\Psi_\alpha^i = c_0 \Psi_{ELMO} + \Psi_{\alpha\alpha}^{i(opt)} \quad (3)$$

where  $\Psi_{\alpha\alpha}^{i(opt)}$  represents the single excitation  $\varphi_\alpha^i \rightarrow \varphi_\alpha^{i(opt)}$ .

The optimized virtual orbital  $\varphi_\alpha^{i(opt)}$  is expressed as a linear combination of all the possible virtual orbitals belonging to a suitable number  $f^i$  of fragments in the neighbourhood of the  $i$ -th fragment:

$$\varphi_\alpha^{i(opt)} = \sum_r^{f^i} \sum_\mu^{N_r^v} c_{\alpha\mu}^{ir} \varphi_\mu^r \quad (4)$$

The coefficients  $c_{\alpha\mu}^{ir}$  can be computed introducing Eq.(4) into Eq. (3) and solving the corresponding secular problem.

Therefore, in this approach we have to solve only  $N$  small secular problems to obtain the set of optimized virtual orbitals that can be then used in expansion (2). As already mentioned, in this way the number of the single excitations is greatly reduced. It should be stressed out again that, due to Eq.(4), the use of optimized virtual orbitals implicitly takes into account a larger number of virtual ELMOs with respect to the usual ELMO-VB calculations.

### C. The transferability of the ELMOs to large systems.

Due to the absence of tails, the ELMOs can be used to build the wavefunction of large molecules by assembling the ELMOs determined on smaller fragments. The partition of the molecule is arbitrarily defined: fragments can range from single bonds up to functional groups or larger units, embedded in a proper neighbourhood. This is the basis of our DENPOL program [14] which is aimed to build up *ab initio* quality electron densities for polypeptides, using the ELMOs determined on model molecules. These molecules are constituted by the 20 aminoacids suitably capped at their C and N termini in different geometrical conformations. The ELMOs determined on these model molecules are properly rotated and transferred to the target polypeptide.

After having performed the ELMOs calculations on the model molecules, the transferring step is trivial. However, for large molecules as polypeptides, the electron density computation is not trivial, due to the large number of the involved ELMOs and to their non orthogonality. In order to realize an efficient algorithm, we have devised the “ELMO-D&C” strategy [17] that exploits the extremely localized nature of the ELMOs taking into account the ‘Divide & Conquer’ methodology.

To discuss this new technique, it is now more convenient to consider the list of the unique atomic functions  $(\chi_\mu)_{\mu=1}^m$  and to express each ELMO as  $\varphi_\alpha^i = \sum_\mu c_{\mu\alpha} \chi_\mu^i$ ,

where some coefficients  $c_{\mu\alpha}$  will be zero according to the localization pattern.

The electron density corresponding to the ELMO wavefunction is given by:

$$\rho(\mathbf{r}) = \sum_{\mu, \nu=1}^{NAO} \chi_\mu(\mathbf{r}) \mathbf{D}_{\mu\nu} \chi_\nu(\mathbf{r})$$

where:

$$\mathbf{D}_{\mu\nu} = 2 \sum_{i,j=1}^N c_{\mu j} [\mathbf{S}^{-1}]_{ji} c_{\nu i}^* \quad (5)$$

with  $\mathbf{S}$  as the overlap matrix between the ELMOs.

To greatly reduce the computational cost associated with Eq. (5), we divide the molecule into  $N_s$  disjoint subunits so that their union provides the whole system. Owing to this fragmentation, we automatically assign a local basis set  $\beta_k$  to the generic  $k$ -th subsystem ( $k = 1, 2, \dots, N_s$ ). Furthermore, following the D&C strategy, we introduce the partition matrix  $\mathbf{P}^k$  for the  $k$ -th subunit:

$$\mathbf{P}_{\mu\nu}^k = \begin{cases} 1 & \text{if } \mu \in \beta_k \text{ and } \nu \in \beta_k \\ 1/2 & \text{if } (\mu \in \beta_k \text{ and } \nu \notin \beta_k) \text{ or } (\mu \notin \beta_k \text{ and } \nu \in \beta_k) \\ 0 & \text{if } \mu \notin \beta_k \text{ and } \nu \notin \beta_k \end{cases}$$

which satisfies the normalization condition

$$\sum_{k=1}^{N_s} \mathbf{P}_{\mu\nu}^k = 1 \quad \forall \mu, \nu$$

and allows to assemble the density matrix as a sum of contributions arising from the different subsystems

$$\mathbf{D}_{\mu\nu} = \sum_{k=1}^{N_s} \mathbf{D}_{\mu\nu}^k \quad \text{with} \quad \mathbf{D}_{\mu\nu}^k = \mathbf{P}_{\mu\nu}^k \mathbf{D}_{\mu\nu}$$

The ‘‘Divide Step’’ allows to introduce a proper approximation in the evaluation of the  $\mathbf{D}_{\mu\nu}^k$  contributions, taking advantage of the extremely localized nature of the ELMOs. In particular, the elements  $\mathbf{D}_{\mu\nu}^k$  can be efficiently computed without the complete set of the  $N$  ELMOs, but considering only the orbitals localized on the  $k$ -th subsystem (‘‘Core ELMOs’’) and the ones that significantly overlap with them (‘‘Buffer ELMOs’’), i.e. orbitals generally belonging to near subunits in the real space. Therefore, for the  $k$ -th subunit:

$$\mathbf{D}_{\mu\nu}^{k \text{ (approx)}} = 2 \mathbf{P}_{\mu\nu}^k \sum_{i,j=1}^{n_k} C_{\mu j} [\mathbf{S}_k^{-1}]_{ji} C_{\nu i}^* \quad (6)$$

with  $n_k$  as the total number considered ELMOs (i.e., the core and buffer ELMOs) for the  $k$ -th subunit and  $\mathbf{S}_k$  as their overlap matrix. It is important to note that  $n_k$  is generally much lower than  $N$ , so the inversion of  $N_s$  matrices  $\mathbf{S}^k$  is clearly less demanding than a single inversion of the complete,  $N$  dimension and sparse matrix  $\mathbf{S}$ . In this way, we realize the ‘‘Conquer Step’’ to efficiently compute the electron density for large molecules, using transferred localized orbitals previously determined on model molecules.

## Applications

### 1. The nature of the ELMO wavefunction

In order to illustrate some features of the ELMOs and to compare them with the LMOs, we present the results obtained on the butadiene molecule in two possible configurations, that is, the planar structure (A) and the one obtained through a 90 degree rotation about the central C-C bond (B), which is characterized by a loss of resonance. The geometries have been obtained at the RHF level using the 6-311G\*\* basis set. The canonical MOs have been localized using the Pipek-Mezey scheme [5] and subjected to tail deletion in order to compare the resulting energy with the one associated with the ELMO wavefunction. The ELMOs have been determined using a ‘Lewis localization scheme’, i.e. each ( $\sigma$  or  $\pi$ ) bond/core is described by a single ELMO defined using only the atomic functions of the corresponding atoms/atom.

Then, the ELMO wavefunction has been relaxed performing just two SCF iterations, using the ELMOs as initial guess. All the energy values are reported in Table 1. The plot of all the ELMOs describing  $\sigma$  bonds are shown superimposed in Fig. 1, while one of the two equivalent ELMOs describing the C-C  $\pi$  bond is shown in Fig. 2 along with the corresponding LMO.

From Figures 1 and 2 it is evident the high degree of localization of the ELMOs and the presence of tails in the LMOs. Even if the magnitude of these tails is very small, their deletion causes a great increase in energy, amounting to 66.6 and 63.4 kcal/mol for the configurations A and B, respectively (see Table 1), with respect to the RHF value. The ELMO wavefunction permits to recover a lot of energy and the energy differences drop to 30.0 and 26.4 kcal/mol, respectively. Moreover, just 1 or 2 SCF cycles using the ELMOs as starting guess are enough to reduce the energy difference by an order of magnitude.

In other papers [11,13] we have already shown that the same results hold also for transferred ELMOs. In particular, we have shown that assembling the ELMO wavefunction for the acetone molecule using ELMOs determined on ethane to describe the C-H and C-C bonds and on formaldehyde to describe the C-O bond, allows to accurately reproduce the charge distribution and other electrostatic properties. Also in this case just one SCF iteration was enough to obtain a wavefunction characterized by an energy of just few kcal/mol above the RHF value [11].

These observations indicate that the ELMOs determined on model molecules can be really considered as good building blocks for assembling the electronic

structure of a target molecule. To further support this idea, we have shown that the electronic charge distributions and the electrostatic potential of some ortho-substituted biphenyl molecules evaluated at the RHF level, can be accurately reproduced using an ELMO wavefunction where only the ELMOs which describe the substituent are optimized, while keeping frozen all the remaining ELMOs, which were previously determined on the unsubstituted biphenyl molecule [13].

In the last example it is evident that the strictly localized nature associated with each ELMO allows to selectively optimize only a part of the molecule. In a similar way, Hierse et. al [39] have shown that using a local representation of the orbitals at the DFT level, it is possible to obtain reliable torsional energy profiles optimizing only the orbitals localized near the bond around which the rotation is performed.

## **2. The relaxation of the ELMOs wavefunction**

As already mentioned, the large difference in energy obtained by the ELMO wavefunction with respect to the RHF value is due to the large reduction in the number of variational coefficients that allows their strict localization. We have shown that using the virtual ELMOs and taking into account the extreme localization of the orbitals, it is possible to perform non orthogonal configuration interaction calculations which permit to recover a large percentage of the energy difference. In addition, a proper definition of optimized virtual orbitals permits to largely reduce the length of the expansion.

At first, we illustrate the use of traditional virtual ELMOs on two small molecules as example: methoxymethylfluoride and butane. All calculations reported in this section have been performed with geometries optimized at the RHF level (the basis set is indicated in the corresponding figure captions).

The ELMOs computations have been carried out using the previously described “Lewis localization scheme”. For each calculation we have also determined, for comparison, the energy associated with the wavefunction built up with LMOs subjected to tail deletion (LMOs-TD). Then, we have considered the expansions that, from now on, we will denote as ELMO-VB(1) and ELMO-VB(2), where one and two virtual ELMOs for each occupied orbital have been selected, respectively. In particular, these orbitals have been chosen according to their lowest orbital energies. In Table 2 we have reported the energy differences with respect to the RHF value, the percentage of the recovered energy with respect to the ELMO

wavefunction and the number of excitations. Also in this case, it is evident the large energetic discrepancy between the ELMO wavefunction and the wavefunction assembled with LMOs-TD. The comparison between the ELMO-VB(1) and ELMO-VB(2) values shows that more than one virtual orbital is necessary to get an evident improvement in the accuracy. This is unfortunately associated with a large increase in the computational requirements. For this reason we have focused our attention on the different contributions arising from the excitations and we have been able to show [22] that only the excitations from an occupied ELMO to the virtual ones localized on the nearest fragments are important.

For this reason, we have topologically classified the fragments as first/second/third-neighbour fragments if they are separated by 0/1/2 bond(s). In Figure 3 we show some examples for the methoxymethylfluoride molecule.

The excitations to be considered are those which satisfy the following topological criteria:

1. excitations from an occupied ELMO describing a bond or a lone pair to a virtual ELMO localized on the first or second neighbour fragment;
2. excitations from an occupied ELMO describing a bond or a lone pair to a virtual ELMO localized on a third neighbour fragment which describes an atomic fragment.

No excitations from the occupied ELMOs which describe the core are considered. Using this criterion, we can reduce the excitations considered in the ELMO-VB(1) and ELMO-VB(2) wavefunctions, obtaining respectively the ELMO-VB(Top\_1) and ELMO-VB(Top\_2) expansions. The results reported in Table 2 clearly show that the topological criterion permits to largely reduce the length of the expansions, still preserving the accuracy of the results. It is also evident the great importance of using at least two virtual orbitals to get satisfactory results.

To further reduce the computational cost while preserving the numerical accuracy, we have developed a method based on the determination of optimized virtual orbitals [23]. These orbitals have been defined in such a way to be particularly suitable to enter in the ELMO-VB expansions; as they are expressed as a linear combination of several virtual orbitals, they intrinsically take into account the effect of many virtual orbitals. Each optimal virtual orbital corresponding to an occupied ELMO is expanded using all the possible virtual orbitals localized on the



first-neighbour fragments. In this way they acquire a little delocalization character but they still remain enough localized to apply the topological criteria reported above. The number of optimal virtual orbitals is equal to the number of occupied orbitals, so the computational cost is equivalent to the ELMO-VB(1) calculation because the time required to determine the optimized virtual orbitals is negligible. The method has been tested on the alanine and serine molecules. The energy results are reported in Table 3 together with the values obtained using the canonical virtual orbitals. It is evident the great improvement offered by the optimal virtual orbitals, which permit to obtain results comparable to the use of a larger number of usual virtual ELMOs. Once again the topological criterion have shown to be a powerful tool to limit the number of excitations to be included. The optimal virtual orbitals have been shown to be useful also in the transferring process. We have assembled the alanine-serine dipeptide ELMO wavefunction using the ELMOs determined on the alanine, serine and formamide molecules. The last one was chosen to model the peptide bond between the alanine and serine. The use of optimized virtual orbitals to relax the wavefunction of the dipeptide has allowed to significantly improve the results; in particular, the electron density distribution and the electrostatic potential were shown to match very well with those obtained at the RHF level.

### **3. 'Chemical insight' from the ELMO wavefunction**

The chemical interpretation of molecular properties, which is crucial to get a proper knowledge of the electronic structure, is naturally recovered by the localized nature of the ELMOs. As an illustrative example, we will show how it is possible to evaluate the different energetic contributions to the torsional barrier of the C-N bond in the  $\text{HCX}(\text{NH}_2)$  ( $\text{X}=\text{O}$ ,  $\text{NH}$ ,  $\text{CH}_2$ ) compounds by performing ELMO calculations at the DFT level.

In fact, although the canonical Hartree-Fock or Kohn-Sham orbitals calculations allow us to obtain a numerical value for the rotational barrier, they are not able to distinguish among the different factors contributing to the overall barrier (delocalization, hyperconjugation, etc.). A chemical interpretation of the electronic properties of formamide and its analogues has been suggested both at the VB by Lauvergnat and Hyberty [40] and at the RHF level by Mo *et al.* [27] which have devised a block-localised wave function method that closely resembles to the ELMO approach. In order to introduce the electronic correlation,

we have applied the DFT-ELMO algorithm following the protocol proposed by Mo *et al.* in the study of the different effects on the rotational energy barrier about the amide bonds of the HCX(NH<sub>2</sub>) (X=O, NH, CH<sub>2</sub>) compounds.

Here we only summarize how the application of the Mo ‘energetic decomposition scheme’ allow us to evaluate the different energetic contributions. A detailed discussion can be found in [27]. All the calculations have been carried out using the BLYP functional [41,42] and the 6-31G\* basis set. The results are presented in Table 4.

First of all, let us consider the molecule in the planar ground state. A traditional DFT calculation, where the orbitals are completely delocalized, obviously includes the resonance between the  $\pi$  C-X electrons and the N lone pair (structure A, see Fig. 4). We can evaluate the resonance energy ( $\Delta E_R$ ) by performing an ELMOs calculation with the  $\pi$  C-X electrons and the N lone pair localized on the C-X bond and nitrogen atom, respectively (structure B). The ELMO which describes the  $\pi$  C-X electrons uses only the  $\pi$ -symmetry atomic orbitals of the C and X atoms and the ELMO of the N lone pair uses the atomic orbital of the nitrogen atom orthogonal to the molecular plane. All the other electrons are described by a proper number of ELMOs which use all the atomic functions. The energy difference between the structures A and B can be ascribed to the resonance term,  $\Delta E_R$ . Through a 90 degree rotation about the C-N bond we obtain the localized structure C. The new increase in the energy can be associated with the steric contribution ( $\Delta E_\sigma$ ). In structure C we can evidence the presence of two hyperconjugative effects. The first one is an interaction between the empty  $\pi_{C=X}^*$  orbital and the antisymmetric combination of the two  $\sigma_{NH}$  orbitals (i.e.  $\pi_{NH_2} \rightarrow \pi_{C=X}^*$ ). The second one is the negative hyperconjugation effect of the nitrogen lone pair and the antibonding C-X bond (i.e.  $n_N \rightarrow \sigma_{CX}^*$ ). By delocalizing all the orbitals on the whole atomic set and keeping the amino group perpendicular to the carbonyl (structure D), we evaluate the total stabilizing hyperconjugation effect ( $\Delta E_H$ ) for the 90 degree rotamer. The separated hyperconjugation contributions can be obtained by two additional ELMO calculations whose energies must be compared with the energy of structure D. The  $\pi_{NH_2} \rightarrow \pi_{C=X}^*$  effect can be evaluated by localizing only the  $\pi$  electrons of the

C=X bond, while the  $n_N \rightarrow \sigma_{CX}^*$  contribution can be obtained by localizing only the N lone pair (see Table 4). As expected, the sum of the two hyperconjugation energy terms ( $\Delta E_H[\text{sum}]$ ), is similar to the total hyperconjugation effect ( $\Delta E_H$ ), suggesting that the two types of hyperconjugation effects can be additive.

Finally, permitting a complete geometry optimization of structure D, we can evaluate the pyramidalization contribution,  $\Delta E_p$  (structure E).

The inclusion of the electronic correlation has shown the same trends with respect to the change of the substituent X, as already observed by Mo et al. [27]. The greatest effect observed at the DFT level with respect to the Mo findings at the RHF level, is an increased value in the resonance contributions (ca. 4-5 kcal/mol) while smaller effects are observed for the steric and hyperconjugative effects. The pyramidalization contribution is practically unaffected by the DFT level. Therefore, there is globally a small increase in the computed torsional barriers (ca. 2-3 kcal/mole).

Furthermore, the promising results obtained by means of the localized orbital DFT-ELMO approach in the study of electronic delocalization phenomena have confirmed the robustness of our algorithm and suggest to extend our strategy for studying other molecules with interesting conjugation properties, such as vinylic systems, benzene and similar cyclic  $\pi$  resonance molecules.

#### 4. The transferability of the ELMOs

The absence of tails, which characterizes the ELMOs, makes them particularly suitable for transfer purposes from model molecules to other systems. In this way, it is possible to build up the electronic structure of large molecules, using a proper database of ELMOs which describe functional groups, aromatic rings and so on. The transfer is quite reliable, provided that the ELMOs are determined on model molecules which contain the same environment of the target molecule.

In our laboratory we have developed DENPOL [14], a program which assembles the electron density of polypeptides using the ELMOs determined on model molecules which are wisely defined to describe the peptide bond and all the different side chains of the 20 aminoacids. These chains are described by a set of ELMOs whose localization scheme is chosen in a proper way. More precisely, the localization pattern is chosen in such a way to describe the largest molecular fragments which can be described independently on the geometrical configuration

of the other fragments. Let us consider, for example, the tyrosine side chain: we will use a set of ELMOs localized on the aromatic ring and another one on the OH substituent. In this way the same set of ELMOs can be transferred, taking into account the different possible values of the torsion angle which connects the oxygen atom to the aromatic ring, just through a proper rotation of the ELMOs.

In order to create the suitable environment for each aminoacid, an opportune set of model molecules was selected. Some examples are reported in Fig. 5.

The possibility to well reproduce the electrostatic properties by means of transferred ELMOs is shown using the peptide Leu-enkephalin. Leu-enkephalin [43], an opioid peptide that has been found to modulate nociception in the spinal cord, is characterized by the sequence Tyr-Gly-Gly-Phe-Leu. Due to its small dimensions, it is possible to perform a RHF calculation for comparison purpose. The electrostatic potentials at the ELMO and RHF levels are plotted in Fig. 6. The results obtained with the ELMO wavefunction are very close to the RHF ones, but they are obtained at a very lower computational cost. The transfer of ELMOs is in fact a very fast step after that the building blocks, which can be used for any peptide, have been determined. Also the polypeptides electron densities can be easily reproduced. Indeed, using DENPOL we have determined the charge distributions for (Gly)<sub>n</sub> and (Leu)<sub>n</sub> in different geometrical arrangement. For the smallest polypeptides we have also carried out the corresponding RHF calculations in order to have a proper reference (see Table 5). To perform quantitative comparisons, we have used the similarity index  $L(a, a')$  introduced by Walker and Mezey [44], which compares point by point two charge distributions in the space bounded by two electron density isosurfaces characterized by the values  $a$  and  $a'$ .

Table 5 shows a very good agreement and points out that reliable electron densities can be computed using the ELMO wavefunction.

For larger polypeptides ( $n=50-100$ ) the electron density computation becomes non trivial, due to the large number of non orthogonal ELMOs involved. For example the (Leu)<sub>100</sub> molecule is characterized by 3105 ELMOs. To overcome this problem, we have proposed the ELMO-D&C method [17] that has been discussed in the Theory Section. The results reported in Table 6 clearly show that this technique can compute the electron density of large systems with a great saving of CPU time, preserving the accuracy of the results.

Recently we have also investigated the utility of the ELMOs in the framework of the QM/MM methods. It is well known that when the two regions (QM and MM) are connected by a covalent bond, the description of the frontier region is not trivial. One of the most popular approach to describe this region has been introduced by the Assfeld's group [18,19,20] and it is denoted as Local SCF (LSCF). It is based on the use of frozen molecular orbital to describe the frontier bond localized on the frontier atoms. In order to determine this two centre orbital, a traditional RHF calculation is carried out on a model molecule followed by an opportune localization procedure. The most localised orbital on the interested centres is subjected to a tail deletion in order to be used into the LSCF calculation. In this context it appears that the ELMOs could be the most suitable orbitals to be transferred. A comparative study on the use of ELMOs and orbitals arising from various localization procedures in the LSCF framework [21] has confirmed this prevision.

In order to obtain accurate results when performing geometry optimization in the framework of the LSCF approach, it was necessary to introduce a special five-parameters force field potential [19]. Recently, it was shown that introducing an additional orbital to describe the core electrons of the frontier MM atom permits to avoid the introduction of the force field term [45]. Also in this case the ELMOs have been one of the natural choices to describe the frontier core orbital.

## Conclusions

ELMOs are the best orbitals that satisfy a pre-selected localization scheme.

The extremely localized nature of the ELMOs is the origin of two important mathematical consequences. The first one is an increasing complexity in the mathematical formalism with respect to the canonical MOs, which unfortunately reflects into convergence difficulties. The coefficients which define the orbitals are more strongly coupled in the equations which define the ELMOs in comparison with the traditional orthogonal MOs. Hence, a great effort for the development of a robust algorithm to determine ELMOs has been necessary. Some problems are still present for the calculation at the DFT level, where more serious convergence problems have been experienced.

The second consequence is a very fruitful use of the techniques based on the ‘Divide&Conquer’ approach, which has permitted, for example, to study very large molecules such as polypeptides.

The most useful advantages in developing approaches based on localized orbitals rely on the possibility to determine them on small model molecules and to transfer them to larger ones. In this way the problem of determining the ELMOs is limited to small systems. Following this approach we have developed the DENPOL program which is able to assemble electronic densities of polypeptides which compare very well with those assembled at the RHF level. Also electrostatic properties are well reproduced.

Of course techniques which relax the transferred ELMOs could be very useful to improve the electronic structure description of the target molecule, permitting to the transferred ELMOs to adapt to the new environment. We have shown that an approach based on the use of virtual ELMOs can be very useful to this aim. At last, ELMOs have recently found useful application in the framework of the hybrid QM/MM methods. Their use in the LSCF approach has generally given better results with respect to the use of traditional tail deleted LMOs.

#### *Acknowledgements.*

We thank Dott. Giuseppe Lanzani for his invaluable technical assistance in the maintenance of our PC Linux Cluster.

## **References**

1. Shurki A (2006) Theor Chem Acc, published on line.
2. Shaik S, Hiberty PC (2004) Rev Comp Chem 20: 1.
3. Boys SF (1960) Rev Mod Phys 32: 296.
4. Edmiston C, Ruedenberg K (1965) J Chem Phys 43: S97.
5. Pipek J, Mezey PG (1989) J Chem Phys 90: 4916.
6. Náray-Szabó G (2000) Comput Chem 24: 287.
7. Smits GF, Altona C (1985) Theor Chim Acta 67: 461.

8. Mc Weeny R (1992) *Methods of Molecular Quantum Mechanics*. Academic, New York.
9. Stoll H, Wagenblast G, Preuss H (1980) *Theor Chim Acta* 57:169.
10. Sironi M, Famulari A (1999) *Theor Chem Acc* 103: 417.
11. Fornili A, Sironi M, Raimondi M (2003) *J Mol Struct (THEOCHEM)* 632: 157.
12. Burrelli E, Sironi M (2004) *Theor Chem Acc* 112: 247.
13. Sironi M, Famulari A, Raimondi M, Chiesa S (2000) *J Mol Struct (THEOCHEM)* 47: 529.
14. Ghitti M, Genoni A, Civera M, Fornili A, Pieraccini S, Sironi M (2005) in: *Proceedings of the International Karlsruhe Nanoscience Workshop*. University of Karlsruhe.
15. Yang W (1991) *Phys Rev Lett* 22:1438.
16. Yang W, Lee T-S (1995) *J Chem Phys* 103: 5674.
17. Genoni A, Ghitti M, Pieraccini S, Sironi M (2005) *Chem Phys Lett* 415: 256.
18. Assfeld X, Rivail J-L (1996) *Chem Phys Lett* 263: 100.
19. Ferré N, Assfeld X, Rivail J-L (2002) *J Comput Chem* 23: 610.
20. Ferré N, Assfeld X (2002) *J Chem Phys* 117: 4119.
21. Fornili A, Moreau Y, Sironi M, Assfeld X (2006) *J Comput Chem* 27:515.
22. Genoni A, Sironi M (2004) *Theor Chem Acc* 112: 254.
23. Genoni A, Fornili A, Sironi M (2005) *J Comput Chem* 26: 827.
24. Diner S, Malrieu JP, Claverie P (1969) *Theor Chim Acta* 13: 1.
25. Daudey JP, Diner S (1972) *Int J Quantum Chem* 6: 575.
26. Bonaccorsi R, Petrongolo C, Scrocco E, Tomasi J (1969) *Theor Chim Acta* 15: 332.
27. Mo Y, Schleyer PVR, Wu W, Lin M, Zhang Q, Gao J (2003) *J Phys Chem A* 107: 10011.
28. Szekeres Z, Surján PR (2003) *Chem Phys Lett* 369: 125.
29. Couty M, Bayse CA, Hall MB (1997) *Theor Chem Acc* 97: 96.
30. Gianinetti E, Raimondi M, Tornaghi E (1996) *Int J Quantum Chem* 60:157.
31. Granovsky AA, <http://classic.chem.msu.su/gran/gamess/index.html>.

32. Schmidt MW, Baldrige KK, Boatz JA, Elbert ST, Gordon MS, Jensen JJ, Koseki S, Matsunaga N, Nguyen KA, Su S, Windus TL, Dupuis M, Montgomery JA (1993) *J Comput Chem* 14: 1347.
33. Guest MF, Bush IJ, van Dam HJJ, Sherwood P, Thomas JMH, van Lente JH, Havenith RWA, Kendrick J (2005) *Mol Phys* 103: 719.
34. Kohn W (1996) *Phys Rev Lett* 76: 3168.
35. Prodan E, Kohn W (2005) *Proc Natl Acad Sci USA* 102: 11635.
36. Dixon SL, Merz KM Jr (1996) *J Chem Phys* 104: 6643.
37. Dixon SL, Merz KM Jr (1997) *J Chem Phys* 107: 879.
38. Löwdin PO (1955) *Phys Rev* 97: 1474.
39. Hierse W, Stechel EB (1996) *Phys Rev B* 54: 16515.
40. Lauvergnat D, Hiberty PC (1997) *J Am Chem Soc* 119: 9478.
41. Becke AD (1988) *Phys Rev A* 38: 3098.
42. Lee C, Yang W, Parr RG (1988) *Phys Rev B* 37: 785.
43. König W (1993) *Peptide and Protein Hormones: Structure, Regulation, Activity*. VCH, Weinheim.
44. Walker PD, Mezey PG (1994) *J Am Chem Soc* 116: 12022.
45. Fornili A, Loos P-F, Sironi M, Assfeld X (2006) *Chem Phys Lett*, submitted.



### Figure Captions:

Figure 1: Superposition of the isosurfaces of the  $\sigma$  ELMOs of the butadiene molecule. (The core orbitals of the carbon atoms are not reported.)

Figure 2: Comparison between the C-C  $\pi$  bond LMO (A, notice the presence of small tails) with the corresponding ELMO (B) in the butadiene molecule.

Figure 3: Examples of neighbour fragments in the methoxymethyl fluoride molecule: the lines connect first (A), second (B) and third (C) neighbour fragments.

Figure 4: Decomposition energy scheme for the rotation barrier about the C-N bond in the  $H CX(NH_2)$  ( $X=O, NH, CH_2$ )

Figure 5: Model molecules used for the determination of the ELMOs to be transferred: (A) N-terminal residue; (B) C-terminal residue; (C) lateral chains for the 20 aminoacids (one model molecule for each residue R) ; (D) peptide bond. The transferred ELMOs are defined on the moiety depicted by the ball & stick model.

Figure 6: RHF (A) and DENPOL (B) Molecular Electrostatic Potentials plotted on the 0.001 a.u. isosurface RHF electron density for the Leu-enkephalin molecule.

Table 1 – Energy values for the butadiene molecule (Basis-Set 6-311G\*\*) .

Method	Energy $_{180^\circ}$ (a.u.) <sup>(a)</sup>	Energy $_{90^\circ}$ (a.u.) <sup>(b)</sup>	$\Delta E_{180^\circ}$ (kcal/mol) <sup>(c)</sup>	$\Delta E_{90^\circ}$ (kcal/mol) <sup>(d)</sup>	$\Delta E_{\text{ROT}}$ (kcal/mol) <sup>(e)</sup>
RHF	-154.946961	-154.938995	0.0	0.0	5.0
LMO-TD	-154.840809	-154.838033	66.6	63.4	1.7
ELMO	-154.899116	-154.896871	30.0	26.4	1.4
ELMO-Relaxed (1) <sup>(f)</sup>	-154.935389	-154.929214	7.3	6.1	3.9
ELMO-Relaxed (2) <sup>(f)</sup>	-154.941188	-154.933604	3.6	3.4	4.8

(a) Energy values for the planar structure

(b) Energy values for the  $90^\circ$  structure

(c) Energy differences with respect to the RHF value for the planar conformation.

(d) Energy differences with respect to the RHF value for the  $90^\circ$  conformation

(e) Rotational barrier between the planar and the  $90^\circ$  structures at different levels of theory

(f) ELMO-Relaxed (#): # represents the number of SCF cycles performed, using the ELMOs as starting Molecular Orbitals

Table 2 – Energy values for the butane (Basis-Set 6-31G) and the methoxymethyl fluoride (Basis-Set 6-31G\*\*) molecules <sup>(a)</sup>

Method	Butane			Methoxymethyl Fluoride		
	NE <sup>(b,c)</sup>	$\Delta E$ (kcal/mol) <sup>(d)</sup>	% <sup>(c,e)</sup>	NE <sup>(b,c)</sup>	$\Delta E$ (kcal/mol) <sup>(d)</sup>	% <sup>(c,e)</sup>
LMO-TD		80.88			85.4	
ELMO		27.94			47.0	
ELMO-VB (1)	289	21.23	24	289	29.7	37
ELMO-VB (2)	578	4.40	84	578	19.2	59
ELMO-VB (Top_1)	161	22.37	20	176	31.3	33
ELMO-VB (Top_2)	322	6.08	78	352	20.0	57

(a) The acronyms are described in the text

(b) Number of excitations

(c) Only for ELMO-VB calculations

(d) Energy difference with respect to the corresponding RHF value

(e)  $\% (ELMO - VB) = 100 \cdot |\Delta E(ELMO - VB) - \Delta E(ELMO)| / |\Delta E(ELMO)|$

Table 3 – Energy values for the alanine and serine molecules (Basis-Set 6-31G\*\*) <sup>(a)</sup>.

Method	Alanine			Serine		
	NE <sup>(b,c)</sup>	$\Delta E$ (kcal/mol) <sup>(d)</sup>	% <sup>(c,e)</sup>	NE <sup>(b,c)</sup>	$\Delta E$ (kcal/mol) <sup>(d)</sup>	% <sup>(c,e)</sup>
LMO-TD		150.7			176.7	
ELMO		74.2			96.3	
ELMO-VB (1)	576	46.3	45	784	55.8	43
ELMO-VB (Top_1)	275	51.5	38	336	59.1	39
ELMO-VB* (1) <sup>(f)</sup>	576	19.6	77	784	25.0	74
ELMO-VB* (Top_1) <sup>(f)</sup>	275	20.8	75	336	26.5	72

(a) The acronyms are described in the text

(b) Number of excitations

(c) Only for ELMO-VB calculations

(d) Energy difference with respect to the corresponding RHF value

(e)  $\% (ELMO - VB) = 100 \cdot |\Delta E(ELMO - VB) - \Delta E(ELMO)| / |\Delta E(ELMO)|$

(f) The \* symbol indicates the use of optimized virtual ELMOs

Table 4 – Individual energy contributions for the  $HCX(NH_2)$  ( $X=O, NH, CH_2$ ) molecules at the DFT-ELMO/BLYP/6-31G\* level. All the values are in kcal/mol.

X	$\Delta E_R^{(a)}$	$\Delta E_\sigma^{(b)}$	$\Delta E_H^{(c)}$	$\Delta E_p^{(d)}$	$\Delta E_{rot}^{(e)}$	$\Delta E_H(n_N \rightarrow \sigma_{CX}^*)^{(f)}$	$\Delta E_H(\pi_{NH_2} \rightarrow \pi_{C=X}^*)^{(f)}$	$\Delta E_H[sum]^{(g)}$
O	29.88	9.63	-13.56	-7.71	18.24	-8.83	-6.42	-15.25
NH	25.56	7.54	-14.34	-7.03	11.73	-9.62	-6.54	-16.16
CH <sub>2</sub>	21.12	3.71	-13.36	-4.55	6.92	-8.75	-6.35	-15.10

(a) Resonance energy

(b) Steric energy

(c) Hyperconjugation energy

(d) Pyramidalization energy

(e) Rotational barrier

(f) Hyperconjugative effects

(g)  $\Delta E_H(n_N \rightarrow \sigma_{CX}^*) + \Delta E_H(\pi_{NH_2} \rightarrow \pi_{C=X}^*)$

Table 5 - Similarity index values between the Restricted Hartree-Fock electron density and the reference ELMO charge distribution for the Gly<sub>10</sub> and Leu<sub>10</sub> polypeptides (Basis-Set STO-4G).

Molecule	L ( 0.001, 10 ) %	L ( 0.1, 10 ) %	L ( 0.01, 0.1 ) %	L ( 0.001, 0.01 ) %
Gly <sub>10</sub> ( $\alpha$ -helix)	98.31	98.64	98.41	98.08
Gly <sub>10</sub> ( $\beta$ -sheet)	98.83	98.90	98.85	98.80

Table 6 – Gly<sub>100</sub>, Leu<sub>50</sub> and Leu<sub>100</sub> polypeptides: similarity index values and comparison of CPU times for the density matrix calculation when different approaches are used (Basis-Set STO-4G).

Molecule	L (0.001, 10) %	CPU time (s)	
		Approximation <sup>(a)</sup>	Nonorthogonal Orbitals <sup>(b)</sup>
Gly <sub>100</sub> ( $\alpha$ -helix)	99.89	74	295
Gly <sub>100</sub> ( $\beta$ -sheet)	99.93	68	299
Leu <sub>50</sub> ( $\alpha$ -helix)	99.87	102	429
Leu <sub>100</sub> ( $\alpha$ -helix)	99.86	563	6658

(a) Density matrix computed exploiting the ELMO-D&C strategy

(b) Density matrix calculated using the traditional formula for non orthogonal orbitals

Figure 1:

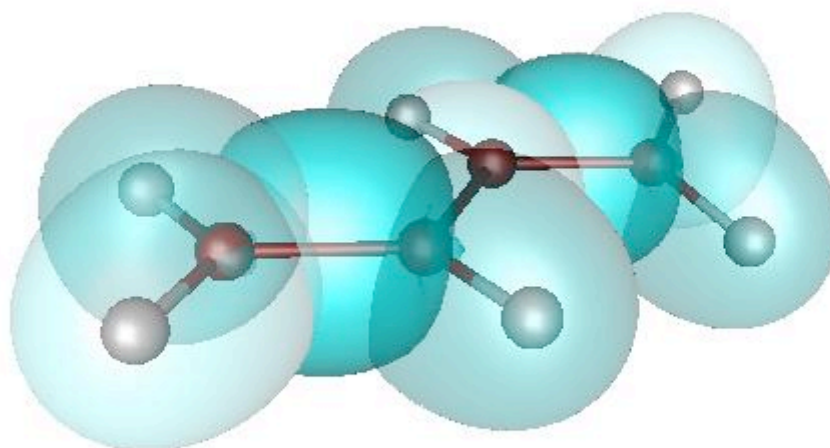




Figure 2:

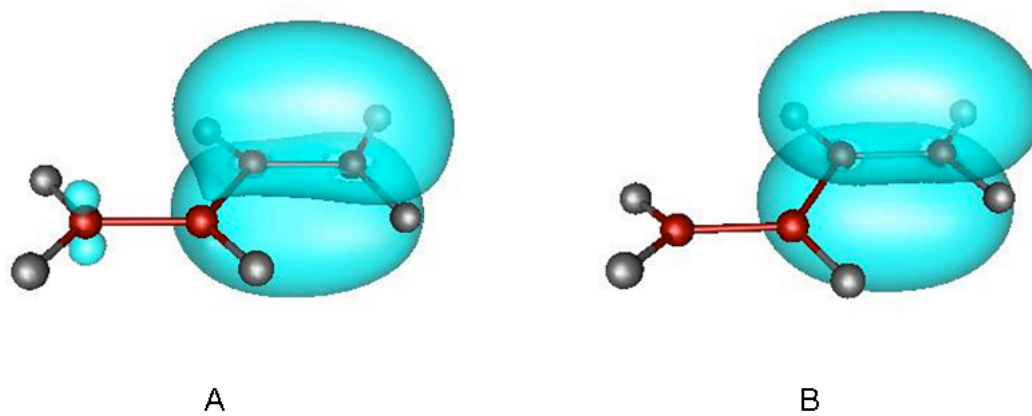


Figure 3:

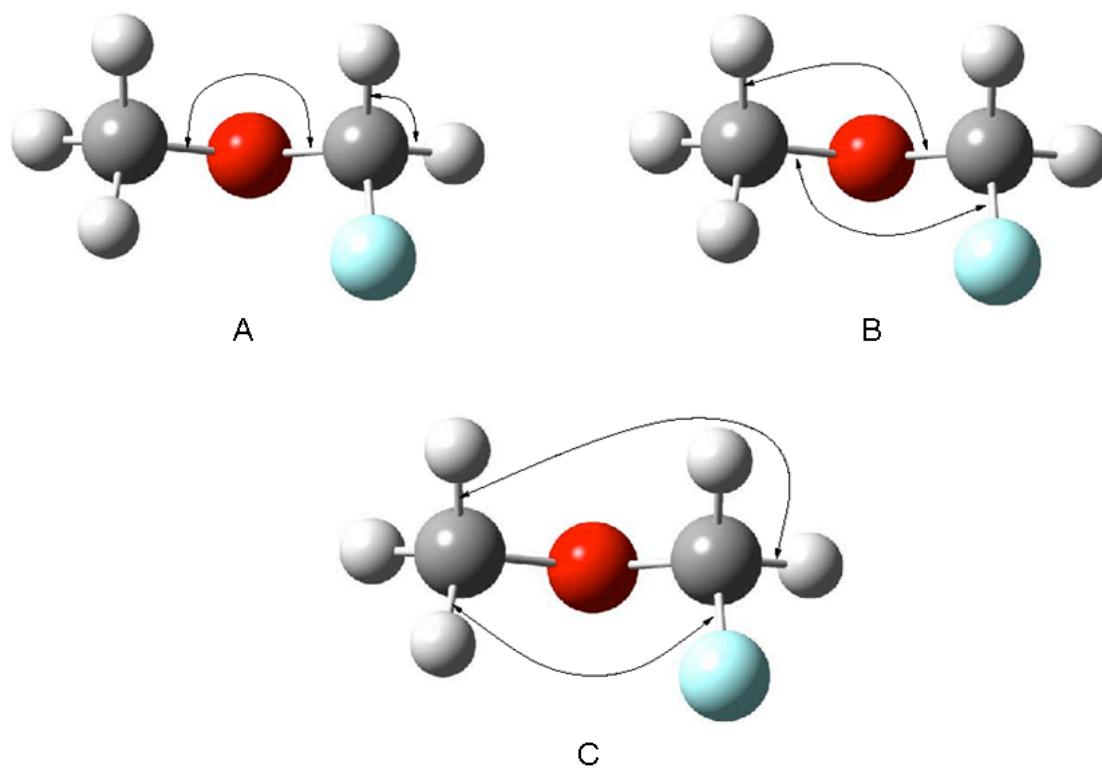


Figure 4:

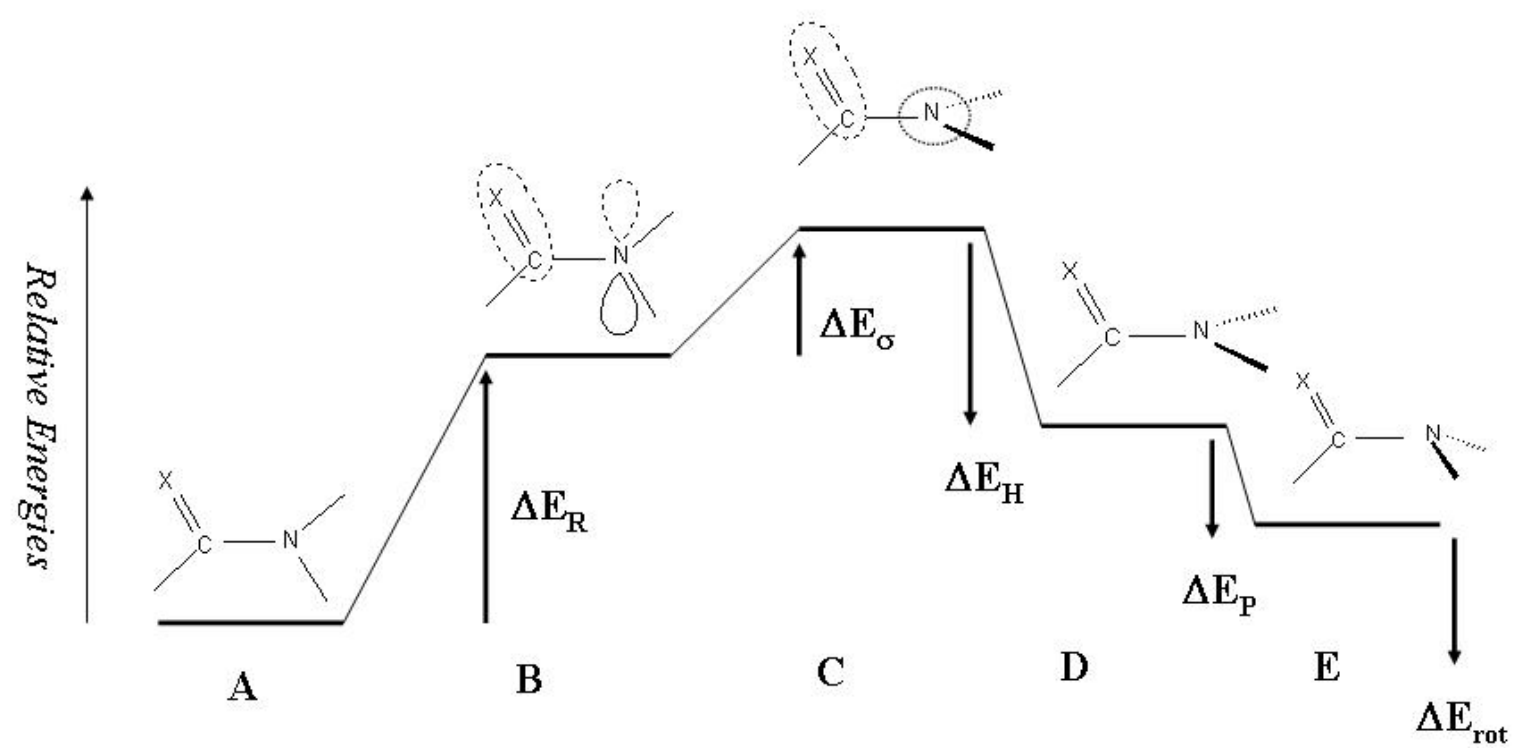


Figure 5:

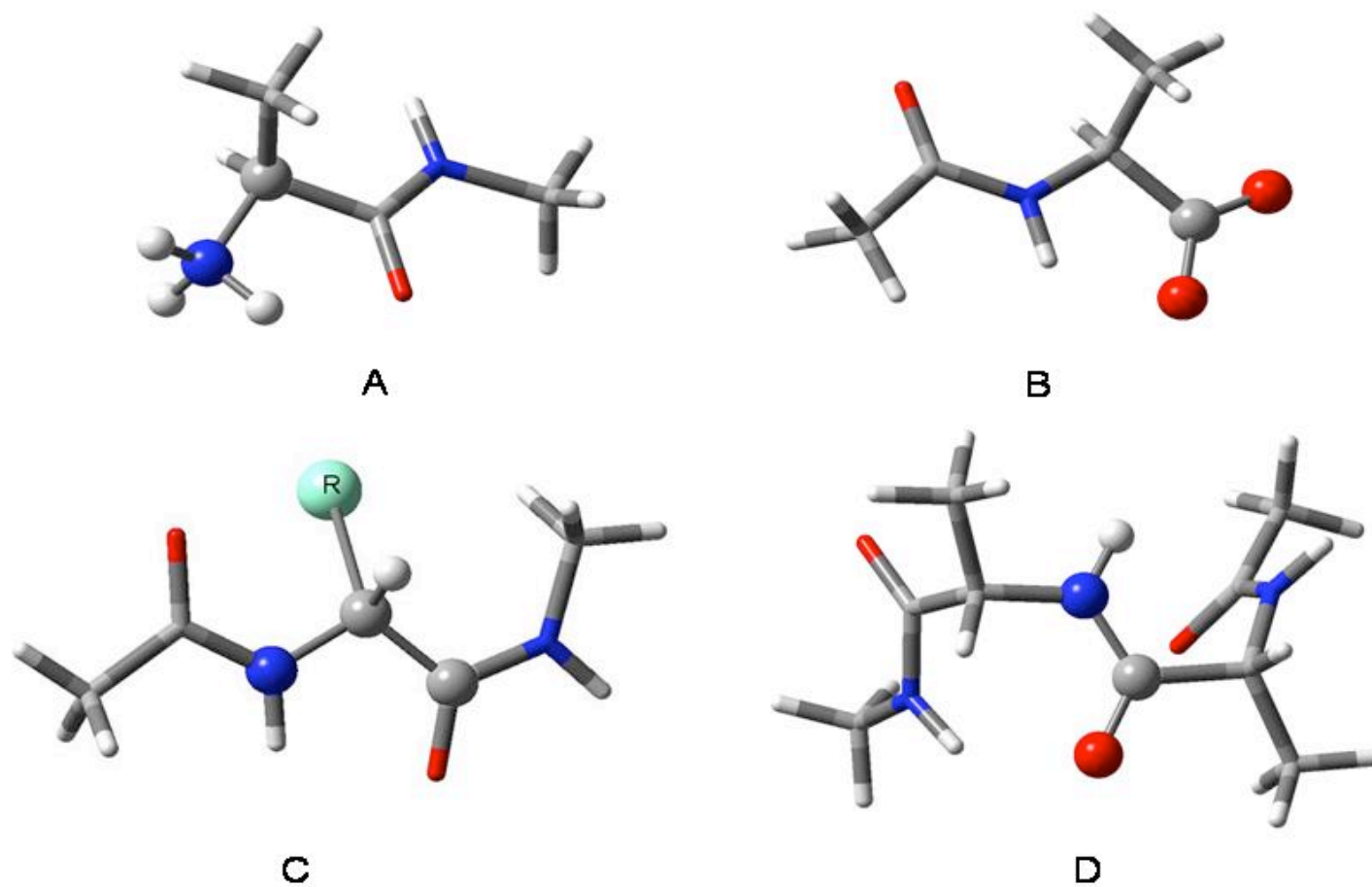
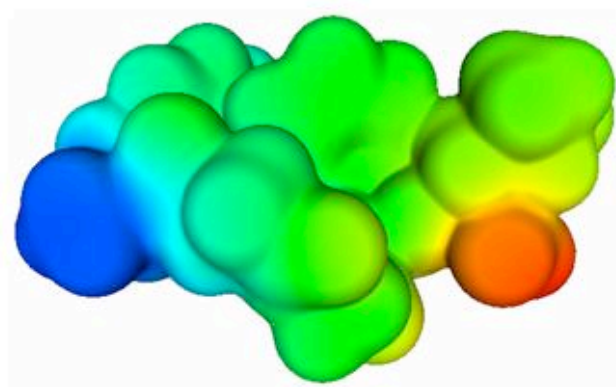
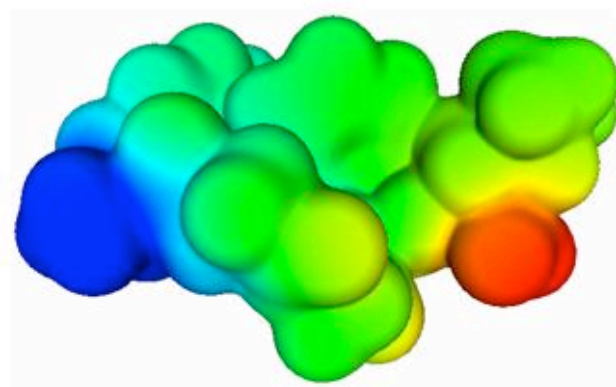


Figure 6:



A



B