Optimal virtual orbitals to relax wave functions built up with transferred extremely localized molecular orbitals

Alessandro Genoni, Arianna Fornili, Maurizio Sironi

Title: Optimal virtual orbitals to relax wavefunctions built up with transferred Extremely Localized Molecular Orbitals.

Authors: Alessandro Genoni (a), Arianna Fornili (a), Maurizio Sironi (a, b).

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

(b) Centre for Biomolecular Interdisciplinary Studies and Industrial Applications, CISI, Via Fratelli Cervi 93, Palazzo LITA, 20090 Segrate (MI), Italy.

Corresponding author: Maurizio Sironi. Dipartimento di Chimica Fisica ed Elettrochimica, Università degli Studi di Milano, Via Golgi 19, 20133 Milano, Italy.

Tel.: +390250314251; Fax: +390250314300; e-mail address: maurizio.sironi@unimi.it

Title: Optimal virtual orbitals to relax wavefunctions built up with transferred Extremely Localized Molecular Orbitals.

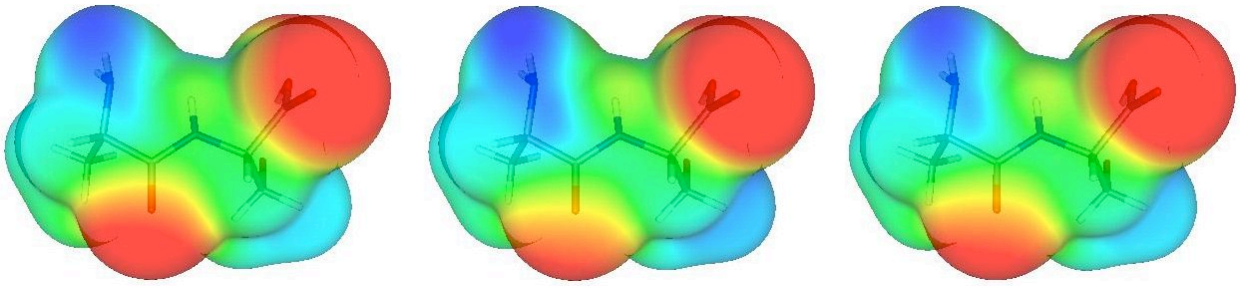
Authors: Alessandro Genoni (a), Arianna Fornili (a), Maurizio Sironi (a, b).

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

(b) Centre for Biomolecular Interdisciplinary Studies and Industrial Applications, CISI, Via Fratelli Cervi 93, Palazzo LITA, 20090 Segrate (MI), Italy.

Corresponding author: Maurizio Sironi. Dipartimento di Chimica Fisica ed Elettrochimica, Università degli Studi di Milano, Via Golgi 19, 20133 Milano, Italy.

Tel.: +390250314251; Fax: +390250314300; e-mail address: maurizio.sironi@unimi.it



Abstract

Extremely Localized Molecular Orbitals (ELMOs), namely orbitals strictly localized on molecular fragments, are easily transferable from one molecule to another one. Hence they provide a natural way to set up the electronic structure of large molecules using a data base of orbitals obtained from model molecules. However, this procedure obviously increases the energy with respect to a traditional MO calculation. In order to gain accuracy, it is important to introduce a partial electron delocalization. This can be carried out by defining proper optimal virtual orbitals that supply an efficient set for non orthogonal configurations to be employed in VB-like expansions.

Abstract

Extremely Localized Molecular Orbitals (ELMOs), namely orbitals strictly localized on molecular fragments, are easily transferable from one molecule to another one. Hence they provide a natural way to set up the electronic structure of large molecules using a data base of orbitals obtained from model molecules. However, this procedure obviously increases the energy with respect to a traditional MO calculation. In order to gain accuracy, it is important to introduce a partial electron delocalization. This can be carried out by defining proper optimal virtual orbitals that supply an efficient set for non orthogonal configurations to be employed in VB-like expansions.

Keywords: Extremely Localized Molecular Orbitals; Optimal virtual orbitals; Valence-Bond approach; Transferability; Molecular fragments.

Introduction

Today one of the most appealing targets for theoreticians is the development of strategies to allow them to study large molecules like those of biological and pharmaceutical importance.

This is why great efforts have recently been made to devise new methods where the computational cost scales linearly with the increasing size of the system [1, 2].

A possible approach in this direction is based on the recognition that chemistry relies on local concepts and that molecules can be set up with a proper collection of smaller entities. This idea has been widely used by various research groups in different contexts. Yang et al. developed the "Divide and Conquer" strategy where a molecule is divided into subsystems each of them characterized by a local Hamiltonian. Once the electronic problem has been solved for each region, the full electron density of the molecule is assembled. This strategy was initially

developed within the framework of Density Functional Theory [3, 4] and it has afterwards been extended to other methods [5-7]. Gadre et al. [8-10] devised the Molecular Tailoring Approach where the molecule is conveniently broken into smaller overlapping fragments of similar size. Calculations at the Hartree-Fock (or MP2) level are then carried out for each subunit in order to obtain the corresponding density matrices later "stitched" together to determine the density matrix for the parent molecule. In the Molecular Electron Density Lego Assembler (MEDLA) approach devised by Walker and Mezey [11], *ab initio* quality electron densities for large systems are built up with a considerable saving of CPU time using a database of electron density distributions of smaller molecular fragments. Mezey et al. [12-14] have extended further this strategy with a new scheme called Adjustable Density Matrix Assembler (ADMA) which requires the creation of a new database of fragment density matrices and which allows the computation of the electrostatic potential for polypeptides and other macromolecules.

Another approach relies on the possibility of defining molecular orbitals (MOs) localized on a few atoms of a molecule (e.g. a functional group), in order to have an entity which can be easily transferred to molecules containing the same group of atoms [15]. It is well known that canonical Hartree-Fock orbitals are not suitable for this purpose. In fact, they are delocalized over the whole system and, even after a localization procedure [16-18], they still preserve non-transferable tails beyond the localization region. It is also worthwhile to note that the deletion of these tails is associated with a dramatic increase in energy.

Different strategies have been adopted to compute orbitals rigorously localized on a few atoms such as Extremely Localized Molecular Orbitals (ELMOs) [19-21], Strictly Localized Molecular Orbitals (SLMOs) [22] or Non Orthogonal Localized Molecular Orbitals (NOLMOs) [23]. In these techniques the orbitals are defined as linear combinations of the only basis functions centred on the atoms belonging to preselected fragments, thus avoiding *a priori* the presence of

tails. Unfortunately, these orbitals are not orthogonal and this makes their determination a non-trivial matter.

In our laboratory we have devised a reliable algorithm to determine ELMOs in the framework of the Hartree-Fock approach [21, 24] and we have recently extended it to the Density Functional Theory [25]. The ELMOs transferability has been tested [24, 26] on a few systems obtaining better results with respect to those achieved using localized MOs submitted to tails deletion. Moreover, ELMOs have turned out to be suitable orbitals for describing the frontier zone in the mixed Quantum Mechanics / Molecular Mechanics methods [24] according to the Local Self Consistent Field (LSCF) strategy developed by the Rivail and Assfeld group [27, 28].

Although these papers have shown the usefulness of the ELMOs in treating a lot of chemical problems, it should be noted that the reduction in the number of variational parameters, used to describe the ELMOs, leads to increase in energy with respect to the Hartree-Fock case. So it is evident that the ELMO wavefunction has to be relaxed in order to increase its accuracy. To accomplish this task, we have recently shown that the virtual orbitals defined in the framework of the ELMOs approach can be used to perform a non-orthogonal configuration interaction (ELMO-VB method) [29]. The localization of the virtual orbitals can be wisely exploited to reduce the expansion.

In order to further improve this approach, in this paper we show that, following similar strategies adopted in the framework of SCF-MI [30-32] and Spin-Coupled [33-36] theories, it is possible to define optimized virtual ELMOs, which constitute an ideal basis to improve the description supplied by the ELMO wavefunction.

Theory

Let us consider a 2*N* electrons closed shell molecule. A localization scheme is then introduced subdividing the system into *f* fragments (which can overlap). Each fragment is described by a suitable number of ELMOs defined using only the atomic orbitals centred on the atoms which belong to the fragment.

So the α-th ELMO for the *i*-th fragment is defined by:

$$\left|\varphi_{\alpha}^{i}\right\rangle = \sum_{\mu=1}^{M_i} C_{\mu\alpha}^{i} \left|\chi_{\mu}^{i}\right\rangle$$

where M_i is the number of basis functions assigned to the *i*-th fragment.

The coefficients $C_{\mu\alpha}^{i}$ are determined minimizing the energy associated with the ELMO wavefunction:

$$\left|\Psi_{ELMO}\right\rangle = \hat{\mathrm{A}}\left[\prod_{i=1}^{f}\prod_{\alpha=1}^{N_i}\varphi_{\alpha}^{i}\,\overline{\varphi}_{\alpha}^{i}\right] = \hat{\mathrm{A}}\left[\Phi\right]$$

where $\hat{\mathrm{A}}$ is the antisymmetrizer and N_i is the number of doubly occupied ELMOs for the *i*-th fragment. It can be shown [19] that the ELMOs must satisfy the following set of coupled equations (one for each fragment):

$$\hat{F}_i \left|\varphi_{\alpha}^{i}\right\rangle = \varepsilon_{\alpha}^{i} \left|\varphi_{\alpha}^{i}\right\rangle \qquad \begin{cases} i = 1,\ldots..,f \\ \\ \alpha = 1,....,M_i \end{cases}$$

where $\hat{F}_i$ is a suitable hermitian operator for the *i*-th fragment built up using the occupied ELMOs of all the fragments, and ε_{α}^{i} are the orbital energies corresponding to the $\left|\varphi_{\alpha}^{i}\right\rangle$ orbitals. In order to improve both the stability and the convergence properties, we have devised a suitable algorithm already described elsewhere [24].

Thus each fragment is characterized by a set of occupied and virtual orbitals which share the same localization scheme. Occupied and virtual orbitals belonging to the same fragment constitute an orthogonal set, while orbitals of different fragments are non-orthogonal.

The energy associated with the ELMO wavefunction is naturally greater than the Hartree-Fock one due to the smaller number of variational parameters. This difference is slightly larger when the ELMOs are determined on smaller systems and then transferred to the target molecule without optimization.

To overcome this problem we have already shown [29] that it is possible to relax the ELMO wavefunction using single excitations from the occupied ELMOs of a fragment to the virtual ELMOs of the other fragments, giving rise to the so called ELMO-VB wavefunction:

$$\left|\Psi_{ELMO-VB}\right\rangle = c_0 \left|\Psi_{ELMO}\right\rangle + \sum_{i,r=1}^{f} \sum_{\alpha=1}^{N_i} \sum_{\mu=1}^{N_r^{virt}} c_{\alpha\mu}^{ir} \left|\Psi_{\alpha\mu}^{ir}\right\rangle \qquad (1)$$

where N_r^{virt} is the number of virtual ELMOs of the *r*-th fragment, $c_{\alpha\mu}^{ir}$ is the weight of the structure $\left|\Psi_{\alpha\mu}^{ir}\right\rangle$ that is connected to a single excitation from the α-th occupied ELMO of the *i*-th fragment to the μ-th virtual orbital of the *r*-th fragment. The structure $\left|\Psi_{\alpha\mu}^{ir}\right\rangle$ is expressed by:

$$\left|\Psi_{\alpha\mu}^{ir}\right\rangle = \hat{A}\left[\Phi\left(\begin{smallmatrix} i \\ \alpha \end{smallmatrix}\right) \varphi_\alpha^i \varphi_\mu^r \Theta_{0,0}^2\right]$$

with $\Phi\left(\begin{smallmatrix} i \\ \alpha \end{smallmatrix}\right)$ as the product of all the doubly occupied ELMOs with the exception of the pair $\varphi_\alpha^i \overline{\varphi}_\alpha^i$ (i.e. $\Phi = \Phi\left(\begin{smallmatrix} i \\ \alpha \end{smallmatrix}\right) \varphi_\alpha^i \overline{\varphi}_\alpha^i$) and $\Theta_{0,0}^2$ as the singlet spin eigenfunction for the electrons that belong to the orbitals $\varphi_\alpha^i \varphi_\mu^r$ which are singly occupied.

It is worthwhile to note that in the ELMO approach the fragments can share atomic orbitals, so resulting in the total number of ELMOs being greater than the total number of basis functions.

Therefore, linear dependence can naturally occur in the ELMO-VB method, which can be overcome by a proper selection of virtual orbitals as described elsewhere [29].

The ELMO-VB expansion could be further improved by introducing higher excitations, but the computational cost would become excessively demanding in terms of CPU time. The main goal of the present approach is to construct compact wavefunctions for a molecule using information obtained from its fragments and in this way to provide results similar to a traditional Hartree-Fock calculation carried out on the whole system.

From our experience, just single excitations can suffice for this purpose, but sometimes several virtual orbitals are needed, thus increasing the length of the VB expansion. Therefore, we have taken into account the possibility to define a small set of optimal virtual orbitals to be used in the construction of the excitations.

For each occupied ELMO of the generic fragment *i*, $\left| \varphi_{\alpha}^{i} \right\rangle$, we can determine an optimized virtual orbital $\left| \varphi_{(\alpha,\, i)}^{opt} \right\rangle$, namely an orbital satisfying these features:

- it has to be a linear combination of all the possible "usual" virtual ELMOs belonging to the $f^{(i)}$ first neighbour fragments to the *i*-th fragment (i.e. fragments whose corresponding ELMOs share some atomic functions with the *i*-th fragment or, in other words, fragments that are not separated by any bonds from the *i*-th fragment; see also the calculations section for a definition of all the topological relationships considered in this paper):

$$\left| \varphi_{(\alpha,\, i)}^{opt} \right\rangle = \sum_{r=1}^{f^{(i)}} \sum_{\mu=1}^{N_r^{virt}} c_{\alpha\mu}^{ir} \left| \varphi_{\mu}^{r} \right\rangle \qquad (2)$$

- it has to minimize the energy associated with the following expansion:

$$\left|\Psi^{\,i}_{\alpha\,ELMO-VB}\right\rangle = c_0\left|\Psi_{ELMO}\right\rangle + \hat{\mathrm{A}}\left[\Phi\left(^{\,i}_{\alpha}\right)\varphi^{\,i}_{\alpha}\,\varphi^{opt}_{(\alpha,\,i)}\,\Theta^{2}_{0,0}\right] \tag{3}$$

Substituting equation (2) into (3) we obtain:

$$\left|\Psi^{\,i}_{\alpha\,ELMO-VB}\right\rangle = c_0\left|\Psi_{ELMO}\right\rangle + \sum_{r=1}^{f^{(i)}}\sum_{\mu=1}^{N_r^{virt}} c^{\,i\,r}_{\alpha\mu}\;\hat{\mathrm{A}}\left[\Phi\left(^{\,i}_{\alpha}\right)\varphi^{\,i}_{\alpha}\,\varphi^{\,r}_{\mu}\,\Theta^{2}_{0,\,0}\right] \tag{4}$$

From this relation it is evident that the coefficients of the linear combination (2) are the structure weights supplied by the ELMO-VB calculation performed using the expansion (4) which involves only the excitations from the occupied orbital $\left|\varphi^{\,i}_{\alpha}\right\rangle$.

Through a series of trivial ELMO-VB computations we therefore obtain the optimized virtual ELMO for each occupied orbital. These optimal orbitals can be used to set up very compact ELMO-VB wavefunctions:

$$\left|\Psi_{ELMO-VB}\right\rangle = c_0\left|\Psi_{ELMO}\right\rangle + \sum_{i,r=1}^{f}\sum_{\alpha=1}^{N_i}\sum_{\mu=1}^{N_r^{virt}} c^{\,i\,r}_{\alpha\mu}\;\hat{\mathrm{A}}\left[\Phi\left(^{\,i}_{\alpha}\right)\varphi^{\,i}_{\alpha}\,\varphi^{opt}_{(\mu,\,r)}\,\Theta^{2}_{0,0}\right]$$

where, unlike equation (1), it is always $N_i^{virt} \le N_i \quad \left(i = 1, \ldots, f\right)$. Moreover, it should be pointed out that the ELMO-VB expansions employing optimized virtual orbitals implicitly take into account a very large amount of usual virtual ELMOs because the former are linear combinations of the latter, as it can be clearly seen from equation (2).

Calculations

In this section, to illustrate the capabilities of the ELMO-VB method and the use of optimized virtual ELMOs, we will present the results obtained on the alanine and serine molecules (both at the 6-31G and at the 6-31G** level) and on the alanine-serine (Ala-Ser) dipeptide (using the standard 6-31G basis set). For all these molecules we have considered geometries optimized at the Restricted Hartree-Fock / 6-31G level.

RHF calculations have been performed using the PC-GAMESS package [37, 38]. A modified version of PC-GAMESS [24] has been used for the ELMOs determination while the ELMO-VB calculations have been carried out with a program written in our laboratory [29,39].

We have first analysed the cases for the alanine and serine molecules. Their optimized geometries are reported in Fig. 1A and 1B, respectively. The ELMOs computations have been performed using the most localized partitioning schemes corresponding to traditional Lewis structures:

- *atomic fragments*, i.e. fragments defined using basis functions centred on only one atom to describe core and lone pair electrons;
- *bond fragments*, i.e. fragments defined using atomic orbitals centred on two atoms to describe bond electrons.

The atomic fragments of the carbon, nitrogen and oxygen atoms are described by 1, 2 and 3 occupied ELMOs, respectively. The bond fragments are characterized by 1 or 2 occupied ELMOs according to their bond order. All the energy values for the alanine-6-31G (alanine-6-31G**) // serine-6-31G (serine-6-31G**) are reported in Table I // II; we observe that the energy associated with the ELMO wavefunction is 74.2 (83.6) // 84.7 (96.3) kcal/mol higher than the RHF value, while the difference increases to 150.7 (152.2) // 175.8 (176.7) kcal/mol when the wavefunction is built up with Pipek-Mezey LMOs [18] subjected to tails deletion (LMO-TD) according to the adopted localization scheme.

To improve the description given by the ELMO wavefunction we have performed three ELMO-VB calculations taking into account one virtual ELMO for each fragment (ELMO-VB [1Fr]), one virtual ELMO for each occupied orbital (ELMO-VB [1Oc]) and two virtual ELMOs for each fragment (ELMO-VB [2Fr]). The selected virtual orbitals are those with the lowest orbital energies. The ELMO-VB [1Fr] expansion for alanine-6-31G (alanine-6-31G**) // serine-6-31G

(serine-6-31G**) consists of 432 // 560 excitations (clearly the number of excitations is independent of the basis set and so only one value for each molecule is reported). This allows to recover 60.0 % (41.0 %) // 57.3 % (37.9 %) of the difference between the RHF and the ELMO energies. A slightly larger gain is achieved with the ELMO-VB [1Oc] wavefunction that yields a 65.2 % (44.6 %) // 63.4 % (42.9 %) decrease with 576 // 784 structures. Finally, the ELMO-VB [2Fr] expansion considers 864 // 1120 excitations and recovers 79.1 % (55.4 %) // 78.2 % (54.1 %) of the energy difference.

We have previously shown [29] that we can exploit the localized nature of the virtual orbitals to introduce a topological selection of the excitations which permits to reduce considerably the number of the included structures still giving roughly the same energy values as the full calculation. This approach is based on the fact that, in order to relax the ELMO wavefunction, it is necessary to introduce a partial delocalization of the orbitals. It is evident that the most important contributions will arise from the atomic functions close to the ELMOs themselves. A considerable saving of CPU time can then be achieved introducing only excitations from an occupied ELMO to virtual orbitals localized on 'near' fragments. In order to efficiently perform this selection, it is useful to classify the orbitals on the basis of their spatial topological relationships [29].

To this purpose the fragments can be topologically classified into:

A) *first neighbour fragments*, i.e. fragments that are not separated by any bonds (for instance, in Fig. 1A, the fragments N_2 and N_2-C_3 or N_2-H_1 and N_2-C_3);

B) *second neighbour fragments*, i.e. fragments separated by only one bond (for instance, in Fig. 1A, the fragments N_2 and C_3 or N_2-H_1 and C_3-C_4);

C) *third neighbour fragments*, i.e. fragments separated by two bonds (for instance, in Fig. 1A, the fragments N_2 and C_4-O_5 or N_2-H_1 and C_4).

The proposed topological selection [29] consists in introducing into the ELMO-VB wavefunction only those excitations which satisfy one of the following criteria:

1) excitations from occupied ELMOs describing a bond or a lone pair to virtual ELMOs localized on the *first and second neighbour fragments*;
2) excitations from occupied ELMOs describing a bond or a lone pair to virtual ELMOs localized on the *third neighbour atomic fragments*.

By following these rules, the number of excitations in the ELMO-VB [1Fr], ELMO-VB [1Oc] and ELMO-VB [2Fr] wavefunctions can be greatly reduced. They involve 203 // 243, 275 //336 and 406 // 486 excitations respectively and allow to recover 54.7 % (33.3 %) // 53.2 % (32.5 %), 59.6 % (38.4 %) // 59.4 % (38.6 %) and 67.7 % (46.2 %) // 71.1 % (48.3 %) of the energy difference between the RHF and the ELMO values (see Tables I and II). These results strongly support the reliability of the topological selection criterion, which allows to significantly reduce the length of the ELMO-VB expansions affecting only slightly the accuracy of the results.

The capabilities supplied by the optimized virtual ELMOs for alanine and serine are illustrated in Table I and II, respectively. For each occupied orbital we have performed an ELMO-VB calculation introducing all the possible single excitations to the virtual orbitals of its first neighbour fragments, thus obtaining the corresponding optimized virtual orbital (see equation (2) in the Theory section). It should be pointed out that these calculations are very fast as they involve only a few dozen excitations.

After the determination of the set of optimized virtual ELMOs, we have generated an ELMO-VB wavefunction built up using single excitations to all the optimal virtual orbitals (ELMO-VB* [1Oc]) or employing only one optimized virtual orbital for each fragment (ELMO-VB* [1Fr]). Particular care has to be paid in choosing only one optimal virtual ELMO for fragments characterized by more than one occupied orbital: in this paper we have selected that optimal

virtual orbital whose corresponding ELMO-VB expansion for optimization (see equation (3)) has the lowest energy value. The ELMO-VB* [1Oc] calculation for alanine-6-31G (alanine-6-31G**) // serine-6-31G (serine-6-31G**) molecule is constituted by 576 // 784 structures and yields a 75.7 % (76.6 %) // 75.1% (74.0%) decrease in energy. In the ELMO-VB* [1Fr] wavefunction the number of excitations is reduced to 432 // 560 excitations and the corresponding energy gain is 71.0% (72.0 %) // 68.6% (67.9 %).

We have also applied the topological selection to the optimized virtual ELMOs. Following the above rules, the number of excitations for the ELMO-VB* [1Oc] is reduced to 275 // 336, allowing to recover 69.1 % (75.1 %) // 67.9% (72.4 %) of the RHF / ELMO energy difference. Similarly the ELMO-VB*[1Fr] expansion is constructed with 203 // 243 structures and provides a 65.0 % (70.3 %) // 61.9 % (66.3 %) decrease in energy.

From the analysis of the results reported in Tables I and II, it is evident that the use of optimal virtual orbitals yields better results than those obtained with usual virtual ELMOs. Moreover, it is important to emphasize the fact that optimized virtual ELMOs, together with the topological selection, give energy lowerings that are always larger than those achieved by means of the analogous complete expansions built up without optimized orbitals, i.e. ELMO-VB [1Fr] and ELMO-VB [1Oc]. Furthermore, as it can be clearly seen from Tables I and II, by employing conventional virtual ELMOs and an equal number of excitations, it is always possible to obtain better energy gains using the 6-31G rather than the 6-31G** basis set. This is probably due to the greater reduction in the number of variational parameters for the ELMO/6-31G** wavefunction than the one for the ELMO/6-31G wavefunction. Thus, in order to increase the decrease in energy associated with the ELMO-VB/6-31G** expansions, more excitations should be considered. Nevertheless, it is important to note that this drawback disappears using optimized virtual ELMOs. In fact for each 6-31G** virtual orbital optimization a slightly greater number of

excitations are taken into account (with respect to the 6-31G case) and so more conventional virtual orbitals are also implicitly considered in the ELMO-VB*/6-31G** wavefunctions (with respect to the ELMO-VB*/6-31G expansions). These results further enhance the usefulness of the optimal virtual ELMO approach.

In the second part of our calculations, we have focused our attention on the possibility to extend this approach to the ELMOs transferred from one molecule to another one. To this purpose we have considered the alanine-serine dipeptide, whose geometry has been optimized at the RHF/6-31G level (see Fig. 1-D). We have carried out an ELMOs calculation according to the most localized partitioning scheme. The same molecular fragments definitions used for the alanine and serine molecules have been adopted. All the energy values for the dipeptide are reported in Table III. It may be observed that the energy associated with the ELMO wavefunction is higher than the RHF value by 160.5 kcal/mol. However, we also note that the use of the Pipek-Mezey LMOs, subjected to tails deletion according to the proposed localization scheme, increases the energy gap up to 316.2 kcal/mol (LMO-TD).

We have then assembled the electronic structure of the dipeptide employing the ELMOs calculated for the alanine and the serine molecules. For the peptide bond we have used ELMOs determined for the formamide molecule. The transferring scheme is outlined in Fig. 2. The orbitals have been properly rotated from these molecules to the target dipeptide molecule and subjected to a simple renormalization to take into account the changes in the bond lengths. It is important to note that the ELMO wavefunctions are constructed using non-orthogonal sets of orbitals and so there is no need to orthogonalize the transferred and normalized orbitals.

In this way we have built up an ELMO wavefunction, labelled ELMO-Tr, whose energy is only 14.2 kcal/mol higher than the one obtained by the ELMOs calculation carried out on the target alanine-serine dipeptide (see ELMO in Table III). This result indicates that the ELMOs

transferability is highly reliable. This property is further enhanced by considering the fact that the geometries of the molecules, used to determine the ELMOs, are quite different from the geometries that these moieties adopt in the target molecule, as it can be seen by comparing Figs. 1A and 1B with Fig. 1D.

In order to relax the ELMO-Tr wavefunction, we have performed ELMO-VB calculations using the transferred virtual orbitals and exploiting the same strategy reported above for the alanine and serine molecules. The ELMO-VB [1Fr] and ELMO-VB [1Oc] expansions consist of 1645 and 2209 structures and recover 55.6 % and 61.8 % of the ELMO-Tr / RHF energy difference. By using optimized virtual orbitals the percentages increase to 63.6 % and 67.6 %.

The introduction of the topological criterion reduces the number of excitations to 448 and 594; the percentage of the energy recovered amounts to 43.7 % and 53.3 % when using the canonical virtual ELMOs and to 58.6% and 61.5% when employing the optimized virtual orbitals. Once again the topological criterion allows to greatly reduce the computational cost while retaining the overall accuracy.

We have also analysed the molecular electron densities provided by the present approach. From a visual comparison it appears that the electron distributions computed at the RHF, ELMO and ELMO-VB levels are globally identical for the alanine and serine molecules (considering both basis sets) as well as for the Ala-Ser dipeptide. In order to compare quantitatively the electron densities, we have calculated the index $L(a, a')$ introduced by Walker and Mezey [11], which compares point-by-point two charge distributions X and Y in the space bounded by the two isosurfaces characterized by the values a and a', that is in the density "shell" $S(a, a', X)$ defined as:

$$S(a, a', X) = \{\mathbf{r}:\ a \le \rho_X(\mathbf{r}) \le a'\}$$

So we have:

$$L\left(a,a'\right)=\left\{L^{*}\left(a,a',X,Y\right)+L^{*}\left(a,a',Y,X\right)\right\}/2$$

where

$$L^{*}\left(a,a',X,Y\right)=1-\left[\sum_{\mathbf{r}\in S\left(a,a',X\right)}\left\{\left|\rho_{X}\left(\mathbf{r}\right)-\rho_{Y}\left(\mathbf{r}\right)\right|/\max\left(\rho_{X}\left(\mathbf{r}\right),\rho_{Y}\left(\mathbf{r}\right)\right)\right\}\right]\Big/ n\left(S\left(a,a',X\right)\right)$$

$$L^{*}\left(a,a',Y,X\right)=1-\left[\sum_{\mathbf{r}\in S\left(a,a',Y\right)}\left\{\left|\rho_{X}\left(\mathbf{r}\right)-\rho_{Y}\left(\mathbf{r}\right)\right|/\max\left(\rho_{X}\left(\mathbf{r}\right),\rho_{Y}\left(\mathbf{r}\right)\right)\right\}\right]\Big/ n\left(S\left(a,a',Y\right)\right)$$

$n\left(S\left(a,a',X\right)\right)$ and $n\left(S\left(a,a',Y\right)\right)$ being the number of grid points falling in the density shell $S\left(a,a',X\right)$ and $S\left(a,a',Y\right)$, respectively.

The values of the similarity measures, reported in Tables IV, V and VI, show that the ELMO charge distributions reproduce very well the corresponding RHF electron densities and that the agreement is further increased by the ELMO-VB calculations. It has to be noted that, for all the electron distributions, the similarity is greater close to the nuclei, as indicated by the L (0.1,10) indices. Furthermore, for the alanine and serine molecules, we can observe that the ELMO/6-31G** electron densities are generally better than the ones yielded by the RHF/6-31G calculations, even if, as a result of the extreme localization of the involved orbitals, this feature changes far from the nuclei (see indexes L (0.001,0.01)). Then, taking into account the alanine-serine dipeptide, it is worthwhile to stress that just transferred ELMOs supply an electron density very close to the Hartree-Fock one.

Finally, in order to highlight finer details of the electronic distributions, we have examined the Molecular Electrostatic Potential (MEP) at the RHF, ELMO and ELMO-VB levels for all the molecules. In particular, we have plotted the MEPs on the corresponding RHF electron density isosurface of 0.001 a.u. and a general similarity has been observed. As an example, in Fig. 3 we

have reported the electrostatic potential maps for the Ala-Ser system; also in this case it is possible to observe that the transfer of ELMOs allows the RHF MEP to be reproduced very well and that the ELMO-VB computations improve the agreement.

Conclusions

The most appealing feature of the ELMOs is the complete absence of tails. Hence, they naturally describe single bonds, lone pairs or functional groups and so they can be used as building blocks to set up the wavefunction of large systems. The present results have clearly indicated that the ELMO wavefunction is more reliable than that supplied by the LMOs with deleted tails. However, due to the unavoidable decrease in the number of variational parameters which define the wavefunction, its energy is higher than the RHF one. Therefore, it is of great importance to develop methods that improve the description of the electronic structure yielded by the ELMO wavefunction itself.

The so called ELMO-VB strategy consists of a non-orthogonal configuration interaction of single excitations from the occupied orbitals of a fragment to the virtual space of the other fragments. However, due to the extremely localized nature of the occupied and virtual orbitals, the expansion can be greatly reduced. A topological selection based on the concept of fragments neighbourhood has been introduced showing that it is possible to reduce greatly the computational requirements and maintain good accuracy.

In order to keep the compactness of the expansion, and to implicitly include at the same time the effect of more virtual orbitals, we have proposed a new strategy which allows to obtain optimal virtual orbitals to be used in a usual ELMO-VB computation. From the reported calculations on two test molecules (alanine and serine) it is possible to observe that, with the same number of selected virtual orbitals, the new approach always yields better results than the traditional one

and, in addition, that the use of optimized virtual ELMOs together with the topological selection is a suitable strategy to get a significant decrease in energy with a relatively small computational cost.

The potential advantage of employing ELMOs relies on the possibility to transfer them from one molecule to another. In this paper we have transferred the ELMOs determined on alanine, serine and formamide to set up the wavefunction for the alanine-serine dipeptide. Although the transferred ELMOs have been determined at a different geometry with respect to the target molecule, the ELMO-VB method has provided energies closer to the RHF value. In this case too, the topological criterion has been effective in reducing the number of configurations to be used without affecting the accuracy of the results.

In conclusion, it should be observed that the ELMO approximation, even when employing non-relaxed transferred ELMOs, allows to reproduce well the Hartree-Fock electron density and Electrostatic Potential of a molecule and that the ELMO-VB strategy further increases the similarity.

Acknowledgement

We would like to thank Dr. Monica Civera for technical support.

References

1. Goedecker, S. Rev Mod Phys 1999, 71, 1085.
2. Li, X.-P.; Nunes, W.; Vanderbilt, D. Phys Rev B 1993, 47, 10891.
3. Yang, W. Phys Rev Lett 1991, 66, 1438.
4. Yang, W. Phys Rev A 1991, 44 7823.
5. Yang, W.; Lee, T.-S. J Chem Phys 1995, 103, 5674.
6. Dixon, S. L.; Merz, K. M., Jr. J Chem Phys 1996, 104, 6643.
7. Dixon, S. L.; Merz, K. M., Jr. J Chem Phys 1997, 107, 879.
8. Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. J Phys Chem 1994, 98, 9165.
9. Babu, K.; Gadre, S. R. J Comput Chem 2003, 24, 484.
10. Babu, K.; Ganesh V.; Gadre S. R.; Ghermani N. E. Theor Chem Acc 2004, 111, 255.
11. Walker, P. D.; Mezey, P. G. J Am Chem Soc 1994, 116, 12022.
12. Mezey, P. G. J Math Chem 1995, 18, 141.
13. Exner, T. E.; Mezey, P. G. J Phys Chem A 2002, 106, 11791.
14. Exner, T. E.; Mezey, P. G. J Comput Chem 2003, 24, 1980.
15. Hierse, W.; Stechel, E. S. Phys Rev B 1996, 54, 16515.
16. Boys, S. F. Rev Mod Phys 1960, 32, 296.
17. Edminston, C.; Ruedenberg, K. J Chem Phys 1965, 43, 597.
18. Pipek, J.; Mezey, P. G. J Chem Phys 1989, 90, 4916.
19. Stoll, H.; Wagenblast, G.; Preuss, H. Theoret Chim Acta 1980, 57, 169.
20. Szekeres, Z.; Surján, P. R. Chem Phys Lett 2003, 369, 125.
21. Sironi, M.; Famulari, A. Theor Chem Acc 2000, 103, 417.
22. Nàray-Szabò G. Comput Chem 2000, 24, 287.
23. Smits, G. F.; Altona, C. Theoret Chim Acta 1985, 67, 461.

24. Fornili, A.; Sironi, M.; Raimondi, M. J Mol Struct (Theochem) 2003, 632, 157.

25. Burresi, E.; Sironi, M. Theor Chem Acc, 2004, 112, 247.

26. Sironi, M.; Famulari, A.; Raimondi, M.; Chiesa, S. J Mol Struct (Theochem) 2000, 529, 47.

27. Assfeld, X.; Rivail, J. L. Chem Phys Lett 1996, 263, 100.

28. Ferrè, N.; Assfeld, X.; Rivail, J. L. J Comput Chem 2002, 23, 610.

29. Genoni, A.; Sironi, M. Theor Chem Acc, 2004, 112, 254.

30. Famulari, A.; Raimondi, M.; Sironi, M.; Gianinetti, E. Chem Phys 1998, 232, 289.

31. Famulari, A.; Specchio, R.; Sironi, M.; Raimondi, M. J Chem Phys 1998, 108, 3296.

32. Specchio, R.; Famulari, A.; Sironi, M.; Raimondi, M. J Chem Phys 1999, 111, 6204.

33. Raimondi, M. Mol Phys 1984, 53, 161.

34. Raimondi, M.; Sironi, M.; Gerrat, J.; Cooper, D. L. Int J Quantum Chem 1996, 60, 225.

35. Clarke, N. J.; Raimondi, M.; Sironi, M.; Gerrat, J.; Cooper, D. L. Theor Chem Acc 1998, 99, 8.

36. Martinazzo, R.; Famulari, A.; Raimondi, M.; Bodo, E.; Gianturco, F. A. J Chem Phys 2001, 115, 2917.

37. Granovsky, A. A. http://classic.chem.msu.su/gran/gamess/index.html.

38. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. J.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S., Windus, T. L.; Dupuis, M.; Montgomery J. A. J Comput Chem 1993, 14, 1347.

39. Genoni, A.; Sironi, M. A general purpose ELMO-VB program; Università degli Studi di Milano, Milan, 2003.

Figure captions

Fig. 1 Optimized geometries at the RHF / 6-31G level: A) alanine; B) serine; C) formamide; D) alanine-serine.

Fig. 2 The molecular fragments transferring for the alanine-serine dipeptide: A) alanine; B) serine; C) formamide; D) alanine-serine. The transferred fragments from alanine, serine and formamide are depicted in grey, black & white and black respectively.

Fig. 3 Molecular Electrostatic Potential maps (blue = positive, red = negative) at different levels for the alanine-serine dipeptide: A) RHF; B) ELMO; C) ELMO-VB* [1Oc] with topological selection. The electrostatic potentials are plotted on the RHF electron density isosurface of 0.001 a.u..

Table I – Energy values for the alanine molecule (6-31G and 6-31G**) [(a)]

						Topological selection				
Method		6-31G		6-31G**			6-31G		6-31G**	
	NE [(b,c)]	ΔE (kcal/mol) [(d)]	% [(c,e)]	ΔE (kcal/mol) [(d)]	% [(c,e)]	NE [(b,c)]	ΔE (kcal/mol) [(d)]	% [(c,e)]	ΔE (kcal/mol) [(d)]	% [(c,e)]
LMO-TD		150.7		152.2						
ELMO		74.2		83.6						
ELMO-VB [1Fr]	432	29.7	60.0	49.3	41.0	203	33.6	54.7	55.8	33.3
ELMO-VB [1Oc]	576	25.8	65.2	46.3	44.6	275	30.0	59.6	51.5	38.4
ELMO-VB [2Fr]	864	15.5	79.1	37.3	55.4	406	24.0	67.7	45.0	46.2
ELMO-VB* [1Fr]	432	21.5	71.0	23.4	72.0	203	26.0	65.0	24.8	70.3
ELMO-VB* [1Oc]	576	18.0	75.7	19.6	76.6	275	22.9	69.1	20.8	75.1

(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) $\%\,(ELMO-VB) = 100\cdot\left|\,\Delta E\left(ELMO-VB\right)-\Delta E\left(\,ELMO\right)\right|\,/\,\left|\,\Delta E\left(\,ELMO\right)\right|$

Table II – Energy values for the serine molecule (6-31G and 6-31G**) [(a)]

						Topological selection				
Method		6-31G		6-31G**			6-31G		6-31G**	
	NE [(b,c)]	ΔE (kcal/mol) [(d)]	% [(c,e)]	ΔE (kcal/mol) [(d)]	% [(c,e)]	NE [(b,c)]	ΔE (kcal/mol) [(d)]	% [(c,e)]	ΔE (kcal/mol) [(d)]	% [(c,e)]
LMO-TD		175.8		176.7						
ELMO		84.7		96.3						
ELMO-VB [1Fr]	560	36.2	57.3	59.8	37.9	243	39.6	53.2	65.0	32.5
ELMO-VB [1Oc]	784	31.0	63.4	55.0	42.9	336	34.4	59.4	59.1	38.6
ELMO-VB [2Fr]	1120	18.5	78.2	44.2	54.1	486	24.5	71.1	49.8	48.3
ELMO-VB* [1Fr]	560	26.6	68.6	30.9	67.9	243	32.3	61.9	32.5	66.3
ELMO-VB* [1Oc]	784	21.1	75.1	25.0	74.0	336	27.2	67.9	26.5	72.4

(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) $\%\,(ELMO-VB) = 100\cdot\left|\,\Delta E\left(ELMO-VB\right)-\Delta E\left(\,ELMO\right)\right|\,/\,\left|\,\Delta E\left(\,ELMO\right)\right|$

Table III – Energy values for the Ala-Ser molecule (6-31G) [(a)]

				Topological Selection		
Method	NE [(b,c)]	ΔE (kcal/mol) [(d)]	% [(c,e)]	NE [(b,c)]	ΔE (kcal/mol) [(d)]	% [(c,e)]
LMO-TD		316.2				
ELMO		160.5				
ELMO-Tr		174.7				
ELMO-VB [1Fr]	1645	77.6	55.6	448	98.3	43.7
ELMO-VB [1Oc]	2209	66.8	61.8	594	81.6	53.3
ELMO-VB* [1Fr]	1645	63.6	63.6	448	72.4	58.6
ELMO-VB* [1Oc]	2209	56.6	67.6	594	67.2	61.5

(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 RHF value

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

Table IV – Computed similarity indexes versus the RHF/6-31G and RHF/6-31G** electron densities for the alanine molecule [a].

Compared electron density	L (0.001, 10) %		L (0.1, 10) %		L (0.01, 0.1) %		L (0.001, 0.01) %	
	6-31G	6-31G**	6-31G	6-31G**	6-31G	6-31G**	6-31G	6-31G**
ELMO [b]	96.81	96.65	98.89	98.67	97.62	97.30	95.65	95.66
ELMO-VB* [1Fr] [b]	97.12	96.82	98.94	98.76	97.77	97.44	96.17	95.88
ELMO-VB* [1Oc] [b]	97.25	96.84	98.94	98.76	97.79	97.44	96.41	95.90
RHF/6-31G	100.0	96.57	100.0	95.80	100.0	96.58	100.0	96.77

(a) Both the ELMO-VB wavefunctions are built up using the topological selection.

(b) Both the 6-31G and the 6-31G** electron densities are taken into account and they are compared with the RHF/6-31G and the RHF/631G** charge distributions, respectively.

Table V – Computed similarity indexes versus the RHF/6-31G and RHF/6-31G** electron densities for the serine molecule [a].

Compared electron density	L (0.001, 10) %		L (0.1, 10) %		L (0.01, 0.1) %		L (0.001, 0.01) %	
	6-31G	6-31G**	6-31G	6-31G**	6-31G	6-31G**	6-31G	6-31G**
ELMO [b]	96.50	96.44	98.88	98.66	97.31	97.03	95.20	95.40
ELMO-VB* [1Fr] [b]	96.86	96.65	98.93	98.73	97.46	97.17	95.81	95.68
ELMO-VB* [1Oc] [b]	97.00	96.68	98.92	98.73	97.49	97.18	96.08	95.74
RHF/6-31G	100.0	96.38	100.0	95.70	100.0	96.36	100.0	96.58

(a) Both the ELMO-VB wavefunctions are built up using the topological selection.

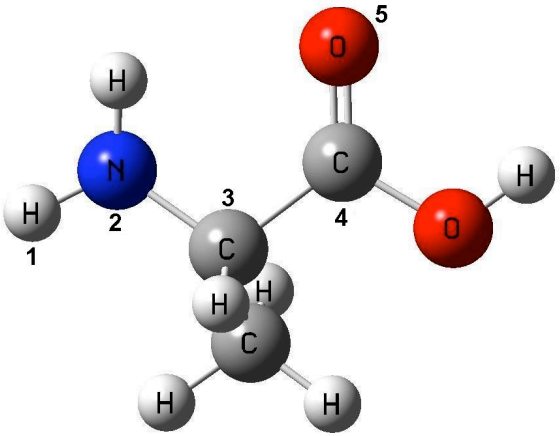
(b) Both the 6-31G and the 6-31G** electron densities are taken into account and they are compared with the RHF/6-31G and the RHF/631G** charge distributions, respectively.

Table VI – Computed similarity indexes versus the RHF/6-31G electron density for the Ala-Ser molecule [a].

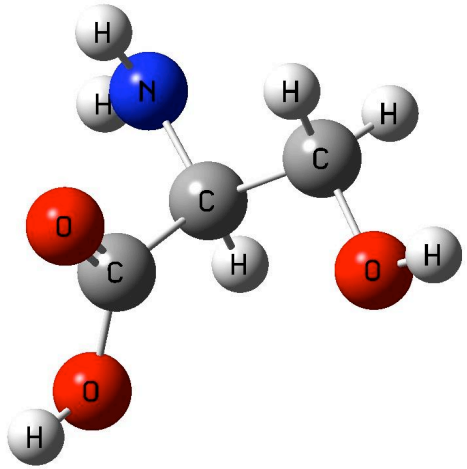
Compared electron density	L (0.001, 10) %	L (0.1, 10) %	L (0.01, 0.1) %	L (0.001, 0.01) %
ELMO-Tr	96.32	98.29	96.80	95.34
ELMO-VB* [1Fr]	96.45	98.34	96.91	95.50
ELMO-VB* [1Oc]	96.48	98.34	96.92	95.55

(a) Both the ELMO-VB wavefunctions are built up using the topological selection.

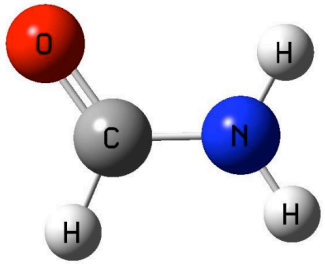
Figure 1 -



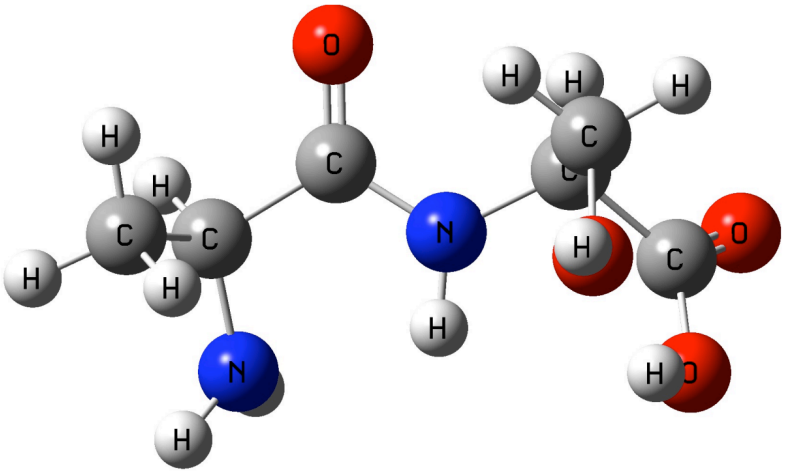
A



B



C



D

Figure 2-

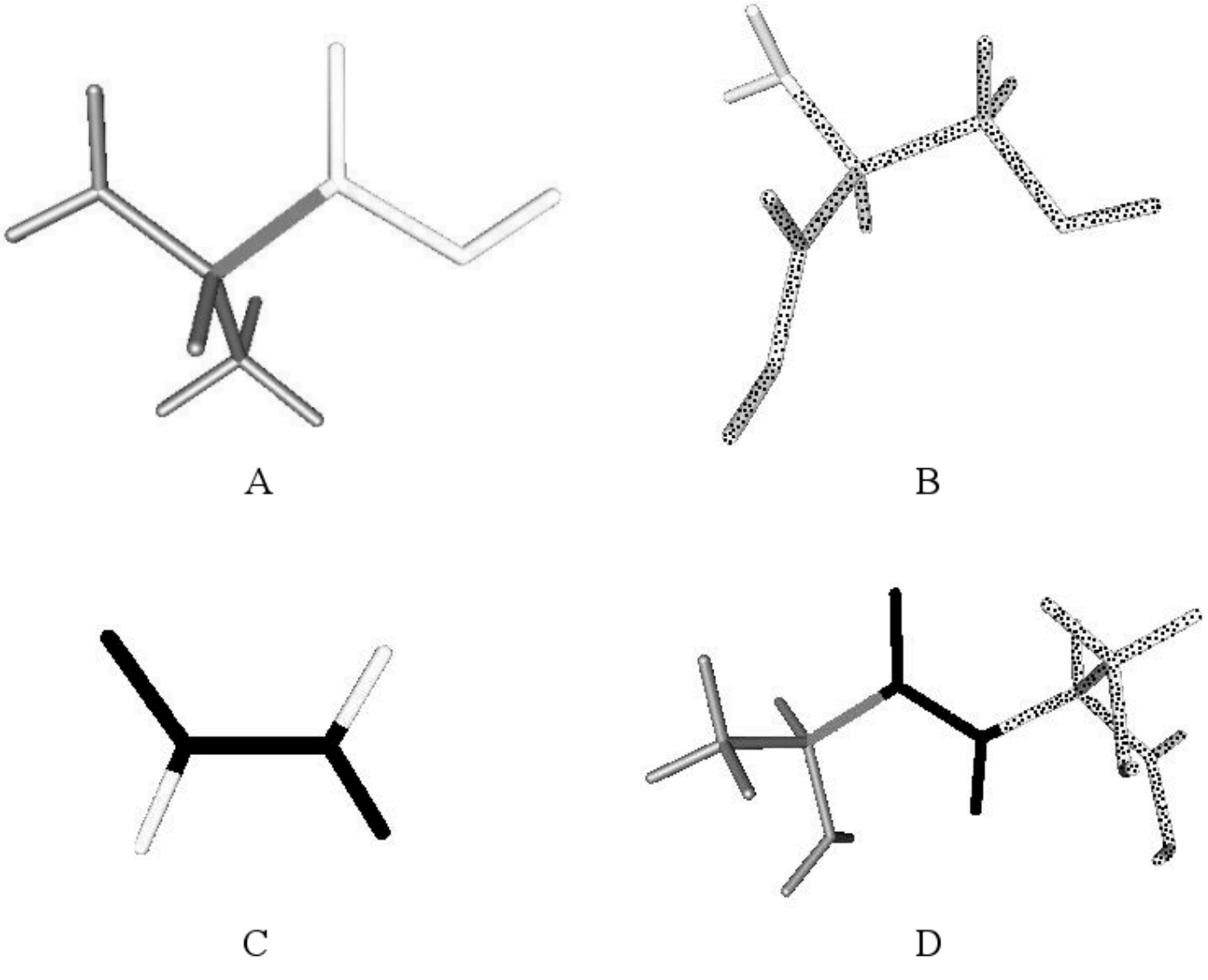
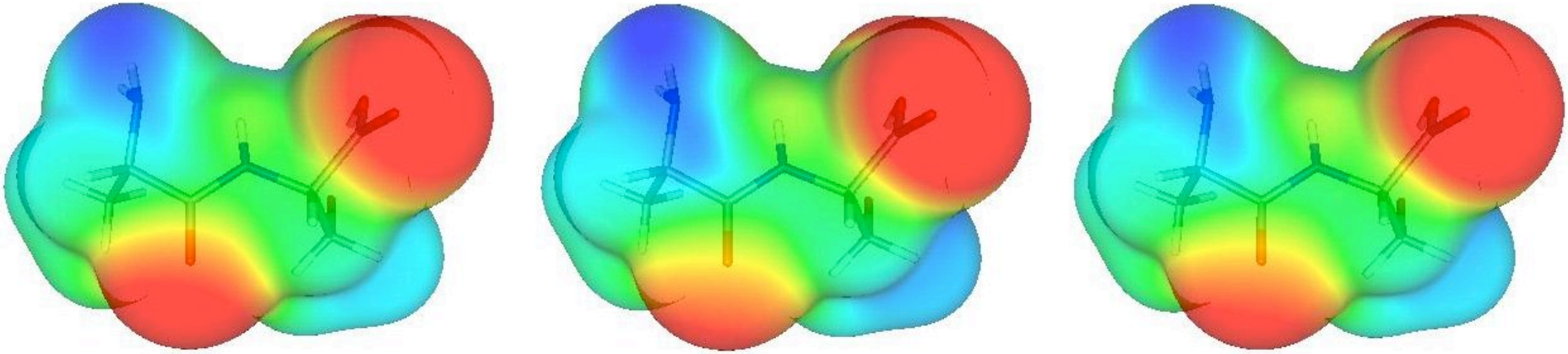


Figure 3-



A B C