

A comparative analysis by means of quantum molecular similarity measures of density distributions derived from conventional *ab initio* and density functional methods

Miquel Solà, Jordi Mestres, Ramon Carbó, and Miquel Duran
Department of Chemistry and Institute of Computational Chemistry, Universitat de Girona,
17071 Girona, Catalonia, Spain

(Received 28 April 1995; accepted 6 October 1995)

A procedure based on quantum molecular similarity measures (QMSM) has been used to compare electron densities obtained from conventional *ab initio* and density functional methodologies at their respective optimized geometries. This method has been applied to a series of small molecules which have experimentally known properties and molecular bonds of diverse degrees of ionicity and covalency. Results show that in most cases the electron densities obtained from density functional methodologies are of a similar quality than post-Hartree–Fock generalized densities. For molecules where Hartree–Fock methodology yields erroneous results, the density functional methodology is shown to yield usually more accurate densities than those provided by the second order Møller–Plesset perturbation theory. © 1996 American Institute of Physics. [S0021-9606(96)02502-4]

I. INTRODUCTION

The first-order density distribution, $\rho(\mathbf{r}_1)$, of a given electronic state is expressed in terms of its wave function Ψ according to¹

$$\rho(\mathbf{r}_1) = N \int \cdots \int |\psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)|^2 ds_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N. \quad (1)$$

The density function in Eq. (1) is a physical observable upon which other molecular properties, directly or indirectly, depend. For instance, the Hohenberg–Kohn theorem,² which is the basis of modern density functional theory (DFT),³ shows that the energy of the ground state can be expressed as a functional of the electron density. Not only the energy, but also any molecular property of the ground state depending of a one-electron operator can be expressed in terms of the one-electron density. Given that the density function is an observable, any theoretical method in the exact limit should reproduce the same density function, and therefore the same molecular properties. Since it is much easier to deal with the first-order electron density, which depends upon three spatial variables only, than with the multivariate wave function, a detailed comparison of different methodologies is usually carried out by making a systematic study of the electron density differences obtained from the methods being compared.⁴ The main goal pursued when comparing electron densities from different DFT methodologies is to discover the disadvantages and benefits of the different available density functionals, in order to assist researchers to build more accurate functionals. Moreover, these studies can also help to understand the successes and failures of DFT in the study of some chemical interactions,⁵ and also to shed light on how nonlocal corrections may influence the calculated electron density. The literature results demonstrate that $\rho^{\text{DFT}}(\mathbf{r}_1)$ con-

structed from SCF-converged Kohn–Sham orbitals is usually more adequate than $\rho^{\text{HF}}(\mathbf{r}_1)$ obtained from ordinary Hartree–Fock (HF) theory.^{3(b),4(a),4(b)}

The fact that $\rho^{\text{DFT}}(\mathbf{r}_1)$ yields in general a more accurate electronic distribution than $\rho^{\text{HF}}(\mathbf{r}_1)$ is also reflected in the significant improvement over Hartree–Fock calculations of molecular properties like equilibrium geometries, dipole moments, and harmonic vibrational frequencies.^{3,6,7} Also, bond dissociation energies, reaction energies and proton affinities are of the same quality as those obtained from post-Hartree–Fock methods if they are computed including nonlocal corrections in the functional.^{7,8} In most cases DFT offers a promising alternative to Hartree–Fock and post-Hartree–Fock methods. This is especially true for molecules where Hartree–Fock performance is very poor, and as a consequence, most correlated methods which use the single-determinant Hartree–Fock wave function as the reference configuration do not reproduce the correct electronic distributions.^{4(a),9}

In this paper we report an analysis of electron density distributions obtained from different methodologies by means of quantum molecular similarity measures (QMSM). The definition used in this work for the QMSM between two molecules $\{I, J\}$ with electron densities $\rho_I(\mathbf{r})$ and $\rho_J(\mathbf{r})$ is given by the integral¹⁰

$$Z_{IJ}(\Theta) = \int \int \rho_I(\mathbf{r}_1) \Theta(\mathbf{r}_1, \mathbf{r}_2) \rho_J(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2, \quad (2)$$

where $\Theta(\mathbf{r}_1, \mathbf{r}_2)$ is a positive definite operator depending on two-electron coordinates. When in Eq. (2) the operator $\Theta(\mathbf{r}_1, \mathbf{r}_2)$ is the Dirac delta function $\delta(\mathbf{r}_1 - \mathbf{r}_2)$ the overlaplike similarity measure is obtained

$$Z_{IJ} = \int \rho_I(\mathbf{r}) \rho_J(\mathbf{r}) d\mathbf{r}. \quad (3)$$

Likewise, the repulsionlike similarity is given by

$$Z_{IJ}(r_{12}^{-1}) = \int \int \rho_I(\mathbf{r}_1) \frac{1}{r_{12}} \rho_J(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2. \quad (4)$$

More general definitions than that those of Eq. (2) based on n th-order density matrices and many-electron Θ operators have also been proposed,¹¹ although to date most applications¹² use Eqs. (3) and (4) in order to compute the similarity between two molecular electron distributions. Once the QMSM has been calculated it is possible to define an euclidean distance between the molecular electronic distributions $\rho_I(\mathbf{r})$ and $\rho_J(\mathbf{r})$ as¹⁰

$$d_{IJ} = [Z_{II} + Z_{JJ} - 2Z_{IJ}]^{1/2}. \quad (5)$$

From definition of Eq. (5) it is found that the following equality, $d_{IJ} = d_{JI}$, holds since $Z_{IJ} = Z_{JI}$. The value of the distance given by Eq. (5) depends on the relative spatial orientation of molecular electron distributions $\rho_I(\mathbf{r})$ and $\rho_J(\mathbf{r})$; in this manner, their mutual orientation has to be optimized in order to maximize Z_{IJ} , which is equivalent to minimize the d_{IJ} value. A final optimized zero distance means that electron densities of electronic distributions $\rho_I(\mathbf{r})$ and $\rho_J(\mathbf{r})$ are completely equivalent, whereas a large value of the distance implies the existence of significant dissimilarities in the two electron density distributions.

To our knowledge, so far comparisons between density distributions have been performed by analyzing density difference contours only at a fixed geometry for all levels of theory,^{4(a),4(b)} and then reflecting only those changes explicitly due to electronic relaxation. The main interest in using QMSM instead of depicting electron density differences between density distributions $\rho_I(\mathbf{r})$ and $\rho_J(\mathbf{r})$, is the fact that with this methodology the analysis can be performed at any desired geometry, and in particular at the optimized geometry corresponding to each methodology employed, thus accounting for both nuclear and electronic relaxation. Therefore, the procedure presented here is deemed to be an extension to the standard analysis of the electron density difference maps.

In this work, comparison of electron densities obtained from different methodologies is carried out by means of QMSM in several small molecules with experimentally determined molecular properties and molecular bonds of various degrees of ionicity and covalency. The analysis performed here aims mainly to compare the electron densities obtained from two density functional formalisms and those obtained from Hartree–Fock and correlated *ab initio* methods, but obviously it can be used to compare any kind of methodologies from which an electron density distribution can be derived.

II. COMPUTATIONAL DETAILS

To minimize basis set effects, which may produce relevant QMSM differences,¹³ the 6-311++G** basis set¹⁴ was used throughout. Geometry optimization of all systems presented in this work was done using the Schlegel method.¹⁵ The GAUSSIAN-92¹⁶ program was used to perform

TABLE I. Euclidean distance matrices (in a.u.) of the formaldehyde molecule computed at the Hartree–Fock optimized geometry, obtained from (a) the exact density and (b) a fitted density.

(a)					
Level	HF	VWN	BP86	MP2	QCISD
HF	0.0000				
VWN	0.1299	0.0000			
BP86	0.0600	0.1129	0.0000		
MP2	0.0425	0.1164	0.0267	0.0000	
QCISD	0.0366	0.1175	0.0300	0.0084	0.0000
(b)					
Level	HF	VWN	BP86	MP2	QCISD
HF	0.0000				
VWN	0.1269	0.0000			
BP86	0.0539	0.1128	0.0000		
MP2	0.0352	0.1161	0.0261	0.0000	
QCISD	0.0297	0.1170	0.0286	0.0063	0.0000

standard¹⁷ Hartree–Fock (HF), second order Møller–Plesset (MP2), singles and doubles quadratic configuration interaction (QCISD),¹⁸ and density functional (DFT) calculations.³ Open-shell systems were studied within the unrestricted formalism. In DFT calculations, two different levels of theory were employed: the local (LDA) and nonlocal (NLDA) approximations. At the LDA level, the parametrization due to Vosko–Wilk–Nusair (VWN)¹⁹ was used. The more sophisticated NLDA level includes the Perdew’s²⁰ nonlocal correlation correction and the Becke’s²¹ nonlocal exchange correction (BP86). The grid used for numerical integration in DFT calculations was the most dense allowed by the GAUSSIAN-92 program (keyword Int=FineGrid).

QMSM were obtained from the GAUSSIAN-92 electron densities using the MESSEM program.²² For MP2 and QCISD calculations, generalized densities²³ were used. Likewise, the DFT electron densities were calculated from SCF-converged Kohn–Sham orbitals. All QMSM were overlaplike and were obtained through use of Eq. (3). In a previous study,²⁴ we showed that overlap measures are more scattered over a range of values larger than repulsion similarities are, and consequently they are more suitable to quantify small changes in electron density distributions. However, the process of maximizing the similarity was carried out using repulsionlike similarity measures as defined by Eq. (4). The reason is due to the fact that the presence of the Coulomb operator smoothes the electron density surface and reduces the cusps of electron density at nuclei, making the process of optimization easier since gradient components are smaller.^{12(c)} An approximate density instead of the exact density was used in order to eliminate the need to evaluate costly four-index integrals as found in Eqs. (3) and (4). Details of this methodology have been given elsewhere.^{12(c),25} The fitting technique used in this work was the so-called PSA method.²⁵ In this particular case, the set of fitting functions was chosen to be the same that the squared molecular s -type renormalized primitive functions. The validity of such approximation can be assessed from the values of Table I,

TABLE II. Euclidean distance matrices (in a.u.) for the molecules studied ordered by increasing molecular weight. Geometrical parameters (Å and degrees) are also included for comparative purposes.

H ₂							
Level	HF	VWN	BP86	MP2	QCISD	R _{HH}	
HF	0.0000					0.735	
VWN	0.0217	0.0000				0.765	
BP86	0.0111	0.0156	0.0000			0.752	
MP2	0.0033	0.0220	0.0095	0.0000		0.738	
QCISD	0.0057	0.0207	0.0063	0.0033	0.0000	0.743	
EXP						0.741 ^a	
LiH							
Level	HF	VWN	BP86	MP2	QCISD	R _{LiH}	
HF	0.0000					1.608	
VWN	0.0376	0.0000				1.603	
BP86	0.0096	0.0368	0.0000			1.609	
MP2	0.0071	0.0382	0.0076	0.0000		1.599	
QCISD	0.0083	0.0382	0.0063	0.0026	0.0000	1.601	
EXP						1.595 ^a	
Li ₂							
Level	HF	VWN	BP86	MP2	QCISD	R _{LiLi}	
HF	0.0000					2.785	
VWN	0.6363	0.0000				2.700	
BP86	0.3590	0.2952	0.0000			2.738	
MP2	0.2862	0.3657	0.0741	0.0000		2.748	
QCISD	0.7001	0.0785	0.3552	0.4268	0.0000	2.692	
EXP						2.67 ^a	
Be ₂							
Level	HF	VWN	BP86	MP2	QCISD	R _{BeBe}	
HF	---					∞	
VWN	---	0.0000				2.381	
BP86	---	1.0927	0.0000			2.447	
MP2	---	4.2223	4.3340	0.0000		2.985	
QCISD	---	4.0596	4.0893	4.0980	0.0000	5.007	
EXP						2.465 ^b	
CH ₄							
Level	HF	VWN	BP86	MP2	QCISD	R _{CH}	
HF	0.0000					1.084	
VWN	0.0846	0.0000				1.097	
BP86	0.0349	0.0731	0.0000			1.099	
MP2	0.0193	0.0762	0.0191	0.0000		1.090	
QCISD	0.0190	0.0781	0.0192	0.0071	0.0000	1.093	
EXP						1.086 ^a	
NH ₃							
Level	HF	VWN	BP86	MP2	QCISD	R _{NH}	HNH
HF	0.0000					1.000	108.3
VWN	0.0919	0.0000				1.021	108.5
BP86	0.0539	0.0906	0.0000			1.024	107.1
MP2	0.0387	0.0891	0.0205	0.0000		1.014	107.3
QCISD	0.0445	0.0948	0.0215	0.0110	0.0000	1.015	107.0
EXP						1.012 ^a	106.0 ^a
H ₂ O							
Level	HF	VWN	BP86	MP2	QCISD	R _{OH}	HOH
HF	0.0000					0.941	106.3
VWN	0.0994	0.0000				0.970	105.1
BP86	0.0544	0.0899	0.0000			0.971	104.2
MP2	0.0420	0.0907	0.0224	0.0000		0.960	103.5
QCISD	0.0392	0.0928	0.0249	0.0063	0.0000	0.959	103.7
EXP						0.959 ^a	103.9 ^a

TABLE II. (Continued.)

HF							
Level	HF	VWN	BP86	MP2	QCISD	R_{HF}	
HF	0.0000					0.897	
VWN	0.1076	0.0000				0.930	
BP86	0.0505	0.0956	0.0000			0.931	
MP2	0.0326	0.0970	0.0230	0.0000		0.916	
QCISD	0.0278	0.0984	0.0265	0.0055	0.0000	0.915	
EXP						0.917 ^a	
LiF							
Level	HF	VWN	BP86	MP2	QCISD	R_{LiF}	
HF	0.0000					1.576	
VWN	0.1263	0.0000				1.570	
BP86	0.2711	0.3316	0.0000			1.599	
MP2	0.2715	0.3345	0.0215	0.0000		1.599	
QCISD	0.2213	0.2893	0.0543	0.0513	0.0000	1.595	
EXP						1.564 ^a	
N ₂							
Level	HF	VWN	BP86	MP2	QCISD	R_{NN}	
HF	0.0000					1.071	
VWN	2.0591	0.0000				1.099	
BP86	2.7158	0.6819	0.0000			1.108	
MP2	3.5765	1.5990	0.9408	0.0000		1.120	
QCISD	2.4805	0.4422	0.2488	1.1846	0.0000	1.104	
EXP						1.098 ^a	
CO							
Level	HF	VWN	BP86	MP2	QCISD	R_{CO}	μ_{CO}
HF	0.0000					1.105	-0.188
VWN	1.5674	0.0000				1.129	0.198
BP86	2.2257	0.6867	0.0000			1.140	0.153
MP2	2.2341	0.6970	0.0259	0.0000		1.140	0.283
QCISD	1.8555	0.3125	0.3914	0.4008	0.0000	1.134	0.082
EXP						1.128 ^a	0.112 ^{a,c}
HCN							
Level	HF	VWN	BP86	MP2	QCISD	R_{HC}	R_{CN}
HF	0.0000					1.058	1.127
VWN	1.4821	0.0000				1.078	1.153
BP86	2.0020	0.5388	0.0000			1.076	1.162
MP2	2.5410	1.1025	0.5797	0.0000		1.068	1.171
QCISD	1.8845	0.4217	0.1241	0.6999	0.0000	1.070	1.160
EXP						1.065 ^a	1.153 ^a
C ₂ H ₂							
Level	HF	VWN	BP86	MP2	QCISD	R_{HC}	R_{CC}
HF	0.0000					1.056	1.183
VWN	0.9574	0.0000				1.073	1.203
BP86	1.3237	0.3814	0.0000			1.071	1.210
MP2	1.6044	0.6639	0.2928	0.0000		1.065	1.216
QCISD	1.3324	0.3922	0.0257	0.2813	0.0000	1.066	1.211
EXP						1.061 ^a	1.203 ^a
NO							
Level	HF	VWN	BP86	MP2	QCISD	R_{NO}	μ_{NO}
HF	0.0000					1.118	-0.196
VWN	2.5712	0.0000				1.148	0.227
BP86	3.6746	1.1700	0.0000			1.162	0.171
MP2	1.4493	1.1717	2.3195	0.0000		1.135	-0.259
QCISD	3.4663	0.9475	0.2287	2.0992	0.0000	1.159	0.105
EXP						1.151 ^a	0.153 ^{a,c}

TABLE II. (Continued.)

O ₂									
Level	HF	VWN	BP86	MP2	QCISD	R _{OO}			
HF	0.0000					1.158			
VWN	4.5999	0.0000				1.204			
BP86	6.3230	1.9721	0.0000			1.223			
MP2	6.3894	2.0547	0.0906	0.0000		1.224			
QCISD	4.5750	0.1394	2.0390	2.1217	0.0000	1.203			
EXP						1.207 ^a			
CH ₂ O									
Level	HF	VWN	BP86	MP2	QCISD	R _{HC}	R _{CO}	HCH	
HF	0.0000					1.094	1.180	116.1	
VWN	1.2834	0.0000				1.121	1.199	116.0	
BP86	2.0541	0.7970	0.0000			1.120	1.212	115.8	
MP2	2.1353	0.8830	0.0932	0.0000		1.105	1.213	116.1	
QCISD	1.8312	0.5712	0.2352	0.3217	0.0000	1.107	1.208	116.0	
EXP						1.116 ^a	1.208 ^a	116.5 ^a	
F ₂									
Level	HF	VWN	BP86	MP2	QCISD	R _{FF}			
HF	0.0000					1.329			
VWN	8.8095	0.0000				1.397			
BP86	12.1487	4.7472	0.0000			1.432			
MP2	10.8068	2.7253	2.1443	0.0000		1.417			
QCISD	10.9238	2.8976	1.9681	0.1799	0.0000	1.418			
EXP						1.417 ^a			
HOOH									
Level	HF	VWN	BP86	MP2	QCISD	R _{HO}	R _{OO}	HOO	HOOH
HF	0.0000					0.943	1.385	102.9	117.4
VWN	4.9949	0.0000				0.977	1.435	100.7	119.7
BP86	8.4992	4.2768	0.0000			0.977	1.477	99.6	120.8
MP2	6.3686	1.5923	2.8116	0.0000		0.965	1.450	99.6	121.5
QCISD	5.8913	1.0208	3.3638	0.5843	0.0000	0.963	1.444	100.3	120.1
EXP						0.950 ^a	1.475 ^a	94.8 ^a	120.0 ^a
<i>cis</i> -FNNF									
Level	HF	VWN	BP86	MP2	QCISD	R _{FN}	R _{NN}	FNN	
HF	0.0000					1.326	1.191	114.6	
VWN	11.7875	0.0000				1.379	1.210	115.1	
BP86	17.9291	11.0480	0.0000			1.421	1.219	116.0	
MP2	12.6845	2.0010	10.6231	0.0000		1.383	1.232	114.6	
QCISD	11.0327	2.4617	12.2111	2.5936	0.0000	1.378	1.224	114.5	
EXP						1.384 ^d	1.214 ^d	114.5 ^d	
<i>trans</i> -FNNF									
Level	HF	VWN	BP86	MP2	QCISD	R _{FN}	R _{NN}	FNN	
HF	0.0000					1.326	1.188	107.5	
VWN	9.0325	0.0000				1.365	1.221	105.8	
BP86	15.4453	9.4255	0.0000			1.409	1.235	104.8	
MP2	11.5801	3.3504	7.0027	0.0000		1.376	1.240	105.2	
QCISD	10.8372	2.3374	7.6613	1.1711	0.0000	1.374	1.229	105.5	
EXP						1.396 ^d	1.230 ^d	105.5 ^d	
FOOF									
Level	HF	VWN	BP86	MP2	QCISD	R _{FO}	R _{OO}	FOO	FOOF
HF	0.0000					1.353	1.229	106.5	85.3
VWN	23.3886	0.0000				1.591	1.180	111.3	88.7
BP86	23.2635	13.6358	0.0000			1.643	1.195	111.9	89.7
MP2	25.2165	21.9137	22.1303	0.0000		1.856	1.130	114.8	90.2
QCISD	19.4891	19.9611	21.6842	24.4330	0.0000	1.483	1.273	107.7	86.8
EXP						1.575 ^d	1.217 ^d	109.5 ^d	87.5 ^d

^aData from Ref. 6(a).^bData from Ref. 30(a).^cNegative values mean that the dipole vector points away from the oxygen atom.^dData from Ref. 29.

TABLE III. Average Euclidean distance matrices (in a.u.) for the three groups of molecules considered: (a) molecules well described at the HF level; (b) intermediate molecules; and (c) molecules with large correlation effects (FOOF not included).

(a)					
Level	HF	VWN	BP86	MP2	QCISD
HF	0.0000				
VWN	0.1507	0.0000			
BP86	0.1056	0.1285	0.0000		
MP2	0.0876	0.1392	0.0247	0.0000	
QCISD	0.1332	0.0988	0.0643	0.0642	0.0000
(b)					
Level	HF	VWN	BP86	MP2	QCISD
HF	0.0000				
VWN	1.6534	0.0000			
BP86	2.3326	0.7093	0.0000		
MP2	2.2568	1.0195	0.7087	0.0000	
QCISD	2.1584	0.5146	0.2090	0.8313	0.0000
(c)					
Level	HF	VWN	BP86	MP2	QCISD
HF	0.0000				
VWN	7.8449	0.0000			
BP86	12.0691	6.2939	0.0000		
MP2	9.5659	2.3447	4.5345	0.0000	
QCISD	8.6520	1.7714	5.4487	1.3301	0.0000

where euclidean distances among Hartree–Fock, VWN, BP86, MP2, and QCISD methods in formaldehyde were computed using exact and fitted densities at the Hartree–Fock optimized geometry. From the values of this table, it can be seen that small differences (usually ranging between 10^{-2} and 10^{-3} a.u.) appear when the exact density is substituted by a fitted density, thus supporting the accuracy of this procedure. Remarkably, the same ordering found with exact QMSM is obtained using the QMSM computed with a fitted density, even when relative differences between distances are quite small.

Bader topological analyses²⁶ were performed through use of ELECTRA program.²⁷ All calculations were run on IBM RISC/6000 350 workstations.

III. RESULTS AND DISCUSSION

We shall begin our discussion by considering the euclidean distance matrices corresponding to the 21 investigated molecules. After that, a series of three diatomic molecules with different bonding character is carefully examined: We discuss the nuclear and electronic relaxation effects on euclidean distances and contours of electron density differences are depicted, and then, Bader analyses of the electron density of these three molecules at the different levels of theory considered are carried out.

A. Euclidean distance matrices

The 21 studied molecules are atomic combinations of the first and second row atoms of the Periodic Table. For each

molecule, the Euclidean distance matrix, together with the computed and experimental geometric parameters are collected in Table II. For the NO and CO molecules, Table II includes the computed and experimental dipole moment.

As commented in the Introduction, the distance between two methodologies over the same molecule is directly related to the dissimilarity between the electronic distributions computed with the two compared methodologies: the larger the distance, the larger the difference in these two electron densities. Therefore, the values of the distance yield a quantitative measure of how similar are two methodologies applied to the molecule under study. Thus it is possible to compare different methodologies, which is the main purpose of the present paper. Unless otherwise noted, here the QCISD generalized density is taken as the reference density, since it is expected to be the closest to the exact density distribution. Therefore, the larger the distance to the QCISD method resulting density, the less accurate can be considered the electron density computed with the method under study.

It is found that, in general, when a system is correctly described at the Hartree–Fock level, the MP2 generalized density is closer to the QCISD density than the density obtained from the DFT methodologies. Molecules like H₂, LiH, CH₄, NH₃, H₂O, and HF can be included in this group. For these systems, the MP2 method seems to be more appropriate than the two tested DFT methodologies. Furthermore, it is also found that MP2 and QCISD geometrical parameters are closer to the experimental values than DFT are. Since a

TABLE IV. Euclidean distance matrices (in a.u.) for the LiF, N₂, and CO molecules computed at the experimental geometry for the different methodologies analyzed.

LiF					
Level	HF	VWN	BP86	MP2	QCISD
HF	0.0000				
VWN	0.1136	0.0000			
BP86	0.0491	0.1019	0.0000		
MP2	0.0349	0.1030	0.0207	0.0000	
QCISD	0.0288	0.1041	0.0245	0.0055	0.0000
N ₂					
Level	HF	VWN	BP86	MP2	QCISD
HF	0.0000				
VWN	0.1251	0.0000			
BP86	0.0510	0.1115	0.0000		
MP2	0.0385	0.1116	0.0205	0.0000	
QCISD	0.0302	0.1134	0.0255	0.0084	0.0000
CO					
Level	HF	VWN	BP86	MP2	QCISD
HF	0.0000				
VWN	0.1245	0.0000			
BP86	0.0525	0.1117	0.0000		
MP2	0.0400	0.1134	0.0232	0.0000	
QCISD	0.0318	0.1138	0.0263	0.0095	0.0000

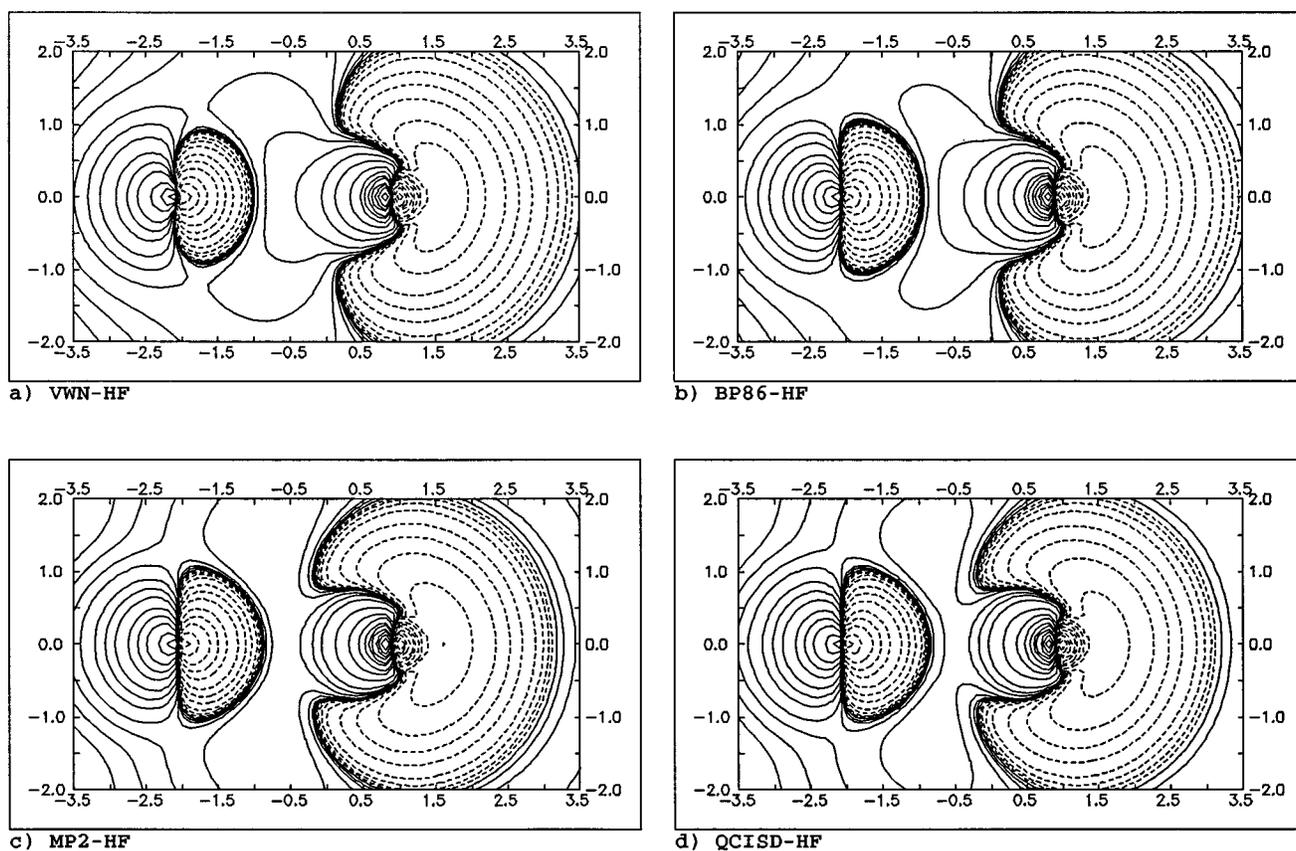


FIG. 1. Plots of $6\text{-}311++\text{G}^{**}$ electron density differences comparing densities obtained from the Hartree–Fock methodology with those computed at the MP2, QCISD, VWN, and BP86 levels, for the LiF molecule at its experimental geometry. In this map the lithium nucleus is on the left. The minimum contour is 1×10^{-4} a.u. and they increase to 2, 4, 8, 20, 40, 80, $\dots \times 10^{-4}$ a.u. Dashed lines correspond to negative values, that is, points where Hartree–Fock density is larger.

number of organic systems are well described at the Hartree–Fock level, it is reasonable to conclude that MP2 is superior to DFT in these cases.

On the other hand, in most systems where Hartree–Fock fails (large dissimilarity between Hartree–Fock and QCISD), the VWN and BP86 densities are closer to QCISD than MP2 is. This occurs basically in the FOOF molecule, where correlation energy is of utmost importance.^{28,29} To a lesser extent it also happens in N_2 , CO, HCN, NO, and O_2 molecules. Thus, when correlation energy becomes essential and the Hartree–Fock density is defective, one gets in most molecules better densities from DFT methodologies than from MP2. In these systems, DFT geometrical parameters are closer to experimental values and QCISD geometries than MP2 geometries are. Unfortunately, this result is not general for all systems. For instance, the *cis* and *trans*-FNNF molecules have both inaccurate Hartree–Fock and BP86 densities despite correlation energy is imperative.²⁹ It is worth mentioning that, in these cases, both MP2 and VWN densities are quite close to the QCISD density. Thus, for these two molecules BP86 nonlocal corrections to the LDA density are quite unsatisfactory.

A particular case is the beryllium dimer. In this case, Hartree–Fock fails to locate a potential energy minimum.^{30(a)}

Correlation energy is thus fundamental because dispersion effects are predominant in this van der Waals complex. Interestingly, QCISD method also fails to provide the correct bond length. As usual, QCISD gives intermediate values between Hartree–Fock and MP2 methodologies,^{4(b),24} because of the common MP2 overestimation of correlation effects. In this case, however, and due to a cancellation of errors, the MP2 bond length is closer to the experimental value than the QCISD bond length. Remarkably, DFT describes quite well the molecular structure of this dimer. It is not at all surprising that a molecule incorrectly described at the Hartree–Fock level becomes better represented by the DFT methodologies than by the post-Hartree–Fock methodologies. In fact, when Hartree–Fock fails, most methods based on the Hartree–Fock wave function as a reference can also be quite inappropriate, because the reference wave function is already incorrect, and in these cases the DFT formalism can be the most successful.⁹

In all cases, the distance between VWN and BP86 densities is smaller than 2.0 a.u., except for F_2 , HOOH, FOOF, and *cis* and *trans*-FNNF molecules for which nonlocal corrections become significant as judged for the large distance between VWN and BP86 methods.

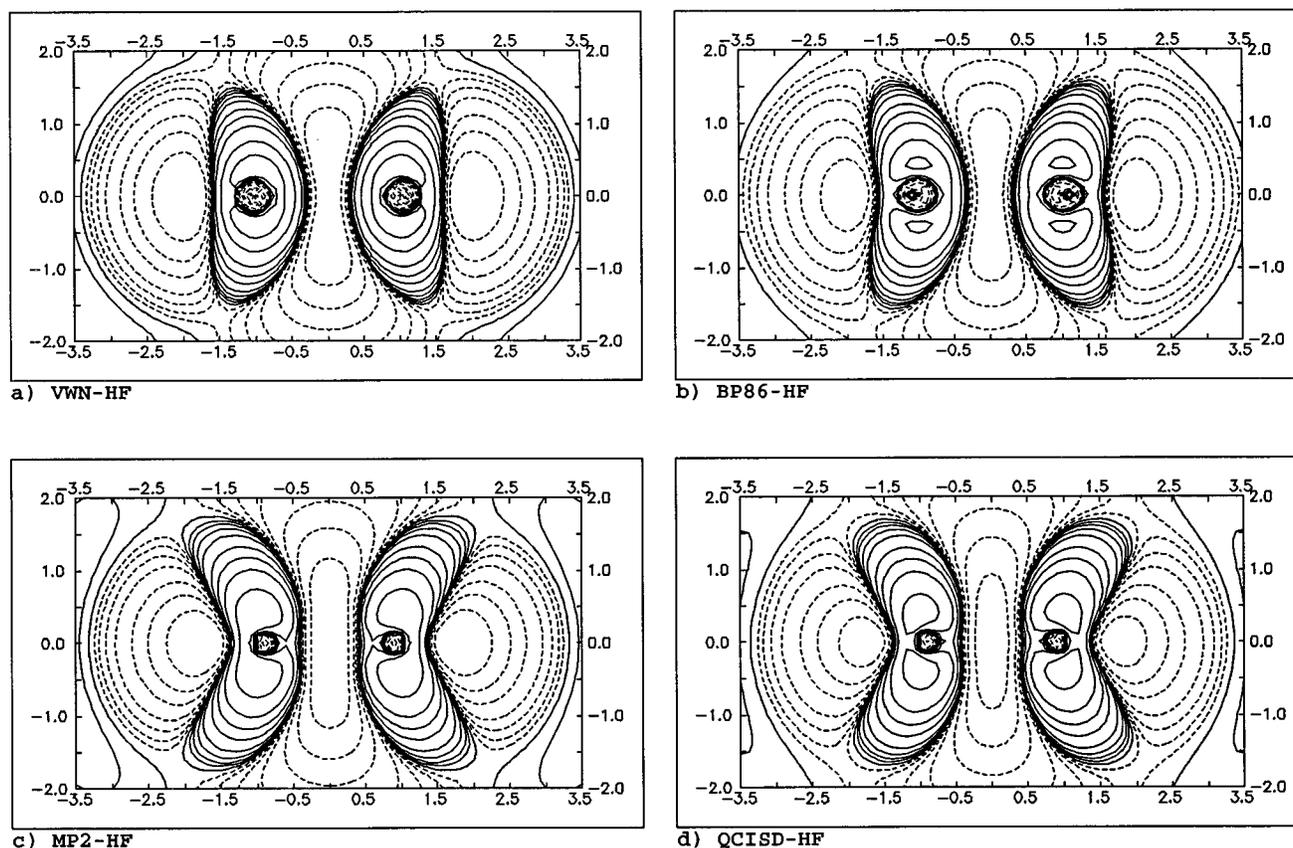


FIG. 2. Plots of $6\text{-}311++\text{G}^{**}$ electron density differences comparing densities obtained from the Hartree–Fock methodology with those computed at the MP2, QCISD, VWN, and BP86 levels, for the N_2 molecule at its experimental geometry. The minimum contour is 1×10^{-4} a.u. and they increase to 2, 4, 8, 20, 40, 80, $\dots \times 10^{-4}$ a.u. Dashed lines correspond to negative values, that is, points where Hartree–Fock density is larger.

The correct sign for small molecular dipole moments, like those of CO and NO molecules, is difficult to predict. It has been shown previously that DFT is successful in computing the proper dipole direction of these two molecules.^{4(b),6,30} From values of Table II, it is found the well-known result that in these two molecules Hartree–Fock yields the erroneous direction for the dipole moment. For the NO molecule, the MP2 method also fails to provide the correct sign to the dipole moment. Conversely, for the CO molecule, MP2 gives the correct direction but clearly exaggerates the dipole moment. In both systems, QCISD produces a value close to the experimental one, showing that this method provides a reliable density distribution for these two molecules. Interestingly, DFT methods produce dipole moments which are better than the MP2 ones, and in the case of the BP86 method they are even as good as those yield by the QCISD procedure.

As an attempt to classify the set of molecules under study depending on the effect of correlation corrections on their density distributions, a principal components analysis (PCA)³¹ has been carried out. This is a well-known technique widely used in pattern recognition and multivariate analysis to reduce the dimensionality of a data set by transforming it into a sort of feature space which reveals the in-

ternal data structure. In this case, the data matrix to correlate is taken as composed by the complete set of molecules (except the beryllium dimer which has not been included in the analysis because distances from Hartree–Fock to the rest of the methodologies are not available) each one described by a row vector, whose elements are the ten different euclidean distances of Table II between the different methodologies in a given molecule. It is found that, qualitatively, it is possible to separate the molecules studied in four main groups: In the first one, we can include molecules which are well described at the Hartree–Fock level, like H_2 , LiH, Li_2 , CH_4 , NH_3 , H_2O , HF, and LiF; the second group contains molecules in which correlation starts to have a rather remarkable influence, like N_2 , CO, HCN, C_2H_2 , NO, and CH_2O ; and finally, to the third group belong those molecules in which correlation energy is compulsory, like O_2 , F_2 , HOOH, *cis* and *trans*-FNNF. Finally, group four contains only the FOOF molecule, which must be considered separated from the rest. Not surprisingly, correlation energy becomes more important with the increase in the number of electrons of the molecule. Quantitative ordering of the studied molecules with respect to the H_2 molecule can be achieved from the multidimensional distance of the principal components space

TABLE V. Bader analyses for the LiF, N₂, and CO molecules at the optimized geometry corresponding to each level studied. Distances to the bond critical point (r_c) referred to Li atom in LiF and to C atom in CO.

Molecule	Method	r_c^a	$\rho(r_c)^b$	$\nabla^2\rho(r_c)^b$	$\lambda_1=\lambda_2^b$	λ_3^b	$ \lambda_1/\lambda_3$
LiF	HF	1.1521	0.0701	0.7213	-0.1615	1.0442	0.1546
	VWN	1.1545	0.0740	0.6999	-0.1626	1.0252	0.1587
	BP86	1.1710	0.0670	0.6334	-0.1438	0.9210	0.1561
	MP2	1.1723	0.0657	0.6424	-0.1426	0.9276	0.1537
	QCISD	1.1687	0.0664	0.6561	-0.1457	0.9475	0.1538
N ₂	HF	1.0115	0.7322	-3.1319	-1.8258	0.5197	3.5130
	VWN	1.0380	0.6732	-2.4649	-1.6946	0.9243	1.8335
	BP86	1.0466	0.6591	-2.3600	-1.6358	0.9116	1.7944
	MP2	1.0585	0.6370	-2.1760	-1.5436	0.9112	1.6940
	QCISD	1.0435	0.6646	-2.4398	-1.6231	0.8064	2.0128
CO	HF	0.7039	0.5173	1.0765	-1.8935	4.8635	0.3893
	VWN	0.7348	0.4911	0.4406	-1.6030	3.6466	0.4396
	BP86	0.7352	0.4759	0.4374	-1.5377	3.5127	0.4377
	MP2	0.7275	0.4663	0.7623	-1.5034	3.7690	0.3989
	QCISD	0.7235	0.4762	0.7730	-1.5769	3.9267	0.4016

^aIn Å.

^bIn a.u.

QCISD. The large distance from the BP86 to the QCISD method in the third group is mainly due to the contributions of the *cis* and *trans*-FNNF molecules.

B. Analysis of the LiF, N₂, and CO molecules

To get more insight into the nature of differences in density distributions obtained from the different methodologies analyzed, we have performed a more accurate analysis of the LiF, N₂, and CO electron densities. We have selected these molecules because, from a Bader's atoms-in-molecules analysis,²⁶ they exhibit different bonding nature: LiF, is a typical case of closed-shell ionic interaction; N₂, is an appropriate example of a molecule with a shared interaction; and finally, CO is a well-known case of an intermediate interaction. Analyses of charge densities in these molecules can put forward the behaviour of the different methodologies in such quite different types of bonding.

Table IV collects the euclidean distance matrices obtained for LiF, N₂, and CO computed at the experimental geometry of these molecules for all levels of theory considered. Since in this case the geometry is kept fixed, it is possible to perform an analysis by means of density difference maps. Figures 1, 2, and 3 show this kind of analysis for LiF, N₂, and CO, respectively. The analysis performed here accounts only for the electronic relaxation, whereas the study carried out in the last section, using the optimized geometries, included electronic and nuclear relaxation.

For LiF, it is found that when taking into account only electronic relaxation [Table IV(a)] a large difference appears between Hartree–Fock and the VWN method. As expected, if both electronic and nuclear relaxation are allowed (Table II) all distances increase. The small difference between Hartree–Fock and correlated methods corresponds now to the VWN density, thus showing that conclusions derived from comparisons between density distributions at a fixed geometry can be quite different from those calculated at the

optimized geometry. Using the experimental geometry, the minimum distance is obtained when comparing MP2 and BP86 methodologies, whereas at the optimized geometry the MP2 density is the density distribution closest to QCISD. In both cases, it is found that the VWN method is that having the larger distances to QCISD, MP2, and BP86 methods. Figure 1 depicts electron density difference contours, in a plane that contains the molecular axis, between VWN, BP86, MP2, and QCISD densities and the Hartree–Fock density, at the Hartree–Fock optimized geometry in all cases. The four maps are quite similar, corroborating again that DFT densities have the common features of *ab initio* correlated densities. The main effect of correlation in this molecule is to push the density out from the region of the fluorine nonshared electron pairs and to accumulate it in the bonding region and in the lithium atom. Clearly, the overall effect of correlation is a reduction in the ionicity of the bond,^{4(a)} i.e., the Hartree–Fock overestimation of the ionic (closed-shell) character of the Li–F bond is reduced by correlated methods.

As far as N₂ is concerned, here too we find that when only electronic relaxation is allowed [Table IV(b)] the large difference in electron densities corresponds to the Hartree–Fock and VWN methods. Despite in all cases distances are quite small, it must be pointed out that VWN gives the large distance to any of the methods analyzed. The reason must be found in the well-known fact that the VWN electron density is too diffuse in the region near the nuclei.^{4(b),24} Interestingly, when the N–N bond length is relaxed, the method which yields larger distances to the other methodologies tested is the Hartree–Fock method. In this case, the VWN produces a density quite close to the QCISD density distribution, only improved by the BP86 methodology. Not only the density, but also the VWN bond length is the closest to the experimental value. On the other hand, despite the MP2 density being initially the nearest to QCISD, when nuclear relaxation is allowed, it becomes among the correlated densities the further away from QCISD. Again, we confirm that conclusions from density analyses at fixed geometry cannot be extrapolated to optimized systems. One can say that large density differences at a fixed geometry do not always imply large structural and density differences in the optimized structures. For this reason, an analysis of density differences at a fixed geometry may provide misleading conclusions.

Electron density difference maps, in a plane that contains the molecular axis, between VWN, BP86, MP2, and QCISD densities and the Hartree–Fock density for N₂ molecule are drawn in Fig. 2. The electron density difference map between Hartree–Fock and VWN reveals the aforementioned VWN reduction in electron density near nuclei. The general picture of all four graphs are quite similar. When correlation is included, there is a decrease of electron density near nuclei, and also in the region of the nitrogen nonshared electron pairs, together with a further diminution in the center of the molecule. The density moves to shells around atomic nuclei. As above, the overall effect of correlation is to increase of the covalent features of the molecule. Interestingly, comparing VWN with the QCISD density difference contours one finds that VWN overestimates the covalent character of N₂.

In this sense, and as found earlier,^{4,24} nonlocal corrections tend to remove this overestimation, thus leading to a density closest to the QCISD density distribution.

The euclidean distance matrices of CO at the fixed experimental geometry and at the optimized geometry for each method follow almost the same trend than those of N₂. For this reason, conclusions obtained with the N₂ molecules apply also for CO. More interesting are the electron density difference maps computed in a plane that contains the molecular axis (Fig. 3). In this case, the behavior of CO is intermediate between those of LiF and N₂. The charge redistribution due to correlation is now more complicated, although as a whole there is a charge transfer from the oxygen atom to the carbon atom, thus resulting in a reduction of the C–O bond ionicity. This conclusion is reinforced by a Mulliken population analysis which gives charges on C atom of 0.109, –0.031, –0.031, –0.024, and 0.015 a.u. for Hartree–Fock, VWN, BP86, MP2, and QCISD methodologies, respectively. As before, VWN and BP86 exaggerate the shift of electron density. A similar conclusion was reached earlier by Wang *et al.*^{4(b)} who studied electron densities in carbon monoxide at its equilibrium bond length, calculated from conventional *ab initio* and density functional methods. These authors also conclude that BP86 partially corrects the VWN density towards the QCISD result.

Finally, in Table V we present an analysis of the density distributions obtained from the different methodologies studied from a Bader's theory point of view.²⁶ The values of the ratio between the perpendicular and the parallel curvatures at the bond critical point ($|\lambda_1/\lambda_3|$) confirm the aforementioned closed-shell (ionic system) interaction for LiF ($|\lambda_1/\lambda_3| < 1$, $\rho(\mathbf{r}_c)$ small, and $\nabla^2\rho > 0$), shared interaction (typically between covalently bound atoms) for N₂ ($|\lambda_1/\lambda_3| > 1$, $\rho(\mathbf{r}_c)$ large, and $\nabla^2\rho < 0$), and intermediate interaction [$\rho(\mathbf{r}_c)$ large but $\nabla^2\rho > 0$] for CO, irrespective of the methodology employed.

The tendency followed by the Hartree–Fock bond lengths, which are in most cases shorter than the correlated bond lengths, is reproduced by distances to the bond critical point. It is found that, when correlation is included, distances to the bond critical points are larger. Furthermore, due to the fact that Hartree–Fock bond lengths are shorter, the density in the bond critical point is usually larger for the Hartree–Fock method as compared to correlated methodologies, except for the VWN method in the LiF molecule, although in this case the VWN Li–F bond length is shorter than the Hartree–Fock bond length. In these three molecules the gradient-corrected BP86 method adjusts the VWN electron density towards the QCISD results. Both charge depletion ($\nabla^2\rho > 0$) and charge concentration ($\nabla^2\rho < 0$) are clearly exaggerated at the Hartree–Fock level, probably as a consequence of the shorter Hartree–Fock bond lengths. On the other hand, DFT and MP2 Laplacian values at the bond critical point are close to the QCISD values, but the DFT Laplacian values of the CO molecule, which are strikingly too low. Nevertheless, this molecule has a very special electron density distribution, with a remarkable charge transfer from C to O, and a large opposing polarization due to the positive

charge on C. These two effects counteract, leading to small dipole moments and a complicated density distribution. Therefore, it is not reasonable to infer from this result that DFT Laplacian values are defective.

IV. CONCLUSIONS

It has been shown that distances obtained from quantum molecular similarity measures can be a useful tool to analyze electron density distribution differences among a series of methodologies. In this way the analysis can be carried out at the optimized geometry corresponding to each methodology. It has been demonstrated that the use of electron density difference contours, which is undeniably practical to illustrate differences at a fixed geometry, can lead to conclusions that are not longer valid at the optimized geometries. It has been concluded that large density differences at a fixed geometry do not always imply large structural and electronic differences at the optimized geometry. Further, it must be pointed out that conclusions derived from the study of a system in a fixed geometry are dependent on the selected geometrical parameters. For instance, one expects that Hartree–Fock density will worsen with respect to correlated densities with the increase in the equilibrium bond length, because of the well-known restricted Hartree–Fock problems to describe dissociation processes correctly.

The analysis performed in this work has shown that, if the single determinant Hartree–Fock wave function is adequate to describe the molecule, electron densities obtained from MP2 are usually superior to those computed with DFT methodologies. On the other hand, when correlation is important, we have concluded that density distributions from DFT are of the same quality, if not better, than MP2 densities. Finally, we have shown that both conventional post-Hartree–Fock and DFT methods correct the overestimated ionicity present in Hartree–Fock electron densities.

ACKNOWLEDGMENT

This work has been founded through the Spanish DGI-CYT Project No. PB92-0333.

¹(a) P.-O. Löwdin, *Phys. Rev.* **97**, 1474 (1955); (b) R. McWeeny, *Proc. R. Soc. London, Ser. A* **232**, 114 (1955); (c) **235**, 496 (1956); (d) **253**, 242 (1959).

²P. Hohenberg and W. Kohn, *Phys. Rev. B* **136**, 864 (1964).

³(a) R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University, New York, 1989); (b) T. Ziegler, *Chem. Rev.* **91**, 651 (1991).

⁴(a) J. Wang, L. A. Eriksson, R. J. Boyd, Z. Shi, and B. G. Johnson, *J. Phys. Chem.* **98**, 1844 (1994); (b) J. Wang, Z. Shi, R. J. Boyd, and C. A. Gonzalez, *ibid.* **98**, 6988 (1994); (c) M. Solà, J. Mestres, R. Carbó, and M. Duran, *QSAR and Molecular Modelling: Concepts, Computational Tools, and Biological Applications* (Prous Science, Barcelona, 1995).

⁵(a) E. Ruiz, D. R. Salahub, and A. Vera, *J. Am. Chem. Soc.* **117**, 1141 (1995); (b) K. Kim and K. D. Jordan, *J. Chem. Phys.* **98**, 10089 (1994).

⁶(a) B. G. Johnson, P. M. W. Gill, and J. A. Pople, *J. Chem. Phys.* **98**, 5612 (1993); (b) C. W. Murray, G. J. Laming, N. C. Handy, R. D. Amos, *Chem. Phys. Lett.* **199**, 551 (1992).

⁷J. Andzlem and E. Wimmer, *Physica B* **172**, 307 (1991).

⁸(a) R. V. Stanton and K. M. Merz, Jr., *J. Chem. Phys.* **100**, 434 (1994); (b) R. van Leeuwen and E. J. Baerends, *Int. J. Quantum Chem.* **52**, 711

- (1994); (c) A. D. Becke, *J. Chem. Phys.* **84**, 4524 (1986); (d) J. Andzlem and E. Wimmer, *ibid.* **96**, 1280 (1992); (e) A. M. Schmiedekamp, I. A. Topol, S. K. Burt, H. Razafinjanahary, H. Chermitte, T. Pfaltzgraff, and C. J. Michejda, *J. Comput. Chem.* **15**, 875 (1994).
- ⁹M. Torrent, M. Solà, and M. Duran (unpublished).
- ¹⁰(a) R. Carbó, M. Arnau, and L. Leyda, *Int. J. Quantum Chem.* **17**, 1185 (1980); (b) R. Carbó and B. Calabuig, *ibid.* **42**, 1681 (1992).
- ¹¹(a) R. Carbó, B. Calabuig, L. Vera, and E. Besalú, *Adv. Quantum Chem.* **25**, 253 (1994); (b) E. Besalú, R. Carbó, J. Mestres, and M. Solà, *Molecular Similarity*, Topics in Current Chemistry Series (Springer, Berlin, 1995), Vol. 173, pp. 31–62; (c) E. Besalú and R. Carbó, *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches* (Kluwer, Dordrecht, 1995), pp. 3–30.
- ¹²(a) M. Solà, J. Mestres, M. Duran, and R. Carbó, *J. Chem. Inf. Comput. Sci.* **34**, 1047 (1994); (b) M. Solà, J. Mestres, M. Duran, and R. Carbó, *J. Am. Chem. Soc.* **116**, 5909 (1994); (c) J. Mestres, M. Solà, M. Duran, and R. Carbó, *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches* (Kluwer, Dordrecht, 1995), pp. 89–111; (d) J. Mestres, M. Solà, R. Carbó, F. J. Luque and M. Orozco, *J. Phys. Chem.* (in press).
- ¹³R. Carbó and Ll. Domingo, *Int. J. Quantum Chem.* **32**, 517 (1987).
- ¹⁴(a) R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, *J. Chem. Phys.* **72**, 650 (1980); (b) T. Clark, J. Chandrasekhar, G. W. Spitznagel, and P. v. R. Schleyer, *J. Comput. Chem.* **4**, 294 (1983); (c) M. J. Frisch, J. A. Pople, and J. S. Binkley, *J. Chem. Phys.* **80**, 3265 (1984).
- ¹⁵H. B. Schlegel, *J. Comput. Chem.* **3**, 214 (1982).
- ¹⁶M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogel, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople, GAUSSIAN 92-DFT, Revision G.1, (Gaussian Inc., Pittsburgh, 1992).
- ¹⁷W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Ab initio Molecular Orbital Theory* (Wiley, New York, 1986).
- ¹⁸J. A. Pople, M. Head-Gordon, and K. Raghavachari, *J. Chem. Phys.* **87**, 5968 (1987).
- ¹⁹S. H. Vosko, L. Wilk, and M. Nusair, *Canad. J. Phys.* **58**, 1200 (1980).
- ²⁰J. P. Perdew, *Phys. Rev. B* **33**, 8822 (1986); **34**, 7406E (1986).
- ²¹A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).
- ²²(a) J. Mestres, M. Solà, E. Besalú, M. Duran, and R. Carbó, MESSEM (Girona, CAT, 1993); (b) E. Besalú, R. Carbó, M. Duran, J. Mestres, and M. Solà, *Modern Techniques in Computational Chemistry: METTEC-95* (ESCOM Science, Leiden, 1995).
- ²³(a) N. C. Handy and H. F. Schaefer III, *J. Chem. Phys.* **81**, 5031 (1984); (b) K. B. Wiberg, C. M. Hadad, T. J. LePage, C. M. Breneman, and M. J. Frisch, *J. Phys. Chem.* **96**, 671 (1992).
- ²⁴M. Solà, J. Mestres, J. M. Oliva, M. Duran, and R. Carbó, *Int. J. Quantum Chem.* (in press).
- ²⁵J. Mestres, M. Solà, M. Duran, and R. Carbó, *J. Comput. Chem.* **15**, 1113 (1994).
- ²⁶(a) R. F. W. Bader, *Acc. Chem. Res.* **18**, 9 (1985); (b) R. F. W. Bader, *Atoms in Molecules: A Quantum Theory* (Clarendon, Oxford, 1990).
- ²⁷J. Mestres, ELECTRA (Girona, CAT, 1994).
- ²⁸(a) K. Raghavachari and G. W. Trucks, *Chem. Phys. Lett.* **162**, 511 (1989); (b) H.-G. Mack and H. Oberhammer, *ibid.* **145**, 121 (1988); (c) D. A. Dixon, J. Andzlem, G. Fitzgerald, and E. Wimmer, *J. Phys. Chem.* **95**, 9197 (1991).
- ²⁹T. J. Lee, J. E. Rice, G. E. Scuseria, and H. F. Schaefer III, *Theor. Chim. Acta* **76**, 81 (1989).
- ³⁰(a) R. O. Jones and O. Gunnarsson, *Rev. Mod. Phys.* **61**, 689 (1989); (b) E. J. Baerends, P. Vernooijs, A. Rozendaal, P. M. Boerigter, M. Krijn, D. Feil, and D. Sundholm, *J. Mol. Struct. (Theochem)* **133**, 147 (1985).
- ³¹S. Haykin, *Neural Networks* (MacMillan College, Don Mills, 1994).