

# A comparative study of isodensity surfaces using *ab initio* and ASA density functions

Xavier Gironés, Lluís Amat, and Ramon Carbó-Dorca

*Institute of Computational Chemistry, University of Girona, Girona, Catalonia, Spain*

*In this article, we report a visual comparison between several of the available methods for constructing electronic density functions. The density forms studied include *ab initio*, atomic shell approximation, and promolecular densities. A graphical comparison is made for six different molecules at different levels of density function values. The differences between the various density functions are analysed by considering a molecular quantum self-similarity measure and the required computational time for all molecules at all computation levels is considered. © 1999 by Elsevier Science Inc.*

**Keywords:** *ab initio densities, atomic shell approximation densities, marching cubes algorithm, promolecular densities, molecular quantum similarity measures.*

## INTRODUCTION

A significant amount of the research in our laboratory is related to molecular quantum similarity measures (MQSM),<sup>1–10</sup> which can be obtained using the integral shown in Eq. (1):

$$Z_{AB} = \iint \rho_A(\mathbf{r}_1) \Omega(\mathbf{r}_1, \mathbf{r}_2) \rho_B(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (1)$$

Where  $\rho_A$  and  $\rho_B$  are the corresponding density functions for two arbitrary molecules A and B and  $\Omega$  is a positive definite operator. In this way, an MQSM,  $Z_{AB}$ , is constructed from a scalar product of the pair of density functions weighted by  $\Omega$ . As the measure  $Z_{AB}$  depends on the relative positioning of both molecules in space, the optimal value of Eq. (1) will be derived from the optimal alignment of the molecules with each other.<sup>11</sup> This optimization procedure, which requires multiple evalua-

tions of the MQSM, and, consequently, the computation of many center integrals, can be very time consuming, depending on the method used to compute the density functions. In addition, it is important to note that the quality of the MQSM depends on the quality of the density functions used in the calculation.

## DENSITY FUNCTIONS FORMS

### *Ab initio* calculations

*Ab initio* calculations using the LCAO-MO approximation are widely used to evaluate density functions.<sup>12</sup> These density functions are formulated as shown in Eq. (2):

$$\rho_A(\mathbf{r}) = \sum_{\mu, \nu \in A} D_{\mu\nu} \chi_{\mu}^*(\mathbf{r}) \chi_{\nu}(\mathbf{r}) \quad (2)$$

where  $D_{\mu\nu}$  is the charge and bond order matrix and  $\{\chi_{\mu}\}$  is the atomic orbitals basis set.

The MQSMs of systems calculated by this procedure are as accurate as the level of theory used. However, this methodology has some disadvantages. Computational requirements are large owing to the amount of calculations needed to construct the density function. Furthermore, when an MQSM must be evaluated, four center integrals must be calculated, because Eq. (1), using expression (2) for  $\rho_A$  and  $\rho_B$ , adopts the form shown in Eq. (3):

$$Z_{AB} = \sum_{\mu \in A} \sum_{\nu \in A} \sum_{\lambda \in B} \sum_{\sigma \in B} D_{\mu\nu} D_{\lambda\sigma} \iint \chi_{\mu}^*(\mathbf{r}_1) \chi_{\nu}(\mathbf{r}_1) \Omega(\mathbf{r}_1, \mathbf{r}_2) \chi_{\lambda}^*(\mathbf{r}_2) \chi_{\sigma}(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (3)$$

As a result, Eq. (3) is computationally time consuming. Molecular systems composed of a large number of atoms are difficult to treat at this level owing to constraints of computer resources. As an MQSM needs to be optimized, the amount of time required to obtain a maximal value for an MQSM can be large, as  $Z_{AB}$  integrals must be repeatedly calculated.

Address reprint requests to: Ramon Carbó-Dorca, Institute of Computational Chemistry, University of Girona, Girona, Catalonia, 17071 Spain. Fax: 34-972-418356. E-mail: director@iqc.udg.es.

Received 24 October 1998; revised 21 January 1999; accepted 13 April 1999.

## Molecular atomic shell approximation

The molecular atomic shell approximation (ASA) approach<sup>13,14</sup> expresses molecular electron densities as linear combinations of 1S Gaussian functions. The resulting density function, constructed in this way, is as shown in Eq. (4):

$$\rho_A^{\text{ASA}}(\mathbf{r}) = \sum_{i \in A} w_i |s_i(\mathbf{r})|^2 \quad (4)$$

where  $s_i(\mathbf{r})$  are normalized 1S spherical functions, that is:  $\int |s_i(\mathbf{r})|^2 d\mathbf{r} = 1$ , and the coefficients  $w_i$  must fulfill the convex conditions shown in Eqs. (5) and (6):

$$w_i \in \mathbf{R} \wedge w_i > 0, \quad \forall i \in A \quad (5)$$

in order to guarantee that the density function is positive definite, and

$$\sum_{i \in A} w_i = 1 \quad (6)$$

to ensure the normalization of ASA density to unity. In this way the statistical probability distribution of expressions such as Eq. (4) is preserved.

When density functions are obtained for molecular systems in this way, then the calculation of MQSM in expression (3) transforms to the simplified form shown in Eq. (7):

$$Z_{AB} = \sum_{i \in A} \sum_{j \in B} w_i w_j \iint |s_i(\mathbf{r}_1)|^2 \Omega(\mathbf{r}_1, \mathbf{r}_2) |s_j(\mathbf{r}_2)|^2 d\mathbf{r}_1 d\mathbf{r}_2 \quad (7)$$

However, two disadvantages arise from this procedure. The first is that the ASA density function needs to be fitted, which requires the *ab initio* calculation to be computed beforehand, providing density functions as in Eq. (2). The other problem is that for each molecule, there are coefficients  $w_i$  that must be fitted to the molecular density (2), and this involves extra computer requirements and time spent. Table 1 lists the times taken to fit some ASA functions.

## Promolecular atomic shell approximation

The promolecular atomic shell approximation (PASA)<sup>14</sup> is based on the same ideas as the ASA methodology, and involves the construction of the density function as a sum of atomic

**Table 1. Time needed (in seconds) to fit ASA to *ab initio* functions obtained from a Gaussian HF/3-21G calculation<sup>a</sup>**

Molecule	Time <sup>b</sup>
Water	13
Allyl alcohol	1 478
Methanol	300
Ethanol	1 675
Benzene	4 815
Glycine	1 750

<sup>a</sup> Frisch *et al.*<sup>17</sup>

<sup>b</sup> Times calculated with a PC Intel Pentium 200 MHz, running under Windows 95. They should be taken as approximate, because Windows 95 runs self-programs in the background that can slow the calculation down.

density contributions, which are expressed in turn as linear combinations of 1S spherical functions as in Eq. (4). The resulting density function is written as in Eq. (8):

$$\rho_A^{\text{ASA}}(\mathbf{r}) = \frac{1}{P_A} \sum_{a \in A} P_a \rho_a^{\text{ASA}}(\mathbf{r}) \quad (8)$$

where  $P_a$  is the atomic number of each atom and  $P_A$  is the total number of electrons in the studied molecule.

In this way, molecular density is a simple sum of atomic densities,  $\rho_a^{\text{ASA}}$  centered at the atomic positions  $\mathbf{R}_a$ , which are defined as in Eq. (9):

$$\rho_a^{\text{ASA}}(\mathbf{r} - \mathbf{R}_a) = \sum_{i \in a} w_i |s_i(\mathbf{r} - \mathbf{R}_a)|^2 \quad (9)$$

with coefficients  $\{w_i\}$  fulfilling convex conditions (5) and (6). In this way, PASA functions automatically include the conditions expressed in Eqs. (5) and (6).

The PASA needs an ASA density function as in Eq. (9). Therefore, for each atom, *ab initio* calculations and fitting processes are used to parameterize 1S function exponents and  $w_i$  coefficients, which are stored for later calculations.<sup>15</sup>

The main advantage of this method is that calculations take little time to be completed. There is no need to calculate the molecular density function at *any* level because atomic density calculations are to be carried out only once. Within the PASA the molecular density function is constructed simply by means of atomic coordinates and ASA densities, and such coordinates can be extracted from any source, including crystallographic data or previous theoretical calculations.































## GRAPHICAL VISUAL COMPARISON BETWEEN DENSITIES

### Computational details

In Tables 2–7, several comparisons are presented between six different molecules calculated at six different computation levels. The choice of molecular systems studied is similar to that used by Mezey,<sup>16</sup> allowing comparison with the results of this earlier work. The molecules used are water, methanol, ethanol, allyl alcohol, benzene, and glycine. The calculation levels are (1) promolecular ASA at two different levels (simple and limit), (2) molecular ASA, and (3) HF/STO-3G, HF/3-21G, and HF/6-31++G\*\*. In all cases, gas-phase geometries are used arising from a full optimization at the HF/6-31++G\*\* level.<sup>17</sup>

A simple ASA calculation is a PASA calculation<sup>14</sup> that involves one Gaussian function for each H, and three Gaussian functions for C, O and N. A limit ASA calculation is again a PASA calculation; however, in this case the number of functions per atom has increased to seven Gaussian functions for H, eight for C, and nine for O and N. It has been found that further increase in the number of functions does not improve the description of the atomic density. In molecular ASA the density function is fitted with as many Gaussian functions as needed, to the LCAO form, as presented in Eq. (3), calculated from a previous computation performed using an *ab initio* procedure.<sup>17</sup> In this case this was chosen to be the HF/3-21G computation level.

**Table 2. Comparison of water molecule density function shapes**

Density Value	ASA			ab initio		
	Simple	Limit	Molecular	STO-3G	3-21G	6-31++G**
0.01						
0.15						
0.22						
0.35						
0.45						

### Image construction

In all approximations, the steps involved in the graphical representation have been chosen in the same way. First, a grid, large enough to envelop the molecule, is defined in

three-dimensional space. At each grid point, a density function value is computed and stored in a file. When all grid points have been calculated, the wireframe model is created, choosing first the isodensity value, which is then computed according to a marching cubes algorithm (see Reference 18

**Table 3. Comparison of methanol molecule density function shapes**





























































Density Value	ASA			ab initio		
	Simple	Limit	Molecular	STO-3G	3-21G	6-31++G**
0.01						
0.15						
0.22						
0.35						
0.45						

Table 4. Comparison of ethanol molecule density function shapes

Density Value	ASA			ab initio		
	Simple	Limit	Molecular	STO-3G	3-21G	6-31++G**
0.01						
0.15						
0.22						
0.35						
0.45						

and Appendix 1). Finally, the input file is rendered by the program GiD.<sup>19</sup>

## RESULTS AND DISCUSSION

As can be seen in Tables 2–7, the lowest and highest density function values correctly describe the surfaces in all cases.

For the middle density function values, bond formation and hydrogen position are not so accurately described with the ASA approaches, as the density function has collapsed over the atoms because only 1S Gaussian functions are used, and the hydrogen density is not well described. However, when working with quantum similarity measures, the major part of

Table 5. Comparison of allyl alcohol molecule density function shapes





























































Density Value	ASA			ab initio		
	Simple	Limit	Molecular	STO-3G	3-21G	6-31++G**
0.01						
0.15						
0.22						
0.35						
0.45						































Table 6. Comparison of benzene molecule density function shapes

Density Value	ASA			ab initio		
	Simple	Limit	Molecular	STO-3G	3-21G	6-31++G**
0.01						
0.15						
0.22						
0.35						
0.45						

the density function surrounds the heavy atoms. These inconveniences are not so important, and do not affect the final results in any significant way. This was demonstrated by performing a comparison employing overlap-like molec-

ular quantum self-similarity measures (MQS-SM),  $Z_{AA}$ . In Eq. (1) the operator  $\Omega(\mathbf{r}_1, \mathbf{r}_2)$  is chosen to be equal to a Dirac's delta function,  $\delta(\mathbf{r}_1 - \mathbf{r}_2)$ , for the six molecules considered. The results are presented in Table 8, and the

Table 7. Comparison of glycine molecule density function shapes

Density Value	ASA			ab initio		
	Simple	Limit	Molecular	STO-3G	3-21G	6-31++G**
0.01						
0.15						
0.22						
0.35						
0.45						



**Table 8. Comparison between overlap-like MQS-SM for the six considered molecules**

Molecule	ASA			<i>Ab initio</i>		
	Simple	Limit	Molecular	STO-3G	3-21G	6-31 <sup>++</sup> G**
Water	80.86	80.52	80.25	79.41	80.24	81.08
Methanol	112.56	112.25	107.89	107.02	108.03	107.74
Ethanol	144.32	144.00	141.05	139.66	141.18	142.24
Allyl Alcohol	175.94	175.60	172.47	170.10	171.61	176.28
Benzene	190.09	190.02	179.59	179.43	180.12	189.81
Glycine	277.16	276.53	273.33	266.53	272.53	274.30

**Table 9. Comparison between overlap-like MQS-SM calculation time<sup>a</sup> for the six considered molecules (in seconds)**

Computational level		Water	Methanol	Ethanol	Allyl alcohol	Benzene	Glycine
ASA	Simple	≈0 <sup>b</sup>	≈0	≈0	≈0	≈0	≈0
	Limit	≈0	≈0	0.17	0.22	0.22	0.44
	Molecular	≈0	0.06	0.33	≈0	0.33	0.38
<i>Ab initio</i>	STO-3G	1	1	3	7	16	12
	3-21G	1	2	4	8	17	13
	6-31 <sup>++</sup> G**	1	21	125	286	576	484

<sup>a</sup> Times calculated with a PC Intel Pentium 200 MHz, running under Windows 95. Times should be taken as approximate, because Windows 95 runs self-programs in the background that can slow the calculation down.

<sup>b</sup> A zero value means that the computer processor has not already updated the clock; a PC, such as the one used, takes 0.056 s to update the system clock.

related times needed to compute them are reported in Table 9.

From Tables 8 and 9, it can be seen that no really significant differences appear between the values of computed overlap-like MQS-SM. However, the times required to compute them differ greatly and become large as the computational level achieves more accuracy.

## CONCLUSIONS

The use of ASA density functions in quantum mechanical computations has opened a new door to fast and precise results, compared with the results obtained using *ab initio* computations. We have shown here that the ASA density function describes molecular shapes in almost the same way as *ab initio* computations, and the MQS-SM example verifies the validity of the use of the ASA density function.

## ACKNOWLEDGMENTS

This research has been partially supported by a CICYT grant (SAF 96-0158), by the Fundació Maria Francisca de Roviralta, and by a EU contract (ENV4-CT97-0508). The authors acknowledge Dr. M. Pastor (Laboratory of Chemometrics, Department of Chemistry, University of Perugia) for his kindness in providing us with the information about the marching cubes algorithm. X. Gironés wishes to acknowledge University of Girona for a predoctoral fellowship.

## APPENDIX-MARCHING CUBES ALGORITHM

Marching Cubes Algorithm (MCA) is used to render isosurfaces in volumetric data.<sup>20,21</sup> The basic notion in MCA consists in that one can define a voxel as the sequence of the pixel values at the eight corners of a cube. If one or more pixels of a cube have values less than the user-specified isovalue, and one or more have values greater than this value, one knows the voxel must contribute to some component of the isosurface. By determining which edges of the cube are intersected by the isosurface, one can create triangular patches that divide the cube between regions within the isosurface and regions outside. By connecting the patches from all cubes on the isosurface boundary, a surface representation is obtained.

### Algorithm details

There are two major components of MCA. The first consists into deciding how to define the section or sections of surface which chop up an individual cube. If each corner is classified as either being below or above the isovalue, there are 256 possible configurations of corner classifications. However if one accounts for symmetry, only 15 configurations remain (see Reference 22a for a figure). Two of these are trivial: all points are inside or outside the cube and do not contribute to the isosurface. For all configurations it is needed to determine where, along each cube edge, the isosurface crosses, and these edge intersection points can be used to create one or more triangular patches for the isosurface.

When surface patches can be created for a single voxel, this process can be applied to the entire volume. One can process the volume in slabs, where each slab is comprised between 2 slices of pixels. Each cube can be treated either independently, or edge intersections between cubes which share the edges can be propagated.

More information about the MCA can be found on the WWW (see Reference 22).

## REFERENCES

- 1 Carbó, R., Leyda L., and Arnau, M. How similar is a molecule to another? An electron density measure of similarity between two molecular structures. *Int. J. Quantum Chem.* 1980, **17**, 1185–1189
- 2 Besalú, E., Carbó, R., Mestres, J., and Solà, M. Foundations and recent advances on molecular quantum similarity. *Topics Curr. Chem.* 1995, **173**, 31–62
- 3 Carbó, R., Besalú, E., Calabuig, B., and Vera, L. Molecular quantum similarity: Theoretical framework, ordering principles, and visualization techniques. *Adv. Quantum Chem.* 1994, **25**, 253–313
- 4 Carbó R., Ed. *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*. Kluwer, Amsterdam, 1995
- 5 Carbó-Dorca, R., and Mezey, P.G., Eds. *Advances in Molecular Similarity*, Vol. 1. JAI Press, Greenwich, Connecticut, 1996
- 6 Cioslowski, J., and Fleischmann, E.D. Assessing molecular similarity from results of ab initio electronic structure calculations. *J. Am. Chem. Soc.* 1991, **113**, 64–67
- 7 Burt, C., Richards, W.G., and Huxley, P. The application of molecular similarity calculations. *J. Comput. Chem.* 1990, **11**, 1139–1146
- 8 Mezey, P.G. Density domain bonding topology and molecular similarity measures. *Topics Curr. Chem.* 1995, **173**, 63–82
- 9 Allan, N.L., and Cooper, D.L. Momentum-space electron densities and quantum molecular similarity. *Topics Curr. Chem.* 1995, **173**, 85–111
- 10 Ponc, R. Similarity models in the theory of pericyclic reactions. *Topics Curr. Chem.* 1995, **174**, 1–26
- 11 Constants, P., Amat, L., and Carbó-Dorca, R. Toward a global maximization of the molecular similarity function: Superposition of two molecules. *J. Comput. Chem.* 1997, **18**, 826–846
- 12 Parr, R.G. *The Quantum Theory of Molecular Electronic Structure*. W.A. Benjamin, New York, 1963
- 13 Constants, P., and Carbó, R. Atomic shell approximation. Electron density fitting program algorithm restricting coefficients to positive values. *J. Chem. Inf. Comput. Sci.* 1995, **35**, 1046–1053
- 14 Amat, L., and Carbó-Dorca, R. Quantum similarity measures under atomic shell approximation: First order density fitting using elementary Jacobi rotations. *J. Comput. Chem.* 1997, **18**, 2023–2039
- 15 ASA coefficients and exponents for an assorted sample of atoms can be seen and downloaded from the WWW site: <http://iqc.udg.es/cat/similarity/ASA/funcset.html>
- 16 Mezey, P.G. *Shape in Chemistry*. VCH Publishers, New York, 1993
- 17 Frisch, M.J., Trucks, G.W., Schlegel, H.B., Gill, P.M.W., Johnson, B.G., Robb, M.A., Cheeseman, J.R., Keith, T., Petersson, G.A., Montgomery, J.A., Raghavachari, K., Al-Laham, M.A., Zakrzewski, V.G., Ortiz, J.V., Foresman, J.B., Cioslowski, J., Stefanov, B.B., Nanayakkara, A., Challacombe, M., Peng, C.Y., Ayala, P.Y., Chen, W., Wong, M.W., Andres, J.L., Replogle, E.S., Gomperts, R., Martin, R.L., Fox, D.J., Binkley, J.S., Defrees, D.J., Baker, J., Stewart, J.P., Head-Gordon, M., Gonzalez, C., and Pople, J.A. *Gaussian-94*, revision E.2. Gaussian, Inc., Pittsburgh, Pennsylvania, 1995
- 18 A Fortran 90 procedure for the MCA can be seen and downloaded from the WWW site: <http://iqc.udg.es/cat/similarity/ASA/mca.html>
- 19 GiD, Geometry and Data, a pre/postprocessor graphical interface. An academic version can be downloaded at the WWW site: <http://gatxan.upc.es>. International Center for Numerical Methods in Engineering (CIMNE), 1998, Barcelona, e-mail: [cimne@etseccpb.upc.es](mailto:cimne@etseccpb.upc.es)
- 20 Lorensen, W.E., and Cline, H.E. A high resolution 3D surface construction algorithm. *Computer Graphics* 1987, **21**, 163
- 21 Watt, A., and Watt, M. *Advanced Animation and Rendering Techniques*. Addison-Wesley, Reading, Massachusetts, 1992
- 22 WWW sites with more information about the marching cubes algorithm: (a) <http://exaflop.org/docs/marchcubes>; (b) [http://panda.uchc.edu/htbit/indiv/software\\_docs/marching\\_cubes.html](http://panda.uchc.edu/htbit/indiv/software_docs/marching_cubes.html)