

QUANTUM MOLECULAR SIMILARITY MEASURES (QMSM) AND THE ATOMIC SHELL APPROXIMATION (ASA)

Pere Constans, Lluís Amat,
Xavier Fradera, and Ramon Carbó-Dorca

Abstract	188
I. Introduction	188
II. Atomic Shell Approximation	190
A. Density Fitted Atomic Shells	191
B. Empirical Atomic Shells	198
III. Similarities in the Atomic Shell Approximation	201
A. HCN/N and NaCN/N Systems	203
B. Spiro Hydantoins Comparison	205
IV. Conclusions	210
Acknowledgments	210
References	210

Advances in Molecular Similarity
Volume 1, pages 187–211
Copyright © 1996 by JAI Press Inc.
All rights of reproduction in any form reserved.
ISBN: 0-7623-0131-7

ABSTRACT

First-order electron density similarity measures for large molecules are straightforward and can be efficiently computed if the *atomic shell approximation* (ASA) is used. Within this approximation the molecular electron distributions are represented by simple superpositions of spherical atomic contributions. A new algorithm to optimally select shells fitting known electron distributions and an empirical scheme to construct molecular densities by summing atomic fragments are presented. The accuracy of both ASA procedures is analyzed comparing approximated and *ab initio* QMSM.

I. INTRODUCTION

Molecules, as quantum objects, are completely described by the set of reduced density matrices arising from successive integration of their attached spin-space N electron wave functions, $\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N)$, being the s order reduced density matrix given by:

$$\begin{aligned} \rho_s(\mathbf{x}_1, \dots, \mathbf{x}_s; \mathbf{x}'_1, \dots, \mathbf{x}'_s) \\ = \binom{N}{s} \int \Psi^*(\mathbf{x}_1, \dots, \mathbf{x}_N) \Psi(\mathbf{x}'_1, \dots, \mathbf{x}'_N) d\mathbf{x}_{s+1} \dots d\mathbf{x}_N \end{aligned} \quad (1)$$

Sets of functions belonging to different molecules could be compared and similarity measures among them mathematically established. Similarities are cognitive relations for ordering and classifying object qualities, and their measure can reveal aspects of accessible human knowledge. The classical understanding of chemical systems as physical, three-dimensional entities can be recovered by means of the diagonal part of the spin independent first-order density matrices, or briefly, the electron densities of probability, which are expressed, removing superfluous indices, as:¹

$$\rho(\mathbf{r}) = N \int \Psi^*(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) \Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) ds_1 d\mathbf{x}_2 \dots d\mathbf{x}_N \quad (2)$$

The spatial electron density function $\rho(\mathbf{r})$ and its derivatives provide the means for a definition of atoms in molecules,² the identification of chemical bonds, and rigorous quantification of chemical concepts as covalent bond order, steric crowding, electronegativity, or bond hardness.³

A quantum molecular similarity measure (QMSM) based on these real space electron densities is generally defined as,⁴

$$z_{AB} = \iint \rho_A(\mathbf{r}_A) \Theta(\mathbf{r}_A, \mathbf{r}_B) \rho_B(\mathbf{r}_B) d\mathbf{r}_A d\mathbf{r}_B \quad (3)$$

where ρ_A and ρ_B are the electron densities of two arbitrary molecules A and B , and Θ is a positive definite operator. Since the set of functions (Eq. 1) and consequently function (Eq. 2) parametrically depends, in the Born–Oppenheimer approximation, on the nuclei coordinates, the measure z_{AB} for any considered molecular geometry is assumed to be taken at the mutual positioning of both molecules which maximizes the integral (Eq. 3). This conceptually simple similarity measure is impractical for drug design purposes because of its computational difficulty. Within the LCAO approach, first-order electron densities are given as double sums over pairs of basis functions in the form,

$$\rho(\mathbf{r}) = \sum_{ij}^n D_{ij} \varphi_i^*(\mathbf{r}) \varphi_j(\mathbf{r}) \quad (4)$$

where D_{ij} are the density matrix coefficients, $\varphi_i(\mathbf{r})$ and $\varphi_j(\mathbf{r})$ are the atomic orbitals, and n is the number of these basis functions. Every evaluation of z_{AB} in the maximization procedure requires $n_A^2 n_B^2$ computations of many center integrals, together with a cumbersome transformation of the elements D_{ij} under molecular rotation. CNDO-like approximations—computations based on a discrete representation of electron densities, computationally more attainable definitions of similarity,⁵ or fittings of electron density to simpler spherical functions⁶—have been proposed with the aim to extend similarity measures based on quantum mechanics to pharmacological design.

Since the First Girona Seminar, where several works were presented exploring this last strategy,⁷ important advances have been done in our laboratory in the representation of electron densities as superposition of spherical atomic shells, eliminating deficiencies, both theoretical and computational, that the simple least-squares fitting (LSF) presents. The theoretical restriction imposed on the set of variational coefficients, i.e. to be non-negative, has led to the development of a fitting scheme for approximating electron densities, the *atomic shell approximation* (ASA), where shells are optimally selected from a nearly complete functional space.⁸ Solving this theoretical constraint in the ASA procedure fixes the computational drawbacks: exponent optimization; nearly linear dependencies; the need for several basis sets to optimally reproduce different calculated densities; and arbitrary assignments of shells in an atom, which could distort the resulting charge distribution within a molecule. Moreover, the ASA opens an avenue for modeling promolecules, i.e., molecular electron representations built on atomic contributions. Therefore, sharp electronic distributions may be diffused by atomic vibrations, or conformational movements may be allowed during the similarity maximization, giving a more realistic vision of molecules. In this latter case, atoms and their attached electrons can be displaced from the original position to construct different conformations. This is, strictly speaking, an extrapolation since the density is initially computed at a single conformational arrangement; thus densities for the rest of the conformations are obtained starting from this initial density. In such a

case, it is likely that the nonphysically reliable density obtained by simple LSF could fail.

Now, at the time of concluding the Second Girona Seminar, one can regard ASA as more than a computational device to approximate first-order QMSM integrals. ASA is an accurate physical model useful to extend QMSM to real problems in pharmacological drug research. The present work is concerned with the ASA and its ability to accurately calculate overlap QMSM based on first-order density functions. The complete ASA fitting scheme will be presented, empirical ASA approaches made by summing atomic fragments of density analyzed, and deviations of approximated QMSM from *ab initio* values quantified.

II. ATOMIC SHELL APPROXIMATION

Electron distributions of atoms in field-free space are spherically symmetric⁹ and expressible in terms of integral transforms over the radial coordinate, such as:

$$\rho_a(\mathbf{r}) = \int_0^{\infty} f_i(\zeta) e^{-\zeta |\mathbf{R}_a - \mathbf{r}|} d\zeta \quad (5)$$

In the case of a Gaussian kernel, the approximation of the integral (Eq. 5) by a finite sum leads to electron densities expressed by a superposition of spherical shells in the form,

$$\rho_a(\mathbf{r}) \approx \sum_i n_i S_i(\mathbf{R}_a - \mathbf{r}) \quad (6)$$

where shells $S_i(\mathbf{R}_a - \mathbf{r})$ are defined as,

$$S_i(\mathbf{R}_a - \mathbf{r}) \equiv \left(\frac{\zeta_i}{\pi} \right)^{3/2} e^{-\zeta_i (\mathbf{R}_a - \mathbf{r})^2} \quad (7)$$

in order to identify coefficients n_i with shell populations. Approximation (Eq. 6) together with the idealization of molecular densities built on spherical atomic shells constitutes the ASA, whose molecular electron distributions appear as:

$$\rho_{ASA}(\mathbf{r}) = \sum_a \sum_{i \in a} n_i S_i(\mathbf{R}_a - \mathbf{r}). \quad (8)$$

This portable representation of electron densities has been widely used when simple functional forms were required, such as the treatment of X-ray crystallographic data,¹⁰ or in molecular shape characterization.¹¹ Equation 8 can also be used to compute molecular wave functions from n *S*-like orbitals.¹² When these representations are applied to QMSM computations, a great simplification is reached with both the number of involved basis functions and integral complexity

being greatly reduced. The following sections show how to obtain the shells S_i and the respective occupations n_i for any molecule, while quantifying at the same time the errors of such approximation by comparison with ab initio QMSM. In Section II.A we present a new algorithm which optimally selects shells from a nearly complete functional space and approximate known molecular electron densities, $\rho(\mathbf{r})$. Section II.B analyzes the construction of $\rho_{ASA}(\mathbf{r})$ based on the approximate additivity and invariance of atomic densities in molecular environments. This rough representation of molecular densities is still useful to compute QMSM with acceptable accuracy when densities are not available, as in the case of large systems, or when they are not worthwhile to compute, as in a first selection of similar compounds in a structural database search.

A. Density Fitted Atomic Shells

Having a discrete or functional representation of the electron density of a system, $\rho(\mathbf{r})$ —the best approach in a least-squares sense— $\rho_{ASA}(\mathbf{r})$, in terms of a complete set of functions $S_i(\mathbf{R}_a - \mathbf{r})$, requires only the lineal minimization of the quadratic error integral function:

$$\varepsilon^2(\mathbf{n}) = \int (\rho(\mathbf{r}) - \sum_a \sum_{i \in a} n_i S_i(\mathbf{R}_a - \mathbf{r}))^2 d\mathbf{r} \quad (9)$$

Nearly complete spaces of Gaussian functions can be generated selecting exponents in a geometric sequence,¹³

$$\zeta_i = \alpha\beta^i \quad (10)$$

together with an implicit dependence of the generators α and β with respect to the basis size n , postulated by Ruedenberg et al.¹⁴ as,

$$\ln \ln \beta = b \ln n + b' \quad (11)$$

and,

$$\ln \alpha = a \ln(\beta - 1) + a' \quad (12)$$

to ensure a successful approach to completeness when n is increased. These *even-tempered* sequences, which are a simple and elegant way to construct truncated basis sets, avoid cumbersome nonlinear optimizations and take control over possible linear dependencies.¹⁴ A simple two-dimensional search over generators α and β gives no significant improvement with respect to a fully variational solution optimizing all the exponent series.¹⁴ The parameters α and β are optimized for different sizes of the basis sets and the constants in Eqs. 11 and 12 are obtained by a linear regression.⁸ The values given by these equations, called *regularized even-tempered* parameters, differ very little from the optimized ones, having the

interesting advantage, besides the theoretical correctness, of allowing different basis sizes and a quality fitting exploration in the implementation of the ASA.

Coefficients \mathbf{n} are subject to the physical constraints derived from the fact that $\rho_{ASA}(\mathbf{r})$ is a density of probability function. These constraints are the normalization condition,

$$\sum_i n_i = N \quad (13)$$

and the set of inequalities,

$$n_i \geq 0 \quad \forall i, \quad (14)$$

assuring a positive valued $\rho_{ASA}(\mathbf{r})$ in its whole domain. Restriction (Eq. 13) can be introduced using a Lagrange multiplier formalism. Then the restricted minimum \mathbf{n}'_0 , denoted by primes, of the quadratic error integral function $\varepsilon^2(\mathbf{n})$ accomplishes the linear equation,

$$\mathbf{S}\mathbf{n}'_0 = \mathbf{t}' \quad (15)$$

where the elements of the overlap matrix \mathbf{S} are,

$$s_{ij} = \int S_i(\mathbf{r})S_j(\mathbf{r})d\mathbf{r} \quad (16)$$

and vector \mathbf{t}' is the sum:

$$\mathbf{t}' = \mathbf{t} + \lambda\mathbf{m} \quad (17)$$

The elements of vector \mathbf{t} are the overlap integral of the $\rho(\mathbf{r})$ to be fitted by the basis functions in the new representation, $S_i(\mathbf{r})$, being:

$$t_i = \int \rho(\mathbf{r})S_i(\mathbf{r})d\mathbf{r} \quad (18)$$

And finally, the elements of \mathbf{m} , taking into account the normalization condition, are given by,

$$m_i = \int S_i(\mathbf{r})d\mathbf{r} = 1 \quad (19)$$

and the Lagrange multiplier λ is given by the products:

$$\lambda = (N - \mathbf{m}^T\mathbf{S}^{-1}\mathbf{t})(\mathbf{m}^T\mathbf{S}^{-1}\mathbf{m})^{-1}. \quad (20)$$

Coefficients solving Eq. 15 can be expressed, in terms of the Cramer's rule, by,

$$n'_{0,i} = (S_{1i}t'_1 + S_{2i}t'_2 + \dots + S_{ni}t'_n) \det |\mathbf{S}|^{-1} \quad (21)$$

where S_{ij} is the cofactor of the element s_{ij} in the metric \mathbf{S} . Since \mathbf{S} is a positive definite matrix, and consequently $\det|\mathbf{S}|$ positive valued, non-negative coefficient values constraints (Eq. 14) are equivalent to:

$$S_{1i}f'_1 + S_{2i}f'_2 + \dots + S_{ni}f'_n \geq 0 \quad \forall i \quad (22)$$

This set of inequalities establishes intricate relationships which, once a system and its attached density function $\rho(\mathbf{r})$ are given, indicates that physically acceptable ASA fitted densities will lie in some subspaces from the nearly complete function space. The ASA algorithm, presented in the following section, is an original way to optimally localize such subspaces, or, in other words, to minimize $\varepsilon^2(\mathbf{n})$ constrained to the set of conditions in Eqs. 13 and 14. The subsequent two sections that follow examine the results of this methodology when applied to atomic and molecular systems, respectively.

Algorithm Scheme

Since the error quadratic integral function $\varepsilon^2(\mathbf{n})$ is a quadratic form, its minimum \mathbf{n}'_0 can be expressed in terms of an arbitrary vector \mathbf{n} by the equation,

$$\mathbf{n}'_0 = \mathbf{n} - \mathbf{S}^{-1}\nabla\varepsilon^2(\mathbf{n}) \quad (23)$$

where the gradient at \mathbf{n} is given by:

$$\nabla\varepsilon^2(\mathbf{n}) = 2(\mathbf{n}\mathbf{S} - \mathbf{t}') \quad (24)$$

Choosing the arbitrary point \mathbf{n} with all the components positive, and taking the direction \mathbf{p} ,

$$\mathbf{p} = \mathbf{S}^{-1}\nabla\varepsilon^2(\mathbf{n}) \quad (25)$$

the shortest *approaching path* from the point \mathbf{n} to the minimum \mathbf{n}'_0 , it is possible to define a new point \mathbf{n}'_1 in \mathbf{p} given by:

$$\mathbf{n}'_1 = \mathbf{n} - \xi\mathbf{p}. \quad (26)$$

The parameter $\xi \in [0,1]$ is the largest step through the descending path that keeps the coefficients positive. Analyzing every component at the intersection,

$$0 = n_i - \xi_i p_i, \quad \forall i \quad (27)$$

it can be defined as a subset of ξ_i values,

$$\xi_k^{(+)} = n_k p_k^{-1} \wedge p_k > 0 \quad \forall k \quad (28)$$

for the positive components of the approaching path \mathbf{p} only, giving the maximum step for the considered component. Obviously, no restriction exists if a component

p_j is negative because the corresponding coefficient $n'_{j,1}$ always will be positive. Then, taking ξ as,

$$\xi \equiv \text{Min}_k (1, \xi_k^{(+)}) \quad (29)$$

forces the new point \mathbf{n}'_1 to have positive or zero components. Since the path \mathbf{p} directly conduces to the minimum, the new set of coefficients will decrease the function $\varepsilon^2(\mathbf{n})$.

At this step of the iterative process the functions with null coefficients and positive slope at \mathbf{n}'_1 are discarded. This is so because they would have negative coefficients in a differential steepest-descent displacement from \mathbf{n}'_1 . Afterwards, a new approaching path is computed:

$$\mathbf{p}_r = \mathbf{S}_r^{-1} \nabla_r \varepsilon^2(\mathbf{n}'_{1,r}) \quad (30)$$

The dimension of the problem has been reduced as indicated in expression (Eq. 30) by the subindices r . In the way previously shown, a new step ξ and a new point $\mathbf{n}'_{2,r}$ are computed. Then after expanding $\mathbf{n}'_{2,r}$ to a whole dimensioned vector \mathbf{n}'_2 , maintaining the original zero values for the discarded functions, a computation of the gradient at this improved \mathbf{n}'_2 is performed, closing the second iteration. The process stops when ξ equals one—the minimum reached in a possible subspace—and when all the slopes of shells with zero occupancies are positive, the conditions of a restricted minimum. In this manner, as shown in Table 1, not only a minimum is found in a problem subspace, accomplishing,

$$\mathbf{n}'_{0,r} = \mathbf{S}_r^{-1} \mathbf{t}'_r \quad (31)$$

Table 1. Schematic Description of the ASA Algorithm^a

-
- Compute Integrals \mathbf{t} and \mathbf{S}
 - Compute λ and \mathbf{t}'
 - Initialize \mathbf{n} and $\nabla \varepsilon^2(\mathbf{n})$
 - DO
 - For i (if $n_i = 0$ and $\nabla_i \varepsilon^2(\mathbf{n}) > 0$ discard shell i)
 - Establish Reduced Dimension \mathbf{t}_r , \mathbf{S}_r and \mathbf{S}_r^{-1}
 - Compute λ_r
 - Compute \mathbf{n}'_r
 - Expand \mathbf{n}'_r to \mathbf{n}'^1
 - If $\xi < 1$ DO Continue
 - If (for i ($n'_i = 0$ and $\nabla_i \varepsilon^2(\mathbf{n}'^1) > 0$)) DO exit
 - End DO
 - Minimum $\mathbf{n}^0 = \mathbf{n}'^1$
-

Note: ^a Nomenclature explained in text.

but also the best subspace, i.e., the best fitting function from all possible combinations of basis set functions, is obtained.

Referring to the computational efficiency of this algorithm, two considerations must be taken into account. First, it is worthwhile to realize that an important computational simplification can be introduced removing constraint (Eq. 13), i.e. using t instead of the more expensive t' , during the localization of compatible subspaces. Since the original density function strictly obeys the electron normalization, any flexible enough fitting expansion will freely reproduce this constraint and, consequently, this imposition does not influence the final selection of functions. Constraint (Eq. 13) can be introduced once this first selection is done, allowing further iterations if necessary. The second consideration refers to Eq. 23, which might yield numerical inaccuracies, reflected in abnormally large values for the gradient components. In such a case the solution could be refined since the compatible subspace is already determined, solving directly the linear system (Eq. 31). Even when the number of matrix inversions to be performed during the iterative procedure is large, the computational cost for this restricted fitting is only slightly greater than the simple LSF. This is because symmetric matrix inversion is a fast process compared to integral evaluation.

Fitting Spherical Systems: The Argon Atom

The closed-shell argon atom has a completely spherical electron distribution, and therefore is a suitable example for testing the flexibility of the restricted ASA function. The density to be fitted was computed at the MP2/6-311G* level of theory. Spanning a nearly complete space with 50 functions generated from even-tempered parameters,⁸ the computed ASA density, composed of 22 shells or selected functions, has an associated quadratic error integral value $\varepsilon^2(\mathbf{n})$ of 6.94×10^{-6} with density scaled to one. Such scaling improves the convergence of the algorithm, especially if the initial fitting space is large. The maximum of the function at the nucleus has a value of 46824.18 au, which is 0.9 units over the ab initio 46823.28, and thus being the greatest local difference. The radial distribution presented in Figure 1 is defined as,

$$D(r) \equiv r^2 \int_0^{2\pi} \int_0^{\pi} \rho(r, \varphi, \phi) \sin \varphi \, d\varphi \, d\phi, \quad (32)$$

for the ab initio and the ASA functions. A complete agreement for the first two shells is found, shells now in the sense of Parr et al.,¹⁵ while some slight differences appear in the outer region of argon. Since $\rho(\mathbf{r})$ decreases rapidly in the neighborhood of the nucleus, one finds that at the distance of 1 au from it the value is only 5.2 au, and values close to zero are found at greater distances. For this reason, this region of large distances has an unnoticeable effect in an unweighted $\varepsilon^2(\mathbf{n})$. This is the reason for the differences at greater distances and not the existence of high quantum number electrons, which prevent neither the spherical symmetry of electron distri-

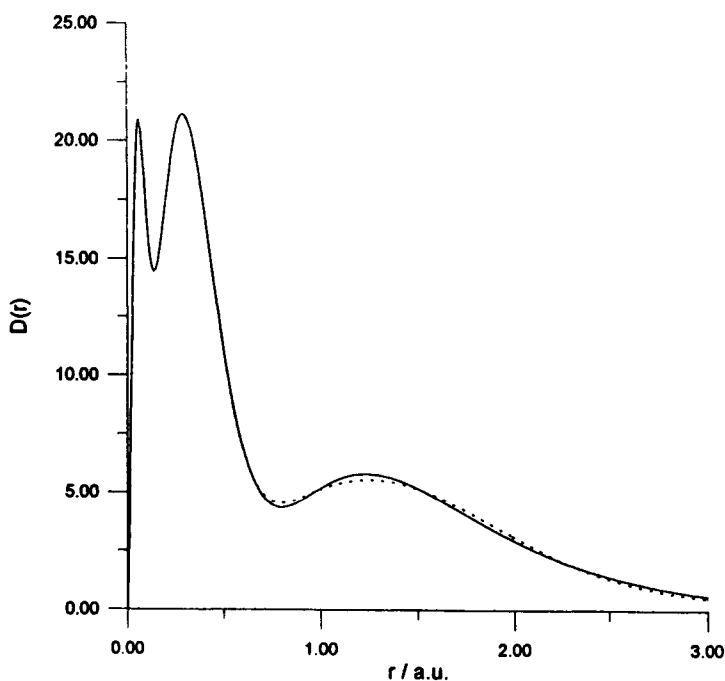


Figure 1. Radial electron distribution $D(r)$ for argon. The MP2/6-311G* distribution is solid line and the ASA is dashed line.

bution in atoms nor their representation by 1S functions. This is also in agreement with the well-established practice of using only 1S Slater or Gaussian functions for spherical orbitals.¹⁶

Fitting Molecular Systems: The Boron Trichloride Molecule

Atoms in molecules no longer have spherical electron distributions. Nevertheless, superposition of spherical atomic shells is still accurate, especially for QMSM purposes, as can be seen in the next example. The electron density for the boron

Table 2. Number of Functions for the Different Densities of the Boron Trichloride Molecule^a

<i>Number of Functions</i>	<i>STO-3G</i>	<i>3-21G</i>	<i>6-21G</i>	<i>6-31G*</i>	<i>6-311G**</i>
Basis functions	32	48	48	72	100
Primitives	96	96	144	184	179
Fitting functions	140	140	140	140	140

Note: ^a The number of initial functions for the ASA fitting is also showed, corresponding to 35 functions per atom.

Table 3. HF Densities for the Boron Trichloride Molecule^a

	<i>STO-3G</i>	<i>3-21G</i>	<i>6-21G</i>	<i>6-31G*</i>	<i>6-311G**</i>
Shells	42	49	61	61	66
Shells on B	9	10	13	13	15
Shells on Cl	11	13	16	16	17
ϵ^2	3.0105E-5	1.7625E-5	6.3197E-6	8.0264E-6	8.0198E-6
Error in S(A,A)	0.0211%	0.0022%	0.0009%	0.0008%	0.0008%

Note: ^a Shells, quadratic errors integrals, and errors in self-similarity.

trichloride molecule, with partial boron–chlorine double bonds, has been computed at different levels of theory at its D_{3h} optimized geometry. The ASA algorithm is independent of these levels of theory since shells are optimally and automatically selected to describe a particular density from a nearly complete space. Table 2 gathers the number of primitives for every basis set whose square is the number of terms in the ab initio density, and the considered basis set size to span a nearly complete space for the ASA fitting, corresponding to 35 functions, generated from parameters in Ref. 8, per atom. Table 3 and Table 4 collect the results of the fitting computations, namely, the number of shells or selected functions and the quadratic integral error $\epsilon^2(\mathbf{n})$, and the error in the self-similarity for an evaluation of the quality of the ASA function. The immediate conclusion from these tables is that when using the ASA there is an important reduction in the number of functions used to express the density function which, together with the fact that these functions are 1S Gaussians, immediately gives an idea of the important reduction in the time needed to compute QMSM. Such simplification does not prevent the generation of QMSM with an acceptable accuracy, as can be seen observing the different errors. As in the previous example, $\epsilon^2(\mathbf{n})$ is computed with density scaled to one and is nearly constant for the different orbital basis sets. The increase in the number of shells when improving the wave function quality is another remarkable aspect of the ASA procedure, showing that it is a *systematic* and *universal* method. Slightly better

Table 4. MP2 Densities for the Boron Trichloride Molecule^a

	<i>STO-3G</i>	<i>3-21G</i>	<i>6-21G</i>	<i>6-31G*</i>	<i>6-311G**</i>
Shells	43	52	61	62	66
Shells on B	10	10	13	14	15
Shells on Cl	11	14	16	16	17
ϵ^2	2.9225E-5	1.6823E-5	5.8198E-6	6.4506E-6	6.3815E-6
Error in S(A,A)	0.0206%	0.0021%	0.0008%	0.0008%	0.0007%

Note: ^a Shells, quadratic errors integrals, and errors in self-similarity.

values for the more precise densities is just a consequence of the optimization of the even-tempered parameters, which were obtained from atomic 6-311G* densities. This selection of shells also gives atomic populations, unambiguously defined in ASA, in agreement with chemical intuition. For the boron atom in the MP2/6-311G** fitting, the atomic population is -0.003 au, in agreement with the expected value. Four acceptable resonant structures can be written down for the boron trichloride molecule, three of them involving double bonds with positive chlorines and the other with partial ionic single bonds with negative chlorines, making the total charge transfer negligible.¹⁷ Exemplifying the importance of a good selection of shells, one can regard the LSF density, computed using the whole 140 function basis set and without positive valued constraint to coefficients (i.e., a lower value in $\varepsilon^2(n)$), which gives a boron charge of -1.10 au) quite far away of what it is expected.

B. Empirical Atomic Shells

A really fast computation of QMSM which could be applied to pattern recognition in 3D structural databases should be extremely simplified and should avoid the need of density computations. Empirical molecular densities can be modeled as simple sums of atomic contributions, having for the so-called *promolecular* electron density:

$$\rho_{EASA}(\mathbf{r}) = \sum_a \rho_{EASA}^a(\mathbf{R}_a - \mathbf{r}), \quad (33)$$

Several functional forms for the shell structure of atoms, $\rho_{EASA}^a(\mathbf{R}_a - \mathbf{r})$, will be analyzed in the present work. The first strategy, based in CNDO-like densities, uses a simple nS STO function per atom, being,

$$\rho_{EASA}^a(\mathbf{R}_a - \mathbf{r}) = q_a |S_a^{l_a}(\mathbf{R}_a - \mathbf{r})|^2 \quad (34)$$

where coefficients q_a are atomic charges, and:

$$S_a^{l_a}(\mathbf{R}_a - \mathbf{r}) = \frac{(2\zeta_a)^{l_a + 1/2}}{\sqrt{4\pi(2l_a)!}} (\mathbf{R}_a - \mathbf{r})^{l_a} e^{-\zeta_a |\mathbf{R}_a - \mathbf{r}|} \quad (35)$$

The radial power term l_a is taken as the row number of atom a in the Periodic Table or, what is nearly the same, the number of maxima in the radial distribution. Exponents ζ_a are taken to exactly reproduce free atom self-similarity values.

A second strategy to enhance atomic densities defines $\rho_{EASA}^a(\mathbf{R}_a - \mathbf{r})$ as a superposition of l_a STO shells in the form:

$$\rho_{EASA}^a(\mathbf{R}_a - \mathbf{r}) = \sum_i^{l_a} m_i |S_a^{l_i}(\mathbf{R}_a - \mathbf{r})|^2 \quad (36)$$

Occupations m_i are the number of electrons commonly associated with the atomic electronic configurations. The set of exponents used are those of Clementi et al.¹⁸ for spherical orbitals.

Similarity measures of a set of fluoro- and chloro-substituted methanes, whose ab initio HF/6-31G** values were already known,⁸ will be reviewed to illustrate the performance of these two empirical approaches. Table 5 presents the similarity values, the ab initio ones in bold, those computed with functional approach (Eq. 34) in italics, and, those with the approach of Eq. 36 in normal type. Results in the

Table 5. QMSM for Fluoro- and Chloro-Substituted Methanes^a

	<i>CH₄</i>	<i>CH₃F</i>	<i>CH₃Cl</i>	<i>CH₂F₂</i>	<i>CH₂Cl₂</i>	<i>CHF₃</i>	<i>CHCl₃</i>	<i>CF₄</i>	<i>CCl₄</i>
CH ₄	31.84	58.78	144.22	58.83	144.23	58.89	144.23	58.92	144.23
	30.56	55.69	128.49	55.75	128.50	55.80	128.50	55.84	128.50
	<i>37.46</i>	<i>68.86</i>	<i>170.32</i>	<i>68.66</i>	<i>169.55</i>	<i>68.44</i>	<i>168.89</i>	<i>68.28</i>	<i>168.31</i>
CH ₃ F		151.11	316.69	150.37	316.87	148.25	317.03	146.60	317.15
		142.37	281.97	141.94	282.16	140.24	282.34	138.97	282.49
		<i>163.54</i>	<i>360.33</i>	<i>161.27</i>	<i>359.59</i>	<i>159.30</i>	<i>358.60</i>	<i>157.66</i>	<i>357.37</i>
CH ₃ Cl			1028.15	318.81	1027.71	319.24	1027.48	319.41	1028.04
			878.59	284.33	878.33	284.80	878.27	284.97	878.92
			<i>1091.08</i>	<i>364.80</i>	<i>1086.24</i>	<i>367.84</i>	<i>1082.27</i>	<i>368.41</i>	<i>1079.13</i>
CH ₂ F ₂				270.43	319.08	258.32	319.55	249.98	319.93
				254.23	284.61	243.77	285.18	236.53	285.67
				<i>289.48</i>	<i>356.80</i>	<i>286.04</i>	<i>355.74</i>	<i>283.38</i>	<i>354.57</i>
CH ₂ Cl ₂					2024.52	319.49	1738.55	319.78	1401.47
					1726.63	284.91	1489.62	285.29	1199.69
					<i>2126.12</i>	<i>354.98</i>	<i>2093.98</i>	<i>353.93</i>	<i>2040.88</i>
CHF ₃						389.77	321.44	386.70	322.01
						366.13	287.24	363.66	287.96
						<i>414.85</i>	<i>353.34</i>	<i>412.76</i>	<i>352.18</i>
CHCl ₃							3020.95	321.74	2694.00
							2574.67	287.56	2303.79
							<i>3146.32</i>	<i>352.25</i>	<i>3109.11</i>
CF ₄								509.07	324.14
								478.04	290.38
								<i>539.88</i>	<i>350.53</i>
CCl ₄									4017.38
									3422.69
									<i>4155.16</i>

Note: ^a Ab initio HF/6-31G* values are in bold, the empirical ASA values using one STO per atom in italics, and EASA with a STO per shell and atom values in medium type.

first approach, with a single nS STO function per atom, show a good agreement with *ab initio* values in case of self-similarities, having a 6% error for CH_4 - CH_4 or a less than a 5% for the CCl_4 - CCl_4 measure, while errors in cross-similarities are larger than 10%. The reason for having more accurate values in self-similarities can be found in the fact that when computing self-similarities there is a perfect matching between the two molecules being compared, which are the same. In this case, the main contribution comes from atoms perfectly superimposed, while contributions from atoms not superimposed are negligible because they are separated by large distances. Given that the exponents in Eq. 34 are taken to reproduce atomic auto-similarities, one can already expect a good result for this case. On the other hand, when dissimilar molecules are compared, one is likely to find pairs of atoms not completely superimposed. These atom pairs are primarily responsible for the greater errors found in this case. The similarity additivity of Eq. 34 is also reflected in the overestimation of all the similarity values, indicating a lack of diffuseness of the atomic densities in molecular environments that this model presents. To better understand this point, one can check that the similarity integral (Eq. 3) increases if charge distribution is concentrated in small areas, being infinity in case of densities collapsed into Dirac deltas. The other approximation, when the density functional form is given by Eq. 36, does not improve the similarity measures in all cases, probably due to the use of nonoptimal exponents to span densities.

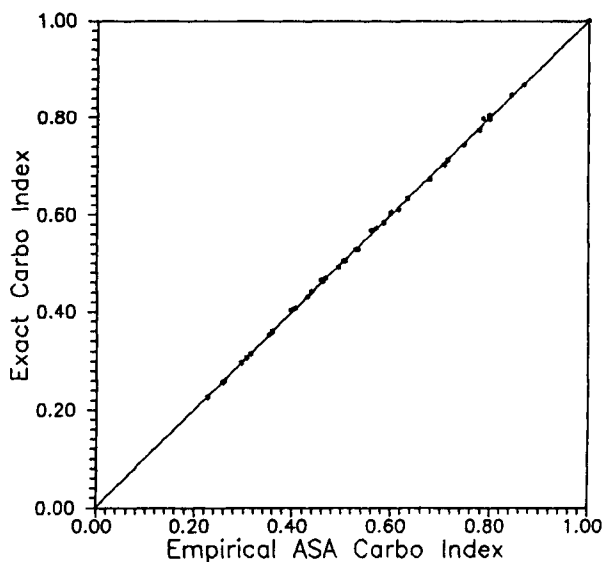


Figure 2. HF/6-31G** versus empirical Carbó indices for the fluoro- and chloro-substituted methanes.

Carbó indices derived from these empirical similarity measures present a better correlation with *ab initio* values as Figure 2 reveals. This agreement can be explained by the systematic deviation which cancels errors in the index computation.

A third strategy using a single 1S GTO function per atom¹⁹ has also been tested with the aim of speeding up similarity maximization. Results are only qualitative and will be presented in next section.

III. SIMILARITIES IN THE ATOMIC SHELL APPROXIMATION

In this section, the performance computing overlap QMSM of several introduced ASA will be analyzed. QMSM for *rigid* molecules are six variable functions, with three of them indicating relative translations and the other three indicating relative orientation. Fixing one of the molecules, molecule A, the similarity function is expressed by,

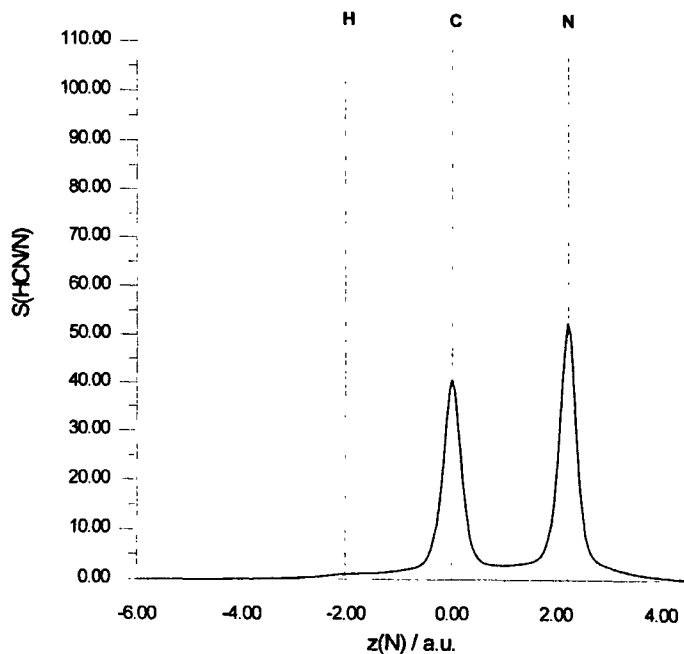


Figure 3. N/HCN Similarity function along the molecular axis. Vertical lines indicate the positioning of molecular atoms.

$$z_{AB}(\Omega) = \int \rho_A(\mathbf{r})\rho_B(\mathbf{r};\Omega)d\mathbf{r} \quad (37)$$

with Ω standing for all six variables. Inside the ASA, similarity measures appear as a sum of isotropic atom–atom contributions, i.e.,

$$z_{AB}(\Omega) = \sum_{ab} z_{ab}(r_{ab}(\Omega)) \quad (38)$$

where the similarity for atomic pairs is given by:

$$z_{ab}(r_{ab}(\Omega)) = \sum_{i \in a} \sum_{j \in b} n_i n_j \int S_i(\mathbf{R}_a - \mathbf{r}) S_j(\mathbf{R}_b(\Omega) - \mathbf{r}) d\mathbf{r} \quad (39)$$

Expression 39 enables a global maximization scheme whose first principles are given in Ref. 8. This scheme is used in all similarity optimizations contained in the present work. Therefore, in Section III.A the similarity function between atomic nitrogen and two linear molecules will be computed at the *ab initio* MP2/6-311G** level of theory, and differences with the approximate functions will be displayed in order to have a vision of the behavior of the ASA atom–atom contributions (Eq.

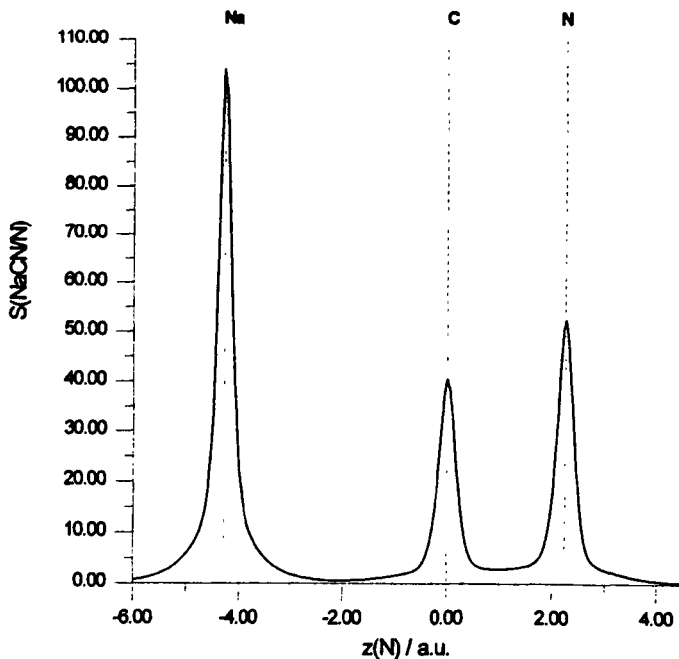


Figure 4. N/NaCN Similarity function along the molecular axis. Vertical lines indicate the positioning of molecular atoms.

39). This will shed some light when, afterwards, in Section III.B the accuracy of the ASA method will be checked in a series of *real* drug design molecules.

Computations of ab initio densities and optimized geometry have been performed using the Gaussian 92 ensemble of programs.²⁰ Program ExSim²¹ has been used to compute ab initio similarities, ASAC²² for fitting the ab initio densities and computation of their similarities, and MolSimil 95²³ for the empirical computations.

A. HCN/N and NaCN/N Systems

Similarity functions for HCN/N and NaCN/N systems only depend on the coordinates of the nitrogen atom with respect to some fixed frame of axis defining the atomic positions of the cyanide molecule having:

$$z_{XCN,N}(r_N) = \int \rho_{XCN}(r) \rho_N(r; r_N) dr \quad (40)$$

If XCN molecules lie along Z axis, the pictures of $z_{XCN,N}(0,0,z_N)$ will be sufficient to show the peculiarities of similarity functions, also present in more complicated

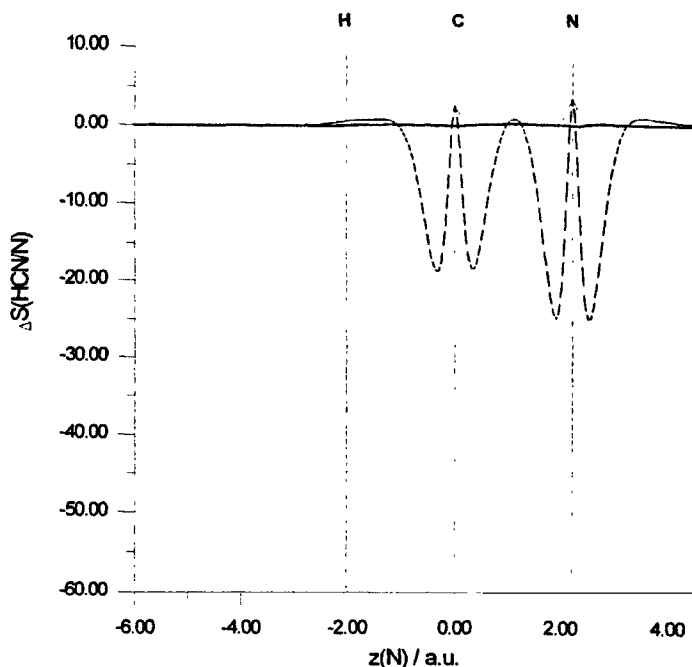


Figure 5. N/HCN ab initio-approximate differences in similarity function. Thick solid line corresponds to ASA computations, fine solid line to Slater empirical approach, and dashed is the empirical Gaussian approach.

systems because of the nearly atom–atom additivity. Figure 3 and Figure 4 represent the similarity function computed at the MP2/6-311G** level of theory for nitrogen vs. hydrogen and sodium cyanide, respectively. The HCN/N function only presents two maxima due to the fact that electron density flows from hydrogen to the electronegative group cyanide. Even if hydrogen was not bonded to an electronegative group, its maximum would appear nearly hidden by the heavier atoms.

The differences with the similarity functions obtained using ASA densities are given for hydrogen and sodium cyanides, respectively, in Figures 5 and 6. Thick lines correspond to the differences between exact and ASA QMSM and are confused with the abscise, showing a nearly complete agreement especially at the maxima. At approximately 1 bohr around carbon and nitrogen coordinates, the maximum difference is found to be 0.2 au in similarity. Fine solid lines correspond to the differences with the empirical function built using Slater-type functions (Eq. 36). They also show a conformity with the exact functions, except at the maxima where they are approximately 10% lower. Dashed lines correspond to the simplest approach analyzed, which consists in a single 1S GTO per atom. These functions

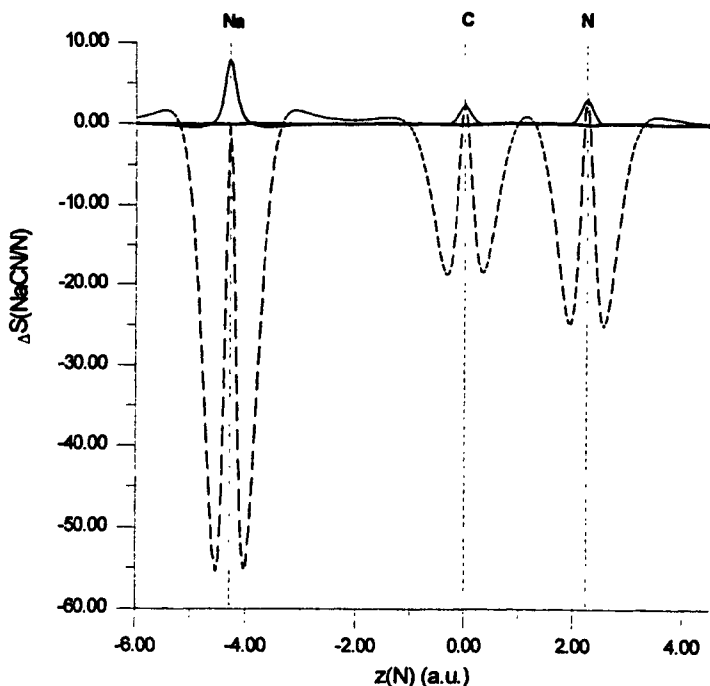


Figure 6. N/NaCN ab initio-approximate differences in similarity function. Thick solid line corresponds to ASA computations, fine solid line to Slater empirical approach, and dashed is the empirical Gaussian approach.

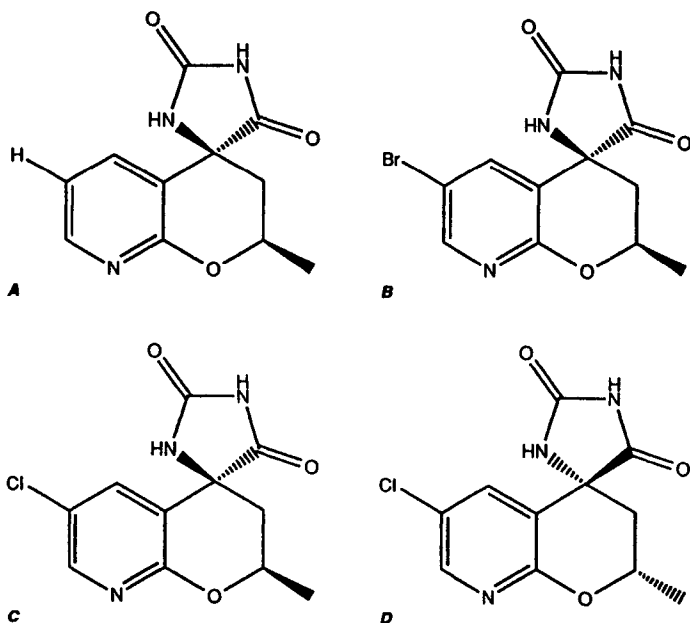


Figure 7. Representation of the four spiro hydantoin aldose reductase inhibitors considered.

are only a qualitative description since a single Gaussian cannot describe simultaneously height and width, thus their use should be restricted to interactive visual matching. Compared molecules usually will be placed at the right maximum arrangement, but the corresponding similarity value will appear highly distorted because of the important errors when nuclei are not perfectly superimposed, the case of most of the nuclei when matching dissimilar molecules.

B. Spiro Hydantoin Comparison

A series of four spiro hydantoin 8-aza-4-chromanones which act as aldose reductase inhibitors²⁴ has been selected to test the performance of the ASA method in a real case of drug design. Their chemical structure is presented in Figure 7.

Ab initio and ASA similarities have been computed at the fully optimized HF/STO-3G geometry. EASA computations were performed with the set of functions in Eq. 36. Similarities and their derived Carbó indices are presented in Table 6 and Table 7, respectively, the ab initio values appearing in bold, the ASA values in normal type, and the EASA ones in italics. Similarity maximization was only performed using ASA and EASA densities, obtaining in both cases the same maxima with just a negligible difference in the final values of Ω . Then, ab initio punctual similarities were performed at the ASA maxima. In order to easily allow

Table 6. Similarities for Spiro Hydantoins^a

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
<i>A</i>	729.840	712.354	454.488	353.080
	729.548	713.971	456.296	354.063
	<i>710.737</i>	<i>630.816</i>	<i>446.103</i>	<i>346.642</i>
<i>B</i>		11053.921	3187.972	2988.291
		11051.215	3194.098	2993.832
		<i>8704.890</i>	<i>2706.978</i>	<i>2510.181</i>
<i>C</i>			1687.997	1294.120
			1687.574	1294.970
			<i>1558.762</i>	<i>1177.520</i>
<i>D</i>				1687.963
				1687.541
				<i>1558.773</i>

Note: ^a Ab initio values are in bold, ASA in medium type, and empirical ASA values in italics.

a comparison of the results, exact–approximate differences and percentual errors are presented in Table 8 and Table 9, respectively, while Table 10 and Table 11 give the errors corresponding to the Carbó indices.

Differences in ASA similarities are mainly originated by the atomic sphericity loss since densities for free atoms are excellently reproduced. This deformation, as commented in Section III.A, is more noticeable when nuclei are not completely

Table 7. Carbó Indices for Spiro Hydantoins^a

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
<i>A</i>	1	0.2508	0.4095	0.3181
		0.2514	0.4112	0.3191
		<i>0.2536</i>	<i>0.4238</i>	<i>0.3293</i>
<i>B</i>		1	0.7380	0.6918
			0.7396	0.6933
			<i>0.7349</i>	<i>0.6814</i>
<i>C</i>			1	0.7667
				0.7674
				<i>0.7554</i>
<i>D</i>				1

Note: ^a Ab initio values are in bold, ASA in medium type, and empirical ASA values in italics.

Table 8. Similarity Differences, Ab Initio-Approximate, for Spiro Hydantoins^a

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
<i>A</i>	0.292 <i>19.103</i>	-1.617 <i>81.538</i>	-1.808 <i>8.385</i>	-0.983 <i>6.438</i>
<i>B</i>		2.706 <i>2349.031</i>	-6.126 <i>480.994</i>	-5.541 <i>478.110</i>
<i>C</i>			0.423 <i>129.235</i>	-0.850 <i>116.600</i>
<i>D</i>				0.422 <i>129.190</i>

Note: ^a ASA values are in medium type and empirical ASA values in italics.

superimposed, having in the previous examples maximum differences of 0.2 au in similarity for carbon and nitrogen atoms. Extrapolating these differences to the present example, one can easily understand the different behavior of self- and cross-similarities, the first ones being more accurate. This also explains, for instance, why z_{BC} has the greatest absolute error, while z_{CD} has the precision of a self-similarity (see Table 8). In the first case the arrangement maximizing the electron density overlap superposes bromine and chlorine atoms, whereas all other atoms appear displaced. Figure 8 shows the molecular superposition for the *B*–*C* pair. By contrast, molecules *C* and *D*, pictured in Figure 9, completely match except for the methyl group and the ring attached at chiral carbons. Nevertheless, the change in chirality

Table 9. Percentage Similarity errors, Ab Initio-Approximate, for Spiro Hydantoins^a

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
<i>A</i>	-0.040 <i>-2.688</i>	0.226 <i>-12.926</i>	0.396 <i>-1.880</i>	0.278 <i>-1.857</i>
<i>B</i>		-0.024 <i>-26.985</i>	0.192 <i>-17.769</i>	0.185 <i>-19.047</i>
<i>C</i>			-0.025 <i>-8.291</i>	0.066 <i>-9.902</i>
<i>D</i>				-0.025 <i>-8.288</i>

Note: ^a ASA values are in medium type and empirical ASA values in italics.

Table 10. Carbó Index Differences, Ab Initio-Approximate, for Spiro Hydantoins^a

	A	B	C	D
A	0	-0.001 <i>-0.003</i>	-0.002 <i>-0.014</i>	-0.001 <i>-0.011</i>
B		0	-0.002 <i>0.003</i>	-0.001 <i>0.010</i>
C			0	-0.001 <i>0.011</i>
D				0

Note: ^a ASA values are in medium type and empirical ASA values in italics.

clearly separates these groups, making the overlap contribution of the relevant atoms negligible.

In the case of EASA similarities, errors obviously come from a poor description of electron densities, which is especially evident for the measures involving the bromine-substituted molecule. However, this simple picture of molecular densities places these molecules at the proper maximum arrangement and gives Carbó indices correctly in one decimal figures.

Regarding the possible application of QMSM in QSAR studies, it is interesting to make a qualitative comparison between the activity values for this set of molecules and some of the QMSM values obtained. Thus, it can be seen that, while *B* and *C* are the most active molecules, the Carbó index is higher for the *C-D* pair than for the *B-C* pair in all the approximations considered, with *D* being an inactive molecule. This result is, at first sight, quite surprising because *B* and *C* share the

Table 11. Percentual Carbó Index Errors, Ab Initio-Approximate, for Spiro Hydantoins^a

	A	B	C	D
A	0	0.239 <i>1.104</i>	0.413 <i>3.374</i>	0.313 <i>3.401</i>
B		0	0.216 <i>-0.422</i>	0.216 <i>-1.526</i>
C			0	0.091 <i>-1.496</i>
D				0

Note: ^a ASA values are in medium type and empirical ASA values in italics.

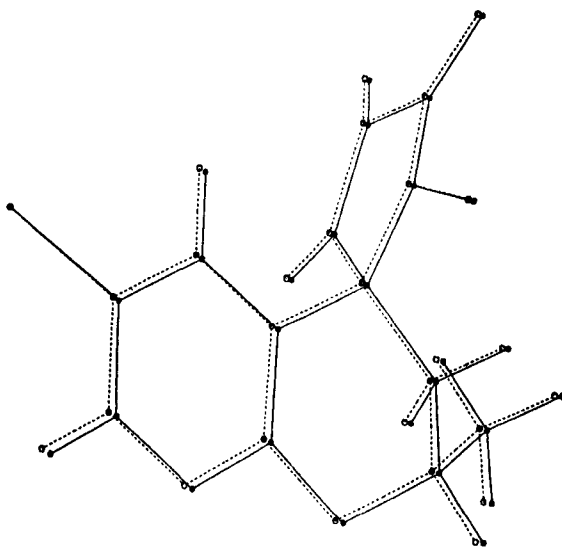


Figure 8. Superposition of the bromine-substituted spiro hydantoin (*B*) with the chloro-substituted (*C*). Pictured by MolSimil 95.

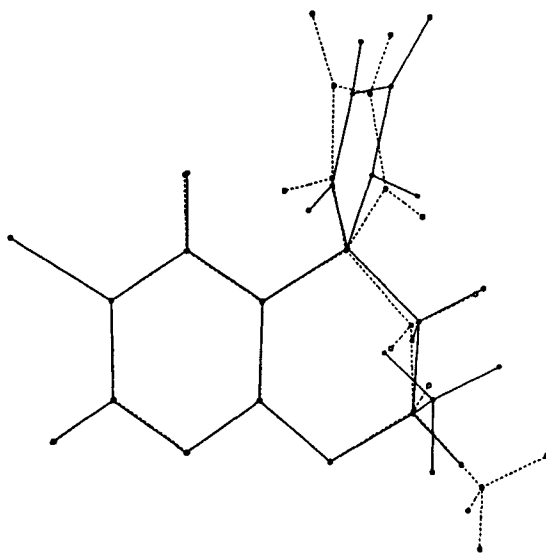


Figure 9. Superposition of the chloro-substituted spiro hydantoin (*C*) and (*D*). Pictured by MolSimil 95.

same structure and differ only in the halogen, while *C* and *D*, although having different halogens, seem to be structurally more different because its five-membered ring cannot be superposed due to the different chirality of the two molecules. However, the low value for the *B*–*C* pair can be attributed to the shifting of the large common substructure slightly out of the maximal superposition, as can be seen in Figure 8. This is forced by the superposition of Br and C1 and because the C–Br and C–C1 distances are slightly different. This arises not from the ASA fitting but rather from the theoretical background consisting in using electronic densities which do not take into account the vibrational motion of atoms.

IV. CONCLUSIONS

The main conclusion of the present work indicates that QMSM based on electron distributions can be accurately computed, even for large molecules. The purpose of this work has been to assess a fast and correct methodology to quantify molecular similarities based on first-order electronic distributions. The ASA, due to its simplicity, brings not only the means to perform fast QMSM computations, but also possible ways of modeling molecules and defining local similarities. Future work will allow nuclear movements and the averaging of electronic distributions by considering harmonic nuclear displacements, thus giving a more real picture of molecules. We expect that within this framework it will be possible to obtain better correlations between QMSM and biological activities in cases such as the spiro hydantoins considered in section III.B. Furthermore, the concept of local similarities could be valuable in the localization of active centers or common patterns in sets of molecules.

ACKNOWLEDGMENTS

P.C. has benefitted from a CIRIT OA/au BQF93/24 fellowship, and L.A. from a "Ministerio de Educación y Ciencia" fellowship. P.C. thanks Dr. M.D. Pujol from the Pharmacological Chemistry Department at the University of Barcelona for her help in selecting an appropriate set of active molecules.

REFERENCES

1. (a) Löwdin, P.O. *Phys. Rev.* **1955**, *97*, 1474–1489. (b) McWeeny, R. *Proc. Roy. Soc. London* **1959**, *A253*, 242–259.
2. Bader, R.F.W. *Atoms in Molecules: A Quantum Theory*; Clarendon Press: Oxford, 1990.
3. (a) Cioslowski, J.; Mixon, S.T. *J. Am. Chem. Soc.* **1991**, *113*, 4142. (b) Cioslowski, J.; Mixon, S.T. *J. Am. Chem. Soc.* **1992**, *114*, 4382. (c) Cioslowski, J.; Mixon, S.T. *J. Am. Chem. Soc.* **1993**, *115*, 1084.
4. (a) Carbó, R.; Leyda, L.; Arnau, M. *Int. J. Quantum Chem.* **1980**, *17*, 1185–1189. (b) Carbó, R.; Calabuig, B. *Int. J. Quantum Chem.* **1992**, *42*, 1681–1693. (c) Carbó, R.; Calabuig, B. *Int. J. Quantum Chem.* **1992**, *42*, 1695–1709. (d) Carbó, R.; Calabuig, B.; Vera, L.; Besalú, E. *Adv.*

- Quantum Chem.* **1994**, *25*, 253–313. (e) Besalú, E.; Carbó, R.; Mestres, J.; Solà, M. *Topics in Current Chemistry* **1995**, *173*, 31–62.
5. Cioslowski, J.; Fleischmann, E.D. *J. Am. Chem. Soc.* **1991**, *113*, 64–67.
 6. Good, A.C.; Richards, W.G. *J. Chem. Inf. Comput. Sci.* **1992**, *33*, 112–116.
 7. (a) Mestres, J.; Solà, M.; Duran, M.; Carbó, R. *J. Comp. Chem.* **1994**, *15*, 1113–1120. (b) Carbó Ed. *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*; Kluwer Academic: Netherlands, 1995.
 8. Constans, P.; Carbó, R. *J. Chem. Inf. Comput. Sci.* **1995**.
 9. Unsöld, A. *Ann. Physik* **1927**, *82*, 355–393.
 10. (a) Coppens, P.; Pautler, D.; Griffin, J.F. *J. Am. Chem. Soc.* **1971**, *93*, 1051–1058. (b) Schwarz, W.H.E.; Lagenbach, A.; Birlenbach, L. *Theor. Chim. Acta* **1994**, *88*, 437–445.
 11. Walker, P.D.; Arteca, G.A.; Mezey, P.G. *J. Comp. Chem.* **1991**, *12*, 220–230.
 12. (a) Paoloni, L.; Giambiagi, M.S.; Giambiagi, M. *Estratto da Atti della Società dei Naturalisti e Matematici di Modena* **1969**, *C*, 89–105. (b) Frost, A.A. *J. Chem. Phys.* **1967**, *47*, 3707. (c) Moncrieff, D.; Wilson, S. *Molecular Physics* **1994**, *82*, 523–530.
 13. Reeves, C.M.; Harrison, M.C. *J. Chem. Phys.* **1963**, *39*, 11–17.
 14. (a) Ruedenberg, K.; Raffeneffi, R.C.; Bardon, D. *Proceedings of the 1972 Boulder Conference on Theoretical Chemistry*; Wiley: New York, 1973, p. 164. (b) Schmidt, M.W.; Ruedenberg, K. *J. Chem. Phys.* **1979**, *71*, 3951–3962. (c) Feller, D.F.; Ruedenberg, K. *Theoret. Chim. Acta* **1979**, *52*, 231–251.
 15. (a) Politzer, P.; Parr, R.G. *J. Chem. Phys.* **1976**, *64*, 4634–4637. (b) Proft, F.; Geerlings, P. *Chem. Phys. Lett.* **1994**, *220*, 405–410.
 16. (a) Huzinaga, S. *J. Chem. Phys.* **1965**, *42*, 1293. (b) Huzinaga, S. *J. Chem. Phys.* **1977**, *67*, 5973–5974.
 17. Pauling, L. In *The Nature of the Chemical Bond and the Structure of Molecules and Crystals*; Cornell University Press: New York, 1960.
 18. (a) Clementi, E.; Raimondi, D.L. *J. Chem. Phys.* **1963**, *38*, 2686. (b) Clementi, E.; Raimondi, D.L.; Reinhard, W.P. *J. Chem. Phys.* **1967**, *47*, 1300–1302.
 19. Besalú, E.; Carbó, R.; Lobato, M. *Sci. Gerund.*, in press.
 20. Frisch, M.J.; Trucks, G.W.; Head-Gordon, M.; Gill, P.M.W.; Wong, M.W.; Foresman, J.B.; Johnson, B.G.; Schlegel, H.B.; Robb, M.A.; Replogle, E.S.; Gomperts, R.; Andres, J.L.; Raghavachari, K.; Binkley, J.S.; Gonzalez, C.; Martin, R.L.; Fox, D.J.; Defrees, D.J.; Baker, J.; Stewart, J.J.P.; Pople, J.A. *Gaussian 92*, Revision B, Gaussian, Inc., Pittsburgh PA, 1992.
 21. Constans, P. *ExSim Program* version 1.0 (CAT, 1995).
 22. Constans, P.; Carbó, R. *ASA Calculations* version 2.0 (CAT, 1995).
 23. Amat, L.I.; Besalú, E.; Carbó, R. *MolSimil 95* (CAT, 1995).
 24. Sarges, R.; Goldstein, S.W.; Welch, W.M.; Swindell, A.C.; Siegel, T.W.; Beyer, T.A. *J. Med. Chem.* **1990**, *33*, 1859–1865.