Molecular Electronic Density Fitting Using Elementary Jacobi Rotations under Atomic Shell Approximation

Lluís Amat and Ramon Carbó-Dorca*

Institute of Computational Chemistry, University of Girona, 17071 Girona, Catalonia, Spain

Received March 25, 2000

Fitted electron density functions constitute an important step in quantum similarity studies. This fact not only is presented in the published papers concerning quantum similarity measures (QSM), but also can be associated with the success of the developed fitting algorithms. As has been demonstrated in previous work, electronic density can be accurately fitted using the atomic shell approximation (ASA). This methodology expresses electron density functions as a linear combination of spherical functions, with the constraint that expansion coefficients must be positive definite, to preserve the statistical meaning of the density function as a probability distribution. Recently, an algorithm based on the elementary Jacobi rotations (EJR) technique was proven as an efficient electron density fitting procedure. In the preceding studies, the EJR algorithm was employed to fit atomic density functions, and subsequently molecular electron density was built in a promolecular way as a simple sum of atomic densities. Following previously established computational developments, in this paper the fitting methodology is applied to molecular systems. Although the promolecular approach is sufficiently accurate for quantum QSPR studies, some molecular properties, such as electrostatic potentials, cannot be described using such a level of approximation. The purpose of the present contribution is to demonstrate that using the promolecular ASA density function as the starting point, it is possible to fit ASA-type functions easily to the *ab initio* molecular electron density. A comparative study of promolecular and molecular ASA density functions for a large set of molecules using a fitted 6-311G atomic basis set is presented, and some application examples are also discussed.

INTRODUCTION

Because of the expensive computational requirements to perform *ab initio* calculations over many-electron systems, namely, for calculations involving four-center integrals, approximated methodologies have become an important device of computational chemistry. For example, many quantum mechanical procedures often employ fitted electronic densities instead of *ab initio* ones. This is the case of some density functional approaches, which use fitted densities to reduce the formal scaling of four-index two-electron integrals.[1−7] However, other applications of fitted electron density functions have been investigated, for instance in the quantum similarity framework.[8−37] In this domain, the atomic shell approximation (ASA)[21−27] has been used in many instances for the evaluation of quantum similarity measures (QSM) in quantum QSPR applications.[28−37] This ASA approach constructs the electronic distribution of a given system as a linear combination of spherical Gaussian-type orbitals, restricting expansion coefficients to be positive definite. This last condition guarantees a probabilistic interpretation of the resultant density function. Recently, a simple but powerful technique based on an elementary Jacobi rotations (EJR) technique[38] in conjunction with a Taylor series expansion was developed to fit electronic density functions.[24] The EJR technique is a norm-conserving orthogonal transformation, and in consequence avoids the inclusion of a Lagrange multiplier to guarantee the proper normalization of electronic density in the objective function. The constraint of preserving positive definite the expansion coefficients in the optimization procedure is merely accomplished by a simple transformation, consisting of defining ASA coefficients as the square module of an underlying variable set.

Nowadays, QSM has evidenced a considerable growth, especially for quantum QSPR studies where large molecular systems are studied.[28−37] This development has been feasible in part because of the utilization of ASA density functions, which allow fast QSM calculations. QSM applied to 3D QSAR analysis habitually assumes that the molecular superposition is defined on the QSM maximum.[16] This fact requires highly repeated computation of QSM and, as a consequence, the need to use a suitable approximation for the electron density function. A promolecular approach[39] was employed in early work to construct molecular ASA density functions. Within this mechanism molecular densities are expressed as sums of the contributions of free atoms. Then, once the atomic coordinates and a fitted atomic basis set are known, the density function of any molecule is easily built. This approximation has been proven to give satisfactory results in QSM applications.[28−37] In this way, some fitted atomic basis sets have been published and presented as a data set on the World Wide Web.[40] Among them, the most relevant are a fitted 3-21G basis set for atoms H to Kr,[23] a fitted Huzinaga basis set for atoms H to Rn,[24] a fitted 6-21G basis set for atoms H to Ar,[25] and a fitted 6-311G basis set for atoms H to Ar.[26] The main advantage of the promolecular approach is that it avoids the need to compute *ab initio*

* To whom correspondence should be addressed. E-mail: director@iqc.udg.es. Phone: +34 972418357. Fax: +34 972418356.

10.1021/ci0000272 CCC: $19.00 © 2000 American Chemical Society
Published on Web 07/18/2000

density functions for the studied molecules. But this simple technique is not sufficient to determine some molecular properties, such as total atomic charges or electrostatic potentials. Moreover, this approach gives an inexact description of chemical bonds because it considers atomic densities constant within the molecular environment.

In the present work, a previously developed EJR fitting algorithm is applied to molecules. Promolecular density functions provide initial guess high-quality molecular electronic densities. This fact permits accurate fitted molecular ASA density functions to be achieved in a fast manner. This computational procedure could be a simple alternative to the previously proposed one,[21,22] which, using a sophisticated methodology, fits *ab initio* electron densities to ASA densities, while satisfying the constraint that expansion coefficients have to be definite positive. The fitted molecular electronic distributions obtained in this way furnish a more realistic description of molecules, permitting the primary electronic properties to be determined with sufficient accuracy.

Here, promolecular and fitted molecular ASA density functions are compared with respect to *ab initio* ones for a representative set of molecules. This analysis is performed by contrasting values of the quadratic error integral function, relative error in the electron−nuclei Coulomb attraction expectation energy, and relative error in overlap-like and Coulomb-like quantum self-similarity measures.

THEORY AND METHOD

Density Functions. The *ab initio* first-order LCAO−MO electron density of a quantum molecular system can be generally written as

$$\rho(\mathbf{r}) = \sum_{\mu,\nu} D_{\mu\nu} \chi_\mu^*(\mathbf{r})\, \chi_\nu(\mathbf{r}) \tag{1}$$

where $\{D_{\mu\nu}\}$ are the elements of the charge−bond order matrix and $\{\chi_\mu\}$ the atomic orbital basis set. At the same time the form of the density function over the MO set can be written as

$$\rho(\mathbf{r}) = \sum_i c_i |\varphi_i(\mathbf{r})|^2 \tag{2}$$

where $\{c_i\}$ are the MO occupation numbers and $\{\varphi_i(\mathbf{r})\}$ the MOs.

The use of fitted densities is very useful to overcome the evaluation of some bottleneck *ab initio* computations, such as four-center integrals. One of these methodologies consists of the ASA framework.[21−27] Formally similar to the MO density defined above and on the basis of various mathematical developments,[17−20] the ASA first-order density function is constructed as a simple linear combination of spherical Gaussian-type functions:

$$\rho^{\mathrm{ASA}}(\mathbf{r}) = \sum_i w_i |s_i(\mathbf{r})|^2 \tag{3}$$

with the restriction that the expansion coefficients $\{w_i\}$ have to be positive definite. A second constraint of $\{w_i\}$ coefficients is related to the normalization condition of $\rho(\mathbf{r})$:

$$\int \rho^{\mathrm{ASA}}(\mathbf{r})\, \mathrm{d}\mathbf{r} = 1 \tag{4}$$

Then, taking a basis set of normalized Gaussian functions, $\{s_i\}$, the expansion coefficients need to fulfill the additional condition

$$\sum_i w_i = 1$$

EJR Fitting Algorithm for ASA Coefficients. The EJR technique was originally employed to calculate electronic energies and wave functions,[41−48] although it can be successfully used in other optimization procedures, such as a fitting methodology. In this respect, EJR has been shown to be an efficient and accurate method for the calculation of fitted densities. The developed fitting algorithm was described in some detail in recent work.[23,24] Then, only a brief review of the main attributes which constitutes the basis of the present study will be examined. Additionally to the implementation of the EJR technique, the computational process was accelerated, expressing the cosine and sine functions of the EJR rotation angle as the first terms of a Taylor series.

The vector containing the ASA coefficients, $\mathbf{w} = \{w_i\}$, is computed by minimizing a quadratic error integral function between the *ab initio* and ASA density functions

$$\epsilon^{(2)} = \int |\rho(\mathbf{r}) - \rho^{\mathrm{ASA}}(\mathbf{r})|^2\, \mathrm{d}\mathbf{r} \tag{5}$$

which can be expressed in a matrix notation as

$$\epsilon^{(2)} = Z + \mathbf{w}^{\mathrm{T}}\mathbf{Z}\mathbf{w} - 2\mathbf{b}^{\mathrm{T}}\mathbf{w} \tag{6}$$

In eq 6 Z denotes an *ab initio* overlap-like quantum self-similarity measure (QS-SM), defined by the integral

$$Z = \int |\rho(\mathbf{r})|^2\, \mathrm{d}\mathbf{r} = \sum_{\mu,\nu} D_{\mu\nu} \sum_{\lambda,\sigma} D_{\lambda\sigma} \int \chi_\mu^*(\mathbf{r})\, \chi_\nu(\mathbf{r})\, \chi_\lambda^*(\mathbf{r})\, \chi_\sigma(\mathbf{r})\, \mathrm{d}\mathbf{r} \tag{7}$$

when $\rho(\mathbf{r})$ is substituted by eq 1. The elements of the matrix $\mathbf{Z} = \{Z_{ij}\}$ and the vector $\mathbf{b} = \{b_i\}$ are given, respectively, by the integrals

$$Z_{ij} = \int |s_i(\mathbf{r})|^2 |s_j(\mathbf{r})|^2\, \mathrm{d}\mathbf{r} \tag{8}$$

and

$$b_i = \int |s_i(\mathbf{r})|^2 \rho(\mathbf{r})\, \mathrm{d}\mathbf{r} = \sum_{\mu,\nu} D_{\mu\nu} \int |s_i(\mathbf{r})|^2 \chi_\mu^*(\mathbf{r})\, \chi_\nu(\mathbf{r})\, \mathrm{d}\mathbf{r} \tag{9}$$

To interpret $\rho^{\mathrm{ASA}}(\mathbf{r})$ as a probability distribution function, it is necessary that all atomic shell occupancies be positive-valued. This restriction is easily accomplished if a new set of coefficients is considered, $\mathbf{x} = \{x_i\}$, and the old ones are defined by the generating rule $\forall i$: $w_i = |x_i|^2$. The application of an EJR transformation, $\mathbf{J}_{pq}(\alpha)$, over the vector $\mathbf{x}$ can be described by the equations

$$\begin{aligned} \dot{x}_p &\leftarrow c x_p - s x_q \\ \dot{x}_q &\leftarrow s x_p + c x_q \end{aligned} \tag{10}$$

where only the elements p and q are modified. The symbols c and s appearing in eq 10 determine the cosine and sine of the EJR rotation angle α. The function to be optimized

corresponds to the variation of $\epsilon^{(2)}$ with respect to the active pair of elements $\{x_p, x_q\}$ after the EJR $\mathbf{J}_{pq}(\alpha)$ is applied on vector **x**:

$$\delta\epsilon^{(2)} = \delta x_p{}^4 Z_{pp} + \delta x_q{}^4 Z_{qq} + 2\delta(x_p{}^2 x_q{}^2) Z_{pq} + 2\delta x_p{}^2 \sum_{i\neq p,q} x_i{}^2 Z_{pi} + 2\delta x_q{}^2 \sum_{i\neq p,q} x_i{}^2 Z_{iq} - 2b_p \delta x_p{}^2 - 2b_q \delta x_q{}^2 \tag{11}$$

The $\delta x_p{}^2$, $\delta x_q{}^2$, $\delta x_p{}^4$, $\delta x_q{}^4$, and $\delta(x_p{}^2 x_q{}^2)$ terms are easily calculated (see ref 23), giving as a result a quadratic equation with respect to s and c:

$$\delta\epsilon^{(2)} = E_{04}s^4 + E_{13}cs^3 + E_{02}s^2 + E_{11}cs \tag{12}$$

To accelerate the computational process, the cosine and sine functions can be expressed as a Taylor series expansion up to second-order terms:

$$\begin{aligned} c &= \cos(\alpha) = 1 - {}^1/_2\alpha^2 + \theta(\alpha^3) \\ s &= \sin(\alpha) = \alpha(1 - {}^1/_6\alpha^2) + \theta(\alpha^4) \end{aligned} \tag{13}$$

Such an approach is efficient only when the objective function is studied near the minimum. Otherwise a more expensive algorithm based on an iterative procedure to find stationary points on eq 12, which was described in some detail in the first paper of this series (ref 23), has to be employed to obtain the optimal EJR angle.

Substitution of eq 13 into eq 12 yields

$$\delta\epsilon^{(2)} = A\alpha^3 + B\alpha^2 + C\alpha \tag{14}$$

where $A = E_{13} - 2E_{11}/3$, $B = E_{02}$, and $C = E_{11}$. Taking a stationary point condition on eq 14

$$\mathrm{d}\delta\epsilon^{(2)}/\mathrm{d}\alpha = 3A\alpha^2 + 2B\alpha + C = 0 \tag{15}$$

and imposing the minimum condition determined by the second derivative, the optimal angle is

$$\alpha^* = \alpha_+ = (1/3A)[-B + (B^2 - 3AC)^{1/2}] \tag{16}$$

so the optimal cosine and sine are given by

$$\begin{aligned} c^* &\approx 1 - {}^1/_2\alpha_+{}^2 \\ s^* &\approx \alpha_+(1 - {}^1/_6\alpha_+{}^2) \end{aligned} \tag{17}$$

Substitution of eq 17 into eq 10 yields a simple expression for the coefficient set variation:

$$\begin{aligned} \dot{x}_p &\leftarrow c^* x_p - s^* x_q = x_p - \alpha_+ x_q + {}^1/_2\alpha_+{}^2(-x_p + {}^1/_3\alpha_+ x_q) \\ \dot{x}_q &\leftarrow s^* x_p + c^* x_q = x_q + \alpha_+ x_p - {}^1/_2\alpha_+{}^2({}^1/_3\alpha_+ x_p + x_q) \end{aligned} \tag{18}$$

which provides the optimal EJR approximate transformation.

Fitting of ASA Exponents. The atomic basis set $\{s_i\}$ is determined by the exponents $\{\zeta_i\}$, which are optimized by minimizing the quadratic error integral function, $\epsilon^{(2)}$, with respect to this nonlinear parameter set. This procedure is essential to obtain accurate fitted density functions with a small number of shells. Starting from a set of $\{\zeta_i\}$ generated by means of an even-tempered sequence, the minimization of exponents is carried out with a Newton method, employing an analytic gradient and a Hessian of $\epsilon^{(2)}$ with respect to $\{\zeta_i\}$.[23] This computational development is only applied in the atomic fitting procedure. In the construction of molecular ASA densities, no variation of atomic exponents is considered at all.

Computational Process. The objective function in the present fitting algorithm is the quadratic error integral defined in eq 5. There are two ASA independent parameter sets, coefficients $\{w_i\}$ and exponents $\{\zeta_i\}$; both are optimized in such a way that the objective function $\epsilon^{(2)}$ is minimized. ASA coefficients are subjected to two simultaneous constraints: they are forced to be positive definite and to accomplish a normalization condition chosen as 1 for atoms and as the number of electrons, N_e, for molecules. The strategy adopted to construct ASA densities is described next.

The first operation consists of fitting an atomic basis set. The essential steps of the atomic fitting algorithm follow.

(1) Set n as the number of fitted atomic shells for a given atom a.

(2) Generate the initial ASA exponents as an even-tempered sequence: $\zeta_i = \alpha\beta^i$, $i = 1, ..., n$.

(3) Compute matrix **Z** and vector **b** from eqs 8 and 9, respectively.

(4) Calculate an initial set of expansion coefficients $\{w_i\}$: $\mathbf{U}^+\mathbf{Z}\mathbf{U} = \Lambda$; $\mathbf{w} = \mathbf{u}^*$.

(5) Optimize the set of exponents $\{\zeta_i\}$ using a Newton algorithm to minimize $\epsilon^{(2)}$.

(6) Optimize the set of coefficients $\{w_i\}$ using the previously described EJR algorithm with an expansion Taylor series procedure for rotation sine and cosine.

(7) If between two successive iterations the condition $\Delta\epsilon^{(2)} < 10^{-6}$ is fulfilled, then stop; otherwise, go to step 5.

This procedure is repeated for an extended number of generating values $\{\alpha,\beta\}$ of even-tempered series, to provide a sufficiently representative grid of starting sets for the exponents $\{\zeta_i\}$. The set of $\{w_i\}$ and $\{\zeta_i\}$ providing the lowest $\epsilon^{(2)}$ is stored.

The next step after the generation of a fitted atomic basis set is the construction of molecular ASA densities. There is a significant point that must be kept in mind when a molecular density is fitted. This procedure needs a previously known molecular *ab initio* density function. Because this fact limits the application of fitted molecular densities, for example, in the practical use of QSM in QSAR analysis, where large molecular systems are studied, a promolecular formalism[39] has been employed in the generation of density functions. For a given fixed nuclear configuration, the promolecular ASA electronic distribution of a molecule A is built as a sum of discrete atomic densities, each centered on its own nucleus:

$$\rho_A^{\mathrm{ASA}}(\mathbf{r}) = \sum_{a\in A} P_a \rho_a^{\mathrm{ASA}}(\mathbf{r}) = \sum_{a\in A} P_a \sum_{i\in a} w_i |s_i(\mathbf{r})|^2 = \sum_{i\in A} \omega_i |s_i(\mathbf{r})|^2 \tag{19}$$

The coefficients P_a can be interpreted as the total electron density on atom a, usually approximated by the atomic

number Z_a. Equation 19 formed as a sum of free atom contributions produces a reasonable approximation to the molecular density.

On the other hand, fitting of ASA functions to molecular *ab initio* electron densities could be implemented in a way similar to that in the atomic case. The proposed procedure for molecules starts from the promolecular ASA density function. This procedure is similar to some self-consistent molecular calculations based on molecular wave functions expanded in terms of spherical Gaussians, which overlaps spherical atomlike charge distributions to generate molecular densities.[1,4] Beginning with this initial guess high-quality $\rho_A^{ASA}(\mathbf{r})$, only the set of ASA coefficients $\{\omega_i\}$ is fitted to *ab initio* $\rho_A(\mathbf{r})$. The first step of the developed strategy for molecular fitting consists in using an unrestricted least-squares procedure to minimize the molecular $\epsilon^{(2)}$ function. The least-squares fitting algorithm uses a Lagrange multiplier to guarantee charge conservation,[1−7,11] but does not restrict the expansion coefficients to be positive definite. In the present scheme, the least-squares fitting algorithm was only implemented as a preliminary and fast adjustment. If nonpositive coefficients are obtained, then the EJR algorithm described in the previous theoretical section is carried out using as the starting point the promolecular ASA density. Normally, if the number of fitted functions is not very large, the use of a least-squares procedure is appropriate, providing the same positive definite $\{\omega_i\}$ coefficients which could be obtained using the EJR algorithm. For molecular fitting, the general strategy of applying expansion Taylor series to describe the cosine and sine functions is generally appropriate because the initial promolecular ASA set of $\{\omega_i\}$ is close to the minimum of $\epsilon^{(2)}$.

Molecular Integrals of Molecular Properties within the ASA Approach. The ASA density functions have been applied to compute approximate integrals of some molecular properties. The main interest of the ASA formalism is the simplicity in the computation of molecular integrals, because they are based just on superposition of 1s Gaussian functions. In the present a normalized 1s GTO of the form

$$|s_i(\mathbf{r}-\mathbf{r}_a)|^2 = (2\zeta_i/\pi)^{3/2}\exp(-2\zeta_i|\mathbf{r}-\mathbf{r}_a|^2) \quad (20)$$

and centered at $\mathbf{r}_a$ has been used. The first integral formula presented corresponds to the so-called overlap QS-SM, which appears in the computation of $\epsilon^{(2)}$. For a given quantum system *A*, this ASA integral is expressed as

$$Z_{AA} = \sum_{i\in A}\sum_{j\in A}\omega_i\omega_j\int|s_i(\mathbf{r}-\mathbf{r}_a)|^2|s_j(\mathbf{r}-\mathbf{r}_b)|^2\,d\mathbf{r}$$
$$= \sum_{i\in A}\sum_{j\in A}\omega_i\omega_j\left(\frac{2\zeta_i\zeta_j}{\pi(\zeta_i+\zeta_j)}\right)^{3/2}\exp\left(-\frac{2\zeta_i\zeta_j}{\zeta_i+\zeta_j}R_{ab}^2\right) \quad (21)$$

But the overlap-like QS-SM integral is not the unique formula to be used as QSM. In the Results and Discussion, Coulomb-like QSM is also used in the pairwise comparison of electron density functions of different molecules. The Coulomb-like QSM between two quantum systems *A* and *B* is defined as

$$Z_{AB}(\mathbf{r}_{12}^{-1}) = \int\int\rho_A(\mathbf{r}_1)\mathbf{r}_{12}^{-1}\rho_B(\mathbf{r}_2)\,d\mathbf{r}_1\,d\mathbf{r}_2 \quad (22)$$

which is approximated within the ASA formalism to

$$Z_{AB}(\mathbf{r}_{12}^{-1})$$
$$= \sum_{i\in A}\sum_{j\in B}\omega_i\omega_j\int\int|s_i(\mathbf{r}_1-\mathbf{r}_a)|^2\mathbf{r}_{12}^{-1}|s_j(\mathbf{r}_2-\mathbf{r}_b)|^2\,d\mathbf{r}_1\,d\mathbf{r}_2$$
$$= 2\sum_{i\in A}\sum_{j\in B}\omega_i\omega_j\left(\frac{2\zeta_i\zeta_j}{\pi(\zeta_i+\zeta_j)}\right)^{1/2}F_0\left(\frac{2\zeta_i\zeta_j}{\zeta_i+\zeta_j}R_{ab}^2\right) \quad (23)$$

where $F_0(x)$ is the zeroth-order incomplete γ function.

Other integrals have been derived and used in this work. For instance the electron−nuclei Coulomb attraction energy, $V(\mathbf{r})$, is defined within the ASA approach as

$$V(\mathbf{r}) = -\sum_c Z_c\int\frac{\rho_A^{ASA}(\mathbf{r}-\mathbf{r}_a)}{|\mathbf{r}-\mathbf{r}_c|}\,d\mathbf{r}$$
$$= -\sum_{i\in A}\omega_i\sum_c Z_c\int\frac{|s_i(\mathbf{r}-\mathbf{r}_a)|^2}{|\mathbf{r}-\mathbf{r}_c|}\,d\mathbf{r}$$
$$= -2\sum_{i\in A}\omega_i(2\zeta_i/\pi)^{1/2}\sum_c Z_c F_0[2\zeta_i R_{ac}^2] \quad (24)$$

and in a similar way, the molecular electrostatic potential of a molecule evaluated at the point $\mathbf{r}_{H^+}$ is defined by

$$V_A(\mathbf{r}_{H^+}) = \sum_c\frac{Z_c}{|\mathbf{r}_c-\mathbf{r}_{H^+}|} - \sum_{i\in A}\omega_i\int\frac{|s_i(\mathbf{r}-\mathbf{r}_a)|^2}{|\mathbf{r}-\mathbf{r}_{H^+}|}\,d\mathbf{r}$$
$$= \sum_c\frac{Z_c}{|\mathbf{r}_c-\mathbf{r}_{H^+}|} - 2\sum_{i\in A}\omega_i\left(\frac{2\zeta_i}{\pi}\right)^{1/2}F_0[2\zeta_i R_{aH^+}^2] \quad (25)$$

RESULTS AND DISCUSSION

Fitted 6-311G Atomic Basis Set. Considerable care has been employed to obtain a compact set of fitted atomic densities, because they constitute the basis for subsequent molecular studies. In this way, a set of ASA functions was fitted to an *ab initio* 6-311G basis set[49,50] for atoms H to Ar.[26] This study is an extension of previous work, where ASA densities were provided for other atomic basis sets.[23−25] Atomic *ab initio* RHF energies and density functions have been calculated using the ATOMIC program.[51] From these spherically symmetric electronic distributions the fitted ASA densities for a different number of atomic shells have been computed. Table 1 lists the fitting results for the atoms involved in the molecular examples studied later on, using for the adjustment from three to seven shells per atom. To assess the quality of the calculated basis set, Table 1 presents the quadratic error integral values $\epsilon^{(2)}$, as well as the errors encountered in the computation of the nuclear attraction potential $V(\mathbf{r})$ and self-similarities Z_{aa} and $Z_{aa}(\mathbf{r}_{12}^{-1})$ values. Coefficients and exponents for this basis set of ASA functions are available for downloading at a World Wide Web site.[40]

As is shown in Table 1, when the number of ASA functions *n* is increased, the $\epsilon^{(2)}$ value quickly decreases. The present result agrees with early findings from a fitted Huzinaga basis set,[24] showing a general tendency to decrease relative errors in Z_{aa} and $V(\mathbf{r})$ values as *n* increases, but not

Table 1. Fitting Results for the 6-311G Basis Set for Atoms H to Ar[a]

		three fitted atomic functions	four fitted atomic functions	five fitted atomic functions	six fitted atomic functions	seven fitted atomic functions
H	$\epsilon^{(2)}$	1.55E−05	4.27E−06	3.76E−07	1.21E−07	6.47E−08
	% Z_{aa}[b]	−0.487	−0.362	0.014	0.007	−0.006
	%$V(\mathbf{r})$[b]	−0.217	−0.220	0.026	0.023	−0.014
	% $Z_{aa}(\mathbf{r}_{12}^{-1})$[b]	−1.010	−0.753	0.052	0.072	−0.041
B	$\epsilon^{(2)}$	1.91E−03	2.04E−04	4.18E−05	1.45E−05	1.36E−05
	% Z_{aa}	0.082	−0.070	−0.068	0.002	−0.024
	%$V(\mathbf{r})$	0.111	−0.267	−0.340	0.049	−0.082
	% $Z_{aa}(\mathbf{r}_{12}^{-1})$	−2.432	−1.754	−1.602	0.098	−0.412
C	$\epsilon^{(2)}$	2.38E−03	3.06E−04	9.84E−05	1.94E−05	1.78E−05
	% Z_{aa}	−0.001	−0.139	−0.133	−0.007	−0.008
	%$V(\mathbf{r})$	−0.109	−0.446	−0.507	0.046	0.038
	% $Z_{aa}(\mathbf{r}_{12}^{-1})$	−3.252	−2.514	−2.325	0.098	0.159
N	$\epsilon^{(2)}$	2.96E−03	4.76E−04	2.08E−04	2.41E−05	1.40E−05
	% Z_{aa}	−0.127	−0.233	−0.216	0.001	−0.015
	%$V(\mathbf{r})$	−0.317	−0.624	−0.663	0.008	−0.005
	% $Z_{aa}(\mathbf{r}_{12}^{-1})$	−3.956	−3.196	−2.936	−0.068	−0.091
O	$\epsilon^{(2)}$	3.74E−03	7.77E−04	2.45E−04	3.09E−05	1.77E−05
	% Z_{aa}	−0.291	−0.354	−0.029	−0.011	−0.011
	%$V(\mathbf{r})$	−0.566	−0.826	−0.106	0.003	−0.011
	% $Z_{aa}(\mathbf{r}_{12}^{-1})$	−4.835	−3.982	−0.818	−0.090	−0.131
F	$\epsilon^{(2)}$	4.71E−03	1.20E−03	2.79E−04	3.59E−05	2.87E−05
	% Z_{aa}	−0.474	−0.491	−0.028	−0.017	−0.014
	%$V(\mathbf{r})$	−0.780	−1.012	−0.094	−0.028	−0.033
	% $Z_{aa}(\mathbf{r}_{12}^{-1})$	−5.525	−4.643	−0.760	−0.211	−0.221
S	$\epsilon^{(2)}$	7.49E−02	5.35E−03	6.64E−04	2.25E−04	5.28E−05
	% Z_{aa}	1.961	0.120	−0.005	−0.018	−0.008
	%$V(\mathbf{r})$	0.626	0.125	−0.105	−0.133	−0.093
	% $Z_{aa}(\mathbf{r}_{12}^{-1})$	−6.792	−1.344	−1.011	−0.847	−0.499
Cl	$\epsilon^{(2)}$	7.88E−02	5.50E−03	6.45E−04	1.83E−04	5.71E−05
	% Z_{aa}	2.043	0.114	−0.007	−0.020	−0.012
	%$V(\mathbf{r})$	1.084	0.122	−0.109	−0.149	−0.113
	% $Z_{aa}(\mathbf{r}_{12}^{-1})$	−4.858	−1.279	−1.003	−0.893	−0.594

[a] Reproduced in part with permission from ref 26. Copyright 2000 Springer. [b] Percent error in Z_{aa}, $V(\mathbf{r})$, and $Z_{aa}(\mathbf{r}_{12}^{-1})$.

in the same proportion than $\epsilon^{(2)}$ varies. With respect to $Z_{aa}(\mathbf{r}_{12}^{-1})$ values, the deviations between ASA and *ab initio* measures are superior to the one-electron properties Z_{aa} and $V(\mathbf{r})$, mainly for the first columns of Table 1, where the errors are significant. But for the basis set employing six and seven fitted functions per atom, the results ameliorate, and could be considered sufficiently accurate in view of the nature of the ASA approximation.

Molecular Fitting. Some illustrative results of promolecular and fitted molecular ASA density functions are presented. As has been already described in the computational process section, first the promolecular ASA density is constructed by means of eq 19, and then, using this as an initial guess, $\rho_A^{\mathrm{ASA}}(\mathbf{r})$, the fitted molecular density, is computed. To simplify the terminology and facilitate the understanding of the present results, ASA calculations at various approximation levels are abbreviated as PASA for *promolecular* ASA densities, and FMASA for *fitted molecular* ASA densities. Additionally, two ASA basis sets have been used in the molecular examples. The notation (3/5/6) refers to an ASA basis set which uses three functions for the H atom, five functions for B, C, N, O, and F atoms, and six functions for S and Cl atoms, whereas in the ASA basis set (4/6/7) the rule of centering four functions on H, six functions on B, C, N, O, and F, and seven functions on S and Cl is followed. This basis set is constructed with the aim to generate sufficiently accurate density functions to be applied in QSM studies with the smallest number of expanded functions. With respect to molecular structures, fully optimized geometries computed using the GAUSSIAN 98 software package[52] have been considered for all studied compounds.

First, a molecular series of halomethane derivatives has been analyzed. *Ab initio* electronic structure calculations have been carried out at the HF level using a 6-311G basis set.[49,50] Table 2 lists the fitting results for these molecules computed for (3/5/6) and (4/6/7) ASA basis sets and using PASA and FMASA densities. The examined parameters for evaluating fitted density functions are the same as those employed for atoms: $\epsilon^{(2)}$ and errors in nuclear attraction potential $V(\mathbf{r})$ and self-similarities Z_{AA} and $Z_{AA}(\mathbf{r}_{12}^{-1})$. The values of $\epsilon^{(2)}$ presented in Table 2 are normalized by the number of electrons, i.e., divided by N_e^2, to provide results comparable to the atomic fitting.

Several features can be spotted in the fitting results shown in Table 2. First, the high quality of the PASA density is confirmed. This is evidenced by the close values of $\epsilon^{(2)}$ to the minimum provided by FMASA density. On the other hand, the use of fitted density functions yields reasonable one-electron properties, as is demonstrated by the remarkably small errors provided by FMASA densities in the computation of Z_{AA} and $V(\mathbf{r})$ values. For instance, the errors in Z_{AA} values for FMASA(4/6/7) calculations are lower than 0.03%, whereas the errors in $V(\mathbf{r})$ values for the same fitted functions are less than 0.15%. With respect to the two-electron property $Z_{AA}(\mathbf{r}_{12}^{-1})$, the errors in FMASA(4/6/7) calculations are less than 0.6% for the set of halomethane derivatives. Although there are high errors present in the Coulomb-like QS-SM evaluation, such parameters are precise enough to be used in QSM applications, avoiding the need to calculate an

Table 2. Fitting Results for Halomethane Molecules

		ASA(3/5/6)		ASA(4/6/7)				ASA(3/5/6)		ASA(4/6/7)	
		PASA	FMASA	PASA	FMASA			PASA	FMASA	PASA	FMASA
CH_4	n	17	17	22	22	CCl_4	n	29	29	34	34
	$\epsilon^{(2)}$	2.70E−04	6.00E−05	3.24E−04	5.87E−05		$\epsilon^{(2)}$	4.77E−05	4.72E−05	2.15E−05	2.06E−05
	% Z_{AA}[a]	−1.488	−0.026	−1.298	0.015		% Z_{AA}	−0.033	−0.010	−0.027	−0.009
	% $V(\mathbf{r})$[a]	0.981	−0.150	1.588	−0.023		% $V(\mathbf{r})$	−0.174	−0.167	−0.128	−0.134
	% $Z_{AA}(\mathbf{r}_{12}^{-1})$[a]	2.788	−0.691	4.646	−0.073		% $Z_{AA}(\mathbf{r}_{12}^{-1})$	−0.782	−0.778	−0.508	−0.524
CFH_3	n	19	19	24	24	$CClF_2H$	n	24	24	29	28
	$\epsilon^{(2)}$	2.82E−04	2.20E−04	2.26E−04	1.45E−04		$\epsilon^{(2)}$	1.14E−04	1.06E−04	6.93E−05	5.75E−05
	% Z_{AA}	−0.486	0.036	−0.446	0.030		% Z_{AA}	−0.097	0.004	−0.087	0.001
	% $V(\mathbf{r})$	0.220	−0.065	0.479	0.010		% $V(\mathbf{r})$	−0.169	−0.092	−0.075	−0.073
	% $Z_{AA}(\mathbf{r}_{12}^{-1})$	0.644	−0.521	1.660	0.016		% $Z_{AA}(\mathbf{r}_{12}^{-1})$	−0.725	−0.552	−0.246	−0.305
$CClH_3$	n	20	20	25	25	CCl_2FH	n	25	25	30	29
	$\epsilon^{(2)}$	1.34E−04	1.08E−04	8.70E−05	5.48E−05		$\epsilon^{(2)}$	8.16E−05	7.72E−05	4.47E−05	3.83E−05
	% Z_{AA}	−0.077	0.0001	−0.066	0.0003		% Z_{AA}	−0.057	−0.002	−0.049	−0.004
	% $V(\mathbf{r})$	−0.038	−0.050	0.052	−0.042		% $V(\mathbf{r})$	−0.159	−0.092	−0.093	−0.091
	% $Z_{AA}(\mathbf{r}_{12}^{-1})$	−0.309	−0.409	0.169	−0.197		% $Z_{AA}(\mathbf{r}_{12}^{-1})$	−0.736	−0.606	−0.364	−0.373
CF_2H_2	n	21	21	26	25	$CClFH_2$	n	22	22	27	26
	$\epsilon^{(2)}$	2.16E−04	1.87E−04	1.54E−04	1.15E−04		$\epsilon^{(2)}$	1.29E−04	1.15E−04	8.11E−05	6.24E−05
	% Z_{AA}	−0.364	0.040	−0.339	0.024		% Z_{AA}	−0.088	0.004	−0.078	0.0004
	% $V(\mathbf{r})$	−0.033	−0.066	0.163	−0.037		% $V(\mathbf{r})$	−0.098	−0.069	−0.006	−0.054
	% $Z_{AA}(\mathbf{r}_{12}^{-1})$	−0.174	−0.510	0.641	−0.171		% $Z_{AA}(\mathbf{r}_{12}^{-1})$	−0.504	−0.488	−0.019	−0.236
CCl_2H_2	n	23	23	28	28	CCl_2F_2	n	27	27	32	31
	$\epsilon^{(2)}$	8.45E−05	7.72E−05	4.56E−05	3.57E−05		$\epsilon^{(2)}$	7.63E−05	7.32E−05	4.19E−05	3.75E−05
	% Z_{AA}	−0.051	−0.002	−0.044	−0.003		% Z_{AA}	−0.061	−0.007	−0.054	−0.006
	% $V(\mathbf{r})$	−0.120	−0.086	−0.058	−0.074		% $V(\mathbf{r})$	−0.203	−0.151	−0.133	−0.108
	% $Z_{AA}(\mathbf{r}_{12}^{-1})$	−0.627	−0.540	−0.263	−0.322		% $Z_{AA}(\mathbf{r}_{12}^{-1})$	−0.855	−0.728	−0.482	−0.422
CF_3H	n	23	23	28	27	$CClF_3$	n	26	26	31	30
	$\epsilon^{(2)}$	1.71E−04	1.53E−04	1.14E−04	9.10E−05		$\epsilon^{(2)}$	1.02E−04	9.52E−05	6.08E−05	5.24E−05
	% Z_{AA}	−0.316	0.037	−0.295	0.025		% Z_{AA}	−0.104	−0.002	−0.094	0.002
	% $V(\mathbf{r})$	−0.200	−0.076	−0.031	−0.052		% $V(\mathbf{r})$	−0.239	−0.140	−0.144	−0.082
	% $Z_{AA}(\mathbf{r}_{12}^{-1})$	−0.699	−0.518	0.012	−0.231		% $Z_{AA}(\mathbf{r}_{12}^{-1})$	−0.928	−0.693	−0.463	−0.331
CCl_3H	n	26	26	31	31	CCl_3F	n	28	28	33	33
	$\epsilon^{(2)}$	6.11E−05	5.86E−05	2.96E−05	2.60E−05		$\epsilon^{(2)}$	5.95E−05	5.81E−05	2.98E−05	2.76E−05
	% Z_{AA}	−0.041	−0.005	−0.034	−0.006		% Z_{AA}	−0.043	−0.009	−0.037	−0.008
	% $V(\mathbf{r})$	−0.156	−0.124	−0.104	−0.107		% $V(\mathbf{r})$	−0.184	−0.160	−0.129	−0.124
	% $Z_{AA}(\mathbf{r}_{12}^{-1})$	−0.743	−0.657	−0.433	−0.440		% $Z_{AA}(\mathbf{r}_{12}^{-1})$	−0.809	−0.755	−0.495	−0.482
CF_4	n	25	25	30	29						
	$\epsilon^{(2)}$	1.44E−04	1.30E−04	9.32E−05	7.72E−05						
	% Z_{AA}	−0.293	0.020	−0.275	0.031						
	% $V(\mathbf{r})$	−0.327	−0.131	−0.173	−0.043						
	% $Z_{AA}(\mathbf{r}_{12}^{-1})$	−1.070	−0.659	−0.431	−0.208						

[a] Percent error in Z_{AA}, $V(\mathbf{r})$, and Z_{AA} $(\mathbf{r}_{12}^{-1})$.

alternative ASA basis specifically fitted for Coulomb integrals. An other characteristic of the present results consists in that the errors in overlap-like QS-SM using FMASA densities always improve the values given by PASA. This trend confirms the high connection between $\epsilon^{(2)}$ and Z_{AA} parameters, which is not obtained in $V(\mathbf{r})$ and $Z_{AA}(\mathbf{r}_{12}^{-1})$ values. Moreover, the deviations between ASA and *ab initio* measures of $V(\mathbf{r})$ and $Z_{AA}(\mathbf{r}_{12}^{-1})$ are generally higher than the error in Z_{AA} values. This fact is consistent with the atomic basis set results documented in Table 1.

In some of the examples presented in Table 2 for FMASA-(4/6/7) densities, the number of fitted functions has decreased with respect to the original PASA. The exclusion of some atomic shells occurs because of the fact that the associated molecular expansion coefficients become zero in the fitting process. It should be noted that the exponents of ASA density functions have been adapted for free atoms, and when these functions are employed in molecules, the saturation of the space of Gaussian functions is more quickly accomplished.

The second example presented for molecular systems concerns benzene and boron trichloride compounds. These molecules were previously fitted to ASA density functions from the *ab initio* HF/6-311G** electronic distribution using an elaborate methodology.[21,22] Basically this old technique could be considered as the limit of the ASA approach. This previous methodology is completely different from the one described here, because it starts from a spanned, nearly complete space of spherical functions, generated from even-tempered parameters. Then, the old algorithm[21] selects the optimal shells by minimizing $\epsilon^{(2)}$, imposing the constraint that expansion coefficients have to be definite positive. In the method of refs 21 and 22, the number of functions involved in the fitting process is very large in comparison with the present method, thus producing the highest computational cost. Table 3 documents the errors for the different approximations with respect to the *ab initio* HF/6-311G** calculations. Fitting results for the nearly saturated basis set, summarized in the third column and denoted as limit ASA, are the best of all ASA approaches. On the other hand, the number of fitted functions for limit ASA is at least 2 times greater than the number of shells in the other ASA densities. Likewise, in the FMASA(4/6/7) fit of the benzene molecule, one function for each hydrogen atom has been eliminated because its expansion coefficient was zero, reducing n from 60 to 54. However, although the quadratic error values are improved, mainly for the BCl_3 molecule, the errors in self-

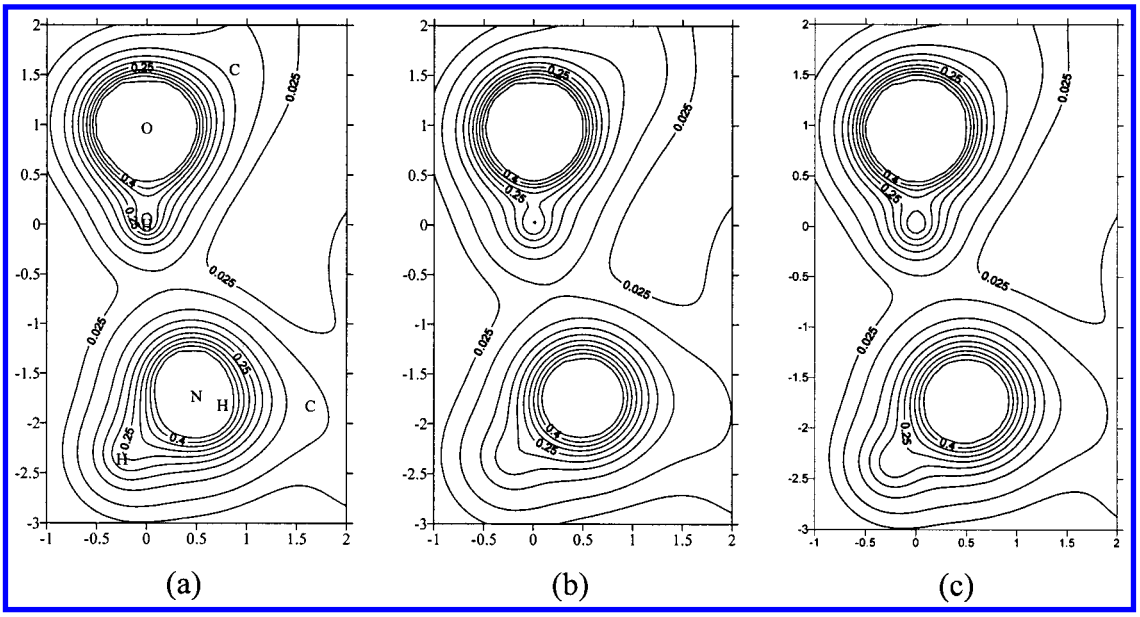

Figure 1. Isodensity contour maps for the intramolecular hydrogen bond of the GABA molecule. Density functions: (a) *ab initio*; (b) PASA(4/6/7); (c) FMASA(4/6/7).

Table 3. Fitting Results for Benzene and Boron Trichloride Molecules

		ASA(3/5/6)		ASA(4/6/7)		limit
		PASA	FMASA	PASA	FMASA	ASA[a]
C_6H_6	n	48	48	60	54	132
	$\epsilon^{(2)}$	8.21E−05	4.87E−05	8.99E−05	4.54E−05	3.91E−05
	% Z_{AA}[b]	−1.674	0.046	−1.545	0.028	0.039
BCl_3	n	23	23	27	27	66
	$\epsilon^{(2)}$	5.55E−05	5.43E−05	2.24E−05	2.13E−05	8.02E−06
	% Z_{AA}	−0.049	−0.006	−0.041	−0.0014	0.0008

[a] From refs 21 and 22. [b] Percent error in Z_{AA}.

similarity behave quite regularly between FMASA and limit ASA. Then, it is therefore plausible that the use of the present ASA densities instead of a nearly saturated basis set could be efficient enough in the computation of similarity measures.

Isodensity Contour Maps. In this section a representation of isodensity contour maps is shown for two molecules: the γ-aminobutyric acid (GABA) and the boron trichloride. This study is an extension of a comparative study of molecular density shape between *ab initio* and ASA densities which was recently reported.[27] The purpose is to graphically show the differences between the studied density functions.

GABA has important physiological effects because it is an inhibitory neurotransmitter in the mammalian central nervous system. The interest of the GABA molecule in this work comes from the intramolecular hydrogen bond, which could give this compound a certain conformation. Next, the description of the hydrogen bond is analyzed for different density functions. Figure 1 shows the isodensity contour maps for *ab initio* HF/6-311G, PASA(4/6/7) and FMASA-(4/6/7) densities in the plane formed by the atoms O−H··· N. These graphs are represented for values of the density function smaller than 0.5 au in the region of the intramolecular hydrogen bond. The main conclusion which could be deduced from Figure 1 is the high similarity between PASA and FMASA maps. On the other hand, the maximal divergences between ASA and *ab initio* densities are located around the bond regions, in particular for the O−C and N−C bonds. This is not surprising since the PASA approach builds molecular electronic distributions from free atomic densities. Moreover, ASA densities describe inaccurately the significant regions of chemical bonds because they use only spherical functions centered on the nuclei. In contrast, it should be noted that there are not significant differences in the description of the hydrogen bond between ASA and *ab initio* graphs.

The second example of an isodensity contour map studies the boron trichloride molecule. The density adjustment for this molecule was studied in the previous section for three ASA approaches fitted to the *ab initio* HF/6-311G** density. This example can help to analyze the effect of the number of fitted functions in the ASA densities. Figure 2 presents the isodensity contour maps for four densities: *ab initio*, PASA(4/6/7), FMASA(4/6/7), and limit ASA. By examining electronic density maps, the differences between the three ASA approaches are not significant. Clearly, from the present results the PASA approach could be conceived as an excellent fast computation of the fitted density function. Additionally, and as was noticed in the GABA maps, the major differences between *ab initio* and ASA densities are encountered around the bond regions.

Representation of Electrostatic Potentials in a 2D Contour Map. Maps of electrostatic potential (MEPs)[53,54] permit information about molecular interactions to be qualitatively obtained, specifically to predict the most probable molecular sites susceptible to an electrophilic attack. In the present work, the electrostatic potential is presented as an example of a molecular property which is not described accurately enough by the PASA approach, and requires the use of FMASA densities to give tangible results. The simple construction of molecular densities by summing contributions of free atoms is not sufficient to describe the negative regions of the MEP, giving only zero or positive values.

A MEP for the ground state of the formaldehyde molecule is shown in Figure 3. Maps obtained from the $V_A(\mathbf{r}_{H^+})$ values and computed using *ab initio* and FMASA(4/6/7) densities

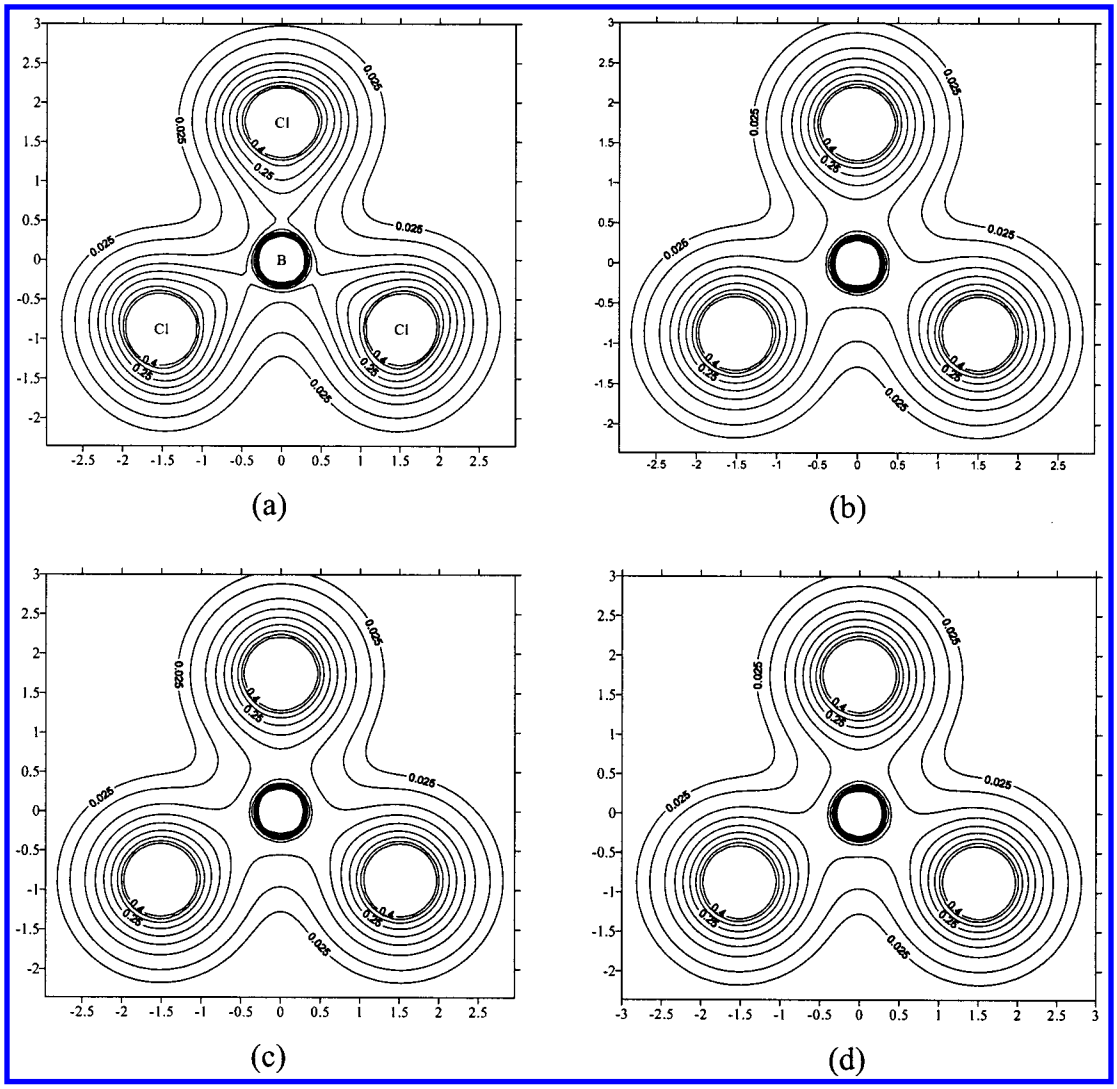

Figure 2. Isodensity contour maps for the boron trichloride molecule. Density functions: (a) *ab initio*; (b) PASA(4/6/7); (c) FMASA-(4/6/7); (d) limit ASA.

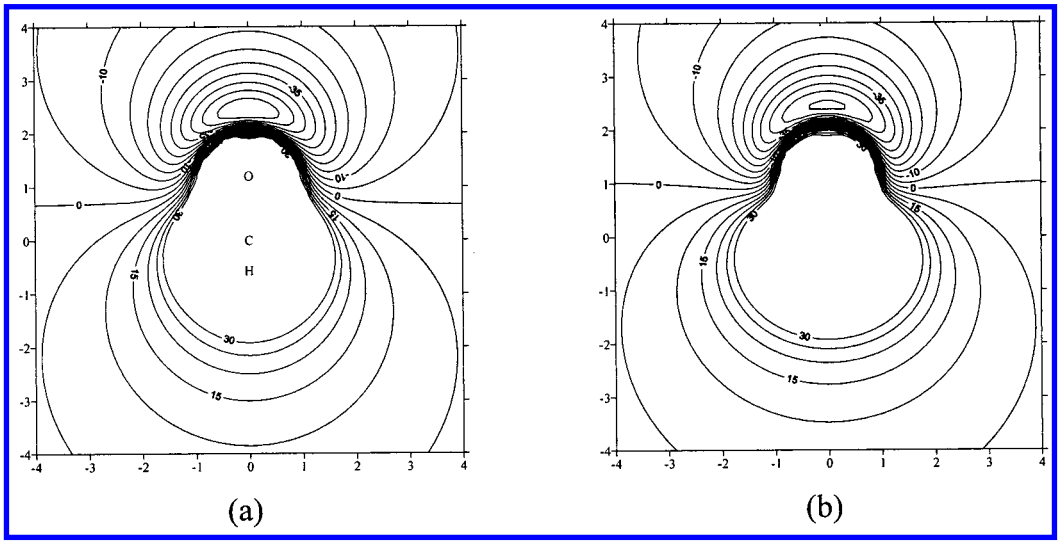

Figure 3. Electrostatic potential contour maps for the H_2CO molecule. Computations for the ground state: (a) *ab initio* and (b) FMASA-(4/6/7).

are represented in the plane perpendicular to the molecular plane containing the carbonyl group. Although for the ground state of the H_2CO molecule the similarity between the exact and approximated MEPs is appreciable, it must be pointed out that ASA has some limitations in the description of MEPs. For instance, the first triplet exited state of the H_2-CO molecule is poorly described even using FMASA densities. This is due to the ASA density structure based on

Chart 1. Molecular Structures of Mannosidase Inhibitors

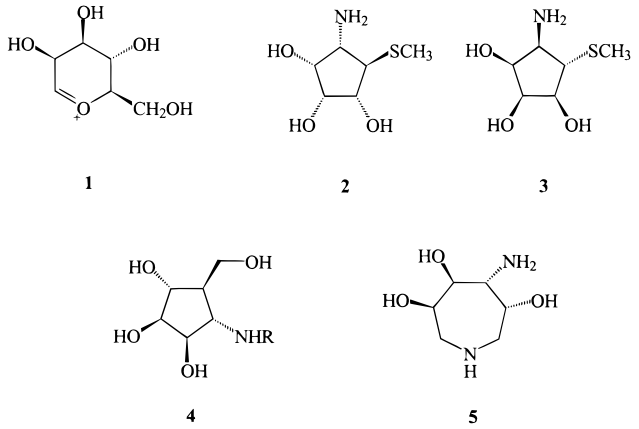

Table 4. Fitting Results for Mannosidase Inhibitors

	n	$\epsilon^{(2)}$	% Z_{AA}[a]	% $Z_{AA}(\mathbf{r}_{12}^{-1})$[a]
flap up mannopyranosyl cation	106	2.86E−05	0.030	−0.158
flap down mannopyranosyl cation	105	2.89E−05	0.031	−0.149
(+)-mannostatin A	115	1.84E−05	0.008	−0.219
(−)-mannostatin A	116	1.87E−05	0.008	−0.214
trihydroxycyclopentylamine	114	2.31E−05	0.024	−0.210
trihydroxyhexahydro-1*H*-azepine	120	2.16E−05	0.020	−0.212

[a] Percent error in Z_{AA} and $Z_{AA}(\mathbf{r}_{12}^{-1})$.

spherical symmetry functions, which impedes complete description of π MO contributions.

QSM Application Example. In this section, a simple application of QSM on molecules is presented. The example concerns some inhibitors of the α-mannosidase. Glycosidase inhibitors are of interest for their diverse biological activities, for example, as antihyperglycemic compounds, inhibitors of tumor metastasis, or antivirals.[55−58] Chart 1 presents the structures of the studied mannosidase inhibitors. The interest to study these carbocyclic compounds by means of QSM comes from the fact that a controversy has appeared in the literature about their resemblance to the mannopyranosyl cation (**1**),[55−58] a proposed intermediate in the reaction catalyzed by the enzyme. Winkler and Holan[55,56] reported the (+)-mannostatin A (**2**), a potent inhibitor of glycoprotein processing, to have a high similarity to structure **1**. The discrepancy appeared when other authors[57,58] proposed the inactive mannosidase inhibitor (−)-mannostatin A (**3**) to be more similar to the mannosyl cation structure than its enantiomeric form **2**. On the other hand, experimental synthesis of two additional compounds has contributed supplementary information: The trihydroxycyclopentylamine (**4**) compound was designed as a mimic of the mannosyl cation structure, and was found to be a potent inhibitor.[59] This experimental finding validates the original proposed mannosyl cation transition state. But an opposite reasoning

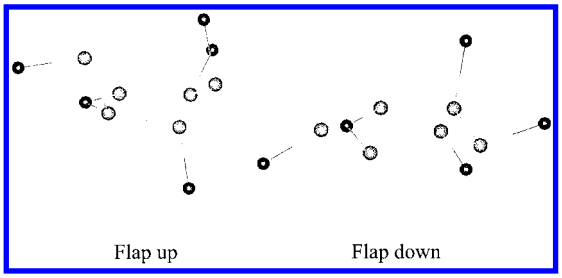

Figure 4. HF/6-311-optimized geometries for flap up and flap down forms of the mannopyranosyl cation. Hydrogen atoms have been omitted.

is evidenced with the trihydroxyhexahydro-1*H*-azepine (**5**) compound, which presents negligible mannosidase activity and was also synthesized as a mimic of the mannosyl cation.[60]

Some preliminary considerations about molecular structures might be indicated prior to QSM analysis being carried out on this problem. As noted in the first paper of Winkler and Holan,[55] mannopyranosyl cation could present two-half-chair forms. Both conformations have been optimized here at the HF/6-311 level of theory, resulting in the finding that the "flap up" form is 4.0 kcal/mol lower in energy than the "flap down" form. Optimized geometries for both conformations are shown in Figure 4. Concerning compound **5**, the hexahydroazepine ring exhibits a high degree of flexibility, having various minimal conformational energy structures. In accordance with previous work of Farr and co-workers,[60] a conformation with the 4-amino group equatorial has been optimized, leading to more stable species than the one with the 4-amino group in the axial position.

The present QSM study for α-mannosidase inhibitors is connected with previous work in which a possible quantification of the Hammond postulate was proposed, to evaluate the structural degree of the transition state advance with respect to the reactants and products by means of QSM.[13,14] Moreover, in another related work, QSM was employed to choose the optimal optimization methodology to construct molecular structures by comparing various theoretical 3D geometries with the experimental one obtained from X-ray analysis.[24,29] Furthermore, the present example shows a situation in which PASA densities appear unable to be used because of the difficulty in describing a delocalized cation within this approach.

After *ab initio* HF/6-311 electronic densities were computed over the studied molecular set, the adjustment of FMASA(4/6/7) densities was carried out. Table 4 documents the quadratic error integral value for the FMASA(4/6/7) densities together with errors in overlap-like and Coulomb-like QS-SM between the ASA and *ab initio* values for the α-mannosidase inhibitors. Remarkable is the small $\epsilon^{(2)}$ value for both half-chair forms of the cation **1**. Subsequent QSM analysis has been performed below using ASA measures of

Table 5. Carbó Index Values Used to Compare Mannosidase Inhibitors

	1 (up)	**1** (down)	**2**	**3**	**4**	**5**
flap up mannopyranosyl cation	1	0.890	0.849	0.876	0.889	0.882
flap down mannopyranosyl cation	0.890	1	0.883	0.863	0.888	0.911
(+)-mannostatin A	0.849	0.883	1	0.900	0.865	0.879
(−)-mannostatin A	0.876	0.863	0.900	1	0.880	0.882
trihydroxycyclopentylamine	0.889	0.888	0.865	0.880	1	0.920
trihydroxyhexahydro-1*H*-azepine	0.882	0.911	0.879	0.882	0.920	1

the Coulomb integral (22). The accuracy of these measures is assessed by the calculated $Z_{AA}(\mathbf{r}_{12}^{-1})$ values, which are generally in good agreement with the *ab initio* ones, giving errors lower than 0.3% as shown in Table 4. Basically, for the usual QSM calculations this precision is sufficient.

Molecular superpositions have been optimized to obtain maximal QSM values.[16] The alignment search procedure is the most time-consuming part of the similarity studies, and necessarily requires the use of ASA densities. It may be worthwhile to mention that the obtained superpositions from QSM are not subjected to external manipulations nor presumptions but are given by the results of the maximal QSM algorithm. Table 5 contains the Carbó indexes, defined as $C_{AB} = Z_{AB}(Z_{AA}Z_{BB})^{-1}$, for the pairwise comparison of α-mannosidase inhibitors. The Carbó index normalizes QSM in the interval {0,1}, indicating that with values closer to 1 there is a greater similarity between compared molecules. Several conclusions could be obtained from the results shown in Table 5. Basically only the first row of Table 5 needs to be analyzed, corresponding to the superpositions of mannosidase inhibitors with the lowest energy flap up half-chair form of the mannosyl cation. The best agreement with the flap up cation **1** is for the molecule **4**, confirming its synthesis as a mimic of mannosyl cation.[59] On the other hand, for the **2** and **3** dissenting molecules the calculated Carbó index for the inactive (−)-enantiomer, $C_{AB} = 0.876$, is somewhat larger than for the active (+)-enantiomer, $C_{AB} = 0.849$. Furthermore, (+)-mannostatin A superimposes better on the flap down mannosyl cation than on the flap up mannosyl cation. In light of the present calculations, the previous results where the (−)-mannostatin A is regarded as the enantiomer more similar to the flap up mannosyl cation can be confirmed, although it has to be pointed out that the small differences between Carbó indexes are not sufficient to discard in a strict sense the alternative hypothesis. The present results only indicate that the proposed intermediate in the hydrolysis of mannopyranosides by α-mannosidase is not an adequate standard of resemblance to use as an indicator of mannosidase inhibition. This fact was demonstrated by the synthesis of compounds **4** and **5** as mimics of mannosyl cation, being active and inactive inhibitors, respectively.

CONCLUSION

This work represents an attempt to apply the EJR technique to fit first-order molecular density functions. The capabilities of the EJR technique justify its use in electron density fitting, for both atoms and molecules. EJR provides a robust computational algorithm for fitting ASA coefficients to *ab initio* density functions. As demonstrated by the presented numerical tests, ASA densities provide quite noticeable agreement with *ab initio* HF method results.

ASA density functions have permitted the extension of QSM to real problems in pharmacological drug design. The main attribute of ASA electron density consists of the fact that expansion coefficients are maintained positive definite, preserving the statistical meaning of density function in the fitted structure. Obtained results demonstrate that ASA densities describe with high accuracy the molecular density shape without computational effort. The potential use of the fitted molecular ASA density functions in the evaluation of some molecular properties, such as electrostatic potentials, may have several applications in computational chemistry. The promolecular approach has also been shown to provide correct density functions in several QSM analyses related to QSAR applications.

ACKNOWLEDGMENT

The present work was supported in part by the Fundació Maria Francisca de Roviralta and a European commission contract, No. ENV4-CT97-0508. This research has been carried out using the CESCA and CEPBA resources, coordinated by C[4].

REFERENCES AND NOTES

(1) Baerends, E. J.; Ellis, D. E.; Ros, P. Self-consistent molecular Hartree−Fock−Slater calculations I. The computational procedure. *Chem. Phys.* **1973**, *2*, 41−51.
(2) Sambe, H.; Felton, R. H. A new computational approach to Slater's SCF-*X*α equation. *J. Chem. Phys.* **1975**, *62*, 1122−1126.
(3) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On some approximations in applications of *X*α theory. *J. Chem. Phys.* **1979**, *71*, 3396−3402.
(4) Delley, B.; Ellis, D. E. Efficient and accurate expansion methods for molecules in local density models. *J. Chem. Phys.* **1982**, *76*, 1949−1960.
(5) Andzelm, J.; Wimmer, E. Density functional Gaussian-type-orbital approach to molecular geometries, vibrations, and reaction energies. *J. Chem. Phys.* **1992**, *96*, 1280−1303.
(6) Gallant, R. T.; St-Amant, A. Linear scaling for the charge density fitting procedure of the linear combination of Gaussian-type orbitals density functional method. *Chem. Phys. Lett.* **1996**, *256*, 569−574.
(7) Goh, S. K.; St.-Amant, A. Using a fitted electronic density to improve the efficiency of a linear combination of Gaussian-type orbitals calculation. *Chem. Phys. Lett.* **1997**, *264*, 9−16.
(8) Carbó, R.; Leyda, L.; 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.
(9) Carbó, R.; Domingo, Ll. LCAO-MO Similarity Measures and Taxonomy. *Int. J. Quantum Chem.* **1987**, *23*, 517−545.
(10) Carbó, R.; Calabuig, B. Quantum similarity measures, molecular cloud description, and structure-properties relationships. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 600−606.
(11) Mestres, J.; Solà, M.; Duran, M.; Carbó, R. On the calculation of *ab initio* quantum molecular similarities for large systems: fitting the electron density. *J. Comput. Chem.* **1994**, *15*, 1113−1120.
(12) Carbó, R.; Calabuig, B.; Vera, L.; Besalú, E. Molecular quantum similarity: theoretical framework, ordering principles, and visualization techniques. *Adv. Quantum Chem.* **1994**, *25*, 253−313.
(13) Solà, M.; Mestres, J.; Carbó, R.; Duran, M. Use of *ab initio* quantum molecular similarities as an interpretative tool for the study of chemical reactions. *J. Am. Chem. Soc.* **1994**, *116*, 5909−5915.
(14) Fradera, X.; Amat, L.; Torrent, M.; Mestres, J.; Constans, P.; Besalú, E.; Martí, J.; Simon, S.; Lobato, M.; Oliva, J. M.; Luis, J. M.; Andrés, J. L.; Solà, M.; Carbó, R.; Duran, M. Analysis of the changes on the potential energy surface of Menshutkin reactions induced by external perturbations. *J. Mol. Struct.: TEOCHEM* **1996**, *371*, 171−183.
(15) Besalú, E.; Carbó, R.; Mestres, J.; Solà, M. Foundations and recent developments on molecular quantum similarity. *Top. Curr. Chem.* **1995**, *173*, 31−62.
(16) Constans, P.; Amat, L.; Carbó-Dorca, R. Toward a global maximization of the molecular similarity function: superposition of two molecules. *J. Comput. Chem.* **1997**, *18*, 826−846.
(17) Carbó-Dorca, R. Tagged sets, convex sets and quantum similarity measures. *J. Math. Chem.* **1998**, *23*, 353−364.
(18) Carbó-Dorca, R. On the statistical interpretation of density functions: ASA, convex sets, discrete quantum chemical molecular representations, diagonal vector spaces and related problems. *J. Math. Chem.* **1998**, *23*, 365−375.
(19) Carbó-Dorca, R. Fuzzy sets and Boolean tagged sets; vector semispaces and convex sets; quantum similarity measures and ASA density functions; diagonal vectors spaces and quantum chemistry. In *Advances in Molecular Similarity*; Carbó-Dorca, R., Mezey, P. G., Eds.; JAI Press: Greenwich, CT, 1998; Vol. 2, pp 43−72.
(20) Carbó-Dorca, R.; Besalú, E. A general survey of molecular quantum similarity. *J. Mol. Struct.: THEOCHEM* **1998**, *451*, 11−23.
(21) Constans, P.; Carbó, R. Atomic shell approximation: electron density algorithm restricting coefficients to positive values. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1046−1053.

(22) Constans, P.; Amat, L.; Fradera, X.; Carbó-Dorca, R. Quantum molecular similarity measures (QMSM) and the atomic shell approximation (ASA). In *Advances in Molecular Similarity*; Carbó-Dorca, R., Mezey, P. G., Eds.; JAI Press: Greenwich, CT, 1996; Vol. 1, pp 187−211.
(23) Amat, L.; Carbó-Dorca, R. Quantum similarity measures under atomic shell approximation: first-order density fitting using elementary Jacoby rotations. *J. Comput. Chem.* **1997**, *18*, 2023−2039.
(24) Amat, L.; Carbó-Dorca, R. Fitted electronic density functions from H to Rn for use in quantum similarity measures: *cis*-diamminedichloroplatinum(II) complex as an application example. *J. Comput. Chem.* **1999**, *20*, 911−920.
(25) Carbó-Dorca, R.; Amat, L.; Besalú, E.; Gironés, X.; Robert, D. Quantum molecular similarity: theory and applications to the evaluation of molecular properties, biological activities and toxicity. In *The Fundamentals of Molecular Similarity*; Carbó-Dorca, R., Ed.; Kluwer Academic Press: Dordrecht, The Netherlands; in press.
(26) Carbó-Dorca, R.; Robert, D.; Amat, L.; Gironés, X.; Besalú, E. Molecular quantum similarity in QSAR and drug design. *Lectures and Notes in Chemistry*; Springer: New York; 2000; Vol. 73.
(27) Gironés, X.; Amat, L.; Carbó-Dorca, R. A comparative study of isodensity surfaces using *ab initio* and ASA density functions. *J. Mol. Graphics Modell.* **1998**, *16*, 190−196.
(28) Lobato, M.; Amat, L.; Besalú, E.; Carbó-Dorca, R. Structure−activity relationships of a steroid family using quantum similarity measures and topological quantum similarity indices. *Quant. Struct.-Act. Relat.* **1997**, *16*, 465−472.
(29) Amat, L.; Robert, D.; Besalú, E.; Carbó-Dorca, R. Molecular quantum similarity measures tuned 3D QSAR: an antitumoral family validation study. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 624−631.
(30) Amat, L.; Carbó-Dorca, R.; Ponec, R. Molecular quantum similarity measures as an alternative to log P values in QSAR studies. *J. Comput. Chem.* **1998**, *19*, 1575−1583.
(31) Robert, D.; Amat, L.; Carbó-Dorca, R. Three-dimensional quantitative structure−activity relationships from tuned molecular quantum similarity measures: prediction of the corticosteroid-binding globulin binding affinity for a steroid family. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 333−344.
(32) Robert, D.; Gironés, X.; Carbó-Dorca, R. Facet diagrams for quantum similarity data. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 597−610.
(33) Robert, D.; Carbó-Dorca, R. Aromatic compounds aquatic toxicity QSAR using quantum similarity measures. *SAR QSAR Environ. Res.* **1999**, *10*, 401−422.
(34) Robert, D.; Gironés, X.; Carbó-Dorca, R. Quantification of the influence of single-point mutations on Haloalkane Dehalogenase activity: a molecular quantum similarity study. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 839−846.
(35) Ponec, R.; Amat, L.; Carbó-Dorca, R. Molecular basis of quantitative structure-properties relationships (QSPR): a quantum similarity approach. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 259−270.
(36) Ponec, R.; Amat, L.; Carbó-Dorca, R. Quantum similarity approach to LFER: substituent and solvent effects on the acidities of carboxylic acids. *J. Phys. Org. Chem.* **1999**, *12*, 447−454.
(37) Amat, L.; Carbó-Dorca, R.; Ponec, R. Simple linear QSAR models based on quantum similarity measures. *J. Med. Chem.* **1999**, *42*, 5169−5180.
(38) Jacobi, C. G. J. Über ein leichtes Verfahren die in der theorie der Säcularstörungen vorkommenden gleichungen numerisch aufzulösen. *J. Reine Angew. Math.* **1846**, *30*, 51−94.
(39) Ruedenberg, K.; Schwarz, W. H. E. Nonspherical atomic ground-state densities and chemical deformation densities from X-ray scattering. *J. Chem. Phys.* **1990**, *92*, 4956−4969.
(40) ASA coefficients and exponents for different fitted atomic basis functions can be downloaded from the WWW site http://iqc.udg.es/cat/similarity/ASA/basisset.html.
(41) Miller, K. J.; Ruedenberg, K. Electron correlation and separated-pair approximation. An application to Berylliumlike atomic systems. *J. Chem. Phys.* **1968**, *48*, 3414−3443.
(42) Silver, D. M.; Mehler, E. L.; Ruedenberg, K. Electron correlation and separated pair approximation in diatomic molecules. I. Theory. *J. Chem. Phys.* **1970**, *52*, 1174−1180.
(43) Mehler, E. L.; Ruedenberg, K.; Silver, D. M. Electron correlation and separated pair approximation in diatomic molecules. II. Lithium hydride and boron hydride. *J. Chem. Phys.* **1970**, *52*, 1181−1205.
(44) Silver, D. M.; Ruedenberg, K.; Mehler, E. L. Electron correlation and separated pair approximation in diatomic molecules. III. Imidogen. *J. Chem. Phys.* **1970**, *52*, 1206−1227.
(45) Carbó, R.; Domingo, Ll.; Peris, J. J. Elementary unitary MO transformations and SCF theory. *Adv. Quantum Chem.* **1982**, *15*, 215−265.
(46) Carbó, R.; Domingo, Ll.; Peris, J. J.; Novoa, J. J. Energy variation and elementary Jacobi rotations. *J. Mol. Struct.: THEOCHEM* **1983**, *93*, 15−33.
(47) Carbó, R.; Domingo, Ll.; Novoa, J. J. Multiconfigurational calculations using elementary Jacobi rotations. *J. Mol. Struct.: THEOCHEM* **1985**, *120*, 357−363.
(48) Carbó, R.; Miró, J.; Domingo, Ll.; Novoa, J. J. Jacobi rotations: a general procedure for electronic energy optimization. *Adv. Quantum Chem.* **1989**, *20*, 375−441.
(49) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72*, 650−654.
(50) McLean, A. D.; Chandler, G. S. Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z=11−18. *J. Chem. Phys.* **1980**, *72*, 5639−5648.
(51) Carbó-Dorca, R. ATOMIC Program 1995, based on Roos, B.; Salez, C.; Veillard, A.; Clementi, E. A general program for calculation of SCF orbitals by the expansion method. IBM Research/RJ518(No. 10901), 1968.
(52) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, Jr., J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. GAUSSIAN 98, Revision A.6; Gaussian, Inc.: Pittsburgh, PA, 1998.
(53) Bonaccorsi, R.; Scrocco, E.; Tomasi, J. Molecular SCF calculations for the ground state of some three-membered ring molecules: $(CH_2)_3$, $(CH_2)_2NH$, $(CH_2)_2NH_2^+$, $(CH_2)_2O$, $(CH_2)_2S$, $(CH)_2CH_2$, and N_2CH_2. *J. Chem. Phys.* **1970**, *52*, 5270−5284.
(54) See for example: Campanario, J. M.; Bronchalo, E.; Hidalgo, M. A. An effective approach for teaching intermolecular interactions. *J. Chem. Educ.* **1994**, *71*, 761−766.
(55) Winkler, D. A.; Holan, G. Design of potential anti-HIV agents. 1. Mannosidase inhibitors. *J. Med. Chem.* **1989**, *32*, 2084−2089.
(56) Winkler, D. A. Molecular modeling studies of "Flap Up" mannosyl cation mimics. *J. Med. Chem.* **1996**, *39*, 4332−4334.
(57) Knapp, S.; Murali Dhar, T. G. Synthesis of the mannosidase II inhibitor mannostatin A. *J. Org. Chem.* **1991**, *56*, 4096−4097.
(58) King, S. B.; Ganem, B. Synthetic studies on mannostatin A and its derivatives: a new family of glycoprotein processing inhibitors. *J. Am. Chem. Soc.* **1994**, *116*, 562−570.
(59) Farr, R. A.; Peet, N. P.; Kang, M. S. Synthesis of 1S, 2R, 3S, 4R, 5R-methyl[2,3,4-trihydroxy-5-(hydroxymethyl)cyclopentyl]amine: a potent α-mannosidase inhibitor. *Tetrahedron Lett.* **1990**, *31*, 7109−7112.
(60) Farr, R. A.; Holland, A. K.; Huber, E. W.; Peet, N. P.; Weintraub, P. M. Pyrrolidine and hexahydro-1*H*-azepine mimics of the 'flap up' mannosyl cation. *Tetrahedron Lett.* **1994**, *50*, 1033−1044.

CI0000272