Modeling Large Macromolecular Structures Using Promolecular Densities

Xavier Gironés, Lluís Amat, and Ramon Carbó-Dorca*

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

Received December 20, 2001

A procedure to easily construct fitted density functions is presented. This methodology, based on the promolecule approach, is able to handle large macromolecular systems, such as proteins. The usual procedure dealing with fitted densities has been improved by adding some restrictions, which allow faster calculations. As a main application example, molecular isodensity contours (MIDCOs) are constructed for two proteins, one of them composed of more than 50 000 atoms. MIDCOs, as a visual representation of the molecular density function, and thus an important descriptor of the molecular charge distribution, constitute a powerful tool in the understanding of molecular systems. MIDCOs are presented for both proteins, allowing exploration of their surfaces, as well as analysis of their shapes. Also, as a quantum mechanical calculation example, molecular quantum self-similarity measures are calculated for several proteins.

1. INTRODUCTION

An important tool for the description of three-dimensional (3D) electronic density functions (DF) is the representation of molecular isodensity contours (MIDCOs), which characterize the molecular shape in a detailed and realistic way. First defined by Mezey,[1−6] MIDCO generation and analysis, which inherently include molecular surface areas and occupied volumes, provide valuable information that eases the understanding of a molecular system; they also assist in performing molecular recognition, drug−receptor interactions, or bulk physical property prediction studies. Due to their importance, several studies have been carried out on MIDCOs and shape analysis,[7−16] leading to the holographic electron-density theorem.[17]

Given a formal nuclear configuration, and assuming a 3D molecular coordinate system, the electronic DF, $\rho(\mathbf{r})$, is a function of the 3D position variable $\mathbf{r}$. Within this coordinate system, $\rho(\mathbf{r})$ assigns a density value at each point $\mathbf{r}$ of the 3D space. This charge DF can be experimentally determined[18,19] or theoretically calculated using an appropriate quantum chemical method provided by, for example, available quantum chemistry software.[20,21]

To faithfully visualize detailed shape features, theoretical models are the most affordable available methodology, due to experimental data from X-ray diffraction,[22] complemented by aspherical atom libraries,[23] is still not accurate enough to provide a reasonable description of the outer molecular region. Since molecular properties and reactivity, as well as biological responses, are strongly connected to the external electronic density, rather than to the core one, it seems natural that an accurate representation should correctly model it. Besides $\rho(\mathbf{r})$-based MIDCOs, other usual methodologies have been described, such as fused sphere models, solvent-accessible surfaces, and motif patterns; however, they only yield to rough and incomplete representations of the molecular shape features.[24−28] MIDCOs, as founded in quantum mechanical ideas, reflect the actual fuzzy characteristics of the molecular shape given by $\rho(\mathbf{r})$; however, one of the major problems in ab initio calculations is the ability to quickly generate these $\rho(\mathbf{r})$ values for large macromolecules, so easy for small molecules.

Ordinary ab initio techniques, with assorted procedures and a wide selection of basis sets,[29] are able to provide accurate $\rho(\mathbf{r})$ values. However, when the studied systems consist of about 150 atoms or more, the computational requirements become unaffordable, with the exception of periodic systems. Many workarounds have been formulated within ab initio techniques to both increase the range of application and reduce computational costs. Between these methodologies, the most relevant are semiempirical methods,[29] which avoid many intermediate calculations approximating with experimental or theoretical data, thus providing good quality $\rho(\mathbf{r})$ values within appealing time limits for those systems where accurate parametrization exists, and linear scaling density functional theory (DFT) methods,[29,30] which sacrifice part of the integral calculation and use approximate algorithms in order to speed up the whole process. However, when dealing with very large macromolecular systems, these procedures are extremely time-consuming, even in present-day multiprocessor hardware.

Several procedures have been developed in order to circumvent the large computational requirements of treating these systems. The molecular electron density approach (MEDLA),[7−10] developed by Walker and Mezey, was the first method that allowed constructing inexpensive and accurate electron densities for large molecules. MEDLA, based on a simple fragment additivity principle, consists of constructing fragment densities, stored in a databank, which can be used as building blocks for building DF for larger molecules by selecting and arranging the interpenetrating fuzzy charge clouds of previously computed density fragments. Other approaches by the same group were also reported.[13,31]

* Corresponding author phone: +34 972418367; fax: +34 972418356; e-mail: quantum@iqc.udg.es.

10.1021/ci010348t CCC: $22.00
Published on Web 05/17/2002

Another possibility is the use of promolecular densities[32−35] as primary building blocks instead of genuine ab initio densities. Promolecular densities are simplified models designed to emulate an ab initio DF, considering that the whole molecular DF is made from discrete atomic, noninteracting, contributions. This procedure, along with the use of lower angular momentum functions, such as spherical 1*S* Gaussian functions that fit the original atomic ab initio DF, allows speeding considerably up any related calculation, such as the necessary grid of $\rho(\mathbf{r})$ around the molecule, needed to represent MIDCOs. Within this scope, detailed MIDCOs have been obtained, indeed for large molecules,[14−16] and it is general enough to permit the inclusion of any element of the periodic table. A related approach, developed to replace hard sphere models, is the Gaussian method for molecular shape[36,37] by Pickup and co-workers, where Gaussian functions are fitted to emulate molecular arrays of van der Waals hard spheres, but not electronic density functions.

Even if promolecular approaches save much computational time, they are still unable to treat large macromolecular structures within reasonable time limits. The intention of this work is to provide a new insight by adding shortcuts when evaluating $\rho(\mathbf{r})$, so the calculation becomes a linear process of the number of atoms present in the molecule instead of a cubic one. In addition, these cutoffs can also be applied when performing some related quantum chemical computations, such as molecular quantum similarity measures (MQSM),[38−44] thus extending the range of application of these methods.

The present work is structured as follows: first, the methodological section exposes the basis of promolecular densities, MIDCO generation, and a brief explanation of MQSM. An application example, dealing with two proteins—proto-oncogene tyrosine kinase,[45] a small protein composed of 872 atoms, and bovine heart cytochrome *c* oxidase,[46] a larger enzyme composed of 56 863 atoms—is presented, in which MIDCOs and similarity measures are performed. Finally, concluding remarks are given.

2. METHODOLOGY

Promolecular Atomic Shell Approximation. To avoid expensive theoretical calculations to obtain electronic DF, the promolecular atomic shell approximation (PASA)[47−49] can be used to simplify such procedures. This approach considers the molecular density as a sum of discrete atomic density contributions, which are taken as 1*S* Gaussian functions and fitted to previously computed atomic ab initio ones. Promolecular ASA densities are mathematically expressed as

$$\rho_{\mathrm{A}}{}^{\mathrm{ASA}}(\mathbf{r}) = \sum_{a\in \mathrm{A}} P_a \rho_a{}^{\mathrm{ASA}}(\mathbf{r}-\mathbf{R}_a) \qquad (1)$$

where P_a is the atomic number of the ath atom. In this way, the molecular density is a simple addition of atomic densities, $\rho_a{}^{\mathrm{ASA}}$, centered at the atomic positions $\mathbf{R}_a$. These atomic densities can be expressed in turn as

$$\rho_a{}^{\mathrm{ASA}}(\mathbf{r}-\mathbf{R}_a) = \sum_{i\in a} w_i |S_i(\mathbf{r}-\mathbf{R}_a)|^2 \qquad (2)$$

where S_i are the normalized spherical 1*S* Gaussian functions and w_i are the coefficients of the linear expansion, which must fulfill the following convexity conditions:

$$w_i \in (0,1],\ \forall\ \boldsymbol{i} \in a \text{ and } \sum_{i\in a} w_i = 1 \qquad (3)$$

to guarantee that the DF is positive definite in its whole range and to normalize the atomic DF in order to ensure that the integral of $\rho_{\mathrm{A}}{}^{\mathrm{ASA}}$ over all space yields the number of electrons of molecule A, thus ensuring the proper definition of the charge distribution function. Once the overall atomic densities are built, each molecular density is constructed by adding, as appropriate, these atomic density elementary pieces.

An improvement can be introduced considering the promolecular nature. In earlier works[14,15] a set of 1*S* Gaussian functions was assigned to each atom forming a molecule. Even if it is of minor importance for small molecules, as the molecular system grows larger, an incredible amount of 1*S* functions may be needed. As an illustrative example, a protein like gene V (PDB code: 1VQB), composed of 1362 atoms, would require about 2700 functions.[15] However, a simple analysis of this protein reveals that only five different atom types are present, namely, hydrogen, carbon, nitrogen, oxygen, and sulfur. Taking into account that in the promolecule each atomic type is described by a unique set of 1*S* Gaussian functions, there is no need to place in memory a repetitive number of copies of the same set for the same type of atom. In this way, gene V protein would have been described only by four sets of 1*S* Gaussian functions (one for H, three for C, N, and O, and four for S), with a total number of 14 functions.

At this point, it needs to be stated that promolecular DF constructed in this way, even fitted, constitute themselves a DF, which may be used in different quantum chemical applications, such as MQSM,[38−44] which are explained later in this section. Due to its simple formulation, PASA DF may not be suitable, or accurate enough, for all kind of calculations; however, in those applicable, the computational performance increase is outstanding compared to the usage of ab initio densities.

Molecular Isodensity Contours (MIDCOs) Using Promolecular Densities. Present-day DF representations of MIDCOs[2], based on the early orbital representations of Mulliken,[50−54] are constructed defining a grid enveloping the studied molecule, whose DF has been previously constructed, and evaluating this DF in each grid point. Needless to say, the grid must be large and fine enough to provide complete and smooth representations of the different isodensity levels.

Even if $\rho(\mathbf{r})$ is defined in the entire 3D space and only becomes zero at infinite distance from the nuclei, this function converges rapidly to zero already at short distances, becoming imperceptible, and thus computationally negligible, at about 5 atomic units (au) from the nearest nucleus.[2] Assuming this fact, and considering the structure of promolecular densities, it seems reasonable that the contribution of an atomic DF does not need to be computed further than 5 au from the respective atomic location.

Therefore, the proposed simplified flow chart for the generation of PASA MIDCOs can be written as follows:

(a) Input atomic coordinates.

(b) Retrieve how many different atomic types are present and allocate the necessary sets of PASA functions.

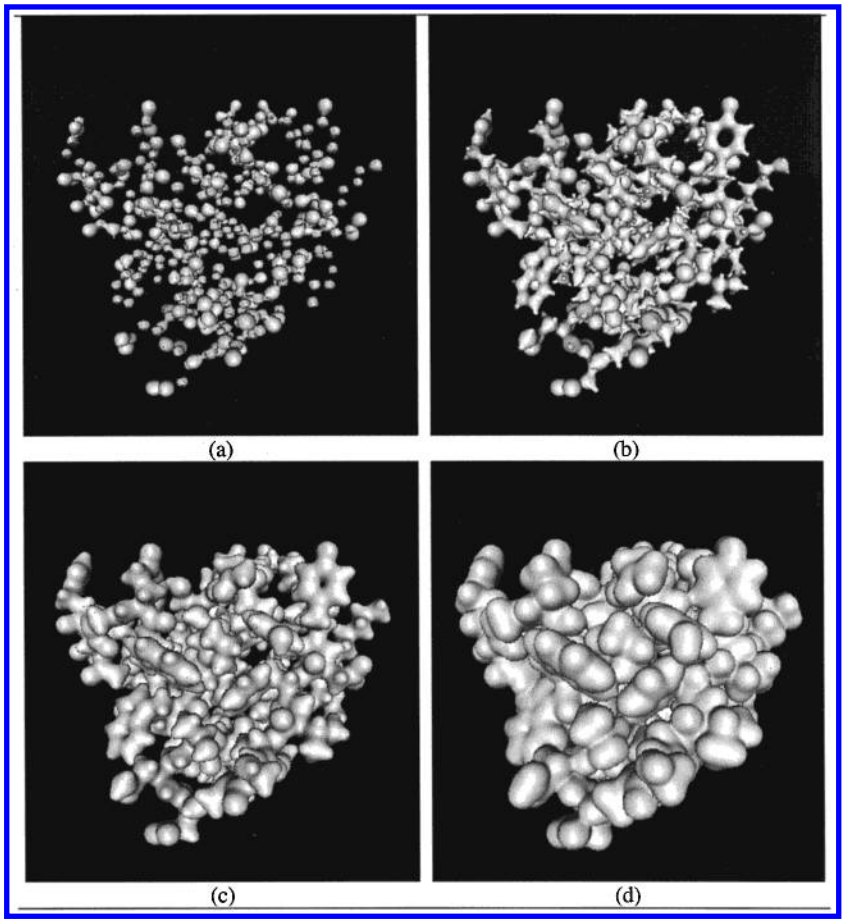

Figure 1. MIDCOs of protein 1ABL at four isodensity levels: (a) 0.250, (b) 0.150, (c) 0.050, and (d) 0.005 (in e/au^3).

(c) Define the molecular grid according to molecular size and desired resolution. All grid points are set initially to zero.

(d) For each atom, evaluate the addressed functions in a (10 au)^3 box surrounding the atom.

(e) Download the grid box to disk using a suitable file format to be imported by any graphics application.

As seen in step d, it has been considered that the contribution of an atom to a point located further than 5 au is practically zero; however, all relevant interactions are contained within the defined box, such as the bonding pattern and outer electronic distribution. This approximation sets the basis of the linear dependence with the number of atoms of the proposed algorithm. This procedure has been coded in Fortran 90[55] and used in the forthcoming examples.

Molecular Quantum Similarity Measures. Given a pair of molecular structures, a possible way to determine their resemblance, accounted for by their native quantum character, can be done using molecular quantum similarity measures (MQSM).[38−44] This methodology states the degree of similarity between two objects by means of their electronic distributions. The simplest MQSM, also known as overlap-like MQSM, is mathematically expressed as

$$Z_{AB} = \int \rho_A(\mathbf{r})\, \rho_B(\mathbf{r})\, d\mathbf{r} \qquad (4)$$

where $\rho_A(\mathbf{r})$ and $\rho_B(\mathbf{r})$ are the respective DF of molecules A and B, and Z_{AB} is the resulting value of the MQSM between A and B, which takes into account the shared amount of density between both structures. In the particular case when A is equal to B, a molecular quantum self-similarity measure (MQS-SM) is defined, accounting for the concentration of the charge in the studied molecular system.

The major use of promolecular ASA DF has been devoted to computing MQSM because its simplicity yields high computational performance and considerable time reductions compared to MQSM calculation based on ab initio densities.[14] Since it was proved that these measures differed by less than 2% from ab initio densities,[14,47] their use was clearly justified in this way.

The computational cost of computing MQSM scales quadratically with the number of functions used, which is roughly proportional to the number of atoms. Even if this quadratic behavior cannot be linearized, the calculation of MQSM can also benefit from a cutoff in order to improve the computational performance when calculating the similarity values. The proposed approximation consists of including only in the calculation the contribution of those functions that are close in space. For this purpose, the same distance threshold of 5 au is also applied.

3. RESULTS AND DISCUSSION

As commented in the Introduction, MIDCOs and MQS-SM will be calculated for two macromolecules using PASA DF. The featuring proteins are proto-oncogene tyrosine kinase (PDB code: 1ABL)[45] and bovine heart cytochrome *c* oxidase (PDB code: 1OCO).[46] 1ABL, which is found in humans, is involved in the transference of phosphorate groups, and its structure was elucidated by theoretical methods. This protein is formed by 872 atoms, but only five

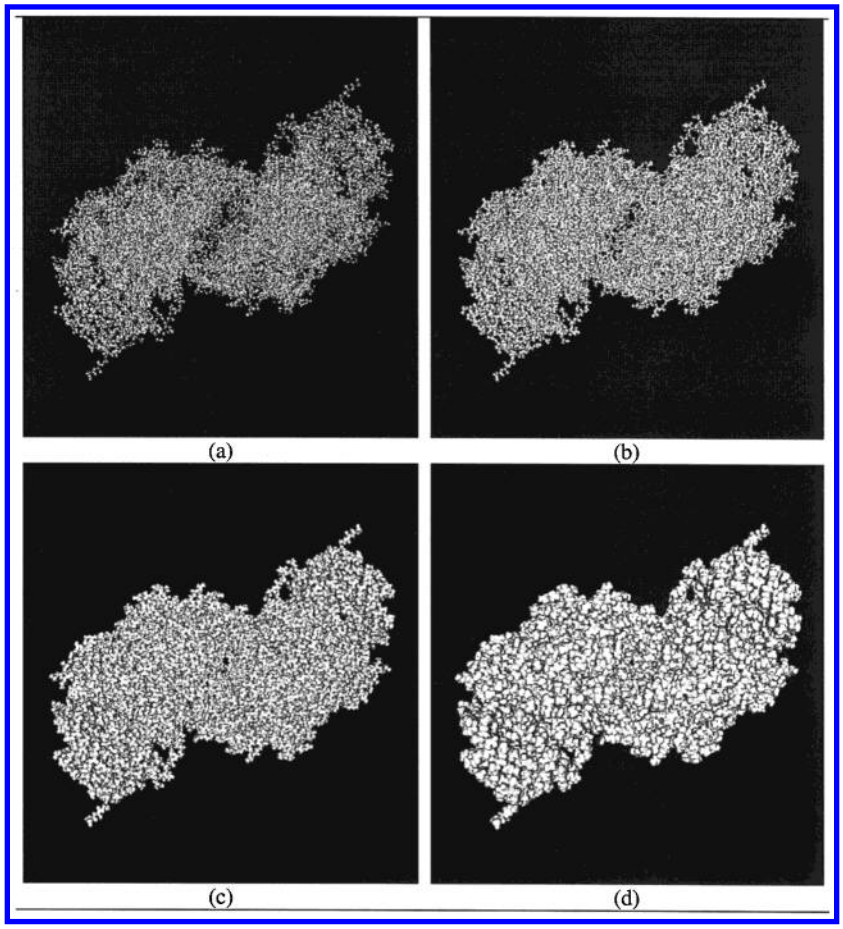

Figure 2. MIDCOs of protein 1OCO at four isodensity levels: (a) 0.250, (b) 0.150, (c) 0.050, and (d) 0.005 (in e/au^3).

different kinds (H, C, N, O, S). On the other hand, 1OCO controls the last step of nutrient molecule oxidation by transferring their electrons to oxygen molecules, thus catalyzing the reaction of oxygen to water and generating energy to later build ATP. This enzyme, which contains metals in a prosthetic group located inside a cavity, is composed of 56863 atoms, but only 10 different types (H, C, N, O, Na, Mg, S, Fe, Cu, Zn).

Molecular Modeling. The molecular geometries were obtained from the Protein Data Bank (PDB).[56] 1ABL, as theoretically derived, was ready for the forthcoming calculations. 1OCO contained solvating waters, which were previously removed; it also lacked hydrogens in its structure, and those were later added using Babel software.[57] PASA DF for both molecules were constructed using previously computed fitting parameters to atomic Hartree−Fock calculations using 6-311G basis set.[49] All these steps, as well as the forthcoming calculations, were carried out on a Intel Pentium III 600 MHz computer with 256 MB RAM running Linux Red Hat 6.2 operating system.

MIDCOs. As a graphical example, some isodensity contours plots are presented in Figures 1 and 2, for 1ABL and 1OCO, respectively, at different density threshold values. 1ABL has been plotted at 0.5 au resolution, whereas 1OCO was plotted at 1 au, due to its larger size, but still providing an acceptable representation. As seen from the plots, as the isodensity level decreases, the following shape characteristics[5] are outlined:

At high values (0.250 e/au^3 and over) only atomic locations and prebonding patterns are present. These plots are in the *localized density range*, which mostly exhibits those density domains collapsed around nuclei or delimiting a well-defined functional group, such as OH or COOH.

At intermediate values (0.150 and 0.050 e/au^3), the prominent density domains are those defining the whole molecular bonding range, as well as outlining a skinny representation of the molecule, similar to other kinds of molecular surface representations, such as solvent-accessible surfaces.[25] At these levels, the volumetric features of density function, as well as initial indications of the occupied space, are evinced.

At low isodensity levels (0.005 e/au^3 and below), effective occupied space and the volumetric characteristics are fully exhibited. At these levels, the unique density domain corresponds only to the full molecule, being only the outer regions visible. A further decrease in the isovalue would yield a quasi-spherical domain, whose edge would be already far from the nuclei.

It is also interesting to consider, regardless of the large number of atoms treated, the time needed to complete the calculation, as well as the dimensions of the grid used. Table 1 lists this information for the previous proteins, as well as for two more, whose size is intermediate between the graphically represented.

From Table 1, a sound relationship between the number of atoms present and the necessary time to compute the grid can be observed. The listed times are not only affordable in present-day computers, but also are very small if the large number of atoms present in the structure are taken into account. As it was commented before, no other methodology

Table 1. Dimension of the Molecular Boxes, Resolution Used, and Time Required to Compute MIDCOs for a Series of Macromolecular Structures[a]

structure	no. of atoms	dimensions of mol box (au)	resolution (au)	time (s)
1ABL	872	65 × 71 × 63	0.5	18
1RVA	8 686	129 × 121 × 147	1.0	24
1QLE	21 358	275 × 237 × 177	1.0	83
1OCO	56 863	299 × 245 × 245	1.0	177

[a] Resolution stands for spacing between points in the grid. 1RVA corresponds to *Eco rv* endonuclease and 1QLE to *Paracoccus denitrificans* cytochrome *c* oxidase. Both structures have been identically treated as 1OCO.

Table 2. Results of MQS-SM (au) Calculations for the Proteins[a]

structure	no. of atoms	MQS-SM (unrestricted)	time (s)	MQS-SM (restricted)	time (s)
1ABL	872	20 994.86	45.76	20 993.00	4.88
1RVA	8 686	218 541.39	4 701.3	218 531.17	475.30
1QLE	21 358	551 956.91	40 200	551 929.00	3 044.8
1OCO	56 863	1 461 628.55	282 480	1 461 626.52	3 283.9

[a] The unrestricted values correspond to the full calculation, whereas the restricted ones correspond to those where the distance threshold has been applied.

has been able to treat, or even construct, DF of such magnitude. Even if the PASA DF consist of functions fitted to ab initio ones, they constitute a suitable starting point for simple calculations involving these vast molecular systems, as it is demonstrated in the next section, where MQS-SMs are computed for these structures.

MQS-SM. A quantum chemical application using PASA DF calculations of MQS-SM will be presented. As previously commented in the Methodology section, a distance threshold of 5 au is also applied, so two functions further than this threshold will not contribute to the measure. However, and for comparative purposes, unrestricted calculations are also performed, so that the time savings can be valued. The results are given in Table 2.

Perusing Table 2 manifests a rough quadratic dependence between the number of atoms and the values of MQS-SM in both approaches. Even if unrestricted MQS-SM values are still affordable in present-day hardware, the application of the distance restriction considerably reduces the calculation time, practically without any loss of precision. As seen in Table 2, differences between both kind of values are below 0.005% in all cases, pointing out the usefulness of the proposed approximation.

According to the definition of MQS-SM, if a quantum object is to be compared to itself, there is no need for structural superposition; however, if the more general case is to be handled, the value of the similarity measure strongly depends on the relative position of both analyzed objects in space.[47,58] Hence, a superposition procedure, as well as repeated samples of the measures, needs to be computed. As quantum similarity measures require intensive calculation, it is interesting to have the possibility to compute them faster within an acceptable precision.

4. CONCLUSIONS

The practical use of promolecular densities derived from atomic shell approximation applied to the construction of density functions of large macromolecular systems has been demonstrated. These fitted densities have been employed to construct molecular isodensity contours of these large systems with both affordable computational and time requirements. As previously demonstrated, molecular surfaces derived from promolecular densities do not significantly differ from those derived from ab initio ones, but decreasing the amount of time and computational effort required to both works to build the density, instantly done within PASA DF, and to construct the MIDCOs. Also, the proposed distance restriction allows turning density evaluation in a linear process dependent only on the number of atoms composing the molecular system.

Even if a fitted representation is used, PASA DF are themselves a kind of DF that truly emulate a higher order DF. In this way, they can be included in some quantum mechanical calculations. As demonstrated, they can provide quantum similarity measures, even for large macromolecular systems, composed of up to 50 000 atoms or more, within reasonable time limits. Furthermore, if the proposed distance restriction is applied, the similarity calculation is still more appealing. Similarity measures computed with distance restrictions have been demonstrated to differ less than 0.005% from those calculated unrestricted.

ACKNOWLEDGMENT

This research has been supported by Investigation Project No. SAF2000-0223-C03-01 from the Spanish Ministerio de Ciencia y Tecnología. X.G. benefited from a predoctoral fellowship from the University of Girona. Financial support from the Fundació Maria Francisca de Roviralta is also acknowledged.

REFERENCES AND NOTES

(1) Mezey, P. G. Molecular Surfaces. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1990; Vol. 1, pp 265−294.
(2) Mezey, P. G. The Topology of Molecular Surfaces and Shape Graphs. In *Computational Chemical Graph Theory*; Rouvray, D. H., Ed.; Nova: New York, 1990; pp 175−197.
(3) Mezey, P. G. Non-Visual Molecular Shape Analysis: Shape Changes in Electronic Excitations and Chemical Reactions. In *Computational Advances in Organic Chemistry (Molecular Structure and Reactivity)*; Ogretir, C., Csizmadia, I., Eds.; Kluwer Academic Publishers: Dordrecht, 1991; pp 261−288.
(4) Arteca, G. A.; Mezey, P. G. Algebraic Approaches to the Shape Analysis of Biological Macromolecules. In *Computational Chemistry*: *Structure, Interactions and Reactivity*; Fraga, S., Ed.; Elsevier: Amsterdam, 1992; Part A, pp 463−487.
(5) Mezey, P. G. *Shape in Chemistry*: *An Introduction to Molecular Shape and Topology*; VCH Publishers: New York, 1993.
(6) Mezey, P. G. Shape analysis. In *Encyclopedia of computational chemistry*; Schleyer, P.v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, UK, 1998; Vol. 4, pp 2582−2589.
(7) Walker, P. D.; Mezey, P. G. Molecular electron density lego approach to molecule building. *J. Am. Chem. Soc.* **1993**, *115*, 12423−12430.
(8) Walker, P. D.; Mezey, P. G. Ab initio quality electron densities for proteins: A MEDLA approach. *J. Am. Chem. Soc.* **1994**, *116*, 12022−12032.
(9) Walker, P. D.; Maggiora, G. M.; Johnson, M. A.; Petke, J. D.; Mezey, P. G. Shape Group Analysis of Molecular Similarity: Shape Similarity of Six-Membered Aromatic Rings Systems. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 568−578.
(10) Walker, P. D.; Mezey, P. G. Realistic, detailed images of proteins and tertiary structure elements: ab initio quality electron density calculations for bovine insulin. *Can. J. Chem.* **1994**, *72*, 2531−2536.
(11) Mezey, P. G. Macromolecular density matrices and electron densities with adjustable nuclear geometries. *J. Math. Chem.* **1995**, *18*, 141−168.

(12) Walker, P. G.; Mezey, P. G. A new computational microscope for molecules: High-resolution MEDLA images of taxol and HIV-1 protease, using additive electron density fragmentation principles. *J. Math. Chem.* **1995**, *17*, 203−234.
(13) Mezey, P. G. Quantum chemistry of macromolecular shape. *Int. Rev. Phys. Chem.* **1997**, *16*, 361−388.
(14) Gironés, X.; Amat, L.; Carbó-Dorca, R. A comparative study of isodensity surfaces using ab initio and ASA density functions. *J. Mol. Graph. Model.* **1998**, *16*, 190−196.
(15) Gironés, X.; Carbó-Dorca, R.; Mezey, P. G. Applications of promolecular ASA densities to graphical representation of density functions of macromolecular systems. *J. Mol. Graph. Model.* **2001**, *19*, 343−348.
(16) Mitchell, A. S.; Spackman, M. A. Molecular Surfaces from the Promolecule: A Comparison with Hartree−Fock Ab Initio Electron Density Surfaces. *J. Comput. Chem.* **2000**, *21*, 933−942.
(17) Mezey, P. G. The holographic density theorem and quantum similarity measures. *Mol. Phys.* **1999**, *96*, 169−178.
(18) *Electron Distribution and the Chemical Bond*; Coppens, P., Hall, M. B., Eds.; Plenum: New York and London, 1982.
(19) Fliszár, S. *Charge Distributions and Chemical Effects*; Springer: New York, 1983.
(20) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, 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, M. T.; Al-Laham, 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.
(21) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(22) Coppens, P. *X-ray Charge Densities and Chemical Bonding*; Oxford: NewYork, 1997.
(23) *International Tables for Crystallography*; Rossman, M. G., Arnold, E., Eds.; Kluwer Academic Publishers: Dordrecht, 2001; Vol. F.
(24) Corey, R. B.; Pauling, L. Molecular models of amino acids, peptides and proteins. *Rev. Sci. Instrum.* **1953**, *24*, 621−627.
(25) Lee, B.; Richards, F. M. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **1971**, *55*, 379−400.
(26) Richardson, J. S. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **1981**, *34*, 167−339.
(27) Agishtein, M. E. Fuzzy molecular surfaces. *J. Biomol. Struct. Dyn.* **1992**, *9 (4)*, 759−768.
(28) Connolly, M. L. In *Network Science*; 1996 (http://www.netsci.org/Science/Compchem/feature14.html).
(29) Foresman, J. B.; Frisch, Æ. *Exploring Chemistry with Electronic Structure Methods*; Gaussian, Inc.: Pittsburgh, 1996.
(30) Scuseria, G. E. Linear Scaling Density Functional Calculations with Gaussian Orbitals. *J. Phys. Chem. A* **1999**, *103*, 4782−4790.
(31) Mezey, P. G. Quantum similarity measures and Löwdin's transform for approximate density matrices and macromolecular forces. *Int. J. Quantum Chem.* **1997**, *63*, 39−48.
(32) Hirshfeld, F. L.; Rzotkiewicz, S. Electrostatic Binding in First-Row Ah and A2 Diatomic-Molecules. *Mol. Phys.* **1974**, *27*, 1319−1343.
(33) Spackman, M. A.; Maslen, E. N. Chemical properties from the promolecule. *J. Phys. Chem.* **1986**, *90*, 2020−2027.
(34) Ruedenberg, K.; Schwarz, W. H. E. Nonspherical Atom Densities and Chemical Deformation Densities from X-ray Scattering. *J. Chem. Phys.* **1990**, *92*, 4956−4969.
(35) Coppens, P. The Structure Factor. In *International Tables for Crystallography*; Shmueli, U., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1996; Vol. B, pp 10−22.
(36) Grant, J. A.; Gallardo, M. A.; Pickup, B. T. A Fast Method of Molecular Shape Comparison A simple application of a Gaussian description of molecular shape. *J. Comput. Chem.* **1996**, *17*, 1653−1666.
(37) Grant, J. A.; Pickup, B. T. A Gaussian description of molecular shape. *J. Phys. Chem.* **1996**, *100*, 2456−2456.
(38) Carbó, R.; Arnau, J.; Leyda, L. 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.
(39) Carbó-Dorca, R.; Besalú, E. A general survey of molecular quantum similarity. *THEOCHEM* **1998**, *451*, 11−23.
(40) Carbó-Dorca, R.; Amat, L.; Besalú, E.; Lobato, M. Quantum Similarity. In *Advances in Molecular Similarity*; Carbó-Dorca, R., Mezey, P. G., Eds.; JAI Press: Greenwich, CT, 1998; Vol. 2, pp 1−42.
(41) Carbó-Dorca, R. Fuzzy sets and Boolean tagged sets; vector semispaces and convex sets; quantum similarity measures and ASA density functions; diagonal vector 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.
(42) Besalú, E.; Carbó, R.; Mestres, J.; Solà, M. Foundations and recent developments on molecular quantum similarity. *Top. Curr. Chem.* **1995**, *173*, 31−62.
(43) Carbó, R.; Besalú, E.; Theoretical foundations of quantum molecular similarity. In *Molecular similarity and reactivity*: *from quantum chemical to phenomenological approaches*; Carbó, R., Ed.; Kluwer: Amsterdam, 1995; pp 3−30.
(44) Carbó-Dorca, R.; Besalú, E.; Amat, L.; Fradera, X. Quantum molecular similarity measures: concepts, definitions and applications to quantitative structure-properties relationships. In *Advances in molecular similarity*; Carbó-Dorca, R., Mezey, P. G., Eds.; JAI Press: Greenwich, CT, 1996; Vol. 1, pp 1−42.
(45) Pisabarro, M. T.; Ortiz, A. R.; Serrano, L.; Wade, R. C. Homology modeling of the Abl-SH3 domain. *Proteins* **1994**, *20*, 203-215.
(46) Yoshikawa, S.; Shinzawa-Itoh, K.; Nakashima, R.; Yaono, R.; Yamashita, E.; Inoue, N.; Yao, M.; Fei, M. J.; Libeu, C. P.; Mizushima, T.; Yamaguchi, H.; Tomizaki, T.; Tsukihara, T.; Redox-Coupled Crystal Structural Changes in Bovine Heart Cytochrome C Oxidase. *Science* **1998**, *280*, 1723−1729.
(47) Amat, L.; Carbó-Dorca R. Quantum Similarity Measures under Atomic Shell Approximation: First-Order Density Fitting using Elementary Jacobi Rotations. *J. Comput. Chem.* **1997**, *18*, 2023−2039.
(48) Amat, L.; Carbó-Dorca, R. Fitted Electronic Density Functions from H to Rn for Use in Quantum Similarity Measures: cis-Diammine-Dichloroplatinium(II) Complex as an Application Example. *J. Comput. Chem.* **1999**, *20*, 911−920.
(49) Amat, L.; Carbó-Dorca, R. Molecular Electronic Density Fitting Using Elementary Jacobi Rotations under Atomic Shell Approximation. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1188−1198.
(50) Mulliken, R. S. Electronic Population Analysis on LCAO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23*, 1833−1840.
(51) Mulliken, R. S. Electronic Population Analysis on LCAO Molecular Wave Functions. II. Overlap Populations, Bond Orders, and Covalent Bond Energies. *J. Chem. Phys.* **1955**, *23*, 1841−1846.
(52) Mulliken, R. S. Electronic Population Analysis on LCAO Molecular Wave Functions. III. Effects of Hybridization on Overlap and Gross AO Populations. *J. Chem. Phys.* **1955**, *23*, 2338−2342.
(53) Mulliken, R. S. Electronic Population Analysis on LCAO Molecular Wave Functions. IV. Bonding and Antibonding in LCAO and Valence-Bond Theories. *J. Chem. Phys.* **1955**, *23*, 2343−2346.
(54) Mulliken, R. S. Criteria for the Construction of Good Self-Consistent-Field Molecular Orbital Wave Functions, and the Significance of LCAO-MO Population Analysis. *J. Chem. Phys.* **1962**, *36*, 3428−3439.
(55) Gironés, X.; Amat, L.; Carbó-Dorca, R. *ASACUBE*, v1.0; Institute of Computational Chemistry: University of Girona, Girona, 2000.
(56) Protein Data Bank, Research Collaboratory for Structural Bioinformatics (http://www.rcsb.org/pdb). Main Citation: Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(57) Walters, P.; Stahl, M. *Babel*, version 1.6; University of Arizona: Tucson, AZ, 1996.
(58) Constants, 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.

CI010348T