Molecular Electron Density Lego Approach to Molecule Building

P. Duane Walker and Paul G. Mezey*,†

Contribution from the Mathematical Chemistry Research Unit, Department of Chemistry, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 0W0

Received June 14, 1993. Revised Manuscript Received September 7, 1993®

Abstract: A new method is presented for the construction of *ab initio* quality approximate electronic charge distributions for large molecules from charge distributions of small molecular fragments. This method is reminiscent to building structures using Lego blocks. The electronic density distribution calculated using the method is quantitatively shown to be very similar to that calculated for entire molecules using conventional *ab initio* packages with standard basis sets such as 6-31G**, while requiring only a fraction of the computational time. The pre-calculated fuzzy electron distributions of base molecular fragments, stored in a data bank, are merged (rotated, translated, and subsequently added together) to calculate these approximate charge distributions for the molecule. The method requires only the Cartesian coordinates of the nuclei for the construction of molecular charge distributions. The speed of the method makes it potentially useful for molecular modeling applications where construction of electron densities of large molecules is especially important. The technique can be used to replace fused sphere models of large molecules with more realistic electron density contours. One additional advantage is that these contours can be displayed for a whole range of density values, from tight, high-density contours, representing the bonding pattern, to loose, low-density contours representing the peripheral shape and size features of molecules.

1. Introduction

Electronic charge distributions of molecules represent important chemical information for the interpretation of the static chemical properties and reactivities of molecules. Molecular isodensity contour surfaces (MIDCO's) provide a commonly used representation of the shapes and sizes of molecular surfaces. The construction and analyses of these surfaces have become an important task in biochemical modeling, quantum pharmacology, and computer-aided drug design (CADD).

Most commonly, molecular charge distributions are calculated using some quantum chemical package of programs such as GAUSSIAN 90.[1] The computational time required for the density calculation in these programs increases rapidly with the number of atomic basis functions. Therefore, the calculation and display of the MIDCO's for molecules of more than forty atoms are not economical using these standard techniques.

Building quantum chemical representations of molecules by combining molecular fragments has been the subject of many studies. Among the earliest approaches, the method of Christoffersen and Maggiora[2] was based on a linear combination of fragment orbitals. In that and related methods,[2-4] the objective was to generate inexpensive yet reasonably accurate molecular orbitals and calculate (both MO and total) energies for large molecules, based on orbitals of smaller molecular fragments.

When building MO's of large molecules from MO's of fragments, ideal additivity of fragment MO contributions cannot be obtained, since both renormalization and an increase in the number of MO's are required.

By contrast, fuzzy electron densities can be assigned to formal molecular fragments in an additive manner: by recombining their fragment densities, using simple addition, the original total electron density is reconstructed.

The above simple principle had little utility in quantum chemical applications, as long as the focus of investigation was on energy values and geometry optimizations. However, the recent increased needs for more accurate modeling of molecular shape, and the advances in the analysis of the shapes of fuzzy, 3-dimensional electronic clouds[5-9] opened up important applications for the above fragment density principle. This idea corresponds to a computational, quantum chemical analog of the synthetic "tinker toy" approach of Michl[10] and the synthetic LEGO approach of Mathias and Stoddart.[11]

In this work, a new, approximate method is presented which calculates the electron density distributions of molecules by combining distributions for smaller fragments present in the molecule. The method is based on the assumption that contribution of a given molecular fragment to the complete molecular electron distribution should be quite similar in different molecules or in different locations of the same molecule, provided that the molecular environments are similar. That is to say, the distribution of electron density of molecular fragments is approximately "transferable". Here, the concept of the electron density of a molecular fragment applies to a single atomic neighborhood in the molecule as well as to large neighborhoods dominated by several nuclei. Earlier work has shown that the regions of the molecular charge distribution in the close vicinity of the nuclei of a molecule are quite invariant to even substantial molecular

† Also at Department of Mathematics, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, S7N 0W0

• Abstract published in *Advance ACS Abstracts,* November 15, 1993.

(1) Frisch, M. J.; Head-Gordon, M.; Trucks, G. W.; Foresman, J. B.; Schlegel, H. B.; Raghavachari, K.; Robb, M. A.; Binkley, J. S.; Gonzalez, C.; Defrees, D. J.; Fox, D. J.; Whiteside, R. A.; Seeger, R.; Melius, C. F.; Baker, J.; Martin, R. L.; Kahn, L. R.; Stewart, J. J. P.; Topiol, S., Pople, J. A. *Gaussian 90*; Gaussian, Inc.: Pittsburgh, PA, 1990.

(2) Christoffersen, R. E.; Maggiora, G. M. *Chem. Phys. Lett.* **1969**, *3*, 419.

(3) Christoffersen, R. E.; Shipman, L. L.; Maggiora, G. M. *Int. J. Quantum Chem.* **1971**, *5*, 143.

(4) Christoffersen, R. E.; Spangler, D.; Hall, G. G.; Maggiora, G. M. *J. Am. Chem. Soc.* **1973**, *95*, 8526.

(5) Mezey, P. G. *J. Comput. Chem.* **1987**, *8*, 462.

(6) Mezey, P. G. Three-Dimensional Topological Aspects of Molecular Similarity. In *Concepts and Applications of Molecular Similarity*; Maggiora, G. M.; Johnson, M. A., Eds. Wiley: New York, 1990.

(7) Walker, P. D.; Arteca, G. A.; Mezey, P. G. *J. Comput. Chem.* **1991**, *12*, 220.

(8) Walker, P. D.; Arteca, G. A.; Mezey, P. G. *GSHAPE 90*; University of Saskatchewan, 1990.

(9) Mezey, P. G. *Shape in Chemistry: An Introduction to Molecular Shape and Topology*; VCH Publishers: New York, 1993.

(10) Michl, J.; Kaszynski, K.; Friedli, A. C.; McMurdie, N. D.; Kim, T. *NATO ASI Ser., Ser. C* **1989**, *273*, 469.

(11) Mathias, J. P.; Stoddart, J. F. *Chem. Soc. Rev.* **1992**, 215.

0002-7863/93/1515-12423$04.00/0

deformations as long as the relative positions of the nuclei within the fragment do not change.[12] Pichon-Pesme *et al.* construct electron density cross sections for large molecules by considering a fragment of the molecule whose constituent atoms satisfy a distance criterion from the plane of the cross section.[13] The terminal ends of the fragment are tied off with hydrogen atoms and a SCF calculation is done on the fragment. This requires multiple SCF calculations to be performed if one needs more than one cross section, a disadvantage when compared to our method described in this paper, that is based on a partitioning of the density matrix and handles the entire density distribution, not just individual cross sections. A fragment approach has also been developed using density functional theory.[14,15] This approach is more fundamental and can be used to generate, in principle, all properties, not only the density. However, this approach only makes the calculation of molecular properties of large molecules possible, it does not make the calculation fast. If one needs to generate the electron density distribution quickly another approach is needed. Bader and co-workers have proposed a theory of "atoms in molecules"[16-18] that uses a compartmentalization of the 3D space to define molecular fragments such as CH_2. They have shown that such compartments of the electron density are approximately transferable, although adjustments of the boundaries are required when recombining such "compartments" in a different molecule. This boundary readjustment is nontrivial and implies some sacrifice of the quantum chemical justification of using such compartments. By contrast, in our work the idea of fragments in molecules will be defined using a fundamentally different criterion. The molecular charge distributions for the molecular fragments described in this work will not have fixed boundaries but rather will be taken as interpenetrating fuzzy charge clouds of electron density.

The distributions for molecular fragments are calculated by considering small parent molecules, with standard as well as some important non-standard molecular geometries, which contain the desired fragment. The method for calculating the electron density distribution for a molecular fragment will be described in detail in section 2. Our method uses a database of molecular fragment density distributions to construct the molecular distributions.

The molecule whose density distribution is being constructed is partitioned into fragments which appear in the database. The density distributions for these fragments are sequentially rotated and translated to account for the arrangement of the fragment in the actual molecule and then added together (superimposed) to model the molecular distribution. The method reminds one of the popular game Lego: building large structures by adding together small building blocks which snap together. In our model, snapping together is accomplished by mutual interpenetration of charge clouds. Hence the method is called the Molecular Electron Density Lego Approach (MEDLA) to molecule building.

To quantitatively calculate the similarity of our generated Lego surfaces with the direct *ab initio* results, we will use the following approach. Carbó developed an index, R_{AB}, to calculate the molecular similarity of two molecules from their electronic density distributions.[19-21]

$$R_{AB} = \frac{\int \rho_A \rho_B \, d\nu}{(\int \rho_A \, d\nu)^{1/2} (\int \rho_B \, d\nu)^{1/2}} \tag{1}$$

Meyer and Richards modified the index to account for the shape similarity of both Van der Waals surfaces and electronic isodensity surfaces.[22,23]

$$S_{AB} = \frac{B_{AB}}{(T_A \, T_B)^{1/2}} \tag{2}$$

where B_{AB} is the number of points falling within surfaces for both molecules and T_A is the number of points falling within the surface of molecule A. When this method was adopted for the case of electron isodensity surfaces, the density threshold value used was chosen to give the approximate Van der Waals surface (usually between 0.001 and 0.002 au[12]).

For all molecules considered in this work, the similarity index S_{AB} was evaluated for the formal molecular body including the entire density range down to a threshold value of 0.001 au.

The computational time required for the method is a small fraction of what would be needed using conventional *ab initio* techniques. The approximate molecular charge distributions closely resemble those calculated using conventional quantum mechanical approaches, but the method requires only seconds of computation time. As well, the memory requirements of the method are much smaller than those required for direct quantum mechanical approaches for the calculation of density distributions of larger molecules. This makes the method potentially useful for molecular modeling applications when studying large molecules. Traditionally the formal surfaces of large molecules are represented by either Van der Waals surfaces (VDWS) or Molecular Electrostatic Potential Surfaces (MEPS). By contrast, the present density Lego approach will allow the interactive use of realistic isodensity surfaces as a more reliable representation for the space filling characteristics of large molecules or molecular fragments. Furthermore, our MEDLA MIDCO's can be easily constructed for a whole range of density values, from the tight, high-density contours representing the bonding pattern, to the loose, low-density contours representing the peripheral shape and size features of molecules.

The paper is organized as follows. In the next section, the method used to calculate the molecular fragments which constitute the molecular database is explained. Section 3 details the molecular electron density Lego approach, and as an illustration, the method is used to construct the molecular charge distribution for β-alanine. Sample charge distributions are calculated and compared for a set of ten molecules using both the MEDLA and conventional techniques in section 4. The last section lists some conclusions and plans for future work.

2. Calculation of Electron Density for a Molecular Fragment

In order to calculate the electron density for a molecule using standard *ab initio* quantum mechanical techniques, the $N \times N$ density matrix **P** is constructed where N is the number of atomic basis functions used. The electron density function $\rho(\mathbf{r})$ is calculated from this density matrix as shown in eq 3,

$$\rho(\mathbf{r}) = \sum_{i=1}^{N} \sum_{j=1}^{N} P_{ij} \varphi_i(\mathbf{r}) \varphi_j(\mathbf{r}) \tag{3}$$

where $\varphi_i(\mathbf{r})$ is the ith atomic basis function. Density matrices have commonly been partitioned into two or more matrices in order to analyze various contributions to the total charge

(12) Arteca, G. A.; Grant, N. A.; Mezey, P. G. *J. Comput. Chem.* **1991**, *12*, 1198.
(13) Pichon-Pesme, V.; Lecomte, C.; Bénard, M. *J. Am. Chem. Soc.* **1992**, *114*, 2713.
(14) Yang, W. *Phys. Rev. Lett.* **1991**, *66*, 1438.
(15) Lee, C.; Yang, W. *J. Chem. Phys.* **1992**, *96*, 2408.
(16) Bader, R. W. F.; Nguyen-Dang, T. T. *Adv. Quantum Chem.* **1981**, *14*, 63.
(17) Bader, R. W. F. *Acc. Chem. Res.* **1985**, *9*, 18.
(18) Bader, R. W. F.; Caroll, M. T.; Chessman, J.; Chang, C. *J. Am. Chem. Soc.* **1987**, *109*, 7068.
(19) Carbó, R.; Leyda, L.; Arnau, M. *Int. J. Quantum Chem.* **1980**, *17*, 1185.
(20) Carbó, R.; Arnau, M. Molecular Engineering: A General Approach to QSAR In *Medicinal Chemistry Advances*; de las Heras, F. G., Vega, S., Eds.; Pergamon Press: Ocford, 1981.
(21) Carbó, R.; Farre, A. J. *Afinidad* **1982**, *39*, 205.
(22) Meyer, A. M.; Richards, W. G. *J. Comput. Aided Mol. Des.* **1991**, *5*, 426.
(23) Good, A. C.; Richards, W. G. *J. Chem. Inf. Comput. Sci.* **1992**, *33*, 112.

density.[24,25] Here our goal is different: the establishment of a molecular fragment density database.

Molecules can be considered as a collection of interpenetrating molecular fragments using some chemical basis for partitioning. In this work, a molecular fragment will be defined to be any subset of the nuclei of a molecule, together with the portion of the electronic density assigned to these nuclei by a partitioning analogous to that in a Mulliken population analysis. Note that in this definition, the nuclei of a formal "fragment" need not even be related to one another by formal chemical bonds; however, in practice the fragments chosen will normally have some chemical "identity" as conventional functional groups.[9] The density matrix for a molecule can be partitioned into $N \times N$ fragment density matrices ($\mathbf{P}^k$) for a molecular fragment of serial index k as shown in eq 4.

$$P^k_{ij} = P_{ij} \text{ if both } \varphi_i \textit{ and } \varphi_j \text{ are atomic basis functions centered on nuclei contained within the molecular fragment}$$

$$= {}^1/_2 P_{ij} \textit{ if } \varphi_i \text{ or } \varphi_j \text{ is an atomic basis function for a nucleus within the molecular fragment but not both}$$

$$= 0 \text{ otherwise} \quad (4)$$

Note that the above scheme follows the spirit of a Mulliken population analysis. Also note that for every partitioning of a molecule into M molecular fragments there will be M fragment density matrices $\mathbf{P}^k$, $k = 1, \ldots M$. The total density matrix $\mathbf{P}$ is the exact sum of the M fragment density matrices,

$$P_{ij} = \sum_{k=1}^{M} P^k_{ij} \quad (5)$$

The electron density $\rho^k(\mathbf{r})$ of a given molecular fragment is then simply,

$$\rho^k(\mathbf{r}) = \sum_{i=1}^{N} \sum_{j=1}^{N} P^k_{ij} \varphi_i(\mathbf{r}) \varphi_j(\mathbf{r}) \quad (6)$$

and the total electron density for the molecule is the sum of the densities for the fragments,

$$\rho(\mathbf{r}) = \sum_{k=1}^{M} \rho^k(\mathbf{r}) \quad (7)$$

Figure 1 shows MIDCO's for some possible molecular fragments of the methane molecule. The original charge density matrix was calculated by using GAUSSIAN 90 with a 3-21G basis set and our program[8] based on relations 4 and 6 to calculate the density distribution. The first "fragment" shown is simply the molecule itself. The second fragment is the "carbon atom" in the methane molecule. Notice that the sp^3 nature of the "atom" is clearly seen in the isodensity surface shown. (This tetrahedral shape is not as prominent for some MIDCO's with different density threshold values, not shown in the figure.) The density for the molecule can be obtained by adding the densities of the carbon fragment and the H_4 fragment shown in Figure 1. This is not simply adding two isodensity surfaces together, but it is actually the addition of the two fuzzy charge distributions and the subsequent recalculation of any desired contour surface for a selected density threshold. The density distributions of the two fragments interpenetrate each other considerably for the MIDCO threshold density value shown in Figure 1.

The recombination of fragments leading to the molecular electron density is exact on the given *ab initio* level. The density

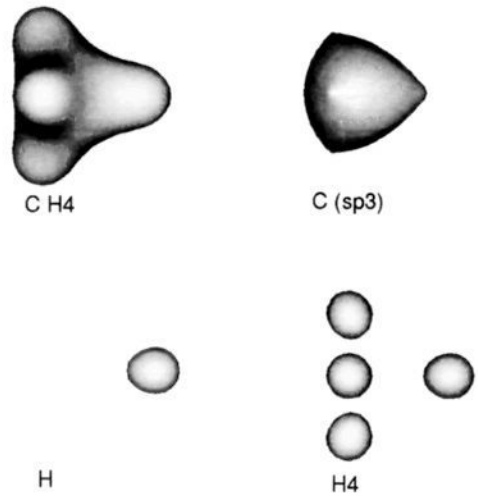

Figure 1. Isodensity surfaces for possible fragments of methane. These densities were calculated using a 3-21G basis set. The threshold value for all isodensity surfaces displayed in this figure was 0.01 au.

Table I. List of Molecular Fragments Used to Construct the Molecules in This Work and the Parent Molecules from Which They Originate

molecular fragment	parent molecule
$CH_2(sp^3)$	propane
$CH_3(sp^3)$	propane
$CH(sp^2)$	propene
OH	methanol
COOH	acetic acid
NH_2	methylamine

of the C fragment plus the density of the H_4 fragment is exactly equal to that of the methane molecule. However, one can add two fragments from different molecules to calculate an approximate density distribution for a larger molecule. This allows the electron density distributions of small molecular fragments obtained from small molecules to be building blocks for approximate distributions of larger molecules. While this procedure is not exact, the closer the environment of the fragments in the original molecules is to that in the new molecule, the closer the approximate distribution is to the actual one. Exploiting these similarities is the basis of the molecular electron density Lego approach.

3. Molecular Electron Density Lego Approach

The first step in developing the MEDLA was to create a database of electronic density distributions for a set of base fragments. In this process general strategies can be followed. One may use very high quality geometry optimization for fragment calculations and include in the data base several special cases for various environments. Alternatively, by acknowledging that in applications for building electron densities of large molecules the main source of error is the multitude of minor distortions and interactions of the fragments, such high accuracy may not always be warranted. In this study we have opted for lower levels of geometry optimizations. In this first application of the method, six fragments are used. Table I gives a list of the fragments and the parent molecule from which they originate. All distributions of the fragments were calculated using the following steps. The geometry of the parent molecule was optimized using the molecular modeling package BIOGRAF.[26] Next the density matrix for this geometry is calculated at the 6-31G** level using GAUSSIAN 90. The fragment density matrix is constructed to calculate the density distribution for the molecular fragment using

(24) Smith, V. H. In *Electron Distributions and the Chemical Bond*; Coppens, P., Hall, M. B., Eds.; Plenum Press: New York, 1982; pp 3–59.
(25) Hall, M. B. In *Electron Distributions and the Chemical Bond*; Coppens, P., Hall, M. B., Eds.; Plenum Press: New York, 1982; pp 205–219.
(26) *BIOGRAF*; Biodesign, Inc.: 199 S. Los Robles Ave., Pasadena, CA 91101, 1989.

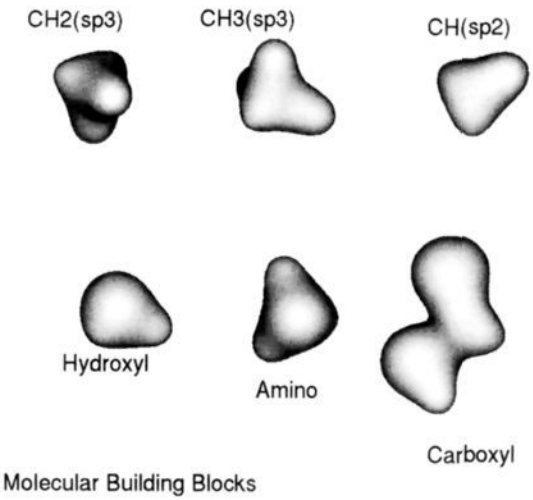

Figure 2. Isodensity surfaces for the six base fragments used to construct the Lego molecular densities in this paper. All fragments were constructed using the method presented in section 2. The basis set used for these fragments was 6-31G**. The threshold value of all surfaces shown was 0.1 au.

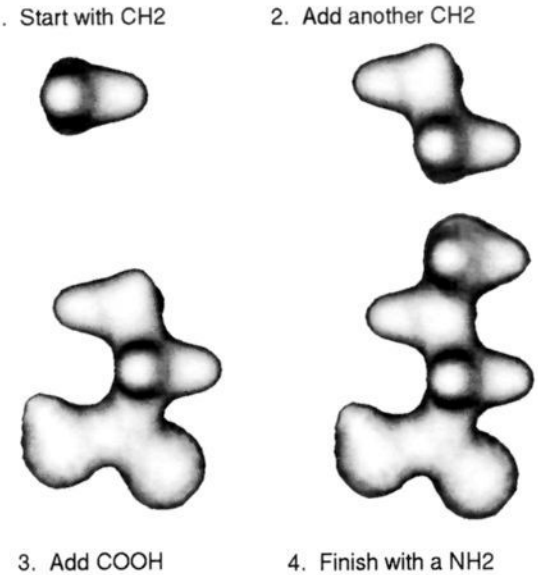

Figure 3. Isodensity surfaces for molecular fragments of β-alanine at intermediate steps of the Lego calculation. This figure shows the sequential method of the Lego method for calculating the density distribution of a molecule. The threshold value for all surfaces shown was 0.1 au.

the method described in the previous section. In this work, the density distribution is stored as a cubic grid of density values. The size and resolution of the grid are 10.0 and 0.1 au, respectively. Also needed in the database are the coordinates of a central atom and three reference points for the fragment, non-coplanar with the central nucleus. In the case of linear or planar fragments, dummy atoms are added as additional reference points. At this resolution and size of the distribution, approximately 4 megabytes of disc space are needed for each fragment in the database. Isodensity contour surfaces for the six fragments used in this work are shown in Figure 2. The parent molecules are chosen to give similar molecular environments for the fragments (the fragments are bonded only to carbon atoms in all parent molecules).

Once the database is created, the program LEGO can be used to calculate the electronic distribution for additional large molecules which can be partitioned into fragments which appear in the database. For these molecules the density distributions calculated using LEGO are constructed using the following steps:

1. The molecular geometry is optimized using the molecular modeling package BIOGRAF or any other method. In this work, little emphasis was placed on geometry optimization, while more attention was paid to the calculation of the density.

2. The molecule is partitioned into molecular fragments which are present in the database. If needed, dummy atoms are added to give nonplanar reference points to the fragment (for example in the carboxyl fragment).

3. The input data on size and resolution for the distribution of the new molecule is read by the program. The resolution can be as high as that used for creating the fragment density distributions in the database. However, to achieve smooth composite isodensity surfaces, a resolution of less than half of that of the base distributions is suggested. All calculations of density distributions for molecules in this work use a resolution of 0.3 au (one third of that of the base fragments).

The next steps are performed sequentially for each fragment present in the molecule.

4. A transformation matrix is constructed which converts the nuclear coordinates of the fragment in the new molecule into the nuclear coordinates of the fragment in the parent molecule. Actually, this transformation is made up of three steps. First there is a translation of the nuclei of the fragment in the new molecule to the origin of the coordinate system, next they are rotated to align the reference vectors with those of the fragment in the parent molecule, and finally the nuclei are translated to their position in the parent molecule.

5. The fragment density $\rho^k(\mathbf{r})$ is calculated for each point $\mathbf{r}$ in the grid as follows. First the coordinates of the point are transformed into a new reference frame using the same transformation as in step 4. This gives the coordinates of the point where the corresponding density value is found for the distribution stored in the database. This density value is then taken for the contribution of the fragment to the density distribution of the new molecule at the original nontransformed coordinates. Note that there will be some loss of information due to the discrete rectangular grid of the stored distributions. This error is reduced somewhat by calculating the distributions for the base fragments using high resolution. If the transformed coordinates are outside the limits of the cubic grid of the base distribution, the contribution of the fragment to $\rho^k(\mathbf{r})$ is assumed to be 0.0.

It is apparent that the method can be used to calculate the approximate density distribution for an entire molecule or any collection of fragments which appear in the molecule. Figure 3 shows isodensity surfaces for various stages of the calculation of the density distribution of the molecule β-alanine. Studies on the conformational preferences and shape changes of various conformers of β-alanine are reported elsewhere.[27,28] First the distribution for one of the CH_2 fragments is calculated, followed by that for the second CH_2 fragment. Next the carboxyl fragment and finally the amino group are added. Note that the density from the individual fragments can be calculated in any order. The chosen order does, however, show the "Lego-like" quality of the method. The ability to construct the distributions for any molecular fragment is potentially useful for much larger molecules. Presently, the method has a practical size limit that the distributions must fit in a cube with edges approximately 45 au long using a resolution of 0.3 au. Any larger distribution would make the real time rendering of the isodensity surfaces impractical at this time. While the present constraints allow for the construction of isodensity surfaces of molecules much larger than those which could be constructed using conventional direct quantum chemical methods, there are, of course, many molecules which are too large to generate their entire density distributions

(27) Ramek, M. *J. Mol. Struct. (Theochem)* **1990**, *208*, 301.
(28) Heal, G.; Walker, P. D.; Ramek, M.; Mezey, P. G., submitted for publication.

Table II. Computational Time Required to Calculate Molecular Charge Distributions for the Molecules Studied Using Both the LEGO Program as Well as Conventional Quantum Mechanical Techniques at the 6-31G** Level of Basis

molecule	CPU time	
	MEDLA (s)	GAUSSIAN/6-31G** (min)
ethane	3.6	11
ethanol	5.1	16
acetic acid	3.6	23
propane	5.1	20
propanol	6.6	31
propanoic acid	5.1	34
1-amino-2-ethanol	6.6	45
β-alanine	6.6	49
2-butene	6.6	40
cyclohexane[a]	10.0	125

[a] Note the grid size for cyclohexane was 75 × 75 × 75 instead of 50 × 50 × 50.

using the method. However, the distribution for a substantial fragment of such a molecule, for example a protein, could still be constructed and analyzed by using the LEGO program.

The sheer size of molecules which can have their density calculated is certainly a benefit of this new approach. Very encouraging are the accuracy of the calculated Lego distributions when compared to those calculated using direct *ab initio* methods at 6-31G** level and the minimal computational time required by the method. For example, the calculation of the entire density distribution for β-alanine using a grid of size 50 and resolution of 0.3 au required only 6.62 s on a KPC Titan workstation using the LEGO program, while the direct calculation using GAUSSIAN 90 and 6-31G** basis sets took over 45 min. In the next section we shall compare the resulting density distributions and computational times for a series of molecules using both the Lego approach and conventional *ab initio* methods.

4. Accuracy and Efficiency of the Molecular Electron Density Lego Approach

In this section the MEDLA is used to calculate the electron distributions for a series of molecules. For comparisons, the distributions were also calculated using a direct quantum mechanical approach at the 6-31G** level. Table II shows the computational time required for the calculation using both methods. It is obvious from this table that as the size of the molecule increases, so does the computational advantage of the LEGO program. The computational time for the Lego method increases approximately linearly with the number of fragments in the molecule, while the time required for conventional techniques increases by a higher power of the number of atomic basis functions.

The computational advantage makes the method appealing, especially since the accuracy of the method is remarkable. Figures 4–13 show isodensity surfaces for the series of molecules constructed using both methods. Table III shows the calculated similarity index S_{AB} for each pair of LEGO/*ab initio* electron density distributions. Both in Table III and from the figures we see that the LEGO generated surfaces have a high degree of similarity with their direct *ab initio* counterparts. Some small discrepancies arise from the transformation of coordinates from the grid in the new molecule into coordinates not exactly found in the distribution grid for the fragment in the parent molecule. The method works equally well for molecules containing the hydroxyl, carboxyl, and amino fragments as it does for the pure hydrocarbons. A couple of the molecules seen here do, however, have a fortuitously high degree of accuracy, since their fragments appear with the same surroundings as the fragments generated for the data base. In the acetic acid molecule (Figure 6), the density due to the carboxyl fragment is identical to that in the data base as the parent molecule used to generate the fragment

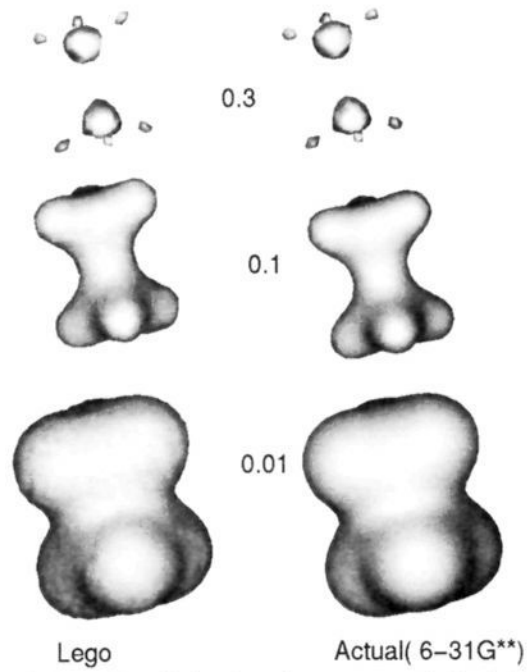

Figure 4. Approximate isodensity surfaces constructed using the LEGO program for three values of threshold density and the corresponding actual 6-31G** isodensity surfaces for the ethane molecule. Atomic units are used throughout.

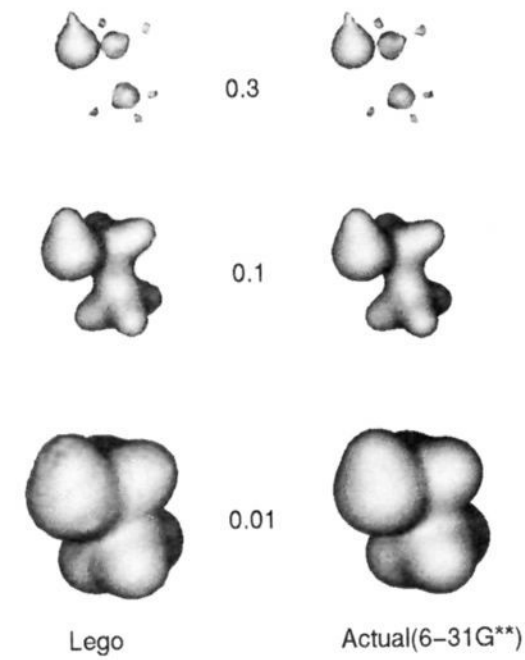

Figure 5. Approximate isodensity surfaces constructed using the LEGO program for three values of threshold density and the corresponding actual 6-31G** isodensity surfaces for the ethanol molecule.

was acetic acid. The same can be said for the CH_2 and one of the CH_3 fragments in the "composite" propane, as propane was the parent molecule for both of these base fragments.

Figure 11 shows that the MIDCO's calculated using the LEGO program agree with the actual 6-31G** surfaces over a wider density range than the ones constructed using direct *ab initio* methods but a smaller basis set, even though the actual computed similarities for the entire distribution are similar. Included in this figure are three isodensity surfaces for β-alanine constructed using both the LEGO program and conventional quantum mechanical methods at both the 6-31G** and the 3-21G levels

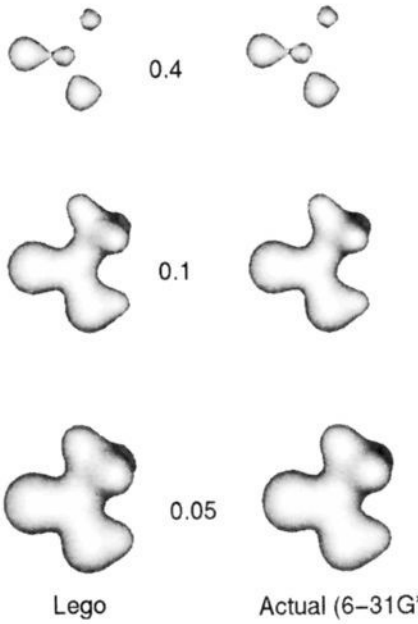

Figure 6. Approximate isodensity surfaces constructed using the LEGO program for three values of threshold density and the corresponding actual 6-31G** isodensity surfaces for the acetic acid molecule.

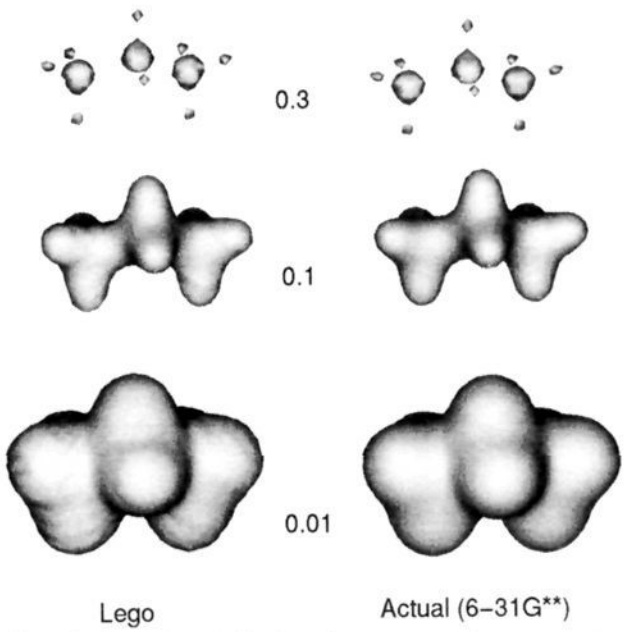

Figure 7. Approximate isodensity surfaces constructed using the LEGO program for three values of threshold density and the corresponding actual 6-31G** isodensity surfaces for the propane molecule.

of basis. Little difference is seen in the 0.04 and 0.1 au isodensity surfaces. However, at a density threshold 0.25 au, where the isodensity surfaces of the LEGO program still match the 6-31G** surfaces closely, the 3-21G surfaces have become significantly different. This suggests that the approximate LEGO generated surfaces are more suitable than the 3-21G conventional surfaces for techniques which examine the high-density regions of the distributions. For this reason, the LEGO method is expected to enhance the scope of methods analyzing bonding patterns, such as the density domain approach to bonding.[29,30]

Two molecules were considered which were not substituted

(29) Mezey, P. G. *Molecular Surfaces.* In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1990; pp 265–294.

(30) Mezey, P. G. *J. Chem. Inf. Comp. Sci.* **1992**, *32*, 650.

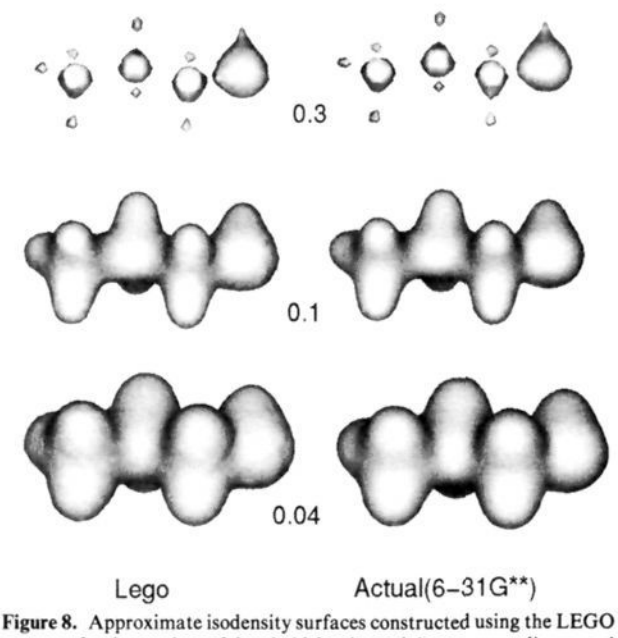

Figure 8. Approximate isodensity surfaces constructed using the LEGO program for three values of threshold density and the corresponding actual 6-31G** isodensity surfaces for the propanol molecule.

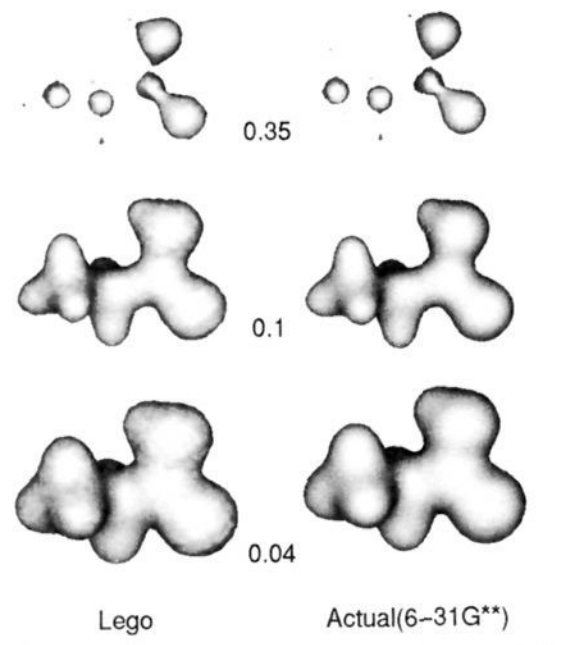

Figure 9. Approximate isodensity surfaces constructed using the LEGO program for three values of threshold density and the corresponding actual 6-31G** isodensity surfaces for the propanoic acid molecule.

straight chain alkanes. In Figure 12, the isodensity surfaces for 2-butene are shown for both the *ab initio* and LEGO calculations. Notice that the C–C double bond is very well represented even at high density (see $\rho = 0.30$ au). Figure 13 shows the results for a cyclic hydrocarbon, cyclohexane. Again, the LEGO generated surfaces are nearly indistinguishable from the direct *ab initio* surfaces. As well, cyclic molecules differ from noncyclic ones as they have one shape change which occurs at relatively low density. At some value of density, the isodensity surfaces change from surfaces topologically equivalent to a torus (see $\rho = 0.10$ au) to surfaces topologically equivalent to a sphere (see $\rho = 0.01$ au). Although not shown in the figure, the density at which this shape change occurs is nearly the same for both methods (0.019 and 0.021 au for the *ab initio* and the MEDLA methods, respectively). This is a promising feature as this shape change

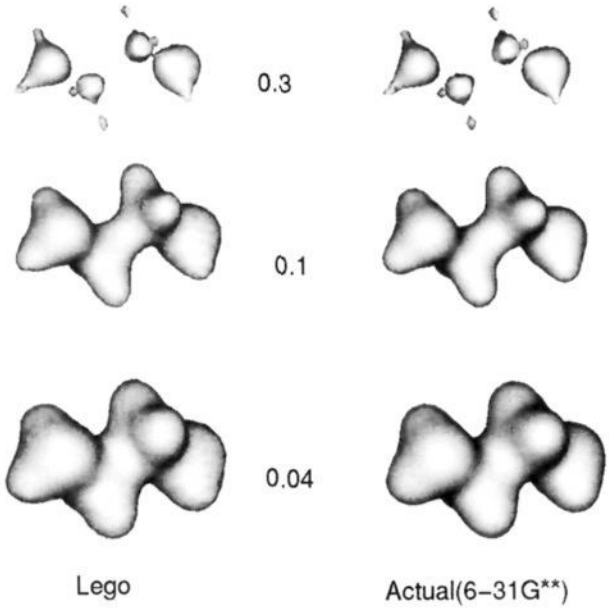

Figure 10. Approximate isodensity surfaces constructed using the LEGO program for three values of threshold density and the corresponding actual 6-31G** isodensity surfaces for the 1-amino-2-ethanol molecule.

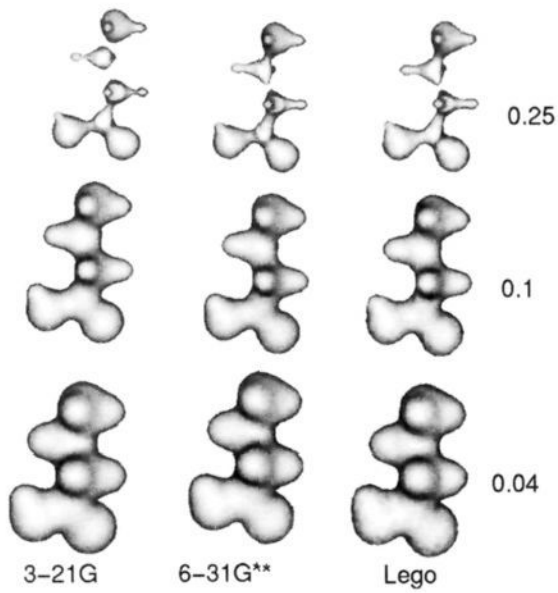

Figure 11. Approximate isodensity surfaces constructed using the LEGO program for three values of threshold density and the corresponding 6-31G** and 3-21G isodensity surfaces for the β-alanine molecule.

is a global feature of the molecule rather than a feature restricted to a single fragment as are the other topological changes seen at higher density. In this figure, we have also included some isodensity surfaces for threshold values which compare approximately to the respective Van der Waals surface for the molecule. The similarities are remarkable. Of course, our goal is not to approximate Van der Waals surfaces but to generate good quality electron densities of large molecules for a whole range of density values. In fact, the importance of fused sphere Van der Waals surfaces is expected to diminish if the calculation of electron densities of large molecules becomes a routine task.

It is important to note that the molecular fragments in the composite molecular models of the examples were bonded to the same set of atoms as they were in the parent molecules. While this is not required, there is some loss of accuracy if the molecular environment is significantly different in the composite molecule.

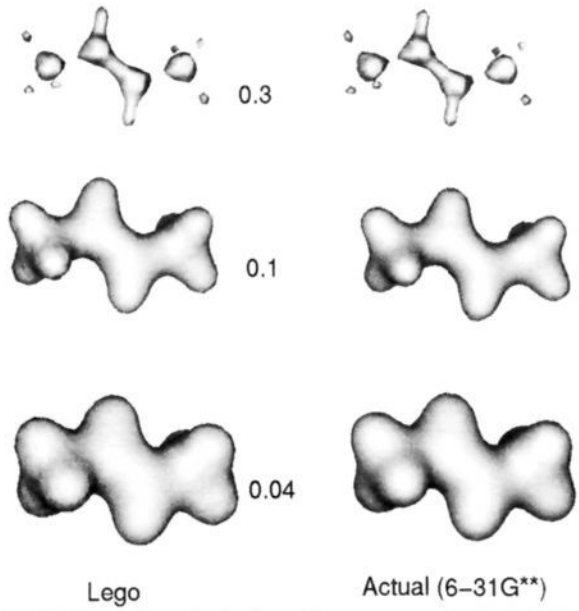

Figure 12. Approximate isodensity surfaces constructed using the LEGO program for three values of threshold density and the corresponding actual 6-31G** isodensity surfaces for the molecule 2-butene.

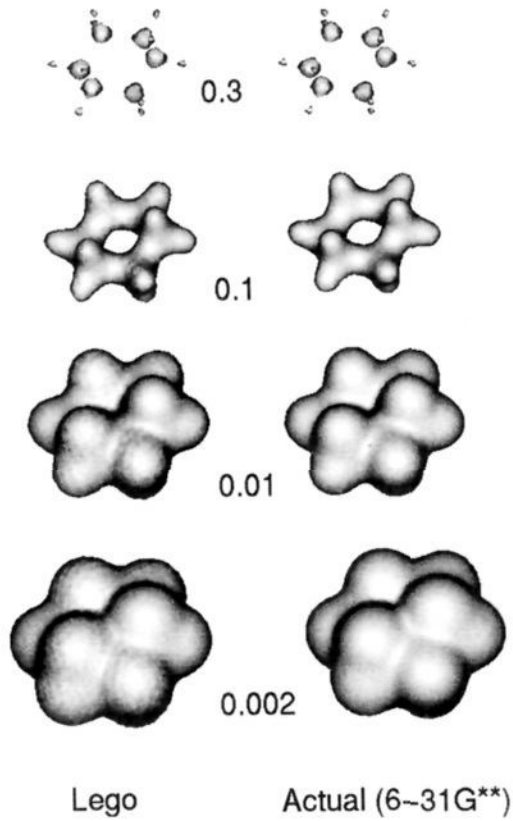

Figure 13. Approximate isodensity surfaces constructed using the LEGO program for four values of threshold density and the corresponding actual 6-31G** isodensity surfaces for the cyclohexane molecule.

The density of a fragment bonded to lithium is expdected to be significantly different from that of the same fragment bonded to fluorine, especially in the range between the bonding atoms. This problem can be easily avoided by adding further "specialized" fragments to the data base. One has to decide on a compromise between the accuracy desired for molecular charge distributions which are calculated and the number of possible fragments one wishes to store in the data base. Using the present size and

Table III. Computed Similarity Indices S_{AB} for Pairs of Molecules Demonstrating the Accuracy of the LEGO Generated Surfaces with Those Calculated from Direct *ab Initio* Results

	similarity index S_{AB} (%)
LEGO/6-31G**	
ethane	96.3
ethanol	96.7
acetic acid	96.0
propanol	97.1
propanoic acid	95.7
ethylamine	97.7
β-alanine	96.7
2-butene	96.9
cyclohezane	99.1
3-21G/6-31G**	
β-alanine	96.6

resolution of the distributions in the data base, 100 fragments would require approximately 400 megabytes of disc space. While this may seem intimidating, one must keep in mind that at present one CD can hold approximately 660 megabytes of information (or 165 fragments). The large number of molecules which can have their molecular distributions calculated with even a set of 100 unique molecular fragments (unique in either makeup or environment) is impressive. Even using a small set of base fragments, the approximate isodensity surfaces generated for the molecules should provide more representative molecular surfaces than the commonly used Van der Waals and similar models.

It should also be mentioned that as of now, the method may miss some global features of the density that are caused by interaction between large molecular fragments. Examples of features not considered are those caused by the formation of hydrogen bonds or by conjugation through a series of alternating double bonds, both of which are properly modeled by the density domain approach.[29,30] One can use the MEDLA method, however, to probe the significance of these interactions, and perhaps calculate these global contributions to a molecular charge distribution. These questions and a critical evaluation of the conventional aromaticity and resonance models are the subject of our current studies.

5. Closing Remarks

In this work the Molecular Electron Density Lego Approach (MEDLA) was proposed and used to calculate the density distributions for a series of molecules based on a combination of fragments of molecular densities pre-calculated and stored in a data base. The calculation of these approximate charge densities required a fraction of the computational time that would be needed for conventional *ab initio* techniques. Even so, the resulting distributions were shown to match closely those calculated using standard methods. As well, the method enables one to have density distributions calculated for much larger molecules than are currently possible using conventional techniques. Finally, the technique can be used to calculate density distributions for a molecular fragment of a very large molecule.

The method was designed mainly to obtain approximate density distributions on a near real time scale. The first goal of the approach was to give molecular modeling packages the ability to display "*ab initio*-like" isodensity surfaces for a whole range of density values without actually performing any quantum mechanical calculation. However, the Lego approach will also be modified to calculate very accurately the density distributions for large molecules. This entails calculating densities for as large a fragment of the molecule as possible using conventional techniques. The molecular environment of the fragment in the parent molecule would be the same as that in the desired molecule for a distance of at least six atomic units. Very accurate density distributions for large molecules could be constructed using this approach. This modification of the LEGO program would extend the original method presented in this paper and would create another efficient computational tool.

One additional use of these LEGO generated electron density distributions is that they could be used to calculate the molecular electrostatic potential distributions for large molecules as well. This will be pursued in later work.

The simplicity, efficiency, and accuracy of the molecular electron density Lego approach of molecular building are expected to provide a very useful method for more accurate and detailed molecular modeling applications.

Acknowledgment. This work has been supported by both individual and strategic grants from the Natural Sciences and Engineering Research Council of Canada.