FUNCTIONAL GROUPS IN QUANTUM CHEMISTRY

Paul G. Mezey

Mathematical Chemistry Research Unit,

Department of Chemistry and Department of Mathematics and Statistics,

University of Saskatchewan,

110 Science Place, Saskatoon, Canada, S7N 5C9

Table of Contents

1. INTRODUCTION

In this report the quantum chemical concept of functional groups will be described, following a topological approach based on the three-dimensional shape of fuzzy molecular bodies and the local shapes of various molecular moieties. This quantum chemical description of functional groups was first proposed in ref. [1] and further elaborated in refs. [2,3], based on the topological shape analysis of molecules [2] and on the *density domain* approach to chemical bonding [4]. Density domains are formal bodies enclosed by molecular isodensity contours; density domains play an important role in molecular shape analysis and serve as the basis of various molecular similarity measures [5]. The proposed quantum chemical model of functional groups also fits within a rather broad, essentially geometrical framework discussed in [6]. Also in ref.[6], a general algebraic structure has been proposed for families of functional groups. The suggested algebraic structure is a mathematical lattice describing the interrelations between functional groups. This general algebraic description is also applicable to the families of quantum chemical functional groups defined by topological shape conditions of the electron density [1-3,5].

A chemical functional group is a collection of nuclei and the associated electron density which occur with a similar nuclear arrangement and a similar local electron density cloud in many molecules. A functional group is a molecular moiety of a specific stoichiometry that typically undergoes similar chemical reactions in most molecules containing this moiety. In chemistry, especially in organic chemistry, the concept of functional groups is a powerful tool used for the characterization and classification of molecules and their reactions; in fact, the presence of a given functional group in a family of molecules is an expression of chemical similarity. Molecules containing common functional groups often exhibit similar physical properties in addition to their similar chemical properties. For example, organic molecules containing the OH group often generate hydrogen bonds between one another, that in turn has the usual consequence of a higher boiling point than that of molecules of similar molecular weight but containing no OH groups. There are important differences, however, depending on the local surroundings of these OH groups, and it is often justified to regard larger moieties, such as the carboxyl group, COOH, or the hydroxymethyl group, CH_2OH, as functional groups.

Apparently, the concept of *similarity* plays an important role in the chemistry of functional groups. Motivated by the recent revival of interest in molecular similarity [7-39], we shall present a systematic approach towards a quantum chemical description of functional groups. There are two main components of the approach described in this report. The first component is shape-similarity, based on the topological shape groups and topological similarity measures of molecular electron densities[2,19-34], whereas the second component is the Density Domain approach to chemical bonding [4]. The topological Density Domain is a natural basis for a quantum chemical model of functional groups, first proposed in [1] and further elaborated in refs. [2,3] and [5].

The conventional chemical notations using either the familiar structural formulas or the more revealing stereodiagrams are especially suited for encoding the identity of molecules. Whereas structural formulas and stereodiagrams given in terms of formal bonds depicted as lines represent only the skeletal structure of molecules; nevertheless, they serve as a "shorthand notation" for molecules and the three-dimensional arrangements of their nuclei. Depending on the chemical process, a given functional group may change or may remain essentially intact during the process. The actual transformations of functional groups (if the given groups are actively participating in the reactions), as well as their simple retention while the reactants turn into products (if the actual reactions affect directly only some other parts of the molecules) are easily modeled by fragments of the bond diagrams. In such models, a functional group is represented by its local bonding pattern, for example, the OH group is represented by the symbol -O-H, indicating the bond within the group as well as a formal broken, "dangling" bond between the functional group and the rest of the molecule. Based on such notations, the various functional groups of a molecule can be recognized. If one is familiar with the typical properties and behavior of the functional groups present, then one can often predict some of the most essential physical and chemical properties of the molecules as well as the main features of their reactions. Evidently, much of our chemical knowledge is formulated and classified in terms of functional groups.

The usual representation of functional groups as a particular local bonding pattern involving a set of atoms is somewhat oversimplified, and only a few attempts have been made to provide a more rigorous, quantum chemical definition and description of functional groups. One of the difficulties is associated with the description of parts of a quantum mechanical molecule, where these parts, as functional groups, possess individual characteristics and a certain degree of autonomy within the molecules. For the quantum chemical description of functional groups, it appears natural to use models and methods also applicable for complete molecules; in particular, it is well advised to avoid models showing artificial boundaries between various parts of a molecule. Models of molecular fragments showing discontinuities of electron density at some of their formal boundaries have artificial features; such discontinuous features are clearly impossible for entire molecules. Subdividing a molecule that appears an easy task for simple, traditional models involving formal chemical bonds depicted as lines is a problem of considerable complexity within a quantum chemical context. In reality, chemical bonding in molecules is not restricted to simple lines, such as the formal lines of bond diagrams; in fact, the actual chemical bonding is due to the electron density distribution throughout a fuzzy molecular body of the electronic charge cloud. In a more realistic model, functional groups require a more precise, quantum chemical definition; in particular, it is of importance to establish the extent by which various parts of the electron density belong to one or to another functional group of the molecule.

In this contribution, an approach to the quantum chemical concept of functional groups will be discussed, with special emphasis on their shapes, similarity and complementarity relations.

2. MOLECULAR FRAGMENTS, CHEMICAL FUNCTIONAL GROUPS

2.1 Semiclassical Concepts of Molecular Bodies and Functional Groups

In principle, the semiclassical concept of a molecular body is rather simple: the molecular body is a fuzzy charge cloud of electron density surrounding a set of essentially pointlike nuclei which vibrate about a specified equilibrium nuclear arrangement. A nuclear arrangement and the surrounding electronic charge cloud form the molecule; there is nothing else there. The shape of the molecule is the shape of the fuzzy, electronic charge cloud, a dynamic object whose motion approximately follows the nuclear vibrations.

The above simple model provides important insight into various molecular properties, physical, chemical and biochemical processes. One should note, however, that this model treats nuclei on two, mutually incompatible levels: as an essentially classical set of pointlike objects with a well-defined equilibrium configuration, and also as a vibrating system of masses exhibiting quantum mechanical uncertainty as well as quantization for their vibrations. According to rigorous quantum mechanics, nuclei are subject to the Heisenberg uncertainty relation, and in a physically correct model of chemistry [40] the concept of "a precise nuclear position in a molecule" is just as meaningless as the concept of "the precise position of an electron in a molecule". In practice, however, the Heisenberg uncertainty for the position of the more massive nuclei is much less severe than it is for the much lighter electrons, and the above semiclassical model is a useful and, computationally, a rather efficient tool in the study of molecular behavior.

One approach to the approximate representation of molecular bodies is based on molecular isodensity contours, MIDCOs, defined with respect to some fixed nuclear configuration K and some electron density threshold a. A MIDCO $G(a,K)$ is defined (in the fixed nuclear configuration approximation) as the collection of all those points $\mathbf{r}$ of the three-dimensional space where the electronic density is equal to the threshold a:

$$G(a,K) = \{ \mathbf{r} : \rho(\mathbf{r},K) = a \}. \qquad (1)$$

If there is no need to include the nuclear configuration K in the notation, one may simply write

$$G(a) = \{ \mathbf{r} : \rho(\mathbf{r}) = a \}. \tag{2}$$

For any continuous electronic density function $\rho(\mathbf{r})$ of a molecule, the set of all points $\mathbf{r}$ fulfilling equation (2) must form a set of a finite number of continuous surfaces. For low values of the electron density threshold a, G(a) is usually a single, closed surface, whereas for high values of threshold a, G(a) is usually a collection of several closed surfaces, each surrounding some of the nuclei of the molecule.

In fact, the complete electronic density of a molecule can be represented by an infinite family of such MIDCOs, a family that contains one set G(a) of continuous surfaces for every threshold value a from the open-closed interval

$$(0, a_{max}], \tag{3}$$

where a_{max} is the electron density at a nucleus of the largest nuclear charge in the molecule. Consider the simplest case, where for a given interval of threshold values there is precisely one surface G(a) for each threshold value a. These MIDCOs are nested within one another, in a manner similar to a series of Russian wooden dolls, where for each pair of a and a' values,

$$a < a', \tag{4}$$

the larger MIDCO G(a) contains the smaller MIDCO G(a').

If the molecular body is represented by a fuzzy electronic charge density, then it is also justified to represent functional groups by electron density charge clouds of similar, fuzzy features. Although most functional groups do not retain their shape and other properties when they actually become disconnected from the rest of the molecule and become independent, separate entities, nevertheless, a consistent framework for their quantum chemical description appears to require a treatment analogous to that of complete molecules. For this end, one has to decide what contribution of the molecular electronic charge density cloud belongs to which functional group of a molecule. In a general scheme, functional groups or other molecular fragments can be treated similarly to entire molecules, leading to the new concept of fragment isodensity contour, FIDCO, analogous to the molecular isodensity contour, MIDCO.

The question of assigning nuclei to various functional groups is much simpler: it is natural to define a functional group as a given subset of the nuclei and their "share" of the electronic density of the molecule. However, even for nuclei, this approach is not fully satisfactory. For example, in some hydrogen-bonded systems the assignment of the bridging hydrogen nucleus to either of the two molecular fragments it links is somewhat ambiguous. In such cases the electronic density may decide, as we shall see in later parts of this report.

2.2 A Fuzzy Set Formalism for Molecular Electronic Density

Molecular bodies do not have actual boundaries; these bodies are fuzzy, three-dimensional charge clouds of the molecular electron density distribution. Molecular shape is the shape of this fuzzy charge cloud. The concept of molecular shape should be clearly distinguished from the concept of molecular conformation; the latter is usually interpreted as a specific nuclear arrangement and the associated three-dimensional pattern of formal chemical bonds. These bonding patterns are usually described by lines of formal chemical bonds of a structural diagram. Since the actual molecular shape is determined by the fuzzy electron distribution, it is of some importance to reflect this fuzziness in models describing molecular shapes and chemical bonding.

For a single, isolated molecule X one may consider the following problem: to what degree do various points $\mathbf{r}$ of the three-dimensional space belong to the molecule? A fuzzy set model of molecules can be introduced that provides a simple and rather natural approach to this problem.

Take the maximum value ρ_{max} of the electron density within a spatial domain D containing the nuclei of molecule X,

$$\rho_{max} = \max \{\rho(\mathbf{r}), \mathbf{r} \in D\}, \qquad (5)$$

and define a fuzzy membership function $\mu_X(\mathbf{r})$ for points $\mathbf{r}$ of the space belonging to molecule X as

$$\mu_X(\mathbf{r}) = \rho(\mathbf{r})/\rho_{max} \cdot \qquad (6)$$

If molecule X is not isolated, for example, if the total electron density at some point $\mathbf{r}$ can be regarded as a sum of electron densities $\rho_X(\mathbf{r})$ and $\rho_Y(\mathbf{r})$ assigned to molecules X and Y, respectively, then the fuzzy membership function of points $\mathbf{r}$ with respect to the two molecules are determined by the relative magnitudes of the individual electron densities.

Consider the general case of a molecular family L of *several* molecules

$$X_1, X_2, \ldots X_i, \ldots X_m, \qquad (7)$$

of electron density contributions

$$\rho_{X_1}(\mathbf{r}), \rho_{X_2}(\mathbf{r}), \ldots \rho_{X_i}(\mathbf{r}), \ldots \rho_{X_m}(\mathbf{r}), \qquad (8)$$

respectively, representing their "share" in the total electron density of the molecular family L.

Following the simpler case of eq.(5), first we take each individual molecule X_i, with their "share" $\rho_{X_i}(\mathbf{r})$ of the complete electron density, and this "share" will be considered as a separate, individual object in the *absence* of all other molecules. The maximum value $\rho_{max,i}$ of the electron density $\rho_{X_i}(\mathbf{r})$ within a spatial domain D_{X_i} containing all the nuclei of molecule X_i is given as

$$\rho_{max,i} = \max \{ \rho_{X_i}(\mathbf{r}),\ \mathbf{r} \in D_{X_i} \}. \tag{9}$$

The point $\mathbf{r}_{max,i}$ is defined as a point $\mathbf{r}$ where this maximum density value $\rho_{max,i}$ is realized:

$$\rho_{X_i}(\mathbf{r}_{max,i}) = \rho_{max,i} \,. \tag{10}$$

Then, in the absence of other molecules, definition (6) of the fuzzy membership function for points $\mathbf{r}$ of the space belonging to molecule X_i becomes

$$\mu_{X_i}(\mathbf{r}) = \rho_{X_i}(\mathbf{r})/\rho_{max,i} \,. \tag{11}$$

If, however, the rest of the molecules of the family L is also considered, then the actual "belonging" of a point $\mathbf{r}$ to molecule X_i is affected by the electron density contributions of all molecules, hence the above membership function is no longer appropriate. If the total electron density $\rho_L(\mathbf{r})$ of the molecular family X_1, X_2, . . . X_j, . . . X_m at point $\mathbf{r}$ is given by

$$\rho_L(\mathbf{r}) = \Sigma_j \, \rho_{X_j}(\mathbf{r}), \tag{12}$$

then the fuzzy membership function $\mu_{X_i,L}(\mathbf{r})$ for points $\mathbf{r}$ of the space belonging to molecule X_i of the molecular family is defined as

$$\mu_{X_i,L}(\mathbf{r}) = \mu_{X_i}(\mathbf{r}) \, [\rho_{max,i} \,/\, \rho_L(\mathbf{r}_{max,i})]. \tag{13}$$

The scaling factor $[\rho_{max,i} \,/\, \rho_L(\mathbf{r}_{max,i})]$ ensures proper accounting of the actual density contributions from various molecules of the family L.

This fuzzy membership function $\mu_{X_i,L}(\mathbf{r})$ can be written in another form:

$$\begin{aligned}\mu_{X_i,L}(\mathbf{r}) &= \mu_{X_i}(\mathbf{r}) \, [\rho_{max,i} \,/\, \rho_L(\mathbf{r}_{max,i})] \\ &= [\rho_{X_i}(\mathbf{r})/\rho_{max,i}\,] \, [\rho_{max,i} \,/\, \rho_L(\mathbf{r}_{max,i})] \\ &= \rho_{X_i}(\mathbf{r}) \,/\, \rho_L(\mathbf{r}_{max,i}) \,, \end{aligned} \tag{14}$$

that is, simply as

$$\mu_{X_i,L}(\mathbf{r}) = \rho_{X_i}(\mathbf{r}) / \rho_L(\mathbf{r}_{max,i}). \qquad (15)$$

The above fuzzy electron density membership functions reflect the relative contributions of the fuzzy, three-dimensional charge clouds of the various molecular electron density distributions to the total electronic density of molecular family L.

2.3 Quantum Chemical Representations of Molecular Bodies and their Subdivisions Using Fragmentation Schemes

A molecule contains a nuclear distribution and an electronic distribution; there is *nothing else* in a molecule. The nuclear arrangement is fully reflected in the electronic density distribution, consequently, the electronic density and its changes are sufficient to derive all information on all molecular properties. Molecular bodies are the fuzzy bodies of electronic charge density distributions; consequently, the shape and shape changes of these fuzzy bodies potentially describe all molecular properties. Modern computational methods of quantum chemistry provide practical means to describe molecular electron distributions, and sufficiently accurate quantum chemical representations of the fuzzy molecular bodies are of importance for many reasons. A detailed analysis and understanding of "static" molecular properties such as "equilibrium" structure, and the more important dynamic properties such as vibrations, conformational changes and chemical reactions are hardly possible without a description of the molecule itself that implies a description of molecular bodies.

If the conventional SCF LCAO *ab initio* method is used for the computation of a wavefunction of a molecule of some fixed conformation K, then the electronic density $\rho(\mathbf{r})$ can be constructed in terms of a set of atomic orbitals $\varphi_i(\mathbf{r})$ (i=1,2,...,n), where n is the number of orbitals. If $\mathbf{r}$ denotes the three-dimensional position vector variable, and if $\mathbf{P}$ is the $n \times n$ dimensional density matrix, then the electronic density $\rho(\mathbf{r})$ of the molecule can be written as

$$\rho(\mathbf{r}) = \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij}\, \varphi_i(\mathbf{r})\, \varphi_j(\mathbf{r}). \qquad (16)$$

This electron density $\rho(\mathbf{r})$ corresponds to the fuzzy "body" of the electronic charge cloud, providing a representation for the shape of the molecule.

In the general scheme described in subsequent sections, a functional group is regarded as a fuzzy body of electronic charge cloud, a fuzzy subset of the electronic charge density cloud of the complete molecule. In this context, a functional group is a special case of a fuzzy fragment of a molecular body, obtained by some subdivision scheme. In this section some aspects of subdivision schemes are discussed.

Any quantum chemical representation of molecular bodies must take into account the fuzzy nature of electron density clouds. This fuzziness can be described in mathematical terms by the fuzzy membership functions discussed in the previous section, or it can be represented geometrically by an infinite family of MIDCOs G(a) where one MIDCO belongs to every threshold value a from the open-closed interval $(0, a_{max}]$. As discussed in section 2.1 of this report, the maximum threshold value a_{max} is the electron density at a nucleus of the largest nuclear charge in the given molecule. In an alternative, topological representation [41-44, see also ref. 2 for a review], one can take advantage of the fact that for non-pathological electronic densities there are only a finite number of topologically different MIDCOs, and it is sufficient to take one MIDCO G(a) from each of the corresponding topological equivalence classes. This allows one to represent the essential shape features of the entire molecular electron density by a finite number of individual MIDCO surfaces.

The natural fuzziness of electron densities of complete molecules is not necessarily retained in all subdivision schemes of molecular electron densities into fragments. One subdivision scheme is based on the theory of catastrophes as applied to the electronic density by Collard and Hall [45]. This method, as further developed by Bader, involves boundary surfaces defined by the condition of zero flux of the electron density gradient across these surfaces [46-51]. This is a natural condition for compartmentalizing a molecular electron density cloud into "atomic regions" within a given molecule, and these compartments have many interesting properties [52-59]. However, these atomic compartments and their combinations into larger molecular fragments also have some properties which are fundamentally different from properties of electron densities of complete molecules: at all locations where such a fragment joins the rest of the molecule, the electron density of the individual fragment itself changes *discontinuously* to zero at the boundary surface. No complete molecule has such discontinuous electron density.

In this study the emphasis is placed on fragmentation schemes and functional group representations where the resulting molecular fragments and functional groups have no electron density discontinuities and show closer analogies with complete molecules. This approach is motivated by the fact that molecular fragments and functional groups often appear as individual objects that are actually transferred from one molecule to another; during the transfer their electron density changes but does not exhibit a discontinuous boundary surface. These, temporarily "independent" fragments show many analogies with complete molecules, including a fuzzy, boundaryless electron density cloud. This suggests that a uniform representation, that applies a fuzzy electron density model to complete molecules as well as to molecular fragments, has some advantages.

Some additional, quantum chemical and computational advantages of fuzzy fragments can be exploited in an approach designed to build electron densities of large molecules, partially motivated by an early approach of Christoffersen and Maggiora for wavefunction generation of large systems [60-62], by some new synthetic principles using molecular building blocks [63,64], and by the quasi-invariance of the medium and high density threshold MIDCOs [65].

An important concern is the efficient detection of local shape changes introduced by chemical changes in remote locations of a molecule. One simple approach [20] applied a truncation method, compatible with the truncation process already used within the shape group methods for molecular shape analysis [41-44].

In the following two sections two approaches will be discussed where molecular fragments are represented by fuzzy electron density models.

2.4 The Pseudo-Density Scheme of Walker

One of the alternative computational methods for diagnosing local shape variations of the electron density was suggested by Walker (P.D. Walker (1992), see also [66]), by introducing a pseudo-density matrix ${}^*\mathbf{P}^k$ of a formal molecular fragment for a subset k of the nuclei of a molecule. The pseudo-density matrix ${}^*\mathbf{P}^k$ is defined by

$$\begin{aligned} {}^*P^k{}_{ij} &= P_{ij} \quad \text{if AO } \varphi_i(\mathbf{r}) \text{ or } \varphi_j(\mathbf{r}) \text{ is centered on a nucleus} \\ &\qquad\qquad \text{that belongs to the subset } k, \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{17}$$

The resulting pseudo-density ${}^*\rho^k(\mathbf{r})$ for the formal molecular fragment involving the subset k of the nuclei is calculated as

$${}^*\rho^k(\mathbf{r}) = \sum_{i=1}^{n} \sum_{j=1}^{n} {}^*P^k{}_{ij}\, \varphi_i(\mathbf{r})\, \varphi_j(\mathbf{r}) \, . \tag{18}$$

Walker's pseudo-density ${}^*\rho^k(\mathbf{r})$ of the formal molecular fragment incorporates an enhanced contribution from the surroundings of the local molecular neighborhood, and it has been suggested as a sensitive diagnostic tool for the detection of shape differences induced by conformational or other changes at other locations within the molecule [66]. The assignment of all interaction matrix elements to the given molecular moiety involving set k of nuclei exaggerates the shape changes by overestimating the role of density interactions, and in fact at places the pseudo-density ${}^*\rho^k(\mathbf{r})$ is somewhat distorted when compared to the actual local shape of the molecule. Nevertheless, these pseudo-densities ${}^*\rho^k(\mathbf{r})$, used as a diagnostic tool, help the detection of interactions affecting shape variations.

An alternative, also distorted, density-based diagnostic tool, analogous to the Walker scheme, is introduced below. If a matrix $^{**}\mathbf{P}^k$ is constructed from all the actual density matrix elements P_{ij} *not* involving any of the nuclei of set k, and if this matrix is used to replace $^{*}\mathbf{P}^k$ in equation (18), then the resulting *pseudo-density complement* $^{**}\rho^k(\mathbf{r})$ of Walker's pseudo-density $^{*}\rho^k(\mathbf{r})$ can also be used for diagnosing interactions affecting local shape changes. Note that distortions of similar magnitude but of opposite sense are shown by $^{**}\rho^k(\mathbf{r})$ and $^{*}\rho^k(\mathbf{r})$, when compared to the actual local shapes of the molecule.

Whereas $^{**}\rho^k(\mathbf{r})$ and $^{*}\rho^k(\mathbf{r})$ are additive in the following sense,

$$^{**}\rho^k(\mathbf{r}) + {}^{*}\rho^k(\mathbf{r}) = \rho(\mathbf{r}), \tag{19}$$

these two pseudo-densities $^{**}\rho^k(\mathbf{r})$ and $^{*}\rho^k(\mathbf{r})$ have fundamentally different definitions, their roles are nonequivalent, and they cannot serve as a basis for a general, additive fragmentation scheme involving more than two parts for a molecule. In particular, if the nuclei of a molecule are classified into m subsets, $m \geq 3$, then for the corresponding local molecular moieties neither the pseudo-densities,

$$^{*}\rho^1(\mathbf{r}),\ {}^{*}\rho^2(\mathbf{r}), \ldots,\ {}^{*}\rho^m(\mathbf{r}), \tag{20}$$

nor their complements,

$$^{**}\rho^1(\mathbf{r}),\ {}^{**}\rho^2(\mathbf{r}), \ldots,\ {}^{**}\rho^m(\mathbf{r}), \tag{21}$$

generate an additive scheme, that is, their sum differs from the electron density $\rho(\mathbf{r})$ of the molecule:

$$^{*}\rho^1(\mathbf{r}) + {}^{*}\rho^2(\mathbf{r}) + \ldots + {}^{*}\rho^m(\mathbf{r}) \neq \rho(\mathbf{r}) \tag{22}$$

and

$$^{**}\rho^1(\mathbf{r}) + {}^{**}\rho^2(\mathbf{r}) + \ldots + {}^{**}\rho^m(\mathbf{r}) \neq \rho(\mathbf{r}) \qquad (22)$$

If the purpose of the analysis is the detection of shape changes, then these pseudo-densities and their complements are good diagnostic tools; however, the pseudo-density scheme is not suitable to build approximate electron densities of large molecules from pseudo-density fragments obtained from smaller molecules.

Walker's pseudo-densities and their complements are designed for enhanced detection of local shape variations and are valuable for local shape analysis. Some more advanced variants of non-additive pseudo-densities are described in ref. [66].

2.5 The Mulliken-Mezey Additive Fuzzy Electron Density Fragmentation Method

A new range of applications becomes possible if a fuzzy density fragmentation method is *additive*. The simplest of these additive fuzzy electron density fragmentation methods was proposed by the author (P.G. Mezey (1992), first applied in ref. [67] and generalized in [5,37]). In a formal sense, the additive density fragmentation method can be regarded as an "atom-group population analysis without integration", and has been motivated by Mulliken's population analysis [68,69]. Consider a molecular electronic density $\rho(\mathbf{r})$ and an arbitrary collection k of the nuclei of the given molecule. According to the simplest version of this method, the k-th fuzzy electron density fragment $\rho^k(\mathbf{r})$ is defined in terms of Mezey's *fragment density matrix,* denoted by $\mathbf{P}^k$.

The elements $P^k{}_{ij}$ of $\mathbf{P}^k$ are defined as follows:

$$\begin{aligned} P^k{}_{ij} &= P_{ij} \text{ if both } \varphi_i(\mathbf{r}) \text{ and } \varphi_j(\mathbf{r}) \text{ are AO's centered on nuclei of the fragment,} \\ &= 0.5\, P_{ij} \text{ if only one of } \varphi_i(\mathbf{r}) \text{ and } \varphi_j(\mathbf{r}) \text{ is centered on a nucleus of the fragment,} \\ &= 0 \text{ otherwise.} \end{aligned} \tag{23}$$

This fragment density matrix $\mathbf{P}^k$ has the same n × n dimensions as that of the density matrix $\mathbf{P}$ of the complete molecule.

Based on the fragment density matrix $\mathbf{P}^k$ for the k-th fragment, the electron density of Mezey's *fuzzy density fragment* $\rho^k(\mathbf{r})$ is defined as

$$\rho^k(\mathbf{r}) = \sum_{i=1}^{n} \sum_{j=1}^{n} P^k{}_{ij}\, \varphi_i(\mathbf{r})\, \varphi_j(\mathbf{r})\,. \tag{24}$$

Consider a molecule and assume that its nuclei are distributed into m mutually exclusive families, where m can take any positive integer value. As follows from the defining equation (23) of the fragment density matrix, the sum of Mezey's fragment density matrices $\mathbf{P}^k$ is equal to the density matrix $\mathbf{P}$ of the molecule:

$$P_{ij} = \sum_{k=1}^{m} P^k{}_{ij} \tag{25}$$

that is,

$$\mathbf{P} = \sum_{k=1}^{m} \mathbf{P}^k. \tag{26}$$

Since the density expressions (16) and (24) are linear in the corresponding density matrices, the sum of fragment densities $\rho^k(\mathbf{r})$ is equal to the density $\rho(\mathbf{r})$ of the molecule:

$$\rho(\mathbf{r}) = \sum_{k=1}^{m} \rho^k(\mathbf{r}). \tag{27}$$

In this fragmentation scheme both the fragment density matrices $\mathbf{P}^k$ and the fragment densities $\rho^k(\mathbf{r})$ are additive, that is, Mezey's scheme is an additive, fuzzy electron density fragmentation scheme. Whereas the method does not in general exaggerate the interactions affecting local shape features, hence for diagnostic purposes it is not as sensitive as Walker's pseudo-density method, however, this additive method avoids the significant distortions of the pseudo-density scheme.

The fuzzy electron density fragment additivity rules (23) - (27) are *exact* at any given *ab initio* LCAO level, hence the reconstruction of the calculated electronic density $\rho(\mathbf{r})$ of the given molecule from the corresponding fuzzy fragment electron densities $\rho^k(\mathbf{r})$ is also *exact*.

Generalizations of this additive fuzzy electron density fragmentation scheme have been proposed in refs. [5,37]. One of these more flexible schemes for assigning interfragment interactions is described below, leading to a generalization [5,37] of the original fuzzy electron density fragment additivity principle:

$$
\begin{aligned}
P^k{}_{ij} &= P_{ij} \quad \text{if both } \varphi_i(\mathbf{r}) \text{ and } \varphi_j(\mathbf{r}) \text{ are AO's centered on nuclei of fragment k,}\\
&= w(k,i,j)P_{ij} \quad \text{if only one of the AO's } \varphi_i(\mathbf{r}) \text{ and } \varphi_j(\mathbf{r}) \text{ is centered on a nucleus of fragment k, where for the weighting factors both conditions } w(k,i,j) > 0, \text{ and } w(k,i,j)+ w(k',i,j) = 1 \text{ hold, and where fragment k' contains the nucleus for the other AO,}\\
&= 0 \quad \text{otherwise.}
\end{aligned} \tag{28}
$$

Some additional conditions must be fulfilled if the function w(k,i,j) is to generate a physically valid additive scheme. One choice relies on a sign-preserving scalar property A(i) that can be assigned to atomic orbitals. For example, appropriately scaled electronegativity can be used as such a scalar property. The choice of

$$w(k,i,j) = A(i)/[A(i)+A(j)], \tag{29}$$

where AO $\varphi_i(\mathbf{r})$ is centered on a nucleus that belongs to the k-th family, generates a valid scheme.

Another, more convenient form of Mezey's fragmentation scheme can be given in terms of membership functions of nuclei within various molecular fragments. The set of all nuclei of the molecule is divided into m mutually exclusive families, denoted by

$$f_1, f_2, \ldots, f_k, \ldots f_m, \tag{30}$$

assocciated with m density fragments,

$$F_1, F_2, \ldots, F_k, \ldots F_m, \tag{31}$$

respectively, where the earlier notation is used for the fragment density functions

$$\rho^1(\mathbf{r}), \rho^2(\mathbf{r}), \ldots, \rho^k(\mathbf{r}), \ldots \rho^m(\mathbf{r}). \tag{32}$$

We denote by $m_k(i)$ the membership function of AO $\varphi_i(\mathbf{r})$ in the set of AOs centered on a nucleus of nuclear set f_k of fragment F_k. Formally, this membership function is defined as follows:

$$m_k(i) = 1 \text{ if AO } \varphi_i(\mathbf{r}) \text{ is centered on one of the nuclei of set } f_k,$$

$$0 \text{ otherwise.} \tag{33}$$

Using these membership functions $m_k(i)$, the elements $P^k{}_{ij}$ of Mezey's $n \times n$ fragment density matrix $\mathbf{P}^k$ for the k-th fragment F_k can be written as

$$P^k{}_{ij} = 0.5\, [m_k(i) + m_k(j)]\, P_{ij}\,. \tag{34}$$

The generalized fragmentation scheme can also be introduced using the membership function formalism of eq. (33), by taking

$$P^k{}_{ij} = [m_k(i)\, w_{ij} + m_k(j)\, w_{ji}\,]\, P_{ij}\,, \tag{35}$$

where the following constraint applies for the weighting factors:

$$w_{ij} + w_{ji} = 1. \tag{36}$$

The original additive fuzzy fragmentation scheme corresponds to the choice of

$$w_{ij} = w_{ji} = 0.5. \tag{37}$$

For the generalized scheme, involving a scalar property $A(i)$, the function w_{ij} can be chosen as

$$w_{ij} = A(i)/[A(i)+A(j)], \qquad (38)$$

an expression analogous to that given in eq. (29). This formulation differs from the previous one by no longer having any need to specify the fragment index k in the weighting factor w_{ij}, since the membership function formalism ensures proper additivity.

The additive fuzzy electron density fragmentation scheme of Mezey is the basis of the Molecular Electron Density Lego Assembler (MEDLA) method [67,70-72], reviewed in section 4. of this report, where additional details and applications in local shape analysis are discussed. The MEDLA method was used for the generation of the first *ab initio* quality electron densities for macromolecules such as proteins [71,72] and other natural products such as taxol [66].

3. THE DENSITY DOMAIN APPROACH TO FUNCTIONAL GROUPS AND LOCAL MOLECULAR PROPERTIES

3.1 The Density Domain Concept

The density domain approach was first proposed [4] as a tool for the description of chemical bonding where the complete shape information of the molecular electron density was taken into account. Density domains are formal bodies of electron density clouds enclosed by MIDCOs defined by eq. (1) [or by eq. (2) if there is no need to specify the nuclear configuration K].

The electronic charge density function $\rho(\mathbf{r},K)$ becomes exactly zero only at infinite distance from the nuclei of the molecule; however, in an approximate sense the electronic density converges to zero very rapidly already at few bohrs from the nuclei. In fact, the electronic density charge cloud is approximately localized near the nuclei. For computational purposes, one may take a small enough threshold a and regard only those regions of the 3D space to belong to the molecule of nuclear configuration K, where the electronic density $\rho(\mathbf{r},K)$ is larger than this threshold.

This approach, however, can be generalized for any threshold value a, and one may consider the fuzzy molecular body as being viewed at various density thresholds, for the whole range of possible densities. An infinite family of formal bodies is obtained, and all these formal bodies, collectively, represent the actual *molecular body*. For each threshold value a, the formal, threshold-dependent molecular body is the *density domain* DD(a,K), defined as the collection of all those points $\mathbf{r}$ of the 3D space where the electronic density is greater than or equal to the threshold a,

$$DD(a,K) = \{ \mathbf{r} : \rho(\mathbf{r},K) \geq a \} . \tag{39}$$

If the specification of the nuclear configuration K is not required, then one may simply write

$$DD(a) = \{ \mathbf{r} : \rho(\mathbf{r}) \geq a \} . \tag{40}$$

Since the electronic density $\rho(\mathbf{r},K)$ is a continuous function of the 3D space variable $\mathbf{r}$, the set of all points $\mathbf{r}$ fulfilling the conditions of defining equation (2) of MIDCOs, that is, the set of boundary points of the density domains defined above, does form a continuous surface.

A pictorial analogy between macroscopic clouds of various densities and molecular charge densities can be used here. A density domain DD(a) is analogous to a cloud we could see if our eyes were adjusted to notice only densities equal to or higher than the threshold a. By readjusting our eyes, different "density domains" of clouds of higher or lower density threshold values could be observed.

Most properties of density domains follow from the properties of MIDCOs. We have seen before that for low values of the electron density threshold a, the MIDCO G(a,K) is usually a single, closed surface, consequently, the density domain DD(a,K) is also a single, connected body. For high values of density threshold a, the MIDCO G(a,K) is often a collection of several closed surfaces, where each closed surface surrounds some of the nuclei of the molecule. Consequently, for such a density threshold, the formal density domain DD(a,K) is in fact a collection of several, disconnected bodies $DD_i(a,K)$.

A density domain DD(a,K) represents a formal *molecular body at an electronic density threshold* a *and nuclear configuration* K. A body DD(a,K) may be a single piece or it may be a collection of several disconnected pieces, the maximum connected components $DD_i(a,K)$ of DD(a,K):

$$DD(a,K) = \bigcup_{i} DD_i(a,K). \tag{41}$$

In precise mathematical terminology, the term "domain" should be applied only to a single maximum component. However, for simplicity, the term "density domain" is used for both the individual maximum connected components $DD_i(a,K)$ and the union DD(a,K) in eq. (41).

The entire electronic density of a molecule can be represented by an infinite family of density domains, a family that contains one set $\{DD_i(a,K)\}$ of density domains for every threshold value a from the open-closed interval $(0, a_{max}]$, where, as in eq. (2), a_{max} can be chosen as the electron density at a nucleus of the largest nuclear charge in the molecule. Due to the nesting properties of MIDCOs for threshold values a and a' fulfilling condition (4), a density domain DD(a,K) of lower density threshold value a contains density domains DD(a',K) of higher threshold values a'.

Both the shape and the size of a density domain DD(a,K) depend on the choice of the threshold value a. Selecting an arbitrary value for a is highly unlikely to result in a valid description and it is better to consider a whole range of possible threshold values a for the density domains DD(a,K). Some important simplifications are possible. First of all, it is not necessary to consider very high a values which occur only in the vicinity of heavy nuclei. By selecting an appropriate small but nonzero lower limit for the threshold values, formal density domains involving two distant, essentially independent molecules can also be eliminated from the analysis. For the purposes of molecular shape analysis, additional and more fundamental simplifications can be obtained by applying topological methods and the GSTE principle [2] (Geometrical Similarity as Topological Equivalence).

If the electron density is regarded as a function $\rho(\mathbf{r},K)$ of the nuclear arrangement K, then the MIDCOs G(a,K) and the enclosed density domains DD(a,K) are also functions of the nuclear configuration K. How these functions vary with K is one of the main questions of chemistry. It is worthwhile to study this question using the nuclear configuration space approach [2,40].

The set of all possible nuclear configurations K of a given stoichiometric family of N nuclei defines a formal nuclear configuration space. This space can be defined so that a proper distance function d(K,K') can be specified that applies to any two nuclear arrangements K and K'. This turns the configuration space into a (3N-6)-dimensional metric space M [40]. This metric configuration space has many properties analogous to those of the ordinary, three-dimensional space, but it also has fundamentally different properties that are a source of many misconceptions. For example, the nuclear configuration space M is not a vector space, and what is especially misleading, it cannot even be turned into one [40].

If the electronic state is specified, than a formal potential energy value can be assigned to each nuclear arrangement K. The molecular potential energy of the given electronic state is a function of the nuclear arrangements K, and it can be pictured as a hypersurface E(K) over the space M [40]. Each chemical species, such as a stable conformation vibrating about a configuration corresponding to an energy minimum, can be associated with an open set, called a *catchment region* of the given energy hypersurface E(K). The catchment region model formally relaxes the classical constraints of rigid nuclear geometry and avoids the conflict with the Heisenberg uncertainty relation [40]. The concept of dynamic molecular shape can be associated with an entire family of $\rho(\mathbf{r},K)$ charge density distributions occurring within the catchment region representing the given stable, vibrating conformation of the molecule [2].

3.2 Chemical Bond as a Density Domain Relation

One of the most deeply held notions of chemistry is the concept of chemical bond. Such bonds, almost always depicted as formal lines between atomic symbols, have their counterpart within the quantum chemical description of electron densities; often [45-51], but not always [54-60], a correspondence can be made between conventional chemical bonds and specific lines defined by the electron density gradient maps. These lines usually interconnect nuclear positions. However, as has been pointed out by Cioslowski [54-60], there are cases where this correspondence fails.

It is clear that the entire electronic density in a molecule has the role of determining the nuclear distribution hence bonding, consequently, chemical bonding cannot be confined to lines in space. It is well understood that bond diagrams represent only an oversimplified, "short-hand" notation for the actual molecular structure, nevertheless, as most successful notations do, chemical bonds as formal lines have acquired an almost unquestioned reputation of their own as if they were truly responsible for holding molecules together.

Formal chemical bonds as lines in space represent only a drastically oversimplified representation of chemical bonding, a mere skeletal model, introduced and in use since the early days of chemistry when there was no hope yet to detect, model, visualize, and understand the intricate, fuzzy, three-dimensional features and the wealth of shape information of molecular electron densities.

A more realistic model is obtained by regarding each molecule as a collection of atomic nuclei embedded in an electron density cloud interacting with the nuclei, where quantum mechanical uncertainty affects the motion of electrons as well as nuclei. The electronic density closely "follows" the nuclear motions, nearly instantaneously readjusting to new nuclear arrangements. In simplistic terms, electronic density is a deformable "glue" holding the mutually repelling nuclei together, where this glue is thick near the nuclei and becomes thinner and thinner further away from nuclear locations, eventually fading out at large distances. The bonding is provided by the *entire body* of this bondaryless "blob" of glue. A description of chemical bonding compatible with the above picture requires a fully three-dimensional model, where the boundaryless, "fading out" features of electronic density is properly represented.

Today we can easily obtain detailed shape information on electron densities, by experiment or by quantum chemical computations, and the best utlization of these methodologies, possibly in combination, is one fertile area of research [73-84]. Yet, there has been little progress in replacing the simplistic bond-line diagrams with more accurate, more descriptive, and better predictive models of chemical bonding.

There is no longer any inherent difficulty in computing reasonably accurate electron densities for small molecules by either of the two main computational approaches: wavefunction methods [85,86] and density functional methodologies [87-89]. With the introduction of the MEDLA technique [67,70], *ab initio* quality electron densities can be computed for virtually any macromolecule, including proteins [66,71,72]. The actual molecular body, the electronic density cloud, has become accessible to rigorous study. Whereas conventional chemical bond diagrams are still useful, they are no longer a choice dictated by necessity, and today reasonably accurate electron density descriptors are easily available for anyone with a workstation. It is surprising how little mainstream chemical thinking has taken advantage of the new computational possibilities, and of the new, deeper understanding accurate electron density computations have already provided. Most of the efforts of contemporary electron density analysis have been focused on justifying the old, skeletal model of chemical bonds as formal lines in space, ignoring the rich details the full, three-dimensional fuzzy electron density has to offer. Today, when detailed electron densities are readily available, the conventional "skeletal" image of chemical bonding using line diagrams is more an impediment than help.

One of the exceptions, that offers an alternative to the conventional bond diagrams is the density domain approach [4,5] to chemical bonding. This approach is based on the following observation: for a given molecule with a specified nuclear configuration K, the infinite family $\{DD(a,K)\}$ of density domains for the range $(0, a_{max}]$ of density thresholds,

$$\{ DD(a,K), a \in (0, a_{max}] \}, \tag{42}$$

contains only a *finite number* of topologically different density domains. The density domains in the above infinite family can be classified into a finite number of topological equivalence classes, representing the essential features of the bonding within the molecule. The topological aspects of the bonding pattern of the three-dimensional electron density can be represented by a finite number of density domains, one taken from each equivalence class. The sequence of critical points of the electronic density can produce the same classification of density domains, however, the location and electron density values of these critical points contain much less shape information than the density domains themselves. It is important to realize that within each equivalence class there are more invariant shape features than the mere pattern of interconnectedness of local regions. These shape features are present in the actual density domains selected to represent each equivalence class [2].

Consider a molecule with a fixed nuclear configuration K and select an initial high density threshold a_{max}. By gradually decreasing this threshold to a small enough value a such that all the nuclei of the molecule are enclosed within a single density domain DD(a,K), one may study the gradual buildup of the electron density in the molecule, and both the sequence of interconnection patterns and the actual shapes and shape changes associated with this imaginary process. By monitoring the topological and essential geometrical changes (where the latter can also be described by topological, in fact algebraic, means [2]), a detailed description of chemical bonding is obtained.

It is useful to indicate which nuclei are enclosed within each maximum connected component, that is, within each density domain $DD_i(a)$. If the nuclei A, B, etc., are enclosed within a density domain component $DD_i(a_j)$ at a threshold a_j, then the detailed notation $DD_i(a_j, K; A, B, \ldots)$ can be used.

As an illustration, we take the simple example of the equilibrium arrangement of the methane molecule from ref.[3]. Using a semiclassical model and a fixed nuclear configuration K, there are only three topologically distinguishable sets of density domains of the methane molecule. At a high density threshold a_1, electronic density is found only near the carbon nucleus, hence there is only a single density domain. That is, at the high density threshold a_1 the set of density domains has a single element,

$$\{DD_1(a_1,K;\ C)\}. \tag{43}$$

At some medium density threshold a_2, symmetry implies that electronic density clouds appear simultaneously around each of the four hydrogen nuclei; one finds that there are five separate density domains. That is, at an intermediate density threshold a_2, the set of density domains has five elements,

$$\{DD_1(a_2,K;C),\ DD_2(a_2,K;H),\ DD_3(a_2,K;H),\ DD_4(a_2,K;H),\ DD_5(a_2,K;H)\}. \tag{44}$$

At some low enough density threshold a_3, the methane molecule has a single density domain that contains all five nuclei C, H, H, H, and H. At this density threshold a_3 the set of density domains has again only a single element,

$$\{DD_1(a_3,K;\ C,H,H,H,H)\}. \tag{45}$$

The shapes of these density domains are characteristic to the set of nuclei enclosed by them, to the nuclear geometry, and also to the location of these density domains within the molecule, collectively represented by the configuration variable K, as well as to the actual density threshold a. The sequence of density domains as a function of density threshold a, augmented with the results of a local shape analysis of these density domains [2], provides a detailed description of chemical bonding within the methane molecule.

A gradual decrease of the density threshold a, starting with a high value [2], is usually accompanied with a series of typical topological changes of the density domains. These changes provide a classification of density domains according to the various ranges of the density threshold parameter a.

The various density ranges form a hierarchy [2], indicated by boldface, underlining, and italics, as well as indentation within the list below. If a subrange in the hierarchy happens to coincide with a range, then its name is given in parentheses after the name of the range.

Localized density range, $[a_{max}, a_g)$: a_{max} is the maximum electron density in the molecule, whereas a_g is the threshold below which only a single density domain is found.

Atomic range, $[a_{max}, a_f)$: a_{max}, is defined above, a_f is the lowest density threshold where two density domains join. Only individual nuclear neighborhoods appear as disconnected density domains, that is, there is precisely one nucleus within each density domain which appears.

Strictly atomic range, $[a_{max}, a_p)$: a_{max}, is defined above, a_p is the lowest density threshold where at least one density domain is no longer convex, as it "reaches out" to join a neighboring density domain. Note that within the strictly atomic range all density domains are convex, and each density domain contains precisely one nucleus.

Prebonding range, $[a_p, a_f)$: both a_p and a_f are defined above. Within the prebonding range each density domain contains precisely one nucleus, but not all density domains are convex.

Functional group range (*bonding range for density domains*), $[a_f, a_g)$: both a_f and a_g are defined above. Within the functional group range at least one density domain contains more than one nucleus, but not all nuclei of the molecule can be found within a single density domain. Those density domains with more than one nucleus represent "complex" functional groups, as contrasted to "simple" functional groups represented by density domains containing a single nucleus. In this density range one finds the various "complex" functional groups as individual entities. Also in this range, the bonding pattern (the actual interconnection pattern) of the density domains is revealed. Consequently, this range is referred to as the functional group range or the bonding range for density domains.

Global density range (molecular density range), $[a_g, a_{min})$: a_g is defined above, whereas a_{min} is a low threshold value below which electron density can be neglected. The molecule is represented by a single, connected density domain for each threshold value within the global density range.

Skinny molecular range, $[a_f, a_c)$: a_f is defined above, whereas a_c is the maximum threshold at and below which all locally nonconvex domains on the surface of density domains are simply connected. In simpler terms, in the skinny molecular range all nuclei are found within a single density domain, but there are formal "neck" regions on the surface of density domains. In the terminology of shape group analysis [2], rings of D_1 type can be found on the surface of density domains.

Corpulent molecular range, $[a_c, a_q)$: a_c is defined above, whereas a_q is the maximum threshold at and below which all density domains are convex. Within the corpulent molecular range no local "neck" regions are found, but the density domains are not convex.

Quasi-spherical molecular range, $[a_q, a_{min})$: both a_q and a_{min} are defined above. For any threshold value within the quasi-spherical range the density domain representing the molecule is convex. If a_{min} is small enough, then in the strict mathematical sense, for very low density thresholds all molecules have convex density domains.

3.3 The "Late - Early Rule"

The gradual decrease of the electron density threshold value reveals many interesting trends. One such trend is called the "Late - Early Rule" [2].

The Late - Early Rule:

If the threshold density a *is gradually decreased in a density domain analysis of a given conformation K of a molecule, then the order of joining of various, individual density domains of a given atom type A to neighboring density domains tends to be the reversed order of their appearance.*

This trend, also not necessarily followed in all instances, can be justified by simple electronegativity considerations. Consider a molecule that contains several nuclei of the same type A, in different local suroundings. In one surrounding, a higher electronegativity neighbor of A is likely to delay the appearence of the density domain of A to a lower density threshold. The density domain about this nucleus A appears *late.* A highly electronegative atom is also likely to have a more extensive charge density cloud about it, implying that an *early* joining of the density domains of this neighbor and this nucleus A is likely. The individual density domain of this nucleus A exists in a narrower threshold interval than the density domains of other nuclei of the same type A having no highly electronegative nuclei within their surroundings. This implies that the threshold interval for the individual existance of a density domain for our selected nucleus A is expected to be contained within the corresponding intervals of other density domains containing the same nucleus type A.

The trend described by the Late - Early rule can be phrased in terms of the expected inclusion relations of those electron density threshold intervals where the density domains of individual nuclei exist as separate entities. A given set of the nuclei of type A is ordered according to the increasing effective electronegativity of the neighbor nuclei $B^{(k)}$ involved in the *first* merger of each of the density domains of nuclei of type A. This ordering is represented by the index k:

$$A^{(k)}, \quad k=1,2,...q, \qquad (46)$$

$$B^{(k)},\ \ k=1,2,...q. \tag{47}$$

Nucleus $B^{(1)}$ is the first nucleus to share its density domain with that of $A^{(1)}$, and nucleus $B^{(q)}$ is the first nucleus to share its density domain with that of $A^{(q)}$; the effective electronegativity of nucleus $B^{(1)}$ is the lowest, whereas the effective electronegativity of nucleus $B^{(q)}$ is the highest.

The interval of density thresholds where a separate density domain exists for nucleus $A^{(k)}$ is denoted by (a_k, a'_k) . Using these notations, the trend expressed by the "Late - Early Rule" is equivalent to the following, ideal sequence of inclusion relations:

$$(a_1, a'_1) \supset (a_2, a'_2) \supset (a_3, a'_3) \supset ..... \supset (a_k, a'_k)..... \supset (a_q, a'_q). \tag{48}$$

Exceptions to the Late-Early Rule are known, however, the rule represents the expected trend.

One example that has been studied in some detail is the ethanol molecule [2]. Let us choose H as nucleus A. The density domain of the OH proton appears relatively late in the process of gradually decreasing the electron density threshold. This observation can be justified by the high electronegativity of oxygen, resulting in a depletion of the electron density at the nearby proton, that has a chance for the formation of a density domain of its own only at a somewhat lower density threshold.

We combine this observation with another: when generating a seqence of density domains by gradually decreasing the electron density threshold a, the most acidic proton of the ethanol molecule loses its separate density domain first. Considering all the minimum values of charge densities separating protons from other nuclei, the minimal electronic charge density is of the highest value between this OH proton and the nearest nucleus, the oxygen nucleus. At first sight, this might appear as a conflict with the known fact that this proton, the most acidic in the molecule, is the most easily donated to proton acceptors. This apparently counterintuitive pattern is a consequence of the high electronegativity and the resulting, more extensive high density domain about the oxygen nucleus. Apparently, the high electron density cloud about the oxygen engulfs this particular proton, resulting in the highest density threshold "bridgehead" between any proton and another nucleus within the ethanol molecule.

3.4 A Density Domain Definition of Functional Groups

The hierarchy of density domains discussed in Section 3.2 has already indicated a natural density domain condition for a functional group. If a family of nuclei has the property within a given molecule that there exists a threshold a such that a corresponding connected density domain contains all of these nuclei while separating them from the rest of the nuclei of the molecule, then these nuclei form the nuclear family of a functional group. The fact that such a separate density domain exists for these nuclei implies that the part of the electronic density dominated by these nuclei forms an entity that has some identity of its own, and some limited "autonomy" within the molecule. This observation justifies taking this density domain as that of a functional group within the molecule. There is still the problem of assigning an electron density contribution to this functional group at lower densities, for example, for density thresholds within the molecular range. This question will be addressed in sections 4.4, 5.3, and 5.4, where local electron density representations and local shape analysis of molecules are discussed.

One of the main advantages of the density domain approach is the introduction of a natural model for a quantum chemical representation of formal functional groups [1-3]. Consider the simplest case: a single connected density domain DD(a,K) and all the nuclei contained within DD(a,K). The boundary MIDCO G(a,K) of the density domain DD(a,K) separates this subset of the nuclei of the molecule from the rest of the nuclei. This fact indicates that the nuclei embedded within DD(a,K), together with a local electronic density cloud surrounding them, represent a sub-entity of the molecule. This sub-entity has an individual identity, since for a range of density threshold values including the value a, the local electron density cloud is separable from the density cloud of the rest of the molecule.

We may consider some chemical examples. Several alcohols, including ethanol and allyl alcohol, have been studied using the density domain shape analysis approach [2,3], and in all these cases a whole range [a',a"] of density threshold values have been found within which the O and H nuclei of the OH group are completely surrounded by MIDCO's, separating these nuclei from all the other nuclei of the molecule. This criterion, the existence of a MIDCO that separates a group of nuclei from all other nuclei of a molecule, is used for the identification and a detailed characterization of chemical functional groups [1-3].

Most actual density domains $DD_i(a_j)$ and the nuclei enclosed by them can be transferred from one molecule to another without altering the topology of the density domains. Whereas some density domains may change their topology in the course of such a transfer if the new environment is sufficiently different, nevertheless, the density domain approach is a criterion derived from physical conditions, specifying those molecular moieties, at the given density threshold a, which can be regarded to have a well defined chemical identity. This approach represents a generalization of the usual chemical concept of functional groups: according to this generalized concept, a collection of all nuclei within a density domain $DD_i(a_j)$, together with the density domain itself can be regarded as a functional group of the molecule at the density threshold a_j.

The actual bonding pattern can be regarded as justification of this approach. Consider a family of MIDCOs for a range of threshold values a that separates a group of nuclei from the rest of the nuclei of the molecule. This indicates a stronger chemical linkage among the local charge density clouds surrounding the nuclei within the given density domain, whereas the linkage between this group and the rest of the molecule is weaker. This is a natural property for a chemically recognizable functional group. Note, however, that in actual computational studies, for example, if the conventional *ab initio* Hartree-Fock-Roothaan-Hall method [90-93] is used within the gaussian basis set formalism of the Gaussian program system [85], the calculated density domains are dependent on the level of basis set and the quantum chemical methods used, hence conclusions whether a group of nuclei represents a functional group may be basis set dependent.

The density domain DD(a,K) can then serve as a natural representative of a formal functional group. Note that this concept is rather general; by this condition, most individual atomic neighborhoods, as well as many parts of molecules seldom exhibiting special chemical properties or reactivities qualify as formal "density domain functional groups". Whereas in some instances the quantum chemical concept of "density domain functional group" deviates somewhat from the usual, intuitive, organic chemistry concept of functional group, the density-based definition provides interesting insight into the roles and "autonomy" of various molecular parts [2,3].

Additional advantages have been pointed out in the Introduction. Since density domains play a major role in molecular shape analysis and in the construction of various molecular similarity measures [5], shape analysis and molecular similarity can be formulated in terms of quantum-chemically defined functional groups. This model is also compatible with a rather general, algebraic-geometrical framework discussed in ref.[6].

3.5 Functional Groups as Fuzzy Subsets of Molecular Electron Density

The early recognition that quantum mechanical uncertainty [94] of electronic arrangement and motion within a molecule [95,96] has a special role in chemistry has been fundamental in the development of quantum chemistry. In general, the uncertainty itself can be represented by fuzzy set methods, applied in a wide range of disciplines [97-103], and both general quantum mechanical [104-108] and quantum chemical applications [40,52,109-111] have led to novel descriptions of physical and chemical properties.

With minor modifications, the fuzzy electron density membership function formalism of molecular families can also be applied to a family of functional groups within a molecule. Consider a molecule X and some electron density threshold a within the functional group range of density. Consider the functional groups appearing as separate density domains

$$DD_1(a,K),\ DD_2(a,K),\ \ldots,\ DD_m(a,K) \qquad (49)$$

at this density threshold a, and for simplicity denote them by the symbols

$$F_1, F_2, \ldots, F_m. \qquad (50)$$

The electron density contribution $\rho^i(\mathbf{r})$ of each functional group F_i is calculated using the Mezey scheme (eqs. 23, 24) for additive fuzzy density fragmentation of the molecular density $\rho_X(\mathbf{r})$, by simply taking the nuclear set k for each fuzzy fragment density as the nuclear set embedded in the corresponding density domain $DD_i(a,K)$ representing functional group F_i.

The fuzzy fragment electron density contributions

$$\rho_{F1}(\mathbf{r}), \rho_{F2}(\mathbf{r}), \ldots \rho_{Fi}(\mathbf{r}), \ldots \rho_{Fm}(\mathbf{r}), \qquad (51)$$

respectively, represent the "share" of each functional group F_i within the total electron density $\rho_X(\mathbf{r})$ of molecule X.

The fuzzy electron density membership functions of functional groups described below reflect the relative contributions of the fuzzy, three-dimensional electron density charge clouds to the total electronic density of molecule X.

For the derivation of appropriate fuzzy membership functions for the electronic densities of functional groups, first we take each individual functional group F_i, with its share $\rho_{Fi}(\mathbf{r})$ of the complete electron density of molecule X, and consider this share as a separate, individual object in the absence of all other functional groups of family $F_1, F_2, \ldots, F_m$ of molecule X. Note that the special value a of the density threshold is used only to identify some of the possible functional groups of molecule X, the very family $F_1, F_2, \ldots, F_m$, and the actual functional group electron densities $\rho_{F1}(\mathbf{r}), \rho_{F2}(\mathbf{r}), \ldots, \rho_{Fm}(\mathbf{r})$ are not restricted to any specific threshold value.

Also note that if a different threshold value a' is chosen, a different assignment of nuclei to individual density domains may result, that is, a different set of functional groups may be identified within the same molecule X. Take the example of the ethanol molecule [3]. At some threshold a, the nuclei of the CH_2 group are separated from all other nuclei by a MIDCO, hence at this density threshold a, CH_2 is manifested as a functional group within ethanol. However, at a somewhat lower electron density threshold a', only two nuclear families are separated by MIDCOs, the nuclei of the CH_3CH_2 and OH moieties. Hence, at this density threshold a', the ethyl group and the hydroxyl group are manifested as functional groups within ethanol, and the fact the methylene group CH_2 is also a functional group is not revealed at the second density threshold a'.

The maximum value $\rho_{max,i}$ of the electron density $\rho_{Fi}(\mathbf{r})$ within a spatial domain D_{Fi} containing all the nuclei of functional group F_i is used as a reference:

$$\rho_{max,i} = \max \{ \rho_{Fi}(\mathbf{r}), \mathbf{r} \in D_{Fi} \}. \qquad (52)$$

A point $\mathbf{r}_{max,i}$ is selected as a point $\mathbf{r}$ where this maximum density value $\rho_{max,i}$ is realized for the given functional group:

$$\rho_{F_i}(\mathbf{r}_{max,i}) = \rho_{max,i} \,. \tag{53}$$

Then, by taking functional group F_i in the absence of other functional groups, and by adapting the approach described in section 2.2 for individual molecules from a molecular family L, the definition of the fuzzy membership function for points $\mathbf{r}$ of the space belonging to functional group F_i is given as

$$\mu_{F_i}(\mathbf{r}) = \rho_{F_i}(\mathbf{r})/\rho_{max,i} \,. \tag{54}$$

The fuzzy membership function defined above reflects the actual electronic charge distribution of functional group within the given molecule X, without directly involving any other density contributions from other functional groups of the molecule.

For practical purposes, an alternative model provides more insight. If the density contributions of the rest of the functional groups of molecule X are also considered, then the actual degree of "belonging" of a point $\mathbf{r}$ to a given functional group F_i is also affected by the electron density contributions of all other functional groups from the family $F_1, F_2, \ldots, F_m$. Consequently, the membership function $\mu_{F_i}(\mathbf{r})$ defined by eq. (54) is no longer appropriate. An alternative fuzzy set approach, including the effects of the electron density contributions of all other functional groups of the molecule is based on Mezey's additive fuzzy electron density fragmentation method.

Molecule X is regarded as a collection of appropriately arranged, mutually interpenetrating electron densities of the functional groups $F_1, F_2, \ldots F_j, \ldots F_m$, where each fuzzy electron density fragment F_j contains the corresponding set of nuclei. As a consequence of the exact additivity property of Mezey's fragmentation scheme, at each point $\mathbf{r}$ the total electronic density $\rho_X(\mathbf{r})$ of molecule X is given as the sum of functional group electron densities:

$$\rho_X(\mathbf{r}) = \Sigma_j \, \rho_{F_j}(\mathbf{r}). \tag{55}$$

If the simultaneous presence of all functional groups $F_1, F_2, \ldots F_i, \ldots F_m$ within molecule X is taken into account, then a new fuzzy membership function $\mu_{F_i,X}(\mathbf{r})$ for points $\mathbf{r}$ of the space belonging to functional group F_i of molecule X can be defined as

$$\mu_{F_i,X}(\mathbf{r}) = \mu_{F_i}(\mathbf{r}) \, [\rho_{max,i} / \rho_X(\mathbf{r}_{max,i})]. \tag{56}$$

The scaling factor $[\rho_{max,i} / \rho_X(\mathbf{r}_{max,i})]$ in the above equation is required in order to reflect the actual density contributions from various functional groups F_i of molecule X.

This fuzzy membership function $\mu_{Fi,X}(\mathbf{r})$ can also be written in a form of a simple density ratio:

$$\mu_{Fi,X}(\mathbf{r}) = \mu_{Fi}(\mathbf{r}) \, [\rho_{max,i} / \rho_X(\mathbf{r}_{max,i})]$$

$$= [\rho_{Fi}(\mathbf{r})/\rho_{max,i}] \, [\rho_{max,i} / \rho_X(\mathbf{r}_{max,i})]$$

$$= \rho_{Fi}(\mathbf{r}) / \rho_X(\mathbf{r}_{max,i}) \, . \qquad (57)$$

This ratio, expressed for each functional group F_i,

$$\mu_{Fi,X}(\mathbf{r}) = \rho_{Fi}(\mathbf{r}) / \rho_X(\mathbf{r}_{max,i}), \qquad (58)$$

is convenient for computational purposes. The fuzzy membership functions $\mu_{Fi,X}(\mathbf{r})$ describe the "degree of belonging" of points $\mathbf{r}$ of the three-dimensional space to the various functional groups $F_1, F_2, \ldots F_i, \ldots F_m$ of molecule X.

3.6 A Fuzzy Set Model for Local Molecular Properties

Local molecular properties can be represented by the properties of functional groups F_i of molecule X, that in turn can be characterized by the fuzzy membership function $\mu_{Fi,X}(\mathbf{r})$ for points $\mathbf{r}$ of the space.

Variations in the local properties can be monitored by calculating the variations of the fuzzy membership function $\mu_{Fi,X}(\mathbf{r})$. If two functional groups F_i and F_j of a molecule X have the same chemical formula, then their differences must be due to their different local surroundings within the molecule X. These differences are necessarily manifested in differences of their fuzzy membership functions $\mu_{Fi,X}(\mathbf{r})$ and $\mu_{Fj,X}(\mathbf{r})$ in molecule X.

Considering a collection of functional groups of the same chemical formula but located in a series of *different* molecules, local similarity measures among these functional groups can be based on their fuzzy membership functions (for a discussion of local similarity measures see sections 5 and 6). These similarity measures provide clues for the similarities and differences in their reactivities, caused by the similarities and differences in their molecular environments.

4. AN APPLICATION OF THE *MEDLA* METHOD FOR THE DIRECT COMPUTATION OF ELECTRON DENSITIES OF FUNCTIONAL GROUPS

4.1 Building Molecular Electron Densities Using the Additive Fuzzy Electron Density Fragmentation Method.

Several versions of an additive fuzzy electron density fragmentation scheme have been proposed in [5,37]. The simplest version of Mezey's scheme, as implemented in [67], can be regarded as an "atom-group population analysis without integration". As pointed out in section 2.3, this simplest version can be thought of as an atom-group generalization of Mulliken's population analysis [68,69], where the integration step is omitted and instead of a formal charge value, a fuzzy electron density fragment is obtained.

The exact additivity of the fragment density matrices $\mathbf{P}^k$ and the fuzzy fragment densities $\rho^k(\mathbf{r})$ defined according to Mezey's scheme (eqs. (23)-(27)) motivates the terminology: additive, fuzzy electron density fragmentation method. The fuzzy electron density fragment additivity rule is exact at the given *ab initio* LCAO level.

The linearity of density expressions (16) and (24) in the corresponding density matrices ensures exact additivity for the fuzzy electron density fragments, as described by eq. (27):

$$\rho(\mathbf{r}) = \sum_{k=1}^{m} \rho^k(\mathbf{r}).$$

The sum of all m fuzzy fragment densities $\rho^k(\mathbf{r})$ is, indeed, equal to the total electron density $\rho(\mathbf{r})$ of the molecule.

4.2. The MEDLA Method

The additive fuzzy electron density fragments, discussed in section 2.5, can also be used to build electron densities for molecules different from the molecules from which the fragments originate. If the local nuclear geometry and the actual local surroundings of a fuzzy electron density fragment in a molecule X coincide with those of a fragment within a different molecule Y, then the same fragment density can also be used as a component to build the electron density of molecule Y. That is, if the local nuclear geometries as well as the local surroundings within reasonably extensive neighborhoods of the two fragments coincide, then the differences between the two fuzzy electron density fragments are expected to be negligible. The same density fragment, as a fuzzy "lego" piece, can then be used for both moleules X and Y. This "lego" principle can then be applied to build electron densities of large molecules from fuzzy fragments obtained from calculations for smaller molecules, where the local nuclear geometry as well as the extended surroundings of the fuzzy fragment in the small "parent" molecule are designed so as to mimic those in the large, "target" molecule.

The application of the additive fuzzy electron density fragments for the building of electron densities of large molecules is called the Molecular Electron Density Lego Assembler method, or MEDLA method [5,37,66,67,70-72].

Some of the fundamental properties of the fragments are important. If our goal is the construction of electron densities of large molecules from pre-calculated fragment densities obtained from smaller molecules, then both additivity and a fuzzy electron density representation appear essential. The first requirement, an additive density fragment scheme, is a natural condition, for both charge preservation and for a simple computational implementation. The second requirement, a fuzzy charge distribution for the fragments, is also natural, since in this representation fragments are analogous to complete molecules. Important computational advantages of the fuzzy representation have also been pointed out [66]. In particular, the severe local errors of combining alternative fragments with boundaries, such as those obtained by fragmentation based on zero flux density gradient surfaces (density doubling and/or density gaps, 100% error), can be avoided.

Both of the natural requirements of additivity and fuzziness are fulfilled by Mezey's fragmentation scheme that has served as the basis of the MEDLA method.

Even if each of the fuzzy electron density fragments $\rho^k(\mathbf{r})$ originate from a different molecule, a good approximation to the total electron density $\rho(\mathbf{r})$ of the target molecule can be obtained using eq. (27):

$$\rho(\mathbf{r}) = \sum_{k=1}^{m} \rho^k(\mathbf{r}).$$

The actual application of the MEDLA method for the construction of electron densities of large molecules requires an electron density fragment databank. In this MEDLA databank the electron densities are stored numerically. Each fragment type, for example, the electron density fragment of a methyl group, is stored in several versions in the databank. These versions differ in the local geometry of the nuclear family (C, H, H, H) where formal bond angles and bond distances may vary, and they also may differ in the share of the local electron density from the interactive effects of various possible surroundings. For example, the protein electron densities reported in [66,71,72] required 21 fragment types, whereas the taxol molecule [66] required several additional fragments from the MEDLA density fragment databank. Each fuzzy MEDLA fragment has been previously obtained from a 6-31G** *ab initio* calculation for a smaller molecule, artificially distorted to match the nuclear geometry and local surroundings of the fragment in the target molecule. Each fragment type is stored in several versions in the databank, for a range of several possible local nuclear arrangements; if a new arrangement is needed, a new fragment can be "custom made" by computing the electron density of the fragment from a small molecule with the required local geometry and surroundings. This new fragment density can then be added to the MEDLA databank. When used for building densities of large molecules, in each case the fragment with matching or nearly matching nuclear geometry can be selected from the databank, and high accuracy can be achieved. The electron density fragments for groups CH_3, CH_2, NH_2, and COOH are typical.

Several numerical tests and detailed comparisons of MEDLA electron densities to electron densities computed by traditional *ab initio* SCF technique using 3-21G and 6-31G** basis sets have shown [67,71] that the MEDLA results are invariably of better quality than the standard 3-21G *ab initio* results, and the MEDLA results are virtually indistinguishable from the standard *ab initio* 6-31G** basis set results obtained with the traditional Hartree-Fock method.

In particular, in addition to direct, point-by-point, numerical density comparisons carried out for several molecules [67,71], the following test results have been obtained:

(a) Detailed comparisons of electron densities of β-alanine obtained by standard *ab initio* SCF calculations using both 3-21G, 6-31G** bases, and MEDLA computations have been performed.

(b) One of the early goals of applied MEDLA studies was the construction of *ab initio* quality electron densities for proteins [71,72]. For this task, a prototype peptide system, glycyl-alanine was tested first, using standard *ab initio* SCF calculations at the 3-21G and 6-31G** level of bases, as well as MEDLA computations.

(c) For nonbonded interactions, the first test involved H-bonding in a helical tetrapeptide, using standard *ab initio* SCF 3-21G and 6-31G** basis set calculations, as well as the MEDLA method.

(d) Additional nonbonded interaction, one between a sulfur atom and a phenyl ring in a molecular fragment from the pentapeptide metenkephalin has been tested in detail, using standard *ab initio* SCF 3-21G and 6-31G** basis set calculations, as well as MEDLA.

In all tests, the MEDLA method performed consistently better than standard *ab initio* SCF 3-21G basis computations, consequently, the claim of "*ab initio* quality" appears justified.

By combining the results of two of these tests, one may conclude that the MEDLA method does not appear to show a bias concerning the joining of various density domains. This is an important concern for the analysis of functional groups.

According to the a=0.007 a.u. (atomic unit) threshold density MIDCO for the hydrogen bonded helical tetrapeptide of test c, the standard 3-21G result does not show a density domain merger at the location of the hydrogen bond, however, this merger is indicated by both the standard 6-31G** and the MEDLA results. On the other hand, according to the a=0.003 a.u. contours for the S - Phe interaction in test d, the standard *ab initio* 3-21G result indicates a bridging of the local density contours where still a gap is indicated by both the standard 6-31G** and the MEDLA computations. Apparently, the MEDLA method outperforms the standard *ab initio* 3-21G computations in an unbiased way: MEDLA indicates a feature where the standard 6-31G** result indicates it, and shows the lack of a feature where it does not appear according to the standard 6-31G** result. Using the standard 6-31G** result as a benchmark, the MEDLA method appears to generate reliable density domain patterns.

4.3 MEDLA Fragment Selection Based on Density Domains

The MEDLA method allows considerable freedom in choosing additive fuzzy electron density fragments; the choice of nuclear family is arbitrary, constrained only by practical, computational limitations. One extreme choice, the case of single nucleus fragments, has been shown to provide good results in the first application of the MEDLA method [67]. Single nucleus MEDLA fragments may have special utility in applying MEDLA as a part of an enhanced X-ray structure refinement process, where Gaussian representations of electron densities can be replaced by more accurate MEDLA estimates, and subsequently refined in an iterative process by gradually matching nuclear locations with the X-ray diffraction data [66]. However, for most applications, probably also for structure refinement purposes, fuzzy density fragments that contain several nuclei are more practical.

One, natural criterion that can be used for MEDLA fragment selection is based on density domains. Since any error of approximation of interfragment electron density interactions is expected to be the smallest if the electron density between the fragments is low, it is natural to choose fragments so that the density is low between nuclei of different fragments. This is achieved by selecting those nuclear families for defining fragments which are separated by MIDCOS from all other nuclei within the parent molecules. This is precisely the condition for density domains within the parent molecules. Consequently, a simple density domain criterion appears as an inherent molecular property for the selection of nuclear families of fuzzy electron density fragments.

4.4 Functional Groups as MEDLA Superfragments

The approach of using density domains of parent molecules as criterion for selecting nuclear families for fuzzy density fragments has additional advantages. If the density threshold for the density domains falls within the functional group range, this guaranties that the density domains are in fact those of formal functional groups. Hence, electron density distributions of functional groups are taken as MEDLA fragments. The building of large molecules from a density fragment database can then approximately follow a simplified version of the intuitive process of synthesis planning by organic chemists, building molecules from functional groups.

In some instances it is justified to use large fragments, involving a large number of nuclei. Some of these fragments can be so large that direct determination of their electron densities from a single "parent" molecule is not feasible. In such cases, the large functional group itself can be built from smaller MEDLA fragments. Whereas such a large functional group is not a complete molecule by itself, the same MEDLA method can be used for its construction that is usually applied for complete molecules. Such large functional groups, themselves bult from MEDLA fragments, are referred to as MEDLA superfragments.

The MEDLA method does not impose any size limitation on the fragments; only the feasibilty of traditional *ab initio* calculations limits the actual size of the fragments and the size of the "coordination shell" around them in the small molecule imitating the actual surroundings within the target molecule. Electron densities of satisfactory accuracy have been obtained in all the test calculations.

4.5 Linear Homotopies of Fuzzy Electron Density Fragments

If the MEDLA electron density fragment database contains a fragment. with the precise nuclear geometry required in the target molecule, then a simple transformation of electron density from the database to the actual position within the target molecule is carried out. In some intances, if there are only very small deviations between the required and already available nuclear arrangements, a linear homotopy (continuous deformation) of the nuclear arrangement and the associated fuzzy electron density fragment can be carried out [66]. For the purpose of direct transfer from the MEDLA database, as well as for small readjustments of the nuclear positions, a useful fragmentation scheme, typically involving four nuclei for the positioning of each fragment, is of special significance [66].

If four nuclei are not coplanar, then they define a tetrahedron. If two tetrahedral nuclear arrangements are congruent, then simple translation and rotation are sufficient to superimpose them exactly. Furthermore, even if the two tetrahedra are not congruent (that is, if one is distorted with respect to the other), then a simple transformation can be used to convert one into the other. Any non-coplanar, tetrahedral arrangement of four nuclei can be obtained by a 3D linear transformation (a linear homotopy) from a reference tetrahedron. Unfortunatey, general polyhedra of five or more nuclei do not have the analogous property, hence, the case of fragment nuclear positioning fully specified by four nuclei is special. Since the electron distribution is dominated by the nuclear arrangement, when a small distortion of the nuclear arrangement changes the electron density distribution, this change can be approximated by applying the same distortion to the density. If a required electron density fragment has a nuclear geometry that does not exactly match that of a corresponding fragment stored in the MEDLA databank, but these geometries are similar, then the same linear homotopy that interconverts the nuclear arrangements can also be applied to the electron density of the fragment of the database, providing a rapid, approximate method for the generation of an approximate electron density for a fragment with the desired nuclear geometry. As long as the geometry changes are small, this linear homotopy method generates good quality approximate electron densities for the required fuzzy density fragments.

Following the notations of ref. [66], the transformation is defined in terms of the coordinates of four reference nuclei A, B, C, and D of the actual fragment in the target molecule and the coordinates of the corresponding four nuclei A', B', C', and D' in the fragment stored in the MEDLA database. Without loss of generality we assume that nuclei A and A' are located at the origin of the coordinate system. We specify three edge-vectors of each of the corresponding two tetrahedra:

$$\mathbf{v}^{(1)} = A \rightarrow B, \tag{59}$$

$$\mathbf{v}^{(2)} = A \rightarrow C, \tag{60}$$

$$\mathbf{v}^{(3)} = A \rightarrow D, \tag{61}$$

$$\mathbf{w}^{(1)} = A' \rightarrow B', \tag{62}$$

$$\mathbf{w}^{(2)} = A' \rightarrow C', \tag{63}$$

$$\mathbf{w}^{(3)} = A' \rightarrow D', \tag{64}$$

respectively. These column vectors form two matrices, **V** and **W**, with elements

$$V_{ij} = v_j^{(i)} \tag{65}$$

and

$$W_{ij} = w_j^{(i)}, \tag{66}$$

respectively. We define a linear transformation $\mathbf{T}$ that converts the points of the ABCD tetrahedron into the corresponding points of the A'B'C'D' tetrahedron by the relation

$$\mathbf{TV} = \mathbf{W}, \tag{67}$$

that is, by

$$\mathbf{T} = \mathbf{W}\ \mathbf{V}^{-1}. \tag{68}$$

If the four nuclei are not coplanar then the inverse matrix $\mathbf{V}^{-1}$ exists.

Whereas in the database the reference nucleus A' is always assumed to be at the origin, the initial location of nucleus A of the actual fragment in the target molecule is arbitrary. The translation placing nucleus A to the origin is denoted by $\mathbf{S}$. The linear transformation $\mathbf{TS}$, a linear homotopy, incorporates both translation and rotation, exactly superimposing the vertices of the two tetrahedra. Furthermore, the

$$\mathbf{p} = \mathbf{TSr} \tag{69}$$

assignment of points $\mathbf{r}$ of the target molecule and points $\mathbf{p}$ of the fragment in the database is general, applicable not only for nuclear locations.

The properly positioned electron density contribution $\rho_{ABCD}(\mathbf{r})$ to each point $\mathbf{r}$ of the ABCD fragment in the target molecule from the A'B'C'D' MEDLA fragment of the database is obtained as follows:

$$\rho_{ABCD}(\mathbf{r}) = \rho_{A'B'C'D'}(\mathbf{TSr}), \tag{70}$$

where $\rho_{A'B'C'D'}(\mathbf{p})$ is the electron density of fragment A'B'C'D' at point $\mathbf{p}$, stored in the MEDLA database.

The $\mathbf{TSr}$ transformation may generate an out-of-range point $\mathbf{p}$ not stored in the database, in this case the

$$\rho_{ABCD}(\mathbf{r}) = 0 \tag{71}$$

convention is used.

Degenerate tetrahedra require special treatment. If the nuclei, A, B, C, and D are coplanar but not colinear, then nucleus D is replaced by a dummy nucleus, noncoplanar with A, B, and C; in this case the chemical fragment involves only three actual nuclei. If A, B, and C are colinear, then C is also replaced by a dummy nucleus, noncolinear with A and B; in this case the fragment involves only two actual nuclei.

The $\mathbf{TS}$ linear homotopy generates no distortion of the tetrahedron ABCD if a fragment with the exact required nuclear geometry is found in the MEDLA database, that is, if the two tetrahedra are congruent. In this case, an undistorted, identical replica of the fragment electron density of the MEDLA database is built into the target molecule. If the fragment nuclear geometries of the target molecule and the MEDLA database exactly coincide, then an identical technique can be applied for fragments involving more than four nuclei. If, however, the nuclear geometries of fragments of more than four nuclei do not coincide, then the TS transformation provides only an approximate density transformation from the MEDLA database to the target molecule; for small deviations from exact coincidence of nuclear geometries the TS linear homotopy method gives satisfactory results.

If high accuracy is required, then the option of generating a new fragment density with the exact required nuclear geometry is always available, and the new density fragment can be added to the database. All distortions due to the TS transformations can be avoided by using the new fragment from the MEDLA database.

4.6 Nonlinear Transformations of Fuzzy Electron Density Fragments

Whereas for the general case of five or more nuclei within the fragment no linear homotopy exists that can interconvert two arbitrary sets of nuclear coordinates, nevertheless, there are infinitely many nonlinear transformations which can accomplish this.

Two simple choices for nonlinear transformations, the DER method, and the WAT method, are described below. In both of these techniques, the main step of the transformation is linear, and nonlinearity is included in a rather transparent way.

In the first of these methods, the Dimension Expansion - Reduction (DER) method, the nuclear position vectors of the 3D Euclidean space are transformed into multidimensional vectors in a nonlinear manner, and the actual geometric transformation is carried out by a simple, *linear* matrix transformation in a multidimensional space, of dimensions $n > 3$, followed by a reduction of dimension to 3D. In the second method, the Weighted Affine Transformations (WAT) method, the transformation is confined to the 3D Euclidean space, and a nonlinearly-weighted average of linear, affine transformations by simplices of nuclear positions is used.

The Dimension Expansion - Reduction (DER) method.

This method performs well if the deviations between the actual and desired nuclear arrangements are small, if the nuclei configurations are general (as in the case of linear homotopy, if there are no four nuclei exactly coplanar and no three nuclei colinear, or if there are, they stay that way), and if the origin of the coordinate system is not too far from the center of mass of the fragment. For large deformations, the nuclear positions are still reproduced exactly, but the nonlinearity of the transformation may excessively deform the electron density at intermediate points. The main advantage of this method is simplicity.

Consider a fragment with n+1 nuclei, denoted by $A_0, A_1, A_2, \ldots, A_n$. For each nucleus A_i, denote the cartesian coordinates by x_i, y_i, and z_i. For simplicity, we shall use a local coordinate system attached to the fragment and assume that nucleus A_0 is located at the origin. Generate an n-dimensional vector $\mathbf{a}^{(i)}$ for each nucleus where the components are the lexicographically ordered unique products of the powers of nuclear position vector components x_i, y_i, and z_i. For example, for dimension n=12, the column vector $\mathbf{a}^{(i)}$ for nucleus A_i has the following components, defined as products of powers of cartesian components x_i, y_i, and z_i :

$$\mathbf{a}^{(i)} = (x_i,\ y_i,\ z_i,\ x_ix_i,\ x_iy_i,\ x_iz_i,\ y_iy_i,\ y_iz_i,\ z_iz_i,\ x_ix_ix_i,\ x_ix_iy_i,\ x_ix_iz_i)' \quad (72)$$

where ( )' denotes transpose. A more general version of this approach is obtained by using other, more general functions $f_4(x_i, y_i, z_i)$, $f_5(x_i, y_i, {}_i), \ldots, f_n(x_i, y_i, z_i)$ as components 4,5,...,n, as long as the components of $\mathbf{a}^{(i)}$ are linearly independent.

Based on these vectors, a simple, nonlinear transformation is defined that places all the n nuclei to the required locations and also transforms the electronic density so that it follows the nuclear distortion. The transformation for the electron density is not unique and is coordinate-dependent for most dimensions, however, the coordinate dependence is small if the distortion of the nuclear arrangement is small. The transformation based on this approach is expected to provide the best results if the dimension of vectors $\mathbf{a}^{(i)}$ is 3 (equivalent to the case of linear homotopy for four nuclei, discussed in section 4.5), or 9 (ten nuclei), or 19 (20 nuclei), and in general, for cases where all coordinate products up to a given overall degree are included as components of the vectors $\mathbf{a}^{(i)}$.

A similar n-dimensional column vector $\mathbf{b}^{(i)}$ is defined for each nucleus B_i of the fuzzy fragment stored in the MEDLA database. The components of this vector are the lexicographically ordered unique products of the nuclear position vector components x'_i, y'_i, and z'_i, where for simplicity we assume that nucleus B_0 is located at the origin of the local coordinate system. For example, for dimension n=9, the components of the column vector $\mathbf{b}^{(i)}$ of nucleus B_i are listed below:

$$\mathbf{b}^{(i)} = (x'_i,\ y'_i,\ z'_i,\ x'_ix'_i,\ x'_iy'_i,\ x'_iz'_i,\ y'_iy'_i,\ y'_iz'_i,\ z'_iz'_i)' \quad (73)$$

Such column vectors for i=1,...n form two matrices, $\mathbf{A}$ and $\mathbf{B}$, with elements

$$A_{ij} = a_j^{(i)} \quad (74)$$

and

$$B_{ij} = b_j^{(i)}, \tag{75}$$

respectively.

If the relative internal coordinates of two sets of n+1 nuclei are compared then the dimensions of matrices **A** and **B** are the same, n. We define a linear transformation **Z** that converts matrix **A** into matrix **B**:

$$\mathbf{ZA} = \mathbf{B}, \tag{76}$$

that is, by

$$\mathbf{Z} = \mathbf{B}\ \mathbf{A}^{-1}. \tag{77}$$

The inverse matrix $\mathbf{A}^{-1}$ exists for most choices of nuclear arrangements except for a set of measure zero that requires special treatment.

Considering the first three rows of matrices **A** and **B**, transformation **Z** is a nonlinear transformation between the nuclear coordinates of the actual set of n nuclei in the target molecule and the corresponding set of n nuclei in the fuzzy density fragment stored in the MEDLA databank. This nonlinear transformation converts the actual nuclear coordinates of the fragment within the target molecule to the exact nuclear coordinates of the fragment within the MEDLA database, and the inverse of this transformation also deforms the corresponding MEDLA electron density fragment in a manner compatible with all the nuclear locations.

Whereas in the four-nucleus case the linear homotopy is unique, we should emphasize that a nonlinear transformation accomplishing an exact conversion of coordinates of more than four nuclei is not unique. However, the proposed scheme, as a special choice, has the advantage of extreme simplicity, requiring a formal *linear* matrix transformation to obtain a prescribed *nonlinear* coordinate distortion of a nuclear arrangement and a compatible, continuous deformation of the entire associated electron density cloud. This is performed using a direct matrix algorithm. The above DER algorithm has been implemented as a computer program [112].

The case of 10 nuclei (dimension n=9) provides a large enough fragment size suitable for the rapid construction of approximate electron densities of large molecules.

The Weighted Affine Transformations (WAT) method.

The basic idea of the WAT method is simple. Consider a fragment with n nuclei, denoted by A_1, A_2, . . . , A_i, . . . , A_n. Note that serial index i starts with 1, and there is no requirement to include the origin as a nuclear position. For each nucleus A_i, $\mathbf{v}^{(i)}$ is the 3D position vector, and the corresponding target nuclear position is denoted by $\mathbf{t}^{(i)}$. Select a nodegenerate simplex, with vertices

$$v^{(p)}, v^{(q)}, v^{(r)}, \text{ and } v^{(s)}, \tag{78}$$

and the corresponding target simplex with corresponding vertices

$$\mathbf{t}^{(p)}, \mathbf{t}^{(q)}, \mathbf{t}^{(r)}, \text{ and } \mathbf{t}^{(s)}, \tag{79}$$

from the target positions. Denote these simplices by (p,q,r,s,V), and (p,q,r,s,T), respectively. With respect to the simplex (p,q,r,s,V), any vector v of the 3D space can be given as an affine combination

$$v = c^{(p)}v^{(p)} + c^{(q)}v^{(q)} + c^{(r)}v^{(r)} + c^{(s)}v^{(s)}, \tag{80}$$

with affine coordinates

$$c^{(p)}, c^{(q)}, c^{(r)}, \text{ and } c^{(s)}, \tag{81}$$

where

$$c^{(p)} + c^{(q)} + c^{(r)} + c^{(s)} = 1. \tag{82}$$

Since our goal is to obtain a transformation that distorts the simplex (p,q,r,s,V) into the target simplex (p,q,r,s,T), it is natural to choose a transformation defined with respect to the simplex pair (p,q,r,s,V), (p,q,r,s,T). This transformation is defined by assigning a vector **t** to each vector v of the 3D space, where **t** has the same set of affine coordinates $c^{(p)}, c^{(q)}, c^{(r)}$, and $c^{(s)}$ with respect to the target simplex (p,q,r,s,T), that is, where

$$\mathbf{t} = c^{(p)}\mathbf{t}^{(p)} + c^{(q)}\mathbf{t}^{(q)} + c^{(r)}\mathbf{t}^{(r)} + c^{(s)}\mathbf{t}^{(s)}. \tag{83}$$

For a single simplex pair (p,q,r,s,V), and (p,q,r,s,T), this transformation is a linear homotopy.

For a set of more than 4 nuclei, the simplices (p,q,r,s,V), and (p,q,r,s,T) can be chosen in several ways; if there is no degeneracy then there are

$$n_s = n*(n-1)*(n-2)*(n-3)/4! \tag{84}$$

different choices for simplex pairs, each pair defining a linear transformation. In the general case, different simplex pairs define different transformations; for the same point v, the set of $c^{(p)}, c^{(q)}, c^{(r)}$, and $c^{(s)}$ affine coordinates can be different. A transformation that is defined in terms of simplices not including a given point pair $(v^{(p)}, \mathbf{t}^{(p)})$ may transform point $v^{(p)}$ into a point different from $\mathbf{t}^{(p)}$. Consequently, a simple average of the results of all of these transformations, another linear transformation, is not suitable in general for transforming the set $\{v^{(i)}\}$ of nuclear positions to the set $\{\mathbf{t}^{(i)}\}$ of nuclear positions and to deform the electronic density accordingly.

However, by introducing a v-dependent, nonlinear, individual weighting for each of these transformations, followed by a weighted averaging where the weights are defined by the distance of point v from the various vertices of the simplices, one can construct an exact transformation that assigns each point of the set $\{\mathbf{v}^{(i)}\}$ to the corresponding point of set $\{\mathbf{t}^{(i)}\}$, while continuously deforming the electronic density that follows the nuclear deformation.

The affine representation given by eq. (80) for point v can be rearranged to give

$$\mathbf{v} = c^{(p)}\mathbf{v}^{(p)} + c^{(q)}\mathbf{v}^{(q)} + c^{(r)}\mathbf{v}^{(r)} + (1 - c^{(p)} - c^{(q)} - c^{(r)})\mathbf{v}^{(s)}, \qquad (85)$$

that is,

$$\mathbf{v} - \mathbf{v}^{(s)} = c^{(p)}(\mathbf{v}^{(p)} - \mathbf{v}^{(s)}) + c^{(q)}(\mathbf{v}^{(q)} - \mathbf{v}^{(s)}) + c^{(r)}(\mathbf{v}^{(r)} - \mathbf{v}^{(s)}). \qquad (86)$$

By collecting the column vectors $(\mathbf{v}^{(p)} - \mathbf{v}^{(s)})$, $(\mathbf{v}^{(q)} - \mathbf{v}^{(s)})$, and $(\mathbf{v}^{(r)} - \mathbf{v}^{(s)})$ into a matrix $\mathbf{S}^{(p,q,r,s,V)}$,

$$\mathbf{S}^{(p,q,r,s,V)} = \text{mat} \left| (\mathbf{v}^{(p)} - \mathbf{v}^{(s)}) \quad (\mathbf{v}^{(q)} - \mathbf{v}^{(s)}) \quad (\mathbf{v}^{(r)} - \mathbf{v}^{(s)}) \right|, \qquad (87)$$

and by collecting the first three affine coordinates $c^{(p)}, c^{(q)}, c^{(r)}$ into a column vector $\mathbf{c}^{(p,q,r,s)}$,

$$\mathbf{c}^{(p,q,r,s)} = (c^{(p)}, c^{(q)}, c^{(r)})', \qquad (88)$$

eq. (86) can be rearranged to give

$$\mathbf{c}^{(p,q,r,s)} = (\mathbf{S}^{(p,q,r,s,V)})^{-1}(\mathbf{v} - \mathbf{v}^{(s)}). \qquad (89)$$

The inverse matrix $(\mathbf{S}^{(p,q,r,s,V)})^{-1}$ of $\mathbf{S}^{(p,q,r,s,V)}$ exists for every nondegenerate simplex (p,q,r,s,V).

By definition of point **t**, the analogous expressions apply for the transformed case:

$$\mathbf{t} - \mathbf{t}^{(s)} = c^{(p)}(\mathbf{t}^{(p)} - \mathbf{t}^{(s)}) + c^{(q)}(\mathbf{t}^{(q)} - \mathbf{t}^{(s)}) + c^{(r)}(\mathbf{t}^{(r)} - \mathbf{t}^{(s)}), \qquad (90)$$

where for the corresponding matrix $\mathbf{S}^{(p,q,r,s,T)}$,

$$\mathbf{S}^{(p,q,r,s,T)} = \text{mat} \left| (\mathbf{t}^{(p)} - \mathbf{t}^{(s)}) \quad (\mathbf{t}^{(q)} - \mathbf{t}^{(s)}) \quad (\mathbf{t}^{(r)} - \mathbf{t}^{(s)}) \right|, \qquad (91)$$

eq. (90) can be written in the form

$$\mathbf{t} - \mathbf{t}^{(s)} = \mathbf{S}^{(p,q,r,s,T)}\, \mathbf{c}^{(p,q,r,s)}\ . \tag{92}$$

Combining eqs. (89) and (92) gives

$$\mathbf{t} - \mathbf{t}^{(s)} = \mathbf{S}^{(p,q,r,s,T)}\, (\mathbf{S}^{(p,q,r,s,V)})^{-1} (\mathbf{v} - \mathbf{v}^{(s)}). \tag{93}$$

By introducing the notations

$$\mathbf{D}^{(p,q,r,s)} = \mathbf{S}^{(p,q,r,s,T)}\, (\mathbf{S}^{(p,q,r,s,V)})^{-1}, \tag{94}$$

and

$$\mathbf{u}^{(p,q,r,s)} = \mathbf{t}^{(s)} - \mathbf{D}^{(p,q,r,s)}\, \mathbf{v}^{(s)}, \tag{95}$$

vector $\mathbf{t}$ can be expressed as

$$\mathbf{t} = \mathbf{D}^{(p,q,r,s)}\, \mathbf{v} + \mathbf{u}^{(p,q,r,s)}. \tag{96}$$

There are n_s simplices (p,q,r,s) and there are n_s such transformations. A set of n_s $\mathbf{v}$-dependent weight functions $w^{(p,q,r,s)}(\mathbf{v})$, one weight function associated with each simplex (p,q,r,s), is chosen so that the weighted average of the n_s transformations exactly assigns the set $\{\mathbf{v}^{(i)}\}$ of nuclear positions to the set $\{\mathbf{t}^{(i)}\}$ of nuclear positions and continuously deforms the electronic density:

$$\mathbf{t} = \sum\nolimits_{(p,q,r,s)} w^{(p,q,r,s)}(\mathbf{v})\, (\mathbf{D}^{(p,q,r,s)}\, \mathbf{v} + \mathbf{u}^{(p,q,r,s)}\ ), \tag{97}$$

where the summation is for all simplices (p,q,r,s), and where

$$\sum\nolimits_{(p,q,r,s)} w^{(p,q,r,s)}(\mathbf{v}) = 1. \tag{98}$$

In order to construct a set of appropriate, $\mathbf{v}$-dependent weight functions $w^{(p,q,r,s)}(\mathbf{v})$, first a $\mathbf{v}$-dependent function $f^{(i)}(\mathbf{v})$ is assigned to each vertex $\mathbf{v}^{(i)}$:

$$f^{(i)}(\mathbf{v}) = \prod\nolimits_{j\ (j\neq i)} d(\mathbf{v}, \mathbf{v}^{(i)}), \tag{99}$$

where $d(\mathbf{v}, \mathbf{v}^{(i)})$ is the distance between points $\mathbf{v}$ and $\mathbf{v}^{(i)}$. If point $\mathbf{v}$ coincides with any of the nuclear positions $\mathbf{v}^{(j)}$, then this continuous $\mathbf{v}$-dependent function $f^{(i)}(\mathbf{v})$ becomes zero, except if j=i, that is, if $\mathbf{v}$ coincides with the nuclear position $\mathbf{v}^{(i)}$.

For each simplex (p,q,r,s), a continuous function $g^{(p,q,r,s)}(\mathbf{v})$ is defined as

$$g^{(p,q,r,s)}(\mathbf{v}) = f^{(p)}(\mathbf{v}) + f^{(q)}(\mathbf{v}) + f^{(r)}(\mathbf{v}) + f^{(s)}(\mathbf{v}). \tag{100}$$

If point $\mathbf{v}$ coincides with any of the nuclear positions $\mathbf{v}^{(j)}$, then $g^{(p,q,r,s)}(\mathbf{v})$ becomes zero, except if j is one of the indices p, q, r, or r, that is, if $\mathbf{v}$ coincides with one of the nuclear positions $\mathbf{v}^{(p)}$, $\mathbf{v}^{(q)}$, $\mathbf{v}^{(r)}$, or $\mathbf{v}^{(s)}$.

The sum of all these $g^{(p,q,r,s)}(\mathbf{v})$ functions is taken as

$$g_{sum}(\mathbf{v}) = \sum_{(p,q,r,s)} g^{(p,q,r,s)}(\mathbf{v}). \tag{101}$$

The $\mathbf{v}$-dependent weight functions $w^{(p,q,r,s)}(\mathbf{v})$ are defined as

$$w^{(p,q,r,s)}(\mathbf{v}) = g^{(p,q,r,s)}(\mathbf{v})/g_{sum}(\mathbf{v}). \tag{102}$$

Using these $\mathbf{v}$-dependent weight functions, the weighting scheme described by eq. (97) ensures that each nuclear position $\mathbf{v}^{(j)}$ is transformed to its counterpart nuclear position $\mathbf{t}^{(j)}$, while the entire electron density is deformed continuously. This method of weighted affine transformations has no origin or coordinate dependence.

The above WAT algorithm has been implemented as a computer program [113].

The weighted affine transformations scheme allows one to introduce small geometry changes relative to the MEDLA database fragments if the database contains no fragment with the exact nuclear geometry K, but h fuzzy electron density fragments $\rho_1(\mathbf{r}), \rho_2(\mathbf{r}), \ldots, \rho_h(\mathbf{r})$, of nearly matching nuclear geometries, $K_1, K_2, \ldots, K_h$, respectively, are available. If all nuclear geometries $K_1, K_2, \ldots, K_h$, of the h different database versions $\rho_1(\mathbf{r}), \rho_2(\mathbf{r}), \ldots, \rho_h(\mathbf{r})$ of fragment electron densities are transformed by the weighted affine transformations method to the same, required nuclear geometry K, than by a weighted averaging of the transformed electron densities $\rho_1'(\mathbf{r}), \rho_2'(\mathbf{r}), \ldots, \rho_h'(\mathbf{r})$, a higher accuracy is obtained. One choice for the weighting factors for the individual transformed MEDLA fragment densities is taken as

$$W_i=U(K, K_1, K_2, \ldots, K_h) / d(K, K_i), \tag{103}$$

where $d(K, K_i)$ is the configuration space distance [40] between the individual database configuration K_i and the actual target configuration K, whereas

$$U(K, K_1, K_2, \ldots, K_h) = 1/(\Sigma_i\ 1/ d(K, K_i)). \tag{104}$$

Using these W_i weighting factors, a MEDLA database electronic density $\rho_i(\mathbf{r})$ with a higher degree of similarity between the given database nuclear configuration K_i and the actual nuclear configuration K has a larger contribution to the resulting weighted MEDLA electron density $\rho_{WM}(\mathbf{r})$,

$$\rho_{WM}(\mathbf{r}) = W_1\, \rho_1'(\mathbf{r}) + W_2\, \rho_2'(\mathbf{r}) + \ \ldots\ + W_h\, \rho_h'(\mathbf{r}). \tag{105}$$

This technique allows an efficient use of an existing MEDLA database, and when combined with shape analysis, it can be used to decide whether it is indeed needed to add a new MEDLA fragment version to the database.

5. LOCAL SHAPE ANALYSIS OF FUNCTIONAL GROUPS

5.1 Local Molecular Fragments in a Global Molecular Environment

The local shapes and shape changes of functional groups are influenced by the global molecular environment. This influence is the primary cause of differences between the reactivities of a given functional group in different molecules. It is of some importance to study these shape differences and to find out the extent of the influence the rest of a molecule can exert on a given functional group. The density domain approach provides a natural criterion for limited shape changes which preserve the identity of a functional group, and it also serves as a tool to explore the range of shape variations possible for a local molecular fragment without becoming a chemically different entity.

5.2 Local Shape Changes Induced by Molecular Environment

The reactivities of functional groups are highly dependent on the molecular surroundings, and the effects of the global molecular environment on the local shape variations can be significant.

The Walker pseudo-density scheme for diagnosing shape changes, discussed in section 2.3, is a sensitive method for detecting the effects of molecular environment. Whereas the magnitudes of formal shape changes obtained by the pseudo-density scheme are exaggerated, the very sensitivity of the method makes it a suitable diagnostic tool. Whereas the numerical changes in the computed pseudo-densities are larger than the actual density changes, the relative magnitudes of formal shape changes of the pseudo-densities are indicative of the relative magnitudes of the actual shape variations in the electron densities of functional groups.

In the following sections the actual electron density variations will be used for local shape analysis.

5.3 Local Shapes of Non-Interacting Functional Groups

The fundamental principle we shall follow in the local shape analysis of functional groups and local molecular moieties is a strict analogy with the shape analysis of complete molecules. Accordingly, instead of molecular isodensity contour (MIDCO) surfaces, the main tool of analysis will be the fragment isodensity contour (FIDCO) surfaces. Some of the ideas and concepts described in this section are illustrated in Figure 1.

We shall use the following notations: A is the actual fragment for shape analysis, B is rest of the molecule, possibly composed from several fragments, B_1, $B_2, \ldots, B_{m-1}$.

Two choices for the representation of a local molecular moiety will be discussed. For the first choice, describing the local shapes of non-interacting functional groups within a molecule, we define a FIDCO for a fragment A in a molecule AB as follows:

$$G_{A\backslash B}(a) = \{\ \mathbf{r} : \rho_A(\mathbf{r}) = a,\ \ \rho_A(\mathbf{r}) \geq \rho_{Bk}(\mathbf{r}),\ \ k=1,...m-1\ \}. \qquad (106)$$

Two, equivalent definitions are given by

$$G_{A\backslash B}(a) = \ G_A(a) \cap \{\ \mathbf{r} :\ \rho_A(\mathbf{r}) \geq \rho_{Bk}(\mathbf{r}),\ \ k=1,...m-1\ \}, \qquad (107)$$

and

$$G_{A\backslash B}(a) = \ G_A(a) \setminus \{\ \mathbf{r} : \exists\ k \in \{1,...m-1\}:\ \rho_A(\mathbf{r}) < \rho_{Bk}(\mathbf{r})\ \}. \qquad (108)$$

Pictorially, in order to obtain FIDCO $G_{A\backslash B}(a)$ of fragment A in molecule AB, it is sufficient to locate those points where the electron density contribution of fragment A is dominant in molecule AB.

The actual shape analysis can be carried out on the "isolated" fragment density contour G(a), where one additional domain type is introduced. These domains represent the connection of fragment A to the rest of the molecule within the actual AB system:

$$D_{-1}(G_{A\backslash B}(a)) = \{\mathbf{r} :\ \mathbf{r} \in G_A(a),\ \exists\ k \in \{1,...m-1\}:\ \rho_A(\mathbf{r}) < \rho_{Bk}(\mathbf{r})\}, \qquad (109)$$

Although the above notation $D_{-1}(G_{A\backslash B}(a))$ refers to the FIDCO $G_{A\backslash B}(a)$, the actual domain $D_{-1}(G_{A\backslash B}(a))$ exists only on the intact $G_A(a)$ contour, and $D_{-1}(G_{A\backslash B}(a))$ appears as a "cover" over the hole(s) of the FIDCO $G_{A\backslash B}(a)$ in molecule AB. The boundary $\Delta D_{-1}(G_{A\backslash B}(a))$, however, is the actual boundary of the FIDCO surface $G_{A\backslash B}(a)$:

$$\Delta D_{-1}(G_{A\backslash B}(a)) = \{\mathbf{r} : \mathbf{r} \in G_{A\backslash B}(a), \exists\, k' \in \{1,...m-1\}: \rho_A(\mathbf{r}) = \rho_{B_{k'}}(\mathbf{r}), \rho_{B_{k'}}(\mathbf{r}) \geq \rho_{B_k}(\mathbf{r}),\ k=1,...m-1\ \}. \tag{110}$$

An alternative, simpler choice for the representation of fragment A in molecule AB is obtained using the composite B of all fragments $B_1, B_2, \ldots, B_{m-1}$:

$$G_{A\backslash\Sigma B}(a) = \{\ \mathbf{r} : \rho_A(\mathbf{r}) = a,\quad \rho_A(\mathbf{r}) \geq \rho_B(\mathbf{r})\ \}, \tag{111}$$

where $\rho_B(\mathbf{r})$ is defined as

$$\rho_B(\mathbf{r}) = \rho_{B_1}(\mathbf{r}) + \rho_{B_2}(\mathbf{r}) + \ldots + \rho_{B_{m-1}}(\mathbf{r}). \tag{112}$$

The local domains of connections between fragment A and the rest B of the molecule AB are defined as

$$D_{-1}(G_{A\backslash\Sigma B}(a)) = \{\mathbf{r} :\ \mathbf{r} \in G_A(a),\quad \rho_A(\mathbf{r}) \leq \rho_B(\mathbf{r})\}, \tag{113}$$

whereas the boundaries of these domains are given as

$$\Delta D_{-1}(G_{A\backslash\Sigma B}(a)) = \{\mathbf{r} :\ \mathbf{r} \in G_{A\backslash\Sigma B}(a),\quad \rho_A(\mathbf{r}) = \rho_B(\mathbf{r})\}. \tag{114}$$

5.4 Local Shapes of Interacting Functional Groups

The local shape analysis can no longer be carried out on an "isolated" FIDCO $G_A(a)$ if the interactions of various molecular fragments in a molecule AB are fully taken into account, beyond simply using these interactions for a truncation of the isolated FIDCO $G_A(a)$. If a detailed description of the interactions is required, then a new contour calculation is needed for the interactive FIDCO $G_{A(B)}(a)$ in molecule AB, where $G_A(a)$ is defined as

$$G_{A(B)}(a) = \{\mathbf{r} : \rho_A(\mathbf{r}) + \rho_B(\mathbf{r}) = a,\quad \rho_A(\mathbf{r}) \geq \rho_B(\mathbf{r})\}. \tag{115}$$

There is no surface defined where the formal "cover" domains of the hole(s) of FIDCO $G_{A(B)}(a)$ would lie, hence no such domains $D_{-1}(G_{A(B)}(a))$ are defined. However, for uniformity, the notation $\Delta D_{-1}(G_{A(B)}(a))$ will be used for the boundaries of the holes on $G_{A(B)}(a)$. These boundaries are defined as

$$\Delta D_{-1}(G_{A(B)}(a)) = \{\mathbf{r} :\ \mathbf{r} \in G_{A(B)}(a),\quad \rho_A(\mathbf{r}) = \rho_B(\mathbf{r})\}. \tag{116}$$

For this second, more accurate choice, the definition of the interactive FIDCO

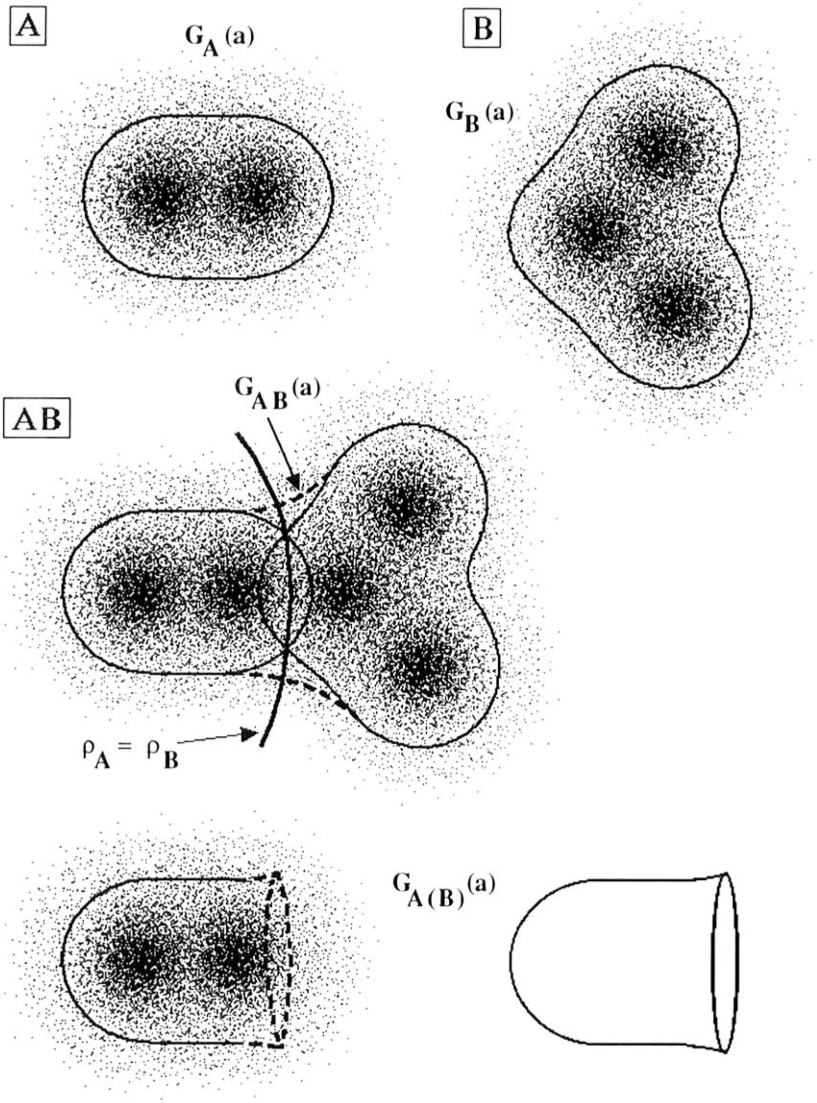

Figure 1. Illustration of the local shape description of non-interacting and interacting functional groups. See text for definitions of symbols.

for a fragment A in a molecule AB describes the local shapes of interacting functional groups within a molecule. This choice for defining interactive FIDCO surfaces $G_{A(B)}(a)$ for local shape analysis requires an additional contour calculation and is computationally more expensive than that of the non-interactive FIDCO model $G_{A\backslash B}(a)$ discussed in secton 5.3.

For both types of FIDCO surfaces, the usual Shape Group method [2] of electron density shape analysis is applicable. The additional formal domain boundaries $\Delta D_{-1}(G_{A\backslash B}(a))$ and $\Delta D_{-1}(G_{A(B)}(a))$ introduce one additional index -1, which can be treated the same way as relative curvature indices. The one-dimensional homology groups obtained by truncations using all possible index combinations are the shape groups of FIDCO surfaces. The (a,b)-parameter maps and shape codes are generated the same way as for complete molecules [2].

6. SHAPE SIMILARITY AND SHAPE COMPLEMENTARITY MEASURES OF FUNCTIONAL GROUPS IN DIFFERENT MOLECULAR ENVIRONMENTS

6.1 Shape Similarity Measures of Functional Groups Based on Local Shapes

The local shape representations by FIDCO sequences is the key to the application of the shape group method for generating shape similarity measures for functional groups and other molecular moieties. The shape of each density domain and each FIDCO $G_{A\backslash B}(a)$ or $G_{A(B)}(a)$ can be described in detail by considering local curvature variations and subdividing the surfaces according to various curvature domains. An algebraic-topological characterization of these shapes can be given, using the Shape Group Method (SGM). The shape group methods [41-44], their physical basis, and their mathematical derivation have been extensively reviewed recently [2] and will not be described in detail here.

With reference to the original derivations [41-44] given for complete molecules, it is sufficient to note that the shape groups of functional groups are the one-dimensional homology groups of FIDCO surfaces truncated according to local curvature criteria and by the boundaries $\Delta D_{-1}(G_{A\backslash B}(a))$ and $\Delta D_{-1}(G_{A(B)}(a))$ of non-interacting and interacting FIDCO surfaces, where these boundaries are defined by the interpenetration condition between the functional group and the rest of the molecule.

Since for the whole range of FIDCO surfaces of a given functional group there are only a finite number of topologically different truncated surfaces, consequently, there are only a finite number of shape groups characterizing the shape of a functional group within a given molecule. Each shape group can be characterized by its Betti number, and the collection of the finite number of different Betti numbers as well as their distribution within an (a,b) parameter plane generate a numerical shape code for the functional group. In the (a,b) parameter plane, a is the electron density threshold for the FIDCO $G_{A\backslash B}(a)$ or $G_{A(B)}(a)$, whereas parameter b is the reference curvature parameter against which the local curvatures of the FIDCOs are compared.

Since the shape group distribution along an (a,b) parameter map of a molecule can be characterized by tthe distribution of the Betti numbers of these shape groups, the map can be represented by a sequence of numbers ordered into a matrix or a vector [2]. The surface truncation involved in the generation of the shape groups may lead to several disjoint surface pieces, and in some applications of the shape group method these surface pieces are characterized separately. For each given (a,b) pair of parameter values, a size-ordering of the surface pieces [33] implies an ordering of their one-dimensional Betti numbers (informally called their "first" Betti numbers) into a sequence:

$$B(1),B(2),\ldots,B(k),\ldots,B(m). \tag{117}$$

An electron density contour surface piece with a larger Betti number usually (but not necessarily) has the larger surface area. If this is the case, then the sequence of Betti numbers in the above ordering is the same as the decreasing sequence of the Betti numbers.

Since the number of surface pieces obtained after truncation according to a set of specified curvature conditions is dependent on the values of parameters a and b, for different (a,b) pairs within an (a,b) parameter map, both the number m of Betti numbers, and their actual values may be different. This non-uniformity of the information representation belonging to different locations within the (a,b)-map can be circumvented by applying a coding-decoding method that relies on the prime factorization of integers. By a reversible transformation it is possible to convert the family of all Betti numbers for each (a,b) location of the parameter map into a single integer number. Such an integer number, a shape ID number, can be assigned to each molecule as well as to each functional group. The shape ID number can be used as a numerical shape code. Since this coding method is uniform for all (a,b) pairs, a uniform representation of shape codes for all molecules and all functional groups is possible [2,33].

Based on this "ID number" approach, the shape similarity of molecules can be evaluated by numerical comparisons of shape codes [2]. The same technique of similarity evaluation, originally developed for complete molecules, can also be applied, in identical form, to the shape codes of functional groups as discussed in earlier parts of this report.

6.2 Local Shape Complementarity Measures for Functional Groups

In most interactions between two reactants, local shape complementarity of functional groups is of importance. A local shape complementarity of molecular electron densities represented by FIDCOs implies complementary curvatures for complementary values of the charge density threshold parameters a. For various curvature domains of a FIDCO, we shall use the notations originally proposed for complete molecues [2]. For example, the symbol $D_{2(b),i}(a, F_1)$ stands for the i-th locally convex domain of a FIDCO G(a) of functional group F_1, where local convexity, denoted by subscript 2(b), is interpreted relative to a reference curvature b. For locally saddle type and locally concave domains relative to curvature b, the analogous subscripts 1(b) and 0(b) are used, respectively.

In general, a locally convex domain $D_{2(b),i}(a, F_1)$ of a functional group F_1, relative to a reference curvature b, shows local shape complementarity with a locally concave domain $D_{0(-b),j}(a', F_2)$ of a complementing functional group F_2, relative to a reference curvature of -b. The threshold values a and a' are also likely to complement each other: the shape complementarity between the higher electron density contours of one functional group and the lower electron density contours of the other functional group is relevant.

For a shape complementarity analysis of functional groups we shall follow the shape complementarity approach described for molecules in ref. [2].

The electron densities of interacting functional groups penetrate each other to some extent. In stronger interactions one expects greater interpenetration than in weaker interactions. For a given pair of interacting functional groups, a common electronic density value a_0 can be chosen for the two isolated functional groups that corresponds to the threshold density of two formal "contact" FIDCO's having tangential contact in the interacting systems. For the purposes of shape complementarity analysis, we approximate the electronic density of the interacting functional groups F_1 and F_2 by the superposition of the electronic densities of the two isolated functional groups. In this case, a pair of FIDCO surfaces $G(a, F_1)$ and $G(a,F_2)$ of the same threshold value a must have one of the three possible relative arrangements:

1. $G(a, F_1)$ and $G(a,F_2)$ have no common points, or
2. $G(a, F_1)$ and $G(a,F_2)$ have a finite number of common points (usually, one common point), or
3. $G(a, F_1)$ and $G(a,F_2)$ have a continuum of common points.

We define the contact density a_0 for a given mutual arrangement of two functional groups F_1 and F_2 as the threshold value that corresponds to the unique electron density threshold of the FIDCOs of case 2.

If a contact density threshold a_0 can be chosen for a given interaction between two functional groups, then the local shape complementarity between $G(a_0, F_1)$ and $G(a_0, F_2)$ is clearly of importance. However, complementarity should also manifest itself within a whole range of density thresholds. One may consider the local shape complementarity of FIDCOs $G(a_0\text{-}a', F_1)$ and $G(a_0\text{+}a', F_2)$ in a density interval containing the contact density threshold a_0,

$$[a_0 - \Delta a,\ a_0 + \Delta a]. \tag{118}$$

Note that the complementarity of the local shapes of those FIDCOs are important where the thresholds deviate from the contact density value a_0 in the opposite sense.

Shape complementarity of functional groups involves matches between locally concave and locally convex domains, and also matches between properly aligned saddle-type domains, that is, between curvature domain pairs of the following combinations:

$$D_{0(b),i}(a_0 - a', F_1)\ ,\ D_{2(-b),i}(a_0+a', F_2); \tag{119}$$

$$D_{1(b),i}(a_0 - a', F_1)\ ,\ D_{1(-b),i}(a_0+a', F_2); \tag{120}$$

and

$$D_{2(b),i}(a_0 - a', F_1)\ ,\ D_{1(-b),i}(a_0+a', F_2). \tag{121}$$

In the molecular case, similar relations formed the basis of a simple model [2] using an (a,b) parameter map approach for complementarity evaluation in a manner analogous to similarity evaluation. The same approach is applicable to fuzzy density fragments of functional groups. The local shape complementarity of matching curvature domains, for example, $D_{2(b),i}(a, F_1)$ of functional group F_1, and $D_{0(-b),j}(a', F_2)$ of functional group F_2, can be tested by computing complementary shape groups. This is achieved by taking complementary truncations for functional groups F_1 and F_2, for example, by truncating the curvature domains $D_{2(b),i}(a, F_1)$ for functional group F_1, and the curvature domains $D_{0(-b),j}(a', F_2)$ for functional group F_2. Using the homology group notation of ref. [2], such a truncation leads to an (a,b) map for the $H^p{}_\mu(a,b)$ shape groups of functional group F_1, and to an (a,b) map for the complementary $H^p{}_{2-\mu}(a,b)$ shape groups of functional group F_2. Typical example is the pair of the (a,b) map of the $H^1{}_2(a,b)$ shape groups of functional group F_1, and the (a,b) map of the $H^1{}_0(a,b)$ shape groups of functional group F_2.

Complementarity of the curvature types for truncation is not sufficient for a direct comparison of the two (a,b)-maps, since one must also take into account the required complementarity of density thresholds a and reference curvatures b. This can be accomplished by a simple transformation of one of the two (a,b)-maps. A central inversion of the (a,b)-parameter map of functional group F_2 with respect to the point $(a_0,0)$ allows one to use a similarity evaluation method for evaluating complementarity. By comparing the centrally inverted (a,b)-map of functional group F_2 to the original (a,b)-map of functional group F_1, and by determining the ratio of matches between the corresponding families of Betti numbers, a numerical complementarity measure is obtained. The central inversion of one of the (a,b)-parameter maps implies a proper match between curvature domain types, density thresholds, and curvature parameters, in agreement with the pairing scheme (119) - (121) specified for FIDCOs $G(a_0\text{-}a', F_1)$ and $G(a_0\text{+}a', F_2)$. The method ensures that the locally convex domains of FIDCO $G(a_0\text{-}a', F_1)$ relative to the reference curvature b are tested for shape complementarity against the locally concave domains of FIDCO $G(a_0\text{+}a', F_2)$ relative to a reference curvature - b, and a similar complementary match is tested when any of the roles within the pairs of FIDCOs, density thresholds, curvature types, or curvature parameters are reversed.

This generalization of the Centrally Inverted Map Method (CIMM) of molecular shape complementarity analysis [2] to FIDCOs of functional groups replaces the problem of complementarity evaluation with a conceptually and computationally simpler similarity evaluation.

7. ENERGY RELATIONS FOR FUNCTIONAL GROUPS AND THEIR INTERACTIONS

7.1 The Fragment Energy Concept and Fragment Interaction Energies

Fundamental relations of density functional theory, the Hohenberg-Kohn theorem [114], the advances made by Kohn and Sham [115], Parr [87,116,117], Levy [118,119], Ludena [88,120], Kryachko [88], March [89], Becke [121-124], Perdew [125], Ziegler [126], Salahub [127,128], Andzelm [129,130], Politzer [131,132], Tachibana [133,134], van Leuven, and Baerends [135], Koch [136], and others have provided both a detailed theoretical foundation and a variety of approximate methods for the computation of relations between electronic density and energy. Whereas most efforts have been focused on complete molecules, the fuzzy fragmentation scheme for the electron density of functional groups imply that the same methodology of density - energy relations is applicable for fuzzy density fragments and for density representations of functional groups.

Density functional theory provides the means for defining fragment energies, based on the following, simple principle: the fuzzy electron density fragment of a functional group, together with the associated set of nuclei, is treated as a complete molecule. The molecular approximations to the E(ρ) electron density - energy functional are equally applicable to functional groups represented by the fuzzy electron density fragments.

Based on this model, the molecular total energy is decomposed into two parts: fragment energies, obtained by a density functional approximation, and interfragment energies describing the interaction energies among the fragments. One area of current research is the development of approximate methods for the computation of the interfragment contributions to the molecular energy, where the consistency of the model is tested by considering alternative fragmentations within a 3D tiling scheme described earlier [66]. A method that is capable of utilizing the fuzzy fragmentation scheme for energy computation is expected to provide a new approach to the description of intermolecular interactions, nonbonding interactions and hydrogen bonds [137], important in the prediction of reactivities.

7.2 Approximate Conformation Optimization from Fragment Models

Whereas density functional theory guaranties that for the ground electronic state of molecules the electron density determines the energy, the actual construction of such energy functions from first principles is a problem of considerable complexity. The electron densities computed by the MEDLA method suggest various approximations to the molecular energy of large systems.

A simple, approximate conformational energy function can be based on a reward-and-penalty function associated with the mergers of fuzzy electron density fragments forming the molecule. According to one alternative, this reward-and-penalty function can be calibrated using conventional *ab initio* computations. A formal density threshold potential function of shape analogous to a Morse potential can be constructed where the distance variable is replaced by the value of the density threshold a where the merger of fuzzy fragments occurs. A minimization of the sum of the pairwise interfragment potentials can be used for geometry optimization where the conformational motions introduce only minor deformations within the fragments selected as functional groups, and the moton is primarily a rearrangement of the mutual positions of the functional groups.

The formal space filling characteristics of molecules are influenced by a merger of electronic density clouds between parts of the molecule not linked directly by formal bonds. The computational experience with protein MEDLA electron densities indicates that these "nonbonded" mergers of MIDCOs tend to occur simultaneously at about the same density threshold a_m, at many locations within the molecule. This trend, pointed out in [66,71], is likely to help the search for favored conformations of proteins, advantageous mutual side chain arrangements and folding patterns.

Based on this observation, the Self-Avoiding MIDCO approach has been proposed as a simple, approximate method for macromolecular conformation analysis [66]. If the threshold value a_m corresponds to the onset of "nonbonded" mergers of MIDCOs $G(K, a_m)$, then the MEDLA method can be applied to generate a large number of MIDCOs $G(K, a_m)$ for a family R of nuclear configurations K. The self-avoiding MIDCO method is used to select from this family R of nuclear configurations those which show favorable "nonbonding" interactions. According to early experience, the optimum value of a_m falls within the range [0.003au, 0.005au] of density thresholds, and a small tolerance limit, $\Delta a \sim 0.001$au, is used for the actual mergers.

For a suitable density tolerance Δa, a simple contact principle [66] can be used for accepting and rejecting nuclear configurations from the family R:

> A given configuartion K is accepted if all nonbonded mergers which appear for the MEDLA MIDCO $G(K, a_m - \Delta a)$ are not yet merged in the MEDLA MIDCO $G(K, a_m + \Delta a)$.

If the above criterion is tested for the family R of nuclear configurations K, a nuclear arrangement K_{mm} with the maximum number of proper mergers can be identified. This configuration K_{mm} is expected to provide a good approximation to a preferred nuclear arrangement of the molecule.

It has been pointed out [66] that the Self-Avoiding MIDCO approach is not a MIDCO version of a hard surface contact model. Parts of a macromolecular electronic density folding back upon itself readjust, that changes the actual local shape of the MIDCO, a feature not present in simple, hard-sphere contact models, such as VdW surfaces. Since the electron densities of molecular parts placed side by side increase due to the mutual interpenetration ("overlap"), the corresponding MIDCO $G(K,a)$ of the given threshold a shows a significant "swelling" near the site of merger. This feature is in fact a manifestation of nonbonded interactions. The Self-Avoiding MIDCO method incorporates some aspects of these interactions into a simple conformation analysis approach, without actually calculating interaction energies [66].

8. SUMMARY

A quantum chemical approach is proposed for the representation of functional groups in chemistry. The approach is based on a simple density domain condition and on the additive, fuzzy electron density fragmenation method that also serves for the rapid caculation of *ab initio* quality electron densities of large molecules. Several aspects of the approach are described, including methods for similarity and complementarity analysis of functional groups.

9. ACKNOWLEDGEMENTS

The orginal research work leading to the methods and results described in this report was supported by both strategic and operating research grants from the Natural Sciences and Engineering Research Council of Canada.

10. REFERENCES

1. Mezey, P.G., *J. Chem. Inf. Comp. Sci.,* **32**, 650 (1992).
2. Mezey, P.G., *Shape in Chemistry: An Introduction to Molecular Shape and Topology* (VCH Publishers, New York, 1993).
3. Mezey, P.G., *Canad. J. Chem.,* **72**, 928 (1994). (Special issue dedicated to Prof. J. C. Polanyi).
4. Mezey, P.G., "Molecular Surfaces". In Lipkowitz, K.B., and Boyd, D.B., eds., *Reviews in Computational Chemistry* (VCH Publishers, New York, 1990).
5. Mezey, P.G., "Density Domain Bonding Topology and Molecular Similarity Measures". In Sen, K., ed., *Topics in Current Chemistry,* Vol. **173**, *Molecular Similarity* (Springer-Verlag, Heidelberg, 1995).
6. Dubois, J.-E., and Mezey, P.G., *Int. J. Quantum Chem.,* **43**, 647 (1992).
7. Carbó, R., Leyda, L., and Arnau, M., *Int. J. Quantum Chem.,* **17**, 1185 (1980).
8. Hodgkin, E.E., and Richards, W.G., *J. Chem. Soc. Chem. Commun.* **1986**, 1342.
9. Carbó, R., and Domingo, Ll., *Int. J. Quantum Chem.,* **32**, 517 (1987).
10. Hodgkin, E.E., and Richards, W.G., *Int. J. Quantum Chem.,* **14**, 105 (1987).
11. Carbó, R., and Calabuig, B., *Comput. Phys. Commun.,* **55**, 117 (1989).
12. Burt, C., Richards, W.G., and Huxley, P., *J. Comput. Chem.,* **11**, 1139 (1990).
13. Johnson, M.A. and Maggiora, G.M., eds., *Concepts and Applications of Molecular Similarity* (Wiley, New York, 1990).
14. Carbó, R., and Calabuig, B., *Int. J. Quantum Chem.,* **42**, 1681 (1992).
15. Carbó, R., and Calabuig, B., *Int. J. Quantum Chem.,* **42**, 1695 (1992).
16. Good, A., and Richards, W.G., *J. Chem. Inf. Sci.,* **33**, 112 (1992).
17. Johnson, M.A., *J. Math. Chem.,* **3**, 117 (1989).
18. Leicester, S.E., Finney, J.L., and Bywater, R.P., *J. Mol. Graph.,* **6**, 104 (1988).
19. Mezey, P.G., *J. Math. Chem.,* **2**, 299 (1988).

20. Arteca, G.A., Jammal, V.B., and Mezey, P.G., *J. Comput. Chem.,* **9**, 608 (1988).
21. Arteca, G.A., Jammal, V.B., Mezey, P.G., Yadav, J.S., Hermsmeier, M.A., and Gund, T.M., *J. Molec. Graphics,* **6**, 45 (1988).
22. Arteca, G.A. and Mezey, P.G., *J. Phys. Chem.,* **93**, 4746 (1989).
23. Arteca, G.A. and Mezey, P.G., *IEEE Eng. in Med. & Bio. Soc. llth Annual Int. Conf.,* **11**, 1907 (1989).
24. Mezey, P.G., "Three-Dimensional Topological Aspects of Molecular Similarity". In Johnson, M.A. and Maggiora, G.M., eds., *Concepts and Applications of Molecular Similarity* (Wiley, New York, 1990).
25. Arteca, G.A. and Mezey, P.G., *Int. J. Quantum Chem. Symp.,* **24**, 1 (1990).
26. Mezey, P.G., *J. Math. Chem.,* **7**, 39 (1991).
27. Mezey, P.G., "New Symmetry Theorems and Similarity Rules for Transition Structures". In Formosinho, S.J., Csizmadia, I.G., and Arnaut, L.G., eds., *Theoretical and Computational Models for Organic Chemistry* (Kluwer Academic Publishers, Dordrecht, 1991).
28. Harary, F. and Mezey, P.G., *Theor. Chim. Acta,* **79**, 379 (1991).
29. Luo, X. and Mezey, P.G., *Int. J. Quantum Chem.,* **41**, 557 (1992).
30. Mezey, P.G., *J. Math. Chem.,* **11**, 27 (1992).
31. Luo, X., Arteca, G.A., and Mezey, P.G., *Int. J. Quantum Chem.,* **42**, 459 (1992).
32. Mezey, P.G., *J. Math. Chem.,* **12**, 365 (1993).
33. Mezey, P.G., *J. Chem. Inf. Comp. Sci.,* **34**, 244 (1994).
34. Mezey, P.G., *Int. J. Quantum Chem.,* **51**, 255 (1994).
35. Carbó, R., Calabuig, B., Vera, L., and Besalu, E., "Molecular Quantum Similarity: Theoretical Framework, Ordering Principles, and Visualization Techniques". In Löwdin, P.-O., Sabin, J.R., and Zerner, M.C., eds., *Advances in Quantum Chemistry ,Vol. 25* (Academic Press, New York, 1994).
36. Mezey, P.G., "Molecular Similarity Measures for Assessing Reactivity". In Carbó, R., ed., *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches* (Kluwer Academic Publ., Dordrecht, The Netherlands, 1995).
37. Mezey, P.G.,"Methods of Molecular Shape-Similarity Analysis and Topological Shape Design". In Dean, P.M., ed., *Molecular Similarity in Drug Design* (Chapman & Hall - Blackie Publishers, Glasgow, U.K., 1995).
38. Walker, P.D. and Mezey, P.G., *J. Comput. Chem.,* in press.
39. Walker, P.D., Maggiora, G.M., Johnson, M.A., Petke, J.D., and Mezey, P.G., *J. Chem. Inf. Comp. Sci.,* **35**, 568 (1995).
40. Mezey, P.G., *Potential Energy Hypersurfaces* (Elsevier, Amsterdam, 1987).
41. Mezey, P.G., *Int. J. Quant. Chem. Quant. Biol. Symp.,* **12**, 113 (1986).
42. Mezey, P.G., *J. Comput. Chem.,* **8**, 462 (1987).
43. Mezey, P.G., *Int. J. Quantum Chem., Quant. Biol. Symp.,* **14**, 127 (1987).
44. Mezey, P.G., *J. Math. Chem.,* **2**, 325 (1988).

45. Collard, K., and Hall, G.G., *Int. J. Quantum Chem.*, **12**, 623 (1977).
46. Y. Tal, Y., Bader, R.F.W., Nguyen-Dang, T.T., Ojha, M., and Anderson, S.G., *J. Chem. Phys.*, **74**, 5162 (1981).
47. Bader, R.F.W., Nguyen-Dang, T.T., *Adv. Quantum Chem.*, **14**, 63 (1981).
48. Bader, R.F.W., Slee, T.S., Cremer, D., and Kraka, E., *J. Amer. Chem. Soc.*, **105**, 5061 (1983).
49. Bader, R.F.W., *Acc. Chem. Res.*, **9**, 18 (1985).
50. Bader, R.F.W., *Atoms in Molecules: A Quantum Theory* (Clarendon Press, Oxford, 1990).
51. Chang, C., and Bader, R.F.W., *J. Phys. Chem.*, **96**, 1654 (1992).
52. Mezey, P.G., *J. Chem. Phys.*, **78**, 6182 (1983).
53. Cremer, D., and Kraka, E., *Croat. Chem. Acta*, **57**, 1265 (1984).
54. Cioslowski, J., *J. Phys. Chem.*, **94**, 5496 (1990).
55. Cioslowski, J., Mixon, S.T., and Edwards, W.D., *J. Amer. Chem. Soc.*, **113**, 1083 (1991).
56. Cioslowski, J., and Fleischmann, E.D., *J. Chem. Phys.*, **94**, 3730 (1991).
57. Cioslowski, J., O'Connor, P.B., and Fleischmann, E.D., *J. Amer. Chem. Soc.*, **113**, 1086 (1991).
58. Cioslowski, J., Mixon, S.T., and Fleischmann, E.D., *J. Amer. Chem. Soc.*, **113**, 4751 (1991).
59. Cioslowski, J., Mixon, S.T., *Can. J. Chem.*, **70**, 443 (1992).
60. Christoffersen, R.E. and Maggiora, G.M., *Chem. Phys. Lett.*, **3**, 419 (1969).
61. Christoffersen, R.E., Shipman, L.L., and Maggiora, G.M., *Internat. J. Quantum Chem.*, **5**, 143 (1971).
62. Christoffersen, R.E., Spangler, D., Hall, G.G., and Maggiora, G.M., *J. Am. Chem. Soc.*, **95**, 8526 (1973).
63. Michl, J., Kaszynski, K., Friedli, A.C., McMurdie, N.D., and Kim, T., *NATO ASI Ser., Ser. C.*, **273**, 469 (1989).
64. Mathias, J.P. and Stoddart, J.F., *Chem. Soc. Rev.*, 215 (**1992**).
65. Arteca, G.A., Grant, N.A., Mezey, P.G., *J. Comput. Chem.*, **12**, 1198 (1991).
66. Walker, P.D., and Mezey, P.G., *J. Math. Chem.*, **17**, 203 (1995).
67. Walker, P.D., and Mezey, P.G., *J. Am. Chem. Soc.*, **115**, 12423 (1993).
68. Mulliken, R.S., *J. Chem. Phys.*, **23**, 1833,1841,2338,2343 (1955).
69. Mulliken, R.S., *J. Chem. Phys.*, **36**, 3428 (1962).
70. Walker, P.D., and Mezey, P.G., *Program MEDLA 93* (Mathematical Chemistry Research Unit, University of Saskatchewan, Saskatoon, Canada, 1993).
71. Walker, P.D., and Mezey, P.G., *J. Am. Chem. Soc.*, **116**, 12022 (1994).
72. Walker, P.D., and Mezey, P.G., *Canad. J. Chem.*, **72**, 2531 (1994).
73. Karle, J., *Proc. Natl. Acad. Sci. USA*, **88**, 10099 (1991).
74. Massa, L., Huang, L., and Karle, J., to be published.
75. Coppens, P., and Hall, M.B., eds., *Electron Distribution and the Chemical Bond* (Plenum, New York and London, 1982).

76. Fliszár, S., *Charge Distributions and Chemical Effects* (Springer, New York, 1983).
77. Purvis III, G.D., and Culberson, C., *Int. J. Quantum Chem., Quantum Biol. Symp.*, **13**, 261 (1986).
78. Kollman, P.A., *J. Amer. Chem. Soc.*, **100**, 2974 (1978).
79. Gillespie, R.J., *Molecular Geometry* (Van Nostrand Reinhold, London, 1972).
80. Gillespie, R.J., and Hargittai, I., *The VSEPR Model of Molecular Geometry* (Allyn and Bacon, Boston, 1991).
81. Wang, J., and Smith Jr., V.H., *Internat. J. Quantum Chem.*, **52**, 1145 (1994).
82. Pichon-Pesme, V., Lecomte, C., Wiest, R., and Benard, M., *J. Am. Chem. Soc.*, **114**, 2713 (1992).
83. Wiest, R., Pichon-Pesme, V., Benard, M., and Lecomte, C., *J. Phys. Chem.*, **98**, 1351 (1994).
84. Pichon-Pesme, V., Lecomte, C., and Lachekar, H., *J. Phys. Chem.*, to be published.
85. Frisch, M.J., Head-Gordon, M.,Trucks, G.W., Foresman, J.B., Schlegel, H.B., Raghavachari, K., Robb, M.A., Binkley, J.S., González, C., DeFries, D.J., Fox, D.J., Whiteside, R.A., Seeger, R., Melius, C.F., Baker, J., Martin, R., Kahn, L.R., Stewart, J.J.P., Topiol, S., and Pople, J.A., *GAUSSIAN 90* (Gaussian Inc., Pittsburgh, PA, 1990).
86. Szabo, A., Ostlund, N.S., *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Macmillan, New York, 1982).
87. Parr, R.G., and Yang, W., *Density Functional Theory of Atoms and Molecules* (Clarendon Press, Oxford, 1989).
88. Kryachko, E.S., and Ludena, E.V., *Density Functional Theory of Many-Electron Systems* (Kluwer, Dordrecht, 1989).
89. March, N.H., *Electron Density Theory of Atoms and Molecules* (Academic, New York,1989).
90. Hartree, D.R., *Proc. Cambridge Phil. Soc.*, **24**, 111, 426 (1928), **25**, 225, 310 (1929).
91. Fock, V., *Z. Physik*, **61**, 126 (1930).
92. Hall, G.G., *Proc. Roy. Soc. London Ser. A*, **205**, 541 (1951).
93. Roothaan, C.C., *Rev. Mod. Phys.*, **23**, 69 (1951), **32**, 179 (1960).
94. Heisenberg, W, *Z. Physik*, **43**, 172 (1927).
95. Löwdin, P.-O., *Adv. Chem. Phys.*, **18**, 365 (1950).
96. Fukui, K., in Löwdin, P.-O., and Pullman, B., eds., *Molecular Orbitals in Chemistry, Physics, and Biology* (Academic Press, New York, 1964).
97. Zadeh, L.A., *Inform. Control*, **8**, 338 (1965).
98. Zadeh, L.A., *J. Math. Anal. Appl.*, **23**, 421 (1968).
99. Kaufmann, A., *Introduction à la Théorie des Sous-Ensembles Flous* (Masson, Paris, 1973).
100. Zadeh, L.A., "Theory of Fuzzy Sets". In *Encyclopedia of Computer Science and Technology* (Marcel Dekker, New York, 1977).

101. Gupta, M.M., Ragade, R.K., and Yager, R.R., eds., *Advances in Fuzzy Set Theory and Applications* (North-Holland, Leyden, 1979).
102. Dubois, D., and Prade, H., *Fuzzy Sets and Systems: Theory and Applications.* (Academic Press, New York, 1980).
103. Sanchez E., and Gupta, M.M., eds., *Fuzzy Information, Knowledge Representation and Decision Analysis* (Pergamon Press, London, 1983).
104. Prugovecki, E., *Found. Phys.*, **4**, 9 (1974).
105. Prugovecki, E., *Found. Phys.*, **5**, 557 (1975).
106. Prugovecki, E., *J. Phys.* A, **9**, 1851 (1976).
107. Ali, S.T., and Doebner, H.D., *J. Math. Phys.*, **17**, 1105 (1976).
108. Ali, S.T., and Prugovecki, E., *J. Math. Phys.*, **18**, 219 (1977).
109. Mezey, P.G., and Maruani, J., *Mol. Phys.*, **69**, 97 (1990).
110. Maruani, J., and Mezey, P.G., *J. Chim. Phys.*, **87**, 1025 (1990).
111. Mezey, P.G., and Maruani, J., *Int. J. Quantum Chem.*, **45**, 177 (1993).
112. Mezey, P.G., *Program DER95* (Mathematical Chemistry Research Unit, University of Saskatchewan, Saskatoon, Canada, 1995).
113. Mezey, P.G., *Program WAT95* (Mathematical Chemistry Research Unit, University of Saskatchewan, Saskatoon, Canada, 1995).
114. Hohenberg, P., and Kohn, W., *Phys. Rev.*, **136**, B864 (1964).
115. Kohn, W., and Sham, L.J., *Phys. Rev.*, **140**, A1133 (1965).
116. Parr, R.G., *Proc. Natl. Acad. Sci. USA*, **72**, 763 (1975).
117. Parr, R.G., *J. Phys. Chem.*, **92**, 3060 (1988).
118. Levy, M., *Proc. Natl. Acad. Sci. USA*, **76**, 6062 (1979).
119. Levy, M., *Phys. Rev. A*, **26**, 1200 (1982).
120. Ludena, E.V., *J. Chem. Phys.*, **79**, 6174 (1983).
121. Becke, A., *Phys. Rev. A*, **33**, 2786 (1986).
122. Becke, A., *J. Chem. Phys.*, **84**, 4524 (1986).
123. Becke, A., *J. Chem. Phys.*, **88**, 1053 (1988).
124. Becke, A., *Phys. Rev. A*, **38**, 3098 (1988).
125. Perdew, J.P., *Phys. Rev. B*, **33**, 8822 (1986).
126. Ziegler, T., *Chem. Rev.*, **91**, 651 (1991).
127. Salahub, D.R., *Adv. Chem. Phys.*, **69**, 447 (1987).
128. Pápai, I., Goursot, A., St.-Amant, A., and Salahub, D.R., *Theor. Chim. Acta*, **84**, 217 (1992).
129. Labanowski, J.K., and Andzelm, J., eds., *Density Functional Methods in Chemistry* (Springer-Verlag, New York, 1991).
130. Andzelm, J., and Wimmer, E., *J. Chem. Phys.*, **96**, 1280 (1992).
131. Politzer, P., *J. Chem. Phys.*, **86**, 1072 (1987).
132. Seminario, J.M., and Politzer, P., *Int. J. Quantum Chem. Symp.*, **26**, 497 (1992).
133. Tachibana, A., *Int. J. Quantum Chem.*, **34**, 309 (1988).
134. Tachibana, A., "Density Functional Theory for Hidden High-T_C Superconductivity." In Hatfield, W.E., and Miller, J.H. Jr., eds., *High Temperature Superconducting Materials* (Dekker, New York, 1988).

135. van Leuven, R., and Baerends, E.J., *Int. J. Quantum Chem.,* **52**, 711 (1994).
136. Hertwig, R.H., and Koch, W., *J. Comput. Chem.,* **16**, 576 (1995).
137. Morokuma, K., *Accts. Chem. Res.,* **10**, 294 (1977).