

SHAPE GROUP STUDIES OF MOLECULAR SIMILARITY: SHAPE GROUPS AND SHAPE GRAPHS OF MOLECULAR CONTOUR SURFACES

Paul G. MEZEY

*Department of Chemistry and Department of Mathematics, University of Saskatchewan,
Saskatoon, Saskatchewan, Canada S7N 0W0*

Received 9 October 1987

Abstract

The earlier, symmetry-independent group theoretical characterization of the shapes of three-dimensional molecular functions, such as electrostatic potentials, electronic charge densities, or molecular orbitals, is extended and compared to a new family of shape descriptors. The incidence graphs and shape graphs, defined by the curvature properties of various molecular contour surfaces, provide an easily visualizable, alternative mathematical technique for a computer-based, non-visual evaluation of molecular shape and molecular similarity. The invariance domains of incidence graphs, shape graphs and shape groups within the dynamic shape space D provide the mathematical basis for the development of a general method for a dynamic description of molecular shapes.

1. Introduction

Characterization of molecular shapes and the recognition and description of similarity among molecules are of major importance in chemistry. Computational methods for the detection of common shape features within families of molecules showing similar biochemical activity have acquired particular significance in modern pharmaceutical research, as well as in pesticide and herbicide chemistry. Molecular shape is an important consideration in drug design and in the development of quantitative structure-activity relationships (QSAR).

The earliest approaches for the description of molecular shape have followed two main strategies: either based on a molecular "skeleton" described by formal bonds, structural formulas and "ball-and-stick" models developed from them, or based on a model of "hard" molecular body, developed from space-filling models, Van der Waals surfaces, or similar descriptions. Whereas both types of models are suitable to describe some internal motions, such as rigid bond rotations, nevertheless they lack the flexibility and deformability of real, non-rigid molecules.

In more recent approaches, various aspects of molecular shape have been described using contour surfaces of calculated quantum chemical properties, most commonly charge density contours and contours of electrostatic potentials [1–29]. By allowing the nuclei to undergo unconstrained configuration changes and by selecting several representative values for the quantum chemical property, e.g. for charge density, it is possible, in principle, to generate a whole sequence of contour surfaces that provides a detailed description. The main problem with the latter approach is interpretation: the contour surfaces are usually displayed on a computer screen and the recognition of shape features and decisions concerning similarity between molecules are based on visual inspection. This is not necessarily the most reliable approach, particularly if many (for example, several thousand) molecules are involved.

It appears desirable to develop alternative methods, where shape recognition and the evaluation of molecular similarity are based on the mathematical (geometrical and topological) properties of the contour surfaces. This can lead to a fully automatic, computer-based analysis of large families of molecules. Such an analysis is possible using the algebraic technique of *symmetry-independent shape groups*, proposed recently [30–33]. In this method, each contour surface is decomposed into domains of various curvature properties, or domains defined by some other physical criteria. The topological relations among these domains define a family of *homology groups*, using the methods of algebraic topology [34–37]. These groups, while independent of the symmetry of the contour surfaces, provide a detailed shape characterization, as well as a mathematical basis for a precise evaluation of molecular similarity. Since in the actual construction each group is defined by the *shape* of the contour surface, in the molecular context these groups are collectively referred to as *molecular shape groups* [30,31].

In the present study, we shall consider the properties of such shape groups in comparison with some alternative tools for the characterization of molecular shape: *incidence graphs*, *shape graphs*, and *dynamic shape graphs*, the latter describing the dynamic features of molecular shape. All these shape descriptors have their common, physical origin in molecular contour surfaces, such as charge density contours or contours of electrostatic potentials, that are used to define the three-dimensional body of molecules. Molecular contour surfaces may be subdivided into domains (cells) based on curvature or other properties. Whereas the subdivision of the *continuous* contour surface into domains is a function of a set of *continuous* parameters (e.g. curvature bounds), nevertheless the resulting tools of shape characterization are groups and graphs of *discrete* mathematics, which is an essential simplification. The fundamental interrelations among these domains and their boundary segments define algebraic-topological relations which can be represented by groups: by molecular shape groups [30,31]. Alternatively, some of these interrelations can be represented by various graphs. The fundamental concepts leading to incidence graphs and to some of the shape graphs may be introduced relying on simple intuitive notions,

without direct reference to algebraic topology and shape groups. For example, the vertices of some incidence graphs correspond to the domains and their boundary segments, whereas an edge between two vertices v_1 and v_2 corresponds to the relation " v_1 is on the boundary of v_2 " or " v_2 is on the boundary of v_1 ". However, a unified treatment and comparison of these shape descriptors are the simplest if one chooses as the starting point the topological model leading to the shape groups [30,31]. Furthermore, an important family of *dynamic shape graphs* is defined directly in terms of shape groups (*vide infra*); hence, some of the tools of algebraic topology are essential for this purpose.

In what follows, we shall give a rather brief review of the construction, definition, and some properties of shape groups that are essential to the development of the method described in the subsequent sections. Note that a more detailed introduction can be found in refs. [30] and [31].

Each shape group $H^p(\nu, a, b)$ is characterized by four parameters, p , ν , a , and b , where the superscript p is a dimension parameter, index ν is characteristic to a family of subsets obtained by a decomposition of the contour surface, with reference to a curvature parameter b , whereas parameter a is the contour value, e.g. the charge density value along an isodensity contour $G(a)$. The precise definitions of parameters b , p , ν , and a brief outline of the shape groups, are given below.

We may regard a small neighborhood of point r on the contour surface $G(a)$ as a function expressed over the tangent plane ${}^2R(r)$ of $G(a)$ at this point r . This function is regarded negative along the normal vector of the tangent plane pointing toward the interior of $G(a)$, it is zero at point r , and positive in the opposite direction. By the very definition of the tangent plane ${}^2R(r)$, this function has a *critical point* at point r . Since the surface $G(a)$ is twice continuously differentiable, the Hessian matrix $\mathbb{H}(r)$ of this function is well defined at point r .

The matrix elements $H_{ij}(r)$ of $\mathbb{H}(r)$ are the second derivatives of a function obtained by regarding the contour surface $G(a)$ as a function defined over its local tangent plane at point r . Since in general the tangent plane ${}^2R(r)$ and the local coordinate system varies with point r , we shall refer to this Hessian matrix

$$\mathbb{H}(r) \tag{1}$$

as the *local Hessian* at point r [30,31]. The local curvature properties of the contour surface $G(a)$ at each point r are characterized by $\mathbb{H}(r)$: the *local canonical curvatures* h_1 and h_2 are the *eigenvalues* of the local Hessian matrix $\mathbb{H}(r)$ [38].

The local curvature properties are distinguished according to a curvature parameter b : the number $\mu = \mu(r, b)$ for each point r is defined as the number of local canonical curvatures that are less than b . In the special case of $b = 0$, μ is the number of negative eigenvalues, that is, μ is the *index* of point r . Each point r of the contour surface $G(a)$ belongs to one of three disjoint subsets A_0 , A_1 or A_2 of $G(a)$,

where $G(a) = A_0 \cup A_1 \cup A_2$, depending on whether none, one or both, respectively, of the local canonical curvatures h_1 and h_2 at point r is(are) smaller than the reference value b [38]. Each of these sets A_0 , A_1 and A_2 may well be disconnected. A maximum connected component of set A_0 , A_1 or A_2 is denoted by D_0 , D_1 or D_2 , respectively. A typical D_μ domain is a two-dimensional subset of the contour surface $G(a)$ for each index $\mu = 0, 1, 2$. We shall use index μ for a general D_μ domain and the index ν for distinguished domains, for example, for those D_ν domains which are eliminated from $G(a)$.

These D_μ domains form the basis of a decomposition of the contour surface $G(a)$ into simply connected subsets: if a D_μ domain is multiply connected, then it is subdivided into simply connected subsets $D_{\mu,k}$. The resulting shape groups are independent of the choice of the actual subdivision.

In the more general case of an arbitrary, fixed curvature parameter b , each point r belongs to one and only one of the $D_{\mu,k}$ domains of some index μ and serial index k , depending on which one of the following three conditions is fulfilled:

$$r \in D_{0,k} \quad \text{for some } k \quad \text{iff} \quad b \leq h_1, h_2, \quad (2)$$

$$r \in D_{1,k} \quad \text{for some } k \quad \text{iff} \quad h_1 < b \leq h_2, \quad (3)$$

$$r \in D_{2,k} \quad \text{for some } k \quad \text{iff} \quad h_1, h_2 < b. \quad (4)$$

In the above notation for the $D_{\mu,k}$ domains, the reference to curvature parameter b is suppressed. If explicit reference is needed, then the $D_{\mu b, k}$ notation may be used.

This decomposition of the contour surface $G(a)$ by the D_μ domains and their $D_{\mu,k}$ subdivisions leads to three types of objects:

- (i) two-dimensional objects, that are either the D_μ domains themselves or their simply connected subdivisions $D_{\mu,k}$;
- (ii) one-dimensional objects, that are boundary line segments of the D_μ domains or their subdividing lines;
- (iii) zero-dimensional objects, that are points where boundary line segments and subdivision lines meet.

The above surface domains, line segments and points are analogous to the faces, edges and vertices, respectively, of a polyhedron. We will use the $C(p, \mu, i)$ notation for these objects, called p -cells, where p is the dimension, μ is the index of the D_μ domain that contains cell $C(p, \mu, i)$, and i is a serial index.

Each cell $C(p, \mu, i)$ has an *orientation* $o(i)$, formally denoted by $+1$ or -1 . For a 2-cell, $o(i) = +1$ may correspond to an anticlockwise orientation, represented by a circular arrow on the cell, indicating an anticlockwise rotation when viewed

from the outside of $G(a)$; then $o(i) = -1$ corresponds to a clockwise orientation. For a 1-cell, orientation can be represented by an arrow along the corresponding line segment, with either, but fixed, assignment of one of the two possible values for $o(i)$, whereas for 0-cells, all orientations may be defined to be the same, e.g. $o(i) = +1$. The *incidence number* $\eta_{ij}(p-1)$ between a p -cell $C(p, \mu, i)$ and a $(p-1)$ -cell $C(p-1, \nu, j)$ is given by

$$\eta_{ij}(p-1) = \begin{cases} \sum_k \gamma_{ij}(p-1, k) & \text{if } C(p-1, \nu, j) \cap C^c(p, \mu, i) \neq \emptyset \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

where for $p = 2$, $\gamma_{ij}(p-1, k)$ is $+1$ if the two arrows point the same way and it is -1 otherwise, and for $p = 1$, $\gamma_{ij}(p-1, k)$ is $+1$ if the arrow of the 1-cell points toward the 0-cell and it is -1 if it points away from it. A 2-cell may be folded, meeting a 1-cell on several occasions; a 1-cell may be a loop starting and ending at the same 0-cell; hence, the $\gamma_{ij}(p-1, k)$ numbers are summed up for all k occasions the two cells meet. In $C^c(p, \mu, i)$, the superscript c refers to the set theoretical closure of cell $C(p, \mu, i)$ in the metric of the three-dimensional Euclidean space. That is, the incidence number $\eta_{ij}(p-1)$ is zero if $C(p-1, \nu, j)$ is *not* on the set theoretical boundary of cell $C(p, \mu, i)$; otherwise it is the sum of individual "encounter numbers" $\gamma_{ij}(p-1, k)$ that may take the values $+1$ or -1 , depending on the relative orientations of these two cells.

A p -chain is a formal linear combination of p -cells

$$c^p = \sum_{i=1} u_i C(p, \mu, i), \quad (6)$$

where we consider only integer coefficients u_i .

The *boundary* Δc^p of a p -chain c^p is defined as

$$\Delta c^p = \sum_{i,j} u_i \eta_{ij}(p-1) C(p-1, \nu, j). \quad (7)$$

A p -cycle is a p -chain with zero boundary. A *bounding p -cycle* is a p -cycle that is the boundary of a $(p+1)$ -chain:

$$c^p = \Delta c^{p+1}. \quad (8)$$

For the addition of coefficients, the p -chains, p -cycles and bounding p -cycles form the Abelian groups C^p , Z^p , and B^p , respectively. The p th *integral homology group* H^p is defined as the difference (quotient) group

$$H^p = Z^p - B^p. \quad (9)$$

By eliminating all $D_{\nu,i}$ domains of some specified index ν and all indices i , for the resulting *truncated contour surface* $G(a, \nu)$, three homology groups, the *shape groups* $H^p(\nu, a, b)$, are obtained, one for each of the dimensions $p = 0, 1$, and 2 . Typically, the shape groups $H^1(2, a, b)$ for intermediate a and b values provide the most chemically interesting information.

These shape groups provide a concise description of the shape characteristics of the given molecular property (e.g. charge density). They are invariant within certain intervals of the two parameters a and b , and those parameter values a_k and b_q where the groups change is characteristic to the given molecular configuration. On the other hand, for a given parameter pair a, b , the shape groups are *invariant* to minor configurational changes within certain domains of a nuclear configuration space [33]. Within these domains, the essential shape features of $G(a, \nu)$, with respect to the specified curvature parameter b , are preserved, that is, the corresponding minor changes of the nuclear configurations do not change the shape classification of the molecule with respect to the curvature parameter b and the physical property described by $G(a)$. Note, however, that for a different curvature parameter b , the invariance domains of shape groups within the nuclear configuration space are in general different. As we shall discuss in more detail in a subsequent section, a complete description of all shape features along the entire configuration space can be given by considering the *product space*

$$D = M \otimes P, \quad (10)$$

where M is a *metric* nuclear configuration space [39,40] and P is the parameter space of parameters a and b . The dimension of M is $3N - 6$ for molecules of N nuclei, whereas the dimension of P is 2 ; hence, the dimension of D is $3N - 4$. Since within space D one may follow the shape variations with changes in nuclear configurations for all values of parameters a and b , space D is referred to as the *dynamic shape space* (*vide infra*).

For a single nuclear configuration, the shape characterization based on D_u domains may be regarded as a *discrete, algebraic shape description* that depends on two *continuous parameters* a and b . This feature allows one to use a concise characterization made possible by the discrete nature of the algebraic model, yet the presence of continuous parameters permits a details description. Particularly important is the fact that the discrete descriptors change only at certain specific values of the continuous parameters, as we shall see later. This is the very reason why the method is advantageous when compared to a straightforward, but cumbersome representation of shapes by a continuous, three-dimensional molecular function itself (e.g. by a three-dimensional charge density where, in fact, the density value corresponds to a

fourth dimension). By considering the entire, chemically accessible range of the two continuous parameters, one, indeed, obtains a *complete* description of the shape of the three-dimensional molecular function (e.g. charge density) for the given nuclear configuration, or for a family of configurations within space M .

It is important to emphasize that the shape groups do not depend on the choice of subdivisions of the D_μ domains, in spite of the fact that the incidence numbers and the set of all p -chains ($p < 2$) do depend on this choice. As we shall see, this does not diminish the usefulness of analyzing the incidence relations directly.

In this study, we propose alternative descriptions and representations of shapes of contour surfaces, using information derived directly from the incidence numbers $\eta_{ij}(p-1)$, on the one hand, and based on neighbor relations of D_μ domains, on the other hand. For each fixed nuclear configuration, these descriptions lead to graph representations that are more directly amenable to pictorial interpretation than the more abstract homology groups. In a subsequent section, we will introduce the concepts of the dynamic shape space and dynamic shape graphs for the description of shapes of conformationally flexible molecules or reacting systems, which will provide a unified framework for shape groups, incidence graphs, and shape graphs.

2. Incidence matrices, incidence graphs, and shape graphs

For dimension p ($p = 1, 2$) and curvature parameter b , the incidence numbers $\eta_{ij}(p-1, b)$ of contour surface $G(a)$, taken as matrix elements, form the *incidence matrix*

$$\eta(p-1, b). \quad (11)$$

Evidently, this incidence matrix does depend on the subdivision of D_μ domains, since the subdivision defines both the 1-cells and the 0-cells. Consequently, any shape description that is based directly on a particular incidence matrix $\eta(p-1, b)$ is subdivision-dependent. However, this ambiguity can be avoided if the subdivisions are standardized in some well-defined way. In particular, two approaches appear natural within the above context of the D_μ domain partitioning, both involving the *minimum number* of subdivision lines, *compatible with some additional conditions*:

- (i) a *minimal triangulation* of $G(a)$ involving all boundary lines of the D_μ domains, or
- (ii) a *minimal subdivision* that partitions every multiply connected D_μ domain into simply connected sets.

In the present study, we shall assume that condition (ii) is applied.

The incidence matrix $\eta(p-1, b)$ defines a graph, the *incidence graph* $g_\eta(p-1, b)$, by taking the absolute values of incidence numbers $\eta_{ij}(p-1, b)$ as the elements of an adjacency matrix $Al(g_\eta(p-1, b))$:

$$A_{ij}(g_\eta(p-1, b)) = |\eta_{ij}(p-1, b)|. \quad (12)$$

The pair of incidence graphs $g_\eta(0, b)$ and $g_\eta(1, b)$ provide a concise representation of *all* the interrelations among all cells of different types, hence all shape features relative to some specified reference curvature value b .

The homology-theoretical boundary of a boundary is zero for any c^p chain,

$$\Delta \Delta c^p = 0. \quad (13)$$

As can be shown (see, e.g. ref. [37]), in terms of incidence matrices the above relation can be given as the following matrix identity:

$$\eta(p-1, b)\eta(p-2, b) = \mathbf{0}. \quad (14)$$

The vertices of each incidence graph $g_\eta(p-1, b)$ are the p -cells and the $(p-1)$ -cells of the actual cell-decomposition of contour surface $G(a)$. The edges of $g_\eta(p-1, b)$ are the cell pairs of nonzero incidence relation. Since each edge of $g_\eta(p-1, b)$ necessarily involves one p -cell and one $(p-1)$ -cell, each incidence graph $g_\eta(p-1, b)$ is necessarily a *bipartite graph* [41], with the two families of vertices of the bipartition precisely the families of p -cells and $(p-1)$ -cells.

Since in both incidence graphs $g_\eta(p-1, b)$ and $g_\eta(p-2, b)$ the vertex family of all $(p-1)$ -cells is common, it is natural to consider *all* incidence relations (orientations disregarded) within a *single* graph. This incidence graph $g_\eta(b)$ is defined as the smallest supergraph of incidence graphs $g_\eta(1, b)$ and $g_\eta(0, b)$. In terms of vertex sets $V(g)$ and edge sets $E(g)$, the incidence graphs $g_\eta(b)$ are defined by

$$V(g_\eta(b)) = V(g_\eta(1, b)) \cup V(g_\eta(0, b)) \quad (15)$$

and

$$E(g_\eta(b)) = E(g_\eta(1, b)) \cup E(g_\eta(0, b)). \quad (16)$$

Evidently, the above construction implies that each incidence graph $g_\eta(b)$ is also a bipartite graph, where the two families of vertices of the bipartition are the set of all 1-cells $\{C(1, \mu, i)\}$, and the union of the families of all 0-cells and 2-cells $\{C(0, \mu, i)\} \cup \{C(2, \mu, i)\}$. In fact, $g_\eta(b)$ can be regarded as a special "tripartite" graph: there are three families of vertices (those of the 0-cells, 1-cells and 2-cells, respectively), with no edge within any family, and with the special restriction that no edge exists between 0-cells and 2-cells. Such incidence graphs and their subgraphs can be characterized in terms of their chromatic numbers [41].

The above incidence graphs $g_\eta(0, b)$, $g_\eta(1, b)$ and $g_\eta(b)$ are based on the incidence relation of homology theory of algebraic topology. However, a different

family of graphs describing the shape of contour surface $G(a)$, with respect to a curvature parameter b and the resulting D_μ domain decomposition, can also be defined using the neighbor relations among the D_μ domains. These graphs will be referred to as *shape graphs*, and the $g_\mu(b)$ notation will be used to distinguish them from incidence graphs $g_\eta(b)$.

In contrast to the incidence graphs, the construction of shape graphs $g_\mu(b)$ does not require any subdivision or triangulation of the D_μ domains of $G(a)$, hence there is no need for choosing an appropriate standardization of subdivisions. Instead, the construction is based directly on the relative distribution of various D_μ domains along the contour surface $G(a)$.

The vertex set of graph $g_\mu(b)$ is the family of D_μ domains:

$$V(g_\mu(b)) = \{D_{\mu,i}\}, \quad (17)$$

where $D_{\mu,i}$ denotes the D_μ set that is the i th maximum connected component of the set A_μ of all points of index μ_b . It is important to keep in mind that the index i in the notation $D_{\mu,i}$ corresponds to a well-defined decomposition of set A_μ into maximum connected components, whereas in eqs. (2)–(4), index k in the notation $D_{\mu,k}$ corresponds to a decomposition of A_μ into simply connected subsets; the latter decomposition may involve some arbitrariness.

The edge set is defined in terms of the following N -neighbor relation, analogous to the "symmetric strong neighbor" relation of potential surface catchment regions of reaction topology [42,40]:

$$N(D_{\mu,i}, D_{\mu',i'}) = \begin{cases} 1 & \text{if } (D_{\mu,i}^c \cap D_{\mu',i'}) \cup (D_{\mu,i} \cap D_{\mu',i'}^c) \neq \emptyset \\ 0 & \text{otherwise,} \end{cases} \quad (18)$$

where the superscript c denotes the set-theoretical closure within the Euclidean metric of the laboratory frame.

The edge set is the family of pairs of D_μ domains with nonzero N -neighbor relation:

$$E(g_\mu(b)) = \{(D_{\mu,i}, D_{\mu',i'}): N(D_{\mu,i}, D_{\mu',i'}) = 1\}. \quad (19)$$

Shape graphs $g_\mu(b)$ gives a detailed shape characterization of contour surface $G(a)$, with respect to reference curvature b .

The typical nonzero N -neighbor relations are those involving the following types of pairs of D_μ domains:

$$(D_{0,i}, D_{1,i'}) \quad (20)$$

and

$$(D_{1,i'}, D_{2,i''}). \quad (21)$$

No nonzero N -neighbor relation is possible between two D_μ domains of the same μ index, since each $D_{\mu,i}$ domain is a maximum connected component of set A_μ of all points of $G(a)$ within index μ .

However, a nonzero N -neighbor relation between a D_0 and a D_2 domain is not impossible. A point $r \in G(a)$ may simultaneously belong to a $D_{0,i'}$ domain and to the closure of a $D_{2,i}$ domain,

$$r \in D_{2,i}^c \cap D_{0,i'}, \quad (22)$$

if at point r both local canonical curvatures $h_1(r)$ and $h_2(r)$ are equal to the reference curvature b . This is a very uncommon possibility, that is in fact a sufficient and necessary condition for the existence of a $(D_{2,i}, D_{0,i'})$ edge within shape graph $g_\mu(b)$.

The above conditions may be expressed directly in terms of local Hessian matrices. The closure $D_{2,i}^c$ of a domain $D_{2,i}$ contains point r , where for the eigenvalues of matrix $\mathbb{H}(r)$ the relations

$$h_1 \leq b = h_2 \quad (23)$$

hold. Since in the case of the strict inequality $<$ in eq. (23) such a point r belongs to some $D_{1,i'}$ domain, these two domains $D_{2,i}$ and $D_{1,i'}$ are N -neighbors, according to definition (18).

Similarly, the closure $D_{1,j}^c$ of a domain $D_{1,j}$ contains points r for which the relations

$$b = h_1, \quad b < h_2 \quad (24)$$

hold; consequently, point r must belong to some domain $D_{0,j'}$, implying that $D_{1,j}$ and $D_{0,j'}$ are N -neighbors.

Clearly, a point r within the closure $D_{2,i}^c$ of a $D_{2,i}$ domain can belong to a domain $D_{0,i'}$ only if for the eigenvalues of $\mathbb{H}(r)$

$$h_1 = h_2 = b \quad (25)$$

holds. Consequently, a pair of domains $D_{2,i}$ and $D_{0,i'}$ are N -neighbors if and only if at a boundary point r the local Hessian matrix has two *degenerate eigenvalues*, which happen to be *equal to the reference curvature value* b . This is a sufficient and necessary condition for an N -neighbor relation between D_0 and D_2 domains. Note that each D_0 domain is a *closed* set; hence, intersections of the type $D_0^c \cap D_2$ are always empty and they cannot contribute to a nonzero N -neighbor relation, according to definition (18).

This result has an interesting consequence regarding the type of shape graphs $g_\mu(b)$ that may occur in chemical applications. For a given reference curvature b , the *typical boundary points* r of a D_μ domain decomposition of the contour surface $G(a)$ are characterized by either one of the two pairs of relations: relations (23) with the strict inequality $<$, or relations (24). That is, typically only the (1, 2) and the (0, 1) combinations occur for the μ indices in the N -neighbor relations on $G(a)$. Consequently, a *typical shape graph* $g_\mu(b)$ is a *bipartite graph*; the two vertex families of the bipartition are the set of all $D_{1,i'}$ domains and the union of the sets of all $D_{0,i}$ and all $D_{2,i''}$ domains. In fact, similarly to the incidence graphs $g_\eta(b)$, a typical shape graph $g_\mu(b)$ is a formal, special tripartite graph. There are three families of vertices:

$$\{D_{0,i}\}, \quad (26)$$

$$\{D_{1,i'}\}, \quad (27)$$

and

$$\{D_{2,i''}\}, \quad (28)$$

where there are no edges involving two vertices from the same set, and in the typical case (but not in all cases), the following special restriction applies: no edge exists between vertices of the first and last of these sets ((26) and (28)).

As implied by conditions (2)–(4), exceptions to the typical case require a special degeneracy of the eigenvalues of a local Hessian matrix (eq. (25)). If such degeneracy occurs, then a (0, 2) combination for the μ indices may exist, and then the resulting shape graph $g_\mu(b)$ may be a graph that is *not* bipartite. However, even in this special case, as long as all three types of D_μ domains are present, $g_\mu(b)$ is still a formal tripartite graph of the same sets of tripartition as given by eqs. (26)–(28), but without the additional restriction. There exists at least one edge between some vertex of the first set, given by eq. (26), and some vertex of the last set, given by eq. (28).

Since all of these graphs can be actually drawn on the contour surface $G(a)$, it is evident that the existence of a homeomorphism between $G(a)$ and a two-sphere 2S is a sufficient (but not necessary) condition for the *planarity* [41] of any one of these graphs.

3. The dependence of incidence graphs $g_\eta(b)$ and shape graphs $g_\mu(b)$ on curvature parameter b and contour parameter a of the contour surface $G(a)$

Since the curvature classification of contour points r of $G(a)$ refers to an orientation with respect to the interior and exterior of $G(a)$, the reference curvature b may take both positive and negative values:

$$-\infty < b < \infty. \quad (29)$$

Just as in the case of shape groups [30–33], the chemically interesting incidence graphs and shape graphs belong to b parameters of absolute values bounded by some constant b_0 . For the special choice of $b = 0$, the incidence graphs and shape graphs describe the interrelations among convex-, concave- and saddle-type domains on $G(a)$. In this special case, the point set of the boundary lines of all D_μ domains consists of all the parabolic points of $G(a)$ [30–33].

For a negative b of a large enough absolute value,

$$b \ll 0, \quad (30)$$

relation (2) holds for eigenvalues h_1 and h_2 at all points $r \in G(a)$; consequently, the entire contour surface $G(a)$ is a single, simply connected D_0 domain. This implies that incidence graph $g_\eta(0, b)$ does not exist, and the incidence graphs $g_\eta(1, b)$, $g_\eta(b)$ and the shape graph $g_\mu(b)$ are trivial graphs of a single vertex.

For a large enough positive value b ,

$$0 \ll b, \quad (31)$$

relation (4) holds for eigenvalues h_1 and h_2 at each point $r \in G(a)$. The entire contour surface is a single, simply connected D_2 domain. Just as in the previous case, incidence graph $g_\eta(0, b)$ does not exist; the incidence graphs $g_\eta(1, b)$, $g_\eta(b)$ and the shape graph $g_\mu(b)$ are trivial graphs.

Consider now the chemically more interesting intermediate values for curvature parameter b , leading to nontrivial graphs. A continuous change in reference curvature b leaves the incidence graphs and the shape graphs invariant within certain (b_1, b_2) intervals; however, the graphs change at the interval limits. The typical change is the addition or elimination of a single vertex, accompanied by the addition or elimination of incident edges, respectively, fulfilling the appropriate conditions for bipartite or formal tripartite graphs; discussed above. Simultaneous additions or eliminations of several vertices are usually indications of a nontrivial symmetry of the contour surface $G(a)$.

The incidence graphs g_η and shape graphs g_μ show a similar dependence on the contour parameter a of $G(a)$. For example, let us assume that the physical property used for shape description is the electronic charge density, that is, $G(a)$ is an isodensity contour of the molecule, characterized by the density value a . Then, for any fixed curvature parameter b , the D_μ domain decomposition, and hence the incidence and shape graphs obtained, depend on this density value a . The graphs are invariant within certain density intervals (a_1, a_2) ; however, they change at the limits a_1 and a_2 of such intervals of invariance. The typical change, just as for parameter b ,

is the inclusion or elimination of a single vertex with its incident edges. Simultaneous multiple vertex changes usually indicate some nontrivial symmetry of $G(a)$. For the electronic charge density contour value a , one may in principle consider the entire range of

$$0 \leq a < \infty. \quad (32)$$

However, very small and very large contour values a provide little useful insight, since they correspond to large, essentially spherical surfaces at large distances from the molecule at their centers, and to a family of small, nearly spherical surfaces near the nuclei, respectively. For all but the extreme values of curvature parameter b , these cases correspond to trivial graphs and completely disconnected graphs, respectively. The chemically interesting cases are those of intermediate charge density values a .

The dependence of these graphs on the two parameters a and b can be represented by the distribution of graphs along an a, b parameter plane P . Such a representation has been introduced earlier for shape groups [31], and in the case of incidence graphs and shape graphs, the analogous considerations apply. In particular, a *complete description of shapes* can be given by specifying the boundaries of those regions of the a, b parameter plane P within which the graphs are invariant.

4. The dependence of incidence graphs $g_\eta(b)$ and shape graphs $g_\mu(b)$ on the nuclear configuration

Molecules are not rigid, geometrical objects, but dynamic, topological entities ([40] and references therein). Hence, a complete characterization of their shapes requires the consideration of entire families of formal nuclear configurations and not just one or a few selected nuclear geometries. However, the contour surfaces $G(a)$ and hence the resulting shape groups H^P , incidence graphs g_η , and shape graphs g_μ , are easiest to define for fixed, formal nuclear geometries. Evidently, these groups and graphs for specified parameter values a and b , as well as their *domains of invariance within the a, b parameter plane P* , do depend on the formal nuclear configuration. Nevertheless, these shape groups and graphs do reflect the topological nature of molecular species, since they remain *invariant* to minor geometry variations within certain domains of the nuclear configuration space.

In reaction topology, a $(3N - 6)$ -dimensional *metric space M* of all internal configurations of any N -atom system can be given [39,40]. This metric space M of metric d_M is a convenient tool for the classification of formal nuclear configurations K into *catchment regions* of any given potential energy hypersurface [43]. These catchment regions represent the various chemical species of a given, fixed stoichiometry and the electronic state of a specified k number of electrons, associated with the given potential energy hypersurface [43,40].

We assume that the parameter plane P of points (a, b) ,

$$P = \{(a, b)\}, \quad (33)$$

is provided with an Euclidean metric d_p ,

$$d_p [(a_1, b_1), (a_2, b_2)] = [(a_1 - a_2)^2 + (b_1 - b_2)^2]^{1/2}. \quad (34)$$

In terms of metric spaces M and P , a *complete, dynamic description of molecular shape* can be given as follows.

Take the *product space* $D = M \otimes P$ of dimensions $3N - 6 + 2 = 3N - 4$,

$$\dim(D) = \dim(M \otimes P) = 3N - 4, \quad (35)$$

and define a metric d_D as

$$d_D = [d_p^2 + d_M^2]^{1/2}. \quad (36)$$

Space D , provided with metric d_D , is the "dynamic shape space", mentioned in the introduction.

For a given family of N nuclei and for the given electronic state of k electrons, the shape groups H^p , incidence graphs g_η , and shape graphs g_μ are invariant within certain domains of the dynamic shape space D . For each of the above shape descriptors, its invariance domains generate a complete partitioning of the product space D . For example, considering the shape groups H^p or the shape graphs g_μ , the specification of the boundaries of their respective invariance domains within D is equivalent to a complete description of the *dynamic shape features of all molecular species* of stoichiometry specified by the N nuclei and of the given electronic state of the k electrons.

For each dimension p of shape groups H^p ($p = 0, 1, 2$) and incidence graphs $g_\eta(p-1)$ ($p = 1, 2$), as well as for incidence graph g_η and shape graph g_μ , the subsets of invariance within the dynamic shape space are denoted by $A_i(X_i)$:

$$A_i(H^{p,i}) = \{\mathbf{d} : \mathbf{d} \in D, H^p(\mathbf{d}) = H^{p,i}\}, \quad (37)$$

$$A_i(g_{\eta,i}(p-1)) = \{\mathbf{d} : \mathbf{d} \in D, g_\eta(p-1, \mathbf{d}) = g_{\eta,i}(p-1)\}, \quad (38)$$

$$A_i(g_{\eta,i}) = \{\mathbf{d} : \mathbf{d} \in D, g_\eta(\mathbf{d}) = g_{\eta,i}\}, \quad (39)$$

and

$$A_i(g_{\mu,i}) = \{\mathbf{d} : \mathbf{d} \in D, g_\mu(\mathbf{d}) = g_{\mu,i}\}. \quad (40)$$

In the above definitions, $\mathbf{d} \in D$ is a point of the dynamic shape space, that is, \mathbf{d} represents a combination of a pair $(a, b) \in P$ of a parameter value a for charge density (or electrostatic potential, or MO, etc.) and reference curvature b , on the one hand, and some specified nuclear configuration $K \in M$, on the other hand. Index i is the serial index of the i th reference entity X_i of a given type: the i th reference group $H^{p,i}$, or the i th reference graph $g_{\eta,i}(p-1)$, or $g_{\eta,i}$, or $g_{\mu,i}$,

$$X_i = H^{p,i} \quad (p = 0, 1, 2), \quad g_{\eta,i}(p-1) \quad (p = 1, 2), \quad g_{\eta,i}, \quad \text{or} \quad g_{\mu,i}. \quad (41)$$

On the right-hand sides of eqs. (37)–(40), $H^p(\mathbf{d})$, $g_{\eta}(p-1, \mathbf{d})$, $g_{\eta}(\mathbf{d})$, and $g_{\mu}(\mathbf{d})$ stand for the actual shape group or graph at the specified point \mathbf{d} of the dynamic shape space D . A general notation $X(\mathbf{d})$ is used for any one of these latter, actual shape descriptors at point \mathbf{d} .

For each point \mathbf{d} within a given set $A_i(X_i)$, the actual group or graph $X(\mathbf{d})$ is identical to a specified reference group or graph X_i . These $A_i(X_i)$ sets are pairwise disjoint and generate a partitioning of the dynamic shape space D :

$$D = \bigcup_i A_i(X_i), \quad (42)$$

$$A_i(X_i) \cap A_{i'}(X_{i'}) = \emptyset \quad \text{if } i \neq i'. \quad (43)$$

A set $A_i(X_i)$ is not necessarily connected, and we may consider its maximum connected components $D_{i,j}(X_i)$

$$A_i(X_i) = \bigcup_j D_{i,j}(X_i), \quad (44)$$

$$D_{i,j}(X_i) \cap D_{i,j'}(X_i) = \emptyset \quad \text{if } j \neq j'. \quad (45)$$

A detailed partitioning of the dynamic shape space D into shape regions by shape descriptor X_i can be given in terms of these maximum connected components,

$$D = \bigcup_{i,j} D_{i,j}(X_i), \quad (46)$$

for each of the shape descriptors X_i specified in (41).

Since any continuous change $\mathbf{d}(j) \rightarrow \mathbf{d}(j')$ leading from one component $D_{i,j}(X_i)$ to another component $D_{i,j'}(X_i)$ of $A_i(X_i)$ must involve points \mathbf{d} characterized by a group or graph $X_{j'}$, different from X_j , the term *domain of invariance* is used for the components $D_{i,j}(X_i)$, rather than for the set $A_i(X_i)$.

Partitioning (46) of the dynamic shape space D is completely specified if the boundaries of the domains of invariance $D_{i,j}(X_i)$ are given.

5. Catchment regions of molecular species and invariance domains of the dynamic shape space

The partitioning (46) of the dynamic shape space D involves all possible chemical species composed from the specified N nuclei, that is, it describes the shapes of all possible conformations of all species of a specified stoichiometry with k electrons in the given electronic state. It is of some interest to investigate the following two problems:

- (i) What shapes are available for a given chemical species, allowing only such variations of the formal nuclear geometry that preserve the chemical identity of the species?
- (ii) Which are those chemical species of the given stoichiometry that may have formal nuclear configurations belonging to some common shape invariance domain $D_{i,j}(X_i)$, or invariance set $A_i(X_i)$?

Within the reaction topology of nuclear configuration space M , the chemical species are represented by catchment regions [43], defined in terms of a given potential energy hypersurface of some specified electronic state [39,40]. These catchment regions, denoted by $C(\lambda, t)$ and identified with chemical species, generate a decomposition of the nuclear configuration space M :

$$M = \bigcup_{\lambda, t} C(\lambda, t). \quad (47)$$

Each catchment region $C(\lambda, t)$ is defined as the collection of all those configurations K of M from where a path of formal steepest descent of the given potential energy hypersurface leads to a common critical point $K(\lambda, t)$ of critical point index λ and serial index t .

Within the above model, the two questions (i) and (ii) can be reformulated in terms of relations between the $D_{i,j}(X_i)$ domains of invariance of the dynamic shape space D and the catchment regions $C(\lambda, t)$ of nuclear configuration space M . A simplified, schematic representation of these two problems (i) and (ii) is shown in fig. 1.

For each catchment region $C(\lambda, t)$, a subset $D(C(\lambda, t))$ is given,

$$D(C(\lambda, t)) = P \otimes C(\lambda, t), \quad (48)$$

that is the set of all points $d \in D$ which involve any point $K \in C(\lambda, t)$ of catchment region $C(\lambda, t) \subset M$, in combination with any values of parameters a and b of parameter plane P .

As answer to question (i) can be given as follows:

A molecular shape characterized by descriptor X_i is available to chemical species $C(\lambda, t)$ if and only if

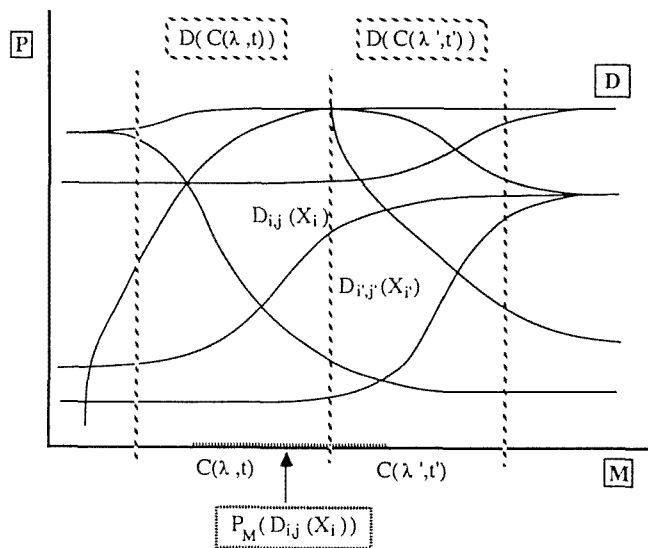


Fig. 1. A schematic representation of the dynamic shape space D , the product space composed from a $(3N - 6)$ -dimensional nuclear configuration space M and the two-dimensional space P of parameters a (e.g. the charge density contour value) and the reference curvature b). In the figure, each of the spaces M and P are schematically represented by one-dimensional coordinate axes. The figure shows the decomposition of the dynamic shape space D into shape invariance domains $D_{i,j}(X_i)$ of shape descriptor X_i (shape group $H^p(v, a, b)$, or incidence graphs $g_{\eta,i}$, $g_{\eta,i}(p - 1)$ ($p = 1, 2$), or shape graph $g_{\mu,i}$). Set $D(C(\lambda, t))$ is the point set of the dynamic shape space available for chemical species represented by catchment region $C(\lambda, t)$. This set overlaps with several shape invariance domains $D_{i,j}(X_i)$. The set $P_M(D_{i,j}(X_i))$ is the projection of shape invariance domain $D_{i,j}(X_i)$ onto the nuclear configuration space M . In the example it overlaps with two catchment regions $C(\lambda, t)$ and $C(\lambda', t')$, indicating that the corresponding shape, as characterized by group or graph X_i , is available to these two chemical species.

$$D(C(\lambda, t)) \cap D_{i,j}(X_i) \neq \emptyset \quad \text{for some index } j. \quad (49)$$

In fact, it is useful to restrict the freedom present on the parameter plane P to some chemically realistic density values (or other property values) a of contour surfaces $G(a)$ and to some intermediate values of curvature parameter b , for example, by considering only a subset $D(C(\lambda, t), (a_1, a_2), (b_1, b_2))$ of set $D(C(\lambda, t))$, defined as

$$D(C(\lambda, t), (a_1, a_2), (b_1, b_2)) = \{d : d \in D, d = a \otimes b \otimes K, a \in (a_1, a_2), b \in (b_1, b_2), K \in C(\lambda, t)\}. \quad (50)$$

Evidently, a molecular shape characterized by descriptor X_i is available to molecular species $C(\lambda, t)$ within some chemically realistic parameter ranges (a_1, a_2) and (b_1, b_2) if and only if

$$D(C(\lambda, t), (a_1, a_2), (b_1, b_2)) \cap D_{i,j}(X_i) \neq \emptyset \quad \text{for some index } j. \quad (51)$$

An answer to question (ii) can be given by considering the projection $P_M(D_{i,j}(X_i))$ of the shape invariance domain $D_{i,j}(X_i)$ on the nuclear configuration space M :

$$P_M(D_{i,j}(X_i)) = \{K : K \in M, \exists (a, b) \in P : a \otimes b \otimes K \in D_{i,j}(X_i)\}. \quad (52)$$

A chemical species $C(\lambda, t)$ may attain a formal nuclear configuration K characterized by shape descriptor X_i within invariance domain $D_{i,j}(X_i)$ if and only if there exists some index j for which

$$P_M(D_{i,j}(X_i)) \cap C(\lambda, t) \neq \emptyset. \quad (53)$$

Conditions (49), (51), and (53) specify all elementary relations between molecular identity and molecular shape within the above topological context.

Considering a given shape invariance domain $D_{i,j}(X_i)$, relation (53) may hold for several *different* chemical species. For example, it may hold for species $C(\lambda', t')$ in addition to species $C(\lambda, t)$. That is,

$$P_M(D_{i,j}(X_i)) \cap C(\lambda', t') \neq \emptyset \quad (54)$$

may also be valid, in addition to (53). Evidently, the simultaneous validity of conditions (53) and (54) is required by any two chemical species $C(\lambda, t)$ and $C(\lambda', t')$ to exhibit a common shape feature, specified by shape descriptor X_i .

6. Global characterization of the dynamic shape space D

For a given shape descriptor X_i , the partitioning (46) of the dynamic shape space D depends on the actual stoichiometry of the N nuclei and on the electronic state of the k electrons. These are the very aspects that define a potential energy hypersurface above the nuclear configuration space M [39,40]. Evidently, a different electronic state, or a different overall stoichiometry leads in general to a different dynamic shape space D .

Instead of comparisons of the shapes of individual molecules, one may address the problem of the comparison of the *families of shapes* available to two different *families* of molecular species, described by two different shape spaces D_A and D_B .

Two such molecular families may share a common stoichiometry, but may have two different electronic states. Alternatively, the stoichiometry may differ between the two families. In either case, with respect to a given type of shape descriptor (group or graph) X_i , the comparison of the two families of molecular shapes available to the two families of molecules is equivalent to a comparison between the shape domain distributions (eq. (46)) of the two dynamic shape spaces D_A and D_B . A concise characterization of the shape invariance domain decomposition (46) of any dynamic shape space D can be given in terms of the *dynamic shape graph* g_D .

The dynamic shape graph g_D is defined by the vertex set

$$V(g_D) = \{D_{i,j}(X_i)\}, \quad (55)$$

containing the shape invariance domains $D_{i,j}(X_i)$ of the dynamic shape space D as elements, and by the edge set

$$E(g_D) = \{(D_{i,j}(X_i), D_{i',j'}(X_{i'})) : N(D_{i,j}(X_i), D_{i',j'}(X_{i'})) = 1\}, \quad (56)$$

containing the pairs of shape invariance domains with a nonzero N -neighbor relation. The comparison of the two distributions of shape invariance domains of the dynamic shape spaces D_A and D_B can be given in terms of their respective dynamic shape graphs g_{D_A} and g_{D_B} .

Each of these graphs g_D characterize an entire dynamic shape space D , that is, all the possible shapes of all possible molecular species that exist along a given potential energy hypersurface. This characterization, as every characterization given in terms of graphs, is concise, and it describes the most essential interrelations, all the possible direct and indirect shape transitions for the given stoichiometry and electronic state. If, however, the detailed analysis of one or a few shape invariance domains $D_{i,j}(X_i)$ is our goal, then a shape analysis can be carried out directly on these $D_{i,j}(X_i)$ sets themselves, using a $(3N - 4)$ -dimensional variant of the D_μ domain decomposition of the boundary hypersurfaces of each set $D_{i,j}(X_i)$ (for the general, n -dimensional D_μ domain decomposition in the $b = 0$ case, see ref. [38]). In terms of such a decomposition, the $(3N - 4)$ -dimensional shapes of the shape invariance domains $D_{i,j}(X_i)$ of space D can also be characterized by *their* shape groups, *their* incidence graphs, and *their* shape graphs. In such a representation, the *dynamic* relations of molecular shape transitions appear as formal *static* shape features of the shape invariance domains $D_{i,j}(X_i)$. Further details of this interesting possibility are beyond the scope of the present study.

EXAMPLE

In fig. 2, various D_μ domain partitionings of three electronic charge density contour surfaces $G(a)$ of a single molecule are given for three charge density values a_1, a_2 , and a_3 , respectively, where

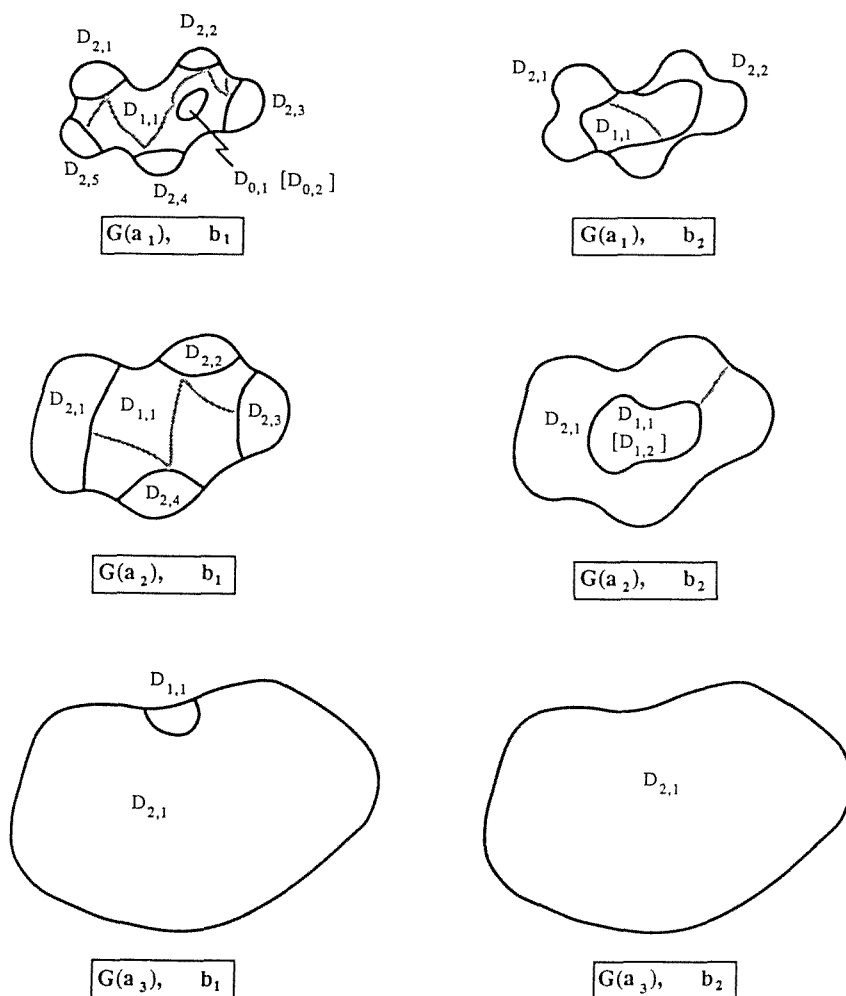


Fig. 2. Three pairs of D_{μ} domain partitionings of electronic charge density contour surfaces $G(a)$ of a molecule, for charge density values $a_1 > a_2 > a_3$, and reference curvatures $b_1 = 0$ and $b_2 > 0$. The far sides of the contour surfaces are assumed to be the mirror images of the sides displayed. The heavy lines are the boundaries of the various D_{μ} domains, whereas the grey lines are those of a minimal subdivision. The domains $D_{0,2}$ and $D_{1,2}$ in diagrams $(G(a_1), b_1)$ and $(G(a_2), b_2)$, respectively, are on the far sides, indicated by placing their symbols in square brackets []; these domains are the mirror images of domains $D_{0,1}$ and $D_{1,1}$, respectively, where the latter are displayed in these diagrams. There are no grey subdivision lines on the far sides of the contour surfaces, except the continuations of lines leading to sets $D_{0,2}$ and $D_{1,2}$ on the above two contours $(G(a_1), b_1)$ and $(G(a_2), b_2)$.

$$a_1 > a_2 > a_3, \quad (57)$$

and for two reference curvatures b_1 and b_2 , where

$$b_1 = 0 \quad \text{and} \quad b_2 > 0. \quad (58)$$

The far side of each contour surface, not shown in the figure, is assumed to be the mirror image of the near side displayed in the figure. The heavy lines are the boundaries of the various D_μ domains, whereas the grey lines are those of a minimal subdivision. Note that the number of subdivision lines used is the minimum needed to subdivide each multiply connected D_μ set into simply connected subsets, and the number of points where these lines join the boundary lines of the original D_μ sets is also the minimum. In two of these diagrams, in $(G(a_1), b_1)$ and $(G(a_2), b_2)$, the domains denoted by $D_{0,2}$ and $D_{1,2}$, respectively, are on the far side of their contour surfaces, a fact indicated by placing their symbols in square brackets []. Domains $D_{0,2}$ and $D_{1,2}$ are the mirror images of domains $D_{0,1}$ and $D_{1,1}$, respectively, where the latter are displayed in these diagrams. There are no grey subdivision lines on the far sides of the contour surfaces, except the continuations of the two lines leading to sets $D_{0,2}$ and $D_{1,2}$, on the above two contours $(G(a_1), b_1)$ and $(G(a_2), b_2)$, respectively. As the parameter value a of the charge density contour surface $G(a)$ decreases from a_1 to a_3 , less and less shape details are distinguishable, as is indicated by the decompositions of these $G(a)$ surface into domains. The two decompositions of any one of these contour surfaces for the two reference curvature parameters b_1 and b_2 , respectively, are clearly distinguishable, since they lead to different domains and different interrelations among them.

In order to proceed with the proposed shape characterization, the $G(a, 2)$ truncated contour surfaces are generated from the $G(a)$ surfaces by the elimination of all the $D_{2,i}$ domains for all serial indices i . The Betti numbers b_p (the ranks) of the resulting shape groups $H^p(\nu, a, b)$ for truncation index $\nu = 2$ and for each of dimensions $p = 0, 1$, and 2 , are listed in table 1. In fig. 3, the corresponding incidence graphs $g_{\eta,i}$ are shown, which contain, as subgraphs, the incidence graphs $g_{\eta,i}(p-1)$ ($p = 1, 2$). In fig. 4, the corresponding shape graphs $g_{\mu,i}$ are given.

The above shape descriptors provide a concise, yet sufficiently detailed characterization of the shape of the electronic charge density of the molecule at the selected contour values a_1, a_2 , and a_3 , with respect to reference curvatures b_1 and b_2 . The determination of these groups and graphs does not require visual inspection, since all their properties can be derived from the contour surfaces, generated by the computer. The method may serve as a basis for a non-visual, computer-based analysis of molecular similarity, applicable to large families of molecules.

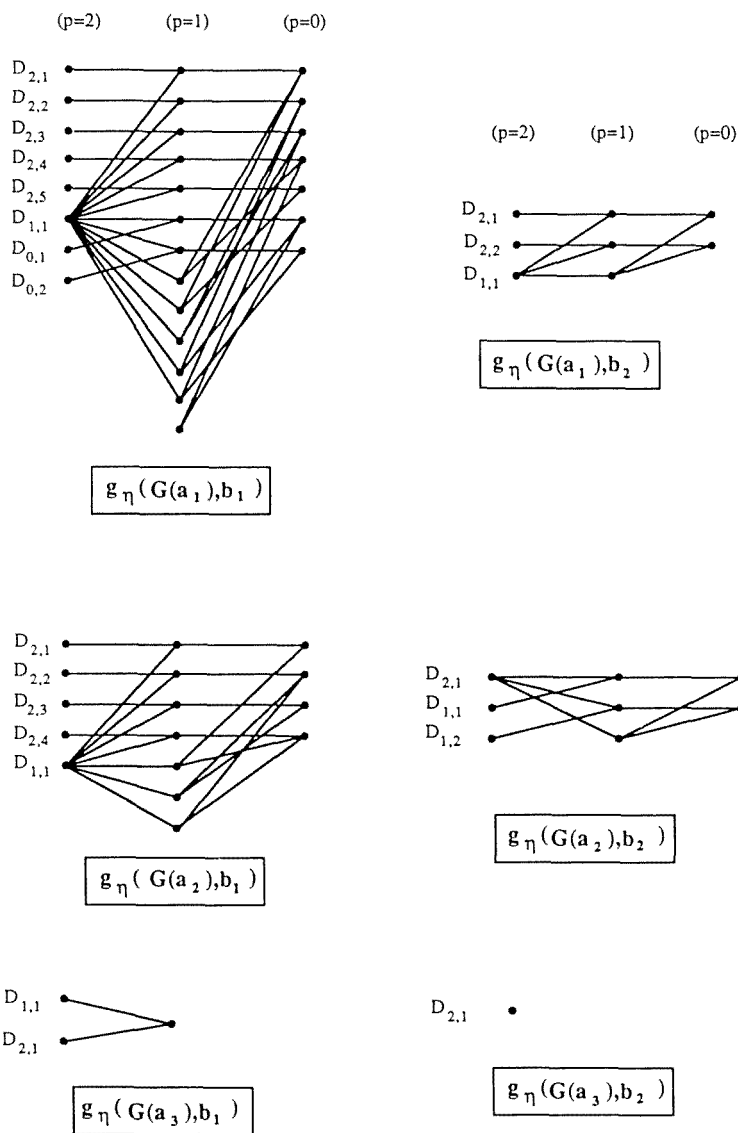


Fig. 3. Incidence graphs $g_{\eta,i}$ of the example shown in fig. 2. The incidence graphs $g_{\eta,i}(p-1)$ ($p=1, 2$) are the subgraphs of the graphs displayed. Note that all these graphs are bipartite. For a discussion of other special properties, see the text.

Table 1

The Betti numbers b_p of the shape groups $H^p(\nu, a, b)$ of the charge density contour surfaces $G(a)$ of the example of fig. 2, given for truncation index $\nu = 2$ and for each of dimensions $p = 0, 1$ and 2.

Shape group	$b_0 (p = 0)$	$b_1 (p = 0)$	$b_2 (p = 2)$
$H^p(2, a_1, b_1)$	1	4	0
$H^p(2, a_1, b_2)$	1	1	0
$H^p(2, a_2, b_1)$	1	3	0
$H^p(2, a_2, b_2)$	2	0	0
$H^p(2, a_3, b_1)$	1	0	0
$H^p(2, a_3, b_2)$	–	–	–

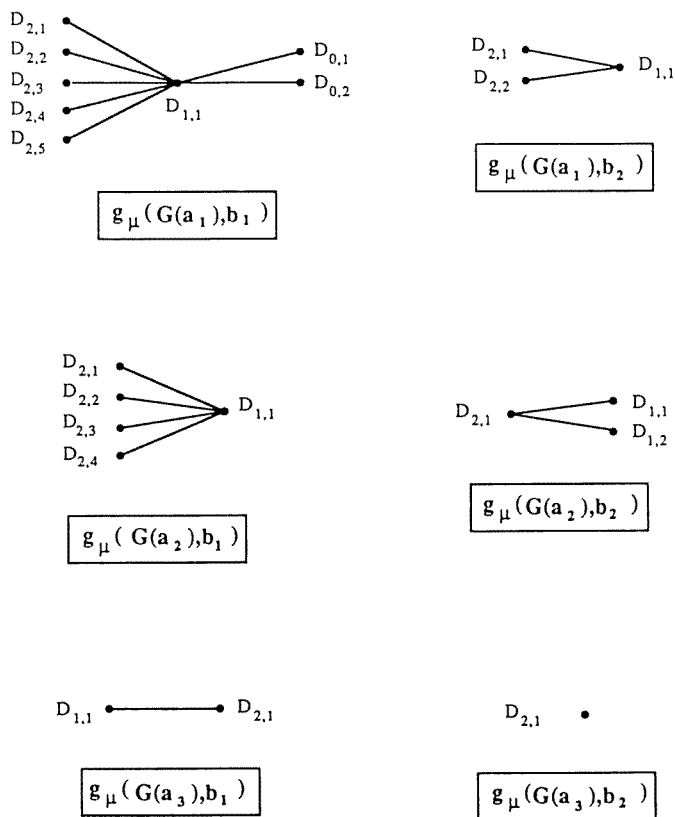


Fig. 4. Shape graphs $g_{\mu,i}$ of the example shown in fig. 2.

Acknowledgement

This study has been supported by a research grant from the Natural Sciences and Engineering Research Council of Canada.

References

- [1] P. Coppens and M.B. Hall, eds., *Electron Distributions and the Chemical Bond* (Plenum, New York and London, 1982).
- [2] W.G. Richards, *Quantum Pharmacology*, 2nd Ed. (Butterworth, London, 1983).
- [3] R. Rein, J.R. Rabinowitz and T.J. Swissler, *J. Theor. Biol.* 34(1972)215.
- [4] E. Scrocco and J. Tomasi, *Topics Current Chem.* 42(1973)95.
- [5] D.M. Hayes and P.A. Kollman, *J. Amer. Chem. Soc.* 98(1976)3335.
- [6] P.A. Kollman, *Acc. Chem. Res.* 10(1977)365.
- [7] E. Scrocco and J. Tomasi, *Adv. Quant. Chem.* 11(1978)115.
- [8] P.A. Kollman, *J. Amer. Chem. Soc.* 100(1978)2974.
- [9] J. Tomasi, On the use of electrostatic molecular potentials in theoretical investigations on chemical reactivity, in: *Quantum Theory of Chemical Reactions*, Vol. 1, ed. R. Daudel, A. Pullman, L. Salem and A. Veillard (Reidel, Dordrecht, The Netherlands, 1979) pp. 191–228.
- [10] C.W. Thornber, *Chem. Soc. Rev.* 8(1979)563.
- [11] E.E. Hodgkin and W.G. Richards, *J. Chem. Soc., Chem. Commun.* (1980) 1342.
- [12] R. Carbó, L. Leyda and M. Arnau, *Int. J. Quant. Chem.* 17(1980)1185.
- [13] G. Náráy-Szabó, *Quant. Chem. Program Exchange* 13(1980)396.
- [14] G. Náráy-Szabó, A. Grofcsik, K. Kósa, M. Kubinyi and A. Martin, *J. Comput. Chem.* 2(1981)58.
- [15] H. Weinstein, R. Osman, J.P. Green and S. Topiol, Electrostatic potentials as descriptors of molecular reactivity, in: *Chemical Applications of Atomic and Molecular Electrostatic Potentials*, ed. P. Politzer and D.G. Truhlar (Plenum, New York, 1981) pp. 309–334.
- [16] J.J. Kaufman, P.C. Hariharan, F.L. Tobin and C. Petrongolo, Electrostatic molecular potential contour maps from ab initio calculations, in: *Chemical Applications of Atomic and Molecular Electrostatic Potentials*, ed. P. Politzer and D.G. Truhlar (Plenum, New York, 1981) pp. 335–380.
- [17] A. Pullman and B. Pullman, Electrostatic molecular potential of the nucleic acids, in: *Chemical Applications of Atomic and Molecular Electrostatic Potentials*, ed. P. Politzer and D.G. Truhlar (Plenum, New York, 1981) pp. 381–405.
- [18] A. Warshel, *Acc. Chem. Res.* 14(1981)284.
- [19] P.H. Reggio, H. Weinstein, R. Osman and S. Topiol, *Int. J. Quant. Chem., Quant. Biol. Symp.* 8(1981)373.
- [20] S. Cox and D. Williams, *J. Comput. Chem.* 2(1981)304.
- [21] Z. Gabányi, P. Surján and G. Náráy-Szabó, *Eur. J. Med. Chem.* 17(1982)307.
- [22] H. Weinstein, R. Osman, S. Topiol and C.A. Venanzi, *Pharmacochem. Libr.* 6(1983)81.
- [23] J. Angyán and G. Náráy-Szabó, *J. Theor. Biol.* 103(1983)777.
- [24] P.E. Bowen-Jenkins, D.L. Cooper and W.G. Richards, *J. Phys. Chem.* 89(1985)2195.
- [25] J.R. Rabinowitz and S.B. Little, *Int. J. Quant. Chem., Quant. Biol. Symp.* 13(1986)9.
- [26] J.C. Culberson, G.D. Purvis III, M.C. Zerner and B.A. Seiders, *Int. J. Quant. Chem., Quant. Biol. Symp.* 13(1986)267.

- [27] G. Náráy-Szabó and P.R. Surján, Computational methods for biological systems, in: *Theoretical Chemistry of Biological Systems*, ed. G. Náráy-Szabó, Studies in Physical and Theoretical Chemistry, Vol. 41 (Elsevier, Amsterdam, 1986) pp. 1 – 100.
- [28] G.D. Purvis III and C. Culberson, *Int. J. Quant. Chem., Quant. Biol. Symp.* 13(1986)261.
- [29] J. Åqvist and O. Tapia, *J. Mol. Graph* 5(1987)30.
- [30] P.G. Mezey, *Int. J. Quant. Chem., Quant. Biol. Symp.* 12(1986)113.
- [31] P.G. Mezey, *J. Comput. Chem.* 8(1987)462.
- [32] P.G. Mezey, *Int. J. Quant. Chem., Quant. Biol. Symp.* 14(1987)127.
- [33] G.A. Arteca and P.G. Mezey, *Int. J. Quant. Chem., Quant. Biol. Symp.* 14(1987)133.
- [34] E.H. Spanier, *Algebraic Topology* (McGraw-Hill, New York, 1966).
- [35] M. Greenberg, *Lectures on Algebraic Topology* (Benjamin, New York, 1967).
- [36] S.-T. Hu, *Elements of General Topology* (Holden-Day, San Francisco, 1969).
- [37] J. Vick, *Homology Theory* (Academic Press, New York, 1973).
- [38] P.G. Mezey, *Theor. Chim. Acta* 54(1980)95.
- [39] P.G. Mezey, *Theor. Chim. Acta* 63(1983)9.
- [40] P.G. Mezey, *Potential Energy Hypersurfaces* (Elsevier, Amsterdam, 1987).
- [41] F. Harary, *Graph Theory* (Addison-Wesley, Don Mills (Ont.), 1972).
- [42] P.G. Mezey, *Theor. Chim. Acta* 60(1982)409.
- [43] P.G. Mezey, *Theor. Chim. Acta* 58(1981)309.