GRAPHICAL SHAPES: SEEING GRAPHS OF CHEMICAL CURVES AND MOLECULAR SURFACES

Frank HARARY

Computing Research Laboratory and Department of Computer Science,
New Mexico State University, Las Cruces, NM 88003, USA

and

Paul G. MEZEY

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

Abstract

Interior seeing graphs of Jordan curves and two-dimensional closed surfaces are proposed for the characterization of the shapes of chemical curves and surfaces. Some of the relevant properties of seeing graphs, in particular, of seeing trees, are described. The proposed applications of seeing graphs provide a molecular shape characterization, by converting the *continuum problem* of a molecular contour surface, for example, that of an isodensity contour of molecular electronic charge distribution, into a *discrete problem* of a graph. The dependence of seeing graphs on the charge density contour value, a continuous parameter, is analysed. The family of seeing graphs occurring within the chemically accessible range of charge density contour values is proposed for a computer-based analysis of molecular similarity. The general method is illustrated with the example of a detailed study on the family of seeing graphs of the electronic charge density of the ethene molecule.

1. Introduction

In recent years, the spectacular advances of computer technology have provided a new basis and additional stimulus to the search for new applications of mathematical methods in chemistry. In addition to the more routine mathematical tools of theoretical and computational chemistry, such as linear algebra, differential equations, group theory, and optimization methods, some of the more modern branches of both continuous and discrete mathematics are also finding important applications. In this study, we shall address a chemical problem that involves both differential geometrical [1,2] and graph-theoretical [3] concepts.

One of the fundamental problems of modern theoretical chemistry and research into computer-based drug design and molecular engineering is the concise yet descriptive characterization of molecular shapes. The three-dimensional shapes of contour surfaces of electronic charge densities, electrostatic potentials, molecular orbitals, or Van der Waals surfaces can be calculated by various quantum chemical or other methods (see, e.g. refs. [4–23]), and can be displayed on a computer screen. Advanced computer graphics methods facilitate the characterization of molecular shapes as well as the recognition of molecular similarity. However, these techniques ultimately rely on visual inspection, which is both an advantage and a disadvantage. Visual inspection is a very powerful tool for grasping some fundamental aspects of shape and for guiding our intuition. However, it is not ideally suited for a tedious, systematic analysis of a large number of fine details, and is certainly not the method of choice if several thousand molecules are involved. It is desirable to develop mathematical methods and computer algorithms that may follow up on an initial, visual inspection and allow for a fully automatic computer analysis of shapes and molecular similarity without direct human intervention.

The fundamental interrelations among the constituent components of a physical system or mathematical model can often be represented by graphs [3,24]. Whereas graph theory provides an essentially discrete mathematical model, continuous features may also be represented, for example, by invoking a dependence of graphs on continuous parameters [25].

For the characterization of the shapes of continuous contour surfaces of three-dimensional molecules, it is natural to apply continuum methods, often in combination with the techniques of discrete mathematics. A family of *shape group* methods, proposed recently [26–30], is based on various decompositions of contour surfaces drawn around molecules, followed by the generation of homology groups of topological objects obtained from these contour surfaces. The resulting groups, the *molecular shape groups,* are independent of the symmetry groups of molecules, and provide a detailed characterization of the shape of both symmetric and asymmetric molecules.

Our purpose is to present an alternative characterization of the shape of continuous chemical curves and molecular surfaces, based on the concept of *seeing graphs* [24]. This characterization is just as easily representable by the computer as the one based on shape groups. Seeing graphs describe a somewhat different aspect of molecular shape, and their conceptual elegance and simplicity are important advantages in applications to the molecular problem.

2.   The seeing graph of a Jordan curve

Here, we shall briefly review the definition and fundamental properties of seeing graphs [24]. It will be shown by a simple proof that every tree is a seeing graph of some Jordan curve. In the next section, the concept of a seeing tree will

be generalized for closed two-dimensional surfaces, such as molecular contour surfaces, important in drug design and computer-based engineering. Consider a Jordan curve J of the plane ^2R, and a set of points

$$\{v_i\}_{i=1,m} \tag{1}$$

called vertices, having the following properties:

(i)   Each vertex v_i is within the interior I of J

$$v_i \in I, \quad 1 \leqslant i \leqslant m. \tag{2}$$

(ii)   Let J_i denote the subset of curve J that can be "seen" from v_i, that is, for every point $P \in J_i$ the interior of the straight line segment ℓ_p interconnecting v_i and P has no point common with J,

$$\text{int } \ell_p \cap J = \varnothing. \tag{3}$$

We set the condition that the union of all subsets J_i is the whole Jordan curve J:

$$\cup J_i = J. \tag{4}$$

(iii)   If all (v_i, v_j) vertex pairs, where v_i and v_j see each other, are connected by edges, then the resulting graph is connected.

(iv) m is the smallest integer for which conditions (i)–(iii) hold.

DEFINITION 1

A graph $G(J)$ with the above properties (i)–(iv) for its vertices and edges is called an *interior seeing graph* (or in short, *seeing graph*) of Jordan curve J.

DEFINITION 2

A tree $T(J)$ obtained from a seeing graph $G(J)$ by eliminating edges while preserving connectedness is called an *interior seeing tree* (or in short, a *seeing tree*) of Jodan curve J.

An illustration of a seeing graph $G(J)$ and a seeing tree $T(J)$ of a Jordan curve J is given in fig. 1. Note that for the given Jordan curve J of the example, condition (iii) of introducing edges for every pair of vertices that see each other cannot lead to a tree for the minimum number $m = 3$ of vertices.

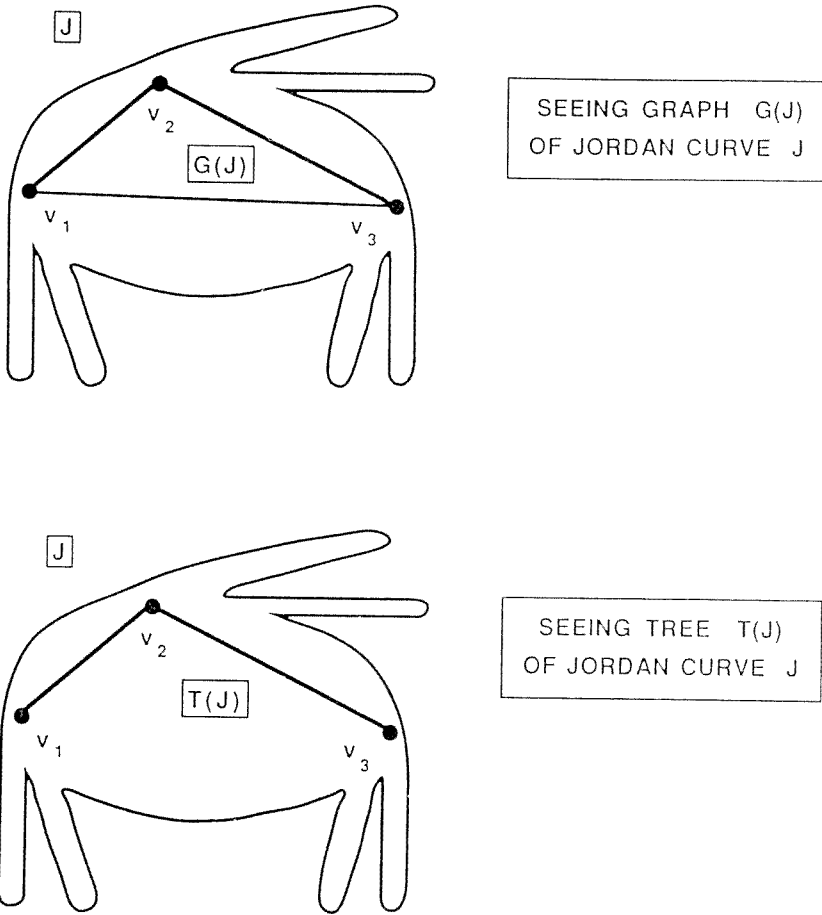We shall prove the following realization result.

Fig. 1. Illustration of the concepts of seeing graph $G(J)$ and seeing tree $T(J)$ of a Jordan curve J. Each pair of bay regions of J requires a separate vertex v_i, and for all possible locations of these vertices, they also see one another. Hence, the seeing graph $G(J)$ of this Jordan curve J (where all vertices that see one another are connected by edges) is not a tree. By eliminating any one of the three edges of $G(J)$, one obtains the seeing tree $T(J)$.

PROPOSITION 1

Each tree T is a seeing tree of some Jordan curve J.

Proof

Take any straight-line-edge representation of an arbitrary tree T in the Euclidean plane ^2E, where no two edges are collinear. The metric of ^2E is denoted by ρ. Define a distance d as

$$d = \min\{\rho(v_i, v_j) : v_i, v_j \in V(T), v_i \neq v_j\}. \tag{5}$$

Take

$$r = d/3 \tag{6}$$

and draw circles S_i with radius r around each vertex v_i:

$$S_i = \{x \in {}^2E, \rho(v_i, x) = r\}. \tag{7}$$

All these circles S_i are disjoint and each has at least one intersection point with an edge e_k of T. These points of intersections are denoted by P_{ik}. Choose a small positive number δ,

$$0 < \delta \ll r \tag{8}$$

and take the open disks D_{ik}

$$D_{ik} = \{x \in {}^2E, \rho(P_{ik}, x) < \delta\}. \tag{9}$$

For a small enough δ, all disks are disjoint.
Eliminate all points of the intersections

$$S_i \cap D_{ik} \tag{10}$$

that open up "2δ-gaps" in each circle at every edge entering the circle. For each edge e_k, define two lines parallel to e_k, at distance δ on both sides of e_k. For a small enough δ value, there are points on each circle S_i which do *not* fall between any such line pair. (This ensures that each vertex v_i of T is needed for the resulting Jordan curve for which T is a seeing tree.) For each edge e_k, retain only those points of the corresponding pair of parallel lines which fall between circles S_i and S_j, where edge e_k joins vertices v_i and v_j.

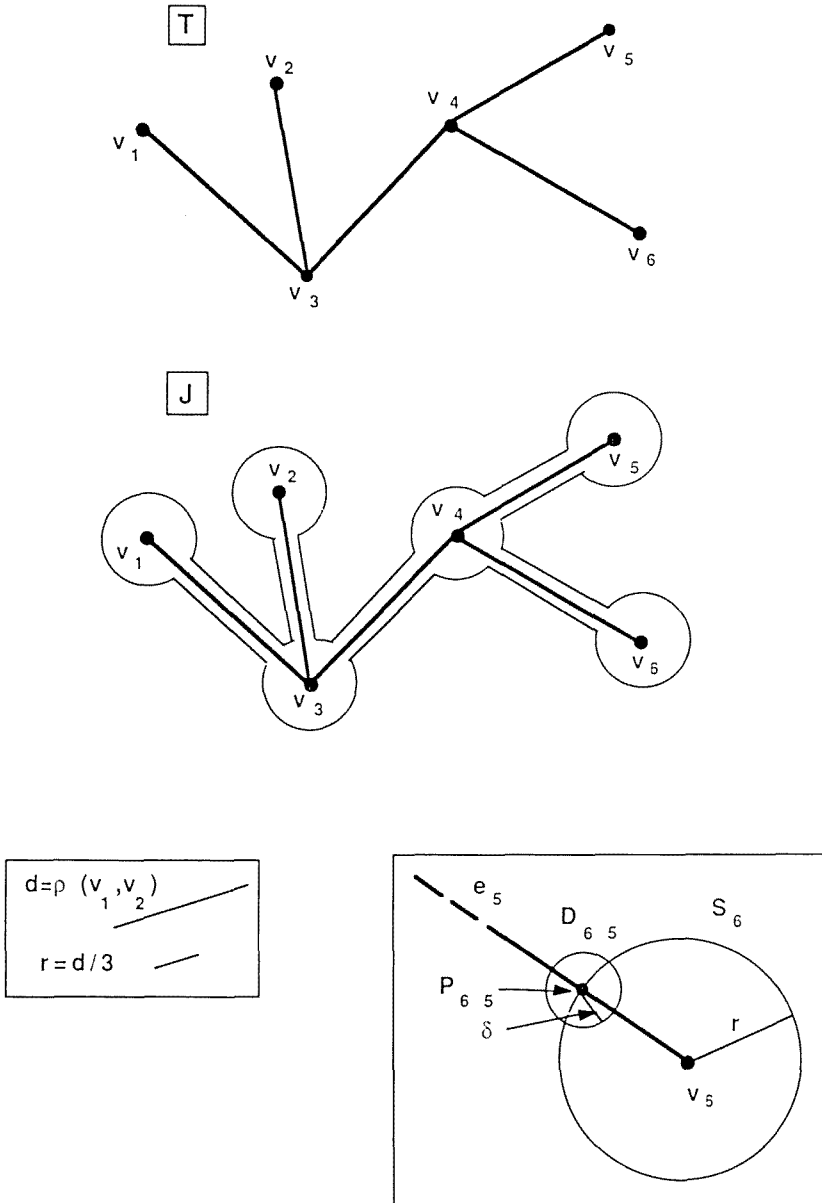CONSTRUCTION OF A JORDAN CURVE J FOR WHICH T IS A SEEING TREE

Fig. 2. Illustration of the construction of a Jordan curve *J* for an arbitrary tree *T*, represented by straight line edges. For the explanation of the notation and further details, see the proof of proposition 1 in the text.

The resulting union of the points retained from the circles and parallel lines form a Jordan curve J, for which the tree T is a seeing tree.

An illustration of the proof of proposition 1 is shown in fig. 2.

3. The seeing tree of a two-dimensional closed surface

The seeing tree of a closed surface is defined analogously to that of a Jordan curve.

Let B be a closed two-dimensional surface that divides the three-dimensional Euclidean space ^3E into two subsets: a bounded set F, referred to as the *interior* of B, and an unbounded set, referred to as the *exterior* of surface B. Consider a family of points called vertices

$$\{v_i\}_{i=1,m} \tag{11}$$

with the following properties:

(i) $v_i \in F, \quad 1 \leqslant i \leqslant m.$ \hfill (12)

(ii)   Let $B_i \subset B$ denote the set of all those points of the surface B which are seen from vertex v_i, that is, for each point

$$p \in B_i \tag{13}$$

there exists a straight line segment ℓ_p that interconnects points v_i and p, and if point p is removed from this line segment ℓ_p, then the remaining set $\ell_p - p$ is contained in F,

$$v_i, p \in \ell_p, \tag{14}$$

$$\ell_p \subset F \cup B, \tag{15}$$

and

$$\ell_p - p \subset F. \tag{16}$$

We set the condition that

$$\bigcup_{i=1}^{m} B_i = B. \tag{17}$$

(iii)   If all vertex pairs (v_i, v_j) for which v_i and v_j see each other within F are connected by edges, then the resulting graph is connected.

(iv)   *m* is the smallest integer number for which a vertex set (11) fulfills properties (i)–(iii).

DEFINITION 3

A graph $G(B)$ having properties (i)–(iv) for its vertices and edges is an *interior seeing graph* (or in short, a *seeing graph*) of closed surface B.

DEFINITION 4

If enough edges are omitted from $G(B)$ so that the resulting graph remains connected, one obtains a tree, the *interior seeing tree* (or in short, *seeing tree*) $T(B)$ of closed surface B.

By replacing the circles S_i of radius r and the parallel lines of distance 2δ with spheres ^2S_i of radii r and cylinders of radius δ, respectively, the proof of proposition 1 can be generalized to the case of two-dimensional surfaces B.

PROPOSITION 2

Any tree T is a seeing tree of some closed surface B.

The concepts of a seeing graph and a seeing tree can be generalized in many ways. For the purposes of molecular shape characterization, *exterior seeing graphs* and *exterior seeing trees* appear useful, as well as bipartite seeing graphs, where the vertex sets of the bipartition correspond to the sets of interior and exterior vertices. Such seeing graphs are advantageous for the description of molecular shape as recognized by an approaching reagent, an aspect of importance in the study of reaction mechansims [31,32]. Seeing graphs may also provide useful alternatives for studying "visibility" problems [33], important, for example, in the design of computer circuits. These and other generalizations will be discussed in a forthcoming study; in the present paper, we shall illustrate the application of interior seeing graphs and interior seeing trees.

4.   Seeing graphs of contour surfaces of molecular electron density as a tool for molecular shape characterization

The shape of an isodensity contour surface B of a molecule of a specified electronic state depends on the density value ρ of the contour and on the mutual arrangement of the nuclei. Within the Born–Oppenheimer approximation, one may give a formal definition to the classical concepts of nuclear geometry K and internal nuclear configuration space M, that is a metric space which can be represented by a manifold with boundary [32]. Consequently, the seeing graph $G(B(\rho, K))$ is also a

function of the electron density contour value ρ, as well as the nuclear configuration K. By considering all possible nuclear configurations for a fixed stoichiometry and all possible density values ρ, one may define a product space D' as

$$D' = \rho \otimes M. \tag{18}$$

(Note that the notation of space D' refers to the fact that it is a subspace of the dynamic shape space D introduced in [25].) Evidently, for a specified electronic state, the seeing graphs $G(B(\rho, K))$ are invariant within certain domains of D', that is, they remain unchanged for certain limited variations of the nuclear configuration K and contour density value ρ. Space D' is partitioned into the invariance domains $D'(G_j)$ of seeing graphs G_j which occur within D':

$$D' = \bigcup_{j=1}^{k} D'(G_j). \tag{19}$$

where

$$D'(G_j) \cap D'(G_k) = \varnothing . \tag{20}$$

The characterization problem is considerably simpler if only a single, fixed nuclear configuration $K \in M$ is studied. A set of seeing graphs G_j occurring for the given nuclear configuration K, as the electronic charge density contour value p changes from $\rho_0 = \rho_{max}$ to $\rho_m = \rho_{min} = 0$, is denoted by $G(K)$:

$$G(K) = (G_1(K), G_2(K), \ldots, G_m(K)). \tag{21}$$

where

$$G_j(K) = G(\rho, K) \quad \text{for} \quad \rho_{j-1} \leqslant \rho < \rho_j . \tag{22}$$

In the above expression, the contour density values ρ_j are critical since the seeing graphs $G(\rho, K)$ undergo changes precisely at these densities. The finite set $G(K)$ provides a shape characterization of the entire, three-dimensional electron distribution of the molecule of the specified nuclear configuration K.

5. Example: Electron density contour surfaces and seeing graphs of the ethene molecule

In fig. 3, selected electron density contour surfaces and characteristic cross-sections are shown for the C_2H_4 molecule. The data for the construction of these plots have been taken from [34,35]. The nuclear configuration K is assumed to be fixed at the equilibrium geometry. The charge density values for the contours are
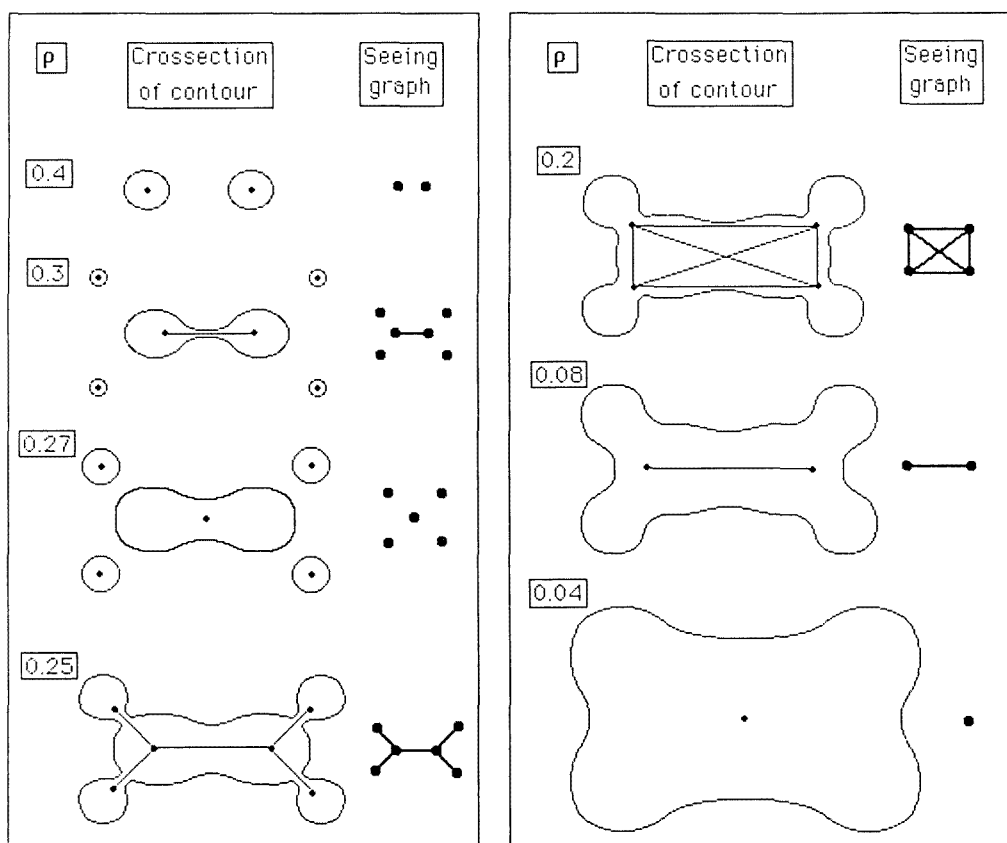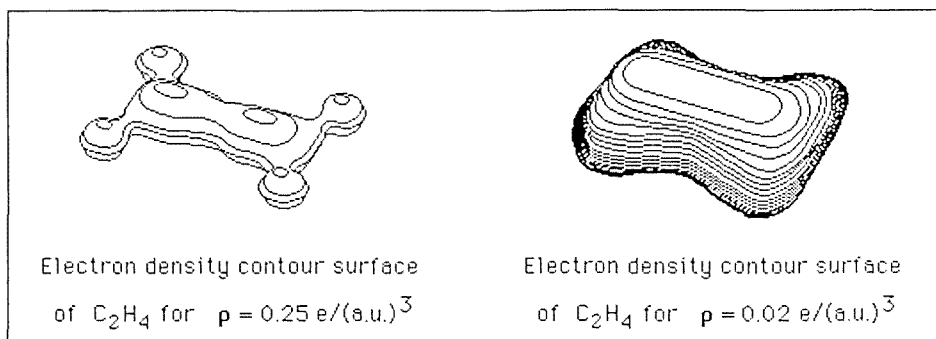
ELECTRON DENSITY CONTOUR SURFACES AND SEEING GRAPHS OF THE
C_2H_4 MOLECULE

Electron density contour surface
of C_2H_4 for $\rho = 0.25$ e/(a.u.)^3

Electron density contour surface
of C_2H_4 for $\rho = 0.02$ e/(a.u.)^3

Fig. 3. Isodensity contour surfaces and their cross-sections calculated for the ethene molecule. Contour plots and data are reproduced with permission from ref. [34]. The contour at a high electronic charge density value $\rho = 0.40$ e/(a.u.)^3 shows only two spherical components about the two carbon atoms; the corresponding seeing graph has only two isolated vertices. As the charge density contour value is gradually decreased, the contours about the four hydrogen nuclei appear and eventually merge with the contour about the carbon nuclei. In the process, more complicated seeing graphs appear as the shape of the contour surface exhibits more fine detail. However, eventually the contour surface at low electronic densities far from the nuclei become simpler and more spherical, and the seeing graph reduced to a single vertex.

ρ = 0.40, 0.30, 0.27, 0.25, 0.20, 0.08, 0.04, and 0.02 in atomic units of electronic charge over bohr^3, $e/(\text{a.u.})^3$.

For the high density contour value of 0.40, there are two, separate spherical contours about the two carbon nuclei, and the four hydrogens do not appear: the highest electron density about the hydrogens is less than this contour value. The seeing graph contains only two vertices and no edges. At the next contour value of 0.30, the four hydrogens appear as four, separated spherical contours, and the two carbon nuclei are surrounded by a single contour. Viewing the entire interior of this contour is not possible from a single location, and the seeing graph of this maximum connected component of the molecular contour has a seeing graph of two vertices, connected by an edge. This edge is the only one in the seeing graph of this molecular contour surface. At the next density value of 0.27, the molecular contour surface is still disconnected and has five components; however, the seeing graph is simpler, since a single vertex is sufficient for viewing the interior of the entire surface component surrounding the two carbon nuclei. The seeing graph has five vertices and no edges. At the density value of 0.25, the contour surface, hence also the seeing graph, is already connected. Since the parts of the contour about the hydrogens are joined to the central part through rather narrow tubes, the seeing graph for this contour density must have six vertices, roughly corresponding to the six nuclei, whereas the five edges interconnecting these vertices mimic the conventional bond pattern of the ethene molecule. At the density value of 0.20, the above tubes are much broader and four vertices are sufficient: one obtains the complete seeing graph K_4. At the lower density value of 0.08, the contour surface shows less detail, it is "swollen" compared to the contours above, and the two-vertex seeing graph K_2 is obtained. At the low density values of 0.04 and 0.02, the contour surfaces are rather "round" and the seeing graph is just a single vertex K_1.

By listing all seeing graphs obtained for a fixed nuclear configuration K, a detailed characterization of the molecular shape can be given. The concept of seeing graphs provides a discrete characterization of a continuous, three-dimensional function, describing the molecular shape. Information on seeing graphs can be easily stored in a computer, and a comparison of molecular shapes for a sequence of compounds

$$X_1, X_2, \ldots, X_p$$

can be accomplished by comparing sets

$$G_1, G_2, \ldots, G_p$$

of their respective seeing graphs. Hence, the study of molecular similarity, with reference to the three-dimensional "body" of molecules, can be formulated in elegant graph-theoretical terms.

Acknowledgements

References

[1] M.P. Do Carmo, *Differential Geometry of Curves and Surfaces* (Prentice-Hall, Englewood Cliffs, 1976).

[2] I.R. Porteous, *Topological Geometry* (Cambridge University Press, Cambridge, 1969).

[3] F. Harary, *Graph Theory* (Addison-Wesley, Reading, MA, 1969).

[4] P. Coppens and M.B. Hall, eds., *Electron Distributions and the Chemical Bond* (Plenum, New York and London, 1982).

[5] S. Fliszar, *Charge Distribution and Chemical Effects* (Springer, New York, 1983).

[6] W. Graham Richards, *Quantum Pharmacology*, 2nd ed. (Butterworth, London, 1983).

[7] R. Franke, *Theoretical Drug Design Methods*, Pharmacochemical Library, Vol. 7 (Elsevier, Amsterdam, 1987).

[8] G.M. Maggiora and R.E. Christoffersen, Quantum mechanical approaches to the study of enzymic transition states and reaction paths, in: *Transition States in Biochemical Processes*, ed. R.D. Gandour and R.L. Schowen (Plenum, New York, 1978).

[9] M.A. Johnson, V. Nicholson, M. Naim and C.-C. Tsai, in: *QSAR in Drug Design and Toxicology*, ed. D. Hadzi and B. Jerman-Blazic (Elsevier, Amsterdam, 1987).

[10] A.J. Stuper, W.E. Brügger and P.C. Jurs, *Computer-Assisted Studies of Chemical Structure and Biological Function* (Wiley, New York, 1979).

[11] Y.C. Martin, *Quantitative Drug Design: A Critical Introduction* (Dekker, New York and Basel, 1978).

[12] I. Motoc, Steric and other structural parameters for QSAR, in: *Steric Fit in Quantitative Structure-Activity Relationships*, ed. A.T. Balaban, A. Chiriac, A. Motoc and Z. Simon, Lecture Notes in Chemistry, No. 15 (Springer-Verlag, Berlin, 1980).

[13] 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, 1979) pp. 191–228.

[14] R. Carbó, L. Leyda and M. Arnau, Int. J. Quantum Chem. 17(1980)1185.

[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] G. Náray-Szabó, A. Grofcsik, K. Kósa, M. Kubinyi and A. Martin, J. Comput. Chem. 2 (1981)58.

[17] A. Warshel, Acc. Chem. Res. 14(1981)284.

[18] P.H. Reggio, H. Weinstein, R. Osman and S. Topiol, Int. J. Quant. Chem. Quant. Biol. Symp. 8(1981)373.

[19] M.L. Connolly, J. Mol. Graph 3(1985)19.

[20] J.C. Culberson, G.D. Purvis III, M.C. Zerner and B.A. Seiders, Int. J. Quant. Chem. Quant. Biol. Symp. 13(1986)267.

[21] S.D. Kahn, C.F. Pau, L.E. Overman and W.J. Hehre, J. Amer. Chem. Soc. 108(1986)7381.

[22] G. Náray-Szabó and P.R. Surján, Computational methods for biological systems, in: *Theoretical Chemistry of Biological Systems,* ed. G. Náray-Szabó, Studies in Physical and Theoretical Chemistry, Vol. 41 (Elsevier, Amsterdam, 1986) pp. 1 – 100.

[23] J. Åqvist and O. Tapia, J. Mol. Graph 5(1987)30.

[24] F. Harary, Math. Comput. Model, in press.

[25] P.G. Mezey, J. Math. Chem. 2(1988)299.

[26] P.G. Mezey, Int. J. Quant. Chem. Quant. Biol. Symp. 12(1986)113.

[27] P.G. Mezey, J. Comput. Chem. 8(1987)462.

[28] P.G. Mezey, Int. J. Quant. Chem. Quant. Biol. Symp. 14(1987)127.

[29] G.A. Arteca and P.G. Mezey, Int. J. Quant. Chem. Quant. Biol. Symp. 14(1987)133.

[30] G.A. Arteca, V. Jammal, P.G. Mezey, J.S. Yadav, M.A. Hermsmeyer and T. Gund, J. Mol. Graph 6(1988)45.

[31] F. Harary and P.G. Mezey, Discr. Appl. Math. 19(1988)205.

[32] P.G. Mezey, *Potential Energy Hypersurfaces* (Elsevier, Amsterdam, 1987).

[33] F. Luccio, S. Mazzone and C.K. Wong, Discr. Math. 64(1987)209.

[34] V.H. Smith, Jr., P.F. Price and I. Absar, Israel J. Chem. 16(1977)187.

[35] V.H. Smith, Jr., p. 3 of ref. [4].