van der Waals Surface Graphs and the Shape of Small Rings

David C. Whitley*

School of Mathematical Studies and Centre for Molecular Design, Portsmouth University, Halpern House, Hampshire Terrace, Portsmouth PO1 2QF, UK

Received July 12, 1997

A van der Waals surface graph is the graph defined on the van der Waals surface of a molecule by the intersections of the atomic van der Waals spheres. This is a discrete invariant of three-dimensional molecular shape. Two applications of these graphs to the study of small ring molecules are described: the basic shapes of rings up to order six are classified, and the results of a substructure search are analyzed, indicating how substructures with a specified three-dimensional shape may be identified.

1. INTRODUCTION

An embedding of a molecule in three-dimensional (3D) Euclidean space determines a van der Waals surface S and a *van der Waals surface graph* G defined on S by the intersections of the atomic van der Waals spheres: the faces of G are the visible parts of the atomic spheres; the edges of G are circular arcs where pairs of atomic spheres intersect on S; and the vertices of G are points where three or more atomic spheres intersect on S. van der Waals surface graphs are discrete invariants of 3D molecular shape that incorporate the topological structure of the molecular surface.[1,2] The purpose of this paper is to describe two applications of these graphs to the study of small ring molecules.

The fundamental properties of van der Waals surface graphs are established rigorously elsewhere,[2] but many of these properties are readily seen by examining any solid model of a van der Waals surface. The edges of G are either closed loops where two atoms meet without intersecting any other atoms (this is common in intersections between hydrogen and nonhydrogen atoms, for instance), or they are arcs joining vertices of G. For a *transversal* embedding[2], (that is, an embedding in 'general position'), the vertices of G are points where exactly three atoms meet on S, because if four or more atoms intersect on S, an arbitrarily small perturbation of the embedding will break this intersection into two or more intersections between triples of atoms. If three atoms meet at a vertex of G, then so do three edges of G; so, for a transversal embedding, the connected components of G are either closed loops or trivalent subgraphs of G. Transversal embeddings also have the property that the associated van der Waals surface graph is unaltered if the embedding is changed by a sufficiently small perturbation, reflecting the intuition that molecular conformations may have the same 'shape' without having identical geometrical positions.[3]

The van der Waals surface S and the graph G both depend on a choice for the atomic van der Waals radii. In the ring molecules studied here, for example, S changes from a set of disjoint atomic spheres, first into a torus and then into a sphere as the van der Waals radii increase from a sufficiently small starting value. The way in which G varies with the radii is related to the Voronoi diagram of the atom centers.[2,4] Throughout this paper a fixed set of atomic radii is assumed. The theoretical results in sections 2−5 are independent of this choice of radii, with the proviso that S is always assumed to be a topological sphere. Where explicit calculations are called for in section 6, the radii from the SYBYL[5] molecular modeling package are used.

The trivalent components of G may be thought of as building blocks of the shape of the molecule, whereas loops of G are the joints where these building blocks are glued together. This paper is concerned with describing the shape of a particular type of molecular building block, a small ring, so attention is restricted here to the situation where G is a connected trivalent graph, in which case the dual graph G^* is a triangulation of S. The faces of G are labeled by the atoms containing them, so G^* is a graph with atom-labeled vertices, which is related to the standard molecular graph G in which vertices represent atoms and edges represent bonds. For many embeddings of small molecules, each atom has a unique face on S, and each pair of bonded atoms has at least one common edge on S. In this case, there is a one-to-one correspondence between the vertices of G and G^*, and G is isomorphic to a subgraph of G^*, so that G is (isomorphic to) a *spanning* subgraph of G^* (a subgraph with the same vertex set[6]).

The assumptions made on the embedding are consolidated in the following definition: a molecular embedding ϕ is a *basic* embedding if

(i) ϕ is transversal,

(ii) S is a topological sphere,

(iii) G^* is a triangulation of S,

(iv) G is a spanning subgraph of G^*.

Conditions (*i*)−(*iii*) are quite weak, and ensure only that the embedding is that of a typical molecular building block. Condition (*iv*) is a little stronger; in particular, it implies that a basic embedding has a unique face on the van der Waals surface for each atom in the molecule.

van der Waals surfaces for two four-membered rings, cyclobutane and cyclopropylamine, are illustrated in Figure 1. Cyclobutane is shown in a puckered embedding[7] with C_2 and C_4 raised above C_1 and C_3, so that C_2 and C_4 intersect

* E-mail: whitley@sms.port.ac.uk.

S0095-2338(97)00427-7 CCC: $15.00
Published on Web 08/08/1998

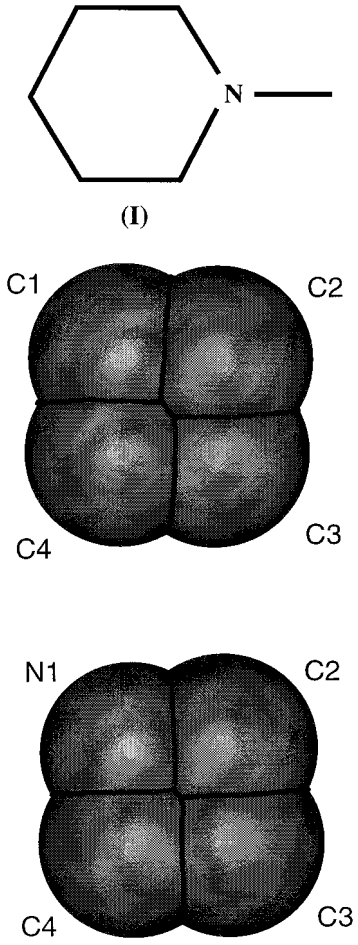

Figure 1. van der Waals surfaces for cyclobutane (top) and cyclopropylamine (bottom).

on the upper, visible part of the surface and C_1 and C_3 intersect on the lower, hidden part. Cyclopropylamine is shown in an embedding where the atom centers are almost coplanar. The smaller radius of the nitrogen atom causes C_2 and C_4 to intersect twice on the van der Waals surface (one intersection is visible in Figure 1 and the other is hidden around the back of the molecule), whereas N_1 and C_3 do not intersect on the molecular surface. The corresponding graphs are shown in Figure 2(b) for cyclobutane and Figure 2(c) for cyclopropylamine. The lower diagrams in Figure 2 show the van der Waals surface graphs G, and the upper diagrams show the dual graphs G^*. Edges of G^* representing atomic intersections that are hidden in Figure 1 are shown as dashed lines in Figure 2.

The dual graph G^* for cyclopropylamine has two edges joining the vertices representing C_2 and C_4, because C_2 and C_4 meet twice on the van der Waals surface. To distinguish embeddings that have such multiple edges from those which do not, recall that a graph is *simple* if two vertices are joined by at most one edge, and no edge joins a vertex to itself.[8] An embedding is defined as *simple* if G^* is a simple graph.

Two basic embeddings have the same *van der Waals shape* if the corresponding graphs G^* are isomorphic as labeled, spherical triangulations. An equivalence class of basic embeddings is a *basic van der Waals shape*, and an equivalence class of simple embeddings is a *simple van der Waals shape*. The basic van der Waals shapes of rings of order ≤ 6 are classified in sections 2−5.

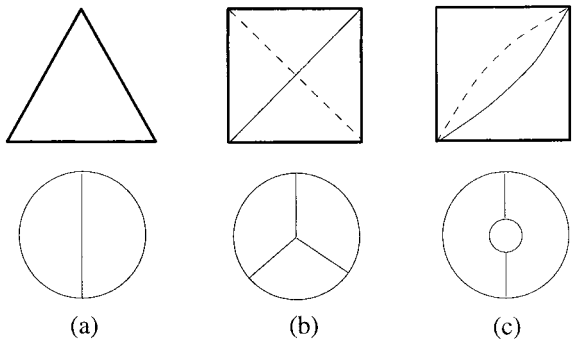

Figure 2. Basic van der Waals shapes for triangles and rectangles.

A related graph is the *van der Waals shape graph*[2] G_S, defined to have a vertex for each atom with a visible face on S, and an edge between two vertices if the corresponding atoms have adjacent faces on S. If two atoms share more than one common edge on S, the corresponding vertices are joined by only a single edge, so that G_S is a simple graph.

If each atom has at least one face on S, and each pair of bonded atoms has at least one common edge on S, then G is a spanning subgraph of G_S, and G_S is obtained from G by adding edges to G between nonbonded atoms that intersect on S. In this case G_S is a natural extension of G, adding 3D information on the embedding to the connectivity described by G. For embeddings that are both basic and simple, G_S and G^* are isomorphic.[2] In this case, the degree sequence of G_S coincides with Arteca and Mezey's sequence of *n-type faces*.[9]

In section 6, the van der Waals shape graphs for the results of a standard, two-dimensional (2D) substructure search are calculated, enabling the 3D shape of the results to be analyzed.

2. EMBEDDINGS OF CYCLIC GRAPHS

Consider the basic van der Waals shapes of molecules whose molecular graphs are isomorphic to C_n, the cyclic graph with n vertices. If ϕ is a basic embedding of C_n, then the dual G^* of the van der Waals surface graph G is a spherical triangulation containing C_n as a spanning subgraph. Let v, e, and f be the numbers of vertices, edges and faces of G, and v^*, e^*, and f^* be the corresponding numbers for G^*. Since G is trivalent and spherical we have $3v = 2e$ and $v - e + f = 2$, so for an n-cycle

$$v = f^* = 2n - 4$$

$$e = e^* = 3n - 6$$

$$f = v^* = n$$

By applying a suitable homeomorphism, we may assume that C_n is placed around a great circle of the sphere. A further $2n - 6$ edges must be added to C_n to complete the triangulation G^*, and $n - 3$ of these edges lie in each of the hemispheres determined by C_n. Thus, the classification of the basic embeddings of C_n reduces to finding the triangulations of an n-gon obtained by adding $n - 3$ edges, then enumerating the spherical triangulations with such an n-gon triangulation in each hemisphere.

The cases $n = 3$ and $n = 4$ are illustrated in Figure 2, where the bottom row shows the van der Waals surface

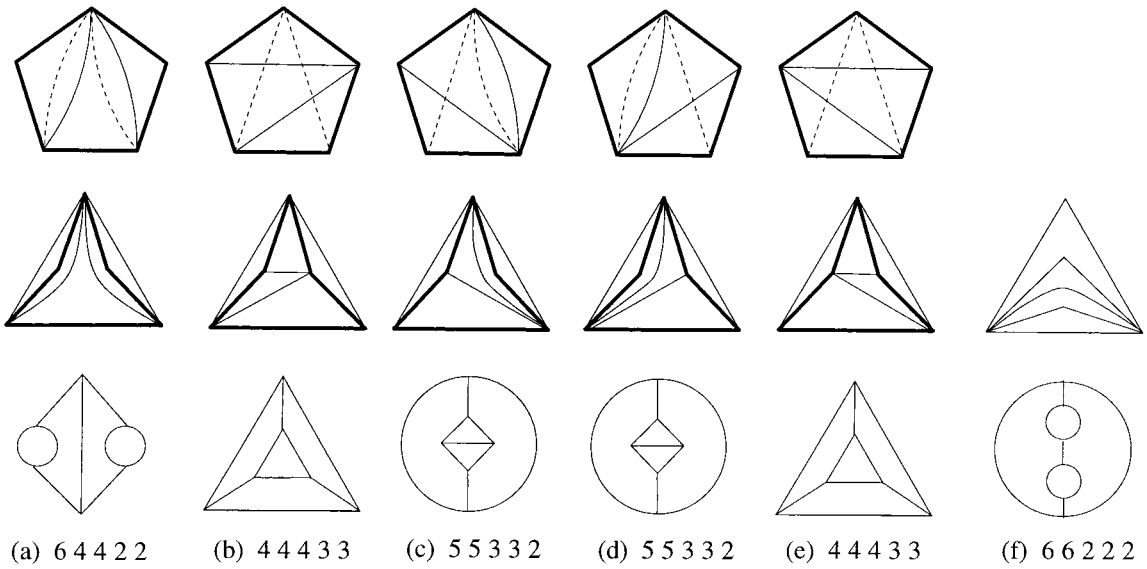

Figure 3. Basic van der Waals shapes for pentagons.

graphs G, and the top row shows G^* with the sphere viewed from the North pole, the subgraph C_n in bold lines around the equator, and edges in the Southern hemisphere in dashed lines. For triangular rings there is a single basic van der Waals shape [Figure 2(a)], in which G divides S into three segments and G^* is itself a triangle. In any triangular ring (cyclopropane, for instance) the atom centers are coplanar, so this shape represents a 'flat' embedding. For rectangles there are two basic van der Waals shapes, one in which each atom intersects every other on S [Figure 2(b)], and one in which two of the atoms do not meet on S [Figure 2(c)]. These are the shapes obtained for the embeddings of cyclobutane and cyclopropylamine in Figure 1.

The cases $n = 5$ and $n = 6$ (the most interesting from a chemical point of view) display a greater variety of shapes, and the classification of these is described in detail later.

3. PENTAGONAL RINGS

Theorem 1. *There are five basic van der Waals shapes for pentagonal rings, including two pairs of enantiomeric van der Waals shapes, one of which is a pair of simple van der Waals shapes.*

Proof. There is only one way to triangulate a pentagon by adding two extra edges: if the vertices of the pentagon are labeled cyclically v_1, ..., v_5, join v_1 to v_3 and v_3 to v_5. Any other triangulation of the pentagon is obtained from this one by cyclically relabeling the vertices. Now suppose that the pentagon is placed around the equator of a sphere, with the vertices equidistant, and with this triangulation on the Southern hemisphere. All possible dual graphs G^* are found by placing the same triangulation on the Northern hemisphere, rotated through angles of $2n\pi/5$, $n = 0$, ..., 4. The five graphs G^* obtained in this way are shown in the first row of Figure 3, using the same conventions as Figure 2. Each graph represents a different van der Waals shape. The second row of Figure 3 shows the same graphs stereographically projected from the South pole onto a plane tangent to the North pole, and the the third row shows the corresponding van der Waals surface graphs G, without the face labels. Each column is labeled by the degree sequence of G^*. The graph G^* in column (a) is symmetric, and represents the same van der Waals shape as its mirror image. The remaining graphs G^* represent two pairs of enantiomeric van der Waals shapes: $G^*(b)$ is the mirror image of $G^*(e)$ and $G^*(c)$ is the mirror image of $G^*(d)$. Only $G^*(b)$ and $G^*(e)$ are simple graphs.

Table 1. Shapes of Common Pentagonal Molecules

shape	molecule
a	1-oxacyclopenta-3-ene
	1-thiacyclopenta-3-ene
	pyrrole
	furan
	thiophene
	1-2-cyclopentadiene
	2*H*-pyrrole
	2*H*-furan
	2*H*-thiophene
b/*e*	cyclopentane
	pyrrolidine
	tetrahyrofuran
	tetrahydrophiophene
	cyclopentene
	pyrroline
	1-azacyclopenta-2-ene
	1-thiacyclopenta-2-ene
c/*d*	1-3-cyclopentadiene
	1-oxacyclopenta-2-ene

The graphs G^* in Figure 3 arise from three trivalent, spherical graphs with $v = 6$, $e = 9$, and $f = 5$, as can be seen from the third row of the figure. There is a fourth graph with these properties, with degree sequence 66222, shown in column (f). However, the dual of this graph, shown in the second row of column (f), contains no embedding of C_5, and therefore does not occur as the van der Waals surface graph of a basic embedding of a pentagonal ring.

To establish which of the graphs in Figure 3 are produced by common pentagonal molecules, the adjacency matrix $A(G^*)$ was calculated for a number of embeddings using an algorithm described elsewhere.[2] The embeddings were either taken from the SYBYL fragment library,[5] or the molecules were built in SYBYL and the embeddings determined by energy minimization, using the default settings in the SYBYL energy minimization routines. The adjacency matrices do

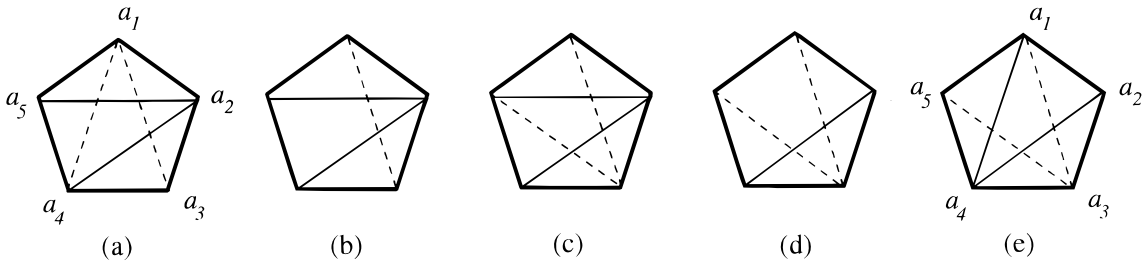

Figure 4. Pseudo-rotation of cyclopentane.

not distinguish between enantiomeric van der Waals shapes: the graphs $G^*(b)$ and $G^*(e)$, for instance, have similar adjacency matrices A. The five pentagonal van der Waals shapes therefore lead to three dissimilar adjacency matrices, one for the symmetric $G^*(a)$, and one for each of the mirror image pairs. The graphs produced by embeddings of 19 molecules are shown in Table 1. Shape (a) is obtained for essentially flat embeddings where the atoms are almost coplanar; shape (b/e) was obtained for molecules in the textbook 'half-chair' (cyclopentane) and 'envelope' (cyclopentene) positions; and shape (c/d) was obtained for just two of the molecules in this sample: 1-3-cyclopentadiene, which is almost flat, and 1-oxacyclopenta-2-ene, which is a slightly twisted envelope.

4. PSEUDO-ROTATION OF CYCLOPENTANE

The pseudo-rotation of cyclopentane[7] can be understood in terms of the different labelings of the half-chair shape graph in Figure 3(b). Suppose that the carbon atoms are labeled cyclically from a_1 to a_5 around the ring, and that the initial embedding has a_1 below and a_2 above the plane defined by a_3, a_4, and a_5. This situation corresponds to the labeling of the half-chair shape graph shown in Figure 4(a). Now suppose that a_1 moves upwards into the plane of a_3, a_4, and a_5, producing an embedding with a_2 above the plane of the other atoms. If a_1, a_3, a_4, and a_5 are symmetrically placed, so that the embedding is invariant under a reflection in the plane passing through a_2 and bisecting the lines joining atoms a_3 and a_1, and atoms a_4 and a_5, this is the textbook 'envelope' position. The envelope is a nontransversal embedding, because the symmetry forces a point of quadruple intersection between the coplanar atoms on the van der Waals surface, producing a quadrilateral face in the dual graph, as shown in Figure 4(b).

Continuing around the path of the pseudo-rotation, a_3 moves downwards to produce a half-chair with a_2 above and a_3 below the plane of atoms a_1, a_4, and a_5 (Figure 4(c)). Next, a_2 moves downwards to produce an envelope with a_1, a_2, a_4, and a_5 coplanar and a_3 below this plane [Figure 4(d)]. Atom a_4 then moves upwards to give a half-chair embedding with a_4 above and a_3 below the plane of a_1, a_2, and a_5 [Figure 4(e)]. Note that the (unlabeled) graph in Figure 4(e) is the (unlabeled) graph in Figure 4(a) rotated clockwise through an angle of $4\pi/5$. The pseudo-rotation continues in this way, passing through 10 envelopes that we may label 2a, 3b, 4a, 5b, 1a, 2b, 3a, 4b, 5a, and 1b, where 1a denotes the envelope with a_1 above the plane of the other atoms, 2b denotes the envelope with a_2 below the plane of the other atoms, etc. Between each envelope embedding, the pseudo-rotation visits 10 half-chair embeddings, corresponding to the 10 different cyclic labelings of the half-chair shape graph in Figure 3.

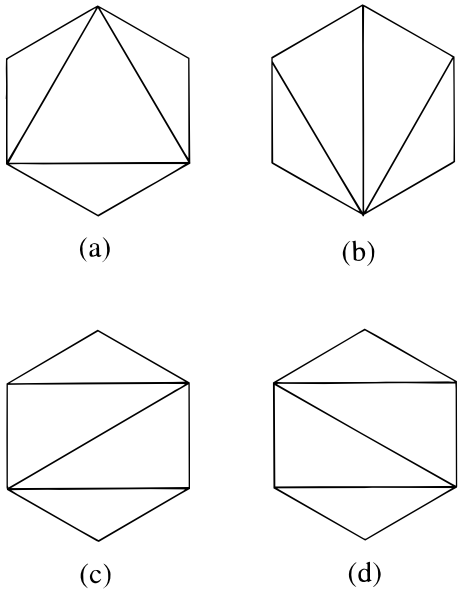

Figure 5. Triangulations of a hexagon.

Alternate half-chairs around the pseudo-rotation are enantiomeric embeddings, whereas every second half-chair embedding has the same (unlabeled) shape graph, rotated through $4\pi/5$.

5. HEXAGONAL RINGS

Theorem 2. *There are 26 basic van der Waals shapes for hexagonal rings, including nine enantiomeric pairs. There are 10 simple, basic van der Waals shapes for hexagonal rings, including four enantiomeric pairs.*

Proof. There are four ways to triangulate a hexagon (Figure 5). The triangulations in Figure 5(c) and Figure 5(d) are mirror images. The dual graphs G^* for a basic embedding of $G = C_6$ are found by taking all possible pairs of these triangulations, drawing one triangulation on the Northern hemisphere and the other on the Southern hemisphere, with the hexagon around the equator, and then rotating the triangulations through angles of $2n\pi/6$, $n = 0$, ..., 5 relative to one another. The $2^4.6 = 96$ possible arrangements produce 26 different graphs G^*. Eight of these graphs are symmetric and the rest form nine enantiomeric pairs. Ten of the graphs are simple, and of these, two are symmetric whereas the others form four enantiomeric pairs. The simple graphs G^* are shown in Figure 6, and the nonsimple ones in Figure 7. For each basic van der Waals shape, these figures show the van der Waals surface graph dual G^* above a planar projection of G^*, labeled by the degree sequence of G^*. The degree sequences of the enantiomeric graphs are marked with an asterisk; only one member of

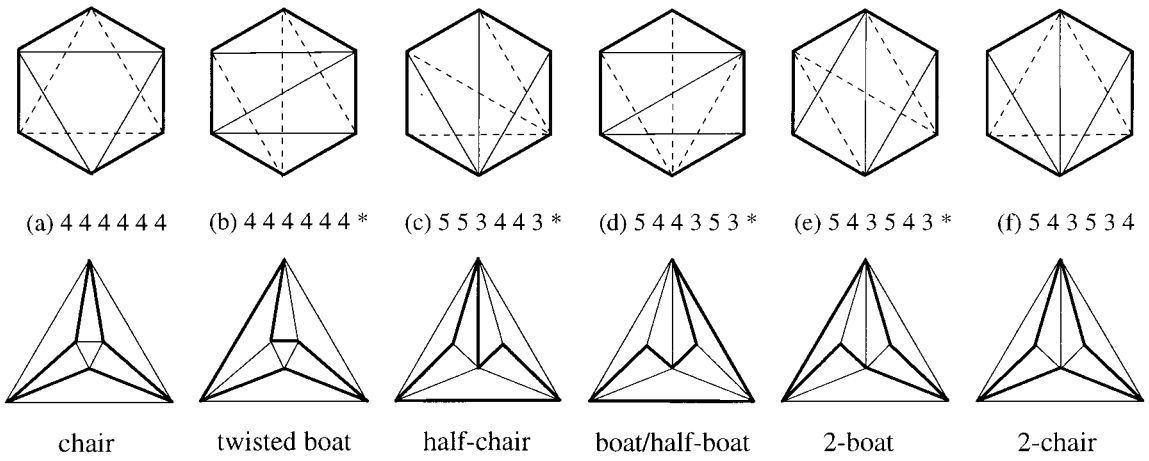

Figure 6. Basic, simple van der Waals shapes for hexagons.

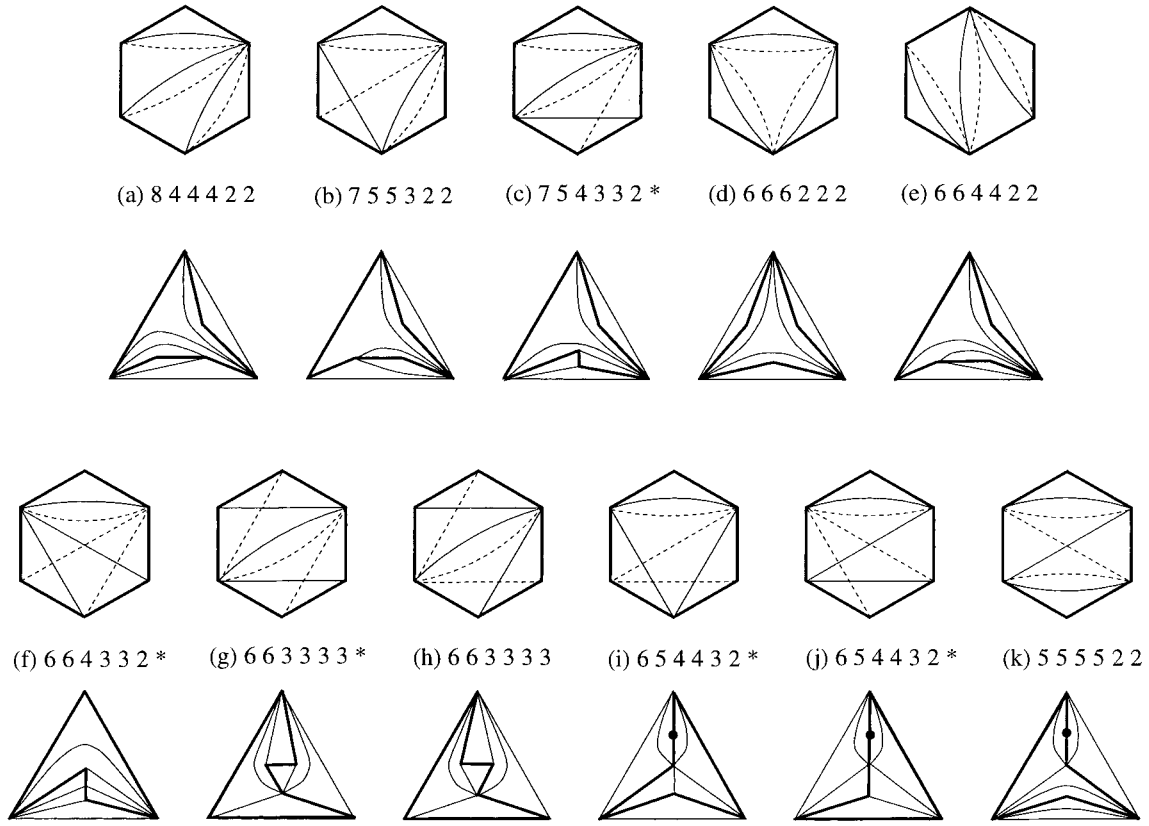

Figure 7. Basic, non-simple van der Waals shapes for hexagons.

each enantiomeric pair is shown. The graphs are drawn using the same conventions as the pentagonal case in Figure 3.

An alternative way to arrive at this result is to note that a van der Waals surface graph for a basic embedding of C_6 is a trivalent, spherical graph with $v = 8$, $e = 12$, and $f = 6$. The 14 graphs with these properties are drawn in Figure 8, labeled by the degree sequences of their duals. The dual graphs with degree sequences 882222, 864222, and 773322 contain no 6-cycle, so do not represent basic van der Waals shapes. The duals of the final two graphs in the figure, with degree sequences 444444 and 554433, are simple graphs containing, respectively, three and seven embeddings of C_6, as indicated by Figure 6. The remaining dual graphs are not simple graphs. The dual graphs with degree sequences 754332 and 664332 each contain two embeddings of C_6, the dual graph with degree sequence 663333 has three embeddings of C_6, the dual graph with degree sequence 654432 has four embeddings of C_6, and the duals of the remaining graphs in Figure 8 each contains a unique 6-cycle and gives a single basic van der Waals shape.

The van der Waals shapes in Theorem 2 that arise from dual graphs G^* with a unique embedding of C_6 can be distinguished by identifying the underlying graph G^* as an abstract graph. The degree sequence of G^* is sufficient to discriminate between these shapes. To distinguish between shapes with the same G^* requires information on the embedding of C_6 in G^*. This embedding defines a cyclic ordering $(v_1 \ldots v_6)$ of the vertices of G^*, and hence a cyclic ordering of the degree sequence $[deg(v_1) \ldots deg(v_6)]$. To remove the dependence on the labeling of the vertices of C_6, define the *degree cycle* of G^* to be $max\{(deg(\alpha v_1) \ldots deg(\alpha v_6)): \alpha \in Aut(C_6)\}$, using the lexicographic ordering

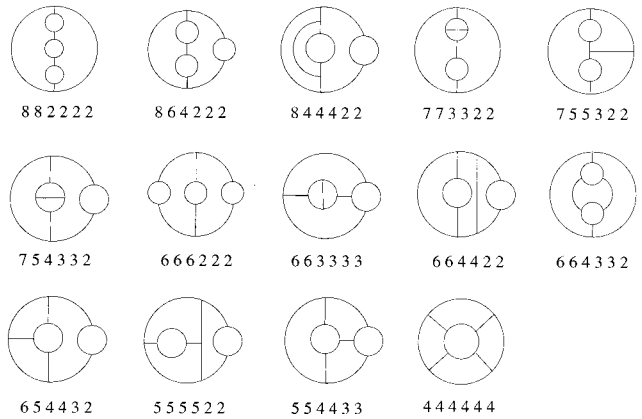

Figure 8. Trivalent, spherical graphs with eight vertices.

Table 2. Shapes of Common Hexagonal Molecules

shape	molecule
6a	cyclohexane (chair)
6b	cyclohexane (twisted boat)
6c	cyclohexene (half-chair)
6d	1,2-cyclohexadiene (half-boat)
6f	1,3-dioxane (chair)
7a	1,2,4-cyclohexatriene (flat)
7d	1,3,5-cyclohexatriene (flat)
7e	1,4-cyclohexadiene (flat)
7k	4H-pyran (flat)

for the degree sequences. In other words, the degree cycle is the largest degree sequence obtained by rotating and reflecting the vertex labels. The four shapes with degree sequence 554433 in Figure 6, for example, have degree cycles 553443, 544353, 543543, and 543534. The degree cycle also discriminates between the two shapes with degree sequence 654432 in Figure 7, but not between the shapes where G^* has degree sequences 444444 and 663333.

The adjacency matrix of G^* can also be used to identify equivalent van der Waals shapes. If $A(v)$ is the adjacency matrix of G^* with rows and columns ordered by $v = (v_1 \ldots v_6)$, then two adjacency matrices A_1 and A_2 represent the same van der Waals shape if $A_1(v) = A_2(\alpha v)$ for some $\alpha \in Aut(C_6)$. In this way the adjacency matrices of G^* separate one of the embeddings of C_6 in the cubic graph from the other two, but the shapes with degree sequence 663333 also have identical adjacency matrices. In this last case, a full description of G^* as a surface graph[10,11] and of the embedding of C_6 in G^* are necessary to identify the shapes. A similar analysis is necessary to discriminate between the two shapes in each enantiomeric pair.

Because graphs G^* with equivalent adjacency matrices have identical degree cycles, and graphs with identical degree cycles have the same degree sequences, grouping embeddings of C_6 by degree sequence, degree cycle, adjacency matrix, and van der Waals shape gives a sequence of increasingly refined classifications of the shape of the embeddings. These invariants may therefore be used as an increasingly stringent set of screens for the shape of an embedding.

The shapes of some embeddings of common hexagonal molecules are listed in Table 2. The embeddings were minimum energy structures produced by SYBYL,[5] and the shapes in the table are numbered by their labels in Figures 6 and 7. The named shapes in Figure 6 were identified from the results in Table 2 and the following section. The nonsimple shapes in Table 2 were all produced by flat embeddings with the atoms almost coplanar.

Table 3. Shape Classes for *N*-Methylpiperidine Embeddings

classifier	total	symmetric	enantiomeric
shape graph	43	5	38
ordered degree sequence	34	4	30
degree sequence	8	–	–
spectrum	9	–	–

6. A SUBSTRUCTURE SEARCH

A potential application of van der Waals surface graphs lies in 3D database searching, and in this section the results of a substructure search for a small hexagonal ring compound, *N*-methylpiperidine (**1**), are classified using van der Waals shape graphs. A 2D substructure search for *N*-methylpiperidine was carried out using the Cambridge Structural Database[12] at the EPSERC's Chemical Database Service.[13] This search produced 865 hits from 160 000 database entries. The adjacency matrices of the van der Waals shape graphs of these embeddings were calculated, and classified according to the invariants shown in Table 3. (Note that these results differ from those reported for a similar exercise,[1] because a more restrictive definition of (**1**) was used here for the substructure search, eliminating many flat embeddings.)

If α is the automorphism of (**1**) given by reflection in the horizontal line through its center, and ϕ' denotes the mirror image of an embedding ϕ, then $\alpha G_S(\phi) = G_S(\phi')$. Given this situation, we call a shape graph $G_S(\phi)$ symmetric if $\alpha G_S(\phi) = G_S(\phi)$ and enantiomeric if $\alpha G_S(\phi) \neq G_S(\phi)$. The 865 embeddings from the substructure search produce 43 shape graphs: five symmetric graphs and 19 enantiomeric pairs. Coarser classifications are obtained by considering the degree sequences and spectra of the shape graphs. These group the embeddings into only eight and nine classes, respectively. In the case of the spectra, the leading eigenvalue was sufficient to distinguish the eight classes. Neither the spectrum nor the degree sequence will detect enantiomeric graphs, but if the degrees of the shape graph vertices are ordered to reflect the cyclic ordering around the ring in **1**, then the resulting *ordered degree sequences* do discriminate between symmetric and enantiomeric graphs. The ordered degree sequences define 34 shape classes: four symmetric and 15 enantiomeric pairs. These results are summarized in Table 3.

Grouping the embeddings by degree sequence, ordered degree sequence, and shape graph produces the refinement of shape classes shown in Table 4. Each row in Table 4 represents a collection of embeddings with the same shape graph, identified by the reference code in the Cambridge Structural Database of a representative molecule containing one of these embeddings. The number of embeddings in each group is shown in parentheses. Several molecules found by the substructure search contain multiple copies of **1**. This happened for two molecules in the table (ABBUMO10 and BODBEB) and labels *a* and *b* are appended to their reference codes to distinguish the resulting shape graphs. A horizontal arrow in the table indicates that the entry to the left is carried over into this entry unchanged. Only four of the shape graphs represent basic embeddings,

Table 4. Groups of Shape Classes for *N*-Methylpiperidine Embeddings

shape graph	ring shapes	ordered degree sequence	degree sequence
ABBUMO10a (597) eq. chair	*a*	ABBUMO10a (642)	ABBUMO10a (712)
BAHCOC (29) eq. twisted boat	*b*		
BUXVEV (16) eq. twisted 2-boat	*b*		
AJMALN10† (67) ax. twisted boat	*b, d*	AJMALN10 (70)	
BODBEBa† (3) ax. chair	*a*		
ABBUMO10b (5) eq. 2-boat	*e*	→	ABBUMO10b (96)
BEWDOW10† (58) ax. boat	*b, d, e*	→	
BOGRIY (24) eq. boat	*d*	→	
BOVCIY (5) eq. 2-chair	*f*	→	
CAVFEK (1) eq. boat	*e*	→	
HEGRAH (3) eq. boat	*d*	→	
BICSIP (18) ax. chair	*a*	→	→
BIVZOV (4) eq. half-chair	*c*	→	BIVZOV (15)
DURYAQ† (1) ax. 2-boat	*d*	→	
KONVAK (10) ax. 2-boat	*d*	→	
BODBEBb (1) ax. twisted boat	*b*	BODBEB (16)	→
DAPLAC10 (12) ax. chair	*a*		
MNICAM10 (3) ax. twisted boat	*b*		
BUZJIP (2) ax. twisted boat	*d*	→	BUZJIP (3)
DBVOBT10 (1) eq. chair	*f*	→	
DUCTIE (1) ax. 2-boat	*d*	→	DUCTIE (3)
JAHYEW (1) ax. 2-chair	*f*	→	
NEWALK10 (1) ax. 2-boat	*e*	→	
HACLIG (2) eq. half-boat	*c*	→	→

and these are labeled with a †. In the nonbasic embeddings the nitrogen atom has two faces on the van der Waals surface, on the 'top' and 'bottom' of the piperidine ring. The listed representative of each group was examined, and a visual description of the embedding is given in the third column. The embeddings described as *2-boats* are boats where the 'prow' and 'stern' are carbon atoms, and the *2-chairs* are chairs where the 'head' and 'foot' are carbon atoms. In both cases, the methyl group is attached to the nitrogen occupying the '2' position (choosing a natural ordering for the ring starting with '1' at an apex).

It is clear from Table 4 that the degree sequences provide too coarse a grouping of the shape graphs, identifying only eight groups, and mixing together chairs and boats. This result is not unexpected, as the raw degree sequence ignores the atom ordering, and hence the molecular structure. The ordered degree sequences take account of the molecular structure, recording the number of intersections each atom makes on the van der Waals surface, but ignoring the identity of intersecting atoms. The grouping of embeddings by ordered degree sequence is very similar to that given by the shape graph itself: there is some mixing of chairs and boats, but the axial/equatorial distinction is maintained.

The second column in Table 4 contains a list of the shape graphs of the piperidine ring present in each group. Only simple, basic shape graphs are obtained for the rings in this sample, and they are labeled as in Figure 6. To compare the shape graph classification with more familiar geometrical descriptions, the torsion angles and ring puckering coordinates for the piperidine rings were retrieved from the Cambridge Structural Database, and the results of analyzing these are shown in Table 5 and Figures 9 and 10.

To compare the torsion angles of two embeddings with the same van der Waals shape requires the automorphism θ providing the equivalence between the corresponding shape graphs G_1^* and G_2^*. If the vertices v' adjacent to each $v \in G_1^*$ are ordered cyclically around v, using a fixed orientation on S, then θ relabels these vertex cycles to match those of G_2^*. To identify equivalent shape graphs, the vertices were labeled 1, ..., 6, and the vertex cycles written as linear integer sequences, using the largest linear representative of the cycle (in the lexicographic sense). These vertex lists

Table 5. Torsion Angles for Piperidine Rings

shape	parameter	1	2	3	4	5	6
a(630)	max	79.14	−23.43	74.22	−21.92	82.31	−21.93
	min	24.02	−72.96	22.97	−79.44	22.30	−79.17
	mean	57.08	−54.02	53.76	−56.45	59.44	−59.83
	sdev	6.50	5.75	6.73	7.39	6.40	6.64
b(81)	max	59.53	−52.64	76.81	57.56	−52.10	56.71
	min	8.95	−85.33	−0.28	−4.93	−77.31	4.55
	mean	34.22	−67.79	31.88	33.13	−67.99	31.46
	sdev	17.02	5.13	18.64	17.90	4.21	17.59
b′(40)	max	84.20	−6.51	−3.19	86.35	−0.60	−1.07
	min	54.28	−66.69	−62.32	56.83	−68.16	−70.35
	mean	69.12	−38.15	−27.24	68.99	−38.66	−27.81
	sdev	6.18	17.89	17.99	7.12	18.27	18.52
c(4)	max	77.74	−44.72	21.85	6.06	25.21	−42.88
	min	70.47	−56.33	13.04	−4.32	7.41	−60.03
	mean	72.86	−52.21	16.33	0.01	17.58	−52.38
	sdev	3.30	5.27	4.11	4.74	7.52	7.77
c′(2)	max	59.79	−0.86	15.35	26.25	40.84	−63.85
	min	56.46	−35.84	−42.34	−17.31	28.57	−72.72
	mean	58.13	−18.35	−13.50	4.47	34.71	−68.29
	sdev	2.36	24.74	40.80	30.81	8.68	6.27
d(50)	max	68.34	12.36	−15.40	60.93	37.35	−49.60
	min	33.71	−6.24	−66.38	−2.89	−0.75	−76.07
	mean	57.04	3.09	−59.02	51.71	8.64	−63.71
	sdev	5.23	5.03	8.31	9.65	6.77	4.45
d′(43)	max	76.93	2.00	−38.42	66.95	10.44	−50.35
	min	56.53	−23.69	−61.31	38.96	−15.61	−65.70
	mean	63.84	−8.00	−52.38	59.51	−3.34	−57.18
	sdev	4.58	5.94	5.35	5.52	5.43	3.43
e(4)	max	70.99	−58.11	12.55	67.89	−41.84	2.16
	min	55.64	−71.98	3.69	37.80	−71.84	−5.96
	mean	59.92	−61.69	8.14	50.05	−54.20	−2.00
	sdev	7.42	6.86	5.09	14.86	14.91	4.57
e′(4)	max	61.53	−0.20	−44.65	57.43	1.22	−46.05
	min	49.45	−11.66	−57.96	49.90	−4.91	−61.58
	mean	57.05	−4.71	−50.29	53.50	−1.40	−54.98
	sdev	5.37	5.27	6.13	3.44	3.03	6.59
f(7)	max	69.93	−44.49	50.01	−30.37	55.75	−52.42
	min	59.93	−62.82	33.92	−44.23	38.47	−65.77
	mean	65.09	−53.13	39.64	−36.42	46.00	−60.73
	sdev	3.85	6.68	5.91	4.68	6.76	4.67

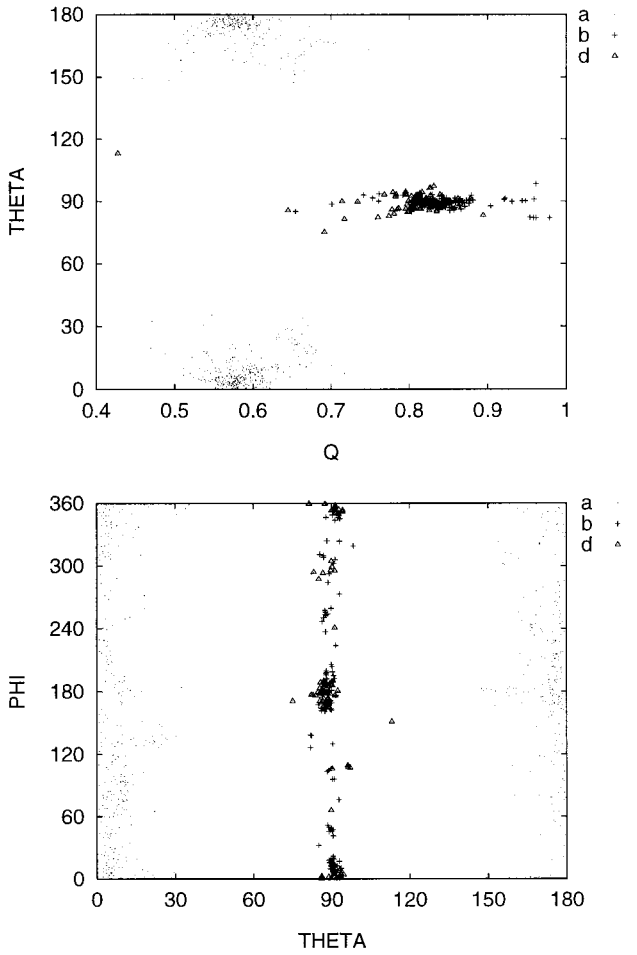

Figure 9. Ring puckering coordinates for major shape classes.

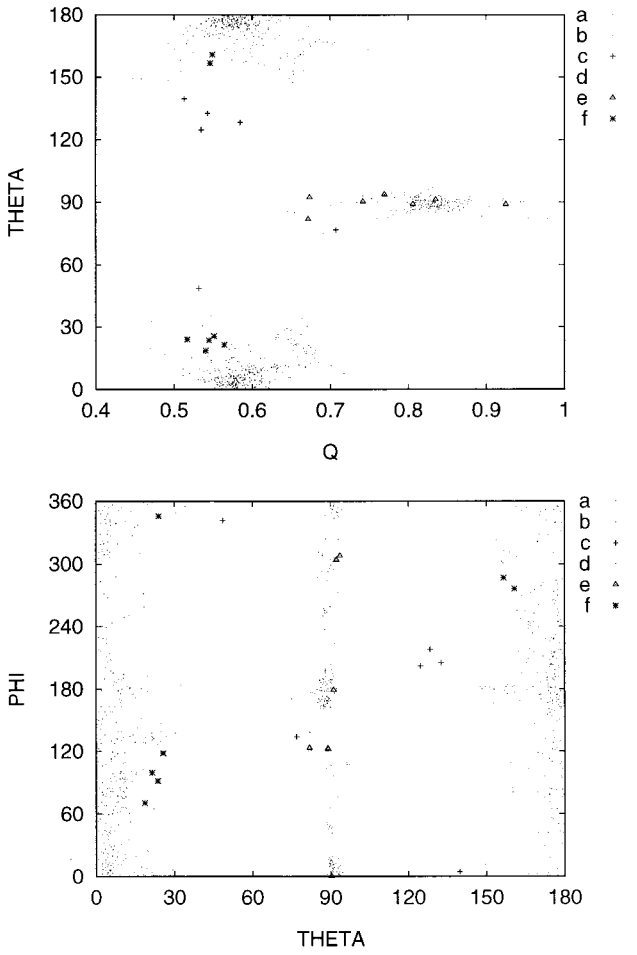

Figure 10. Ring puckering coordinates for all shape classes.

were terminated with a 0 and concatenated into a single integer sequence describing the embedding. Each element of *Aut*(*G*) was then applied to the embedding, and from the resulting set of integer sequences the largest, s_{max}, was chosen to represent the embedding. Equivalent van der Waals shape graphs have identical sequences s_{max}, so identifying the shapes present in a set of embedddings reduces to sorting the sequences s_{max}, and by tracking the automorphisms θ involved, the torsion angles of different embeddings can be correctly matched.

This procedure revealed the 10 shape classes indicated in Table 5, where the number of embeddings with each shape is shown in parentheses. Mirror image shape graphs are indicated with a prime. Over 97% of the rings have shape graphs *a* (chairs, the overwhelming majority), *b* and *b′* (twisted boats), and *d* and *d′* (boats). Data for the torsion angles of the embeddings with each shape graph are shown in columns 1−6. The mean values for shape graphs *a*, and *d* and *d′* should be compared with those for the textbook chair (60, −60, 60, −60, 60, −60) and boat (60, 0, −60, 60, 0, −60). The mean values for shape graphs *b* and *b′* are near (30, −60, 30, 30, −60, 30) and (60, −30, −30, 60, −30, −30), the mirror images of the halfway point between two successive boats in the pseudo-rotation of cyclohexane − the textbook twisted boat. Shape graphs *c* and *c′* both have one torsion angle with mean near zero, corresponding to the four coplanar atoms in a textbook half-chair embedding. The average values for shapes *e* and *e′* are consistent with those for shapes *d* and *d′* (after reversing the order of the torsion angles in *e*), indicating that all of these represent boat-like embeddings. Finally, comparing *f* with *a*, we see that the average shape in *f* is a slightly flattened chair.

The ring puckering coordinates[14] for an *n*-atom ring define a nonlinear projection of the $(3n - 6)$-dimensional conformation space into a space of dimension $n - 3$. Each hexagonal ring is represented by a point in a 3-dimensional space with spherical polar coordinates (Q, θ, ϕ). These are plotted for the shape classes identified in Figure 6, making no distinction between mirror images. Figure 9 shows the ring puckering coordinates for the members of the three largest shape classes: the chairs *a* form two clusters near the poles $\theta = 0, \pi$ of the sphere $Q = 0.6$, whereas the twisted boats *b* and boats *d* lie close to the equator $\theta = \pi/2$ of the sphere $Q = 0.8$. The ring puckering coordinates of all the embeddings are plotted in Figure 10, with the half-chair, 2-boat, and 2-chair embeddings identified to emphasize their positions relative to the main chair and boat clusters.

7. CONCLUDING REMARKS

The results presented here on classification of ring shapes and 3D substructure searching provide two applications of van der Waals shape and surface graphs, but many other methods used for analyzing molecular graphs may be applied to these graphs to obtain results incorporating the 3D shape of a molecular embedding. The ordered degree sequences already defined for *N*-methylpiperidine can be generalized

to embeddings of other molecules *G* by writing the degrees of the vertices of the van der Waals shape graph in the Unique SMILES ordering,[15] then taking the maximal sequence obtained by applying automorphisms of *G*. Invariants that are independent of *Aut*(*G*), such as connectivity indices[16] in quantitative structure−activity relationships and fingerprints[17] in similarity searching, can be applied directly to either van der Waals shape or surface graphs.

When a sequence of increasingly discriminating invariants is available, for instance the degree sequence, ordered degree sequence, and van der Waals shape graph used in §6, a hierarchical clustering of embeddings is obtained. Furthermore, any level in the hierarchy may be employed as a screen to select embeddings with 3D shapes similar, at this level, to a target embedding.

ACKNOWLEDGMENT

Dr. John Bradshaw of Glaxo Wellcome, and Professor Martyn Ford, Dr. Watcyn Wynn, and Dr. Neil Hoare of Portsmouth University, have made significant contributions throughout the development of this work and their encouragement and advice is gratefully acknowledged. Thanks are also due to a referee for his comments on an earlier version of this paper. Use of the EPSERC-funded Chemical Database Service at Daresbury is acknowledged. Financial support was provided by Glaxo Research and Development Ltd. and BBSRC grant number 322/PAC02714.

REFERENCES AND NOTES

(1) Whitley, D. C.; Ford, M. G. The application of molecular topology to drug design − topological descriptions of molecular shape. In *Molecular Similarity in Drug Design*; Dean, P. M., Ed.; Blackie: Glasgow, 1995; pp 269−290.

(2) Whitley, D. C. Van der Waals surface graphs and molecular shape. *J. Math. Chem.* **1998**, *23*, 377−397.

(3) Mezey, P. G. *Shape in Chemistry: An Introduction to Molecular Shape and Topology*; VCH: New York, 1994.

(4) Okabe A.; Boots B.; and Sugihara K. *Spatial Tesselations: Concepts and Applications of Voronoi Diagrams*; Wiley: Chichester, 1992.

(5) *SYBYL Molecular Modelling System (Version 6.4)*; Tripos Associates: St. Louis, MO.

(6) Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1969.

(7) Eliel, E. L.; Wilen, S. H. *Stereochemistry of Organic Compounds*; Wiley: Chichester, 1994.

(8) Wilson, R. J.; Watkins, J. J. *Graphs: An Introductory Approach: A First Course in Discrete Mathematics*; Wiley: Chichester, 1990.

(9) Arteca, G. A.; Mezey, P. G. Shape characterization of some molecular model surfaces. *J. Comp. Chem.* **1988**, *9*, 554−563.

(10) Jones, G. A.; Singerman, D. Theory of maps on orientable surfaces. *Proc. London Math. Soc.* **1997**, *37*, 273−307.

(11) White, A. T. *Graphs, Groups and Surfaces*; North-Holland: Amsterdam, 1973.

(12) Allen, F. H.; Kennard, O. 3D search and research using the Cambridge Structural Database. *Chemical Design Automation News* **1993**, *8*, *1*, 31−37.

(13) Fletcher, D. A.; McMeeking, R. F.; Parkin D. The United Kingdom Chemical Database Service. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 746−749.

(14) Cremer, D.; Pople, J. A. A general definition of ring puckering coordinates. *J. Am. Chem. Soc.* **1975**, *97*, *6*, 1354−1358.

(15) Weininger, D.; Weininger, A.; Weininger, J. L. SMILES II. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97−101.

(16) Trinajstic, N. *Chemical Graph Theory*, volumes I and II; CRC: Boca Raton, FL, 1983.

(17) James, C. A.; Weininger, D. *Daylight Theory Manual*; Daylight Chemical Information Systems, Inc., Irvine, CA, 1995.

CI970427Y