CHEMICAL STEREOGRAPHS: AN EXTENSION OF CHEMICAL GRAPHS FOR REPRESENTING STEREOCHEMICAL AND CONFORMATIONAL STRUCTURES

Mark JOHNSON*
Computational Chemistry, Upjohn Laboratories, Kalamazoo, MI 49001, USA

Chun-che TSAI
Department of Chemistry, Kent State University, Kent, OH 44242, USA

and

Victor NICHOLSON
Department of Mathematical Science, Kent State University, Kent, OH 44242, USA

Abstract

Chemical stereographs are presented as vehicles for representing qualitative three-dimensional features of molecules that put stereochemical and conformational distinctions in a common graph-theoretic formalism. They extend the concept of a chemical graph by adding "tetrads", each qualitatively characterizing the three-dimensional arrangement of four atoms with respect to its clinicity and handedness components. The characterization is sufficiently precise to distinguish synperiplanar, synclinal (gauche), anticlinal and antiperiplanar relationships between vicinal atoms of various conformers. Collectively, the tetrads constitute the embedding graph which presents new possibilities in displaying the stereochemical and conformational features of a molecule. A chemical graph and one of its possible embedding graphs constitute a chemical stereograph. Potential applications of chemical stereographs in the areas of structural representation, molecular symmetry analysis, and stereo-specific substructure searching are discussed.

1. Introduction

Molecules are three-dimensional entities, and complete mechanistic treatments of the interactions of molecules must take this into account by careful attention to the relative 3D coordinates of each of the atoms. On the other hand, there remains an area of the logic of molecular structure, classification and interaction that is best understood and most clearly communicated at a qualitative level that does not require, and can be obscured by, molecular representation specifying relative 3D coordinates of the atoms [1]. This point is amply demonstrated by the structural formulas so frequently used to identify molecules under discussion.

*Part of this work was presented by M.J. at the 1989 PacifiChem Mathematical Chemistry Minisymposium. The work was funded by The Upjohn Company.

At whatever qualitative level is being addressed, there is a need for vividly relevant diagrams that communicate the qualitative information of the problem. Examples include the common structural diagrams for molecules, Fisher projections for representing handedness in a sequence of chiral centers, Newman projections for indicating handedness and conformational distinctions, and ribbon diagrams for indicating "supersecondary" structure in proteins. These diagrams are almost invariably highly informative from a chemical viewpoint; otherwise, they are short lived. Thus, it is somewhat of a paradox that they are usually quite uninformative from a computational viewpoint. This is because, computationally, they are only bit-mapped images of two or more colors displayed on a sheet of paper or on a computer screen. If one wishes the computer to move beyond the simple storage and retrieval of these bit-mapped images, then various perception algorithms are required to interconvert these bit-mapped images into an analytical representation on which meaningful operations relevant to chemistry can be defined.

One of these analytical representations which has played an important role in many areas of chemistry is the mathematical concept of a graph. A graph is simply a set of points, called vertices, and another set consisting of pairs of these points, called edges. Although, simple in concept, graphs have found remarkable use in chemistry, especially in the form of a chemical graph [2]. As a chemical object, the vertices represent atoms and the edges represent the bonds defining the constitution of a molecule. As a mathematical object, the chemical graph is a graph in which the vertices and edges have been further distinguished by assigning each vertex and edge an appropriate atom and bond label. There are, to be sure, a number of "chemical" constraints imposed on this assignment, usually associated with valency. For example, a vertex labeled with a "C" cannot share edges with (be bonded to) more than four other vertices. Moreover, if one of its shared edges is labeled "double bond", then that vertex labeled "C" cannot share edges with more than three other vertices.

Chemical graphs are useful concepts in mathematical and computational chemistry because of the mathematical concepts, relationships and operations that can be used to relate graphs. The isomorphic and substructure relationships are two particularly important examples. However, it must always be remembered that these relationships are defined formalistically in terms of vertices, edges and label assignments without regard to chemistry. This can lead to difficulties. For example, 1, 2-dichlorobenzene can be represented by a chemical graph as a circuit of alternating single and double bonds in two mathematically nonisomorphic ways. Although such discrepancies can be used to advantage (see, for example, [3]), they obviously can create difficulties.

These discrepancies can often be circumvented in numerous ways. For example, we might say that two aromatic systems A and B are *equivalent* if there exists a chemical graph representation of A that is isomorphic to a chemical graph representation of B. Such a definition uses the mathematical concept of graph isomorphism to define a relevant concept of chemical equivalence. Alternatively, we could replace the alternating circuit of single and double bond labels with aromatic bond labels.

Then, aromatic systems A and B are equivalent if their chemical graphs are mathematically isomorphic. These two definitions of chemical equivalence accomplish the same end. Moreover, both use the same definition of mathematical isomorphism between graphs. However, the latter definition of topological chemical equivalence is simpler and more directly related to the mathematical definition of graph equivalence.

This issue of chemical equivalence versus mathematical equivalence underlies the focus of this paper. Here, we seek a general definition of stereochemical equivalence. However, what mathematical definition of equivalence do we use to define stereochemical equivalence?

A number of approaches have been tried. For a review, see [4]. Many in use in chemical-structure databases involve some type of extension of the atom and bond labels of the chemical graph. Two new classes of labels are defined. The first is a unique numbering of the atoms which needs not, but often reflects some sort of prioritization of the atoms somewhat along the line of the CIP (Cahn–Ingold–Prelog) rules [5]. The second is a parity label assigned to the atoms or bonds defining the chiral centers. Since chirality in $\mathbb{R}^3$ requires at least four points, a label assigned to a single vertex or edge cannot represent chirality without additional statements clearly indicating the set of atoms involved in the parity label assignment. In the cases of asymmetric carbon centers and double bonds, these additional statements can be quite simple. Moreover, these special cases handle a large majority of the compounds in many chemical databases. However, as one moves to higher coordinate centers or to the specification of the dihedral angles about a carbon–carbon single bond, a much more elaborate set of statements is required to unravel this parity assignment. This becomes especially true if one wishes to distinguish the various subarrangement of ligands that can be distinguished at a higher coordinate center. Although such subarrangements can, in principle, be computationally unraveled, the algorithms for doing so can become complex and unattractive from a mathematical viewpoint.

If one's only requirement is to determine if two labeled graphs used to represent stereochemistry are isomorphic, as in the case of compound registration, matters are greatly simplified by numbering the atoms in a canonically unique manner. Many approaches have been proposed [6–9]. In this case, the parity assignment does not need to be unraveled because equivalent stereochemical representations will always have the same assignment. Moreover, in the case of the numerous dihedral angles that may share a common bond, the parity assignment can represent the dihedral angle of whichever pair of vicinal atoms comes first in the canonical numbering [6]. However, in terms of a general treatment of stereochemistry from a mathematical chemistry viewpoint, canonical numbering schemes add an additional layer of complexity on the already bothersome problem of mathematically unraveling parity assignments.

A somewhat different approach is being promoted in the development of a standard data exchange format for chemistry [4]. In this case, the parity label is associated with the set of atoms that actually define the chiral center, thereby

eliminating the need to "decode" the atoms associated with the parity label. The parity assignment is made with respect to a standardized numbering associated with each "type" of chiral center. From a mathematical and computational viewpoint, both generality and simplicity are achieved in this manner. Moreover, only eight "types" (square, rectangle, anti-rectangle, tetrahedron, square-bipyramid, trig-bipyramid, octahedron, and cube) seem to cover most of the stereochemical distinctions required for most chemical databases. These features commend this approach as a data exchange format since the major requirement is an unambiguous transfer of structural information. Yet this approach has some problems from the viewpoint of mathematical chemistry. Although closed from a practical point of view, the set of possible types is not mathematically closed. In addition, it is not clear how one compares a sub-arrangement of atoms in one type with a subarrangement of atoms in another type, or how one uses types to extend the concept of a generalized transformation in computer-aided organic synthesis [10–15].

The use of "types" of three-dimensional arrangements has also been proposed by Balasubramanian [16] in an interesting group-theoretic treatment of stereochemistry. In this case, the nonterminal vertices of the chemical graph are distinguished as roots and assigned a type. The type is characterized in terms of group operations. The author refers to the chemical graph with the types assigned to the nonterminal vertices as a "molecular stereograph". We also use the term "molecular stereograph", but without employing group-theoretic terminology. In addition, our use incorporates dihedral information not represented in the Balasubramanian molecular stereograph. Although the Balasubramanian extension of the chemical graph does not incorporate dihedral information, the possibility of such a further extension would seem to exist. Whether or not either or both of these approaches might then be viewed as derivatives of the other has yet to be determined.

The augmentation of the chemical graph with group-theoretic concepts in obtaining representations of stereochemistry is a wide and productive area of research [17–22]. Here, we can only point out a few fundamental distinctions from our proposed approach. In most cases, stereoisomers are represented as permutations on a reference assignment of a set of ligands to the positions defined by a molecular skeleton. This facilitates both the enumeration of stereoisomers (see [23] for a review) and the examination of stereoisomerization modes [24, 11, 25 and cited references]. However, from a molecular similarity point of view, this creates difficulties when one wishes to define the similarities and differences between representations defined with respect to different molecular skeletons. Groups defined in terms of the automorphisms of a chemical graph have also been used to identify the chiral centers of a molecule and to generate all the stereoisomers of a molecule [19]. Interestingly, the same authors use a version of the parity labeled graph discussed earlier to actually represent the molecular configuration [18].

In a series of articles, Dreiding and coworkers [26, 27] propose representing the stereochemical structure of a molecule using "multiplexes", which can be viewed as an extension of the chemical graph. A chemical graph specified how pairs of

atoms are related in terms of their being bonded or not bonded. A multiplex specifies how pairs, triplets or even quadruplets of atoms are related. In particular, quadruplets of atoms can be used to specify the handedness of the four atoms bonded to an asymmetric carbon. Our approach falls into this last genre, with the twist that the quadruplets are organized into doublets relating one pair of atoms with another pair. These doublets are then represented by edges of a second graph, called the embedding graph. An embedding graph combined with the chemical graph constitutes a mathematical representation of the structural formula that provides a mathematically simple definition of a stereo-specific substructure.

Although the purpose of this study is to detail exactly how these two graphs, the chemical graph and the embedding graph, reflect the stereochemical structure of a molecule, the basic idea is quite easy to illustrate. Figure 1 shows eight qualitative stereochemcial positions associated with the conformers of 1, 2-disubstituted

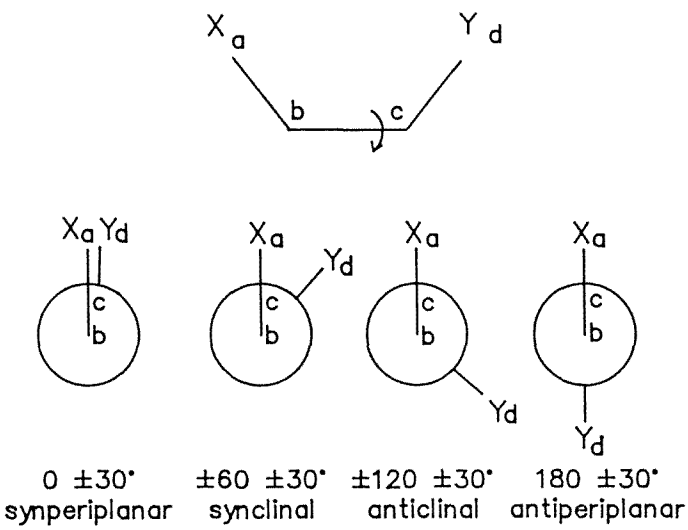


Fig. 1. Conformers of 1, 2-disubstituted ethanes.

ethane. Figure 2 decomposes these positions into their clinicity and handedness components. The details of this decomposition will be given in section 2.

Figure 3 depicts an indexed conformation of 1, 1, 2-trisubstituted ethane, together with its chemical graph and embedding graph. The chemical graph is a familiar representation of the bonds of the molecule. The only thing new is the "embedding graph". We see that each "vertex" of the embedding graph specifies a pair of indices of the chemical graph. Consequently, each "edge" of the embedding graph specifies the relationship between two pairs of atoms in the chemical graph. For example, in the center of the embedding graph, we see that the atom pair *bc*

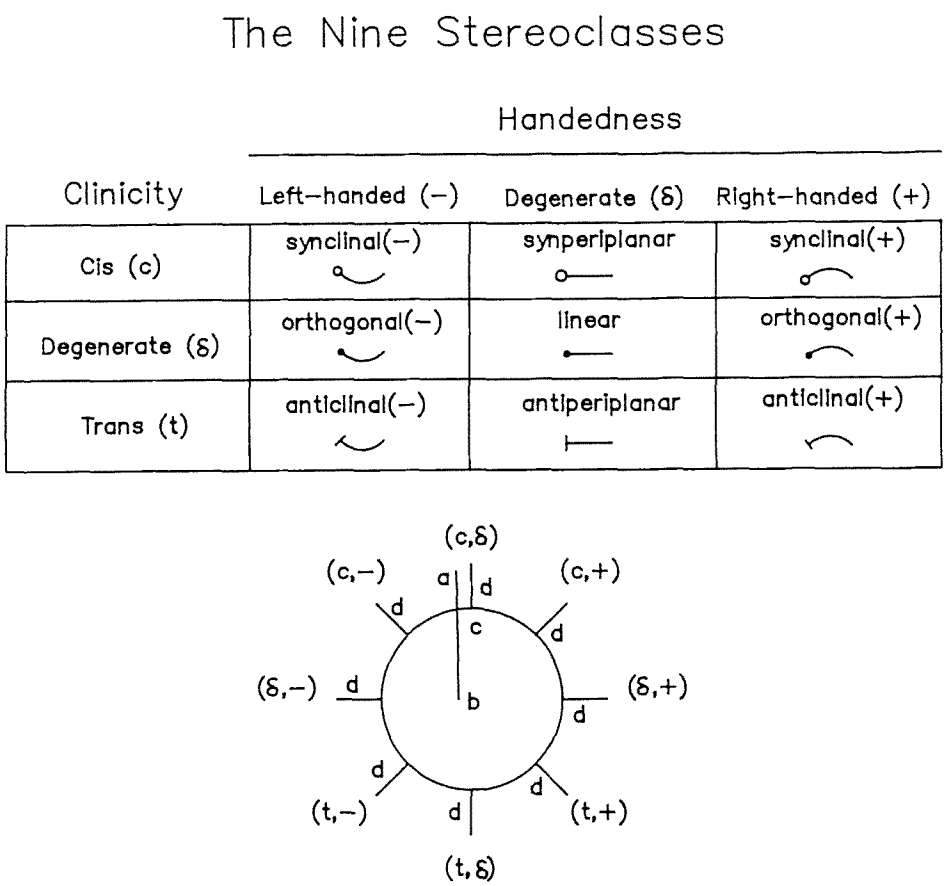


Fig. 2. A decomposition of nine qualitatively distinct arrangements of two vicinal atoms into their clinicity and handedness components.

is joined to the atom pair *ad* by a "synperiplanar" edge as defined in fig. 2, reflecting the fact that the dihedral angle $\angle abcd$ is roughly 0°. The cis clinicity is represented by the small circle of the edge, while the degenerate handedness is represented by drawing the main line of the edge straight. Similarly, the atom pair *bc* is joined to the atom pair *de* by an "anticlinal(−)" edge, indicating that the dihedral angle $\angle ebcd$ is between −90° and −180°. Here, the trans clinicity is represented by the short cross mark, and the negative or left-handedness is represented by a "left-turning" line as one walks from the *bc* pair to the *de* pair.

The dihedral angle $\angle abce$ formed by the three geminal atoms *a*, *c*, and *e* and the central atom *b* is also well defined. It is the angle formed by the intersection of the *abc*-plane with the *ebc*-plane. We can visualize this angle by replacing the *d* in fig. 2 with an *e*. In this case, atom *e* would actually lie in front of atom *c* rather than behind atom *c*, as is the case for atom *d*. The dihedral angle $\angle abce$ can be seen to be between

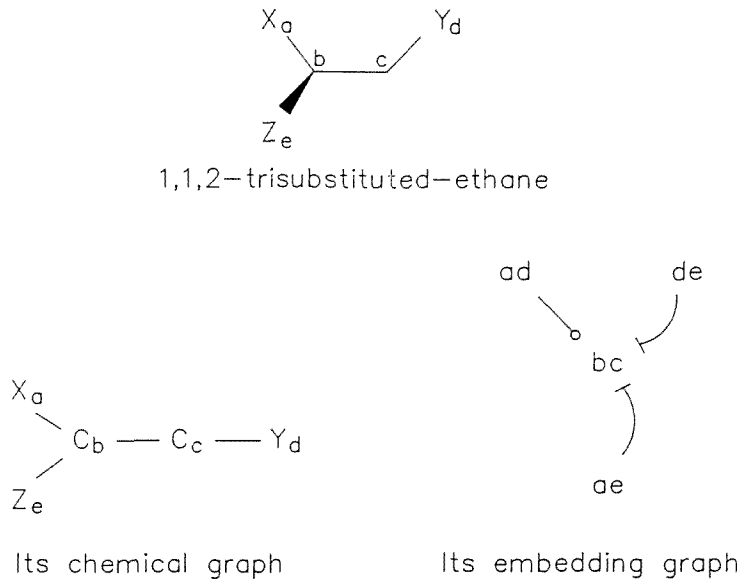


Fig. 3. The chemical graph and embedding graph of a conformation of 1, 1, 2-trisubstituted ethane.

90° and 180°. This is indicated in the embedding graph by the "anticlinal(+)" edge connecting the atom pair *bc* to the atom pair *ae*. Among other things, the "anticlinal(+)" gives an orientation to the four points that is equivalent to knowing the handedness of atom *b*. If ligand *Z* is reflected in the plane of the paper, the sign of the latter two dihedral angles $\angle ebcd$ and $\angle abce$ would change. This change would be reflected by changes in the handedness of the corresponding edges in the embedding graph.

Figure 3 is a brief illustration of a conformational stereograph. It specifies both the handedness of the chiral center and the qualitative size of the bond angle $\angle abe$ through the biedge (*bc, ae*). It specifies the orientation and qualitative magnitude of the two dihedral angles $\angle abcd$ and $\angle ebcd$ through the biedges (*bc, ad*) and (*bc, ed*). Thus, the stereograph provides a discrete representation of the qualitative three-dimensional arrangement of atoms without the use of coordinates, symmetry arguments and permutation groups, reference structures and types, or atom and bond parity labels requiring clarifying specifications of the atoms involved.

This brief illustration also leaves many points unanswered. For example, why was the dihedral angle $\angle abcd$ included, but the dihedral angle $\angle abec$ was not? Why was the dihedral angle $\angle abcd$ corresponded to the biedge (*bc, ad*) and not to the biedge (*ab, cd*)? These details will be clarified as the concept of a stereograph is developed. Section 2 begins by developing the concepts of clinicity and handedness for an ordered set of four points in three-dimensional space without reference to chemical graphs. Section 3 brings the chemical graph and the embedding graph together into a definition of a chemical stereograph and shows how to determine if two stereographs are equivalent.

The concept of isomorphism underlying this definition of equivalence is further developed in section 4, which discusses applications of chemical stereographs in structural representation, molecular symmetry analysis, and stereo-specific substructure searching. A glossary of terms can be found in the appendix.

2. Tetrads

We begin by defining the concepts of clinicity and handedness without reference to a chemical graph. There are a number of reasons for doing so. First, it frees the underlying geometric concepts of stereochemistry from their diverse expressions arising out of the specialized nomenclature and notational needs associated with the study of particular structural families and stereochemical problems. Second, it distinguishes the unique role played by the chemical graph in making stereochemical distinctions. Third, it enables one to see more clearly how, without the use of coordinates or permutations, a very primitive stereochemical concept can be repeatedly applied in diverse chemical contexts to develop very discriminating representations of molecular structure.

There are a number of qualitative statements one can make regarding the arrangement of any four points $\{a, b, c, d\}$ in three-dimensional space (herein denoted by $\mathbb{R}^3$). These statements are often made with respect to an imposed ordering of the points, much as is done when one orders the four ligands of an asymmetric carbon atom under the Cahn–Ingold–Prelog rules or when one orders the four atoms of a dihedral angle. We indicate this ordering by writing the four points as $a < b < c < d$ or by saying a precedes b precedes c precedes d. The ordered set $a < b < c < d$ will be called a *tetrad*. We shall call a the *first end*, b the *first center*, c the *second center*, and d the *second end*.

Let $T = a < b < c < d$ and assume that the four points of T are embedded in $\mathbb{R}^3$. Then, we can associate a clinicity $C(T)$ and a handedness $H(T)$ value with T as follows. We always associate the dihedral angle $\angle abcd$ with T. This dihedral angle is the angle formed by the intersection of the plane containing points a, b and c with the plane containing points b, c and d. If the set $\{a, b, c\}$, the set $\{b, c, d\}$, or both sets of points are collinear, as would arise in the cases involving a linear triple bond, the dihedral angle $\angle abcd$ is undefined. In this case, we say that the dihedral angle is *degenerate* and we write $C(T) = \delta$ and $H(T) = \delta$ to indicate that both its clinicity and handedness are degenerate. If neither set is collinear, there is a well-defined dihedral angle $\angle abcd$ associated with the tetrad T. To visualize this angle, we look down the line segment bc as indicated in fig. 2. Various combinations of clinicity and handedness values are indicated in that figure. These are defined algebraically by

$$C(T) = \begin{cases} c & \text{if } 0 \le |\angle abcd| \le 90 - \varepsilon, \\ \delta & \text{if } 90 - \varepsilon < |\angle abcd| < 90 + \varepsilon, \\ t & \text{if } 90 + \varepsilon \le |\angle abcd| \le 180 \end{cases}$$

and

$$H(T) = \begin{cases} + & \text{if } \varepsilon \le \angle abcd \le 180 - \varepsilon, \\ \delta & \text{if } 0 \le |abcd| < \varepsilon \text{ or if } 180 - \varepsilon < |\angle abcd| \le 180, \\ - & \text{if } -180 + \varepsilon \le \angle abcd \le -\varepsilon. \end{cases}$$

The value of ε is somewhat arbitrary, but should be chosen so that the dihedral angles seldom fall near the edges of the defining cutoff intervals.

The ordered pair $(C(T), H(T))$ consisting of the clinicity $C(T)$ and the handedness $H(T)$ of the tetrad T will be called the *stereocity* of that tetrad. It should be noted that every tetrad embedded in $\mathbb{R}^3$ has a well-defined stereocity, even if its dihedral angle is degenerate.

Although our use of figs. 1 and 3 in illustrating and developing the concept of a tetrad and its stereocity suggests that chemical bonds are somewhat basic to the definitions, this is not the case. The preceding algebraic definition is applicable to every set of four points in $\mathbb{R}^3$ which has somehow been prioritized or ordered. However, we shall later see how the bonding structure of a molecule enables us to select which sets of four points are of interest, and how these points might be ordered.

Four points $\{a, b, c, d\}$ can be ordered in any one of 4! or 24 possible ways. Each ordering defines a tetrad and can be thought of as a perspective from which to view the point set $\{a, b, c, d\}$. As we have just seen, each tetrad or "perspective" can have any one of nine possible stereocities. The 24 tetrads with their associated stereocities give a fairly detailed characterization of the arrangement of the four points which is independent of changes in scale and location. We shall refer to this characterization as the *tetral arrangement* of four points.

As a characterization of the arrangement of four points in space, the tetral arrangement is a much finer classification of the arrangements of four points in space than the CIP classification of all nonplanar arrangements into achiral, left-handed or right-handed [28]. The lower resolution of the CIP characterization arises because the ordering of or "perspective on" the four points is not treated as an inherent part of the characterization. Rather, the ordering is introduced only if the four points are in a chiral arrangement. As a consequence, the CIP classification must treat cis–trans distinctions as planar concepts involving the handedness of two triangles in the plane. In this way, planar arrangements of four points are classified as "achiral in the plane", cis, and trans. The "achiral in the plane" arises if two or more of the four points are not distinguished. By making the ordering an inherent part of the characterization, the distinction or nondistinction of the four points due to symmetry considerations ceases to be an issue; the ordering itself distinguishes the points.

It is usually most efficient to treat a problem at the lowest level of resolution that can differentiate any issues that must be resolved in the problem. From this view, the CIP rules often suffice. We reduce the resolution of the tetral arrangement classification by excluding many of the tetrads that might be associated with four points. This will be done in the next section when the chemical graph is reintroduced into the discussion.

The number of possible tetral arrangements of four points in $\mathbb{R}^3$ is an open problem. Its solution will require a definition of when two tetrads embedded in $\mathbb{R}^3$ are equivalent.

Before giving a definition, it helps to see what happens to the handedness of the $(c, +)$ tetrad in fig. 2 when we change the ordering of the two centers and/or the ordering of the two ends. Suppose we change only the ordering of the two ends. To ascertain the handedness of the dihedral angle in this case, we again look down the bc line segment from b to c. However, now we must note the direction our eye moves as it proceeds from point d to point a, because d comes before a in the interchanged ordering. In this case, our eye goes to the left. Thus, an interchange in the ordering of the two ends changes the handedness. Similarly, if we interchange the ordering of two centers, we must look down the bc line segment from point c to point b. In effect, we must look down the bc line segment in fig. 2 from the opposite side of the paper. In this case, our eye moves to the left as it proceeds from point a to point d. Consequently, an interchange in the ordering of the two centers changes the handedness. However, if we interchange both the two ends as well as the two centers to obtain the dihedral angle $\angle dcba$, we must look from the opposite side of the paper and move our eyes from d to a. In this case, our eyes move to the right just as they did originally. It follows that if we make both interchanges, i.e. interchange the ends and also interchange the centers, then the assigned handedness is unchanged.

Now we define when we wish to say two tetrads are handedly equivalent. Let $A_< = a < b < c < d$ and $B_< = w < x < y < z$ be two tetrads with point sets $A = \{a, b, c, d\}$ and $B = \{w, x, y, z\}$. As we have seen, if the points in A and the points in B are points in $\mathbb{R}^3$, then both $A_<$ and $B_<$ have well-defined stereocities. Any one-to-one mapping $f: A \to B$ that corresponds *centers to centers* and *ends to ends* will be called a *structural correspondence* between $A_<$ and $B_<$. There are only four possible structural correspondences, given by

(f_1) $a \leftrightarrow w$, $b \leftrightarrow x$, $c \leftrightarrow y$ and $d \leftrightarrow z$,

(f_2) $a \leftrightarrow w$, $b \leftrightarrow y$, $c \leftrightarrow x$ and $d \leftrightarrow z$,

(f_3) $a \leftrightarrow z$, $b \leftrightarrow x$, $c \leftrightarrow y$ and $d \leftrightarrow w$,

(f_4) $a \leftrightarrow z$, $b \leftrightarrow y$, $c \leftrightarrow x$ and $d \leftrightarrow w$.

Correspondence f_1 interchanges no points, while correspondence f_4 makes two interchanges. Thus, f_1 and f_4 are called *even* correspondences. On the other hand, correspondences f_2 and f_3 are called odd since they make one interchange. We shall say $A_<$ is *handedly* equivalent to $B_<$ with respect to the structural correspondence $f: A \to B$ if one of three cases holds:

(1) $A_<$ and $B_<$ both have degenerate handedness;

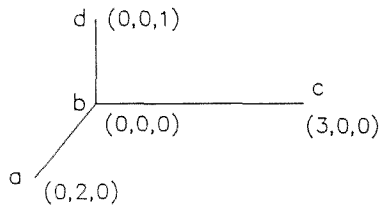
(2) $A_<$ and $B_<$ have the same nondegenerate handedness and f is even;

(3) $A_<$ and $B_<$ have opposite nondegenerate handedness and f is odd.

Every two nondegenerate tetrads $T = a < b < c < d$ and $T' = w < x < y < z$ are handedly equivalent with respect to two of the four possible structural correspondences

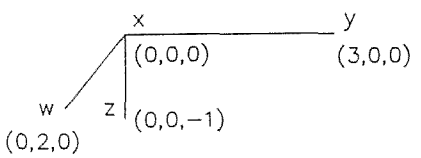
and not handedly equivalent with respect to the other two. If T and T' have the same nondegenerate handedness, then T and T' are handedly equivalent with respect to the two even structural correspondences, but not with respect to the two uneven ones. If T and T' have opposite nondegenerate handedness, then T and T' are handedly equivalent with respect to the two odd correspondences, but not with respect to the two even ones. Thus, we must rule out some of the possible structural correspondences if we are to have a definition of stereo equivalence of tetrads that distinguishes handedness.

If the two tetrads are embedded in $\mathbb{R}^3$ as indicated by the two tetrads in fig. 4, this can be simply done by requiring that the structural correspondences preserve *relative*



a<b<c<d
left-handed

a



w<x<y<z
right-handed

b

	f_1	f_2	f_3	f_4
a	w	z	w	z
b	x	x	y	y
c	y	y	x	x
d	z	w	z	w
handedly equivalent?	no	yes	yes	no
preserves relative distances?	yes	no	no	no

c

Fig. 4. The four structural correspondences between two tetrads embedded in $\mathbb{R}^3$.

distances. (Ignore the coordinate line segments used to "visualize in $\mathbb{R}^3$" each set of four points and concentrate only on the tetrad ordering and relative distances of the points.)

In the figure, only correspondence f_1 preserves relative distances because only f_1 corresponds the smallest distance in fig. 4(a), which is between points b and d, to the smallest distance in fig. 4(b), which is between points x and z. Since the two tetrads are not equivalent under f_1, the two tetrads are not handedly equivalent, as indeed they clearly cannot be.

We define two tetrads T and T' embedded in $\mathbb{R}^3$ to be *equivalent in* $\mathbb{R}^3$ if they have identical clinicities and they are handedly equivalent with respect to a structural correspondence that preserves relative distances. This definition is a "geometric" definition of equivalence. Its use of distance implies a use of coordinate representations of the points. We seek a coordinate-free formalism for stereochemistry. Thus, we shall restrict the possible structural correspondences using constraints defined in terms of the connectivity or bonding in the chemical graph rather than constraints defined in terms of relative distances.

3. Chemical stereographs

Consider the period of time a set of atoms exists as a particular molecule in three-dimensional space. During this period of time, the pairwise distances of the molecule are in continual flux. However, there are many constants or invariants in this flux. In particular, the distance between two bonded atoms stays within a small circumscribed range that defines their bond distance. If by chance this range is exceeded, the bond is broken and the set of atoms changes its molecular identity. A chemical graph of a molecule can be viewed as a drawing of that set of invariants of this flux which define the chemical bonds of the molecule.

There are other invariants of this flux that are not encoded in the chemical graph. For example, the dihedral angle given by a cis arrangement of two vicinal atoms associated with a double bond varies within a small circumscribed range about 0°. As another example, any three geminal atoms of an sp^3 carbon atom stay within a relatively well-defined arrangement. These examples describe invariants associated with the fluxional motion of four atoms which Floersheim et al. [27] refer to as "mobility restrictions". In the proposed formalism, these stereochemical invariants are represented as tetrads with assigned stereocities. What we shall term the embedding graph of a molecule can be viewed as a "drawing" of a set of tetrads which define various stereochemical invariants of the molecule. Whereas the edges of a chemical graph have values such as "single" and "aromatic", the edges of the embedding graph have values such as "synclinal(−)" and "antiperiplanar".

3.1. TETRADS DEFINED ON CHEMICAL GRAPHS EMBEDDED IN $\mathbb{R}^3$

We begin by defining an important subset of tetrads based on the chemical graph of the molecule. To do this, we shall refer to the two ends and two centers of a tetrad as the *tetral structure* of the tetrad. The four tetrads $a < b < c < d$, $d < b < c < a$, $a < c < b < d$ and $d < c < b < a$ have the same tetral structure because they have identical centers and ends. On the other hand, $a < b < c < d$ and $b < a < c < d$ differ in their

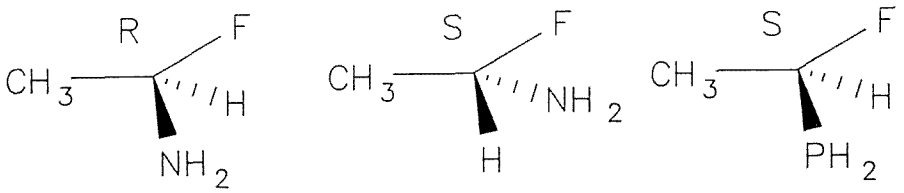
tetral structure because the first tetrad has points *a* and *d* for ends, whereas the second has points *b* and *d* for ends.

Two types of tetrads can be distinguished by how their tetral structures are defined: vicinal and geminal. The *vicinal* tetrad is perhaps the most obvious. Its ends are the two vicinal atoms and its centers are the two atoms of the center bond. A vicinal tetrad is classified as either *free* or *fixed*, depending on whether or not the bond connecting the centers is free to rotate.

Now suppose a molecule is flopping around in $\mathbb{R}^3$. At each instant, any vicinal tetrad has a well-defined stereocity according to the conventions set forth in fig. 2. If the vicinal tetrad is fixed, the stereocity will remain the same for all instants in time over which the molecule maintains its identity. Thus, a fixed tetrad has a well-defined value for a molecule, or expressed another way, is a motion invariant of a molecule.

On the other hand, the stereocity of a free vicinal depends on which instant in time we are talking about. However, over any interval of time in which the molecule exists as a single conformer, a free vicinal will, generally speaking, have a well-defined stereocity. That is, for all instants in time that the molecule exists as that particular conformer, the stereocity assigned that tetrad will remain unchanged. Thus, a free vicinal tetrad is a motion invariant of a conformer, but not of the associated molecule.

Our choice for the geminal tetrads is less obvious, and reflects the work by Dreiding and Wirth [26] on chirons. Consider the handedness of the asymmetric carbon atoms in the following 1-substituted-fluoroethanes. Using the Cahn–Ingold–Prelog (CIP) ordering of the atoms, molecules **1** and **2** are assigned opposite handedness, which



1 2 3

Scheme 1.

symbolically suggests to the uninitiated some type of stereochemical difference. This is as it should be. On the other hand, molecules **1** and **3** are assigned opposite handedness, which again symbolically suggests to the uninitiated some type of stereochemical difference. However, in this case, the amino and phosphino groups occupy the same relative positions in the two molecules. Thus, the CIP rules do not lead to a symbolically straightforward method of comparing the relative positions of the amino and phosphino groups in molecules **1** and **3**. The problem does not lie in the choice of the ordering of the atoms. The problem lies in the fact that the relative positions of the amino and phosphino groups

are defined with respect to the central atom and the remaining three ligands, whereas the CIP handedness is assigned with respect to the four ligands.

Although one could easily consider tetrads whose four points constitute the four geminal atoms surrounding an asymmetric carbon, this problem of distinguishing between the absolute stereochemical assignment (the R, S assignment) from the relative stereochemical positioning is best avoided for our ultimate purpose of incorporating stereochemistry into a formalism suitable for molecular similarity analysis of diverse structures. Thus, we define the four atoms of a *geminal* tetrad to consist of the central atom and three geminal atoms. In addition, we require that every geminal tetrad has its associated central atom as one of its centers. We shall classify a geminal tetrad as either *free* or *fixed*, depending on whether or not its central atom is invertible. For example, since amines generally have low inversion barriers, geminal tetrads in which the central atom is an sp^3 nitrogen generally will be free. On the other hand, geminal tetrads in which the central atom is an sp^3 carbon generally will be fixed.

Again, suppose a molecule is flopping around in $\mathbb{R}^3$. At each instant, a geminal tetrad has a well-defined stereocity according to the conventions set forth in fig. 2. If the tetrad is fixed, the stereocity will remain the same for all instants in time over which the molecule maintains its identity. Thus, like a fixed vicinal tetrad, a fixed geminal tetrad has a well-defined value for a molecule, i.e. is a motion invariant of that molecule.

On the other hand, the stereocity of a free geminal tetrad depends on which instant in time we are talking about. For example, if the central atom of the geminal tetrad is the nitrogen of an amine, the handedness of the associated geminal tetrad would change each time the amine undergoes an inversion. However, over any interval of time in which the molecule exists as a single conformer, a free geminal tetrad will, generally speaking, have a well-defined stereocity. Thus, a free vicinal tetrad is a motion invariant of a conformer, but not of the associated molecule.

Prelog and Helmchen [28] discuss many of the issues raised here regarding the definitions of vicinal tetrads and geminal tetrads using the terms helical and tripodal stereogenic units. However, there are some critical differences that should be noted besides the differences in the way ordering or precedence is taken into account and the resulting resolution of the stereochemical characterizations that is obtained. Their definition of a stereogenic unit is "three-dimensional"; the different types of units are defined in terms of chirality centers, planes, and axes. Moreover, the terms "helical" and "tripodal" have strong geometric connotations which we are removing in order to obtain a mathematics of stereographs which is free of three-dimensional concepts and operations.

The concepts of vicinal and geminal tetrads are easily generalized to handle more complicated stereochemical distinctions that can arise. A vicinal tetrad is easily generalized by replacing the centers of the tetrad with two atoms at the end of any path, such as the path defined by the two double bonds of allene. Such extensions are necessary if one is to use chemical stereographs to differentiate the stereoisomers of substituted allenes. Geminal tetrads can be generalized in a similar manner, i.e. replacing the bonds with paths that intersect only at the "central" atom. The concepts of fixed and free can also be defined in this more general setting, since they simply distinguish tetrads whose

assigned stereocity remains unchanged during the existence of a molecule from tetrads whose assigned stereocity can change during that existence. These generalizations are given simply to suggest the variety of ways tetrads might be defined in terms of the chemical graph of a molecule. However, since geminal and vicinal tetrads encompass by far the largest part of the stereochemistry of small molecules, we shall restrict our attention to such tetrads in the remainder of this study for reasons of simplicity.

3.2. A LOCAL ORDERING CONVENTION FOR REDUCING THE NUMBER OF TETRADS

Recall that the tetral structure of a tetrad is a specification of its centers and ends. As noted earlier, there are always four tetrads that have a given pair of centers x and y, and a given pair of ends w and z. These are $w < x < y < z$, $z < x < y < w$, $w < y < x < z$, and $z < y < x < w$. Any one of these tetrads determines the stereocity of the other three by virtue of the four structural correspondences in section 2. Thus, we only need to include one of them in the embedding graph. The question is: Which one?

The CIP rules attempt to obtain a canonical ordering that can be agreed upon by investigators. As mentioned in the introduction, canonical naming schemes are useful for identifying compounds, but have received little use in molecular similarity analysis due to, we feel, their mathematical and algorithmic complexity. In fact, molecules are still being envisioned which have nonequivalent atoms that the CIP rules fail to distinguish [29,5]. Instead, we shall use the following ordering convention which assumes that the atoms are alphanumerically ordered by their indices. This ordering convention is "local" like the one in [5]. As such, it is easily evaluated and well-defined in all cases. We shall see that it still provides an "absolute" designation of handedness in many situtations.

(1) The atoms of a vicinal tetrad are always ordered from one end to the other in the direction consistent with the alphanumeric ordering of the indices of the centers.

(2) The atoms of a geminal tetrad are ordered so that the central atom is the first center. If one of the ends is a terminal atom and the other end is not, that terminal atom is the second end. If both ends are terminal atoms and one is a hydrogen atom, that hydrogen atom is the second end. Otherwise, the remaining two ends are ordered alphanumerically by their indices.

In many cases, this ordering convention generates an "absolute" handedness designation. For example, suppose we have a geminal tetrad with central atom y, other center x, terminal end w, and nonterminal end z. This situation would arise if the central atom were the asymmetric carbon in 3-pentanol, the nonterminal end and other center were the two geminal carbons, and the terminal end were the alpha hydrogen. From the second statement in rule (2), the two ends are ordered according to $z < w$. From the definition of a tetrad, the centers must lie between the two ends in the ordering of the four points. It follows that $z < y < x < w$ is the only tetrad possible having the supposed tetral structure. Since we arrived at this ordering without recourse to the alphanumeric

ordering of the vertices of the chemical graph, the assigned stereocity associated with the arrangement of these four atoms will be independent of the investigator. A similar statement can always be made with regard to vicinal tetrads. To see why, we simply note that the two possible orderings are *even* permutations of each other and consequently must be assigned identical stereocities. This independence of the stereocity assignment from the alphanumeric ordering of the atoms of a tetral structure is summarized in the following *ordering invariance* rule:

> The only stereocity assignments that are dependent on the alphanumeric ordering of the atom indexing are those for geminal tetrads whose ends are (a) both nonterminal atoms, (b) both terminal atoms, neither of which is a hydrogen, or (c) both hydrogen atoms.

3.3. REPRESENTING THE TETRADS OF A CHEMICAL GRAPH AS AN EMBEDDING GRAPH

We have been using the term "embedding graph" for the set of tetrads associated with the chemical graph. We have not actually shown that the information in this set of tetrads can, in fact, be represented by a graph, that is, by a set of vertices and a set of edges.

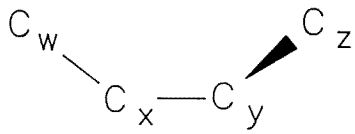
We begin by representing the tetral structure of each tetrad $a < b < c < d$ as an ordered pair (bc, ad) in which the first pair, b and c, of points defines the centers of the tetrad and the last pair, a and d, defines the ends. This ordered pair will be called a *biedge*.

The biedge (bc, ad) is in standard form because its centers and ends are both ordered alphanumerically. This is not essential. The biedges (bc, da), (cb, ad) and (cb, da) also represent the tetral structure of the tetrad $a < b < c < d$. Whenever convenient, we shall write biedges in standard form.

If we are considering only vicinal and geminal tetrads, it is an easy matter to correspond any biedge to a unique tetrad by means of our ordering convention. Thus, we have a one-to-one correspondence between the set of vicinal and geminal tetrads of a chemical graph with alphanumerically labeled vertices and the set of biedges associated with that same graph.

We now let the set of biedges define the set of directed edges of the embedding graph and let the collection of center pairs and end pairs of the biedges define its vertex set. We shall call this graph a *generic embedding* graph. An *embedding* graph is a generic embedding graph in which each biedge is assigned a stereocity.

Figure 5 illustrates these definitions for the simple case of a synclinal conformation of *n*-butane. Assuming the points in the chemical graph are ordered in accordance with the alphanumeric ordering of their indices, the only vicinal tetrad constructable from the chemical graph is given by $w < x < y < z$. There are no geminal tetrads. Consequently, its generic embedding graph consists of a single biedge (xy, wz), indicated by drawing a directed line starting at a dot placed next to the center pair xy and terminating at the end pair wz.



A synclinal conformation
of the skeleton of n−butane.

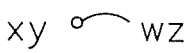
$$C_w - C_x - C_y - C_z$$

Its hydrogen−reduced chemical graph.

$$w<x<y<z$$

Its only tetrad.

Its generic embedding graph.



Its embedding graph.

Fig. 5. The hydrogen-reduced graph and an associated embedding graph of a synclinal conformation of normal butane.

To obtain the embedding graph, we replace each directed line in the generic embedding graph with a corresponding stereocity symbol from fig. 2. In this notation, the marked end is placed nearest the center pair. The mark is a circle if the clinicity is cis, a dot if the clinicity is degenerate, and a short line creating a tee if the clinicity is trans. In addition, as one "stands" on the side of the paper with the printed arc and "walks" along the arc from the center pair over to the end pair, one will "turn to the right" if the handedness is plus and "turn to the left" if the handedness is minus. Otherwise, the handedness is degenerate and the line is straight. In fig. 5, the single line in the generic embedding graph is replaced by the synclinal(+) symbol.

Figure 3 is now understandable for the most part. The vicinal biedge (*bc*, *ad*) is associated with a synperiplanar tetrad with centers *b* and *c* and ends *a* and *d*. We see

from the chemical graph that this tetrad is a vicinal tetrad. By the ordering convention, the centers are ordered alphanumerically according to $b < c$ and consequently the point set must be ordered $a < b < c < d$. The geminal biedge (bc, ae) is associated with an anticlinal(+) tetrad with centers b and c and ends a and e. We see from the chemical graph that this is a geminal tetrad. By the ordering convention, the centers are ordered so that the center corresponding to the central atom of the tetrad comes first, i.e. $b < c$. Since the two ends are not distinguished by the ordering convention, they are ordered alphanumerically. Thus, the corresponding tetrad is $a < b < c < e$.

Although we can understand the information conveyed in the embedding graph, our reasons for including some biedges in the embedding graph and not including others are not yet clear. The biedges (ab, cd) and (cd, ae) are not included because their centers and ends do not represent tetral structures of either vicinal or geminal tetrads. However, the biedges (be, ac) and (ab, ce), which have tetral structures corresponding to geminal tetrads, were also excluded. The reason for doing so will be addressed after we formally define a chemical stereograph.

3.4. DEFINITION OF A CHEMICAL STEREOGRAPH

Let $G = (V, E)$ denote a chemical graph in which the vertices in V have atom assignments and the edges have bond assignments. Let $\mathcal{T}$ denote a set of tetrads defined on G which are in a one-to-one correspondence with a set BE of biedges. Then, the triple (V, E, BE) will be called a *chemical stereograph* with chemical graph (V, E) and embedding graph (BV, BE), where BV is the set of center pairs and end pairs derived from the biedges in BE.

The generality of this definition merits emphasis. Although we have largely been restricting our attention to vicinal and geminal tetrads, no such restriction is made or implied in the definition of a chemical stereograph. Any of the 24 possible orderings of any set of four atoms of the underlying chemical graph might constitute a tetrad corresponding to a biedge in BE. The only requirement is that there exists a one-to-one correspondence between the tetrads in $\mathcal{T}$ and the biedges in BE. This correspondence will usually come in the form of some type of correspondence convention. Section 3.2 illustrated how a correspondence convention can be constructed when the tetrads in $\mathcal{T}$ are either vicinal or geminal tetrads. This convention easily generalizes to the case of generalized vicinal and generalized geminal tetrads.

The definition of a chemical stereograph also generalizes to the case in which the vertices in V need not represent atoms and the edges in E need not represent bonds. For example, the vertices might represent characteristic points on the α-helices and β-strands defined by the secondary structure of a protein and the edges denote various ways these characteristic points might be pairwise associated [30,31]. We will usually refer to a chemical stereograph by the generic term stereograph unless we wish to emphasize the particular tie a chemical stereograph has to its underlying chemical graph.

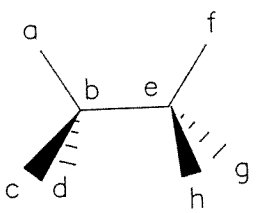
3.5. MOLECULAR AND CONFORMER STEREOGRAPHS

A chemical graph with p vertices gives rise to p choose 4 or $(p(p-1)(p-2)(p-3)/4!)$ ways of selecting the four atoms of a tetrad. Each such subset gives rise to 4 choose 2 or 6 possible tetral structures. Since each tetral structure can be ordered 4 ways, it is clear that there is a large number of possible tetrads that might be used to characterize the three-dimensional structure of a molecule. By means of the four structural correspondences in section 2, the latter four orderings were shown to convey the same information. As a consequence, only one ordering is needed. However, we are still left with a large number of tetral structures. Section 3.1 excluded all but a fraction of the tetral structures by restricting our attention to vicinal and geminal tetrads. In this section, two factors will be considered in further reducing the number of tetrads under consideration.

The first factor relates to the amount of structural information available. If the stereocity of a tetrad is not constant for a molecule, that tetrad can be excluded from consideration. It follows that more tetrads are used to represent conformers than to represent their corresponding molecules because more is known about the dihedral angles of conformers than about their corresponding molecules. This distinction is reflected in the distinctions between molecular and conformer stereographs.

However, even the number of tetrads in a molecular stereograph can be quite large. A pentacoordinate atom has 5 choose 3 or 10 ways of choosing the sets of 3 geminal atoms of its possible geminal tetrads. In addition, there are 3 choose 2 or 3 ways of selecting the two ends of the tetral structures involving 3 geminal atoms. Thus, there are 30 tetral structures in the total characterization of the geminal tetrads of a pentacoordinate center. There is an obvious desire to establish and remove redundancies that remain in the complete set of geminal and vicinal tetrads. This section describes a rule for excluding most of these redundant tetrads.

Let $\mathcal{T}(S)$ denote the set of tetrads associated with the chemical stereograph S as a result of the correspondence convention. Two geminal tetrads in $\mathcal{T}(S)$ are *siblings* in S if they involve the same central atom. Since the atoms of a molecule exist in hybridization states with characteristic symmetry groups, the stereocity of most geminal tetrads can be deduced from a knowledge of the stereocity of any of its siblings. (This may not be the case for the generalized geminal tetrads mentioned at the end of section 3.1.) This sibling redundancy can be seen in structure **4**. For example, we



4

Scheme 2.

know from the geometry of an sp^3 asymmetric center that the positions of atoms a, b, d, and e determine the position of atom c relative to its geminal atoms. Thus, the stereocity of the geminal tetrad $a < b < e < d$ determines the stereocity of the siblings $a < b < e < c$ and $a < b < c < e$.

A similar result holds for vicinal tetrads. Two vicinal tetrads are *siblings* if they have the same centers. The stereocity of two *nonsibling* geminal tetrads sharing their respective centers with a vicinal tetrad combined with the stereocity of that vicinal tetrad determines the stereocity of all siblings of that vicinal tetrad. For example, if we know the stereocity of the nonsibling geminal tetrads $a < b < e < d$ and $f < e < b < g$ in structure **4** as well as the stereocity of the vicinal tetrad $a < b < e < f$, then we can deduce the stereocity of the vicinal siblings $a < b < e < g$ and $c < b < e < h$. It is this redundancy amongst sibling vicinal tetrads that underlies Wipke's [6] development of computer names for conformers.

Now suppose S is a stereograph with chemical graph G. A geminal or vicinal tetrad T is *represented* in S if there is a biedge of S that corresponds to a sibling of T. It follows from the preceding discussion that if T is represented in S, then the stereocity of T can be deduced from S even if T is not directly corresponded to a biedge of S. Returning to fig. 3, we see that the geminal tetrad $T = c < b < a < e$ has a tetral structure that does not correspond to any biedge in the embedding graph, i.e. (ab, ce) is not a biedge of the embedding graph. However, T is a sibling of the geminal tetrad $a < b < c < e$ corresponding to the biedge (bc, ae). Thus, T is represented in the embedding graph even though it does not correspond to an edge of the embedding graph.

We will be primarily interested in stereographs of chemical entities which are either molecules or conformers. A stereograph of a molecule will be called a *molecular* stereograph if it satisfies two conditions. First, all of its biedges must correspond to *fixed* geminal and *fixed* vicinal tetrads. As discussed in the introduction to section 3, only fixed tetrads have the same assigned stereocity over all conformational states of a molecule. Second, every fixed geminal and fixed vicinal tetrad of that molecule must be represented by a biedge. This second condition simply says that if a vicinal or geminal tetrad has a single, and consequently well-defined, stereocity over all conformational states of a molecule, then that bit of structural information is represented by at least one biedge in the molecular stereograph. A stereograph of a molecule is called a *conformational* stereograph of that molecule if all of its biedges represent geminal and vicinal tetrads and if every geminal and vicinal tetrad of that molecule is represented by a sibling tetrad.

The stereograph S in fig. 3, which is represented by the chemical graph and embedding graph, is not a molecular stereograph. This is because the carbon–carbon bond is free to rotate. Consequently, $a < b < c < d$ is a free vicinal tetrad. Since it is represented by, in fact corresponds to, the biedge (bc, ad) of the embedding graph, S cannot be a molecular stereograph. However, S is a conformational stereograph. Every edge of its embedding graph corresponds to either a geminal or vicinal tetrad. Moreover, every geminal or vicinal tetrad defined on the chemical graph is represented by a biedge of S.

The informational redundancy in a family of sibling tetrads can be used to dramatically reduce the number of tetrads that must be included in the stereographic representation

of a molecule or conformer. However, diminishing informational redundancy can reduce informational accessibility, especially if rather arbitrary rules are used to eliminate redundancies. We attempt some sort of middle ground by introducing the following *sibling selection rules* for filtering out "redundant" sibling geminal and vicinal tetrads.

Assign each atom of a tetrad, geminal or vicinal, a value of 0 if it is a hydrogen, 1 if it is a halogen, 2 if it is a generic ligand, and 3 otherwise. Let the score of the tetrad be the sum of its two *end* values. Accept only those geminal siblings with the lowest score and accept only those vicinal siblings with the highest score. Of the geminal siblings with the lowest score, accept only those with the highest number assigned to the noncentral center.

To determine which geminal and vicinal tetrads are to be corresponded to biedges in the stereograph of 1, 1, 2-trisubstituted ethane, we shall use the indexing of its chemical graph as given in fig. 3. Under the scoring system, the three ligands *a*, *d*, and *e* are assigned scores of 2 and the two carbon atoms *b* and *c* are assigned scores of 3. Table 1 gives all of the tetrads defined on the chemical graph and their assigned scores. There are only

Table 1

The geminal and vicinal tetrads of the chemical graph in fig. 3, grouped into sibling families and assigned scores. The biedges are given in standard form

Tetrad	Biedge	Type	Sibling family	Score
$a < b < c < e$	(bc, ae)	geminal	b	4
$a < b < e < c$	(be, ac)	geminal	b	5
$c < b < a < e$	(ab, ce)	geminal	b	5
$a < b < c < d$	(bc, ad)	vicinal	bc	4
$e < b < c < d$	(bc, ed)	vicinal	bc	4

two sibling families, the geminal family defined by the central atom *b* and the vicinal family defined by the center pair *bc*. Had we included the hydrogen, then there would have been another geminal family of siblings having atom *c* for its central atom. Table 1 indicates that geminal tetrad $a < b < c < e$ has the lowest score, so its biedge (bc, ae) is included in the stereograph. Since both sibling vicinal tetrads have the same score, both of their biedges (bc, ad) and (bc, de) are included in the stereograph. In this case, the geminal tetrads could not be further resolved by the number assigned to the noncentral atoms. However, in a molecule having a bromochlorofluoromethyl group, the latter number ensures that the associated geminal tetrads have nonterminal atoms for centers.

3.6. SOME EXAMPLES OF MOLECULAR AND CONFORMER STEREOGRAPHS

Figures 6–8 depict some chemical stereographs satisfying the sibling selection rules that illustrate a variety of points. Figure 6 gives the molecular stereographs of

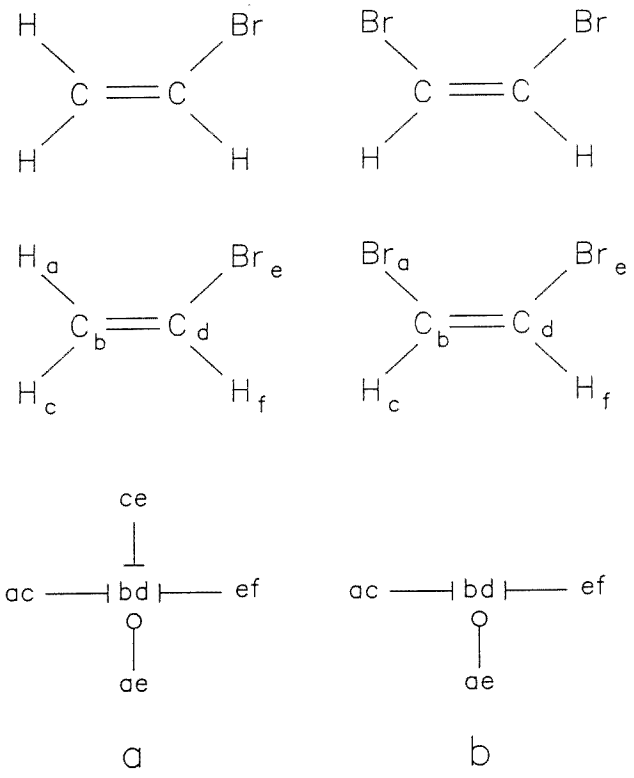


Fig. 6. The molecular stereograph of bromoethene and cis-dibromoethene.

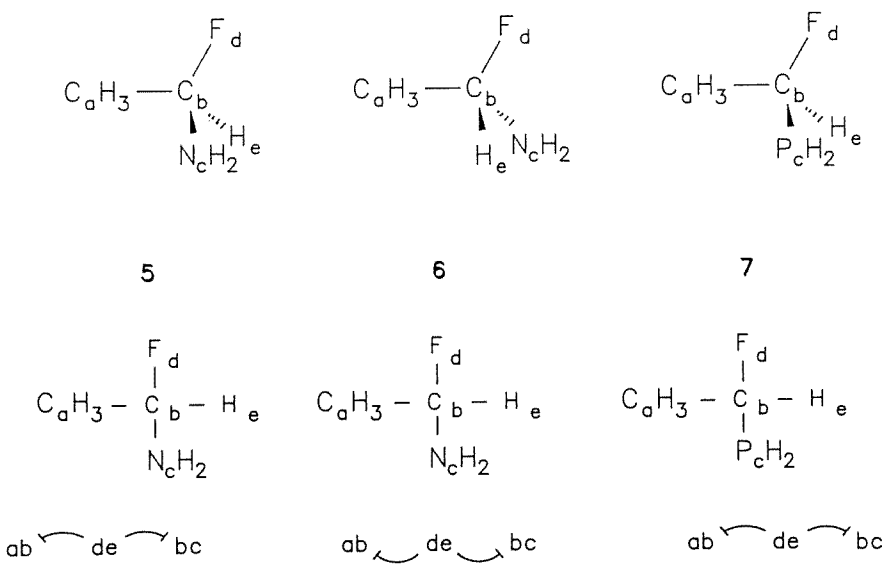


Fig. 7. The molecular stereograph of three 1-substituted fluoroethanes.

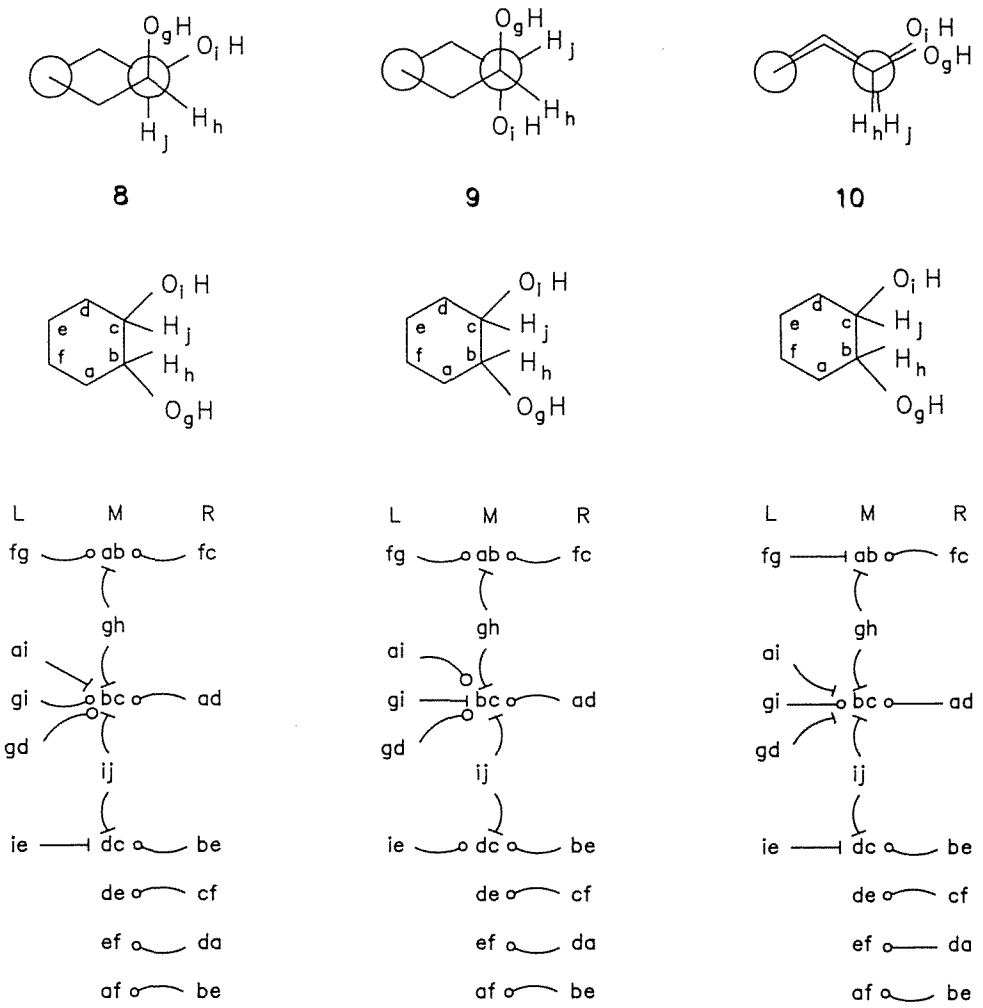


Fig. 8. Conformational stereographs of three dihydroxycyclohexanes.

bromoethylene and cis-dibromoethylene. The chemical graph of cis-dibromoethylene in fig. 6(b) is aesthetically drawn to suggest the cis arrangement. However, the drawing does not mean that the chemical graph itself possesses any information concerning the cis versus trans arrangement. The information resolving the cis–trans distinction is entirely contained within the synperiplanar tetrad (*bd, ae*).

As molecular stereographs, both stereographs in fig. 6 must contain biedges representing all fixed geminal tetrads. We see from both stereographs that the geminal tetrads are all antiperiplanar. Since the vicinal tetrad is also planar, it is easily deduced that all six atoms of dibromoethylene are planar.

Figure 6 also illustrates how our sibling selection rules govern the inclusion of biedges in the stereograph. For bromoethylene, the geminal biedge (*bd, ef*) has a score

of 1, whereas the sibling biedge (*df, be*), which shares center *d* for the central atom, has a score of 4. Consequently, (*df, be*) is not included in the stereograph. Similarly, the vicinal biedge (*bd, cf*) has a score of 0, whereas its sibling (*bd, ce*), which shares the center pair *bd*, has a score of 1. Consequently, (*bd, cf*) is not included in the stereograph.

Figure 7 returns to the stereochemical problem posed by the molecules **1**–**3**, the 1-substituted-1-fluoroethanes. In this figure, the methyl, amino and phosphino groups are treated as ligands whose indices are associated with the central atom. We first note that the ordering invariance rule applies to the two tetrads sharing atoms *d* and *e* as ends. A comparison of the first two stereographs in fig. 7 clearly indicates that the two molecules differ only in their handedness. A comparison of the first and third stereographs reveals that the amino and phosphino groups indexed by *c* are in the same position relative to the atoms indexed by *b*, *d* and *e*. Thus, the phosphino group is in the same relative position as the amino group.

This comparison was facilitated by the fact that the ordering invariance rule applied. In the next section, we will illustrate how such comparisons are made on 1, 1-disubstituted-1-fluoroethanes in which the hydrogen is replaced by another halogen. In this case, the ordering invariance rule does not apply.

In fig. 6, all general and vicinal tetrads are fixed so that both stereographs are molecular as well as conformational stereographs. On the other hand, the stereographs in fig. 7 are molecular, but not conformational stereographs.

Figure 8 presents some conformational stereographs associated with three 1, 2-dihydroxycyclohexanes. Again, we note that all depicted tetrads satisfy the ordering invariance rule. In the discussion of the stereographs in fig. 8, it is informative to compare the accessibility of the structural information contained in the Newman projection with the accessibility of similar information in the stereographs.

We begin by showing how the stereographs differentiate the stereochemistry of the chiral centers. These distinctions are confined to the geminal tetrads in the M columns. Clearly, the left-handed (*dc, ij*) geminal tetrad with central atom *c* in the middle stereograph differs from the corresponding right-handed tetrads of the first and third stereographs. On the other hand, all three (*bc, gh*) geminal tetrads with central atom *b* are right-handed. It follows that the relative handedness of the three molecules is the same at the chiral center of carbon *b*, but not at carbon *c*. These observations are also quite apparent from the Newman projections.

The ring conformations are represented by tetrads whose ends are given in the R column. Both the Newman projections and the stereograph representations clearly indicate that **8** and **9** agree in their ring conformation, but differ from **10** in that regard. The Newman projections present these distinctions as distinctions in overall shape, chair versus boat. The stereographs present these distinctions as sequences in handedness changes in the dihedral angles as one moves around the ring. In the chair conformation, this sequence is left, right, left, right, left, right. For the boat conformation, this sequence is right, degenerate, left, right, degenerate, left. In every case, the vicinal tetrads defining the ring have cis clinicity.

The arrangement of the two hydroxyl groups relative to each other is distinguished by the single (*bc*, *gi*) tetrad in column L. The synclinal(+), antiperiplanar, and synperiplanar arrangements are easily distinguished by the stereocity of the respective tetrads for the three structures. The Newman projections are designed to make such observations transparent.

Finally, the arrangement of the hydroxyl groups relative to the arrangement of the ring atoms is given by the remaining vicinal tetrads (*ab*, *fg*), (*bc*, *ai*), (*bc*, *gd*), and (*cd*, *ie*). These vicinal tetrads have both a ring atom and an oxygen atom as ends. The stereographs clearly differ in the stereocities of these tetrads. To discuss this difference, we shall say a tetrad is *intermediate* to two other tetrads if the former shares a center with each of the latter two tetrads and the latter two tetrads share no centers with each other. The first stereograph has an antiperiplanar tetrad (*bc*, *ai*) and a synclinal(−) tetrad (*bc*, *gd*) intermediate to a synclinal(+) tetrad (*ab*, *fg*) and another antiperiplanar tetrad (*cd*, *ie*). The second has two synclinal(−) tetrads intermediate to two synclinal(+) tetrads. Finally, the third has an anticlinal(−) tetrad and an anticlinal(+) tetrad intermediate to two antiperiplanar tetrads.

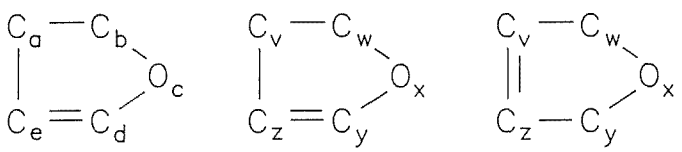
Collective statements of this type are difficult to make from the Newman projection for a number of reasons. Although the Newman projection is designed to convey the nature of particular dihedral angles, it is not designed to do this for very many dihedral angles in a single diagram. For example, the natures of the dihedral angles $\angle fabg$ and $\angle icde$ are not at all obvious from the Newman projection. Moreover, the Newman projection gives a holistic view of structure. The collective statements of the preceding paragraph require that observations on individual dihedral angles be teased apart and then rearranged into a logical structure. This teasing apart and structural rearrangement is essentially what a stereograph is all about. Each isolated tetrad together with its stereocity is analogous to an elemental statement about stereochemical structure. Thus, the stereograph can be thought of as a way of structurally arranging elemental stereochemical statements.

The comparisons with the Newman projection are made to emphasize the point that different representations of structure make different things obvious. Here, we have used the term "obvious" in reference to human perception. If we are to use the computer as an "assistant" in molecular similarity analysis, we require representations that make chemically meaningful facts "obvious" with respect to "computer perception". The stereograph seems to be well suited in the latter regard.

3.7. DETERMINING WHEN TWO CHEMICAL STEREOGRAPHS ARE EQUIVALENT

All the comparisons between stereographs in figs. 6–8 have involved tetrads in which the ordering invariance rule applies. When this rule does not apply, one must use a method of comparison that takes into account different ordering conventions. To do this, we look first at how one ultimately determines when two chemical graphs are equivalent. In a vague sense, one can say two chemical graphs are equivalent if one can define a matching between the atoms that preserves the bonding arrangement. More specifically, let $G = (V, E)$ and $G = (V', E')$ be two chemical graphs. Then, G and G'

are isomorphic if there exists a one-to-one mapping $f: V \rightarrow V'$, which preserves atom and bond types, and such that uv is a bond of G, i.e. $uv \in E$, if and only if $f(u)f(v)$ is a bond of G', i.e. $f(u)f(v) \in E'$. Clearly, the one-to-one correspondence given by $a \leftrightarrow v, b \leftrightarrow w, c \leftrightarrow x, d \leftrightarrow y, e \leftrightarrow z$ defines an isomorphism between the chemical graphs **11** and **12**. This correspondence preserves atom types, i.e. atoms with indices c and x are both oxygens. It preserves bonds and their types, i.e. indices de define a double bond in **11** and the corresponding pair of indices yz define a double bond in **12**. Thus, **11** and **12** are isomorphic. Clearly, no such isomorphism exists between **11** and **13**.



11 12 13

Scheme 3.

The same concept of isomorphism extends directly to stereographs. Roughly speaking, two stereographs are equivalent if one can define a matching between the atoms that preserves atom type, preserves bonds and their types, and preserves tetrads and their stereocities. More specifically, let $S = (V, E, BE)$ and $S' = (V', E', BE')$ be two stereographs with their associated correspondence conventions. A one-to-one function $f: V \rightarrow V'$ is a *biedge isomorphism* between S and S' if f is an isomorphism between the chemical graphs (V, E) and (V', E') of S and S' and if $b_f = (f(a)f(b), f(c)f(d))$ is a biedge of BE' whenever $b = (ab, cd)$ is a biedge of BE and vice versa. Let $T(b)$ and $T(b_f)$ be the tetrads that are corresponded to the biedges b and b_f under the respective correspondence conventions of S and S'. The stereographs S and S' are *isomorphic* if there exists a biedge isomorphism $f: V \rightarrow V'$ such that for every biedge b in BE, $T(b)$ and $T_f(b)$ are isomorphic as tetrads with respect to f restricted to the point set of T.

To see how the definition works, consider the two hydrogen-reduced molecular stereographs of molecule **14** in fig. 9 which, as stereographs of the same molecule, must be isomorphic. There is only one chemical graph isomorphism f which preserves atom types. It is given by $a \leftrightarrow z$, $b \leftrightarrow x$, $c \leftrightarrow w$, $d \leftrightarrow v$, and $e \leftrightarrow y$. Under this isomorphism, the right-handed tetrad $T = c < b < a < d$ is mapped to the left-handed tetrad $T' = v < x < z < w$. However, because the ordering invariance rule does not apply, this does not mean the tetrads are not handedly isomorphic. To make this determination, we must compare the T'-ordering $v < x < z < w$ of the set $\{v, w, x, z\}$ to the ordering of the set induced by the restriction $f|P(T)$ of f to the point set $P(T) = \{a, b, c, d\}$. Since $T = c < b < a < d$, the ordering induced on $\{v, w, x, z\}$ is given by $f(c) < f(b) < f(a) < f(d)$ or $w < x < z < v$. This latter ordering is an odd permutation of the T'-ordering. Thus, T and T' are handedly isomorphic with respect to $f|P(T)$, where $f|P(T)$ denotes the function f restricted to the point set $\{a, b, c, d\}$ of T. Since T and T' also have identical

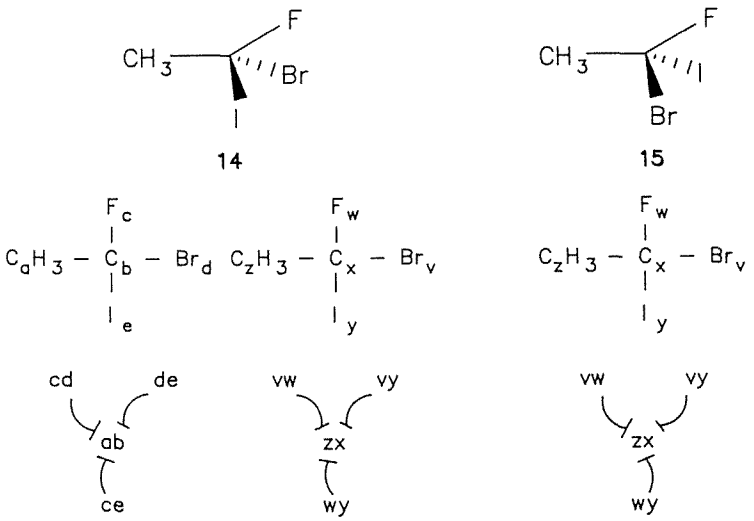


Fig. 9. The molecular stereographs of two 1, 1, 1-bromofluoroiodoethane isomers.

clinicities, they are isomorphic with respect to $f \mid P(T)$. Because this is true for each of the three biedges of the stereographs of **14**, the stereographs are isomorphic.

To show that the first stereograph of molecule **14** is not isomorphic to the stereograph of **15**, we note first that there is only one chemical graph isomorphism, which must be the one in the preceding paragraph since the two molecules are stereoisomers. This time, the right-handed tetrad $T = c < b < a < d$ is mapped to the right-handed tetrad $T' = v < x < z < w$. As before, the T'-ordering $v < x < z < w$ of the point set $P(T')$ is an odd permutation of the ordering of $P(T')$ as defined by f, namely $f(c) < f(b) < f(a) < f(d)$ or $w < x < z < v$. It follows that T and T' are not sterically isomorphic and, consequently, the two stereographs are not isomorphic with respect to this chemical graph isomorphism. Since there is no other chemical graph isomorphism, the two stereographs are not isomorphic.

In comparing these two stereographs, we have implicitly been using our ordering convention for the correspondence convention required for each stereograph. The definition of stereograph equivalence does not require that the two stereographs be constructed using the same correspondence convention. The definition only requires that one know the ordering of the tetrad that corresponds to each biedge.

4. Application areas of stereographs founded on the concept of isomorphism

The utility of a representation reflects the variety and significance of its applications. Here, we shall briefly discuss three that aid in the evaluation of the significance of the rules for including and excluding tetrads from the molecular and conformational stereographs.

4.1. STRUCTURAL REPRESENTATION

A structural representation determines the class of those chemical entities having the same or isomorphic representations. This class is called an equivalence class. Such equivalence classes and their representation underlie how we recognize and talk about the chemical entities of interest to us. They have a long history [32] and well developed mathematical theory [33]. Thus, we should begin by noting that the concept of a stereograph isomorphism sets up the necessary equivalence classes for any chemical entity that one might represent by a stereograph.

The canonical sequences discussed in the introduction provide a general method of assigning a unique name to an equivalence class. The methods in the cited references can be used to assign a unique canonical sequence to a chemical stereograph. In particular, if we view a stereograph as a special case of a complex, the canonicalization scheme in [34] is directly applicable. This canonical sequence can, in turn, be converted to a single number or molecular ID if desired. Thus, stereographs can serve as a basis for registering and "naming" [35] either molecules or conformers.

Although canonical sequences are useful concepts in computational chemistry, they do not have the visual impact of the stereographic diagrams displayed in the figures. We have seen how the sibling selection rules aid the visual perception of a stereograph by reducing the number of tetrads. Our diagrams were designed to convey an understanding of a stereograph and how its stereochemical elements function together. If one's primary goal is to bring a holistic picture of the molecule or conformer to mind, then further reductions in the redundancies in the stereographs should give rise to perceptually more efficient stereochemical and conformational diagrams.

We have focused on two types of molecular entities: structural formulas and conformers. In doing so, we have let the vertices and edges of the chemical graph denote atoms and covalent bonds, respectively. Such interpretations are not essential to either stereographs or to the formalism being developed. The vertices, edges and tetrads of a stereograph should represent whatever structural elements best characterize the chemical entity under study. However, if the underlying graph of the stereograph is no longer a chemical graph, the redundancies amongst a family a sibling tetrads may no longer suffice as a basis for reducing the number of included tetrads. In addition, if one considers tetrads other than geminal and vicinal tetrads, the ordering convention developed in section 3.2 may no longer suffice for the correspondence convention that maps each biedge of the embedding graph to a unique tetrad.

4.2. MOLECULAR SYMMETRY ANALYSIS

We have used the concept of an isomorphism to determine if two chemical stereographs S and S' are isomorphic. In this section, we consider the case in which S and S' denote the same stereograph.

An atom x is automorphic to an atom y in a particular stereograph S if there exists an isomorphism between S and itself that maps x to y. One can show that if x is auto-

morphic in S to y, then y is automorphic to x in S. Thus, we shall simply say x and y are *automorphic* in S.

One can show that methyl hydrogens are always automorphic to each other in the molecular stereograph. For example, the stereograph of monosubstituted methane is given in fig. 10(a). To show that H_a is equivalent to H_b in that molecule stereograph,

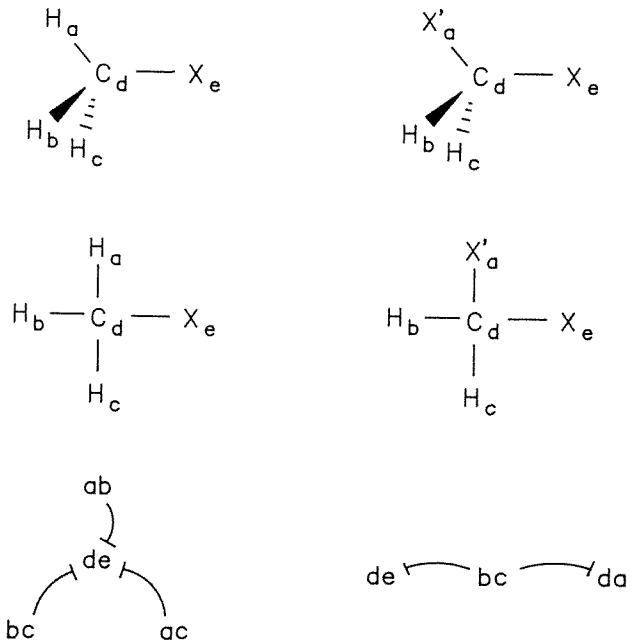


Fig. 10. The molecular stereographs of monosubstituted and disubstituted methane, illustrating the concept of hydrogen equivalence.

consider the biedge isomorphism f given by $a \to b, b \to c, c \to a, d \to d, e \to e$. Under f, the left-handed biedge (de, ab) with associated ordering $a < d < e < b$ is mapped to the biedge (de, bc) having the same handedness. The ordering of the point set $\{b, c, d, e\}$ under the latter biedge is given by $b < d < e < c$. The ordering induced by f on this set is given by $f(a) < f(d) < f(e) < f(b)$ or $b < d < e < c$. Since f does not change the ordering and since the two biedges have the same handedness, the two biedges are handedly isomorphic. Similarly, the left-handed biedge (de, bc) is mapped to the biedge (de, ac) having opposite handedness. The ordering of the point set $\{a, c, d, e\}$ under the latter biedge is given by $a < d < e < c$. The ordering induced by f is given by $f(b) < f(d) < f(e) < f(c)$ or $c < d < e < a$, an odd permutation of $a < d < e < c$. Thus, the latter two edges are handedly isomorphic. Similarly, one shows that the right-handed biedge (de, ac) is handedly isomorphic to the left-handed biedge (de, ab). Since all of the tetrads have trans clinicities, f is an automorphism. Thus, H_a is equivalent to H_b.

On the other hand, the two hydrogens in disubstituted methane are not equivalent if the two ligands are not identical. In this case, there are only two biedge isomorphisms, namely the identity mapping and the mapping f given by $a \rightarrow a$, $b \rightarrow c$, $c \rightarrow b$, $d \rightarrow d$, $e \rightarrow e$. Under f, the biedge (de, bc) is mapped to itself. This biedge orders its point set according to $b < d < e < c$. However, the mapping orders this point set according to $f(b) < f(d) < f(e) < f(c)$ or $c < d < e < b$, an odd permutation of the ordering $b < d < e < c$. Thus, (de, bc) is not handedly isomorphic to itself under f. It follows that f is not an automorphism. Consequently, H_a is not equivalent to H_b in disubstituted methane.

Resolving issues of atom equivalence mathematically or computationally requires the higher resolution of stereochemistry afforded by the ordering explicitly incorporated into the tetrad. If one requires chiral centers to assign stereochemical features to atoms, then two methylene hydrogens are necessarily treated symmetrically. This is not the case when using molecular stereographs.

The set of sibling tetrads included in the molecular stereographs of fig. 10 were determined by the sibling selection rules in section 3.6. The particular set of tetrads that satisfied these rules gave rise to a simple and direct relationship between the chemical equivalence of two hydrogens and the automorphic equivalence of vertices on a molecular stereograph. Such would not have been the case used had the sibling selection rules been designed to exclude all but one tetrad from each sibling family. For example, had only the biedge (de, ab) been included in the first embedding graph of fig. 10, then one could not directly show that H_c is equivalent to either of the other two hydrogens, since H_c would not be in a tetrad while the other two hydrogens would. This does not mean that hydrogen equivalence could not be defined in molecular stereographs involving only one sibling from each sibling family. It just means that the connection between the chemical and mathematical concepts underlying hydrogen equivalence would be less direct.

It is easy to establish that the set of automorphisms on a stereograph constitute a group. These automorphic groups are defined for complexes in general in [27], but are relevant to our special case of stereographs. Here, we only note the existence of this important research area and its potential relationship to the group-theoretic approaches briefly discussed in the introduction.

4.3. STEREO-SPECIFIC SUBSTRUCTURE SEARCHING

Another variant of the concept of an isomorphism underlies a stereo-specific definition of a substructure. Specifically, we call S a *labeled substereograph* of a stereograph S' if every vertex, edge and tetrad in S is also a vertex, edge and tetrad of S'. Now let S and S' be any two stereographs. Then, S is a *substereograph* of S' if S is isomorphic to a labeled substereograph of S'.

Our sibling selection rules lead to a less direct relationship between the chemical concept of a stereo-specific substructure and the mathematical concept of a substereograph. The problem is illustrated in fig. 11. The substructure of the query is represented by structure **16**. The remaining structures represent three molecules in a database. The

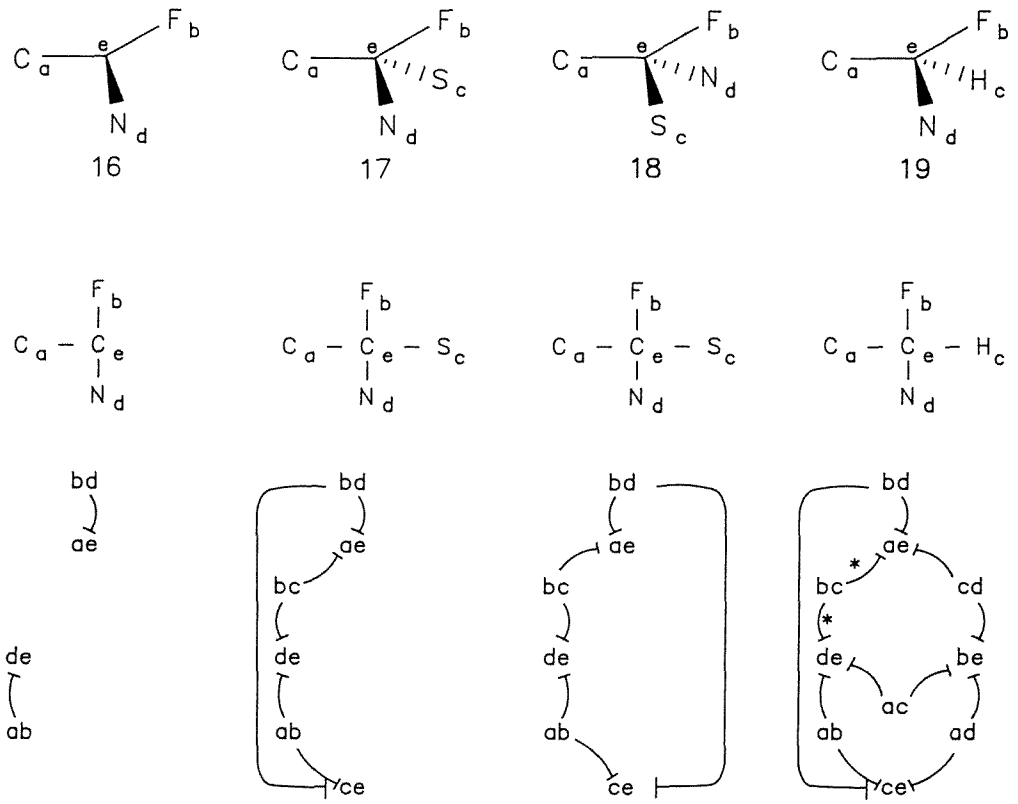


Fig. 11. Molecular stereographs of some substituted 1, 1-aminofluoroethanes. The complete molecular stereograph for **19** is depicted. Only the asterisked biedges would be included in the molecular stereograph satisfying our scoring rules.

methyl hydrogens have not been depicted for these molecules, since they are irrelevant to the argument and their inclusion would only complicate the diagrams. A correct stereo-specific substructure search would list out molecules **17** and **19**, but not **18**.

The molecular stereographs corresponding to substructure query **16** and molecules **17** and **18** were constructed to satisfy the sibling selection rules. However, the molecular stereograph associated with molecule **19** was constructed so that each fixed geminal and vicinal tetrad is not only represented by a biedge in the molecular stereograph, but also corresponds to a biedge. Such a molecular stereograph is called a *complete* molecular stereograph. The complete embedding graph of molecule **17** is identical to that of **19**. The corresponding complete embedding graph of substructure **16** is obtained from that of **19** by simply deleting all biedges except (*be, ad*) and the two biedges that are shown for **16** in the figure. The complete embedding graph of molecule **18** is obtained by reversing the handedness of each tetrad in the embedding graph of **19**.

It is easy to see how the complete molecular stereograph of **16** is a substereograph of the complete molecular stereographs of **17** and **19**, but is not a substereograph of the

the graph-theoretic concept of a stereograph in a manner that should prove useful in other areas of mathematical chemistry.

Acknowledgements

The referees are to be thanked for a number of helpful suggestions, critical comments and useful references. This work was supported by The Upjohn Company.

Appendix: Glossary of terms

Tetrad: An ordered set $a < b < c < d$ of four points. Points a, b, c, and d are referred to as the first end, first center, second center, and second end. If the tetrad is embedded in $\mathbb{R}^3$, angle $\angle abcd$ is the dihedral angle of the tetrad.

Clinicity of a tetrad: A property of a tetrad which is cis, degenerate or trans as its associated dihedral angle is less than, roughly equal to, or greater than 90°.

Handedness of a tetrad: A property of a tetrad which is left-handed, degenerate, and right-handed as its associated dihedral angle is negative, roughly equal to either 0° or 180°, or positive.

Stereocity of a tetrad: A specification of the clinicity and handedness of a tetrad.

Tetral structure: A specification of the two ends and two centers of a tetrad. The four tetrads $a < b < c < d$, $d < b < c < a$, $a < c < b < d$, and $d < c < b < a$ all have the same tetral structure.

Vicinal tetrad: A tetrad whose ends are two vicinal atoms and whose centers define the center bond.

Geminal tetrad: A tetrad whose points define three geminal atoms and a central atom and such that the central atom is one of the tetrad's centers.

Free/fixed tetrad: A tetrad whose stereocity can/cannot change from one conformer to the next.

Ordering convention: A convention based on the chemical graph for selecting one of the four tetrads having the same tetral structure.

Biedge: An ordered pair (ab, cd) which specifies the tetral structure with centers a and b and ends c and d of a tetrad defined with respect to a graph.

Correspondence convention: A convention for setting up a one-to-one correspondence between a set of tetrads and a set of biedges defined with respect to the graph.

Generic embedding graph: The directed graph whose edges constitute a set of biedges defined with respect to another graph, usually a chemical graph.

Embedding graph: A generic embedding graph in which stereocities have been assigned to each of its biedges.

Chemical stereograph: A chemical graph together with an associated embedding graph involving only geminal and vicinal tetrads.

Molecular stereographs: Roughly speaking, a chemical stereograph which specifies the stereocities of all fixed vicinal and geminal tetrads.

Conformer stereograph: Roughly speaking, a chemical stereograph which specifies the stereocities of all vicinal and geminal tetrads.

Sibling tetrads: Two geminal tetrads are siblings if they share the same central atom. Two vicinal tetrads are siblings if they share the same central bond.

Sibling selection rules: A set of rules for deciding which tetrads in a set of sibling tetrads are actually corresponded to, as opposed to being simply represented by, biedges in a chemical stereograph.

References

[1] A. Ugi, Chimia 40(1986)340.
[2] A.T. Balaban (ed.), *Chemical Applications of Graph Theory* (Academic Press, New York, 1976).
[3] M. Randić, J. Amer. Chem. Soc. 99(1977)444.
[4] J.M. Barbard, A.P.F. Cook and B. Rohde, in: *Chemical Information Systems Beyond the Structure Diagram*, ed. D. Bawden and E. Mitchell (Ellis Horwood, Chichester (1990), pp. 29–41.
[5] L. Gann and J. Gasteiger, in: *Software-Entwicklungen Chemie 1*, ed. J. Gasteiger (Springer, Berlin, 1986), pp. 17–33.
[6] W.T. Wipke and T.M. Dyott, J. Amer. Chem. Soc. 96(1974)4834.
[7] H. Beierbeck, J. Chem. Inf. Comput. Sci. 22(1982)215.
[8] W.C. Herndon and J.E. Leonard, Inorg. Chem. 22(1983)554.
[9] A.T. Balaban, O. Mekenyan and D. Bonchev, J. Comp. Chem. 6(1985)562.
[10] J. Blair, J. Gasteiger, C. Gillespie, P.D. Gillespie and I. Ugi, in: *Computer Representation and Manipulation of Chemical Information*, ed. W.T. Wipke, S.R. Heller, R.J. Feldman and E. Hyde (Wiley, New York, 1974), pp. 129–145.
[11] I. Ugi, M. Wochner, E. Fontain, J. Bauer, B. Gruber and R. Karl, in: *Concepts and Applications of Molecular Similarity*, ed. M.A. Johnson and G.M. Maggiora (Wiley, New York, 1990), pp. 239–288.
[12] A. Weise, J. Prakt. Chem. 322(1980)761.
[13] S. Fujita, J. Chem. Inf. Comput. Sci. 27(1987)120.
[14] C.S. Wilcox and R.A. Levinson, in: *Artificial Intelligence Applications in Chemistry*, ed. T.H. Pierce and B.A. Hohne (ACS Symp. Series 316, ACS, 1989), pp. 209–230.
[15] M.A. Johnson, E. Gifford and C.-C. Tsai, in: *Concepts and Applications of Molecular Similarity*, ed. M.A. Johnson and G.M. Maggiora (Wiley, New York, 1990), pp. 289–320.
[16] K. Balasubramanian, Theor. Chim. Acta 51(1979)37.
[17] A. Feldman, J. Organ. Chem. 24(1959)1556.
[18] J.G. Nourse, J. Amer. Chem. Soc. 101(1979)1210.
[19] J.G. Nourse, J. Amer. Chem. Soc. 101(1979)1216.
[20] J. Brocas, M. Gielen and R. Willem, *The Permutational Approach to Dynamic Stereochemistry* (McGraw–Hill, New York, 1983).
[21] I. Ugi, J. Dugundji, R. Kopp and D. Marquarding, *Perspectives in Theoretical Stereochemistry* (Springer, Berlin, 1984).
[22] J.W. McLead, J. Chem. Inf. Comput. Sci. 26(1986)77.
[23] D.H. Rouvray, Chem. Soc. Rev. 3(1974)355.
[24] P. Gillespie, P. Hoffmann, H. Klusacek, D. Marquarding, S. Pfohl, F. Ramirez, E.A. Tsolis and I. Ugi, Angew. Chem. Int. Ed. 10(1971)687.
[25] J.G. Nourse, J. Amer. Chem. Soc. 99(1977)2063.

[26] A.S. Dreiding and K. Wirth, Match 8(1980)341.
[27] P. Floersheim, K. Wirth, M.K. Huber, D. Pazis, F. Siegerist, H.R. Haegi and A.S. Dreiding, in: *Symmetries and Properties of Non-Rigid Molecules: A Comprehensive Survey*, ed. J. Maruani and J. Serre (Elsevier, Amsterdam, 1983), pp. 59–80.
[28] V. Prelog and G. Helmchen, Angew. Chem. Int. Ed. 21(1982)567.
[29] E.F. Meyer, Jr., J. Comput. Chem. 1(1980)229.
[30] E.M. Mitchell, P.J. Artymiuk, D.W. Rice and P. Willett, J. Mol. Biol. 212(1990)151.
[31] G.M. Maggiora, P.G. Mezey, B. Mao and K.C. Chou, Biopolymers 30(1990)211.
[32] D.H. Rouvray, in: *Concepts and Applications of Molecular Similarity*, ed. M.A. Johnson and G.M. Maggiora (Wiley, New York, 1990), pp. 15–42.
[33] R. Rosen, in: *Concepts and Applications of Molecular Similarity*, ed. M.A. Johnson and G.M. Maggiora (Wiley, New York, 1990), pp. 369–382.
[34] K. Wirth, J. Chem. Inf. Comput. Sci. 26(1986)242.
[35] M. Randić, J. Chem. Inf. Comput. Sci. 24(1984)164.
[36] M.A. Johnson, J. Math. Chem. 3(1989)117.
[37] G.M. Maggiora and M.A. Johnson, in: *Concepts and Applications of Molecular Similarity*, ed. M.A. Johnson and G.M. Maggiora (Wiley, New York, 1990), pp. 1–14.