

Partial Orderings in Chemistry

D. J. Klein^{*,†} and D. Babić[‡]

Texas A&M University—Galveston, Galveston, Texas 77553-1675, and Rugjer Bošković Institute, HR-10001 Zagreb, Croatia

Received December 31, 1996[®]

The cosmopolitan relevance of partially ordered mathematical structures in chemistry is argued. Many examples are briefly noted, including those involving chemical periodicities, reactivities, aromaticities, electronegativities, molecular branching, molecular shapes, symmetries, complexities, curve fittings, and more. A few fundamental theorems concerning metrics (or distance functions) on partially ordered sets are noted, first for the intuitively appealing “scaled” posets, then for the more general “transformed” posets. Interspersed along the way are a few examples which are developed to a greater extent, including Randić–Wilkins periodicity for alkanes; the general concept of aromaticity; molecular branching; least-squares fittings; and (most extensively) molecular shapes, chiralities, and symmetries. In each of these types of examples clarifications, alternative views, and extensions of previous works result.

1. INTRODUCTION

In chemistry (and elsewhere) there now and again is encountered a characteristic to which various investigators ascribe different numerical values for the same molecule. As examples (in chemistry) one might mention “electronegativity” or “aromaticity” or “acidity” or “reactivity”. Often the ambiguity in ascribed numerical values does not seem to be due to imprecision in experiment—but rather the ambiguity is imagined to be due to some conceptual “shortcoming” which if only rectified would lead to “true” or “correct” values for the particular characteristic. But there is another possibility—namely that such characteristics are not strictly numerical quantities. Perhaps many of these characteristics do not lead to a total ordering of the objects (e.g., molecules) being so characterized. That is, the various ascriptions of numerical values by different investigators would be just different partly faithful representations of an underlying “partial ordering”. Perhaps then the idea of partial ordering is a fundamental concept cosmopolitantly implicit throughout much of chemistry—and perhaps too it is of relevance throughout many other scientific fields.

Thence the general idea of partial orderings should be considered, developing some rather general features and noting a fair number of chemical examples, ranging somewhat beyond even the examples already alluded to. Indeed such partial-ordering chemical ideas are found implicitly in the works of a number of authors usually in the context of a special application, e.g., for degrees of molecular “branching”, for degrees of “chirality”, or for molecular “similarity”, or for aromaticity, or for chemical “periodicity”. And beyond the bounds of chemistry the idea of partial orderings should have application, the abstract idea presumably being traceable in some form back to W. Leibniz, as noted in G. Birkhoff’s seminal book¹ giving much prominence to the mathematics of posets (and refinements therebeyond). G.-C. Rota has proposed² that partial orderings are at the heart of the whole field of combinatorics (and indeed this idea seems to be becoming accepted). In chemistry the relevance

of partially ordered sets has been particularly persistently studied by Ernst Ruch and co-workers^{3–5} as well as by Milan Randić and co-workers.^{6–12} For some other work see refs 13–15. In Eigen & Winkler’s popular book¹⁶ *Laws of the Game* (concerning the nature of science and the principles of nature) ideas concerning partial orderings are given some note.

Here attention is directed: first to several types of partial orderings of presumed chemical relevance; and second to some general types of numerical functions on them. Sections 2 and 3 introduce some initial concepts and first examples. Much is developed in sections 4–8 for the so-called “scaled” (or “scaleable”) partial orderings, including the oft-realized case of “mimicks”, overall expanding somewhat on earlier work.¹⁵ For these mimicks section 5 indicates how linear dependence questions thereon may be addressed, and section 6 applies these techniques in the context of the example of “aromaticity” viewed as a partial ordering. Section 7 concerns metrics (or distance functions) on scaled partial ordered sets, and section 8 considers applications to molecular shapes, particularly as regards symmetry and chirality measures for shapes. Thereafter (in sections 9–14) it is described how many of the concepts may be extended to the more general case of partially ordered sets with a general type of “transformational” structure, again leading to metrics and consequent dissimilarity measures on partially ordered sets. Section 10 discusses molecular branching in the framework of transformed partial orderings; sections 11–13 focus further on certain discretized characterizations of form and shape (to obtain some novel modifications to earlier introduced measures of form and shape); and finally section 14 indicates a possible relevance for curve fitting. Many particular results are placed in a more general framework—and often the earlier particular results are extended.

2. PARTIAL ORDERINGS—POSETS

An informative approach to partial orderings starts with a simple yet nontrivial example, which involves a set of (bounded connected) geometric “regions” of space, such as for a molecule identified say as the region encompassed by its van der Waals surface. Such a set of regions is partially

[†] Texas A&M University—Galveston.

[‡] Rugjer Bošković Institute.

[®] Abstract published in *Advance ACS Abstracts*, May 15, 1997.

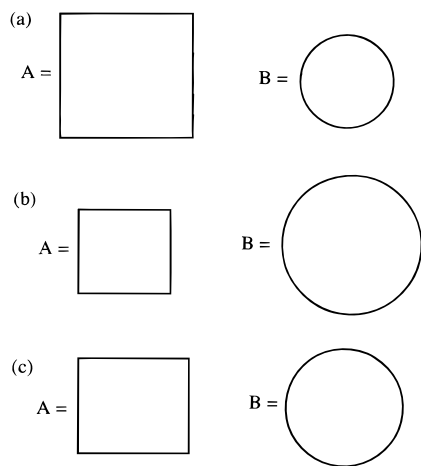


Figure 1. Comparisons of a square (including its interior) and a disk: (a) the square covers the disk; (b) the disk covers the square; and (c) neither can cover the other.

ordered by the operation of “covering”: one region A covers another B if the boundary of A can be translated and rotated so that it entirely encloses B . A convenient two-dimensional illustration is given in Figure 1, where two regions (one circular and the other square) are compared. Clearly the set of regions is no more than partially ordered—there are pairs of regions neither of which covers the other, though sometimes one covers the other, and whenever both members of a pair cover the other, they must be equivalent. If A covers B and B covers C , then A covers C .

A more formal definition of a partial ordering may be made. A set \mathcal{P} of objects is said to be *partially ordered* if there is a pairwise relation between elements of \mathcal{P} , such that this relation, here denoted by \geq , satisfies three conditions

$$\begin{aligned} A &\geq A \\ A &\geq B \text{ and } B \geq A \Rightarrow A = B \\ A &\geq B \text{ and } B \geq C \Rightarrow A \geq C \end{aligned} \quad (2.1)$$

where $A, B, C \in \mathcal{P}$. A partially ordered set is also called a *poset*. The previous paragraph's example involving the covering operation conforms to these conditions. Another standard mathematical example is that of the subset relation (usually denoted by \subseteq), and yet another is the subgraph relation (also usually denoted by \subseteq)—and the significance of this partial ordering (and associated Möbius functions) for both chemistry^{13,14} and statistical mechanics¹⁷ has been noted elsewhere. Another partial ordering, that of “majorization”, has been^{3–6} emphasized for use in chemistry and physics, though it is even more widely studied¹⁸ in other contexts—it will be touched upon in section 10 here.

There are a few common points of additional notation. The relation $A \geq B$ is also equivalently expressed as $B \leq A$, and also if $A \geq B$ and $A \neq B$, then $A > B$. A standard sort of diagrammatic representation of \mathcal{P} (particularly for discrete \mathcal{P}) is the so-called *Hasse diagram*, obtained as a diagram of nodes and line segments such that each node corresponds to an element $A \in \mathcal{P}$ arranged such that the node for any next (i.e., immediately) greater element $B \in \mathcal{P}$ is located *above* the A node and the two nodes are connected by a line segment. For example, the Hasse diagram of the poset of subsets of the set $\{1,2,3\}$ is shown in Figure 2.

Partial orderings may be viewed as a first weakening of the usual *total* ordering, which is such that for every pair of

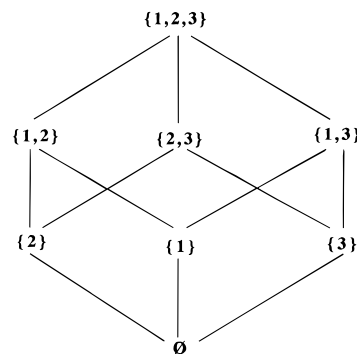


Figure 2. Hasse diagram for the poset of subsets of the parent set $\{1,2,3\}$ under the subset partial ordering.

elements $A, B \in \mathcal{P}$ it must be that $A \geq B$ or $B \geq A$ (or both, if $A = B$). The standard total ordering is that of “greater than or equal to” on the real numbers \mathcal{R} . And of key relevance is the fact that generally totally ordered sets may be represented by such real numbers in a manner which is *faithful*, in the sense that there is a mapping f from \mathcal{P} to \mathcal{R} such that for every $A \in \mathcal{P}$ there is a unique corresponding $f(A) \in \mathcal{R}$ with $A \geq B$ if $f(A) \geq f(B)$. In such a case $f(\mathcal{P})$ is the representation of \mathcal{P} , and the property can reasonably be identified with f . Then the property might be termed *numerical*. Indeed, Campbell¹⁹ has argued that properties being so numerical is a key prerequisite for doing (at least much of) science. However, the circumstance of nontotally ordered posets should not be precluded if found to be of use—and already a number of examples of some interest have been noted. In Birkhoff's words (on page 2 of ref 1) “The world around us abounds with examples of partially ordered sets”, and he goes on to give a number of mathematical examples such as the subset partial ordering on the set of subsets of a parent set and the divisor partial ordering on the set of positive integers. In the next section a class of chemically fundamental posetic examples are proposed.

3. CHEMICAL PERIODICITY, MULTIPOSETS, AND ISOTONICITY

Rather fundamental in chemistry is the idea of “periodicity”. The most significant example of chemical periodicity is that of Mendeleev as applied to the chemical elements. But there are many other examples of such systematic regularities (or periodicities), though often not presented so explicitly as in a “periodic table”. In some cases even other than that of Mendeleev there are explicitly presented “periodic tables”—one such example being that of Randić and co-workers’^{7–11} periodic tables of alkanes. These tables are constructed utilizing the numbers p_2 and p_3 of paths of lengths 2 and 3 in the hydrogen-deleted carbon network of the alkanes of a given composition. A two-dimensional grid is constructed with p_2 and p_3 values plotted in horizontal and vertical directions, and the different realizable molecular networks are located at the nodes of the grid, as in Figure 3 for the case of the nonanes. The general idea is that various physical properties vary in a gradual monotonic manner along the horizontal and vertical grid lines. Such results are illustrated in Figure 4 for the cases of molar entropy (at 25 °C) and of the boiling point (at atmospheric pressure). Many other properties exhibit similar behaviors, though the directions of monotonicity may be different for different properties.

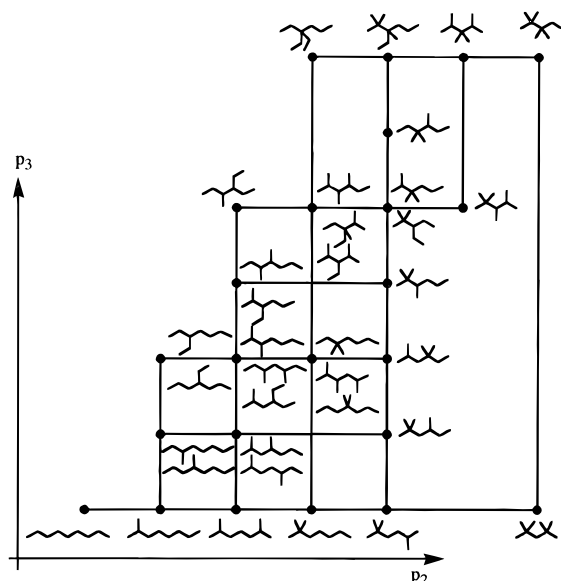


Figure 3. The Randić–Wilkins grid for nonanes.

Such periodic tables are reminiscent of the Hasse diagrams of partial orderings. Generally it is proposed¹⁵ that a periodic table is a *multiposet*, viewable as a network with nodes representing chemical species and the same edges of the network to appear in different associated Hasse diagrams identified with different properties of the chemical species. That is, the same network in different orientations yields different Hasse diagrams (each identified to different partial orderings). For example, the two Hasse diagrams identified with the entropy and boiling points of the (network of the) nonane periodic table are as in Figure 5. There too may be other Hasse diagrams identified with other physical properties. But note should be made of a slight complicating factor, where in some periodic tables (as in the present) more than one chemical species may occupy the same position (e.g., as 2- and 3-methyloctane in the table of nonanes), whence the second (so-called antisymmetric) definitional condition

(concerning the implication that $A = B$ if both $A \geq B$ and $B \geq A$) is not met. With the lack of satisfaction of this condition the set is not strictly a partial order but is more properly described as a *quasi-order*. But also rather than viewing the elements of the poset to be equivalence classes of structures rather than individual structures—that is, the species occupying the same position in a Hasse diagram are described as equivalent, whence there results a proper partial ordering on the set of equivalence classes.

In general chemical periodicity may be viewed as an identification of which different properties identify with which different Hasse diagrams (i.e., which different partial orderings) there are associated to the periodic table under question. Mendeleev's periodic table offers a more elaborate example of multiposet structure, with relations sometimes following between A and B families and sometimes not, as discussed elsewhere.¹⁵

Often properties under question may be “numerical” so that partial-ordering faithful mappings from the partially ordered set \mathcal{P} to the real numbers \mathcal{R} typically are of interest. In a more precise language a mapping f from a poset \mathcal{P} to \mathcal{R} is *isotonic* if

$$A, B \in \mathcal{P} \text{ and } A \geq B \Rightarrow f(A) \geq f(B) \quad (3.1)$$

Thence to quantitatively fit or expand a property identified as faithful to a poset \mathcal{P} one might desire a set of isotonic functions. In the case that the periodic table is obtained in terms of certain primitive properties (such as p_2 and p_3 in the considered case of nonanes) it is natural to seek (for the purpose of making property expansions) standard sets of isotonic functions for each poset \mathcal{P} associated with the multiposet (i.e., with the periodic table). In this regard the methods of the next two sections are relevant.

4. SCALED POSETS, SCALEABILITY, AND MIMICKS

Many posets \mathcal{P} have additional mathematical structure. One particular type¹⁵ is associated with the feature that the

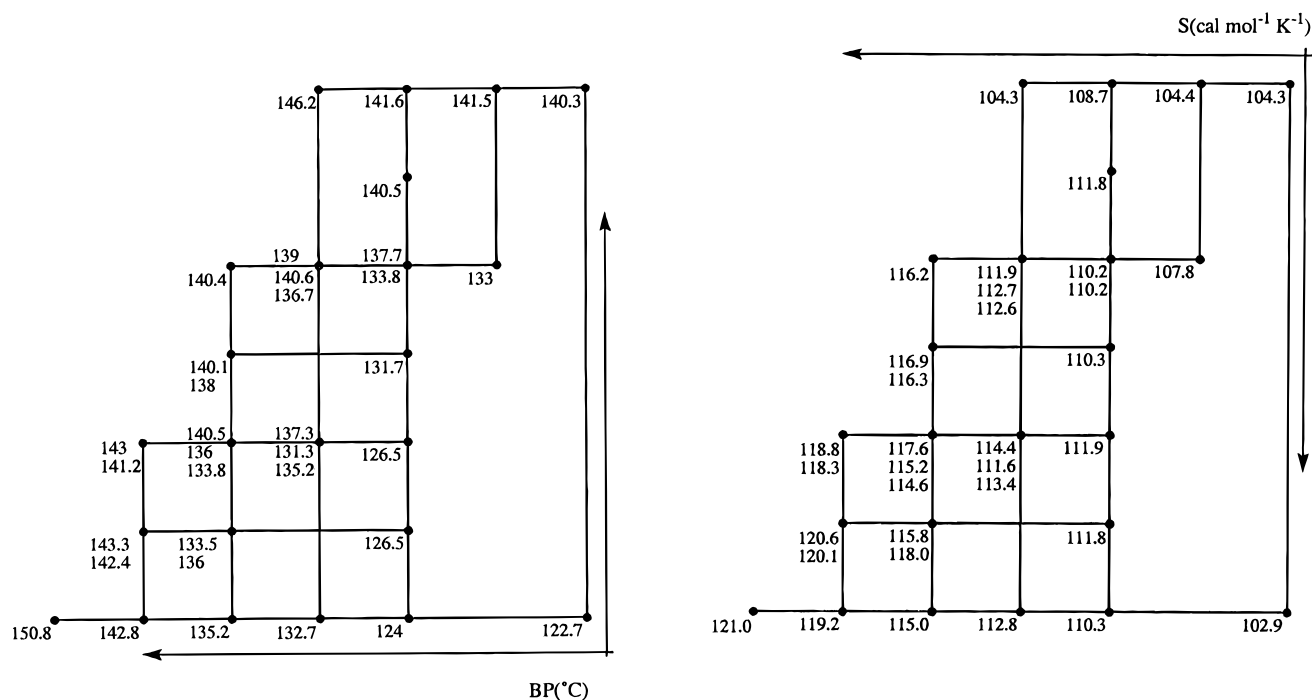


Figure 4. Property values for nonanes entered on the Randić–Wilkins grid: (a) boiling points (°C) and (b) molar entropies (cal/mol·K). The arrows indicate directions of increase of the properties.

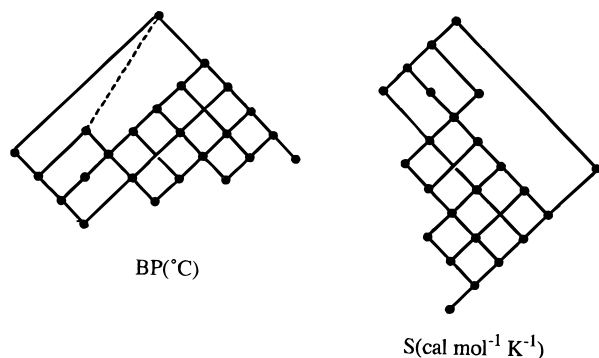


Figure 5. The Randić–Wilkins grids reoriented to Hasse diagrams appropriate for (a) boiling point and (b) molar entropies.

positive reals \mathcal{R}_+ “multiply” onto the members of \mathcal{P} to give other members of \mathcal{P} . More precisely a poset \mathcal{P} is a *scaled* poset if there is a scalar multiplication by members of \mathcal{R} such that

$$a \in \mathcal{R}_+ \text{ and } A \in \mathcal{P} \Rightarrow a \cdot A \in \mathcal{P}$$

$$a, b \in \mathcal{R}_+ \text{ and } A \in \mathcal{P} \Rightarrow (ab) \cdot A = a \cdot (b \cdot A)$$

$$a, b \in \mathcal{R}_+, A, B \in \mathcal{P}, a \geq b, \text{ and } A \geq B \Rightarrow a \cdot A \geq b \cdot B \quad (4.1)$$

An example of such a scaled poset is again provided in the context of the covering relation of figure 1—the multiplication by a scalar c corresponds to ordinary magnification by a factor c . Indeed this example extends directly to the treatment of molecular conformations as represented by nuclear positions and as explained in more detail soon in sections 7 and 8—and too it extends to biotaxonomic applications.²⁰ Other important examples of scaled posets involve comparisons of matrices (and their eigenspectra), as indicated elsewhere.¹⁵

Next there is a slightly more extensive class of posets which “in essence” still have the feature of being scaled. More particularly a poset \mathcal{P} is said to be *scaleable* if first, \mathcal{P} is a subset of another poset \mathcal{P}_+ which is itself scaled; and second, this embedding of \mathcal{P} in \mathcal{P}_+ is faithful in the sense that whenever $A, B \in \mathcal{P}$ then A, B are in the same relation as in \mathcal{P}_+ . Thus in the context of the covering partial ordering, a discrete set of regions though not scaled would remain scaleable. But there are less trivial examples as soon to be seen.

Indeed it may be argued (for example) that the discrete periodicity posets of the preceding section are in some sense scaleable. Given a poset \mathcal{P} there generally are several isotonic functions from \mathcal{P} to \mathcal{R}_+ of interest, say collected into a vector $f \equiv (f_1, f_2, \dots, f_M)$, and one may identify with each $A \in \mathcal{P}$ a set $f(A) \equiv (f_1(A), f_2(A), \dots, f_M(A))$. The set $\mathcal{P}' \equiv \{f(A) : A \in \mathcal{P}\}$ is partially ordered under the relation \geq' defined as

$$f(A) \geq' f(B) \text{ if } f_i(A) \geq f_i(B), \quad i = 1 - M \quad (4.2)$$

This new poset \mathcal{P}' imitates \mathcal{P} in some manner and is termed a *mimick* of \mathcal{P} . Presumably \mathcal{P}' is a more faithful mimick the greater the dimension of the space of the functions f .

And the mimick \mathcal{P}' is scaleable, when \mathcal{P}' is identified as having elements

$$h \cdot f(A) \equiv (hf_1(A), hf_2(A), \dots, hf_M(A)) \equiv f(h \cdot A), \quad A \in \mathcal{P}, \quad h \in \mathcal{R}_+ \quad (4.3)$$

Of course, if the $A \in \mathcal{P}$ are molecules and the $f_i(A)$ are numerical molecular properties, the $h \cdot A$ so defined generally do not represent any realizable molecule. But quite generally mimicks are seen to be scaleable.

Within the context of this scaleable mimick, questions as concerns dependences amongst the different f_i then arise and the question of linear dependences in particular. The idea of poset dimensionality is involved in the general dependence question: the *dimension* of a general poset \mathcal{P} is the smallest number of isotonic functions on \mathcal{P} such that the resulting mimick is isomorphic to \mathcal{P} . There are^{21,22} a number of equivalent descriptions of posetic dimension, along with other results, but a related (narrower) concern oft of importance is that of linear dependence and space dimension. To treat this, one generally may utilize powerful matrix-algebraic tools of analysis, e.g., schemes exhibiting some similarity to *principal components analysis*²³ of statistics, with the statistical sample set being identified to \mathcal{P} . Such analyses then aim at an understanding of the suite of correlated properties making up the description of the poset—such ideas being discussed some in the next section.

In the case of Randić’s periodic table of alkanes one can take $f \equiv (p_2, p_3)$, and the dimensionality of the poset is 2. One can surmise that in a number of such cases the posetic dimension d of a mimick is just what a chemist might term the dimension of an associated periodic table, with many other properties given in terms of (its d -fold) position in the table.

Now (as also emphasized previously¹⁵) it seems that often in chemistry it is naught but the mimicks that are available. This might be viewed to be the case with electronegativity, with acidity, with oxidizing power, with various reactivities, etc. A more elaborated example of aromaticity is found in section 6, and also see section 10 on branching.

5. LINEAR DEPENDENCES

Linear algebraic techniques are naturally anticipated to be of value in dealing with poset mimicks which are expressed in terms of vectorial lists of properties for different molecules. These techniques usually entail fundamental assumptions which it is appropriate to note and sometimes appropriately tailor to the situation at hand. An initial such assumption has to do with the vector space. Corresponding to each property f_i one may utilize vectors $|i\rangle$ with A th components $(A|i)$ proportional to the property values $f_i(A)$ for the A th molecule, $A \in \mathcal{P}$. That is, $|i\rangle$ may be viewed as in correspondence with a sequence of numbers with something proportional to $f_i(A)$ in the A th position. These different $|i\rangle$, through the process of making linear combinations, span a vector space, which can have a dimension not necessarily matching the number of components. Now in so forming linear combinations of these vectors (and their components) it is desirable to ensure that they are on “equal footing”. First in order not to bias results toward the largest molecules one might recognize that many indices are size-extensive and so introduce for such properties a factor inversely

proportional to a size measure $|A|$. Next since the different types of indices may have different types of units it is appropriate to rescale them so that the vectors $|f_i\rangle$ become "normalized" (in the sense that their inner products $\langle f_i | f_i \rangle$ as of the next paragraph are 1, though also one could instead arrange that the standard deviation or range is unity). Thence the A th component of the vector $|i\rangle$ is reasonably taken as

$$(A|f_i) \equiv n_i f_i(A)/|A|^{\sigma(i)}, \quad A \in \mathcal{P} \quad (5.1)$$

where n_i is a normalization factor, $\sigma(i)$ is 1 or 0 as property f_i is or is not size extensive, and \mathcal{P} is identified to a set of molecules for which the $f_i(A)$ are the "bare" properties. General vectors $|u\rangle$ then appear as properly comparable linear combinations of the $|f_i\rangle$.

After having so characterized the linear space, a next question concerns an *inner product*, which between vectors $|u\rangle$ and $|v\rangle$ is taken to be of the form

$$(u|v) \equiv \sum_A (A|u) \cdot w_A \cdot (A|v) \quad (5.2)$$

with w_A some positive "relevance" weighting. To make the results depend less strongly upon the sample set \mathcal{P} chosen, one might take w_A inversely proportional to $|\mathcal{P}|$ (at least if all the indices f_i are available for every molecule in \mathcal{P}). Further, not to overly weight highly sampled sequences of similar molecules, one could conceivably include in w_A a factor diminishing the weight given to A if many smaller molecules of similar structure are included in \mathcal{P} —but this we do not currently worry about. Thence the inner product currently proposed is

$$(u|v) \equiv |\mathcal{P}|^{-1} \sum_{A \in \mathcal{P}} (A|u) \cdot (A|v) \quad (5.3)$$

Evidently the inner product makes averages (over $A \in \mathcal{P}$).

Questions of possible linear dependences amongst the spanning vectors may be addressed in terms of the *overlap* matrix \mathbf{M} with elements

$$M_{ij} \equiv (f_i|f_j) \quad (5.4)$$

There are *eigenproperty* vectors $|\lambda\rangle$ defined in terms of columnar eigenvectors \mathbf{C}_λ with eigenvalue λ and components $C_{\lambda i}$, thusly

$$|\lambda\rangle \equiv \sum_i C_{\lambda i} |f_i\rangle, \quad \text{for } \mathbf{M} \mathbf{C}_\lambda = \lambda \mathbf{C}_\lambda \quad (5.5)$$

(The distinctive notation for $|\lambda\rangle$ and \mathbf{C}_λ is of use since they generally have different numbers of components.) Then for any null-eigenvalue eigenvector \mathbf{C}_ϕ to \mathbf{M} the corresponding eigenproperty exhibits a linear dependence. That is, making use of eqs 5.3 and 5.5, one finds

$$\sum_A w_A |(A|\phi)|^2 = (\phi|\phi) = \sum_{ij} C_{\phi i}^* M_{ij} C_{\phi j} = \mathbf{C}_\phi^\dagger \mathbf{M} \mathbf{C}_\phi = 0 \quad (5.6)$$

so that

$$\sum_i C_{\phi i} f_i(A) = (A|\phi) = 0 \quad (5.7)$$

For linear dependence with the allowance of the possibility of the addition of a constant (as is usual in typical statistical analyses), one may include in the set of spanning vectors a *unital* vector $|f_0\rangle$, which has all components $f_0(A) = 1$. The

maximum-eigenvalue eigenvectors of \mathbf{M} naturally give combination indices which most strongly correlate with the partial ordering. These are generally different than those from principal components analysis which may in effect use a slightly different inner product and include the effects of f_0 only by taking modified "residual" properties δf_i with $|\delta f_i\rangle$ constrained to be orthogonal to $|f_0\rangle$. That is, instead of an analysis in terms of properties f_i , principal components analysis is pursued in terms of property residuals $\delta f_i = f_i - (f_0|f_i)^{-1} f_0$. A general framework for elucidation of interrelations between such different linear dependence analyses is found elsewhere.²⁴

One natural use of the linear-dependence analysis is to select a smaller set of relevant indices, which then in turn might be used (much as position in a periodic table) to predict other properties of interest. If there are d larger eigenvalues to \mathbf{M} , then one might seek a set of d of the original indices f_i which best span this large-eigenvalue subspace. This can be neatly done by looking at the corresponding submatrices of \mathbf{M} . That is, for each set \mathcal{S} of d aromaticity indices take the submatrix $\mathbf{M}(\mathcal{S})$ restricted to this subset and determine its eigenvalues, which for a better representation of the full matrix \mathbf{M} should match closely to the larger more significant eigenvalues of \mathbf{M} . More particularly (by standard minimax principle) the ordered eigenvalues $\lambda'_1 \geq \lambda'_2 \geq \dots \geq \lambda'_d \geq 0$ of $\mathbf{M}(\mathcal{S})$ must be bounds for the d largest ones $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d > 0$ of \mathbf{M} , thusly $\lambda_i \geq \lambda'_i$ for $i = 1 - d$. Thence one has a natural measure of the *closeness of match* (for $\mathbf{M}(\mathcal{S})$ to \mathbf{M} or for \mathcal{S} to the full set of indices)

$$m(\mathcal{S}) \equiv |\Pi_i \{\lambda'_i / \lambda_i\}|^{1/d} \quad (5.8)$$

where the eigenproperties $|\lambda_i\rangle$ and $|\lambda'_i\rangle$ are chosen to be normalized (i.e., with $(\lambda_i|\lambda_i) = (\lambda'_i|\lambda'_i) = 1$). A best set \mathcal{S} is then that with the maximum $m(\mathcal{S})$, the other properties being more accurately linearly representable (via near-null-eigenvalue eigenvectors) in terms of those in \mathcal{S} .

6. AROMATICITY

As one example of a partial ordering realized as a mimick one might consider the case of *aromaticity*. Indeed, there are several measures of aromaticity (e.g., as discussed in several reviews²⁵):

- *as excess stability over a local bond-energy-type additivity for heats of formation;
- *as total excess NMR chemical shifts over that ascribed to local atomic environments;
- *as a diamagnetic susceptibility exaltation;
- *as a bond-length deviation sum away from localized (pure) double and single bonds;
- *as excess stability under the influence of various double-bond addition reactions;
- *as an extreme bond-length deviation away from uniform π -bond delocalization;
- *as an extreme of a free valence on one of the π -centers; etc.

An analysis for linear dependence would identify (presumably useful) predictive correspondences amongst the different aromaticity indices. Indeed there have been workers^{26–28} who have taken an integrative approach to seek for correlations within different sets of aromaticity indices. Parkyani and Boniface²⁸ and Krygowski *et al.*²⁹ advocate new optimal

Table 1. Larger Eigenvalues and Eigenvectors to **M**

λ	0.0002	0.0003	0.0014	0.0057	0.0048
$I_{5,6}$	0.1505	-0.2740	0.0531	0.1586	0.4884
RC	-0.5237	0.4239	0.4847	-0.0849	-0.2601
ΔN	-0.1031	-0.1341	0.0195	-0.0304	0.1988
DRE	0.0068	0.0178	-0.1353	0.2283	-0.1199
HSRE	-0.1802	-0.1742	0.3145	-0.2770	0.4258
ΔH_A	0.1456	0.6491	-0.5193	0.0184	0.2125
ΔH_f	0.0653	-0.0217	-0.0696	0.0530	-0.0091
ΔH_f^{MNDO}	-0.0627	0.1072	-0.0755	-0.0713	-0.1593
χ_M	-0.0939	-0.0938	0.0944	0.7454	0.0130
Λ	0.1210	-0.0201	-0.0829	-0.4943	0.1441
α_{aniso}	-0.2798	-0.5023	-0.4555	-0.1578	-0.4464
$-^{15}\text{N}$	0.0179	0.0360	-0.0198	-0.0234	-0.0829
f_0	0.7270	-0.0222	0.3751	-0.0653	-0.4012

linear combinations as preferred aromaticity indices. Katritzky *et al.*²⁶ Jug and Köster,²⁷ and Krygowski *et al.*²⁹ have performed “principal components analyses” with selected samples of molecules and have identified the space of aromaticity indices to be at least two-dimensional. The larger eigenvalues to the (non-negative definite) correlation matrix yield linear-combination indices which these workers view as of more relevance. But the 0-eigenvalue linear combinations may be argued to be of relevance in yielding linear dependences amongst the indices (the linear dependences being only approximate if an eigenvalue is only near 0). Moreover, these 0-eigenvalue combinations may be combined in different amounts with the maximum-eigenvalue combinations to yield new combinations which individually give equivalent large correlations—and from such a variety of combinations there can be sought one to maximize an “optimal” index’s similarity to one of the primitive indices. But the principal components analysis rather than the initial properties deals with what is better described as their components orthogonal to a reference “unital” or “constant” vector, and one may wish not to impose such a constraint in an analysis of the linear dependences.

The analysis for linear dependences discussed in the preceding section might be applied in the present case of aromaticity. The data of Katritzky *et al.*²⁶ for 12 aromaticity indices, as given there for a set \mathcal{P} of 16 molecules, can be analyzed in the current framework. These authors’²⁶ aromaticity indices may be viewed to be from three different groups. The three geometrically based indices are closely related to the average deviation of bond orders, to the coefficient of variation of bond orders, and to the lowest bond order ΔN , respectively. Their five energetic quantities are the Dewar resonance energy DRE, the Hess–Schaad resonance energy HSRE, the experimental heat of atomization ΔH_A , the experimental heat of formation ΔH_f , and the MNDO heat of formation ΔH_f^{MNDO} —all these are size extensive, so we divide by $|A|$. Their four magnetic indices are the diamagnetic molar susceptibility χ_M , diamagnetic susceptibility exaltation Λ , nitrogen-15 chemical shifts ^{15}N , and the one related to anisotropic polarizability α_{aniso} —here the first two are divided by $|A|$. A common problem occurring also with this data set is that occasionally a property value for one of the molecules is missing. In such cases to carry out the linear algebraic analysis discussed in connection with eqs 5.1–5.8 we take for the missing value the average value of the property for all the molecules for which it is available. Then making this analysis, eigenvalues and eigenvectors to **M** are found as reported in Table 1, where again the small eigenvalues (say <0.1) indicate linear dependences, which, e.g., may be used to make predictions

for one property (appearing in the combination) in terms of the others. With the presumption of dimensionality $d = 3$ or 4 (including the constant property) maximal closenesses of match may be determined as

$$m(\mathcal{S}) = 0.3524, \quad d = 3 \text{ for } \mathcal{S} = \{f_0, ^{15}\text{N}, \Delta H_f\}$$

$$m(\mathcal{S}) = 0.4302, \quad d = 4 \text{ for } \mathcal{S} = \{f_0, ^{15}\text{N}, \Delta H_f, \Delta N\} \quad (6.1)$$

Here the listed properties which best span the space of Katritzky *et al.*’s aromaticity indices are of different types: magnetic, energetic, and geometrical types, roughly in consonance with their earlier principal components analyses. From Table 1 it is seen that these lists correlate with the properties making larger contributions to the larger-eigenvalue eigenvectors to **M**. Some of the degree of independence may be due to the feature that the indices of Katritzky *et al.*²⁶ include several properties (e.g., the molar susceptibility, the heats of formation, and especially the heat of atomization) which often are not strictly viewed as aromaticity indices, being total values rather than deviations from a simpler (localized) description—i.e., these total indices undoubtedly include other “0-order” effects due to localized (σ -) bonding. In any event the leading properties of eq 6.1 are determined to be those from which most reliably the other aromaticity indices may be determined, through the combinations indicated in the eigenvectors to **M**. The partial ordering manifests itself through the linearly independent set, with one molecule A being more aromatic than another B if every one of these indices indicates a greater aromaticity.

Here then aromaticity is seen in a positive light as of interest and utility. The view here (and more briefly indicated elsewhere^{15,30}) is that aromaticity is not a single numerical index but rather a collection of posetically interrelated property deviations (away from 0-order expectations) for which nevertheless useful interrelations and correlations occur. That is, aromaticity is a suite of anomalously valued properties showing correlations amongst the anomalies.

7. METRICS ON SCALED POSETS

Often a useful type of function on a set (such as a general poset) of objects is a distance function or metric. Formally a *metric* on a set \mathcal{P} is standardly defined to be a function from $\mathcal{P} \times \mathcal{P}$ onto the non-negative reals such that, for arbitrary $A, B, C \in \mathcal{P}$,

$$d(A, B) = d(B, A) \geq 0$$

$$d(A, B) = 0 \Leftrightarrow A = B$$

$$d(A, B) + d(B, C) \geq d(A, C) \quad (7.1)$$

The prototypical example of a metric is the usual Euclidean distance on an M -dimensional coordinate space—i.e., for positions **x** and **y** with components x_i and y_i , $i = 1 \rightarrow M$, the distance between these two positions is

$$d(\mathbf{x}, \mathbf{y}) \equiv \{\sum_i |x_i - y_i|^2\}^{1/2} \quad (7.2)$$

And another class of (Minkowski) distance functions on coordinate space are of the same form except that the exponents of 2 and 1/2 appearing in (7.2) here are replaced

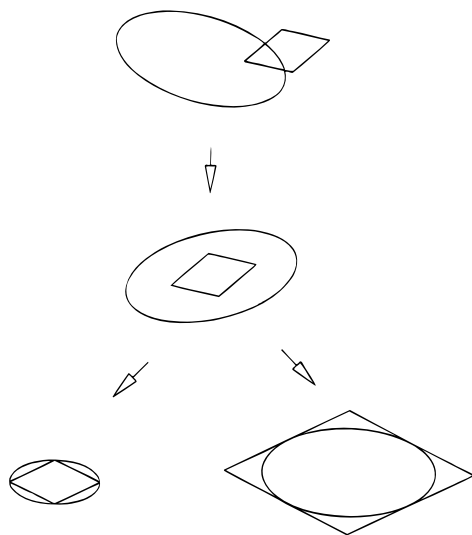


Figure 6. The comparison of an ellipse A and a rhombus B . First the rhombus and ellipse are translated and rotated to bring them into a maximum extent of correspondence. Then two different dilations are made: at the bottom left one finds $c(A \uparrow B) \cong 0.5$ and at the bottom right one finds $c(B \uparrow A) \cong 2.8$. Thus $d_f(A, B) = \log\{2.0 \cdot 2.8\}$ and $d_s(A, B) = \log\{0.5 \cdot 2.8\}$.

by α and $1/\alpha$ with $\alpha \geq 1$ —i.e., these modified functions still satisfy the three conditions of (7.1) formally demanded of a metric.

Distance functions allow comparisons between pairs of elements of \mathcal{P} , but too there are other types of “comparator” functions. One such additional *comparator* on a scaled poset \mathcal{P} is defined as

$$c(A \uparrow B) \equiv \min\{x \in \mathcal{R}_+ : x \cdot A \geq B\}, \quad A, B \in \mathcal{P} \quad (7.3)$$

where the minimum is over all positive scalars. With particular reference to the “covering” partial ordering (of Figure 1) this comparator is just the minimum dilatation factor (≥ 1 or ≤ 1) by which A can be multiplied such that the result covers B .

These dilatation-factor comparators give rise to distance functions, the first of which is enunciated in the following theorem.

Theorem A. For a scaled poset \mathcal{P} the function

$$d_f(A, B) \equiv |\log c(A \uparrow B)| + |\log c(B \uparrow A)|$$

is a metric on \mathcal{P} .

The distance function of this theorem (established in ref 15 or also as a special case of theorem C in section 9 here) is called the *form* distance of \mathcal{P} . It gives a way to measure the distance (or the degree of distinction or of dissimilarity) between two elements A and B , which might be called *forms* in the context of our covering poset of geometric forms.

A second result is developed which in the context of the covering poset involves grouping together all the forms of the same *shape*, which is here taken as the form made independent of scale. More particularly for a scaled poset define $A, B \in \mathcal{P}$ to be *scale* equivalent if

$$c(A \uparrow B) \cdot c(B \uparrow A) = 1 \quad (7.4)$$

That is, the scale factors for A to cover B and for B to cover A are reciprocals. This partitions \mathcal{P} into (scale) equivalence classes (which consist of forms of the same shape for the covering poset case). Then the following theorem can be shown.¹⁵

Theorem B. For a scaled poset \mathcal{P} the function

$$d_s(A, B) \equiv \log\{c(A \uparrow B) \cdot c(B \uparrow A)\}$$

is a metric on the set of scale equivalence classes of \mathcal{P} .

This d_s is called the *shape* distance for \mathcal{P} and gives a way to measure the dissimilarity between shapes (independently of scale). An example of two forms illustrating the form and shape distances between a rhombus and ellipse is indicated in Figure 6. That is, these distance functions compare two forms or shapes in terms of the amounts of magnification needed to bring them into coincidence to the extent that one completely covers the other.

It is to be emphasized that the metrics of theorems A and B apply for any scaled poset, independently of the particular case of geometric forms in terms of which they have been presented here. Again such scaled posets include some matrix-comparison posets (discussed in ref 15 rather than here) and the wide array of mimicks of section 4.

8. FORM AND SHAPE—CHIRALITY AND SYMMETRY

Comparisons of forms or shapes especially in a molecular context should be of much relevance in chemistry, e.g., as emphasized by Mezey.³⁴ The form of a molecular conformation might simply be taken as its van der Waals surface or as a particular electron-density contour. But, however, obtained they can be compared by the covering partial ordering mentioned in section 2 in connection with figures 1 and 6. Indeed this is the example from which some of the nomenclature of the preceding section is motivated. And indeed some of the ideas then hark back to earlier papers—particularly the magnification factors $c(A \uparrow B)$ are identifiable in Mezey³¹ and also in a graph-theoretic sense in refs 15 and 32. But beyond these references there is a large literature of molecular shape comparison, within different broad views of comparing molecular electronic densities, of comparing molecular volumes (e.g., as result from the union of atomic van der Waals spheres), or of nuclear conformations. As lead references see, e.g., refs 33 or 34 or 35, respectively. In the current sort of partially ordered context ref 15 attends some to each of these possibilities. Here focus is made to the case of (three-dimensional) forms and shapes, with the preceding section identifying ways to measure the distance between (i.e., the dissimilarity between) different pairs of forms or shapes, and the topic is to be returned to in section 11.

One sort of application of the current distance functions yields a measure of a molecule's chirality. A molecule is *achiral* if it (as a geometric form) can be superimposed faithfully on its mirror image, and otherwise it is *chiral*. Of course a molecule could be almost achiral in the sense that only a very slight change in its nuclear conformation might be needed to bring it to an achiral form, and as Weinberg and Mislow³⁷ have emphasized, the lack of chirality is the “typical” case, so that a measure of nearness to achirality is of more typical interest than “pure” chirality by itself. A natural way to measure the *degree of chirality* χ of a (molecular) form F is as the form distance between F and its mirror image F'

$$\chi(F) \equiv d_f(F, F') \quad (8.1)$$

Indeed such chirality measures have been of much interest recently, and various other possibilities (similarly defined

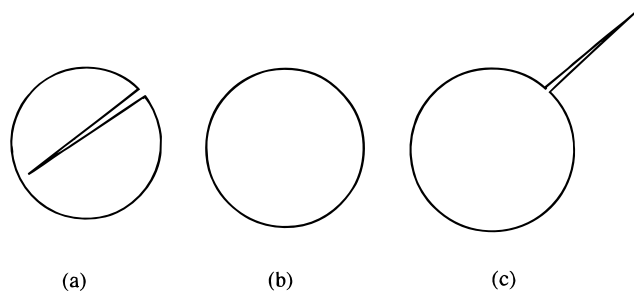


Figure 7. Three forms in (a), (b), and (c) with “large” form distances between any pair, although one might imagine them to be similar in form (and shape), at least if the cut in (a) and the whisker in (c) are of very small widths.

but using other distance functions) have been discussed in overviews by Zabrodsky and Avnir³⁶ and by Weinberg and Mislow.³⁷

More generally than chirality as it relates to mirror image symmetry, the question of symmetry with respect to an arbitrary point group arises. Just as for achirality the circumstance of symmetry is an exception, as emphasized by Avnir and co-workers^{36,38}, so that again the nearness to a symmetry is of more typical interest. The symmetry group is denoted \mathcal{G} , and its elements $G \in \mathcal{G}$ act on a molecular form F to yield new forms $G \cdot F$. The union of all these forms denoted by

$$\mathcal{G}F \equiv \bigcup_{G \in \mathcal{G}} G \cdot F \quad (8.2)$$

then has the symmetry of \mathcal{G} . And it is natural to take the degree of \mathcal{G} -symmetry of F to be the distance between F and $\mathcal{G}F$, taken as a reference form. But there is some ambiguity in that different results are obtained with different centers of rotation or with different planes of reflection for the elements of \mathcal{G} that is, there really are different isomorphic choices for \mathcal{G} depending on the “centering” of the point group operations with respect to the form F . Thence the *degree of \mathcal{G} -symmetry* $\delta_{\mathcal{G}}(F)$ is best taken as the distance minimized over the set $\{\mathcal{G}\}$ of differently centered isomorphic point groups

$$\delta_{\mathcal{G}}(F) \equiv \min_{\{\mathcal{G}\}} d_{\mathcal{G}}(F, \mathcal{G}F) \quad (8.3)$$

This can entail a minimization over several degrees of freedom: three translational degrees of freedom locating a center of rotation relative to F ; two rotational degrees of freedom orienting the principle axis of rotation relative to F ; and sometimes an additional rotational degree of freedom locating a secondary axis of rotation. This scheme is related to that discussed by Avnir and co-workers^{36,38}, who however deal directly with just the nuclear conformations.

Both the symmetry and chirality measures require adjustments in the manner of centering of symmetry groups to attain a minimal value for a distance. The means by which to constructively and efficiently carry out such adjustments are discussed by Avnir and co-workers^{36,38} and also by Weinberg and Mislow³⁷ in the context of other distance measures, but most of these considerations still apply with the present form distance.

A point of some conceivable interest is how the degree of chirality compares with the degree of symmetry with respect to a two-component “chirality” group \mathcal{C} , consisting of the identity along with either the inversion or a reflection. To

understand this consider the definition of $\chi(F)$, and more basically of $d_{\mathcal{C}}(F, F')$, for which F' is moved about to a particular location (and orientation) to bring it into a close correspondence with F so that a minimal magnification of F' will cover F . Note that this magnification covers both F and F' , so that it in fact also covers the union of F and F' so located. That is, $c(F \uparrow F) = c(F \uparrow F \cup F')$. Now this F' may be viewed as the result of the nonidentity transformation (inversive or reflective) \hat{i} of \mathcal{C} followed by a translation τ and rotation R , thusly $F' = R\tau\hat{i} \cdot F$. But since R and τ have been chosen to achieve a minimum, the result taking both of these to be the identity (whence $F \cup F'$ becomes $\mathcal{C}F$) must be greater. Thus $c(F \uparrow F) \leq c(F \uparrow \mathcal{C}F)$ and upon relating these minimal magnification factors involved in defining either $d_{\mathcal{C}}(F, F')$ or $d_{\mathcal{C}}(F, \mathcal{C}F)$ one has

$$\chi(F) \leq \delta_{\mathcal{C}}(F) \quad (8.4)$$

In a number of cases equality seems to apply.

In fact rather generally one could identify a degree of symmetry to a degree of chirality. Since every achiral object exhibits a point-group symmetry with an improper rotation, one could examine the degrees of symmetry for all possible such groups \mathcal{G} . That is, the degree of chirality asymmetry would be

$$\chi_{\text{asym}}(F) \equiv \min_{\{\mathcal{G}\}} d_{\mathcal{G}}(F, \mathcal{G}F) \quad (8.5)$$

where the minimum is over all centerings and orientations of \mathcal{G} and over all possible such \mathcal{G} . This set of \mathcal{G} we identify as the groups \mathcal{S}_p with improper p -fold rotation axes with p a prime or 1—the constraint to just primes or 1 is acceptable since all other point-groups containing an improper rotation contain one of these groups as a subgroup. This seems a daunting search, but presumably for most molecular forms one could in practice forego examination of larger values of p .

A point of some note is that the present form distance and scale distance can be quite sensitive to certain “small” defects in a shape. E.g., the shapes in figure 7a–c are quite distant (in terms of these distance functions) regardless of how narrow in width the cut in a or whisker in c might be. A related sort of approach making these two shapes quite similar may be developed in terms of the considerations of the next section.

9. TRANSFORMED POSETS

The ideas and general results outlined for scaled posets may be considerably generalized. In place of the dilatations in the scaled case one considers a general group \mathcal{T} of transformations to change the elements of \mathcal{P} into one another. Indeed one may allow \mathcal{T} even to be a *monoid* (such satisfying all the axioms of a group except that inverses need not exist). Further one presumes a *monoid norm* μ which is a function from \mathcal{T} to the non-negative reals \mathcal{R}_+ such that

$$1 \equiv \text{identity in } \mathcal{T} \Rightarrow \mu(1) = 0$$

$$1 \neq T \in \mathcal{T} \Rightarrow \mu(T) > 0$$

$$S, T \in \mathcal{T} \Rightarrow \mu(S) + \mu(T) \geq \mu(S \cdot T) \quad (9.1)$$

Then a corresponding function $d_{\mathcal{T}}$ on $\mathcal{P} \times \mathcal{P}$ may be defined

as

$$d_{>}(A,B) \equiv \min_{T \cdot A \geq B} \mu(T) + \min_{T' \cdot B \geq A} \mu(T') \quad (9.2)$$

where the minimizations are taken over all $T, T' \in \mathcal{T}$. (And if no T or T' exists such that $T \cdot A \geq B$ or $T' \cdot B \geq A$, then the corresponding minimum for $\mu(T)$ or $\mu(T')$ is taken as ∞ .) As an example in the special case of the covering poset with \mathcal{T} the group of dilatations, the monoid (or group) norm on \mathcal{T} may be taken to be $\mu(T) = |\log\{m(T)\}|$, $T \in \mathcal{T}$ where $m(T)$ is the magnification factor for the dilatation $T \in \mathcal{T}$, and the function $d_{>}$ turns out to be the form distance d_f . Now another related function $d_{<}$ on $\mathcal{P} \times \mathcal{P}$ may be defined

$$d_{<}(A,B) \equiv \min_{T \cdot A \leq B} \mu(T) + \min_{T' \cdot B \leq A} \mu(T') \quad (9.3)$$

where again minimizations are taken over all $T, T' \in \mathcal{T}$. When \mathcal{T} is a group so that T and T' have inverses and it also occurs that $\mu(T^{-1}) = \mu(T)$, these two functions $d_{>}$ and $d_{<}$ become identical, as is the case with the dilatations on the covering poset of forms. Now rather generally one has the following theorem.

Theorem C. Let \mathcal{P} be a transformed poset with a transformation monoid \mathcal{T} and with a monoid metric μ satisfying eq 9.1. Then the functions $d_{>}$ and $d_{<}$ defined by eqs 9.2 and 9.3 are metrics on \mathcal{P} .

To prove the result for $d_{>}$ first note that with the definition as in (9.2) in terms of the monoid norm the function $d_{>}$ clearly takes only non-negative values. If $A, B \in \mathcal{P}$ such that $d_{>}(A,B) = 0$, then the two terms on the right-hand side of (9.2) must be 0, so that by the first two conditions of (9.1) one obtains $A \geq B$ and $B \geq A$, whence $A = B$. Now it remains to prove that $d_{>}$ satisfies the triangle inequality, so that we consider

$$d_{>}(A,B) + d_{>}(B,C) \equiv \min \mu(S) + \min \mu(S') + \min \mu(T) + \min \mu(T')$$

with the minimums taken such that $S \cdot A \geq B$, $S' \cdot B \geq A$, $T \cdot B \geq C$, and $T' \cdot C \geq B$, respectively. Then $TS \cdot A \geq T \cdot B \geq C$, and $S'T' \cdot C \geq S' \cdot B \geq A$, so that $d_{>}(A,B) + d_{>}(B,C) \geq \mu(TS) + \mu(S'T')$. But then $d_{>}(A,B) + d_{>}(B,C) \geq \min \mu(U) + \min \mu(U')$ where U and U' are chosen to be such that $U \cdot A \geq C$ and $U' \cdot C \geq A$. Thus $d_{>}(A,B) + d_{>}(B,C) \geq d_{>}(A,C)$, and a quite similar approach applies for $d_{<}$, so that the theorem is proved.

It may be noted that in place of a monoid norm one may often also deal with a special type of distance function ρ on the monoid \mathcal{T} . Such a special type of metric ρ on \mathcal{T} is one which satisfies the condition of being \mathcal{T} -invariant

$$\rho(TT_1, TT_2) \equiv \rho(T_1, T_2), \quad T_1, T_2, T \in \mathcal{T} \quad (9.4)$$

Then a \mathcal{T} -invariant metric ρ determines a monoid norm μ , because given such a ρ one may verify that $\mu(S)$ defined as $\rho(S, 1)$ turns out to satisfy the conditions of eq 9.1. In seeking a converse starting from a given monoid norm μ one might seek to define ρ via

$$\rho(T_1, T_2) \equiv \min\{\mu(S_1) + \mu(S_2) + \mu(S_3) + \mu(S_4) | S_1 T_1 S_2 = S_3 T_2 S_4\} \quad (9.5)$$

and indeed such seems to apply for many monoids. As already noted by Batagelj³⁹ such is possible for \mathcal{T} being a

group if also the norm satisfies $\mu(T^{-1}) = \mu(T)$. Thence the theorem C is related to this group-theoretic result.

There are many examples of transformed posets or more generally of *transformable* posets which may be embedded in a larger transformed poset \mathcal{P}_+ . The first such are that of scaled and scaleable posets already discussed. Another general such poset is the “shiftable” poset of probability distributions as briefly discussed elsewhere.¹⁵ Yet another general type of transformed poset is that involving the subsets of a set X (i.e., using the \subseteq partial ordering)—when the monoid \mathcal{T} is generated by transformations T_a in correspondence with the elements of $a \in X$ such as to adjoin that corresponding element to any subset of X . That is,

$$T_a \cdot A \equiv A \cup \{a\}, \quad a \in X, \quad A \subseteq X \quad (9.6)$$

and the monoid norm μ takes a value for a transformation counting the number of elements added. Alternatively one could consider the transformations deleting elements from sets, or one could take \mathcal{T} to include both types of transformations (either adding or deleting elements of X). A related case for the poset of subgraphs of a graph G involves transformations in correspondence with the vertices of G and are such as to give the induced graph resulting upon addition to the vertices of the subgraph.

Given a poset with a maximum element and a Hasse diagram, one may consider the monoid generated by transformations carrying an $A \in \mathcal{P}$ one step in the upward “direction” in the Hasse diagram and a monoid norm counting the number of upward steps. Such applies to many of the “periodic-table-related” posets of section 3 and to some of the mimicks of section 4. If the poset has a minimum element, then transformations in the downward direction may be viewed to constitute \mathcal{T} . With only the much weaker assumption that the Hasse diagram is connected, \mathcal{T} may be taken to consist of both upward and downward transformations with the monoid norm again counting steps up or down. One very important case falling under this category involves that of sequences (as occur with polypeptides, proteins, etc.) with a partial ordering being that of subsequences—the sequences then may be manipulated with insertions and deletions (as then generate \mathcal{T}) and the monoid norm involves the number of such primary *indels*—see, e.g., see ref 40.

It seems that in many cases one can view the transformations as defining a partial ordering \mathcal{P} . The transformation monoid \mathcal{T} has in general two submonoids: $\mathcal{T}_{>}$ whose elements increase the “size” of an element of \mathcal{P} ; and $\mathcal{T}_{<}$ whose elements decrease the size of an element of \mathcal{P} (with size here being something rather general). E.g., for the form-covering case, $\mathcal{T}_{>}$ and $\mathcal{T}_{<}$ are the sets of dilatations with magnification factors $m \geq 1$ and $m \leq 1$, respectively.

10. MOLECULAR BRANCHING

Another case of what can be viewed as an example of a partial ordering is that of molecular branching. Indeed such “branching” has long been a recognized feature of importance, particularly for the different isomeric alkanes, where this just distinguishes the different isomeric species. E.g., commercial oil refineries process the alkanes to generate more highly branched species with higher “octane numbers” (and greater profit). The recognition of such relevance of branching goes back some time—one nice discussion (though likely not the first) is found in the monograph of Tatevski

*et al.*⁴¹—and currently branching (and consequent steric interference) is the central issue in consideration of “dendrimers”.⁴² With the recognition of the relevance of molecular branching there then have been proposed^{43–50} numerous “branching indices”, often taken to be of a graph-theoretic nature. For a hydrogen-deleted graph G these include the following:

*the count $p_2(G)$ of length-two paths;⁴⁴

*the negative (or inverse) of Randić’s connectivity index⁴⁵

$$\chi(G) \equiv \sum_{a \sim b} (\delta_a \delta_b)^{-1/2} \quad (10.1)$$

where the sum is over all edges of G and δ_a denotes the degree of a vertex a ;

*the negative (or inverse) of the “Wiener” index $W(G)$,^{47–49} where

$$W(G) \equiv \sum_{a < b} d_{a,b} \quad (10.2)$$

with the sum over all pairs of vertices of G and $d_{a,b}$ the shortest-path distance between a and b ;

*the information-theoretic branching function⁴⁶

$$I(G) \equiv \sum_i \{ \delta_i(G)/2q \} \log_2 \{ \delta_i(G)/2q \} \quad (10.3)$$

where q ($=\sum_i \delta_i(G)/2$) is the number of edges, and \log_2 denotes the base-2 logarithm function. One might guess that some of these indices might be isotonic functions for an underlying poset and that the collection of such functions together constitutes a mimick (such as generally described in section 4).

But in the case of molecular branching something more specific than the mimick is possible.^{5,6} It is reasonable to view the branching of a graph G to be determined from the distribution of numbers of sites of different degrees. Let this distribution of degrees δ_j be ordered such that

$$\delta_1 \geq \delta_2 \geq \delta_3 \geq \dots \geq \delta_N \quad (10.4)$$

where N is the number of vertices of G . Then letting $a_i(G)$ denote the number of vertices of degree i , there are precisely $a_i(G)$ of the δ_j which take the value i , and $\sum_i \delta_i = \sum_k k \cdot a_k(G)$ is twice the number of edges of G , this sum for trees being $2(N-1)$. Now one might view a transformation moving edges around in G (without changing it so as to make it disconnected) to increase the branching if one $a_k(G)$ increases at the expense of another $a_j(G)$ with $j < k$. Such a transformation increases the number of more highly branched vertices at the expense of the number of less highly branched vertices. An example is indicated in Figure 8a and indeed some such transformations (usually more numerous listed through specialization, or generalization) have already been considered^{46–49} as criteria for increasing branching. One may specify a set of graphical transformations to be taken as the monoid defining a *vertex-branching* poset \mathcal{P} . That is, if two graphs G and G' are such that

$$a_i(G) = a_i(G'), \text{ except } a_j(G) = a_j(G') - 1 \text{ and } a_k(G) = a_k(G') + 1 \text{ with } k > j \quad (10.5)$$

then the *branching extent* of G' is immediately greater than that of G in a Hasse diagram. Denoting this branching extent of G as $\beta[G]$, one then writes $\beta[G'] \geq \beta[G]$ when G and G' are so related or indeed if they are more distantly related

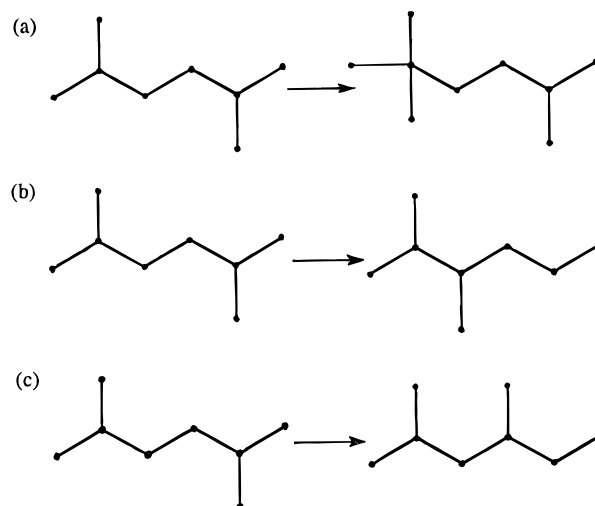


Figure 8. Three transformations increasing different types of branching.

along a path decreasing from G' to G in the Hasse diagram. But one can then reexpress^{5,6} the characterization of this partial ordering in a manner without mention of the Hasse diagram (or associated elementary transformations):

$$\beta[G'] \geq \beta[G] \text{ if } \sum_{i=1}^k \delta_i(G') \geq \sum_{i=1}^k \delta_i(G), \quad k = 1 - N - 1 \quad (10.6)$$

Indeed this branching extent provides a special case of the so-called *majorization* partial ordering on distributions of a rather general nature^{5,6,18}, which has been in fact considered some time back in a purely mathematical context⁵¹ and early on too in an economic context.⁵²

In a chemical context a modified view of branching has sometimes been taken⁴⁹ wherein additional transformations leaving the a_i fixed have been considered to increase branching. Such a transformation is indicated in Figure 8b—and indeed there is some rationale for this: under the transformation the branching has become more localized though not to the extent of collapsing it more extensively on a single site—that is, the transformation has brought the various branches into closer proximity, which, were it to be so close as to collapse onto a single site, would already be agreed to increase branching. This suggests a modified approach defined in terms of edge degrees $\delta_e \equiv \delta_{\{j,k\}} \equiv \delta_j + \delta_k$ for edge $e \equiv \{j,k\}$ or also in terms of $a^*_i(G)$ as the number of δ_e taking the value i . Then the *edge-branching extent* $\beta^*[G']$ of graph G' is greater than that $\beta^*[G]$ of G if the vertex degrees of (10.5) are replaced by edge degrees—or also it can be viewed in terms of (10.4) with a^*_i replacing a_i . Indeed even further extended transformations have been considered where even the δ_e remain fixed, while however the degree counts for even larger subgraphs (paths) are shifted around, as in figure 8c. In such circumstances the branching extent of the preceding paragraph would be more properly described as the “vertex-branching extent”. But even Bonchev, in whose papers such more extended transformations have been considered, now seems to view⁵⁰ such extended transformations of less relevance, and so here we do not further pursue such modifications.

But now one can seek to interrelate various numerical branching indices to one of the branching extent partial orderings. For example it is readily seen that p_2 is isotonic:

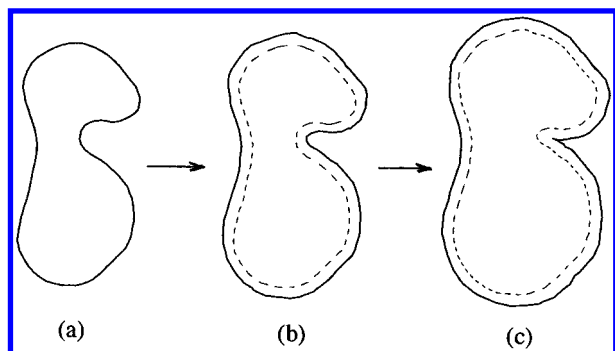


Figure 9. The enlargement transformation relevant for the Hausdorff distance function: the original form A is transformed to $T_r \cdot A$, and this is again transformed in the same manner to obtain $T_{2r} \cdot A = T_r \cdot T_r \cdot A$.

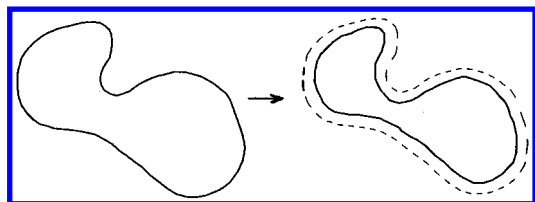


Figure 10. The diminishment transformation for a dual type of Hausdorff distance function: the original form A on the left is transformed to $T_{-r} \cdot A$ on the right.

one notes that for a vertex of degree i there are $i(i+1)/2$ length-2 paths centered thereon, so that $p_2(G) = \sum_i i(i+1) \cdot a_i(G)/2$, and increasing an a_k at the expense of an a_j with $k > j$ clearly leads to an increase in p_2 . The result for $I(G)$ is a direct standard result⁵³ for a probability distribution, here identified as $P_i(G) \equiv a_i(G)/2q$.

11. RETURN TO FORM AND SHAPE—CHIRALITY AND SYMMETRY

Even for posets already introduced the general transformational approach offers further possibilities for the definition of a distance function and consequent other quantities, e.g., as degrees of symmetry or chirality. For the case of the geometric shapes one may introduce an alternate transformation monoid \mathcal{T} , say that whose elements extend the boundaries of the shape by different amounts. That is, for $a \in \mathcal{R}_+$ and a shape A in Euclidean space \mathcal{E} (on which there is defined a Euclidean metric d_E) define a transformation

$$T_r \cdot A \equiv \{x \in \mathcal{E} \mid d_E(x, a) \leq r, \exists a \in A\} \quad (11.1)$$

For example in Figure 9 the shape in part a becomes enlarged to that in part b and upon even further enlargement becomes as that in part c. In this case the shapes of 7a–c can be quite similar. The metric d_H resulting here is that of Hausdorff,⁵⁴ usually described in a different way: the distance between two forms A and B is the sum of two smallest numbers r and s such that first every point \mathbf{x} of A has at least one point \mathbf{y} of B within a (Euclidean) distance r , and second every point \mathbf{y} of B has at least one point \mathbf{x} of A within a distance s . Again for our purposes it is understood that the two forms A and B are to be moved around in space relative to each other to achieve the minimum value for $d_H(A, B)$. Here r and s identify the transformations T_r and T_s by which A and B might need to be grown in order to cover the other form. Moreover this metric has been proposed by several workers^{37,55} to be utilized for the measurement of

chirality (especially of molecules). They propose a chirality measure $\chi_H(F)$ of a form F to be proportional to the Hausdorff metric between F and its mirror-image form F' in such a way as to be independent of scale: letting $d(F)$ be the maximum (Euclidean) distance between two points of F , they define the (Hausdorff) *chirality measure*

$$\chi_H(F) \equiv d_H(F, F')/d(F) \quad (11.2)$$

Too one can imagine parallel definitions of *degree of symmetry*, say as

$$\delta_{\mathcal{CH}}(F) \equiv \{\min_{\{G\}} d_H(F, G)/d(F)\} \quad (11.3)$$

Again this approach is somewhat like that of Avnir and co-workers,⁵⁶ though differing in being scale independent (in the sense that any given form, no matter how magnified, yields the same value for $\delta_{\mathcal{CH}}$), and second this earlier proposal is for the nuclear positions (rather than for the molecular form as the region within a closed surface in three dimensions).

But there are yet other possibilities for distance functions along with associated chirality and symmetry measures. At least a couple of possibilities entail the introduction of a transformation which is something like a “generalized inverse” to that of (9.1): for any $r \in \mathcal{R}_+$ and any form F with a surface $S(F)$ define

$$T_{-r} \cdot A \equiv \{x \in A \mid d_E(x, a) \geq r, \forall a \in S(A)\} \quad (11.4)$$

That is, T_{-r} “shrinks” a shape in by an amount r , as in Figure 10. Then a first possibility involves utilizing these T_{-r} , $r \in \mathcal{R}_+$, to define a new distance analogous to $d_<$ in eq 9.3, the Hausdorff distance d_H corresponding to $d_>$ of eq 9.2. Another possibility more nearly parallel to the modification to our earlier form distance in section 7 to obtain a shape distance. That is, a “tighter” Hausdorff-type distance $d_{\text{th}}(A, B)$ between A and B might be the sum of two real numbers r and s such that r is the minimum such that $T_r \cdot A$ covers B and s is the minimum such that $T_s \cdot B$ covers A . Here r and s may now be negative (or positive), so that evidently d_{th} is tighter than d_H in the sense that d_{th} always yields values which are at least as great as d_H , though when used in a chirality measure the result is the same because of the symmetric relation between F and F' .

Even other plausible metrics are conceivable, several of which are reviewed by Weinberg and Mislow,⁵⁷ and some of these can be viewed in terms of transformed posets. For instance, if v denotes a *volume measure* on three-dimensional forms (or an area measure on two-dimensional forms), then one can define a metric distance between two forms A and B as

$$d_v(A, B) \equiv \min v(\{x \in A \cup B \mid x \notin A \cap B\}) \quad (11.5)$$

where the forms are moved around (i.e., translated and rotated) so as to achieve a minimum value. Here one may view the transformations involved to move small bits out of one form so as to either encompass or be encompassed in the other form, and the measure of the extent of such an overall transformation is just the net amount of small bits so moved. Then a chirality measure is

$$\chi_v(F) \equiv d_v(F, F')/v(F) \quad (11.6)$$

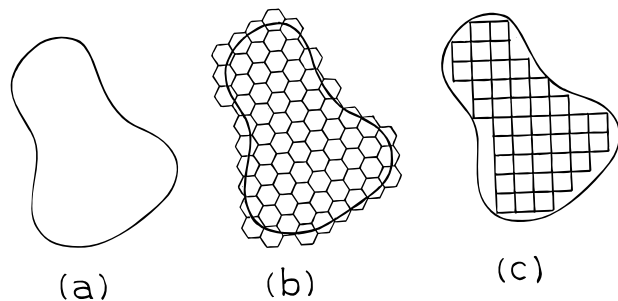


Figure 11. Form discretization in terms of animals: (a) the original form; (b) a circumscribing polyhex animal, as in Randić and Razinger;⁶⁰ and (c) an inscribed square-cell animal, as in Mezey.⁶²

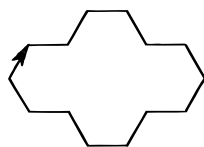


Figure 12. The binary code assignment for a polyhex animal. Starting from the "arrowed" edge and proceeding clockwise there results the code $[+-++-++-++-++-]$ with $+$ and $-$ for right and left turns, respectively.

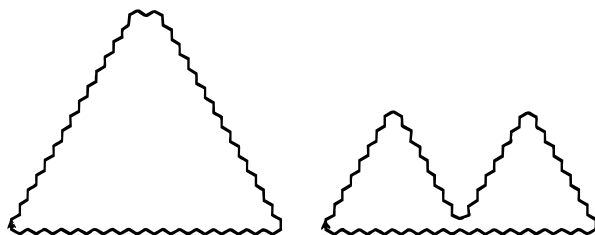


Figure 13. Two circumscribed polyhex animals around a near triangle and a "bitriangle" such that there result similar Hamming codes, namely both with codes $[(+-)_7 + \star (+-)_6 + \star - \star (+-)_6 + \star - (+-)_7 + (+-)_5 + +]$ differing only in the positions designated by the stars (and where the subscripts count repeated patterns). Note that even at much higher resolution there would be only four positions of difference in the codes for the underlying forms.

where the division by $\nu(F)$ makes this chirality measure independent of scale—and the same as that of Gilat.⁵⁸

Finally a few comments might be made as to measures for the symmetry of electronic wave functions. This problem is somewhat different than that of comparing shapes or nuclear conformations or electron densities—because wave functions can be symmetry adapted to irreducible representations other than the totally symmetric and still be viewed as possessing the group symmetry. Thence the different irreducible symmetry projections (and their weights) are of natural relevance. The symmetry weights are non-negative and sum to 1 (for a normalized wave function) so that they represent a probability distribution, and a "degree of symmetry adaptation" has been considered⁵⁹ in terms of an information-theoretic functional. But too the sets of symmetry weights can be partially ordered under majorization and metrics considered thereon.

12. DISCRETIZED FORM AND SHAPE—RANDIĆ AND RAZINGER

Further yet there are discretization schemes to deal with form and shape. For instance, one might compare via the covering operation two-dimensional forms F in the Euclidean plane with polyhex forms P which are forms partitioned into regular hexagonal tiles and encode information about F in terms of a minimal P covering F . See, e.g., Figure 11, and

also the article of Randić and Razinger⁶⁰ considering this construction (for chemical purposes). Also very much the same approach to shapes is proposed by Harary and Mezey^{61,62} where however they utilize the square lattice and instead look for maximal square-tiled forms covered by the parent form. Such connected sections of uniformly sized cells cut (by a Jordan curve) from a regular grid are often termed *animals*. (Then no two cells a and b share but a single point p unless they each share a p -containing edge with another cell c adjoined in the animal to both a and b .) See, e.g., Eden or Read,⁶³ but this terminology is fairly standard in statistical mechanics too. Given a class of such animals they can be viewed as graphs and can be compared directly, whereupon the covering operation is also equivalent to the subgraph operation, and a relevant monoid of transformations is that generated by circumscriptions of an animal by new cells around the boundary and also perhaps by a generalized inverse transformation deleting all cells on the boundary. The current results then provide metrics on the set of polyhex forms, and such may be used as a sort of "fuzzy metric" on the originating parent forms encoded into the animal forms. Moreover, such proposed metrics though clearly the discrete analogues of the Hausdorff metrics of the preceding section seem not to have been previously suggested. And another possibility is the use of the analogue of the "volume measure" metric of eq 10.5, whereby one looks for a maximum-sized isomorphic subgraphs H_1 and H_2 of a pair of the animal graphs G_1 and G_2 being compared and takes the distance between G_1 and G_2 to be $|G_1| + |G_2| - 2|H|$. Indeed just this is already studied and used³⁵ for comparing general connected graphs.

For the case of polyhex animals the "shape dissimilarity measure" of Randić and Razinger⁶⁰ presumes the cyclic sequence of turns (left or right) in proceeding around the polyhex to specify the polyhex as a cyclic sequence of 0s and 1s as in ref 64, or $+$ s and $-$ s as illustrated in Figure 12. Then they take the Hamming distance⁶⁵ between two such cyclic sequences to be the distance between the two corresponding shapes, this distance being the minimum number of 0s and 1s in a sequence that need to be changed to the other label so as to transform one cyclic sequence to the other. But this is not a "shape metric" in our current sense in that it is (strongly) dependent on scale, and indeed it may be argued not even to be a reasonable "form metric" in our current sense. That is, their Hamming-distance dissimilarity measure will count two polyhex forms as similar if they have similar overall sequences (when cycled around so as to match as much as possible), whence this measure does not pay much attention to short changes in a sequence though this can give great changes in direction of the walk around the perimeter and thence to a great dissimilarity of form (while their measure gives a high degree of similarity). For instance, the two forms of Figure 13 when covered by (small) hexagons yield much the same code except at the few corners (each 120° corner-turn involving an even number of 60° turns around the hexagonal-lattice animal, so that all the longer straight-line runs of the underlying forms have alternating 0's and 1's which for the most part can be brought into coincidence). On the other hand, the current discrete subgraph-based form metric avoids much of this sort of problem.

But most of these approaches have another problem in that one parent form may have two different optimal discretizations, both having the same number of cells but

represented by quite different graphs. Of course, one may argue that such strict degeneracies are an exception, but in fact they indicate that (rather frequently) two but slightly different parent shapes may have quite different discretizations. The problem may be illustrated with the example of two minimal polyhex coverings of a single rectangle with the orientation between the rectangle and the lattice having polyacenic strips oriented along either the horizontal or vertical sides of the underlying rectangle—presumably the two consequent polyhexes have a similar global shape but can have a great difference via either the measure of Randić and Razinger⁶⁰ or by the subgraph-based form metric. Here this problem can be usually largely avoided if one makes a canonical orientation say with a polyacenic direction chosen to be in the direction of the axis for the maximum moment of inertia. If part of the reason of the discretization is to make such computations of inertias more amenable, then after a trial orientation and discretization one might evaluate the inertia matrix for the polyhex and choose its major axis as a direction for a refined orientation, whence the process is iterated to convergence.

Randić and Razinger¹² go on to suggest that their scheme be used to define a chirality measure. But there are several difficulties, a couple of which are at least partly addressed by our preceding remarks. The questions of the extension to three dimensions and to the choice of the size of the cells are addressed in Mezey's scheme and discussed in the next section.


13. DISCRETIZED FORM AND SHAPE—MEZEY

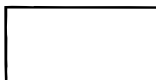
Using the square-cell lattice animals, Mezey^{61,62} follows a somewhat different approach making use of a whole sequence of animals with cells of different sizes. Given a (two-dimensional) form F , an animal with cells of side s is termed an *interior-filling* animal if there is no other inscribable animal with a greater number of equal-sized cells. These interior-filling animals then are partitioned into classes $\mathcal{C}(F,n)$ according to their numbers n of cells. Then the *grade of shape similarity* $s(F_1,F_2)$ of forms F_1 and F_2 is the smallest value c such that

$$\mathcal{C}(F_1,n) \cap \mathcal{C}(F_2,n) = \emptyset, \text{ all } n \geq c \quad (13.1)$$

(where \emptyset is the empty set). This is a "shape" index (or shape invariant) in that it does not depend on scale, and Mezey (assuming that it takes values ≥ 2) proposes to use it in making a chirality measure. But this grade of similarity can be argued to be a poor measure for the comparison of shapes, as can be illustrated by way of some simple examples. First for the two forms F_1 and F_2 which themselves are essentially just the one- and two-cell animals one has the result indicated in Figure 14 from which it follows that $s(F_1,F_2) = 0$ (or maybe 1), while the intuitively rather different (polymer-like) form F_3 of Figure 15 turns out to have a quite high grade of similarity to F_1 .

A modification of this type of Mezey's measure seems to avoid most such problems. In this modified approach for a form F consider square-cell animals which circumscribe F such that every cell contains some part of F . In fact start with an animal $A_0(F)$ with a single square cell of minimal

$F_1 =$


$F_2 =$




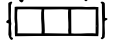

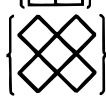
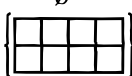
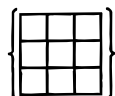

n	$\mathcal{C}(F_1,n)$	$\mathcal{C}(F_2,n)$
1		\emptyset
2	\emptyset	
3	\emptyset	
4		\emptyset
5		\emptyset
6	\emptyset	\emptyset
7	\emptyset	\emptyset
8	\emptyset	
9		\emptyset
10	\emptyset	\emptyset
11	\emptyset	

Figure 14. A tabulation of Mezey's first 11 classes $\mathcal{C}(F,n)$ for the case that $F = F_1$ is the one-cell animal and for the case that $F = F_2$ is the two-cell animal. For $n \geq 12$ the sets $\mathcal{C}(F_1,n)$ and $\mathcal{C}(F_2,n)$ contain even more faithful representations of F_1 and F_2 and so surely are disjoint.

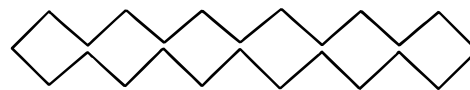


Figure 15. A polymer-like form which by Mezey's grade of similarity is very similar to the form F_1 of Figure 14.

size such as to minimally cover F . Then let $A_1(F)$ be the animal obtained by dividing $A_0(F)$ up into a 2×2 set of cells and deleting any new (smaller) cell not overlapping with any part of F . Continue this halving of the cell sides with subsequent deletions to obtain a sequence of animals $A_0(F), A_1(F), A_2(F), A_3(F), \dots$. Then define a new grade of shape similarity between F and F' as the first value of n such that $A_n(F)$ is distinct as a graph from $A_n(F')$. Indeed 2^{-n} should be related to a sort of distance between F and F' . That is, n is a number of doubling magnifications (or the "resolving power") needed to distinguish between the shapes of F and F' —and just as in our consideration of scaled posets, such magnification factors should be logarithms which yield distances. More particularly imagining that all F are scaled so that the sides of $A_0(F)$ are of length 1 and the sides of the cells in $A_n(F)$ are of length 2^{-n} , one considers the first value of n such that $A_n(F) \neq A_n(F')$ while $A_{n-1}(F) = A_{n-1}(F')$. Then all points of $A_n(F)$ are within $\sim 2^{-n}$ of all the points of $A_{n-1}(F) = A_{n-1}(F')$ which covers $A_n(F')$, so that all points of $A_n(F')$ are within $\sim 2^{-n}$ of $A_n(F)$. A similar statement applies interchanging the roles of F and F' , and as a consequence the Hausdorff distance between $A_n(F)$ and $A_n(F')$ is roughly $\sim 2^{-n}$. That is, 2^{-n} can be viewed as a (scaled) discretized Hausdorff distance between F and F' . See Figure 16. This process of halving the scales of a covering animal is similar

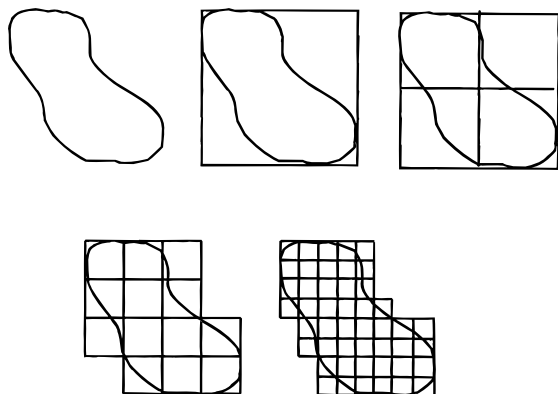


Figure 16. An example for a form F (which is the first shown shape) of the assignment of the first few members of a sequence $A_0(F)$, $A_1(F)$, $A_2(F)$, $A_3(F)$

to that used in some other discretization schemes,⁶⁶ though the earlier considered choice of the initial cell $A_0(F)$ as an arbitrary (i.e., nonoptimally chosen) initial square covering F does not yield well-defined invariants. Further, instead of animals based on the square lattice it could be based on the triangular lattice but not the hexagonal lattice—since a regular-hexagonal cell cannot be broken up into smaller regular-hexagonal cells.

Note that there are nice computationally amenable features of the present new proposal. The sliding around (translating and rotating) of different animals to optimally match a form is confined to the simple single-cell animal. Whereas in the approach of Mezey (or Randić and Razinger) there is such a problem at every stage of resolution. Moreover, after the (relatively simple) first stage in the present approach, the sizes of the cells at subsequent stages are thereafter fixed, whereas in the other approaches the sizes are not. There remain some problems in that one structure may have more than one way to fit the initial cell $A_0(F)$ around F —or perhaps more significantly, given such a form F with two fittings for the initial cell there are likely to be two ways to make (infinitesimally) small perturbations of F to F_1 and F_2 such that each has just one of the two optimum fittings, whence F_1 and F_2 likely would be described as more different than they should be. Such a problem seems often to be eliminated if instead of a single globally optimum one-cell animal, one accepts a set of different optimal one-cell animals, each such being a locally minimum size (under the influence of small rotations and translations).

Finally the current scheme readily extends to three-dimensional forms, such as are of primary interest for comparing molecular shapes. One simply makes use of animals made out of uniformly sized cubic cells. (Even higher-dimensional extensions are straightforward.) Of course Mezey's approach^{61,62} similarly extends, though the problems with optimum alignments at each possible scale becomes even more severe with one more translational and two more rotational degrees of freedom in choosing optimum alignments between animal and parent form. Overall there seems to be a decided preference for the present approach in measuring what is here termed "form" or "shape" dissimilarities.

But finally it may be of interest to note that the present discretized scheme leads to ready computation of a number of additional geometric indices. Just counting the number $\#(A_n)$ of cells in the n th stage animal A_n yields a measure of area (in two-dimensions) or volume (in three-dimensions)

$$\mu(A_n) \equiv \#(A_n)/2^n \quad (13.2)$$

Moreover, if one imagines that $\mu(A_n)$ approaches a well defined limit, then the count appearing here scales as $\#(A_n) \approx C \cdot (2^{-n})^\varphi$, and one might formally define the quantity

$$\varphi_n \equiv \log\{\mu(A_n)/\mu(A_{n+1})\}/\log 2 \quad (13.3)$$

as an n th stage *form dimension*. Indeed this approach parallels somewhat one frequent approach to a geometric form's *fractal dimension*

$$\varphi \equiv \varphi_\infty \equiv \lim_{n \rightarrow \infty} \varphi_n \quad (13.4)$$

See, e.g., Mandelbrot in ref 67.

14. CURVE FITTING

One important nontrivial example of a transformed poset involves the rather widespread circumstance of curve fitting. Most simply one may consider a set of data points $(x_i, y_i) i = 1 \rightarrow N$ to which there are various straight-line approximants $f_{a,b}$ for $a, b \in \mathcal{R}$ with

$$f_{a,b}(x_i) \equiv ax_i + b \quad (14.1)$$

The *residual* magnitudes $|y_i - f_{a,b}(x_i)|$, or their squares, are imagined to be shuffled around so that they form a nondecreasing sequence $\mathcal{S}_{a,b}$

$$\delta_1 \leq \delta_2 \leq \dots \leq \delta_{N-1} \leq \delta_N \quad (14.2)$$

The various lines (as specified by the $f_{a,b}$) may be identified to the elements of a poset \mathcal{P} with the partial ordering such that

$$f_{a,b} \geq f_{a',b'} \Leftrightarrow \{\delta_i \geq \delta'_i \mid \delta_i \in \mathcal{S}_{a,b}, \delta'_i \in \mathcal{S}_{a',b'}, i = 1 \rightarrow N\} \quad (14.3)$$

Here then there is generally no "best-fit" line, with a set of residuals each smaller than for every other line. The standard "least-squares" fit is but one isotonic function on \mathcal{P} —it is that associated with the distance from a hypothetical "0-error" $\delta'_i = 0$ line using a Euclidean metric $\{\sum_i (\delta_i - \delta'_i)^2\}^{1/2}$. But there are other possible metrics on this set \mathcal{P} of lines, and in particular ones associated with the transformation group \mathcal{T} such that

$$T_{c,d} \cdot f_{a,b} \equiv f_{ac,b+ad}, \quad f_{a,b} \in \mathcal{P}, \quad T_{c,d} \in \mathcal{T} \quad (14.4)$$

with a monoid norm, e.g.,

$$\mu(T_{c,d}) \equiv \alpha |\log(c)| + \beta |d| \quad (14.5)$$

where α and β are (fixed) positive numbers characterizing the norm. Then via eqs 8.2 and 8.3 this leads to metrics $d_>$ and $d_<$ on \mathcal{P} , such metrics so far seeming to have been little utilized.

15. PROSPECTS, ETC.

Posets here have been emphasized as a rather general cosmopolitan occurrence in chemistry, and indeed several of the circumstances such as those involving shapes, or branching, or curve fitting should have much wider applications too. The Randić–Wilkins periodic table of alkanes and of the concept of "aromaticity" illustrate the important

ideas of multiposets and of poset mimicks—and the discussed linear algebraic analysis of aromaticity illustrates how in the context of posets correlations between different aromaticity indices can be identified (though the techniques should carry over to other mimicks). Several of the indicated applications, e.g., as concerns molecular-form dissimilarity and degrees of symmetry or chirality, are of much current interest but hopefully here have been elucidated—and at least some new chirality, symmetry, and shape indices have been identified. There have been indicated a number of posetic tools, including linear algebraic analyses for mimicks, and also the metrics and dissimilarity measures for diverse examples of scaled posets, or more generally of transformed posets, here illustrated in terms of the concepts of “molecular branching”, of molecular form and shape, and of the fitting of curves. It thus seems that partial orderings may be a fundamental natural mathematical structure of much potential utility in chemistry. Perhaps they make up one of C. A. Coulson’s “primitive patterns of understanding”.⁶⁸

ACKNOWLEDGMENT

Special acknowledgement is made to discussions with Professors Ruch and Randić whose studies have led the way toward the realization of the relevance posets have for chemistry. Professor Randić is surely a worthy receptor of the Skolnik award, for his work in this area of posets and his work on many other chemical graph-theoretic problems, including some in the chemical documentation area too. Finally the authors acknowledge the Welch Foundation of Houston, TX, for financial support.

REFERENCES AND NOTES

- (1) Birkhoff, G. *Lattice Theory*; American Mathematics Society: Providence, RI, 1948.
- (2) Rota, G.-C. On the Foundations of Combinatorial Theory. I. Theory of Möbius Functions. *Zeit. Wahrscheinlichkeitstheorie Verw. Geb.* **1964**, *2*, 340–368.
- (3) Ruch, E. The Diagram Lattice as Structural Principle; *Theor. Chim. Acta* **1975**, *38*, 167–183.
- (4) Ruch, E.; Mead, A. The Principle of Increasing Mixing Character and Some of Its Uses. *Theor. Chim. Acta* **1976**, *41*, 95–117. Ruch, E.; Schraner, R.; Seligman, T. H. The Mixing Distance. *J. Chem. Phys.* **1978**, *69*, 386–392. Ruch, E.; Leshe, B. Information Extent & Information Distance. *J. Chem. Phys.* **1978**, *69*, 393–401. Ruch, E. Direction Distance. *Acta Appl. Math.* **1993**, *30*, 67–93.
- (5) Ruch, E.; Gutman, I. The Branching Extent of Graphs. *J. Comb. Inf. System Sci.* **1979**, *4*, 285–295.
- (6) Gutman, I.; Randić, M. Algebraic Characterization of Skeletal Branching. *Chem. Phys. Lett.* **1977**, *47*, 15–19.
- (7) Randić, M. On comparability of structures. *Chem. Phys. Lett.* **1978**, *55*, 547–551.
- (8) Randić, M.; Wilkins, C. L. Graph Theoretical Ordering of Structures as a Basis for Systematic Searches for Regularities in Molecular Data. *J. Phys. Chem.* **1979**, *83*, 1525–1540. Randić, M.; Wilkins, C. L. On a graph-theoretical basis for the ordering of structures. *Chem. Phys. Lett.* **1979**, *63*, 332–336. Randić, M.; Wilkins, C. L. Graph-Theoretical Analysis of Molecular Properties: Isomeric Variations in Nonanes. *Intl. J. Quantum Chem.* **1980**, *18*, 1005–1027. Randić, M.; Trinajstić, N. On Isomeric Variations in Decanes. *MATCH* **1982**, *13*, 271–290.
- (9) Randić, M. Survey of Structural Regularities in Molecular Properties. I. Carbon-13 Chemical Shifts in Alkanes. *Intl. J. Quantum Chem.* **1983**, *23*, 1707–1722.
- (10) El-Basil, S.; Randić, M. Equivalence of Mathematical Objects of Interest in Chemistry & Physics. *Adv. Quantum Chem.* **1992**, *24*, 239–290.
- (11) Randić, M.; Trinajstić, N. Isomeric variations in alkanes: boiling points of nonanes. *New J. Chem.* **1994**, *18*, 179–189.
- (12) Randić, M.; Razinger, M. Molecular Shapes & Chirality. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 429–441.
- (13) Essam, J. W.; Kennedy, J. W.; Gordon, M.; Whittle, P. The Graph-like State of Matter. 8. LCGI Schemes & the Statistical Analysis of Experimental Data. *J. Chem. Soc., Faraday II* **1977**, *73*, 1289–1307.
- (14) Kennedy, J. W.; Gordon, M. Graph Contractions & a Generalized Möbius Inversion. *Ann. N. Y. Acad. Sci.* **1979**, *319*, 331–348.
- (15) Klein, D. J. Chemical Graph-Theoretic Cluster Expansions. *Intl. J. Quantum Chem. S* **1986**, *20*, 153–171.
- (16) Klein, D. J. Similarity & Dissimilarity in Posets. *J. Math. Chem.* **1995**, *18*, 321–348.
- (17) Eigen, M.; Winkler, R. section 8.2 In *Laws of the Game*; Knopf, A. A.; Ed.; New York, 1981.
- (18) Essam, J. W.; Fisher, M. E. Some Basic Definitions in Graph Theory. *Rev. Mod. Phys.* **1970**, *42*, 271–288. Essam, J. W. Graph Theory & Statistical Physics. *Disc. Math.* **1971**, *1*, 83–112. Domb, C. Graph Theory & Embeddings. In *Phase Transitions & Critical Phenomena III*; Domb, C., Green, M. S., Eds.; Academic Press: New York, 1974; pp 1–94.
- (19) See, e.g.: Marshall, A. W.; Olkin, I. *Inequalities: Theory of Majorization & Its Applications*; Academic Press: New York, 1979. Alberti, P. M.; Uhlmann, A. *Stochasticity & Partial Order*; VEB Deutscher Verlag: Berlin, 1981.
- (20) Campbell, N. *What is Science*; Dover Pub.: New York, 1952; Chapter 6.
- (21) Lele, S. R. Some Comments on coordinate-free and scale invariant Methods in Morphometrics. *Am. J. Phys. Anthropol.* **1991**, *85*, 407–418. Lele, S. R.; Richtsmeier, J. T. On comparing biological Shapes: Detection of influential Landmarks. *Am. J. Phys. Anthropol.* **1992**, *87*, 49–65. Radhakrishna Rao, C.; Suryawanshi, S. Statistical analysis of shape of objects based on landmark data. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 12132–12136.
- (22) Dushnik, B.; Miller, E. W. Partially Ordered Sets. *Am. J. Math.* **1941**, *63*, 600–610.
- (23) Trotter, W. T., Jr. A Note on Dilworth’s Embedding Theorem. *Proc. Am. Math. Soc.* **1975**, *52*, 33–39. Lovasz, L.; Nešetřil, J.; Pultr, A. On a Product Dimension of Graphs. *J. Comb. Th. B* **1980**, *29*, 47–67. Hedetniemi, S. T.; Laskar, R. On Irredundance, Dimension, & Rank of Posets. *Ann. N. Y. Acad. Sci.* **1989**, *576*, 235–240.
- (24) Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psych.* **1933**, *24*, 417–441 and 498–520.
- (25) Löwdin, P.-O. On Linear Algebra, the Least Squares Method, & the Search for Linear Relations by Regression Analysis in Quantum Chemistry & Other Sciences. *Adv. Quantum Chem.* **1992**, *23*, 83–126.
- (26) See, e.g.: Garratt, P. J. *Aromaticity*; McGraw-Hill: New York, 1971. Lewis, D.; Peters, D. *Facts & Theories of Aromaticity*; MacMillan & Co.: New York, 1975. Balaban, A. T. Is Aromaticity Outmoded? *Pure Appl. Chem.* **1980**, *52*, 1409–1429. Herndon, W. C.; Nowak, P. C. Nonplanarity & Aromaticity in Polycyclic Benzenoid Hydrocarbons. In *Advances in Theoretically Interesting Molecules*; JAI Press: 1992; pp 113–141. Simkin, B. Y.; Minkin, V. I.; Glukhovtsev, M. N. The Concept of Aromaticity in Heterocyclic Chemistry. *Adv. Heterocyclic Chem.* **1993**, *56*, 303–428.
- (27) Katritzky, A.; Barczynski, P.; Musumarra, G.; Pisano, D.; Szafran, M. Aromaticity as a Quantitative Concept. 1. A Statistical Demonstration of the Orthogonality of ‘Classical’ and ‘Magnetic’ Aromaticity. *J. Am. Chem. Soc.* **1989**, *111*, 7–15.
- (28) Jug, K.; Köster, A. M. Aromaticity as a Multi-Dimensional Phenomenon. *J. Phys. Org. Chem.* **1991**, *4*, 163–169.
- (29) Parkanyi, C.; Boniface, C. Aromaticity: A New Multilinear Approach. *Bull. Soc. Chim. Belg.* **1990**, *99*, 587–594.
- (30) Krygowski, T. M.; Ciesielski, A.; Bird, C. W.; Dotschy, A. Aromatic Character of the Benzene Ring Present in Various Topological Environments in Benzenoid Hydrocarbons. Nonequivalence of Indices of Aromaticity. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 203–210. Krygowski, T. M.; Cyrański, M.; Ciesielski, A.; Swirska, B.; Leszczyński, P. Separation of the Energetic and Geometric Contributions to Aromaticity. 2. Analysis of the Aromatic Character of Benzene Rings in Their Various Topological Environments in the Benzenoid Hydrocarbons. Crystal & Molecular Structure of Coronene. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1135–1141. Cyrański, M.; Krygowski, T. M. Separation of the Energetic and Geometric Contributions to Aromaticity. 3. Analysis of the Aromatic Character of Benzene Rings in Their Various Topological & Chemical Environments in the Substituted Benzene Derivatives. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1142–1145.
- (31) Klein, D. J. Aromaticity via Kekule Structures & Conjugated Circuits. *J. Chem. Ed.* **1992**, *69*, 691–694.
- (32) Mezey, P. G. Semisimilarity of Molecular Bodies: Scaling-Nesting Similarity Measures. *Intl. J. Quantum Chem.* **1994**, *51*, 255–264.
- (33) Randić, M.; Kleiner, A. F.; De Alba, L. M. Distance/Distance Matrices. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 277–286.
- (34) Carbo, R.; Besalu, E.; Amat, L.; Fradera, X. Quantum molecular similarity measures (QMSM) as a natural way leading towards a theoretical foundation of quantitative structure-property relationships. *J. Math. Chem.* **1995**, *18*, 237–246. Carbo, R.; Besalu, E.; Amat, L.; Fradera, X. Quantum molecular similarity measures (QMSM) & indices (QSMI). *J. Math. Chem.* **1996**, *19*, 47–56.
- (35) Mezey, P. G. *Shape in Chemistry*; VCH Press: New York, 1993.

- (35) MacGregor, J. J.; Willet, P. Use of a Maximal Common Subgraph Algorithm in the Automatic Identification of the Ostensible Bond Changes Occurring in Chemical Reactions. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 137–140. Johnson, M. Relating Metrics, Lines & Variables Defined on Graphs to Problems in Medicinal Chemistry. In *Graph Theory & Its Applications to Algorithms & Computer Science*; Alavi, Y., et al. Eds.; John Wiley & Sons: New York, 1985. Barnard, J. M. Substructure Searching Methods: Old & New. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 532–538.
- (36) Zabrodsky, H.; Avnir, D. Measuring Symmetry in Structural Chemistry. *Adv. Molec. Struct. Res.* **1995**, *1*, 1–31.
- (37) Weinberg, N.; Mislow, K. Distance functions as generators of chirality measures. *J. Math. Chem.* **1993**, *14*, 427–450.
- (38) Zabrodsky, H.; Peleg, S.; Avnir, D. Continuous Symmetry Measures. *J. Am. Chem. Soc.* **1992**, *114*, 7843–7851.
- (39) Batagelj, V. Norms and Distances over Finite Groups. *J. Comb., Inf. System Sci.* **1995**, *20*, 243–252.
- (40) Sankoff, D. Matching Sequences under deletion/insertion constraints. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 4–6. Kruskal, J. B. An Overview of Sequence Comparison: Time Warps, String Edits, & Macromolecules. *SIAM Rev.* **1983**, *25*, 201–237.
- (41) Tatevskii, V. M.; Benderskii, V. A.; Yarovoi, S. S. *Rules & Methods for Calculating Physico-Chemical Properties of Paraffinic Hydrocarbons*; Pergamon Press: New York, 1961.
- (42) See, e.g.: Newkome, G. R.; Moorefield, C. N.; Baker, G. R. Building Blocks for Cascade Polymers. Part 9. Building Blocks for Dendritic Polymers. *Aldrich. Acta* **1992** (1992) 1–23. Tomalia, D. A.; Durst, H. D. Genealogically Directed Synthesis: Starburst/Cascade Dendrimers & Hyperbranched Structures. *Top. Curr. Chem.* **1993**, *165*, 194–313.
- (43) Kovats, E. sz. Zusammenhänge zwischen Struktur und gaschromatographischen Daten organischer Verbindungen. *Zeit. Anal. Chem.* **1961**, *181*, 351–364.
- (44) Gordon, M.; Scantlebury, G. R. Non-Random Polycondensation. *Trans. Faraday Soc.* **1964**, *60*, 604–621.
- (45) Randić, M. On characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
- (46) Bonchev, D.; Trinajstić, N. Information theory, distance matrix, and molecular branching. *J. Chem. Phys.* **1977**, *67*, 4517–4533. Bonchev, D.; Knop, J. V.; Trinajstić, N. Mathematical Models of Branching. *MATCH* **1979**, *6*, 21–47. Bonchev, D.; Mekanyan, O.; Trinajstić, N. Topological Characterization of Cyclic Structures. *Intl. J. Quantum Chem.* **1980**, *17*, 845–893.
- (47) Barysz, M.; Knop, J. V.; Pejaković, S.; Trinajstić, N. Characterization of Branching. *Polish J. Chem.* **1985**, *59*, 405–432.
- (48) Bertz, S. H. Branching in Graphs & Molecules. *Disc. Appl. Math.* **1988**, *19*, 65–83.
- (49) Polansky, O. E.; Bonchev, D. The Wiener Number of Graphs. *MATCH* **1986**, *21*, 133–186. Bonchev, D. Topological order in molecules. 1. Molecular branching revisited. *J. Mol. Struct. (Theochem)* **1995**, *336*, 137–156.
- (50) Bonchev, D.; Mekanyan, O. Comparability Graphs & Molecular Properties. *J. Chem. Soc., Faraday Trans. II* **1984**, *80*, 695–712.
- (51) Muirhead, R. F. Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters. *Proc. Edinburgh Math. Soc.* **1903**, *21*, 144–157.
- (52) Dalton, H. The measurement of the inequality of incomes. *Econom. J.* **1920**, *30*, 348–361.
- (53) Shannon, C. E.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois Press: Urbana, IL, 1949.
- (54) Hausdorff, F. In *Set Theory*; Chelsea, New York, 1957; pp 166–168.
- (55) Buda, A. B.; Auf der Heyde, T.; Mislow, K. On Quantifying Chirality. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 989–1007. Buda, A. B.; Mislow, K. A Hausdorff Chirality Measure. *J. Am. Chem. Soc.* **1992**, *114*, 6006–6012.
- (56) Zabrodsky, H.; Avnir, D. Continuous Symmetry Measures. 4. Chirality. *J. Am. Chem. Soc.* **1995**, *117*, 462–473.
- (57) Weinberg, N.; Mislow, K. A unification of chirality measures. *J. Math. Chem.* **1995**, *17*, 35–53.
- (58) Gilat, G. Chiral coefficient – a measure of the amount of structural chirality. *J. Phys. A* **1989**, *22*, L545–550. Gilat, G. On quantifying chirality – Obstacles & problems towards unification. *J. Math. Chem.* **1994**, *15*, 197–205.
- (59) Klein, D. J.; Cantu, A. A. Symmetry Adaptation and the Relation between Spin-Free & Space-Spin Formulations of Electronic Structure. *Intl. J. Quantum Chem.* **1974**, *8*, 223–233.
- (60) Randić, M.; Razinger, M. On Characterization of Molecular Shapes. *J. Chem. Inf. Comput. Sci.* **1996**, *35*, 594–606.
- (61) Harary, F.; Mezey, P. G. Chiral and Achiral Square-cell Configurations; the Degree of Chirality. *New Developments in Molecular Chirality*; Mezey, P. G., Ed.; Kluwer: Dordrecht, 1991; pp 241–256. Mezey, P. G. A Global Approach to Molecular Chirality. *New Developments in Molecular Chirality*; Mezey, P. G., Ed.; Kluwer: Dordrecht, 1991; pp 257–290.
- (62) Mezey, P. G. The Degree of Similarity of 3-Dimensional Bodies: Application to Molecular Shape Analysis. *J. Math. Chem.* **1991**, *7*, 39–49. Mezey, P. G. Similarity Analysis in 2 & 3-Dimensions Using Lattice Animals and Polycubes. *J. Math. Chem.* **1992**, *11*, 27–45.
- (63) Eden, M. A Two-Dimensional Growth Process. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Berkeley, 1961; pp 223–239. Read, R. C. Contributions to the Cell Growth Problem. *Can. J. Math.* **1962**, *14*, 1–20.
- (64) Klein, D. J.; Herndon, W. C.; Randić, M. On the Classification of Polyhex Polymers. *New J. Chem.* **1988**, *12*, 71–76.
- (65) Hamming, R. W. Error detecting & error correcting codes. *Bell. Syst. Tech. J.* **1950**, *29*, 147–160.
- (66) Karfunkel, H. R.; Eyraud, V. An Algorithm for the Representation & Computation of Supermolecular Surfaces & Volumes. *J. Comput. Chem.* **1989**, *10*, 628–634.
- (67) See, e.g.: Mandelbrot, B. *The Fractal Geometry of Nature*; Freeman, W. H.: San Francisco, 1983.
- (68) March, N. H. Private communication.

CI9601776