

Fundamental Difference That Exists between Synthetic versus Analytic Chemical Nomenclature

Seymour B. Elk

Elk Technical Associates, 321 Harris Place, New Milford, New Jersey 07646

Received June 19, 1996[®]

A problem of a fundamental nature in the assignment of a consistent expandable system of chemical nomenclature is that every “synthetic” nomenclature algorithm will eventually, as the allowable permutation of atoms and means of connecting them (i.e., bonds) grows large, produce internal contradictions. This forces setting of limitations to the indiscriminate use of any synthetic system for canonically naming a given molecule. Although such a problem does not arise for an “analytical” algorithm, by definition, such a system is not expandable. A geometrical analog which underlies this problem is noted in an examination of the smallest geometric entity that involves “shape” (the simplex in two dimensions) and the implications to chemistry in the topological difference that exists between SAS and ASA triangle comparisons.

In 1980¹ we noted, but did not thoroughly examine, that there exists a, **far more than just semantical**, difference between a nomenclature formulated by annellating modules to a previously formulated composition of modules, therein called a “synthetic nomenclature”, and one in which we nomenclature² a given molecule by analyzing the entire entity in terms of its modules, therein called an “analytic nomenclature”. Fourteen years later,⁵ in an article giving a simplified nomenclature for cata-condensed polyhex systems, we erred by using a synthetic system for canonically naming a concatenation of polyhexes. This was pointed out⁶ and an erratum page⁷ published. In order to prevent others from similarly assigning an inappropriate nomenclature to chemical molecules, it is desirable that the underlying foundation of the intrinsic difference that exists between a synthetic vs an analytic nomenclature be probed.

As a first step in such an examination, the large permutation of geometrically distinct ways in which atoms can be joined together in the formation of a molecule is noted. The concomitant result of this is that chemical nomenclatures are, for the majority of cases, geometry-based systems. Second, as far as we can ascertain, the (probably unconscious) personal bias of the authors of elementary and even advanced geometry textbooks seems to always be “analytic” in nature. Consequently, they overlook the fact that there exists a subtle, but **major**, difference in the applicability of the different types of comparisons that geometers make between figures, which extends to even the most elementary comparison of simplexes in 2D-space. In particular, let us focus on a property that is topologically inherent in the side–angle–side (SAS) congruence relationship for triangles vs the “nearly-similar”⁸ angle–side–angle (ASA) relationship. This difference arises because of the different topological descriptors that we have fused together in our mental picture of what we mean by a particular figure, i.e., images that we isomorphically interchange. For example, as well as the more traditional definitions in terms of a boundary-defined union of segments of dimension $n - 1$ and the single n -dimensional content-defined figure that the boundary encloses, for a simplex of dimension n , we have an important additional means of

uniquely describing (i.e., defining) an n -simplex in terms of the $n + 1$ “vertex” points. Although such a “definition” is unique for simplexes, for any combination of segments larger than a simplex it is inadequate inasmuch as this set of points could describe multiple figures. In particular, a *triangle* may be defined as a set of three points in an Euclidean space,¹⁰ but a set of four points in a plane can be connected in multiple ways (Figure 1).

At this point, we note that when making an SAS triangle comparison, we begin with a side and at its end point measure off the given angle. Although we could pick an angle in any plane relative to a given three-dimensional (or higher) embedding space, the specification of an angle uniquely determines a plane. Now, because of this planarity, when the second given length is measured and marked off, the figure formed is constrained to lie in the single plane of the given angle. By contrast, for the ASA formulation, since there are two angles, whose *only* requirement is that they have a common side, there is **no logical necessity** that the configuration formed will be coplanar and thus be part of a closed homoloidal figure. Furthermore, even if we were to include the unstated heuristic that the two angles are coplanar, we have the added assumption that these two angles are on the same side of the given side in the given plane, rather than forming a concave sector (“Z”-shaped) of a larger closed union of segments (Figure 2).

An additional item of significance is noted by “analyzing” any three-dimensional figure. Now, in at least half of the possible combinations, specification of ASA will involve four vertices of the “embedding polyhedron” which are noncoplanar. On the other hand, specification of SAS uniquely determines a set of three points, which, in any homoloidal space, determines exactly one plane. The result of this is that, even in large polyhedra, there will always be a subregion of “duplicativity”¹¹ between any two polyhedra having successive side–angle–side equality that is not found in polyhedra having successive angle–side–angle equality. Furthermore, for the 3-simplex (tetrahedron), we note that face angles (F) are the three-dimensional counterparts of the sides (S) in a 2-simplex (triangle) and that dihedral angles (D) correspond to the triangle’s angles (A). Confining focus to simplexes in Euclidean space, we observe that analogous

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

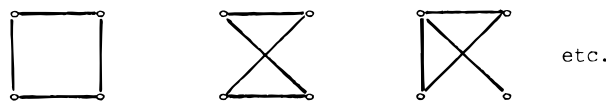


Figure 1.

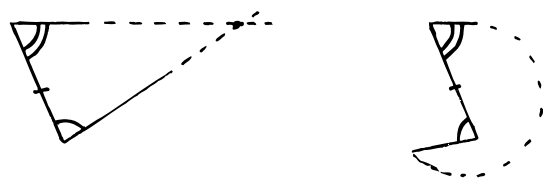


Figure 2.

to SAS vs ASA triangle comparisons, there exist the corresponding FDF vs DFD comparisons of two tetrahedra. Here we note the same limitations as to size of embedding space, namely, FDF determines a 3D-space, whereas DFD may extend into higher space.

We ascribe the omission of this observation not only to the “analysis” vs “synthesis” bias but also to the fact that most textbooks have a focus that is limited to “plane” geometry, in contradistinction to the scientifically more meaningful perspective that the geometry relevant to science is not pigeon-holed in a single dimension, and certainly not in one as small as a plane. In fact, the geometrical picture that is relevant to analysis in chemistry is predetermined by the dimension of the embedding space of the molecule under consideration.¹⁰ Note this is in contradistinction to the corresponding picture for synthesis, in which it is often important to note that the embedding space has the capacity to increase as each additional atom is appended to the previously considered aggregation of atoms. Such an idea had found expression earlier in the formulation of the class of molecules now referred to as “helicenes”,¹² in which the coplanar aggregation of benzene modules was forced into the third dimension by the appropriate positioning of yet another benzene ring to specified combinations of such modules, as well as the similar type extension that we proposed which yields the potential new class of “helicenes”.¹³

Despite our ability to make logical extrapolations of the above paragraph to various esoteric combinations, such as four-dimensional configurations,¹⁴ etc., most of the time—because of our heuristic representation of molecules as points in a physical space—there is the practical limitation that representation of molecules be confined to, at most, a three-dimensional embedding space. Consequently, the main questions of applicability in chemistry will be whether we want to use a three-dimensional embedding space or just a two (or even one)-dimensional portion of it in describing various chemical moieties.

Because of the general truism that the lower the dimension of the embedding space required to uniquely represent a moiety the simpler the description of that moiety will be and thus the easier it shall be to assign a canonical name to it, one “optimal” goal might be to represent every distinct molecule in a zero-dimensional space, i.e., by a unique number. Despite that, to a large extent, this may be accomplished by various “number theory” schemes, such as our use of extended Matula numbers,¹⁵ as well as, when attention is limited to acyclic compounds, by Randic’s compact codes,¹⁶ etc., most chemists rightly consider that, in terms of pragmatism, this has been achieved at too high

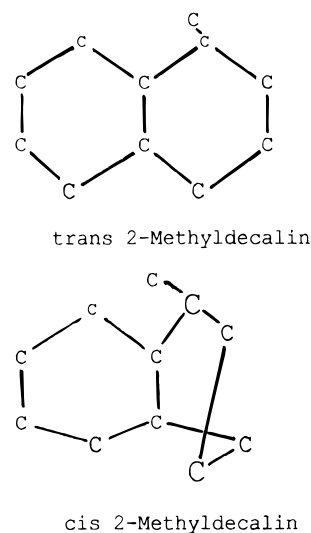


Figure 3.

a price. Instead let us return to a more traditional criteria for an “ideal” nomenclature,¹⁷ as well as to the added property of divisibility for indexing purposes recognized by Goodson.¹⁹

Observe that the very first attribute desired (see item labeled a in ref 17) involves some one-dimensional entity. Furthermore, Polton²⁰ advises of the desirability of such linearity not only for manual manipulation of nomenclature but, as it becomes increasingly dominant, for computer processing. Additionally, we note that although the linear character string so chosen may contain all sorts of information, such a string must internally be further linearly ordered according to some previously *delineated* algorithm.

Progressing now to representation of chemical moieties using a one-dimensional space, one encounters the notations developed by Dyson²¹ and Wiswesser,²² to name but a few. Because of inherent limitations in such a scheme, Goodson makes a distinction between what may be called a “nomenclature” vs a “notation”.²³ Additionally, we note that many limited linear nomenclatures have been evolved over the years, which are primarily of value *either* when an entire class of moieties under consideration is, in fact, actually linear *or*, as is more usual, when such a class is isomorphic with some prescribed linear structure. Although there do exist some *actually geometrically linear* classes of molecules, such as the polyacetylenes, the polyallenes, the acenes, etc., these are the exception. Most systems referred to as “linear”, such as the “straight-chain” alkanes, are misnomers. The vast majority of molecules over three carbons in length involve tri- and tetra-substituted carbon atoms, which form respectively approximately 120° or 109° 28.5’ angles, rather than the 180° angle which defines geometric linearity. Consequently most “organic chemistry” nomenclature schemes, especially the standard IUPAC system,²⁴ ignore the actual geometry angles involved and perpetuate the use of such misnomers on the premise that the molecules involved are isomorphic with an idealized linear system. However, it should be noted that the price paid by such a “simplification” is the sacrifice of realistic Euclidean distance in favor of maintaining only graph theoretic distance—even though space interactions, which are well known in chemistry, are completely ignored by this standard system.²⁵ One needs only look at *cis*- vs *trans*-2-methyldecalin (Figure 3) to notice

the false impression given by using such a nomenclature. Instead of the minimal difference implied by the similarity of names between these two molecules, there is a chemical difference that, in many respects, is every bit as great as between "nearly similar" functional groups, namely, we note the "inherent three-dimensionality" of the cis isomer vs the "effective two-dimensionality" of the trans isomer.

For pragmatic purposes, in most (if not all) single-number systems of nomenclature, the first parameter selected traditionally has been the length of the longest chain of atoms—either restricted to carbon atoms as in IUPAC²⁴ or including all non-hydrogen atoms as in nodal nomenclature.^{26–28} In every such case, this pseudolinearity is simultaneously both a great simplifying factor and a potential source of gross misinterpretation and thus misinformation. In other words, we have to be prepared to pay the price of this simplification and be vigilant about any extension beyond the given set for which the system was originally formulated. This is equivalent to asserting that **every** synthetic nomenclature scheme shall encounter an impasse when new formulations of matter arise that are not "neatly" described by the existing system. On the other hand, unless the present system encompasses the attributes we wish to name, **every** analytic system will be limited to what has already been codified and thus will not be unambiguously expandable to nomenclating any new formulation of matter.

REFERENCES AND NOTES

- (1) Elk, S. B. A Nomenclature for Regular Tessellations and its Application to Polycyclic Aromatic Hydrocarbons. *MATCH* **1980**, 8, 121–158.
- (2) The verb "nomenclate" is not listed in even the larger American dictionaries, such as *Webster's Unabridged Dictionary*,³ but is given in the *Oxford Dictionary*.⁴ Here, however, it is a direct synonym of "to name". To the contrary, in scientific usage, this verb has the connotation of either assigning a name to a given entity within the guidelines of a pre-existing canonical nomenclature, or, if there is not such a nomenclature establishing one.
- (3) *Webster's Third New International Dictionary (Unabridged)*; G & C Merriam Co.: Springfield, MA, 1993. Note that if such a word were to be one that is presently accepted, it should be listed on p 1534.
- (4) *The Oxford English Dictionary*, 2nd ed.; The Clarendon Press: Oxford, U.K., 1989; Volume X, p 470.
- (5) Elk, S. B. A Simplified Algorithm Using Base 5 Numbers to Assign Canonical Names to Cata-Condensed Polybenzenes. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 637–640.
- (6) Muller, W. R.; Szymanski, K.; Knop, J. V.; Trinajstić, N. A Remark on the Naming of Cata-Condensed Benzenoids with Base 5 Numbers. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 759–760.
- (7) Erratum, ref 5. Elk, S. B. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 786.
- (8) To say that two relationships are "nearly anything", such as "nearly similar", "nearly equal", etc., is an oxymoron. Either they have the stated relationship in toto or they do not. The presence of parts that have the desired comparison, while possibly of interest by itself, in no way affects the totality. Basak⁹ demonstrated that, if we allow for the designation of "nearly equal", then BLACK = WHITE.
- (9) Skolnik Award Symposium. 208th American Chemical Society Meeting, Washington, DC, Aug. 23, 1994.
- (10) Elk, S. B. Topologically Different Models to be Used as the Basis for Ring Compound Taxonomy. *J. Chem. Inf. Comput. Sci.* **1985**, 25, 17–22.
- (11) The term "duplicativity" encompasses both congruence and symmetry. Note that it is a more appropriate term for all planar triangle comparisons since rarely is the distinction between congruence and symmetry in a plane made. The rationale for this "inadequacy" is that in a three-dimensional physical world, the transformation of reflection produces an image that is superimposable on the object; consequently, reflection is deemed to be a permissible transformation in a plane but not in a three-dimensional space where image and object are distinct.
- (12) Martin, R. H., The Helicenes. *Angew. Chem., Int. Ed. Engl.* **1974**, 13, 649–660.
- (13) Elk, S. B. Enumeration of the Set of Saturated Compounds that, in Theory, Could be Formed by the Linear Fusion of Regular Pentagonal Modules, Including the Logical Extension to "Helicenes". *J. Mol. Struct. (THEOCHEM)* **1989**, 201, 75–86.
- (14) Elk, S. B. Interpretation of Kuratowski's Theorem in Graph Theory as both a Topological Abstraction and a Chemical Reality. *J. Chem. Inf. Comput. Sci.* **1990**, 30, 69–72.
- (15) Elk, S. B. Expansion of Matula Numbers to Heteroatoms and to Ring Compounds. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 233–236.
- (16) Randić, M.; Nikolic, S.; Trinajstić, N. Compact Codes: On Nomenclature of Acyclic Chemical Compounds. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 357–365.
- (17) An ideal nomenclature should have the following attributes:¹⁸ (a) linear character string (permits lexicographic ordering), (b) formula gives rise to unique name, (c) name permits retrieval of formula, (d) coding process simple, preferably without computer, (e) decoding process simple, (f) coding process not dependent on chemical intuition, (g) names brief, (h) names pronounceable, and (i) names easily comprehensible to chemists.
- (18) Read, R. C.; Milner, R. S. A New System for the Designation of Chemical Compounds for the Purpose of Data Retrieval. I. Acyclic Compounds. Report to University of West Indies, 1968; revised as Research Report CORR-78-42, Department of Combinatorics and Optimization, University of Waterloo, Ontario N2L 391, Canada.
- (19) Goodson, A. L. Graph-Based Chemical Nomenclature. 1. Historical Background and Discussion. *J. Chem. Inf. Comput. Sci.* **1980**, 20, 167–72.
- (20) Polton, D. J. *Chemical Nomenclature and the Computer*; Research Studies Press Ltd.: Taunton, England, 1993; p 2.
- (21) Dyson, G. M. *A New Notation and Enumeration System for Organic Compounds*; 2nd ed.; (Longmans, Green & Co., Ltd.: London, England, 1949.
- (22) Wiswesser, W. J. *A Line-Formula Chemical Notation* Crowell: New York, 1954.
- (23) See ref 19, p 167.
- (24) *International Union of Pure and Applied Chemistry, Nomenclature of Organic Chemistry: Section A*; Pergamon Press: Oxford, U.K., 1979.
- (25) Elk, S. B. Defining Stereochemistry as a Function of Distance. *MATCH* **1994**, 31, 89–97.
- (26) Lozac'h, N.; Goodson, A. L.; Powell, W. H. Nodal Nomenclature -- General Principles. *Angew. Chem., Int. Ed. Engl.* **1979**, 18, 887–899.
- (27) Lozac'h, N.; Goodson, A. L. Nodal Nomenclature II. Specific Nomenclature for Parent Hydrides, Free Radicals Ions and Sunstituents. *Angew. Chem., Int. Ed. Engl.* **1984**, 23, 33–46.
- (28) Goodson, A. L. Application of Graph-Based Chemical Nomenclature to Theoretical and Preparative Chemistry. *Croat. Chem. Acta* **1983**, 56, 315–324.

CI960080J