

Uniparametricity—Why a Canonical Ordering (and a Consistent Nomenclature) for Organic Compounds Continues To Evade Us

Seymour B. Elk

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

Received November 1, 1996[®]

Formulation of a “consistent” system of structure, taxonomy, and nomenclature for chemical molecules continues to elude us precisely because we are not asking the right questions. Advancing knowledge of “ever more curious” molecules has far outstripped our ability to evolve a comprehensive mathematically-consistent system. We are thus left NOT with the question “How can we create an ideal system?” BUT rather with the question “What is the inherent (mathematics) problem that we have not faced up to, as we try to approximate a systematic (algorithmic) way of describing, naming, or merely listing all compounds that we have created either in the laboratory or in theory?”

1. GENERAL INTRODUCTION

Almost all systems of cataloging and ordering of objects (in more mathematical terms we call such objects “elements of a set”) are based on a concept we hereby call “uniparametricity”. One parameter is selected as primary, and all set elements are then arranged according to that parameter. In everyday situations, the two most common parameters chosen are (1) the set of integers and (2) the set of letters of the alphabet. For both of these sets there is a prescribed ordering. From a heuristic perspective, however, there does exist a major difference between these two ordering schemes. This becomes evident when describing the various concepts of interchangeability: Upon comparing two sets of elements described by equal integers, there is usually no difference between “equal” set elements. For example, in a methyl group in chemistry we assert that there is no difference between the three hydrogens; i.e., there is unlimited interchangeability. Comparison of two sets whose elements are described by equal letters, on the other hand, usually does not have this same sense of equality. In a dictionary, for instance, we note that a word may have multiple meanings; however, this does not mean that these various definitions are *in any way* interchangeable. Because of this difference, we create a sequence rule that adds a second parameter to be used when there is equality between two or more set members that were ordered by using the primary rule. For example, we usually use as the second parameter the, admittedly artificial, criterion of which of these definitions is, at the time of compilation, the most commonly used. In a similar manner, when two different words have identical spellings, but different pronunciation, a dictionary has no means of canonically ordering them and again resorts to some, usually the same, artificial criterion. The situation gets even worse in a telephone directory for a large city wherein the logic of the selected ordering of the various persons with a common name, say John Smith (no middle initial or name), often is based on some arbitrarily chosen second parameter such as division of the city into wards, alphabetizing of street names, date when telephone was installed, etc.

2. APPLICABILITY TO CHEMISTRY

Despite all of the various studies of chemical taxonomy and nomenclature that have been done, it is painfully evident that we have, to date, been unsuccessful in devising a usable canonical ordering of the entire set of chemical compounds. What we have achieved is a canonical ordering of some very limited subclasses of molecules as well as a complete (albeit inconsistent and unexpandable) ordering, by fiat, of all molecules.

Examination of this problem from a mathematical perspective reveals that this is a special case of a more general problem:

To tabulate an infinite (or just a large finite) number of elements of a set, that are related by more than one parameter, in a systematic way.

In other words, upon reflection, we note that the molecules that we have successfully ordered, such as the “straight chain” alkanes, have a readily recognizable parameter that we may set into a one-to-one relationship with the set of integers. However, for the case of all chemical molecules or for the more general case of set orderings, the system we wish to order consists of several important parameters—rather than a single one; consequently, the problem becomes much more difficult.

One option that is available, BUT **ONLY** for finite sets, is referred to as the “brute force” method: Focus on first one parameter, go through all possibilities, and then repeat this process for the second parameter, then the third parameter, etc. Note that this technique is of limited value for finite sets and is useless for infinite sets. Instead, let us focus on tabulations that are equally applicable to finite and infinite sequences. Before doing so, however, it should be noted that, not so coincidentally, the brute force method of finite tabulations is exactly the philosophy that organic chemists have traditionally used in definitive reference works, such as The Ring Index¹ and its successors, the Parent Compound Handbook² and the Ring Systems Handbook.³

At this point we observe that a property of greater mathematical significance than whether a given set contains a finite or an infinite number of members is

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

Despite that we shall often be talking about an infinite number of set elements, we assert that the answer to this question is “Yes” whenever these set elements can be systematically arranged in a reproducible order. One such ordered infinite set is the set of natural numbers. We shall use the term “complete” to designate those sets whose elements may be so arranged. Having formulated such a definition, we note that the set of organic compounds⁴—whether we limit attention to compounds that we can actually create in the laboratory or to all that have been theoretically proposed—is an example of an incomplete, but still finite, set. The comparison of this set with the infinite, but complete, set of natural numbers is now noted. In other words, we may readily see a canonical ordering for one specially selected infinite set but are stymied for a selected finite one.

Let us now partition the set of incomplete sets (sets in which, previously unanticipated, new elements are allowed to be added) into “open” vs “closed”, where a closed set is one in which new elements are added **ONLY AFTER** listing all of the “original” elements. One result of adding new elements to a closed set is that the added elements are not in the “correct” sequence with respect to the original set members. For example, the new compounds listed in each of the supplements to The Ring Index have an ordering of their own unrelated to the original ordering. This is a logical consequence whenever one is dealing with closed sets. The more desirable, but harder to handle, open case occurs whenever the ordering is continually updated to collate new elements at points “in the middle”.

3. DECIMAL ORDINAL ORDERING OF SET MEMBERS

A simple pragmatic way to label an open infinite (or finite) set is with a decimal system of ordinal numbers. Note that such a system allows for an infinitely detailed set of insertions—**ONCE** an algorithm setting a priority order has been established. A decimal outline is created, wherein as each entry is made it is assigned a decimal number that is sequential if the new entry is at the same level as the previous entry and which has an extra digit if it is subordinate to one of the previous entries. Such an outline is mathematically sufficient **PROVIDED** that we have a means of ordering any two elements that we may envision. This means that for the telephone directory the different John Smiths are, in fact, ordered by a pre-established algorithm (such as alphabetized street names and then lowest house number if on the same street). In this way, every new name fits in place between two previously listed entries.

The mathematical generalization of the telephone directory is the insertion of an ordinal number in a file, say, between items numbered 1.3.4. and 1.3.5. Here, we envision that, as with decimals, there is a limitless number of implied 0.'s. Consequently, between 1.3.4.0. and 1.3.5.0. we simply use 1.3.4.1., etc. Note that there is no limit to the number of indentation levels and that, by this method, every element in the set may be collated at the appropriate point in the scheme. The open-endedness of such an ordering is not a problem, because, as Georg Cantor in developing the foundations of transfinite sets a century ago demonstrated (in a series of articles that originally appeared in *Mathematische Annalen*)⁵ “every denumerable set of denumerable sets is itself denumerable”; consequently, so long as the analogy is valid, there shall exist a unique correspondence between

| ROMAN (ENGLISH) | GREEK |
|-----------------|---------|
| aardvark | ἀβάρια |
| aardwolf | ἀβάντα |
| abacus | ἀβρό |
| abalone | ἀβλεψία |
| abandonment | ἀβυσσος |

| CYRILLIC (RUSSIAN) | HEBREW |
|--------------------|-------------|
| аааЖУР | א ב ח |
| аБРиКóс | ה ט ז א |
| аБСУРД | י יו ב א |
| аБСUЕСС | ל ה ו פ ב א |
| аВѢ'Н С | т ב ו ב א |

Figure 1.

the set of all organic compounds and the set of natural numbers. Our objective thus reduces to the following: “How do we devise an algorithm that spells out the desired alignment of set members in a “one-to-one” correspondence?”.

Before we are ready to undertake the analogous cataloging of chemical molecules, we examine a limitation that influences our listing of words in a dictionary or names in a telephone directory; namely, these orderings are valid *only so long as* all of the set elements are limited to the use of exactly one alphabet. In other words, that we have exactly a one parametric set on which to base order. For example, when a news maker arises in a country that uses a different alphabet (or worse, a pictographic form of writing, such as many of the Oriental languages), the media is inconsistent in spelling the name until an agreement is formally (or informally) made. A comparable situation exists in the nomenclature of chemistry, when either a new combination of atoms or a different isomerism is discovered. A recent example of this being the way to designate atoms enclosed within a fullerene.

4. A MULTILANGUAGE ORDERING ANALOGY

Because the analogous system is so much simpler than that of the relevant chemistry, let us carry through the much simpler problem posed by the use of different alphabets in order to get an idea what type problems we must solve **OR EVADE** in the case of chemical compounds.

Figure 1 is a list of the first five nouns in dictionaries of four different languages, each using a different alphabet. In order to be able to tabulate these words in an “orderly sequence” we would need either (a) an agreed ordering of the alphabets, say Roman, Greek, Cyrillic, and Hebrew—as was the arbitrary order chosen in Figure 1, or (b) an agreed algorithm for sequencing the alphabets, such as the 1 (mod 4) entry is Roman, the 2 (mod 4) entry is Greek, the 3 (mod 4) entry is Russian, and the 0 (mod 4) entry is Hebrew (Figure 2), or (c) a transformation of all entries into a common alphabet with an assumed equivalence of letters (Figure 3), or etc. At this point it should be noted that mathematically our objective has been refined to that of mapping a finite incomplete open set into an infinite complete open set, and that irrespective, which of the different ways is selected, our ultimate objective is to create a unique ordering by means of some agreed upon canon.

Now, let us consider a multiparameter system—wherein at least two of the parameters are allowed to increase without

aardvark
 àβaria
 ⚡⚡⚡⚡
 ⚡⚡⚡
 aardwolf
 àβávrá
 ⚡⚡⚡⚡⚡
 ⚡⚡⚡
 abacus
 àβró

Figure 2.

aardvark
 aardwolf
 abacus
 àβaria
 ⚡⚡⚡⚡
 abalone
 abandonment
 àβávrá
 ⚡⚡⚡
 àβró

Figure 3.

0,0 — 0,1 — 0,2 — 0,3 — 0,4 — ...
 1,0 — 1,1 — 1,2 — 1,3 — 1,4 — ...
 2,0 — 2,1 — 2,2 — 2,3 — 2,4 — ...
 3,0 — 3,1 — 3,2 — 3,3 — 3,4 — ...
 4,0 — 4,1 — 4,2 — 4,3 — 4,4 — ...

Figure 4.

bound. For such a system we seek the formulation of a “pairing scheme” that reorders the elements so that the first element is paired with the number 1, the second element with the number 2, etc. Such a pairing is called a “one-to-one relation with the set of natural numbers”. Note that such a correspondence cannot be made to accommodate a system in which we first list all single rings up to an unlimited size; next list all combinations of two rings—each up to an unlimited size; etc., as we would not be able to restart the assignment of meaningful positional numbers when we progressed from one to two rings, from two to three rings, etc.

Returning to the case of a multiparameter infinite system, wherein at least two of the parameters may individually increase without bound, such as in Figure 4, it shall become incumbent upon us not just to find a canonical ordering by rearranging the ordering of the elements so that all finite elements of any parameter will be sequenced before allowing any of the parameters to increase without bound but equally importantly to find such an order with the added specification of **scientific relevance**.

Probably, the most familiar cases of ordering multiple parameters, each of which may increase without bound, are those formed using Cartesian coordinates—in the plane, or in three-space, etc. Here the n-tuplet of coordinates are listed according to some preselected parameter. Two simple choices, illustrated for integer values of the coordinates in Figures 5 and 6, respectively, are

the algebraic sum of the coordinates (1)

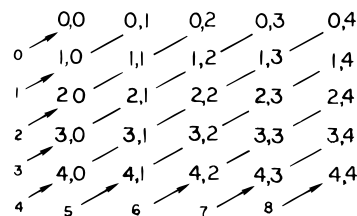


Figure 5.

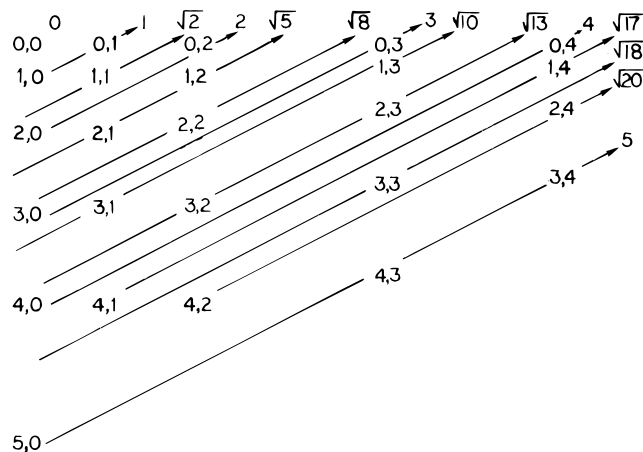


Figure 6.

and

the root-sum-square of the coordinates (2)

Of these two choices, we note that the first has no readily recognizable correlation to any important physical quantity, while the second relates to the idea of distance. In other words, we have a basis (**WHICH MAY OR MAY NOT BE RELEVANT**) for correlating to a physically salient property using the second choice.

In both of the above choices, however, the correspondence with a single parameter, while useful, is not explicit enough for many cases. An additional ordering has to be appended whenever two or more points are expressed by the same value of the first parameter. In simplistic pragmatic terms, this lack of ability to uniquely sequence the various set elements occurs more frequently for “small” combinations in the first example (adding the coordinates—Figure 5) than it does in the distance-ordered second example (Figure 6). If we further arbitrarily limit our attention in the second example to that subset with $x > y$, the first such cases are $d(4,3) = d(5,0)$; $d(5,5) = d(7,1)$, etc. The elimination of ambiguity is effected by the adoption of an additional sequencing rule appended to the previously chosen algorithm, namely the following: “When two or more pairs have the same value of the first parameter (distance, for example) then these are ordered according to some specified second parameter (such as the smaller first coordinate), etc.”.

5. UNIPARAMETRICITY IN SCIENCE AND MATHEMATICS

So long as we are able to describe our system using a single parameter that has a correlation to some idea in physics **OR**, in those cases where it is necessary to use two or more such parameters, if there is a logical reason why there is a preferred ordering of these parameters, i.e., one of the parameters is of greater priority than the others, etc., we are still in the realm of “science” (from the Latin word for

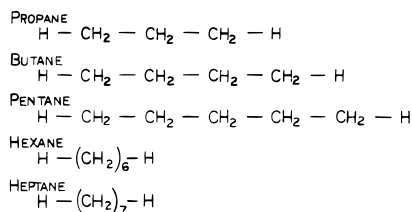


Figure 7.

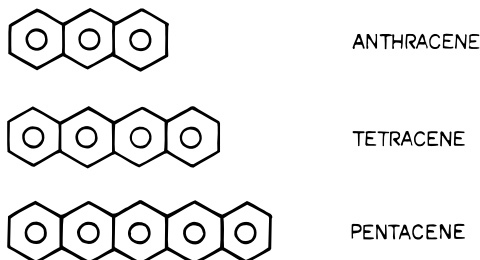


Figure 8.

“knowledge”). As illustration of this latter case, consider polar coordinates: Here the coordinate of the radius is given priority over that of the angle—even though mathematically there is no justification for such a priority. Nevertheless, there does exist a chemistry “justification”; namely, the existence of spherical force fields about point atoms.

This latter idea suggests the importance of spherical coordinates in three-dimensional space. Namely, lacking any more compelling priority, there is a scientifically-suggested (but certainly NOT mandated) selection of primacy implied by ordering first according to radial distance. After having made this selection, however, one might question whether or not there is a consistent scientific basis that justifies the further ordering of importance of the polar angle θ before the azimuthal angle ϕ . This is notwithstanding the fact that there does exist a mathematical logic; namely, as we increase dimension in a polar system, there is a polar angle, but no azimuthal angle in two dimension. Consequently, listing the polar angle first is our heuristic choice for a three-dimensional embedding space.⁶

An important observation to be made before progressing is that there has evolved two different classes of models that are of use in enumerating chemical structure and nomenclature:

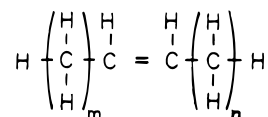
- (1) those which are truly ONLY a one parameter system and which can be used for only a limited set of molecules, such as the “straight chain” alkanes (Figure 7)—even though none of these molecules is geometrically a straight chain,⁷ the acenes (Figure 8),⁸ etc.

and

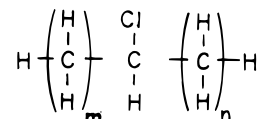
- (2) combinations of various one parameter systems, which are yoked together, often incorrectly and are then applied to, sometimes inappropriate, structures.⁹

The utility of any description or nomenclature of the second class is dependent on our ability to unambiguously order what appears to be disjoint properties with a single “unifying parameter”. For example, although it is a short step to setting up a “logical” basis for ordering other molecules, which are, in fact, two parameter systems, such as the monounsaturated straight chain alkenes, or the straight chain monochlorinated alkanes (Figure 9), etc., a question that we should pose to ourselves is

SOME “TRIVIAL” TWO PARAMETER SYSTEMS



$$\text{FOR ANY } \text{C}_{m+n+2} \text{H}_{2m+2n+4} \\ \text{WITH } m, n \geq 0$$



$$\text{FOR ANY } \text{C}_{m+n+1} \text{H}_{2m+2n+3} \text{Cl} \\ \text{WITH } m, n \geq 0$$

Figure 9.

What price in logic do we pay when we try to order BOTH of these properties SIMULTANEOUSLY?

In other words,

Is the algorithm that establishes priority based on fiat, or is there some inherent scientific relationship that actually does order these different properties in a logical manner?

The answer to this question is often further complicated by our inadequate precision in the use of terms. This has led to an entire series of articles describing the orismological problems¹⁰ encountered in any system of taxonomy. Furthermore, all such systems involve a trade-off between how complex the ordering algorithm (to meet preceived physical correlation) can be and how useful it will be. The more sophisticated any new system of ordering may be, the more it **MAY** be able to accomplish; **HOWEVER, ALSO** the more removed it most probably will be from earlier, firmly entrenched, techniques;¹¹ consequently, the harder it will be to gain acceptance for, what many see as, a “Rube Goldberg”.¹³

6. THE MATULA SYSTEM: A UNIPARAMETRIC NOMENCLATURE—AT TOO HIGH A PRICE (?)

During the past six years, we have perfected a uniparametric nomenclature scheme for organic compounds based on a method proposed by David Matula in 1967 for canonically naming rooted trees.¹⁴ Unlike the usual (*intuitive*) choice of reference (a linear chain of unsubstituted, singly-bonded carbon atoms), Matula noted that the linear chain might be discarded, and, in its place that the counting scheme be based on prime factorization. The result of this was the assignment of a unique integer to every rooted tree. Application of this scheme to organic chemistry required merely selecting which carbon atom should be designated as the “root” in order to “optimally”¹⁵ name that particular alkane.

Matula’s “nomenclature” remained a curiosity for 23 years until we recognized its applicability not only to alkanes but also to *any* one parameter system, in particular, polybenzenes and polymantanes.¹⁶ This was carried further so that any molecule, which could be adequately described by a graph, might now be interpreted as a substituted uniparametric

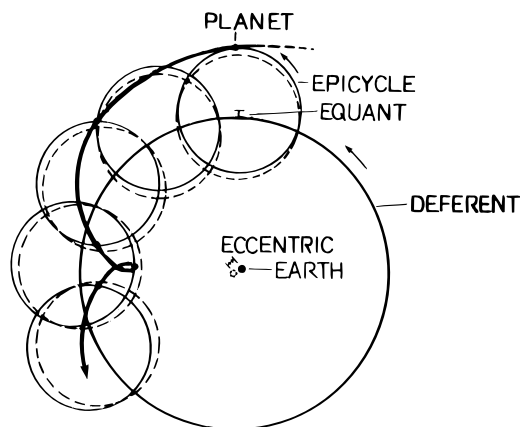


Figure 10.

system with the basic system named using the single parameter and then adding the substitutions in a similar manner.¹⁷

7. FUTURE APPLICABILITY

A question that is applicable not only to use of Matula type nomenclature but also to any other new system that is presented which chemists must answer is whether the price in terms of heuristics (the unfamiliarity of, say, prime number techniques vs the well-established intuitive idea of several linear scales) is too great to use such a "complicated" system, or whether the merits—especially its computer-friendly properties—will outweigh the evident liability.

In the event that mathematical chemists are able to formulate a true correlation between all elements of a large enough set of chemical moieties and a one parameter ordering, we feel that it is worth the hassle to try to have it gain acceptance. However, in most of the physical measuring systems so far devised we have not established such a priority; consequently, unless a reasonably supportable physics criterion can be formulated, the very need for a second parameter has undercut any claim to such a system truly being science. This is equivalent to saying that if some physics-based parameter, such as the center of gravity, moment of inertia, etc. can be found, the process "may" be justified. AND THEN AGAIN IT MAY NOT! Usually, we are limited to pragmatic choices that "fit" presently-familiar compounds. In such instances it is essential that we recognize our limitations. In the worst case scenario, an ordering is imposed in order to fulfill a preconceived (or maybe I should say an ill-conceived) idea. Even though we may find it hard to admit, such orderings are often no better than the appending of an epicycle to an eccentric, etc., a la the pre-Keplerian Ptolemaic system of astronomy¹⁸ (Figure 10).

Despite this very valid objection, the complexity of the problem has seduced us into adopting exactly such an approach in devising a system of nomenclature that we futilely hope is going to be broad enough to encompass all of the known and most of the newly-being discovered molecules. From a realistic perspective, it is extreme egotism to think that we could anticipate, and thus devise a system to account for, all of the new moieties that are to be discovered in the future.¹⁹ It is, therefore, incumbent upon us, as mathematical chemists, to search for sets that fit some relevant aspect of physics that can be used for ordering as

many possible compounds as we can either create in the laboratory or merely just describe mathematically.

REFERENCES AND NOTES

- (1) Patterson, A. M.; Capell, L. T.; Walker, D. F. *The Ring Index*; Chemical Abstracts Service, American Chemical Society: Washington, DC, 1940; 2nd ed.; 1960; three supplements in 1963, 1964, and 1965.
- (2) *Parent Compound Handbook*; Chemical Abstract Services, American Chemical Society: Washington, DC, 1976.
- (3) *Ring Systems Handbook*; Chemical Abstract Services, American Chemical Society: Washington, DC, 1984.
- (4) The set of organic compounds capable of being formed is limited by the number of carbon atoms in the universe, which, even using worse case rough calculations concerning the size of atoms and the size of the universe, is substantially less than a googol (10^{100}). In other words, we have a very large, BUT FINITE, number. The set of compounds that have to-date been theoretically proposed is, in like manner, also a finite number; albeit a much larger one. The reason for this limitation is that no polymer is, in actuality, an infinite chain of atoms—despite that this is the formula often erroneously stated; e.g., polyethelene is not truly $(\text{CH}_2)_\infty$, rather $\text{Q}(\text{CH}_2)_n\text{Q}'$, where the Q's are chain-ending atoms (probably H) and n is some large finite number. Similarly, every physical example of graphite is only a large but not an infinite two-dimensional array of sp^2 hybridized carbon atoms, neither is any diamond an infinite three-dimensional array.
- (5) Young, J. W. *Lectures on Fundamental Concepts of Algebra and Geometry*; Macmillan Company: New York, 1939; p 86.
- (6) Similarly, even though it is not relevant in chemistry, had we been dealing in a physical system described using a higher dimensional space, we would list in order polar angle, azimuthal angle, third angle created in a fourth dimension, etc.
- (7) Elk, S. B. *Connotations Inherent in the Concept of 'Straight' in Geometrical and Molecular Models*. *MATCH* **1988**, 23, 19–33.
- (8) Clar, E. *Polycyclic Hydrocarbons, I & II*; Academic Press: New York, 1964.
- (9) Elk, S. B. Topological Considerations Subtly Inherent in the Formation and Subdivision of Fused vs. Bridged Ring Compounds. *J. Chem. Inf. Comput. Sci.* **1987**, 27, 70–73.
- (10) (a) Elk, S. B. Orismology (the Science of Defining Words) and the Geometrical Foundations of Chemistry. 1. Introduction. *THEOCHEM* **1994**, 313, 199–206. (b) Elk, S. B. Orismology (the Science of Defining Words) and the Geometrical Foundations of Chemistry. 2. Inherent Topology. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 325–330. (c) Elk, S. B. Orismology (the Science of Defining Words) and the Geometrical Foundations of Chemistry. 3. What Constitutes a Ring? *THEOCHEM* **1995**, 358, 119–124. (d) Elk, S. B. Orismology (the Science of Defining Words) and the Geometrical Foundations of Chemistry. 4. Polycyclic Compounds. *J. Chem. Inf. Comput. Sci.* **1996**, 36, 385–392. (e) Elk, S. B. Orismology (the Science of Defining Words) and the Geometrical Foundations of Chemistry. 5. The Heuristics of Primacy. *J. Chem. Inf. Comput. Sci.* Accepted for publication. (f) Elk, S. B. Orismology (the Science of Defining Words) and the Geometrical Foundations of Chemistry. 6. Parent Compounds. pending.
- (11) Max Planck in his autobiography¹² advises the following: "New scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die and a new generation grows up that is familiar with it."
- (12) Planck, M. *Scientific Autobiography and Other Papers*; Williams & Norgate: London, 1950; p 33–34.
- (13) Marzio, P. C. *Rube Goldberg—His Life and Work*; Harper & Row: New York, 1973.
- (14) Matula, D. W. *SIAM Review*. **1968**, 10, 273. (Presented at S.I.A.M. Meeting, Santa Barbara, CA, Nov. 67).
- (15) To the mathematics community the optimal name for any entity is a unique integer. This is exactly the output of applying the Matula algorithm¹⁴ to the configuration of any alkane.
- (16) (a) Elk, S. B. A Canonical Ordering of Polybenzenes and Polymantanes Using a Prime Number Factorization Technique *J. Math. Chem.* **1990**, 4, 55–68. (b) Elk, S. B. A Problem with the Application of Matula's Method of Prime Numbers and Rooted Trees for Canonical Nomenclature of Alkanes. NYAS, Graph Theory Day, 1989; Vol. XVIII, pp 40–43.
- (17) Elk, S. B. Expansion of Matula Numbers to Heteroatoms and to Ring Compounds. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 233–236.
- (18) *Larousse Astronomy*; de la Cotardiere, P., Ed.; Facts on File Publications: New York, 1986; p 11.
- (19) Elk, S. B. The Limitations of Describing Chemical Compounds By Virtue of their Graph Theoretical Representation by a Computer. *MATCH* **1988**, 23, 9–18.