communications in mathematical
match
and in computer chemistry

no. 36, October 1997

ISSN 0340-6253

MATCDY (36) 157-164 (1997)

ORISMOLOGY (THE SCIENCE OF DEFINING WORDS) AND THE GEOMETRICAL FOUNDATIONS OF CHEMISTRY. 6. PARENT COMPOUNDS.

Seymour B. Elk
ELK TECHNICAL ASSOCIATES
321 Harris Place
New Milford, New Jersey 07646

ABSTRACT - Orismology, the science of defining words, plays a major role in understanding concepts in chemistry. Its effects are so pervasive that how different words are denoted, as well as the unstated -- but implied -- connotations that they also carry, often influences the direction in which both laboratory and theoretical chemistry research progresses. An examination of the role that synthetic geometry (including graph theory, topology, etc.) plays not only in describing the physical foundations that underlie the organization and description of chemical moieties, but of equal importance, in helping us to sort out 'just what is our perception of these moieties' is undertaken. This sixth article of the sequence is an examination of 'what do we mean when we say that one compound is a "parent" of another?', and more importantly how do we organize the entire set of chemical compounds so that we formulate a consistent system of taxonomy.

In a series of earlier articles with the common title of "Orismology (The Science of Defining Words) and the Geometrical Foundations of Chemistry"[1], we have examined the implications of several of the words that are important in our understanding of modern chemistry with special emphasis on chemical structure, taxonomy, and nomenclature. Continuing our focusing of attention on the logical development of chemical taxonomy, we now wish to probe which of the millions of known compounds it is either desirable or else expedient to designate as the "starting" point, or "parent", of other compounds.

Although the literal meaning of the word "parent" when applied to compounds in chemistry should be that associated with the reaction:

A + other reactants --> Z + other products

where A is the "parent" and Z the "child", over the years there has evolved a more sophisticated meaning for this term. Despite that in both biology and in common parlance, the primary meaning of the term "parent" is both "binary" and "single generational", such is NOT the orismological development of this term in chemistry. This can be readily seen in the above equation. Note that there is neither need for a single other reactant or other product, nor is there the need or the implication of a one-pot reaction.

Because of the above, a second definition included in most dictionaries[2] is: "the material or source from which something is derived". Cahn & Dermer[3] advise that the original idea behind a "chemical parent" was that of a substance from which a second substance was formulated. In other words, to some the concept of "natural products" was implicit in the usage of the term "parent compound"; while to others "formulation" can be interpreted as mathematical vs. laboratory formulation. That there is ambivalence as to which meaning is intended is shown by these same authors when earlier in their treatise (before they have defined

"parent") they discuss the compound whose formal name is: bis(_-cyclopentadienyl)iron(II), common name = ferrocene; with the statement: "Ferrocene may be used as a parent ..."[4]

One would not be amiss to assert that which of these two diverse usages is intended by the term "parent compound" is dependent on the audience using it. This is illustrated further in a recent article in *Chemical & Engineering News*[5], wherein the writer advises:

> "Three simple analogs of the natural product leinamycin retain the DNA-cleaving activity of the more complex parent compound."

In other words, to the natural products chemistry community the older meaning of the word is paramount.

In contradistinction to this perspective, we observe that although ethanol is usually the source from which ethane is produced and thus one might logically assert that ethanol is "more natural" than ethane; when it comes to chemical nomenclature, the name ethane is deemed to be the parent compound name and ethanol is merely a derived product. The implications of this is that parent *names* do not necessarily reflect "chemical parentage"; i.e., reactant vs. product. Consequently, in order to add the preconceived connotation that is desired for purposes of chemical taxonomy and nomenclature, Fletcher, Dermer and Fox[6] formally define "parent compound" in terms of a substitutive nomenclature[7-8] in which the substituents are atoms or groups of atoms that have replaced hydrogen atoms in a hydrocarbon that is designated as the "parent". It is this latter definition which forms the basis for most of the common usage of the word in "organic" chemistry[9]. In other words, there is an implied "primacy" (see Reference 1, Part 5) based on simplicity of mathematical descriptors, rather than chemical availablity, in that set of compounds which we shall call "parent compounds".

The formal term "parent compound", although used sporadically much earlier, became of significance when Chemical Abstracts decided that an entirely new compilation of known organic compounds was desired, rather than further updatings of The Ring Index (2 Editions)[10] as well as three Supplements thereto[11]. The purpose of this tome, published in March 1977, was spelled out in the introduction as:

> "The *Parent Compound Handbook* is intended as a major reference work for chemists and for those who use *Chemical Abstracts*.

It was divided into two major parts (Part 1. Parent Compound File -- delineated four different types of "parents": Cage Parents, Acyclic Stereo Parents, Cyclic Stereo Parents and Ring Parents, while Part 2. Index of Parent Compounds described seven different indices to use in locating and characterizing organic compounds). An up-dated and detailed description, as well as instructions how to most efficiently use this Parent Compound Handbook, was published three years later. A delineation of its purpose is:

> "... The PCH is both a current-awareness service and a major reference work on ring systems and natural products for those who use Chemical Abstracts (CA) and its associated indexes...."[12]

In other words, the connotations implied by the term *parent compound*, which have become today's de facto standard, relate more to correlation with the various indexes of Chemical Abstracts than to either some specific chemical or mathematical attribute.

Other important remarks concerning our present usage of the term "parent compound" include:

(1) Many (if not most) reactions can be made to go in the reverse direction when the concentrations and external factors are favorable; consequently, any definition in terms of reactants vs. products is apt to lead to inconsistency.

(2) The term "parent compound" is not used in I.U.P.A.C.'s Definitive Rules for the Nomenclature of Inorganic Compounds[13].

(3) The term "parent" is prevalent in "organic" chemistry and always *seems* to involve some sort of **one parameter**[14] system.

Focusing on this last item, in a recent report[15], we introduced the term "uniparametricity" in order to denote a system in which exactly one parameter is paramount in canonically ordering the elements of a specified set. Furthermore, one may observe that the various topological indices[16] were formulated in to order to organize a set of elements according to some single parameter that has been given a position of primacy[17].

The heuristics of primacy influences the choice of organization to be used when trying to establish a canonical nomenclature for organic chemistry. For example, let us focus of some of the systems that have evolved over the years -- with special emphasis on standardization. Beginning with the nomenclature system that has evolved under the auspices of the International Union of Pure and Applied Chemistry, we note that the first section of the nomenclature book for organic chemistry[18] is for Hydrocarbons and that the first subsection is for saturated, unbranched, acyclic hydrocarbons. In other words, even though it was not specifically stated as such, a one parameter system was chosen (probably intuitively, rather than by conscious design) as the starting point to which other compounds will be referenced. Similarly, upon examining other nomenclature systems that have been proposed, we note that, although the parameter which has been selected as *the primary one* may vary, it is, most frequently, a linear one. For

example, in nodal nomenclature[19] the choice of primary parameter is a covalently bonded linear chain of connected "nodes", where a node is either a non-hydrogen atom (both carbon and heteroatoms included) or a larger aggregation of atoms that function as a unit under some perspective, such as a benzene ring. Furthermore, this chain is considered without regard to bond multiplicity.

We feel that the importance of linearity per se, in contradistinction to uniparametricity, has been greatly exaggerated. Polton[20], for example, advises:

> "For computer processing, chemical structures must be condensed into a linear form, a string of characters."

A linear form is just one, admittedly intuitive and very convenient, form of uniparametricity. A second uniparametric system that we have used and helped develop involves the use of prime factorization and the subsequent creation of nomenclatures based on Matula[21] numbers. From a theoretical "Weltanschauung" such a system should not be overlooked, despite that, from a pragmatic perspective, it is highly unlikely that the chemistry community is ready now, or at any time in the foreseeable future, to agree to accept such a nomenclature scheme.

Returning to the ideas that underlie linearity, we see that IUPAC[18] and those who followed this line of thought such as Lozachʼh[19], Dyson[22], etc., next addend a second linear string to the originally designated primary linear string, etc. thereby beginning the development of a complex algorithm that shall be used to nomenclate[23] a given molecule. It is, however, precisely this need to addend a second parameter that leads to the consistency problems encountered. The reason for such problems is the fact that there is no mathematically proscribed method for such addending. To the contrary, whatever specific bias is suggested by the given sample will be built into the system.

REFERENCES AND NOTES

1. Part 1. Introduction, *J.Mol.Struct. (THEOCHEM)* **1994**, 313, 199-206.
   Part 2. Inherent Topology, *J.Chem.Inf.Comput.Sci.* **1994**, 34, 325-330.
   Part 3. What Constitutes a Ring, *J.Mol.Struct. (THEOCHEM)* **1995,** 358, 119-124.
   Part 4. Polycyclic Compounds, *J.Chem.Inf.Comput.Sci.* **1996**, 36, 385-392.
   Part 5. The Heuristics of Primacy, accepted for publication in *J.Chem.Inf.Comput.Sci.*, **1997**, 37.

2. **Webster's New Collegiate Dictionary**, G & C Merriam Co; Springfield, Mass., 1981, p.828.

3. R. S. Cahn & O. C. Dermer, **Introduction to Chemical Nomenclature**, 5-th Ed., Butterworths, London, 1979, p.43.

4. Ibid, p. 32.

5. *Chemical & Engineering News*, Feb 26, 1996, p.38.

6. J. H. Fletcher, O. C. Dermer, R. B. Fox, **Nomenclature of Organic Compounds - Principles and Practice**, Adv.Chem.Ser. 126, Am.Chem.Soc.; Washington, D.C., 1974, p. 71.

7. Cahn & Dermer[8] describe several fundamentally distinct types of nomenclature that are common to organic chemistry. These include: (1) substitutive; (2) conjunctive (also called "additive"); (3) combined substitutive - additive; (4) radicofunctional; (5) replacement; (6) subtractive; (7) trivial; and (8) semi-trivial.

8. Ibid #3, pp. 55-57.

9. Excluding the boron compounds, which have many of the attributes that we originally (historically) had associated with "organic" vs. "inorganic" chemistry, we have been unable to find any consistent usage of the term "parent compound" in the more traditional domain allocated to "inorganic chemistry".

10. A. M. Patterson, L. T. Capell, D. F. Walker, "The Ring Index" (1-st Ed.), Am. Chem. Soc., Washington, D. C., 1940. Second Edition, 1960

11. Supplements published in 1963, 1964 and 1965.

12. J. E. Blake et al, "Parent Compound Handbook -- Successor to The Ring Index", *J.Chem.Inf.Comput.Sci.*, 20, 1980, 162-167.

13. International Union of Pure and Applied Chemistry, **Nomenclature of Inorganic Chemistry**: 2-nd Edition; Butterworths: London, U.K., 1970.

14. S. B. Elk, "A Canonical Ordering of Polybenzenes and Polymantanes Using a Prime Number Factorization Technique", J.Math.Chem., 4, (1990) 55.

15. S. B. Elk, "A Canonical Ordering of Polybenzenes and Polymantanes Using a Prime Number Factorization Technique", J.Math.Chem., 4, (1990) 55.

16. See Footnote #2 of Reference #1, Part 2

17. Ibid #1, Part 5, page 8 of manuscript

18. International Union of Pure and Applied Chemistry, **Nomenclature of Organic Chemistry**: Section A; Pergamon Press: Oxford, U.K., 1979.

19. N. Lozac'h, A. L. Goodson, W. H. Powell, "Nodal Nomenclature - General Principles", Ange.Chem.-Int.Ed., 18, (1979), 887-899.

20. D. J. Polton, **Chemical Nomenclature and the Computer**; John Wiley: New York 1993, p. 2.

21. a. D. W. Matula, "A Natural Rooted Tree Enumeration by Prime Factorization", SIAM Rev. **1968**, 10, 273.
    b. S. B. Elk, "A Canonical Ordering of Polybenzenes and Polymantanes Using a Prime Factorization Technique", *J.Math.Chem.* **1990**, 55-68.
    c. S. B. Elk, "Expansion of Matula Numbers to Heteroatoms and to Ring Compounds", *J.Chem.Inf.Comput.Sci.* **1995**, 35, 233-236.

22. G. M. Dyson, "Some New Concepts in Organic Chemical Nomenclature": Subcommittee Report, Chemical Structures Assn., 1983, p. D4/21.

23. Denotations and connotations of the verb to "nomenclate" are described in detail in Footnote #11 of Part 4 of this series of reports (Reference #1).