The Atom Connectivity Matrix (ACM) and Its Characteristic Polynomial (ACMCP)*

By LEONARD SPIALTER
Chemistry Research Laboratory, ARL, Wright-Patterson Air Force Base, Ohio
Received February 24, 1964

INTRODUCTION

Chemical pictographs or molecular diagrams are the natural language in which chemists describe molecules and reactions thereof. Unfortunately, such pictographs do not lend themselves to direct filing and comparison primarily because dimensions, orientation, and angles are not important fundamentals in the usual case. The essential components are the constituent atoms and the bonds between them (in the current view of molecular structure). The need to develop a reference capability for pictographs has led to many schemes and manipulations thereof based on some sort of accounting or atom-bonding balance sheet.

One family of such schemes includes those which consider molecules as being connected collections of "readily-recognizable" local assemblies of atoms such as chains, *n*-membered rings, functional groups (carbonyl, carboxylic acid, etc.), and the like. To such local assemblies are assigned characteristic symbols (digits, letters, or marks) so that the complete molecule is then simply represented by a collection of such symbols with suitable additional symbolism to describe their interconnections. To this particular family belong all the linear cypher notations such as those of Wiswesser,[1] Dyson,[2] IUPAC,[3] Bouman,[4] Gruber,[5] Gordon, *et al.*,[6] Silk,[7] and the like. The advantages of such systems include their compactness, ready correlation with molecular structure and current naming systems, and apparent ease of fragment searching or handling of partial indeterminates. On the debit side are the requirements for services of a trained intermediary for "molecular dissection," multiplicity of rules as to seniority or hierarchy of presentation and local conventions, necessity for a glossary or dictionary to define symbols, use of arbitrary nonchemically significant symbols and marks which are not meaningful or recognizable in all languages and which may conflict with standard chemical symbols, non-universality in that only currently known structures have defined rules and new types of structures require rules supplementation, searching difficulties (ameliorated only partially by the permuted line entries concept[8]), and lack of good ordering conventions for a list of notations for different molecules. Another drawback is the lack of compatibility with the most widely used, chemically significant, and fundamental indexing parameter even for compounds of unknown structure, the molecular formula (although a suggested means of deriving this from the subject notations is implicit in a recent thesis and publication by Garfield[9]).

It must be recognized that all of the above schemes are simply extrapolations of the usual verbalizable chemical nomenclature with essentially all of the advantages and limitations thereof. It is good for relatively rapid communication between chemists knowledgeable in the rules, but fails when language barriers must be crossed or when one universally clear, unique, and acceptable name is to be placed in an ordered file.

Alternate and, possibly, increasingly more popular methods of describing molecular pictographs are those based on some sort of topological mapping of the individual atoms and the bonds between them. Into this category fall the "accounting" techniques of Ballard and Neeland,[10] and Gluck and Rasmussen,[11] the Polish tree-based method of Hiz,[12] and the matrix-based concepts of Ray and Kirsch,[13] and Sussenguth.[14]

It appears likely that such a representation based only on atoms and interconnecting bonds is the simplest mathematical transformation of a chemical pictograph. By "simplest" is meant that a minimal number of rules is required, essentially no more additional symbols than the usual chemical ones are utilized, and a clear-cut one-to-one correspondence with the chemical pictograph is obtained.

The present paper is devoted to a concept which belongs to the latter class of representations. It is initially based on a matrix construct which is then converted by standard algebraic techniques to a derivative form which has

* Portions of this paper were presented before the Division of Chemical Literature, 145th National Meeting of the American Chemical Society, New York, N. Y., Sept. 11, 1963, and before the Chemical Information and Data System (CIDS) Phase IV Workshop, U. S. Army Missile Command, Huntsville, Ala., Nov. 12–14, 1963.

(1) W. J. Wiswesser, "A Line Formula Chemical Notation," T. Y. Crowell Co., New York, N. Y., 1954.

(2) G. M. Dyson, "A New Notation and Enumeration System for Organic Compounds," 2nd Ed., Longmans, Green and Co., New York, N. Y., 1958.

(3) International Union of Pure and Applied Chemistry, "Rules for I. U. P. A. C. Notation for Organic Compounds," John Wiley and Sons, Inc., New York, N. Y., 1961.

(4) H. Bouman, *J. Chem. Doc.*, **3**, 92 (1963).

(5) W. Gruber, *Angew. Chem.*, **61**, 429 (1950); "Die Genfer Nomenklatur in Chiffren," Beihefte *Angew. Chem. Chem.-Ing. Tech.*, No. **58**, 1950.

(6) M. Gordon, C. E. Kendall, and W. H. T. Davidson, "Chemical Ciphering," Royal Institute of Chemistry Monograph, 1948.

(7) J. A. Silk, *J. Chem. Doc.*, **3**, 189 (1963).

(8) P. F. Sorter, C. E. Granito, J. C. Gilmer, A. Gelberg, and E. A. Metcalf, *ibid.*, **4**, 56 (1964); 145th National Meeting of the American Chemical Society, New York, N. Y., Sept. 8–13, 1963, Abstracts 12G.

(9) E. Garfield, "An Algorithm for Translating Chemical Names to Molecular Formulas," Institute for Scientific Information, Philadelphia, Pa., 1961.

(10) D. L. Ballard and F. Neeland, *J. Chem. Doc.*, **3**, 196 (1963); *Chem. Eng. News*, **41**, 129 (April 15, 1963); 144th National Meeting of the American Chemical Society, Los Angeles, Calif., March 31–April 5, 1963, Abstracts 9F.

(11) *Chem. Eng. News*, **41**, 35 (Dec. 9, 1963).

(12) H. Hiz, *J. Chem. Doc.*, **4**, 173 (1964).

(13) L. C. Ray and R. A. Kirsch, *Science*, **126**, 814 (1957).

(14) E. H. Sussenguth, Jr., U. S. Army-Sponsored Conference on Chemical Information and Data System (CIDS), U. S. Army Missile Center, Huntsville, Ala., Nov. 12–14, 1963; 147th National Meeting of the American Chemical Society, Philadelphia, Pa., April 5–10, 1964, Abstracts 3F.