(4) Hosoya, H. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332.
(5) Randić, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
(6) Balaban, A. T. *Theor. Chim. Acta* **1979**, *53*, 355.
(7) Razinger, M.; Chretien, J. R.; Dubois, J. E. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 23.
(8) Stuper, A. J.; Brugger, W. E.; Jurs, P. C. *Computer-Assisted Studies of Chemical Structure and Biological Function*; Wiley: New York, 1979.
(9) Program runs on Apple IIe home computer. Revised version involving prime number weights is available upon request to noncommercial users.
(10) Figueras, J., a preprint, private communication.
(11) Szymanski, K., Müller, W.R., Knop, J. V., Trinajstić, N., a preprint, private communication.
(12) Knop, J. V.; Müller, W. R. Jericević, Z.; Trinajstić, N. *J. Chem. Inf. Comput. Sci.* **1983**, *21*, 91.
(13) Triansjtić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1983; Vol. II, Table 4, 153.
(14) Randić, M., submitted for publication in *Croat. Chem. Acta*.

Compact Molecular Codes†

MILAN RANDIĆ*

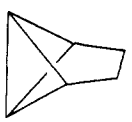
Department of Mathematics and Computer Science, Drake University, Des Moines, Iowa 50311, and Ames Laboratory, Iowa State University, Ames, Iowa 50011

Received May 22, 1985

In this paper we introduce structural codes that are easy to derive for most molecular skeletal forms of chemical interest yet satisfy numerous desirable properties, being linear, unique, reconstructable, derivable and decodable by hand, brief, based on familiar symbols, easily comprehensible, and efficient. Codes in general imply a resolution of the following problems: (1) canonical numbering of atoms; (2) graph isomorphism; (3) discernment of the symmetry of the structure (graph). Our approach resolves these problems in a remarkably simple way, at least for the examples selected. The approach is based on an extension of the *N*-tuple codes of Knop and co-workers, which apply only to trees (acyclic graphs). By excising selected vertices in a polycyclic graph, one arrives at subspanning trees for the polycyclic graph for which *N*-tuple codes of Knop et al. are adopted. Subsequently, such an incomplete code is augmented by the listing of adjacencies for the vertices, which represent ring closures. This paper presents numerous illustrations of the compact codes and discusses the rules that govern construction of the compact codes and the relative ease of the search for the codes. In order to more clearly show the relative simplicity of the new codes, we end with a comparison of the compact codes with a selection of alternative codes currently in use.

INTRODUCTION

The history of chemical nomenclature and the search for codes with desirable qualities is old and continuing. As early as 1881, Friedrich Konrad Beilstein[1] initiated a nomenclature system that is still of interest and serves as a basis for the naming of numerous structures. In 1900, Adolph von Bäyer[2] suggested the nomenclature for bridged bicyclic molecules, which is still the basis for the systematic naming of compounds like norbornane etc. Already, the extension of the nomenclature to tricyclic systems pointed to some difficulties. Besides *digits* used to indicate the number of carbon atoms in individual bridges, one needs *labels* to indicate the particular bridges in polycyclic structures. For example



is named (by IUPAC rules) tricyclo[3.1.0.0^{2,6}]hexane. Observe two *kinds* of uses of digits: 3.1.0.0 indicates *structural* data, the number of carbon atoms in the four branches of the structure, and 2,6 is a *label* referring to selected carbon atoms.

Much progress followed the early interest in chemical nomenclature. Coding is important not only for chemical documentation but also for enumeration of isomers and the construction of graphs. Finally, structural codes are of interest in structure–property and structure–activity studies.[3] Recently, Read[4] reviewed desirable qualities for codes for chemical structures. These are listed in Table I. Various codes proposed in the past satisfy to some degree several of the suggested desirable features, but no code has been found that would satisfy all the requirements satisfactorily. Not all the requirements are, however, equally important, nor can they be resolved with similar efforts. Of the attributes required of codes, according to Goodson,[5] the ones most difficult to comply with are that names be based on linear character strings to permit lexicographic ordering and that names be brief. This means that codes should be short and that standard symbols (e.g., digits, letters, and other common mathematical or typographical symbols, such as brackets, slashes, asterisks, etc.)

Table I. List of Requirements of Codes as Proposed by Read[4]

(1) codes should be a linear string of symbols
(2) coding algorithm should produce a unique code
(3) structure should be recoverable by a clearly defined process
(4) coding should be simple; preferably, it should be possible to code a compound by hand (without the use of a computer)
(5) decoding process should be simple, preferably one that can be carried out by hand
(6) coding process should not depend on chemical intuition or properties of chemicals
(7) coding should not depend on any list of names or other nonsystematic items
(8) codes should be brief
(9) codes should be pronounceable
(10) symbols used should be familiar (available on standard typewriter or computer keyboard)
(11) codes should be easily comprehensible
(12) coding and decoding algorithms should be efficient

† Dedicated to Professor Vladimir Prelog on the occasion of his 80th birthday. Ames Laboratory is operated by the Iowa State University for the U.S. Department of Energy, under Contract W-7405-Eng-82, Division of Basic Sciences. This work is supported in part by the Office of the Director.

* Address correspondence to the author at Drake University.

0095-2338/86/1626-0136$01.50/0