

Compact Codes: On Nomenclature of Acyclic Chemical Compounds[†]

Milan Randić*

Department of Mathematics and Computer Science, Drake University, Des Moines, Iowa 50311

Sonja Nikolić and Nenad Trinajstić

The Rugjer Bošković Institute, P.O.B. 1016, HR-41001 Zagreb, The Republic of Croatia

Received August 24, 1994[®]

The aim of this report is to demonstrate that the compact codes recently introduced have the potential to develop into a full new system of chemical coding. Our proposal is based on the N-tuple concept developed by Knop and his co-workers and its transformation into the M-tuple which consists of a string of numbers, each representing either primary, secondary, tertiary, or quaternary carbon atom. The conversion from N-tuple to M-tuple is simple, but the new code allows chemically more meaningful interpretation. The proposed novel scheme was successfully applied to acyclic chemical systems such as alkanes and compounds with heteroatoms and multiple bonds.

INTRODUCTION

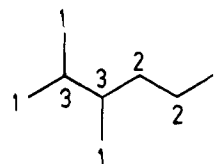
With the growth of chemistry the need for *standard* names was already apparent in the last century when several attempts, in particular by Beilstein in St. Petersburg,¹ were made to arrive at practical chemical nomenclature. More recently the area of designation of chemical compounds attracted attention of people with various interests and different backgrounds: organic synthetic chemists, chemical documentalists, computer scientists, theoretical chemists, mathematical chemists, and even mathematicians.² On one side, this not only indicates that the subject may be more complicated than it would seem to uninitiated persons^{3,4} but also gives promise that talent from such diverse backgrounds may eventually produce useful and meritorious solutions, particularly if useful elements from one scheme are combined with useful elements from another. Professor Balaban and his group are particularly active in this area.⁵ They developed a new procedure for the ordering and generation of trees and alkanes using the GENLOIS program.

We outline here one such novel approach to designation of chemical compounds, which follows from the mathematical code, called the N-tuple code, introduced by Knop and co-workers⁶⁻⁹ for computer enumeration and generation of trees (i.e., acyclic structures). We feel that the task of developing a system for the designation of chemical compounds should be separated into two phases: (1) deriving the scheme for naming acyclic structures (trees) and (2) extending such a scheme to polycyclic compounds. The reason for this subdivision is that each of the two phases involve *different* aspects of the coding process, and while one author may have an excellent way for extending codes from acyclic to polycyclic systems, her/his initial acyclic codes need not be optimum. On the other hand, another person may have an elegant system of names for trees but a clumsy extension to polycyclics. In both cases one could improve the situation by combining the best of each system into a novel one. For instance, one can try to have a system based on acyclic codes of one scheme with an elegant generalization to polycyclics due to an entirely different

source. The scheme introduced here represents a "chemical" modification of the graph-theoretical N-tuple code⁶ which results in novel systems for designation of acyclic structures. Extension of the N-tuple code to polycyclic systems can be achieved by using the concept of the subspanning tree, as had already been outlined.¹⁰⁻¹³

AN OUTLINE OF THE NOVEL ACYCLIC NOMENCLATURE

We start with the N-tuple code for trees (acyclic graphs) as proposed by Knop and co-workers which we will illustrate on a simple graph of 2,3-dimethylhexane (the graph-theoretical degrees of vertices are indicated):



As is customary in chemical graph theory^{14,15} one suppresses hydrogen atoms and considers the carbon skeleton alone. Hydrogen atoms are implied by the valency saturation of carbon atoms. The task is to produce 8-tuple, a sequence which will *uniquely* characterize the molecular graph corresponding to the carbon skeleton of 2,3-dimethylhexane.

The N-tuple code consists of a string of numbers, each representing the degree *d* of a vertex in a tree or subtrees. To obtain the N-tuple code, one first has to identify the vertices of the highest valency and select among them one that will result in a code that produces lexicographically the largest number. After the starting vertex is located, that vertex and adjacent edges are removed. The subtrees thus produced are examined. Typically, this means looking for the largest chains, and, if several of the same length appear, their codes are derived and combined in such a way that the result corresponds to the lexicographically highest number.

Observe that the process *induces* numbering of vertices, each place in the sequence belongs to a single vertex (atom), and its ordinal number becomes the canonical label for the vertex.

[†] Dedicated to Professor A. T. Balaban (Bucharest), the winner of the ACS Division of Chemical Information 1994 Hermann Skolnik Award.

[®] Abstract published in *Advance ACS Abstracts*, February 1, 1995.

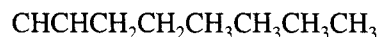
Table 1. Alkanes C₂–C₇ and Their Codes

molecule	N-tuple	M-tuple	formula	contraction	polish notation
ethane	1 0	1 1	CH ₃ CH ₃	(CH ₃) ₂	2(CH ₃)
propane	2 0 0	2 1 1	CH ₂ CH ₂ CH ₃	CH ₂ (CH ₃) ₂	(CH ₂)2(CH ₃)
butane	2 1 0 0	2 2 1 1	CH ₂ CH ₂ CH ₂ CH ₃	(CH ₂) ₂ (CH ₃) ₂	2(CH ₂)2(CH ₃)
2-methylpropane	3 0 0 0	3 1 1 1	CHCH ₃ CH ₃ CH ₃	CH(CH ₃) ₃	(CH)3(CH ₃) ₃
pentane	2 1 1 0 0	2 2 2 1 1	CH ₂ CH ₂ CH ₂ CH ₂ CH ₃	(CH ₂) ₃ (CH ₃) ₂	3(CH ₂)2(CH ₃)
2-methylbutane	3 1 0 0 0	3 2 1 1 1	CHCH ₂ CH ₂ CH ₂ CH ₃	CHCH ₂ (CH ₃) ₃	(CH)(CH ₂)3(CH ₃)
2,2-dimethylpropane	4 0 0 0 0	4 1 1 1 1	CCH ₃ CH ₃ CH ₃ CH ₃	C(CH ₃) ₄	(C)4(CH ₃)
2-hexane	2 1 1 1 0 0	2 2 2 2 1 1	CH ₂ CH ₂ CH ₂ CH ₂ CH ₂ CH ₃	(CH ₂) ₄ (CH ₃) ₂	4(CH ₂)2(CH ₃)
2-methylpentane	3 1 1 0 0 0	3 2 2 1 1 1	CHCH ₂ CH ₂ CH ₂ CH ₂ CH ₃	CH(CH ₂) ₂ (CH ₃) ₃	(CH)2(CH ₂)3(CH ₃)
3-methylpentane	3 1 0 1 0 0	3 2 1 2 1 1	CHCH ₂ CH ₂ CH ₂ CH ₂ CH ₃	CH(CH ₂ CH ₃) ₂ CH ₃	(CH)2(CH ₂ CH ₃) ₂ (CH ₃)
2,3-dimethylbutane	3 2 0 0 0 0	3 3 1 1 1 1	CHCHCH ₃ CH ₂ CH ₂ CH ₃	(CH) ₂ (CH ₃) ₄	2(CH)4(CH ₃)
2,2-dimethylbutane	4 1 0 0 0 0	4 2 1 1 1 1	CCH ₂ CH ₂ CH ₂ CH ₂ CH ₃	CCH ₂ (CH ₃) ₃	(C)4(CH ₂)3(CH ₃)
2-heptane	2 1 1 1 1 0 0	2 2 2 2 2 1 1	CH ₂ CH ₂ CH ₂ CH ₂ CH ₂ CH ₂ CH ₃	(CH ₂) ₅ (CH ₃) ₂	5(CH ₂)2(CH ₃)
2-methylhexane	3 1 1 1 0 0 0	3 2 2 2 1 1 1	CHCH ₂ CH ₂ CH ₂ CH ₂ CH ₂ CH ₃	CH(CH ₂) ₃ (CH ₃) ₃	(CH)3(CH ₂)3(CH ₃)
3-methylhexane	3 1 1 0 1 0 0	3 2 2 1 2 1 1	CHCH ₂ CH ₂ CH ₂ CH ₂ CH ₂ CH ₃	CHCH ₂ (CH ₂ CH ₃) ₂ CH ₃	(CH)(CH ₂)2(CH ₂ CH ₃)(CH ₃)
3-ethylpentane	3 1 0 1 0 1 0	3 2 1 2 1 2 1	CHCH ₂ CH ₂ CH ₂ CH ₂ CH ₂ CH ₃	CH(CH ₂ CH ₃) ₃	(CH)3(CH ₂ CH ₃)
2,2-dimethylpentane	4 1 1 0 0 0 0	4 2 2 1 1 1 1	CCH ₂ CH ₂ CH ₂ CH ₂ CH ₂ CH ₃	C(CH ₂) ₂ (CH ₃) ₄	(C)2(CH ₂)4(CH ₃)
2,3-dimethylpentane	3 2 1 0 0 0 0	3 3 2 1 1 1 1	CHCHCH ₂ CH ₂ CH ₂ CH ₂ CH ₃	(CH) ₂ CH ₂ (CH ₃) ₄	2(CH)(CH ₂)4(CH ₃)
2,4-dimethylpentane	3 1 2 0 0 0 0	3 2 3 1 1 1 1	CHCH ₂ CHCH ₂ CH ₂ CH ₂ CH ₃	CHCH ₂ CH(CH ₃) ₄	(CH)(CH ₂)(CH)4(CH ₃)
3,3-dimethylpentane	4 1 0 1 0 0 0	4 2 1 2 1 1 1	CCH ₂ CH ₂ CH ₂ CH ₂ CH ₂ CH ₃	C(CH ₂ CH ₃) ₂ (CH ₃) ₂	(C)2(CH ₂ CH ₃)2(CH ₃)
2,2,3-trimethylbutane	4 2 0 0 0 0 0	4 3 1 1 1 1 1	CCHCH ₃ CH ₂ CH ₂ CH ₂ CH ₃	CCH(CH ₃) ₅	(C)(CH)5(CH ₃)

Some 20 years ago Hiz¹⁶ considered codes, called cyphers, for trees that have similar character, adopting the so-called Polish notation for logic. [The so-called Polish reverse notation was introduced by Polish mathematician-logician Lukasiewicz, who is also responsible for one of the basic 16 binary operations of mathematical logic. "Reverse" indicates the fundamental aspect of the notation, in which arguments are preceded by indicating the operation first. The label "Polish" presumably suggests difficulties or inertness on the part of English speaking communities to learn some multisyllabic names originating elsewhere. Hence, the Polish notation is in fact the Lukasiewicz notation!] The cyphers for trees of Hiz, however, lacked the important ingredient introduced by Knop *et al.*⁶ the criterion for selecting one of many alternative sequences as canonical, which would then result in fewer alternative names to be screened and a single name to be found as canonical. In the above case of a tree corresponding to the carbon skeleton of 2,3-dimethylhexane, one easily finds that the longest chain starts at the tertiary carbon at the left, the degree of which is 3. This is to be followed by the sequence: 2 1 1 0 (which stands for vertices of the chain having valency 3 2 2 1, respectively, i.e., representing another tertiary carbon, two secondary carbon atoms, and a primary carbon atom). Upon reaching the end of the chain one backtracks to the previous branching and records those atoms, then backtracks again and indicates at each branching point all atoms involved, listing chains associated with the largest codes first. Hence, we obtain for the 2,3-dimethylhexane tree as the N-tuple code: 3 2 1 1 0 0 0 0. Each number (as in the case of the code by Hiz) represents one atom, the space between the digits corresponds to a bond. Each 0 indicates a termination of a chain.

We propose the following modification of the N-tuple code of Knop *et al.*: Instead of writing down (d – 1) for all atoms except the first, we will write d for all atoms. Thus, the above N-tuple code for the 2,3-dimethylhexane tree then becomes 3 3 2 2 1 1 1 1. The modification, called M-tuple, is computationally trivial, but the new code allows a simpler interpretation. We can now read each 1 as a *primary* carbon atom or CH₃ (methyl) group; each 2 as a *secondary* carbon atom or CH₂ (methylene) group; each 3 as a *tertiary* carbon

atom or CH (methyne) group; and each 4 as a *quaternary* carbon atom or C. Therefore, the above M-tuple code can immediately be rewritten in "chemical" language as



which can be further abbreviated as



or, if one adopts the Polish notation as 2(CH)2(CH₂)4(CH₃). Remember that the formula is *unique* and that the atoms have *unique* labels!

NOVEL CODES FOR SMALLER ALKANES

To illustrate the novel approach better, in Table 1 we have listed N-tuple and M-tuple codes for smaller alkanes.

In the first column of Table 1, the traditional names of alkanes are given, their N-tuple and M-tuple codes are given in the second and third column, and in the fourth column the alkane formulas are explicitly executed by replacing each numeral in the M-tuples according to the convention from the above: 1 by CH₃, 2 by CH₂, 3 by CH, and 4 by C. The derived formulas have the appearance of familiar chemical structures. This is an important advantage that very few nomenclature systems possess; it is one of the requirements that Read² explicitly stipulated in his list of desirable properties of a chemical code. This requirement is particularly important for chemical nomenclature, which, as Goodson¹⁷ emphasized, differs from codes, being even more restrictive, as nomenclature is used in verbal communication as well. This is not so much the case with the codes, which are intended for computer processing and manual manipulations. The same is true to some extent for the nodal nomenclature of Lozac'h *et al.*¹⁸ as well as for the subsequent graph-based chemical nomenclature of Goodson.^{17,19,20} The last two columns of Table 1 show contracted formulas with customary subscript representation of the repeating units (such as methylene or methyl) and the alternative form based on Polish reverse notation in which functor always precedes its arguments. Functor represents the operation to be

Table 2. The M-tuple Codes and Novel Formulas for All Octane Isomers

octane isomer	M-tuple code	chemical formula
n-octane	2 2 2 2 2 2 1 1	(CH ₂) ₆ (CH ₃) ₂
3-ethylhexane	3 2 2 1 2 1 2 1	CHCH ₂ (CH ₂ CH ₃) ₃
4-methylheptane	3 2 2 1 2 2 1 1	CH(CH ₂ CH ₂ CH ₃) ₂ CH ₃
3-methylheptane	3 2 2 2 1 2 1 1	CH(CH ₂) ₃ CH ₃ CH ₂ (CH ₃) ₂
2-methylheptane	3 2 2 2 2 1 1 1	CH(CH ₂) ₄ (CH ₃) ₃
2,5-dimethylhexane	3 2 2 3 1 1 1 1	CH(CH ₂) ₂ CH(CH ₃) ₄
2,4-dimethylhexane	3 2 3 2 1 1 1 1	(CHCH ₂) ₂ (CH ₃) ₄
3,4-dimethylhexane	3 3 2 1 1 2 1 1	(CH) ₂ CH ₂ (CH ₃) ₂ CH ₂ (CH ₃) ₂
3-ethyl-2-methylpentane	3 3 2 1 2 1 1 1	(CH) ₂ (CH ₂ CH ₃) ₂ (CH ₃) ₂
2,3-dimethylhexane	3 3 2 2 1 1 1 1	(CH) ₂ (CH ₂) ₂ (CH ₃) ₄
2,3,4-trimethylpentane	3 3 3 1 1 1 1 1	(CH) ₃ (CH ₃) ₅
3-ethyl-3-methylpentane	4 2 1 2 1 2 1 1	C(CH ₂ CH ₃) ₃ CH ₃
3,3-dimethylhexane	4 2 2 1 2 1 1 1	CCH ₂ (CH ₂ CH ₃) ₂ (CH ₃) ₂
2,2-dimethylhexane	4 2 2 2 1 1 1 1	C(CH ₂) ₃ (CH ₃) ₄
2,2,4-trimethylpentane	4 2 3 1 1 1 1 1	CCH ₂ CH(CH ₃) ₅
2,3,3-trimethylpentane	4 3 1 1 2 1 1 1	CCH(CH ₃) ₂ CH ₂ (CH ₃) ₃
2,2,3-trimethylpentane	4 3 2 1 1 1 1 1	CCHCH ₂ (CH ₃) ₅
2,2,3,3-tetramethylbutane	4 4 1 1 1 1 1 1	C ₂ (CH ₃) ₆

performed on arguments, thus 2CH₂ in the formula of butane indicates occurrence of two methylene groups in succession.

Observe from Table 1 that only in two cases the resulting formula is identical to standard chemical formula for the compound, that is, in the case of ethane and 2,2-dimethylpropane. In *all* other cases the formulas in Table 1 differ from the standard formulas by placing a single CH₃ group *at the end* rather than *at the beginning* of the code. Nevertheless, it is trivial to rewrite the formula if one insists on doing it. The new system of writing these formulas has the following advantages:

(a) Recollect that the N-tuple code and M-tuple code induce *numbering* of atoms,¹⁰ hence the position of a group (or atom) in the formula implies an atomic label. These codes are to introduce a novel term, subst-isomorphic to chemical formulas of Table 1, i.e., there is a 1:1 substitution correspondence between them.

(b) Novel concatenation of individual symbols allows additional contractions of formulas.

(c) For many purposes, CH₃ groups are chemically of the least interest (merely indicating end-sites in various branches). In a searching process, then, it is better to start with some other atom, not the end-atom, and in the proposed current formula carbon atoms are sequenced according to their functionality: Quaternary atoms precede tertiary atoms, which precede secondary atoms, and the primary atoms are the last, unless the branching pattern dictates otherwise.

Hence one may use CH₂CH₃CH₃ as the linear formula for propane rather than CH₃CH₂CH₃. The advantages, of course, will be more visible when the approach is extended to larger alkanes and other acyclic compounds having multiple bonds and heteroatoms (*vide infra*).

HIGHER ALKANES

In order to illustrate further the novel codes and names we present in Table 2 the M-tuples and novel formulas for all octane isomers.

It is not difficult to derive these formulas. They are not listed because someone may have a hard time to derive one on her/his own; they are listed to illustrate an *additional* property of the M-tuple codes—the possibility of *ordering* structures. We started with numerically the smallest code and produced a sequence of lexicographically increasing M-tuple codes. As one can see, first we have structures with one tertiary carbon, then with two and so on, and when all possibilities having tertiary carbons have been exhausted, the first structure having quaternary carbon emerges. One can see this regularity well-illustrated in Table 4 of Knop *et al.*⁶ where all alkane-trees having $n = 11$ vertices (159 undecane-trees) are shown as printed by computer. The ordering of structures as output was not pointed out in that work. The sequence starts at the second column, then at the bottom of the third column, at the top of the fourth column, etc. One can thus trace the “metamorphosis” of individual graphs

Table 3. Nodal Nomenclature of Some Higher Alkanes and the Corresponding Formulas Based on Compact Codes

compd no.	nodal name	compact code-based formula
1	(6)hexanodane	(CH ₂) ₄ (CH ₃) ₂
2	(5.1 ³)hexanodane	CH(CH ₂ CH ₃) ₂ CH ₃
3	(10.1 ² 1 ⁷ 1 ⁸)tridecanodane	(CH) ₂ (CH ₂) ₄ CH(CH ₃) ₅
4	(9.3 ⁴ 1 ⁵)tridecanodane	(CH) ₂ (CH ₂) ₃ (CH ₃) ₂ (CH ₂ CH ₂ CH ₃) ₂
5	(13.5 ⁷ 2 ⁵ 1 ¹⁴ 1 ¹⁵)docosanodane	(CH) ₃ CH ₂ CH(CH ₂) ₂ (CH ₂ CH ₃) ₂ (CH ₂) ₅ (CH ₃) ₃ (CH ₂) ₂ CH ₃
6	(8.3 ⁴ 2 ⁵ 1 ¹²)tetradecanodane	(CH) ₃ (CH ₂ CH ₂ CH ₃) ₂ CH ₂ (CH ₃) ₃
7	(10.3 ⁵ 3 ⁵ 1 ¹² 1 ¹¹ 1 ¹⁴ 1 ¹⁴)henicosanodane	C ₃ CH ₂ (CH ₃) ₃ (CH ₂) ₂ CH(CH ₃) ₂ (CH ₂) ₃ (CH ₂ CH ₃) ₂ (CH ₃) ₂
8	(13.6 ⁷ 6 ⁷ 2 ¹⁵ 1 ⁹ 1 ¹¹ 1 ¹⁵ 1 ¹⁷ 1 ²¹ 1 ²³)tetratriacontanodane	C(CH ₂ CH) ₂ (CH ₂ CH ₃) ₂ CH ₃ ((CH ₂ CH) ₂ CH ₂ (CH ₃) ₃) ₃
9	(12.5 ⁶ 3 ⁵ 2 ¹³ 1 ²¹)tricosanodane	(CH) ₄ CH ₂ (CH ₂ CH ₂ CH ₃) ₂ (CH ₂) ₅ CH ₃ (CH ₂) ₃ (CH ₃) ₃
10	(12.4 ⁶ 3 ¹³ 2 ⁹ 1 ¹⁴ 1 ¹⁸)tricosanodane	(CH) ₃ (CH ₂) ₃ CHCH ₂ (CH ₂ CH ₃) ₂ (CH ₂) ₄ CH ₃ CH ₂ CH(CH ₃) ₂ CH ₂ (CH ₃) ₂

Table 4. The M-tuple Code and Its “Chemical” Translation for an Unusually Large Alkane Having 58 Carbon Atoms (Figure 2)^a

M-tuple													“chemical” translation
4	3	3	2	1	1	1							C(CH ₃) ₂ CH ₂ (CH ₃) ₃
4	2	4	1	1	1								CCH ₂ C(CH ₃) ₃
2	2	3	2	4	2	3	2	2	4	3	3	2	(CH ₂) ₂ CHCH ₂ CCH ₂ CH(CH ₂) ₂ C(CH) ₂ CH ₂ (CH ₃) ₃
2	4	1	1	1									CH ₂ C(CH ₃) ₃
2	2	2	2	2	1								(CH ₂) ₅ CH ₃
2	2	4	2	1	1	1							(CH ₂) ₂ CCH ₂ (CH ₃) ₃
2	2	1	1										(CH ₂) ₂ (CH ₃) ₂
2	2	4	2	1	1	1							(CH ₂) ₂ CCH ₂ (CH ₃) ₃

^a Each row in the table indicates exhaustion of a particular branching segment.

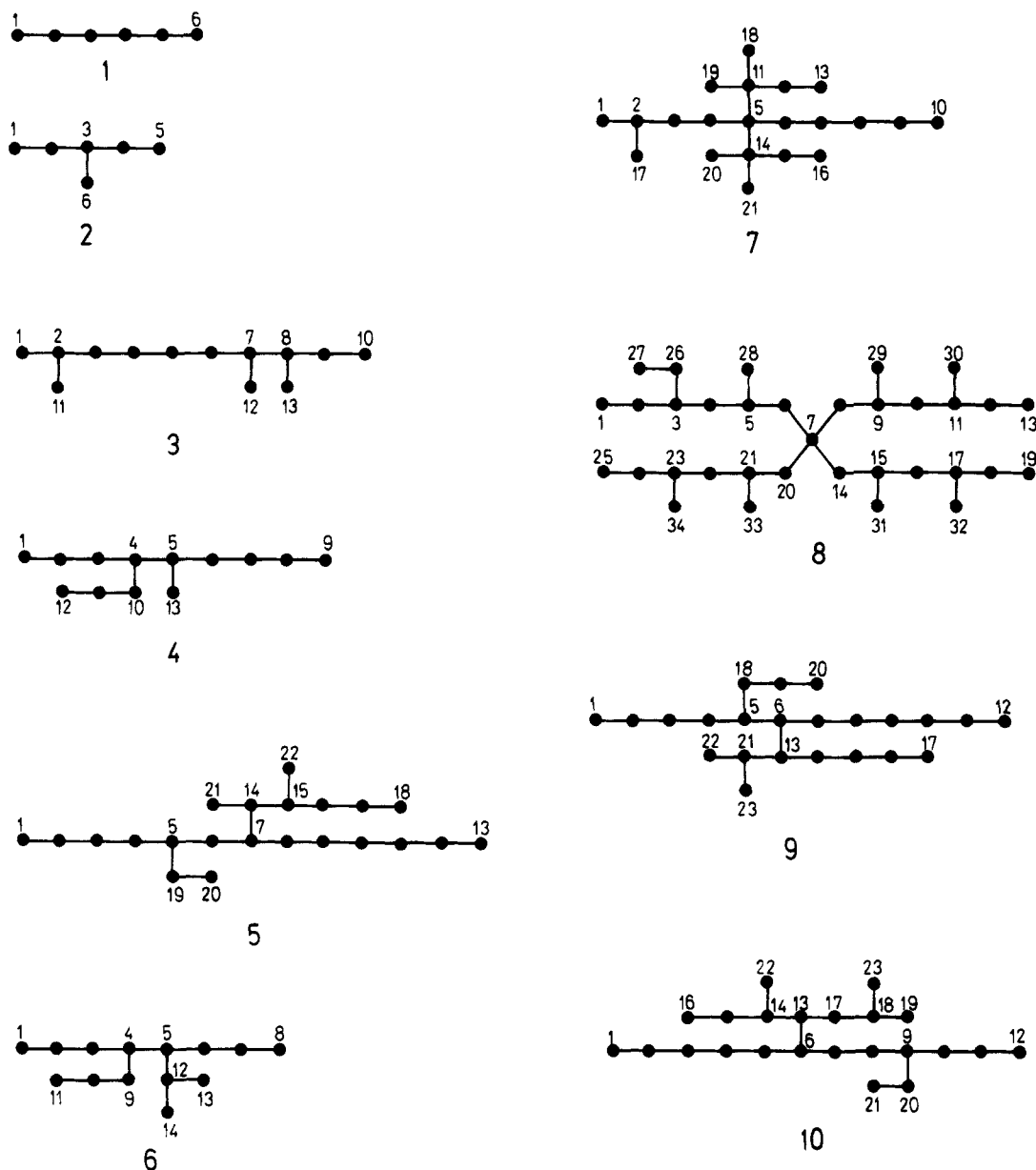


Figure 1. Trees corresponding to some higher alkanes whose nodal nomenclatures and chemical formulas, based on M-tuple codes, are compared in Table 3.

Table 5. The M-tuple Codes and the Corresponding Chemical "Translates" for a Sample of Klein's Hydrocarbons^a

11	$C_3(CH_3)_3C(CH_3)_3CH_2CHC(CH_3)_3C(CH_3)_6$	4	4	4	1	1	1	4	1	1	1	2	3	4	1	1	1	4	1	1	1	1	1	1
12	$CCHC(CH_3)_3(CH_2)_2C(CH_3)_3C(CH_3)_3CH_2C(CH_3)_6$	4	3	4	1	1	1	3	3	4	1	1	1	4	1	1	1	2	4	1	1	1	1	1
13	$C_3(CH_3)_3C(CH_3)_3CHCH_2C(CH_3)_3CH_2C(CH_3)_6$	4	4	4	1	1	1	4	1	1	1	3	2	4	1	1	1	2	4	1	1	1	1	1
14	$CCHCCH_2C(CH_3)_3CH_2C(CH_3)_3CH_2C(CH_3)_3C(CH_3)_6$	4	3	4	2	4	1	1	1	2	4	1	1	1	2	4	1	1	1	4	1	1	1	1

Alternative (Abbreviated) Forms

11	$C_2(C(CH_3)_3)_2CH_2CHC(CH_3)_3)_2(CH_3)_2$
12	$CCH(CH_3)_3(CH_2)_2C(CH_3)_3)_2CH_2C(CH_3)_6$
13	$C_2(C(CH_3)_3)_2CH(CH_2C(CH_3)_3)_2(CH_3)_3$
14	$CCHC(CH_2C(CH_3)_3)_2C(CH_3)_6$

^a Highly branched alkanes are not superimposable on a tetrahedral lattice.

along the columns (i.e., as output), with abrupt changes occurring at places where regularities in code changes are exhausted and a vertex with a higher valency is introduced.

Many coding schemes have been proposed in the past,^{4,8,21} but very few comparisons have been made to show the advantages and disadvantages of individual schemes. Part of the problem is that authors have their collection of structures given as illustrations. Varied rules, which fre-

quently are not as simple as claimed or perceived by their authors, already present a barrier to a less familiar user to generate additional codes to be used in comparisons. We therefore suggest that codes be tested on a collection of representative compounds, and in the case of acyclic structures we think that 18 octanes offer a sufficiently varied class of acyclic compounds exhibiting considerable structural variations, to be used for comparisons.²²

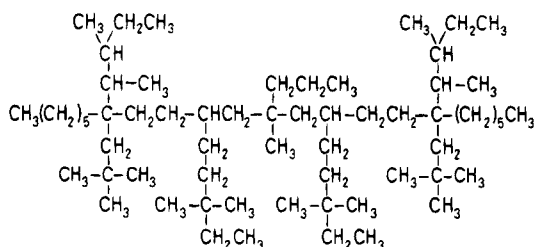


Figure 2. Alkane having 58 carbon atoms (see Table 4 for its complete M-tuple code).

To illustrate further application of novel codes to larger alkanes, in Tables 3–5 we consider selected alkane skeletons, depicted in Figures 1–3.

Structures shown in Figure 1 serve as an illustration of the nodal nomenclature of Lozach *et al.*,¹⁸ which is included in Table 3. Observe that correct nodal names involve numbering of vertices, and the derived labels enter the codes. Our derived structural formulas are *unique*, except for the association of groups. The latter, however, when expanded always produce the same structural formula! The basic difference of the N-tuple approach and the nodal approach (which in this respect parallels standard chemical nomenclature) is in the first step: N-tuple is based on the numerically largest initial string of valencies, the nodal name is based on the longest chain.

In Figure 2 and Table 4 we consider a “monster” alkane, a 58 carbon atom case, introduced by Lin²³ as the illustration of his approach to basic cognizer of molecular shape, fragments, and atomic environments. Clearly the formula for this compound will be lengthy, it has 58 characters for carbon atoms only! In order to facilitate derivation of the formula, we have listed in Table 4 portions of the M-tuple code line by line, in which in each line are the entries leading from a branching vertex to the terminal group. Terminal group here is represented by a vertex having all backtracking back to itself, i.e., it is adjacent to terminal vertices of degree 1. Finally, in Table 5 and Figure 3 we have examined the smallest of Klein’s hydrocarbons.²⁴ These are highly

branched alkanes that cannot be embedded on a tetrahedral lattice, though their every proper subtree can. In addition to their chemical formulas we also give alternative abbreviated forms in which we allow occurrence of nested brackets, but one should observe that although association of a string of symbols need not be unique, the code is unique when expanded. Abbreviated forms are only suggested for convenience, the standard simplification consists only of grouping together successively repeating symbols, as shown in the upper part of Table 5.

ACYCLIC COMPOUNDS WITH HETEROATOMS

Extension of the M-tuple codes to heteroatomic systems is straightforward because in the code each symbol (digit) represents a vertex of the molecular graph. Just as we have replaced 1, 2, 3, and 4 by CH₃, CH₂, CH, and C in case of alkanes, the digits 1, 2, and 3 can be replaced by NH₂, NH, and N in case of nitrogen, and digits 1 and 2 can be replaced by OH and O in case of oxygen. Other heteroatoms can similarly be introduced directly into the code, Si, P, and S having roles analogous to C, N, and O. We can also adopt the customary signs for CC double and triple bonds and, of course, for CO double bonds by placing = and ≡ signs as appropriate between the atoms. Observe that in the M-tuple code the space between successive digits belongs to a *bond* between the atoms (unless atom at the left is terminal, i.e., it indicates the end of a particular branch). Hence the = sign between bonded atoms can be viewed as a functor in the Polish notation, the argument being limited to the succeeding atomic symbol. With these preliminaries one can write down line formulas for most of acyclic chemistry! In Table 6 we give a few illustrations that have already been considered by Read,² in order to show a considerable similarity between the two systems. The heteroatomic systems considered are depicted in Figure 4.

Indeed, one should not necessarily view *alternative* schemes (such as the scheme of Read and the scheme here outlined) as competitive because they may serve different

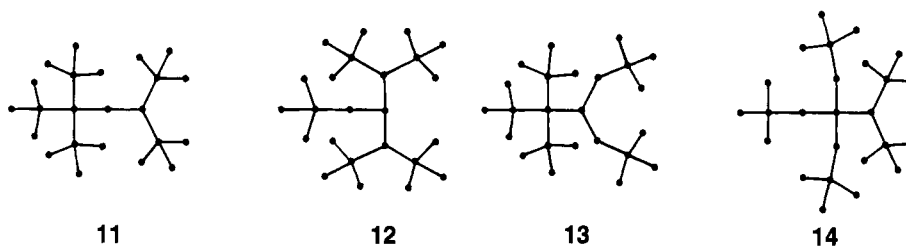


Figure 3. Trees corresponding to Klein’s hydrocarbons. The four smallest trees which cannot be embedded on a tetrahedral lattice.

Table 6. Selection of Branched Heteroatomic Compounds and Comparison of Read’s Designation for Them and Our Chemical Names Based on the M-tuple Code

compd	Read’s code	compact code
15	N(C=OCH ₃) ₃	CN(C=O·CH ₃) ₂ =OCH ₃
16	CH(CH ₃) ₂ (CH ₂ OH)	CHCH ₂ OH(CH ₃) ₂
17	C(CH ₃)(CH=O)(CH ₂ CH ₃)((CH ₂) ₃ CH ₃)	CCH=O(CH ₂) ₃ CH ₃ CH ₂ (CH ₃) ₂
18	C=O(CH ₃) ₂	C=O(CH ₃) ₂
19	CH(OCH ₃)((CH ₂) ₄ OH) ₂	C=O(CH ₂) ₄ CH(CH ₂) ₄ OH·OCH ₃ =O
20	C(CH ₃)(CH=O)((CH ₂) ₃ CH ₃)((CH ₂) ₄ CH ₃)	CCH=O(CH ₂) ₄ CH ₃ (CH ₂) ₃ (CH ₃) ₂
21	CH(OH)(CH ₂ OH)·(CH(OH)) ₂ NHCH(OH)·CH(OH)(CH ₂ OH)	CHC ₂ NH(CH ₂) ₂ CH ₂ ·OHCH ₃ (=O) ₂ CH ₂ (OH) ₂
22	C(NH ₂)(=NH)(S(CH ₂) ₂ CH ₃)	CS(CH ₂) ₂ CH ₃ =NHNH ₂
23	NO(CH ₂ CH ₃) ₃	N(CH ₂ CH ₃) ₂ O
24	P((CH ₂) ₂ CH ₃) ₂ CH(CH(Cl ₂ CHL ₂)P(CH ₂ CH ₃) ₂)	CCHP(CH ₂ CH ₂ CH ₃) ₂ P(CH ₂ CH ₃) ₂ CHL ₄
25	N(CH ₂ CH ₃) ₂ CH=CHN(CH ₂ CH ₃) ₂	NCH=CHN(CH ₂ CH ₃) ₄
21	C(O·CH ₂ ·CH ₃) ₂ (C=O·O·CH ₂ ·CH ₃)(CH ₂ ·C=O·O·CH ₂ ·CH ₃)	C ₂ OCH ₂ CH ₃ =OCH ₂ COCH ₂ CH ₃ =O

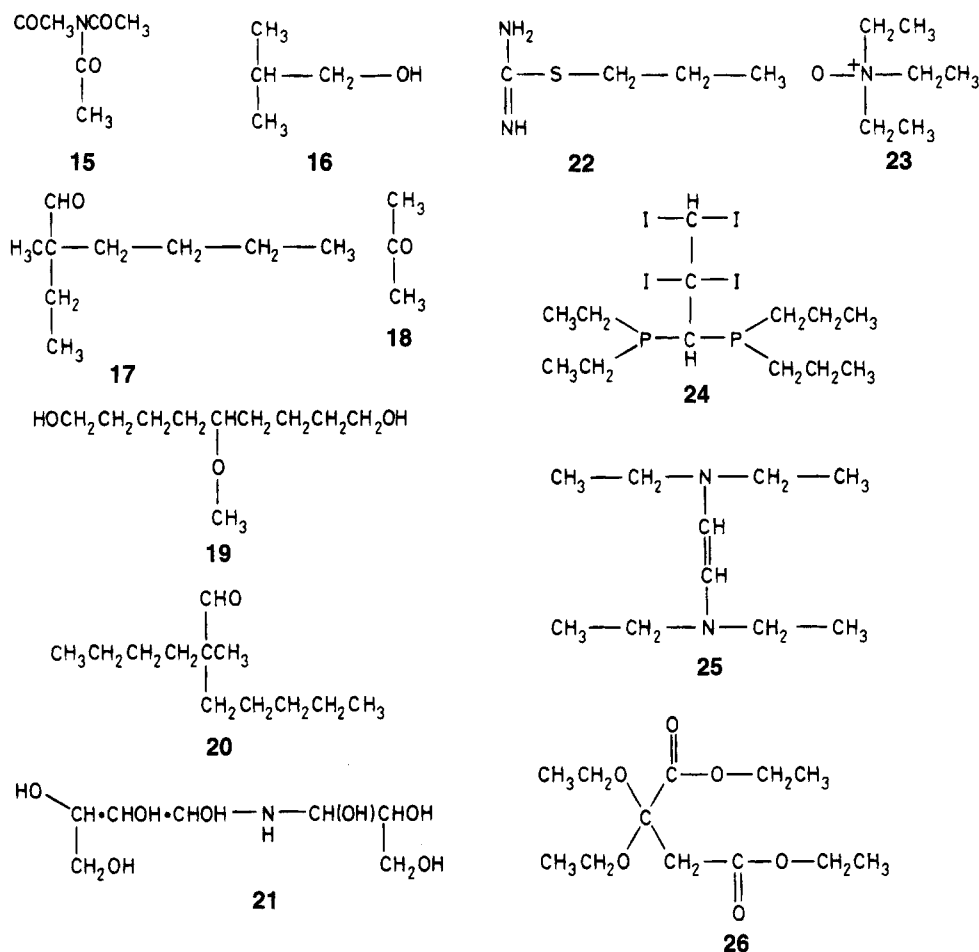


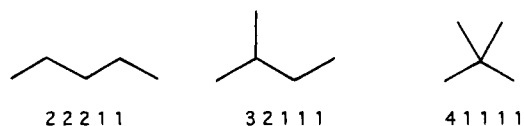
Figure 4. Heteroatomic acyclic compounds considered in Table 6.

purposes or eventually can be combined into a more general approach. The current status of the "problem of chemical coding" as perceived here is that of an *evolving system*. As with evolution of other more complex systems (created by Nature or by humans), each step may contribute to a better organization; smaller changes accumulate and hopefully result in a system that will survive. Current pressure felt by large organizations on having a system that works *now* may eventually become counterproductive in freezing the evolution at a stage that is far from ideal, or even worse, in perpetuating forms that have limited adaptability. From this point of view IUPAC prescribed forms²⁵ as well as WLN code²⁶ appear, in analogy with the evaluation of life, as some prehistoric creatures that have limited capacity to adapt to unforeseen possibilities of the future developments in chemistry. They will eventually become extinct, despite their current artificial support. Purely mathematical schemes will eventually evolve, examples of which are the nodal nomenclature of Lozac'h and collaborators,^{4,18} graph-based chemical nomenclature of Goodson,^{17,19,20} Read's designation of chemical compounds,² the compact codes,^{10,11,13} and other schemes that place mathematical basis above immediate chemical need or tradition.

APPLICATION

The codes should not only be judged by their appearance! They serve various needs, such as use in *construction* of molecular forms of given empirical formulas, for *ordering* and *classification* of compounds in subclasses that may exhibit some structural resemblance and have similar proper-

ties, for *search* and *substructure searches* in documentation, etc. We here will illustrate use of our M-tuple codes for hand construction of all acyclic isomers of $C_2H_5NO_2$, the problem considered by Serov *et al.*²⁷ to illustrate their "mathematical synthesis and analysis of molecular structures", yet another approach to the broad task of construction and computer manipulation of structures that is based on graph-theoretical concepts. In $C_2H_5NO_2$ we have five non-hydrogen atoms, and the only possible acyclic molecular graphs are the following forms (isomers of pentane)



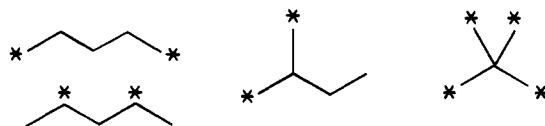
with the corresponding M-tuple codes shown. Next we list possible replacement codes:

- 1 can be replaced by CH_3 , NH_2 , or OH
- 2 can be replaced by CH_2 , NH , or O
- 3 can be replaced by CH or N
- 4 can be replaced by C

In addition, we have to allow occurrence of the functors designating multiple bonds, in particular the functions $=$ and \equiv to represent double and triple bonding. By not considering cyclic structures we need not be concerned here with the role of aromatic CC bond (bond of nonintegral multiplicity). Also here we will not consider ionic species and in particular occurrence of N^+ (quadrivalent positive nitrogen) or COO^-

(negatively charged deprotonated carboxylic acid group). Finally, a word of caution: A number of elements can arise in more than a single valence state. For example, carbon can, besides its normal four-valency, also occur as divalent and even hexavalent, although these are rare and special situations. For nitrogen, besides the normal three-valency, there are number of instances of five-valency. In the present illustration we will exclude these additional possibilities, that, if desired, can be subsequently monitored.

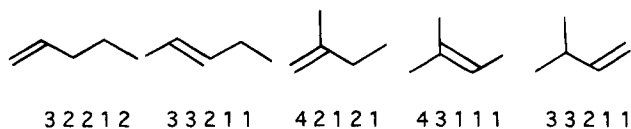
With the above preliminaries we initiate the construction. Because the initial tree graphs have indistinguishable vertices and introduction of heteroatoms corresponds to labeling of a graph with selected set of colors (satisfying the valency restrictions given previously), we may in a systematic production of all substitution codes consistent with the constraints derive duplicates. There are alternative ways to prevent occurrence of duplicates or their subsequent elimination. Here, it suffices to observe equivalence of vertices and perform the substitution only when a vertex belonging to an equivalence class appears. Because the underlying graphs are so simple, one can detect the equivalence by inspection, but in more complex situations there are schemes for detecting equivalence and symmetry properties of molecular graphs.²⁸⁻³⁷ For the three trees we have marked equivalent vertices by (*):



Let us start with the code 2 2 2 1 1 and possible substitution groups: CH₂, NH, and O for 2 and CH₃, NH₂, and OH for 1, remembering that we have to use two carbons atoms, one nitrogen, and two oxygens, combined with five suppressed hydrogens. We can write down various hypothetical substituted systems and then eliminate those that do not satisfy the restriction on C, N, O, and H stoichiometry. In order to be systematic and not overlook useful combinations, one can follow a strictly alphabetic ordering of substituents. The beginning of such list is shown below with the comment on feasibility

structure	feasibility	reason
CH ₂ CH ₂ CH ₂ CH ₃ CH ₃	no	too many C atoms
CH ₂ CH ₂ CH ₂ CH ₃ NH ₂	no	too many C atoms
CH ₂ CH ₂ CH ₂ NH ₂ NH ₂	no	too many C atoms
CH ₂ CH ₂ NHOHOH	no	too many H atoms

Continuing this process one can see that no saturated compounds are possible, so we have to examine systematically multigraphs, shown below based on the above trees:



In this problem we need to examine only trees having one double bond. This can be seen from the last nonfeasible structure shown above: CH₂CH₂NHOHOH, which if considered in a hypothetical hydrogen abstraction reaction will produce CH=CHNHOHOH which *satisfies* our valency restrictions and is one of the many possible acyclic structures

Table 7. Acyclic Structures of the Formula C₂H₅NO₂^a

M-tuple	formula code	conventional formula
3 3 2 1 1	CHCHNHOHOH CHCHONH ₂ OH CHCHOOHNH ₂ CHNCH ₂ OHOH CHNOCH ₃ OH CHNOOHCH ₃ NCHCH ₂ OHOH NCHOCH ₃ OH NCHOOHCH ₃	HO-CH=CH-NH-OH HO-CH=CH-O-NH ₂ NH ₂ -CH=CH-O-OH HO-CH=N-CH ₂ -OH HO-CH=N-O-CH ₃ CH ₃ -CH=N-O-OH HO-N=CH-CH ₂ -OH HO-N=CH-O-CH ₃ CH ₃ -N=CH-O-OH
4 2 1 2 1	CCH ₂ NH ₂ OOH CCH ₂ OHNHOH CCH ₂ OHOHNH ₂ CNHCH ₃ OOH CNHOHCH ₂ OH CNHOHOCH ₃ COCH ₃ NHOH COCH ₃ ONH ₂ CONH ₂ CH ₂ OH CONH ₂ OCH ₃ COOHCH ₂ NH ₂ COOHNHCH ₃	O=C(OH)-CH ₂ -NH ₂ HN=C(OH)-CH ₂ -OH O=C(NH ₂)-CH ₂ -OH O=C(OH)-NH-CH ₃ CH ₂ =C(OH)-NH-OH O=C(CH ₃)-NH-OH NH=C(OH)-O-CH ₃ O=C(NH ₂)-O-CH ₃ CH ₂ =C(OH)-O-NH ₂ O=C(CH ₃)-O-NH ₂ CH ₂ =C(NH ₂)-O-OH NH=C(CH ₃)-O-OH
4 3 1 1 1	CCHOHNH ₂ OH CCHNH ₂ OHOH CNCH ₃ OHOH CNOHCH ₃ OH	NH ₂ -C(OH)=CH-OH OH-C(OH)=CH-NH ₂ CN-C(OH)=N-CH ₃ CH ₃ -C(OH)=N-OH
3 3 2 1 1	CHCHNHOHOH CHCHONH ₂ OH CHNCH ₂ OHOH CHNOCH ₃ OH NCHCH ₂ OHOH NCHOCH ₃ OH	HO-CH(OH)-CH=NH NH ₂ -CH(OH)-CH=O HO-CH(OH)-N=CH ₂ CH ₃ -CH(OH)-N=O HO-N(OH)-CH=CH ₂ CH ₃ -N(OH)-CH=O

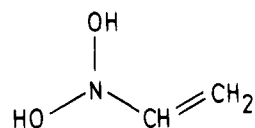
^a Continuation of the list given in the text.

having the formula C₂H₅NO₂. All other structures can then be thought of as being derived from the corresponding saturated structure by abstraction of two hydrogens, thus inducing one double bond between the unsaturated valencies on adjacent atoms. The presence of a double bond changes the N-tuple and M-tuple codes. The linear M-tuple code 2 2 2 1 1 becomes now 3 2 2 1 2, leading to the following substitutions:

structure	usual notation
CHCH ₂ NHOHO	O=CH-CH ₂ -NH-OH
CHCH ₂ ONH ₂ O	O=CH-CH ₂ -O-NH ₂
CHCH ₂ OOHNH	NH=CH-CH ₂ -O-OH
CHNHCH ₂ OHO	O=CH-NH-CH ₂ -OH
CHNHOCH ₃ O	O=CH-NH-O-CH ₃
CHNHOOHCH ₂	CH ₂ =CH-NH-O-OH
CHOCH ₂ NH ₂ O	O=CH-O-CH ₂ -NH ₂
CHOCH ₂ OHNH	NH=CH-O-CH ₂ -OH
CHONHCH ₃ O	O=CH-O-NH-CH ₃
CHONHOHCH ₂	CH ₂ =CH-O-NH-OH
CHOOCH ₃ NH	NH=CH-O-O-CH ₃
CHOONH ₂ CH ₂	CH ₂ =CH-O-O-NH ₂
NCH ₂ CH ₂ OHO	O=N-CH ₂ -CH ₂ -OH
NCH ₂ OCH ₃ O	O=N-CH ₂ -O-CH ₃
NCH ₂ OOHCH ₂	CH ₂ =N-CH ₂ -O-OH
NOCH ₂ CH ₃ O	O=N-O-CH ₂ -CH ₃
NOCH ₂ OHCH ₂	CH ₂ =N-O-CH ₂ -OH
NOOCH ₃ CH ₂	CH ₂ =N-O-O-CH ₃

This exhausts the search for structures that start with a double bond and are linear chains; in all, 18 possibilities. Completion of the construction which follows with linear chains with the double bond in the center, and branched acyclic forms, is shown in Table 7.

The results agree with those of Serov and collaborators (1976), except for the order in which structures occur and a single missing structure in their report:



The total number of possible acyclic isomers of formula $C_2H_5NO_2$ is 49, as stated by Serov *et al.*,²⁷ but their Table 3 contains one incorrect structure, that is, $CH_3CH_2NO_2$, which either signifies a pentavalent nitrogen (which we excluded from forming combinations) or a cyclic structure. In addition, their Table 3, which shows 50 structures, instead of 49, has a duplicate (fourth structure in the first column and the second structure in the second column are identical). These discrepancies may have been introduced in preparation of their table and need not suggest that the authors developed a faulty algorithm.

DISCUSSION

We have only briefly illustrated the potential of the N-tuple codes of Knop and collaborators,⁶ which have been here "chemically" modified so that they appear in more familiar form to chemists. The N-tuple and M-tuple codes are very compact (i.e., brief), and, being in one-to-one correspondence with atoms, allow immediate interpretation as chemical formulas. The advantages of this may be hinted: Use of this nomenclature allows immediate translation into computer codes by simple statements on substitution of valencies. Hence, the codes deserve another attribute, besides being compact, they can be viewed as *direct* or, if one wishes, *instant* codes. Because their use by computer is instantaneous, no preprocessing is required. Moreover, the codes are so simple to derive that the work can be delegated to administrative personnel! This should be contrasted with other codes, most of which require familiarity with sometimes lengthy rules involving ranking of atoms, linking subgroups, insertion of various typographical symbols, ordering of symbols, iterative steps, clustering, special vocabulary, qualification of branches, etc. There is some organization of the N-tuple codes, but the dominating principle of producing maximal lexical code possible for a structure suffices to resolve all the ambiguities. The rule, thus, is simple: the price to be paid is in searching for the canonical name (code). We have seen similar situations with the molecular codes based on the smallest binary number, when rows of the adjacency matrix are read from left to right and from top to bottom,³⁸ and other canonical forms,^{39,40} but in practice it appears that fewer rules more than compensate for work involved in the search for the codes. Furthermore, a simple rule of *structural* origin reflects some structural characteristics which are lost in systems of nomenclature based on a list of requirements (determined by a committee!), because these tend to be *ad hoc* and introduced at demand, i.e., *a posteriori*, even if sometimes they can anticipate future structures. But novel structures emerge unexpectedly, as demonstrated so many times in the past and in the present: Nobody except, of course, Doering and Roth⁴¹ anticipated bullvalene with its unique (room temperature) structure. The case points to a special class of flexible systems for which we still have to develop suitable codes, codes that will reflect

their unique characteristics. Even less dramatic structures, like helicenones⁴² with their dynamic capabilities ought to be encoded so that during computer (or manual) processing of codes these important structural features can become apparent. Here should also be mentioned buckminsterfullerene, a C_{60} carbon compound which has the geometry of truncated icosahedron⁴³ and appears to be difficult structure to code.^{44–46} Little progress has yet been made in that direction although in the recent paper Read⁴⁷ outlines advantages for a nonexplicit description of an aromatic CC bond, which can be deduced from unsaturated valencies of atoms involved. A similar approach may be used for fluxional structures. Finally, one has to resolve the problem of coding stereochemistry, and, although we have not touched on this, several stereochemical coding schemes have been described in the literature^{48,49} that may be incorporable into the present codes.

In conclusion, the purpose of the present paper is to demonstrate that the compact codes recently introduced¹⁰ have the potential to develop into a full new system of chemical coding. By this we do not necessarily suggest that the present scheme has to be adopted without further modifications (something that committees are good at). We merely wish to indicate that the "process of deriving chemical coding" has entered another phase in its evolution, the phase that can be termed "mathematical", and that, as we and others have demonstrated, leads to, so to speak, "higher forms" of structural codes that ought to succeed existing schemes, which, although at present of great use and satisfying the needs moderately well, are from an evolutionary point of view not viable and will sooner or later be recognized as archaic.

Finally, we should add that the N-tuple codes of Knop *et al.*⁶ and M-tuple codes may be generalized by relaxing the constraints in their rules. For instance, one can consider an alternative selection of a *root*. Once a root is selected, the procedure uniquely determines the code.^{6,9} In another study we consider as an alternative to the present rule that produces the maximum code to select as a root the "central" vertex (or vertices, there being in the case of a tree at most two central vertices).⁵⁰ This work will be reported elsewhere.

ACKNOWLEDGMENT

S.N. and N.T. were supported by the Ministry of Science and Technology of the Republic of Croatia by Grant No. 1-07-159. We are thankful to Professor A. T. Balaban and Dr. Alan L. Goodson for their constructive comments on the early version of the manuscript.

REFERENCES AND NOTES

- (1) Luckenbach, R. The Beilstein Handbook of Organic Chemistry: The First Hundred Years. *J. Chem. Inf. Comput. Sci.* **1981**, 21, 82–83.
- (2) Read, R. C. A New System for the Designation of Chemical Compounds. 1. Theoretical Preliminaries and the Coding of Acyclic Compounds. *J. Chem. Inf. Comput. Sci.* **1983**, 23, 135–149. See, also: Karaburnaliev, S.; Ivanov, J.; Mekenyan, O. Coding of Chemical Structures based on a Line Notation. *Comput. Chem.* **1994**, 18, 189–193.
- (3) Wolman, Y. *Chemical Information*, 2nd ed.; Wiley: Chichester, 1988.
- (4) Polton, D. J. *Chemical Nomenclatures and the Computer*; Research Studies Press: Chichester, 1993.
- (5) Balaban, A. T. Topological Indices and Their Uses: A New Approach for the Coding of Alkanes. *J. Mol. Struct. (Theochem)* **1988**, 165, 243–253. Balaban, T.-S.; Filip, P. A.; Ivanciuc, O. Computer Generation of Acyclic Graphs Based on Local Vertex Invariants and

- Topological Indices. Derived Canonical Labelling and Coding of Trees and Alkanes. *J. Math. Chem.* **1992**, *11*, 79–105.
- (6) Knop, J. V.; Müller, W. R.; Jeričević, Ž.; Trinajstić, N. Computer Enumeration and Generation of Trees and Rooted Trees. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 91–99.
- (7) Knop, J. V.; Müller, W. R.; Szymanski, K.; Nikolić, S.; Trinajstić, N. In *Computational Chemical Graph Theory*; Rouvray, D. H., Ed.; Nova Science Publ.: Commack, NY, 1990; pp 9–32.
- (8) Trinajstić, N.; Nikolić, S.; Knop, J. V.; Müller, W. R.; Szymanski, K. *Computational Chemical Graph Theory: Characterization, Enumeration and Generation of Chemical Structures by Computer Methods*; Horwood/Simon & Schuster: Chichester, 1991.
- (9) Müller, W. R.; Szymanski, K.; Knop, J. V.; Trinajstić, N. Bit-tuple Notation for Trees. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 960–961.
- (10) Randić, M. Compact Molecular Codes. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 136–148.
- (11) Randić, M. Compact Codes. 2. Bicyclic Saturated Hydrocarbons. *Croat. Chem. Acta* **1986**, *59*, 327–343.
- (12) Randić, M.; Nikolić, S.; Trinajstić, N. Compact Molecular Codes for Polycyclic Systems. *J. Mol. Struct. (Theochem)* **1988**, *165*, 213–228.
- (13) Nikolić, S.; Trinajstić, N. Compact Molecular Codes for Annulenes, Aza-annulenes, Annulenoannulenes, Aza-annulenoannulenes, Cyclazines and Aza-cyclazines. *Croat. Chem. Acta* **1990**, *63*, 155–170.
- (14) Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1983; Vols. I and II.
- (15) Trinajstić, N. *Chemical Graph Theory*; 2nd revised ed.; CRC Press: Boca Raton, FL, 1992.
- (16) Hiz, H. A Linearization of Chemical Graphs. *J. Chem. Docum.* **1964**, *4*, 173–180.
- (17) Goodson, A. L. Graph-Based Chemical Nomenclature. I. Historical Background and Discussion. *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 167–172.
- (18) Lozac'h, N.; Goodson, A. L.; Powell, W. H. Die Nodal Nomenklatur—Allgemeine Prinzipien. *Angew. Chem.* **1979**, *91*, 951–964. See, also: Lozac'h, N.; Goodson, A. L. Nodal Nomenclature II—Specific Nomenclature for Parent Hydrides, Free Radicals, Ions and Substituents. *Angew. Chem., Int. Ed. Engl.* **1984**, *23*, 33–46. Lozac'h, N. Principles for the Continuing Development of Organic Nomenclature. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 180–185.
- (19) Goodson, A. L. Graph-Based Chemical Nomenclature. II. Incorporation of Graph-Theoretical Principles into Taylor's Nomenclature Proposal. *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 172–180.
- (20) Goodson, A. L. Application of Graph-Based Chemical Nomenclature to Theoretical and Preparative Chemistry. *Croat. Chem. Acta* **1983**, *56*, 315–324.
- (21) Kirby, E. C. Coding and Enumeration of Trees That Can Be Laid Upon a Hexagon Lattice. *J. Math. Chem.* **1991**, *11*, 187–197.
- (22) Randić, M.; Trinajstić, N. Viewpoint 4—Comparative Structure—Property Studies: The Connectivity Basis, *J. Mol. Struct. (Theochem)* **1993**, *284*, 209–221.
- (23) Lin, C. Chemical Inference Based on SEFLIN. I. Basic Cognizance of Molecular Shape, Fragments and Atomic Environment of Organic Compounds. *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 47–51.
- (24) Klein, D. J. Rigorous Results for Branched Polymer Models with Excluded Volume. *J. Chem. Phys.* **1981**, *75*, 5186–5189.
- (25) E.g.: *IUPAC Nomenclature of Organic Chemistry*; Pergamon Press: Oxford, 1979.
- (26) Wiswesser, W. J. *A Line-Formula Notation*; Crowell: New York, 1954. See, also: Agarwal, K. K.; Gelernter, H. L. A Computer-Oriented Linear Canonical Notational System for the Representation of Organic Chemistry with Stereochemistry. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 463–479.
- (27) Serov, V. V.; Elyashberg, M. E.; Gribov, L. A. Mathematical Synthesis and Analysis of Molecular Structures. *J. Mol. Struct.* **1976**, *31*, 381–397.
- (28) Randić, M. On Discerning Symmetry Properties of Graphs. *Chem. Phys. Lett.* **1976**, *42*, 283–287.
- (29) Shelly, C. A.; Munk, M. E. Computer Perception of Topological Symmetry. *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 110–113.
- (30) Jochum, C.; Gasteiger, J. Canonical Numbering and Constitutional Symmetry. *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 113–117.
- (31) Lloyd, E. K. Graphs Corresponding to Intramolecular Rearrangements. *Math. Chem. (Mülheim/Ruhr)* **1979**, *7*, 255–271.
- (32) Balasubramanian, K. Symmetry Groups of Chemical Graphs. *Int. J. Quantum Chem.* **1982**, *21*, 411–418. See, also: Balasubramanian, K. Computational Technique for the Automorphism of Groups of Graphs. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 621–626.
- (33) Randić, M.; Katović, V.; Trinajstić, N. In *Symmetries and Properties of Non-Rigid Molecules: A Comprehensive Survey*; Maruani, J., Serre, J., Eds.; Elsevier: Amsterdam, 1983; pp 399–408.
- (34) Davis, M. I.; Ellzey, M. L., Jr. Technique for Determining the Symmetry Properties of Molecular Graphs. *J. Comput. Chem.* **1983**, *4*, 267–275.
- (35) Brocas, J.; Gielen, M.; Willem, R. *The Permutational Approach to Dynamic Stereochemistry*; McGraw-Hill: New York, 1983.
- (36) Flurry, R. L., Jr. Groups of Simple Graphs. *Int. J. Quantum Chem.* **1984**, *25*, 309–320.
- (37) Randić, M.; Oakland, D. O.; Klein, D. J. Symmetry Properties of Chemical Graphs. IX. The Valence Tautomerism of P_7^{3-} Skeleton. *J. Comput. Chem.* **1986**, *7*, 35–54.
- (38) Randić, M. On Recognition of Identical Graphs Representing Molecular Topology. *J. Chem. Phys.* **1974**, *60*, 3920–3928. See, also: Randić, M. On Canonical Numbering of Atoms in a Molecule and Graph Isomorphism. *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 171–180.
- (39) Hendrickson, J. B.; Toczko, A. G. Unique Numbering and Cataloguing of Molecular Structures. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 171–177. See, also: Balaban, A. T.; Mekenyan, O.; Bonchev, D. Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities (HOC Procedures). I. Algorithms for Finding graph Orbits and Canonical Numbering of Atoms. *J. Comput. Chem.* **1985**, *6*, 538–551.
- (40) Herndon, W. C.; Leonard, J. E. Canonical Numbering, Stereochemical Descriptors and Unique Linear Notation for Polyhedral Clusters. *Inorg. Chem.* **1983**, *22*, 554–557. See, also: Herndon, W. C. In *Chemical Applications of Topology and Graph Theory*; King, R. B., Ed.; Elsevier: Amsterdam, 1983; pp 231–242.
- (41) Doering, W. von E.; Roth, W. R. A Rapidly Reversible Degenerate Cope Rearrangement. *Tetrahedron* **1963**, *19*, 715–737.
- (42) Newman, M. S.; Lednicer, D. The Synthesis and Resolution of Hexahelicene. *J. Am. Chem. Soc.* **1956**, *78*, 4765–4770.
- (43) Kroto, H. W.; Heath, J. R.; O'Brien, S. C.; Curl, R. F.; Smalley, R. E. C_{60} : Buckminsterfullerene. *Nature* **1985**, *318*, 162–163.
- (44) Eckroth, D. The Correct von Baeyer Name for (Buckminster)fullerane. *J. Chem. Educ.* **1993**, *70*, 609–611.
- (45) Rücker, C.; Rücker, G. Nomenclature of Organic Polycycles Out of the Computer—How to Escape the Jungle of the Secondary Bridges. *Chimia* **1990**, *44*, 116–120.
- (46) Balaban, A. T.; Babić, D.; Klein, D. J. W. R. Hamilton, His Circuits and IUPAC Nomenclature for Fullerenes. *J. Chem. Educ.*, in press.
- (47) Read, R. C. A New System for the Designation of Chemical Compounds. 2. Coding of Cyclic Compounds. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 116–128.
- (48) Cahn, R. S.; Ingold, C. K.; Prelog, V. Specification of Molecular Chirality. *Angew. Chem., Int. Ed. Engl.* **1966**, *5*, 385–415.
- (49) Wipke, W. T.; Dyott, T. M. Simulation and Evaluation of Chemical Synthesis. Computer Representation and Manipulation of Stereochemistry. *J. Am. Chem. Soc.* **1974**, *96*, 4825–4834. Choplin, F.; Marc, R.; Kaufmann, G.; Wipke, W. T. Computer Design of Synthesis of Phosphorus Chemistry: Automatic Treatment of Stereochemistry. *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 110–118. Razinger, M.; Perdih, M. Computerized Stereochemistry: Coding and Naming Configurational Stereoisomers. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 290–296. Bonchev, D.; Balaban, A. T.; Liu, X.; Klein, D. J. Molecular Cyclicity and Centricity of Polycyclic Graphs. I. Cyclicity Based on Resistance Distances and Reciprocal Distances. *Int. J. Quantum Chem.* **1994**, *50*, 1–20.
- (50) Harary, F. *Graph Theory*, 2nd printing; Addison-Wesley: Reading, MA, 1971; p 35.

CI940098Z