L66 B6 A B- C 1B ITJ B- CT56 BNJ B2OR CG DG& GG D- BT5OJ EG

Mol formula	Mol weight
C30H27CL4N1O2	575.362

Apart from the possibilities in substructure searching and molecular weight calculation based on WLN, the result of our analysis is also useful as the chemical base for the evaluation of structure/property and structure/activity relationships. At the moment we are calculating lipophilic constants based on the work of Nys and Rekker,[17] with the aid of DEAN's. The absence or presence of each individual DEAN of the second level representation could possibly be used as a bit screen.

Our further research will be directed toward comparison of the tables of the analysis of starting material and end products, and using the difference as a mean for automatic encoding or determination of the reactions.

ACKNOWLEDGMENT

The authors wish to express their gratitude to the management of Gist-Brocades Research for the use of their IBM-1130, and to Mr. G. J. Keijser and Mr. A. van der Woude for their valuable discussions concerning this work.

LITERATURE CITED

(1) Osinga, M., and Verrijn Stuart, A. A., "Documentation of Chemical Reactions. I. A Faceted Classification," *J. Chem. Doc.*, **13**, 36–39 (1973).
(2) Clinging, R., and Lynch, M. F., "Production of Printed Indexes of Chemical Reactions. I. Analysis of Functional Group Interconversions," *J. Chem. Doc.*, **13**, 98–102 (1973).
(3) Hyde, E., Matthews, F. W., Thomson, L. H., and Wiswesser, W. J., "Conversion of Wiswesser Notation to a Connectivity Matrix for Organic Compounds," *J. Chem. Doc.*, **7**, 200–204 (1967).
(4) Hyde, E., and Thomson, L. H., "Structure Display," *J. Chem. Doc.*, **8**, 138–146 (1968).
(5) Thomson, L. H., Hyde, E., and Matthews, F. W., "Organic Search and Display Using a Connectivity Matrix Derived from Wiswesser Notation," *J. Chem. Doc.*, **7**, 204–209 (1967).
(6) Campey, L. H., Hyde, E., and Jackson, A. R. H., "Interconversion of Chemical Structure Systems," *Chem. Brit.*, **6**, 427–430 (1970).
(7) Granito, C. E., Roberts, S., and Gibson, G. W., "The Conversion of Wiswesser Line Notations to Ring Codes. I. The Conversion of Ring Systems," *J. Chem. Doc.*, **12**, 190–199 (1972).
(8) Granito, C. E., "CHEMTRAN and the Interconversion of Chemical Substructure Systems," *J. Chem. Doc.*, **13**, 72–74 (1973).
(9) Osinga, M., "Automatic Encoding of Chemical Reactions," Paper presented at the NATO/CNA Advanced Study Institute on Computer Representation and Manipulation of Chemical Information, Noordwijkerhout, June 4–15, 1973.
(10) Adamson, G. W., Cowell, J., Lynch, M. F., McLure, A. H. W., Town, W. G., and Yapp, A. M., "Strategic Considerations in the Design of a Screening System for Substructure Searches of Chemical Structure Files," *J. Chem. Doc.*, **13**, 133 (1973).
(11) Adamson, G. W., Creasey, S. E., and Lynch, M. F., "Analysis of Structural Characteristics of Chemical Compounds in the Common Data Base," *J. Chem. Doc.*, **13**, 158 (1973).
(12) Crowe, J. E., Lynch, M. F., and Town, W. G., "Analysis of Structural Characteristics of Chemical Compounds in a Large Computer-Based File. Part I. Noncyclic Fragments," *J. Chem. Soc. C*, 990 (1970).
(13) Adamson, G. W., Lynch, M. F., and Town, W. G., "Analysis of Structural Characteristics of Chemical Compounds in a Large Computer-Based File. Part II. Atom-Centered Fragments," *J. Chem. Soc. C*, 3702 (1971).
(14) Adamson, G. W., Lambourne, D. L., and Lynch, M. F., "Analysis of Structural Characteristics of Chemical Compounds in a Large Computer-Based File. Part III. Statistical Association of Fragment Incidence," *J. Chem. Soc., Perkin Trans. 1*, 2428–2433 (1972).
(15) Adamson, G. W., Cowell, J., Lynch, M. F., McLure, A. H. W., Town, W. G., and Yapp, A. M., "Analysis of Structural Characteristics of Chemical Compounds in a Large Computer-Based File. Part IV. Cyclic Fragments," *J. Chem. Soc., Perkin Trans. 1*, 863 (1973).
(16) Adamson, G. W., Creasey, S. E., Eakins, J. P., and Lynch, M. F., "Analysis of Structural Characteristics of Chemical Compounds in a Large Computer-Based File. Part V. More Detailed Cyclic Fragments," *J. Chem. Soc., Perkin Trans. 1*, in press.
(17) Nys, G. G., and Rekker, R. F., "Statistical Analysis of a Series of Partition Coefficients with Special Reference to the Predictability of Folding of Drug Molecules. The Introduction of Hydrophobic Fragment Constants (*f* values)," *Chim. Therap.*, **5**, 521–35 (1973).

A Numerical Identifier for the Chemical Elements, Expressing Their Periodic Relationships†

KENNETH W. LOACH

Department of Chemistry, State University College of Arts and Science, Plattsburgh, New York 12901

Received September 10, 1974

Each chemical element can be assigned two numbers related to its periodic table position. The separate or packed numbers are effective identifiers of the element and its periodic relationships. A Fortran subroutine LEMENT has been written that rapidly converts an element symbol into the equivalent numerical identifiers.

Processing of chemical information usually requires the storage and retrieval of chemical element identities. The two common identifiers are the traditional element symbol and the atomic number. The symbol is familiar and the number is easily manipulated, but neither lend themselves readily to the systematic or automatic recognition of periodic relationships. Searches of either identifier for classes of elements (*e.g.*, any transition element, any group Va element) requires considerable additional processing and (usually) multiple tests of each element identifier.

There is a need for an element identifier that is compact, easily manipulated, and which allows the ready recognition of periodic relationships. This can be achieved by assigning

† Presented at the 168th National Meeting of the American Chemical Society, Division of Chemical Literature, Atlantic City, N.J., Sept. 11, 1974.

Table I. NROW and NGRP Values of the Representative Elements

NROW \ NGRP	02	03	04	05	06	07	08	01	(00)
18	H							He	
17	Li	Be	B	C	N	O	F	Ne	
16	Na	Mg	Al	Si	P	S	Cl	Ar	
15	K	Ca	Ga	Ge	As	Se	Br	Kr	
14	Rb	Sr	In	Sn	Sb	Te	I	Xe	
13	Cs	Ba	Tl	Pb	Bi	Po	At	Rn	
12	Fr	Ra							
(11)									

Table II. NROW and NGRP Values of the Transition Elements

NROW \ NGRP	04	05	06	07	08	09	10	11	02	03	(00)
05	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	
04	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	
03	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	
02	Ac										
(01)											

Table III. NROW and NGRP Values of the Inner-Transition Elements

NROW \ NGRP	−14	−13	−12	−11	−10	−09	−08	−07	−06	−05	−04	−03	−02	−01
−07	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
−08	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lw
(−09)														

to each element two signed two-digit numbers, NROW and NGRP, related to the element's periodic table coordinates. The proposed numbers are shown in Tables I–III.

Each element can be uniquely designated by the pair of numbers NROW and NGRP stored separately or by the signed four-digit packed number NATM, where

$$NATM = NROW * 100 + NGRP$$

The separate numbers each require 2 decimal or 8 binary digits minimum storage space, and the packed number requires 4 decimal or 16 binary digits, so neither has any advantage in storage requirements. The packed number would be the best choice, if most searching were for specific elements, because it would allow faster recognition of single elements and yet still allow generic searches. The separate numbers would be more efficient if generic searches were frequent.

The proposed numerical identifier is suitable for computer-processing but could also be used in other information systems such as printed indices, card files, and other traditional systems. The NROW and NGRP numbers can be generated from the element symbols manually or by computer (see Experimental Section).

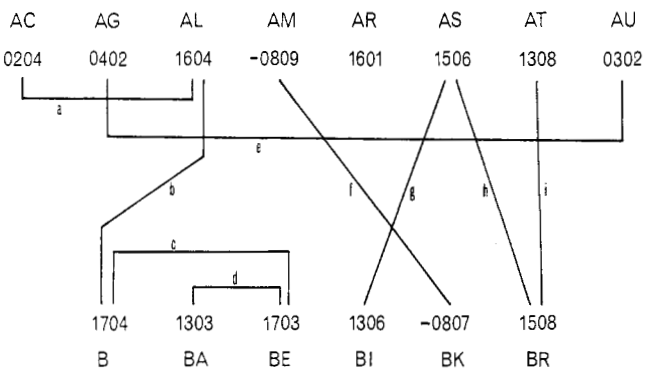

Figure 1. Typical periodic relationships of a set of elements from the NROW and NGRP values. (a) group III, NGRP = 04; (b) group IIIa, NGRP = 04 and NROW > 10; (c) consecutive pair in same periodic row; (d) group IIa, NGRP = 03 and NROW > 10; (e) group Ib, NGRP = 02 and 10 > NROW > 0; (f) actinides, NROW = −08; (g) group Va, NGRP = 06 and NROW > 10; (h) 4th row representative elements, NROW = 15; (i) halogens, NGRP = 08 and NROW > 10.

TYPICAL RELATIONSHIPS

In the proposed scheme, all representative elements have NROW >10, all transition elements have NROW > 00 and all inner transition elements have NROW < 00. Representative, transition, and inner transition elements in the same period have their NROW values separated by intervals of 10 (*e.g.*, Cs, Hf, Ce are all in the same horizontal period; they have (NROW- 10), NROW, and (NROW + 10), respectively, equal to 03).

Corresponding A and B subgroups have the same NGRP values for the representative and transition elements. These elements can readily be distinguished from the inner-transition elements by the difference in sign of the respective NGRP values. Group VIII elements can readily be distinguished as a set (since NGRP > 8) or classified into three subgroups by their distinct NGRP values.

The NROW and NGRP values of the representative elements increase in parallel with the general electronegativity trends. This may be a useful basis for some searching or classification strategies. The other elements have parallel NROW and NGRP trends, although this may be of less value for these elements.

An example of the ready recognition of periodic relationships is shown in Figure 1, where the first 14 elements in symbol alphabetical sequence are listed, together with their NATM values. The lines connecting the elements indicate some of the simple periodic relationships that can be readily recognized by inspection of the NROW and NGRP patterns.

GENERALIZED ELEMENT-CLASS NUMBERS

NROW, NGRP, and NATM values can be assigned to classes of elements. These class identifiers could be useful in specifying Markush structures, generalized chemical fragments, and the like. For this purpose the NROW values 11, 01, and −09 are reserved for the general representative,

transition, and inner-transition elements, respectively, and the NGRP value 00 for any element group. Thus classes of elements can readily be given numerical identifiers, *e.g.*

NATM	
1100	representative element
0100	transition element
−0900	inner-transition element
1700	second row representative element (Li-Ne)
0500	first row transition element (Sc-Zn)
−0700	lanthanide (Ce-Lu)
−0800	actinide (Th-Lw)
1108	halogen (F-At)
0102	group Ib element (Cu-Au)

and so on for other similar classes of elements.

EXPERIMENTAL

Computer programs were run in Fortran IV on a Burroughs 3500 computer (XFORTS complier). A subroutine LEMENT[1] has been written that will supply the NROW and NGRP values for any one- or two-letter element symbol argument (H to Lw). The subroutine matches letters in descending order of probability of occurrence so as to match the argument in the minimum time.

A frequency of occurrence table for chemical element symbols was generated by randomly sampling an index of molecular formulas[2] (index page numbers were generated with a random number generator GRBG). When LEMENT was applied to the sample element list (53 elements from 617 compounds), a frequency-weighted average of only 3.73 character-tests was required for identification of any element symbol.

ACKNOWLEDGMENTS

I wish to thank Lawrence Scacciaferro for practical assistance in this work. The work was supported by a Faculty Research Fellowship and Grant-in-Aid from the Research Foundation of the State University of New York.

LITERATURE CITED

(1) A listing of LEMENT subroutine is available from the author.
(2) Chemical Abstracts 8th Collective Formula Index, 1967–1971.

The Connectivity Stack, a New Format for Representation of Organic Chemical Structures

YOSHIHIRO KUDO

JEOL, Toyko, Japan

SHIN-ICHI SASAKI*

Miyagi University of Education, Sendai, Japan

Received September 16, 1974

A new type of format to represent the enumerated structural formula consistent with given information is discussed.

A system for the automated structure elucidation of organic compounds,[1] CHEMICS,[2] has been developed by the authors. One of the functions involved in the system is the enumeration[3] of all possible chemical structures (the informational homolog) which are consistent with given structural information. Although several studies[4,5] concerning the enumerations of the structures have already been published, the method developed by the authors is more widely applicable and more completely guaranteed compared to others. In this paper a new type of format to represent the enumerated structural formula is discussed.

From a practical point of view the format to be adopted in CHEMICS has to contain "dynamic" as well as "static" features. That is, it has to make execution time as short as possible and at the same time ensure complete and unique enumeration. Enumeration methods, such as in the heuristic DENDRAL,[4] and the one by Balaban,[5] are elegant but fix the limits of their scopes for application in terms of types of structures. On the other hand, description methods, such as IUPAC/Dyson notation,[6] CA/Morgan notation,[7] and Wiswesser line notation,[8] treat all types of compounds, but have great regard for unambiguous and unique representation of individual structures.

The connectivity stack** is a sequence of elements of the connectivity matrix. When the elements of the matrix are a_{ij}'s, the stack $a_1, a_2, a_3, \ldots, a_k, \ldots$, uniquely corresponds to one of the matrices after establishment of a correspondence rule. Here, the function of correspondence rules will be explained in, for practical reasons, two cases:

Case 1

$$k = N(i - 1) - i(i + 1)/2 + j \qquad (i < j)$$

(N stands for the order of the matrix)

Case 2

$$k = i + (j - 2)(j - 1)/2 \qquad (i < j)$$

Usually sets of equivalent stacks are possible to represent one structure. For example, in the case of propane, three, 101, 011, and 110, corresponding to three numberings, 1–2–3, 1–3–2, and 2–1–3, respectively, which are numbered to the carbon atoms, exist. In the case of propane, the same

* Author to whom correspondence should be addressed.

** In general, the word "stack" means a set of data of which the elements are sequentially arranged (in contrast with, *e.g.*, a tree arrangement), and where conversion always takes place at the end of the set (this in contrast with, *e.g.*, queue).