

Replacement-Nodal-Subtractive Nomenclature and Codes of Chemical Compounds<sup>†</sup>

OTTO R. GOTTLIEB\* and M. AUXILIADORA C. KAPLAN

Instituto de Química, Universidade de São Paulo, 05508 São Paulo, SP, Brazil

Received January 7, 1985

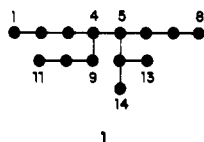
Any chemical compound can be designated by replacement-nodal-subtractive (RNS) nomenclature and codes. In application to organic compounds, the system indicates all information on heteroatoms by the replacement method, on the relative position of all atoms except the hydrogens by nodal numbering, and on unsaturation by the subtractive method. No unfamiliar concepts and symbols are used. Thus, the resulting name or its codified version can be directly transcribed into the corresponding structural formula without consultation of glossaries and is compatible with computer storage. Programs can be conceived for the retrieval of carbon skeletons and structural subunits including heterocycles and functional groups.

## INTRODUCTION

The organic chemist is usually quite satisfied with the IUPAC system of organic nomenclature, in spite of the recognized existence of three problems: (1) Application requires memorizing or consulting an enormous, ever increasing number of terms and rules.<sup>1</sup> (2) The resulting names require complex programs for computerized storage and retrieval. (3) The names are based on the English language and translation into other idioms requires many equivocal decisions. Any new proposal will consequently elicit interest only if designed to eliminate these objections.

In this sense, the general principles of nodal numbering<sup>2</sup> are quite satisfactory. Any structure, described in terms of a graph, has its nodes numbered in a unique and sequential manner. For this purpose, graphs are classified into acyclic, monocyclic, polycyclic, and assemblies of several cyclic or cyclic plus acyclic types. Selected fundamental rules for nodal numbering of each type are transcribed in the following four paragraphs. The original paper<sup>2</sup> must be consulted for additional rules concerning chiefly questions of seniority.

**Acyclic Graphs.** The main chain is the longest unbranched chain of nodes. It is numbered first, from one end to the other, the lowest possible locants being assigned to the longest branches. Branches are numbered successively in order of decreasing length. For example, numbering of 1 leads to the

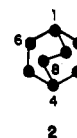


descriptor [8.3<sup>4</sup>2<sup>5</sup>1<sup>12</sup>], which indicates an unbranched chain of eight nodes with branches of three, two, and one node occurring respectively at positions 4, 5, and 12.

**Monocyclic Graphs.** Numbering proceeds sequentially from any node in either direction until all nodes are numbered.

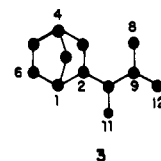
**Polycyclic Graphs.** Numbering begins at one of the nodes of the main ring (i.e., the monocyclic ring containing the greatest number of nodes) to which the main bridge (i.e., the longest unbranched chain of nodes both ends of which are linked to the main ring) is attached. It proceeds from this bridgehead in the direction that gives the lower locant to the other bridgehead. The main bridge is numbered sequentially after the main ring, beginning with the node of the main bridge

connected to the node in the main ring having the locant 1. The secondary bridges are numbered successively in the same manner. For example, numbering of 2 leads to the descriptor



[06.2<sup>1,4</sup>], which indicates a ring [symbolized by 0 (zero)] of six nodes sustaining a bridge of two nodes between the bridgeheads numbered 1 and 4.

**Assemblies of Cyclic and Acyclic Graphs.** A structure graph is called an assembly if it is composed of more than one cyclic module or at least one acyclic module and at least one cyclic module. The descriptor of each module retains the numbering that would be used to describe it if it were an isolated graph. The principal module is determined by the following criteria, which are applied successively until a decision is reached: (a) largest number of nodes; (b) cyclic module preferred to acyclic module; (c) largest number of rings or branches (side chains); (d) largest main ring or longest chain; (e) longest bridge or branch (side chain); (f) lowest locants for attachment of bridges or branches (side chains). The definitive numbering of an assembly begins with the principal module and proceeds successively through each chain of modules attached to the principal module. The original numbering is retained for the principal module. The other modules are renumbered sequentially by adding to each of their original locants a number equal to the total number of nodes that have already been assigned definitive locants. For example, numbering of 3 leads



to the descriptor [(06.1<sup>1,4</sup>)2:10(4.1<sup>2</sup>)], which indicates the existence of a cyclic module (06.1<sup>1,4</sup>) and an acyclic module (4.1<sup>2</sup>). The principal (cyclic) module encompasses seven nodes; for definitive numbering of the acyclic module, seven must be added to all its original locants. Thus, the connection of the modules involves nodes 2:10.

Addition of the chemical information to the descriptor, however, as proposed in "Nodal Nomenclature", part II,<sup>3</sup> is less satisfactory. Insaturation/saturation is indicated by

<sup>†</sup> Presented Oct 20, 1984, at the Symposium on Chemical Nomenclature, XVI Latinamerican Congress of Chemistry, Rio de Janeiro, Brazil.

subtractive and additive methods, the latter complementary to the use of such trivial terms as arene and axene. In a further combination of systems, at intermediate positions of a structure heteroatoms or groups of heteroatoms are defined as replacement and hence considered to be part of the skeleton, while at terminal positions they are defined as "substituents". Their representation by substitutive prefixes and suffixes necessitates definition of many designations and knowledge of many empirical rules. Besides, the differentiation of structural graph and substituent(s) destroys the homogeneity of the nodal treatment. Neither structure nor functionality is uniformly the main classification criterion for the substance. Fortunately, however, the substitutive method is not essential to nodal nomenclature.<sup>3</sup> Indeed, only by abandoning it altogether can the potentiality of nodal numbering, as the basis of a logical, definitive, and "closed" system of chemical nomenclature, be fully exploited.

### PROCEDURE

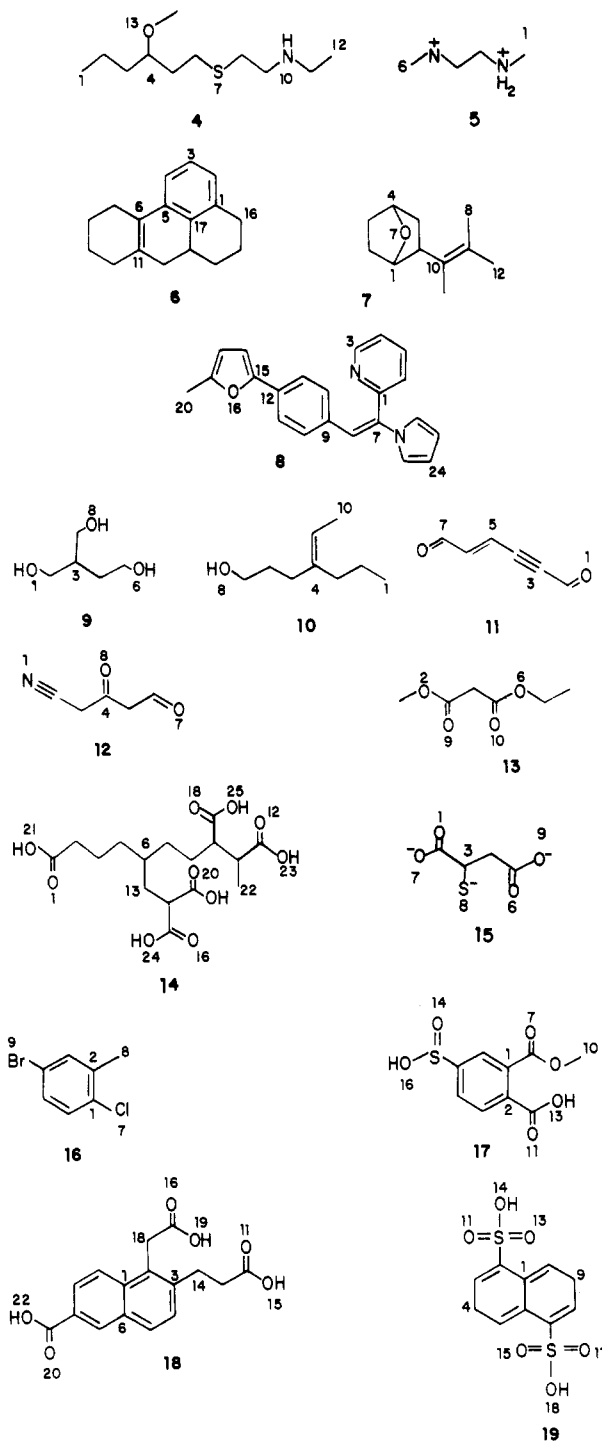
In the proposed procedure, nodal status is assigned to all carbon and heteroatoms of a molecule. Hydrogens are not represented on the nodal graph, unless located on atoms with free valences or ionic charges. The additional load for the nodal descriptor, which may thus occur, is of little concern. Much more important than the simplicity of the name itself is the ease of its deduction and of its interpretation. Nodal numbering is performed as originally described,<sup>2</sup> with the additional provision that, existing a choice, first attention is paid to multiple bonds. As far as possible, they should be included in the main nodal sequence. Double bonds thus frequently fall between nodes possessing contiguous numbers, a fact that simplifies their notation.

Chemical information is added to the descriptor solely by replacement and subtractive principles.<sup>1</sup> In organic nomenclature the prefix carba is omitted, and only heteroatoms, equally designated by "a" terms,<sup>1</sup> such as bora, aza, oxa, fluora, and alumina, are quoted as prefixes in order of increasing atomic numbers. Unsaturation, in relation to the fully hydrogen saturated parent, designated by ene and yne, as well as free radicals and ions, is quoted as suffixes. Double bonds in aromatic systems, free radicals with delocalized free valences, or ions with delocalized charges are named by selecting the localized form characterized by the smallest locants. Stereochemical information is added, listing the localization of *R* or *S* chirality in the prefix and of *E* or *Z* geometry, jointly with the respective locants for double bonds, in the suffix. As a general rule, numbering is performed by assigning the lowest possible locants to the descriptor, the prefix, and the suffix, in this order of decreasing preference.

### APPLICATION

On application to substances that lack "substituents", this procedure is substantially coincident with the previous proposal,<sup>3</sup> and the original paper can be consulted for examples such as 10-aza-13-oxa-7-thia[12.2<sup>4</sup>]tetradecane (4), 2,5-diaza[6]hexan-5-ium-2-ylum (5), tetracyclo[016.1<sup>1</sup>50<sup>6</sup>.110<sup>13</sup>.17]-heptadecane-1,3,5(17),6(11)-tetraene (6), 7-oxabicyclo[(06.1<sup>1</sup>.4<sup>2</sup>2:10(4.1<sup>2</sup>)]dodecan-9-ene (7), and 2,21-diaza-16-oxatetracyclo[(06)1:7(2)8:9(06)12:15(05)17:20(1)7:21(05)]-pentacosan-1,3,5,7,9,11,13,15(19),17,22,24-undecaene (8). The original versions of the names assigned to substances 4 and 8 were slightly modified to list the heteroatoms in order of increasing atomic numbers and to avoid the introduction of additive nomenclature<sup>1</sup> as an additional method for the location of saturated and unsaturated centers.

On application to substances with "substituents",<sup>3</sup> modified names result as exemplified: 1,6,8-trioxa[6.2<sup>3</sup>]octane against [4.1<sup>2</sup>]pentane-1,4,5-triol<sup>3</sup> (9); 8-oxa[8.2<sup>4</sup>]decan-4(9)*E*-ene

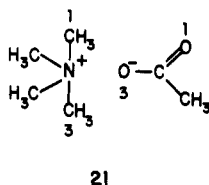
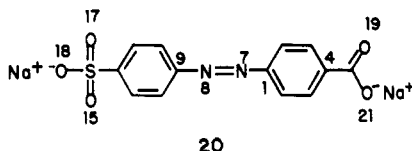


against [7.2<sup>4</sup>]nonan-4(8)-en-1-ol<sup>3</sup> (10); 1,8-dioxa[8]octane-1,5*E*,7-trien-3-yne against 1,6-dioxo[6]hexan-2-en-4-yne<sup>3</sup> (11); 1-aza-7,8-dioxa[7.1<sup>4</sup>]octane-4(8),6-dien-1-yne against 1-nitrilo-3,5-dioxo[5]pentane<sup>3</sup> (12); 2,6,9,10-tetraoxa[8.1<sup>3</sup>1<sup>5</sup>]-decane-3(9),5(10)-diene against 3,5-dioxo-2,6-dioxa[8]octane<sup>3</sup> (13); 1,12,16,18,20,21,23-26-decaoxa[12.4<sup>6</sup>2<sup>9</sup>1<sup>4</sup>1<sup>2</sup>1<sup>10</sup>1<sup>11</sup>.1<sup>15</sup>1<sup>17</sup>1<sup>19</sup>]hexacosane-1,11,15,17,19-pentaene against [10.3<sup>5</sup>1<sup>8</sup>1<sup>9</sup>1<sup>12</sup>]hexadecane-1,10,13,14,16-pentoic acid<sup>3</sup> (14); 1,6,7,9-tetraoxa-8-thia[6.1<sup>2</sup>1<sup>3</sup>1<sup>5</sup>]nonane-1,5-diene-7,8,9-triide against 2-sulfido[4]tetrane-1,4-dioate<sup>3</sup> (15); 7-chloro-9-bromoacyclo[(06)1:7(1)2:8(1)4:9(1)]nonane-1,3,5-triene against 5-bromo-2-chlorocyclo[(06)1:7(1)heptan(1-6)arene<sup>3</sup> (16); 7,9,11,13,14,16-hexaoxa-15-thiacyclo[(06)1:8(4)2:12(3)5:15-(3)]hexadecane-1,3,5,7,11,14-hexaene against 7-oxo-5-sulfinio-8-oxacyclo[(06)1:7(3)2:10(1)]decan(1-6)aren-10-oic acid<sup>3</sup> (17); 11,15,16,19,20,22-hexaoxabicyclo[(010.0<sup>1</sup>.6<sup>3</sup>)3:14-(4.1<sup>2</sup>)2:18(3.1<sup>2</sup>)8:21(3)]docosan-1,3,5,7,9,11,16,20-octaene

against bicyclo[010.0<sup>1,6</sup>]3:11(3)2:14(2)8:16(1)]hexadecan-(1-10)arene-13,15,16-trioic acid<sup>3</sup> (**18**); 11,13,14,15,17,18-hexaoxa-12,16-dithiabicyclo[(010.0<sup>1,6</sup>)2:12(3.1<sup>2</sup>)7:16(3.1<sup>2</sup>)]-octadecan-2,5,7,10,11,12,15,16-octaene against 3,8-dihydrobicyclo[010.0<sup>1,6</sup>]decarene-5,10-disulfonic acid<sup>3</sup> (**19**).

### LINE FORMULAS

All these names contain at least one redundancy. Clearly, 2,5-diaza and 7,8-dioxa state twice that two nitrogens and two oxygens are present in **5** and **12**, respectively; the descriptors [6.2<sup>3</sup>] and [(010.0<sup>1,6</sup>)3:14(4.1<sup>2</sup>)2:18(3.1<sup>2</sup>)8:21(3)] already indicate **9** and **18** to be an acyclic octane and a bicyclic docosane, respectively. For this reason it is proposed that, while in spoken language the complete names can be retained, they may be written in abridged expressions with the usual *R* and *S* symbols for indication of chirality, as well as with chemical symbols for the "a" terms, as prefixes. The *Z* and *E* symbols for indication of double bond geometry, as well as the bidented symbol |<sup>≡</sup> for the double bond and the tridented |<sup>≡</sup> for the triple bond, are listed as suffixes. Information on radicals (symbol e), negative charges (symbol n) and positive charges (symbol p) is also given in suffixes. The resulting line formulas are illustrated by the following examples: 7,8-H-2,5-N-[6.1<sup>2</sup>1<sup>2</sup>]2,5-p (**5**); 1,6,8-O[6.2<sup>3</sup>] (**9**); 1-N-7,8-O[7.1<sup>4</sup>]4(8),6-|<sup>≡</sup>-1-|<sup>≡</sup> (**12**); 11,13,14,15,17,18-O-12,16-S[(010.0<sup>1,6</sup>)2:12-(3.1<sup>2</sup>)7:16(3.1<sup>2</sup>)]2,5,7,10,11,12,15,16-|<sup>≡</sup> (**19**); 7,8-N-15,17,18,19,21-O-16-S[(06)1:7(2)8:9(06)12:16(3.1<sup>2</sup>)4:20-(3)]1,3,5,7,9,11,13,15,16,19-|<sup>≡</sup>-18,21-n;Na-p;Na-p (**20**); 2-N-



[3.1<sup>2</sup>1<sup>2</sup>]2-p;1,3-O[3.1<sup>2</sup>]1-|<sup>≡</sup>-3-n (**21**). The designation (1R,11S,16R,18S,22S,23R)-8,19-N-12,25-O[(0.21.2<sup>1,11</sup>-1<sup>15,19</sup>0<sup>1,18</sup>0<sup>2,7</sup>0<sup>8,22</sup>0<sup>16,23</sup>)9:25(1)]2,4,6,9(25),14-|<sup>≡</sup> is deduced for the highly complex heptacyclic strychnine (in Portuguese estriquinina) with astounding ease. Chemists from any nation would be able to perform the deduction of strychnine's structural formula from this expression with ease.

### FUTURE DEVELOPMENTS

Computer storage of such line formulas seems perfectly possible. In computer retrieval it will suffice to search for the presence of both locants of each |<sup>≡</sup> pair among the prefixes in order to identify diatomic and triatomic groups with double bonds, such as N=N, S=O, C=N, C=O, C=S, C=C,

C=C=O, and C=C=C. Analogously each |<sup>≡</sup> pair will reveal the presence of C≡N and C≡C groups. A search for adjacent numbers of each locant for an element with undefined ligands in the list of such groups leads to the identification of triatomic groups, such as Cl-C=O, N-C=O, O-C=O, S-C=O, O-C=N, C=N-O, and tetraatomic groups, such as O-C=N-O, C=C-C=O, and C=C-C=C, as well as more extended conjugated and aromatic systems. A search for unused locants among the prefixes will lead to the identification of monoatomic groups such as O in alcohols and ethers, N in amines, and S in thiols and sulfides. Since the descriptor indicates the lowest locant and allows calculation of the highest locant of each cycle, a search for the presence of the extreme and the intermediate locants of such a unit(s) will identify heterocycles. Thus, although the descriptor codifies the gross structure of the molecule, analytical separation of the carbon skeleton from its functionality is nevertheless possible.

### CONCLUSIONS

Read<sup>4</sup> summarized the desirable attributes of a chemical coding system as follows: The symbols used in the codes should be familiar. The codes should be linear strings of symbols, easily comprehensible to chemists, brief, and pronounceable. The coding process should not depend at all on chemical intuition or on knowledge of the properties of chemical compounds, should not need to have recourse to any list of names or other nonsystematic items, and, hence, should be "closed", i.e., applicable to *any* compound, including previously unknown, unusual, and complicated substances. The coding and decoding processes should yield unique results, be simple, and, preferably, be performable by hand.

RNS nomenclature possesses all these attributes, even the one concerning pronounceability. Only the codified versions of the names, intended for computer storage, are of course not suitable for oral communication. Brevity, considered to be incompatible with most of the other attributes,<sup>4</sup> does not fare too badly, as evidenced by the comparison of 1*H*,6*H*-imidazo[1,5-*a*][4,1]benzoxazepine (latest *Chemical Abstracts* index name) and T B576 BN DN HO CH 1H J (Wiswesser line notation)<sup>5</sup> with 3,5-N-13-O[014.0<sup>1,5</sup>0<sup>6,11</sup>]2,6,8,10,14-|<sup>≡</sup> (RNS designation). Indeed, it will be difficult to devise a system of nomenclature that is simpler, more appealing to the chemist, and more suitable for international communication.

### REFERENCES AND NOTES

- (1) IUPAC "Nomenclature of Organic Chemistry"; Pergamon Press: Oxford, 1979; Sections A-F and H.
- (2) Lozac'h, N.; Goodson, A. L.; Powell, W. H. "Nodal Nomenclature—General Principles". *Angew. Chem., Int. Ed. Engl.* **1979**, *18*, 887–899.
- (3) 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.
- (4) 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.
- (5) Chemical Abstracts Service "Ring Systems Handbook"; American Chemical Society: Washington, DC, 1984; 1984 ed., Ring Systems File I (RF1-RF 32530).