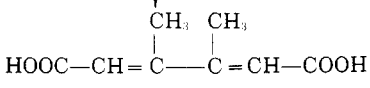
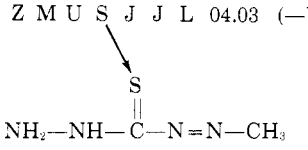
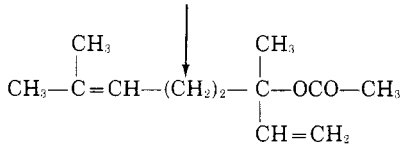
2) 1 2 3 4 5 6 7 8 9 10
Q V D U L U L D V Q 05.04 07.06 (D=U—{—U=D—)



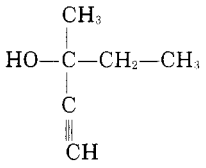
3) 1 2 3 4 5 6 7
Z M U S J J L 04.03 (—U—J=J—)



4) L U L D L 2 X L D D O V L (—U=D—{—X—D=D)
03.02 08.07 10.07



5) Q X L C C L 2
03.02 05.02



LITERATURE CITED

(1) Hyde, E., F. W. Matthews, Lucille H. Thomson, and W. J. Wiswesser, J. CHEM. Doc. 7, 200 (1967).
(2) Thomson, Lucille H., E. Hyde, and F. W. Matthews, J. CHEM. Doc. 7, 204 (1967).

# 107 Years of Line-Formula Notations (1861–1968)*

WILLIAM J. WISWESSER
Fort Detrick, Frederick, Maryland 21701

Received May 21, 1968

**Within seven years after the birth of structural chemistry in 1861, virtually all of the main ideas relating to line-formula conventions were devised and published in the leading chemical journals of a century ago. No basically new practices appeared for some 79 years. Then, within an identically brief period of just seven years (1947–1954), virtually all of the fundamental features of structure-delineating chemical notations appeared in the international chemical literature. The key characteristics of the old conventions and new systems are surveyed.**

"The recent international interest in chemical notation has made it seem profitable and desirable to examine the historical records for a guiding background." This was the opening remark for a report on "The History of Chemical Notation" presented 18 years ago at the 118th Meeting of the American Chemical Society in Chicago.[1] That report in turn quoted an opening remark on "Chemical Notation and Nomenclature" by Samuel William Johnson that has an amusing echo today because his remark now is nearly 100 years old:[2]

> "Beginners in Chemistry are liable to much confusion and embarrassment from the fact that there are now in use two distinct systems of Chemical Notation and several forms of nomenclature."

Johnson was referring to notation developments that were then just 10 years old. More recent details on the "Origin of the Line-Formula Method" were given in 1954,[3] and these historic developments again were reported to the American Chemical Society in 1962, this time with a century-old perspective.[4]

The key idea of structural chemistry was popularized in 1860, when the leading chemists of the world attended the first International Chemical Congress at Karlsruhe[5] to resolve their confusions about atoms, molecules, and equivalents. At the close of this 4-day session, Stanislao Cannizzaro clarified the concept of molecules with his reprints on "the message of his old teacher, Avogadro,

* Presented before the Division of Chemical Literature Symposium on Notation Systems, 155th Meeting, ACS, San Francisco, Calif., April 1968.

who was then dead and forgotten."[6] Alexander Butlerov thereupon predicted that the future task of the chemist was to determine the atomic arrangements in these molecules. He was the first to use the word "structure" in its modern chemical meaning.[7]

## THE LINE-FORMULA CONVENTION

Ever since this first recognition of structural importance, chemists have identified the molecular configurations in text copy simply by delineating the symbols for the corresponding molecular segments, one after another as connected. These are the "line-formula" descriptions.

The exact convention relating to "linear expression of formulae" was clearly defined by A. D. Mitchell 20 years ago:[8] "Each full point (or period mark) is regarded as separating two atoms which are directly linked in the main chain of a compound, and atoms or groups attached to each of these atoms are written immediately after it and before the next full point."

Josef Loschmidt led this new era of structural chemistry with his publication of *Chemische Studien* (Vienna, 1861), which contained 368 remarkably astute graphical diagrams, including scores of benzene ring postulations[9] (four years before Kekule "introduced" this ring to the chemical world). His text included a few "rationelle Formel" such as $C_2H_5,O,C_2H_5$ for ether, $C_2H_5,O,CH_3$ for methyl ethyl ether, and $C_2H_5,O,C_2H_4,O,C_2H_5$ for the diethyl ether of ethylene glycol. In 1863, L. Carius showed a few other comma-separated, line-formula descriptions like CHHH,O,H and $C_2H_5,O,H$,[10] and Emil Erlenmeyer omitted the punctuation with familiar forms like $C_3H_7OH$, $C_3H_7I$ and $C_3H_7OC_3H_7$.[11] Meanwhile H. Hübner had used the conventional period punctuation in the first line-formula descriptions to appear in a chemical journal,[12] with his discussion of the $CH_2CN.CO.Br$ and $CH_2Br.CO.CN$ isomers.

In 1866, August Kekulé,[13] Henry Debus,[14] H. L. Buff,[15] Erlenmeyer,[16] and E. Frankland and B. F. Duppa[17] completed the popularization of this line-formula technique with scores of examples like $CH_3.CO.OH$, $CH_3.CO.CH_3$, $NC—CH_2—CO_2H$, $C_6H_5—CH_2—Br$, and $C_6H_5.SO_2.OH$. Debus described ethylene simply as $CH_2.CH_2$ and acetylene as CH.CH; the unsaturation mark that Erlenmeyer introduced at this same time was picked up two years later by C. Graebe[18] in his $C_6H_5—N=N—C_6H_5$ description of azobenzene, and by A. W. Hofmann[19] in his explanation of $—N=C=S$ and $—S—C\equiv N$ isomers. Finally, when H. Wichelhaus introduced the word "Valenz" (as a simplification of equivalence or quantivalence) in 1867, he suggested the punctuating refinement of periods to set off "side-group" terminations and hyphens to show directly connected groups, as in his $CO.OH-CH_2-CO.OH$ and $CH_3-CH.OH-CO.OH$.[20] Thus, within the short period of just seven years after the birth of structural chemistry in 1861, virtually all of the main ideas relating to line-formula descriptions were conceived and published. No basically new practices appeared for some 79 years. Then, within an identically brief period of just seven years (1947–1954), virtually all of the fundamental features of structure-delineating chemical notations appeared in the international chemical literature.

Examples of these early variations of "rational formulas" were given with explanatory comments in two other recent reports on chemical structure information processing; hence, they need not be repeated here.[21, 22]

G. Malcolm Dyson inspired the modern development of chemical notations with his 1947 manual on "A New Notation and Enumeration System for Organic Compounds."[23] An intriguing feature about these first Dyson ciphers was that they were restricted to capital letters, on-line numerals, and 10 punctuation marks. A single letter represented the benzene ring, and five other letters represented similar aromatic skeletons that were used as building blocks for higher elaborations of polycyclic systems. In addition to his manuals, Dyson presented a number of papers on his notation.[24-33]

F. L. Taylor in 1947 developed a mathematically elegant method for enumerating topologically similar positions in all ring systems—carbocyclic or heterocyclic, aromatic or saturated.[34] Dyson's greatly enlarged second edition incorporated this Taylor system of ring enumeration after A. M. Patterson pointed out the "Possibilities for a Combined System of Notation and Nomenclature for Organic Compounds."[35]

The G-K-D ciphers of Gordon, Kendall, and Davison appeared in 1947 and 1948 with their publications on "A New Systemization of Chemical Species"[36] and "Chemical Ciphering: A Universal Code as an Aid to Chemical Systematics."[37] At that time, all three were employed in the research department of the Dunlop tire firm in Birmingham, England. They evidently designed their system for early versions of computers, for Davison shortly thereafter reported on "Programs and Equipment for Sorting Gordon-Kendall-Davison Punched Cards for Any Structurally Defined Group."[38] Again in 1957, Davison and Gordon reported on "Sorting of Chemical Groups Using Gordon-Kendall-Davison Ciphers."[39] Davison and Dyson's critical views of each other's approaches were aired in the Letters column of *Chemistry & Industry* in 1954.[40]

The Gruber Notation was first reported in 1949,[41] and for a time attracted the interest of examiners looking for a suitable international chemical notation. Dr. Gruber pleaded with the IUPAC representatives, before and after their Stockholm meeting in 1951, not to be hasty in their selection of an internationally recommended notation. He provided some stimulating suggestions for classification prefixes and prime marks for another notation,[42] but did not pursue further development of his own. In February of this year, he responded to our news about this Notation Symposium with his best wishes, and informed us that he will be 82 years old in June!

In 1949, the IUPAC Commission on Codification, Ciphering, and Punched Card Techniques announced its interest in seeking an internationally suitable chemical notation and invited designers to submit their proposals for review in 1951. It enumerated 11 desirable characteristics or "desiderata for an internationally acceptable chemical notation."[43] These were: simplicity of use, ease of printing and typewriting, conciseness, recognizability, ability to generate a unique chemical nomenclature, compatibility with the accepted practices of inorganic chemical nomenclature, uniqueness, generation of an unambiguous and useful enumeration pattern, ease of manipulation by machine methods—for example, by punched cards, exhibi-

tion of associations (descriptiveness), and ability to deal with partial indeterminates.

The six essential qualities that Dyson cited at this same time seem more to the point, in consideration of all the subsequent developments away from chemical nomenclature: First he cited conciseness, then linear expression, uniqueness, adaptability for mechanical manipulation, general simplicity, and recognizability.[35]

John A. Silk responded to the IUPAC Commission's invitation with his "New System of Organic Notation," distributed privately in 1951. More than a decade later, he published additional details and improvements in his "Linear Notation for Organic Compounds."[44,45] Silk had been inspired by Dyson's London lecture of 1946. In 1949, he wrote an article reviewing methods for ring systems, but did not publish it. His interest again returned in 1957 when he served in an ICI company committee to review and comment on the tentative draft of the IUPAC-Dyson notation. He got busy again during the winter of 1960–61 to write the papers published in the *Journal of Chemical Documentation*. J. G. Cockburn, a member of the *British Chemical Abstracts* staff, responded at the same time (1951) with his "Newcastle System." Silk met him around that time and recalls that Cockburn devised his system more as a chemical shorthand than a systematic notation and was not proposing to develop it.

In August 1951, these linear notations of Dyson, Gordon, Kendall and Davison, Gruber, Silk and Cockburn were briefly reviewed by the IUPAC Commission at Massachusetts Institute of Technology, along with a notation that was based on the "Principle of Least Effort."[46] This notation, first described in 1950, later was reported in *Chemical and Engineering News*,[47] in the British ASLIB Proceedings,[48] and finally was published as a small manual.[42]

F. R. Benson meanwhile had shown how this American notation could be used for "Recording and Recovering Chemical Information with Standard Tabulating Equipment"[49]; and at this same ACS Meeting in 1953, the author described tabulating applications in a toxicity registry.[50]

E. G. Smith started using this notation with standard punched-card equipment in 1952 and in 1954 wrote a faculty report on its use in "A Punched Card Catalog of the Physical Properties of Some Common Organic Chemicals." In 1960, he reported its attractions in "Machine Searching for Chemical Structures."[51] He was so encouraged with his experiences in encoding some 50,000 structures that he volunteered to start revising the 1954 manual. After seven years of hard labor—laboring mainly with a controlling committee of notation users—he completed a manual that incorporates a number of his own ideas and provides far more comprehensive coverage than the 1954 edition. This new manual[52] is published by the McGraw-Hill Book Company and is available with additional computer-generated services from the Institute for Scientific Information.[53]

H. T. Bonnett started studying this notation in February 1953, after his interest was aroused by F. R. Benson. Bonnett and D. W. Calhoun in 1961 described the "Application of a Line Formula Notation in an Index of Chemical Structures"[54]; and at the same time, Gelberg, Nelson, Yee, and Metcalf reported on "A Program Retrieval of Organic Structure Information via Punched Cards."[55] Both groups were collaborating with E. G. Smith, using the notation that the author called to the attention of the ACS Division of Chemical Education as a "New Tool for Teaching Structural Chemistry" at this same national meeting of ACS.[56]

Alan Gelberg decided to experiment with this "least effort" notation in December 1959 at the Industrial Liaison Office of Edgewood Arsenal. He was instrumental in demonstrating the capability of the notation and the power of its permuted indexing to a number of other subsequent users.

H. Winston Hayward in November 1961 announced a "New Sequential Enumeration and Line Formula Notation System for Organic Compounds."[57] By this time the IUPAC Commission on Codification, Ciphering, and Punched Card Techniques had adopted the Dyson notation as the basis for a provisional international notation, had published a tentative version in 1958,[58] and a final version in 1961.[59] Hayward later reported on some experiences with his notation in a summer training project,[60] and additional developments are included in this symposium.[61] Dyson also will be reporting on "Modifications and Abbreviations Recommended for Computer and Visual Handling of the IUPAC Notation" in the next paper,[62] followed by H. F. Dammers and D. J. Polton's "Use of the IUPAC Notation in Computer Processing of Information on Chemical Structures."[63]

Joshua Lederberg, in Stanford University's School of Medicine in Palo Alto, introduced a "Notational Algorithm for Tree Structures" called DENDRAL-64 (*dendr*itic *al*gorithm, 1964), "A System for Computer Construction, Enumeration and Notation of Organic Molecules as Tree Structures and Cyclic Graphs."[64,65] Part II, "Topology of Cyclic Graphs," appeared a year later[66] and his Systematics of Organic Molecules, Graph Topology and Hamiltonian Circuits, A General Outline of the DENDRAL System" issued in 1966. Unique DENDRAL notations can be generated from the nonunique synonyms through Lederberg's computer programs.

No new and unique chemical notation system has come to our attention since 1964, when the "Survey of Chemical Notation Systems" was published.[43] Meanwhile Sorter, Granito, Gelberg, and their former associates at Edgewood Arsenal reported on encyclopedic notation indexes made by "permuting" the symbols of our line notation with computer programs.[67-69]

F. A. Landee in 1964 also reported on valuable computer applications, such as a "Checker" program that calculates a molecular formula from the WLN and prints out the discrepancies when this does not agree with the input formula.[70] Errors are about equally divided between notations and formulas! Landee carried these notation ideas to Moscow in 1965, in a paper on "Computer Methods of Handling Files of Chemically Oriented Information."[71] Other impressive notation-programming developments from his Computations Research Laboratory were reported elsewhere[72] and in this symposium.[73]

Gibson, Granito, Renard, and Metcalf in 1965 provided a brief introduction to "The Wiswesser Line-Notation," principally for chemistry students and teachers.[74] A more comprehensive introduction to the various applications

of this WLN was presented at Edgewood Arsenal in October 1966, and these proceedings have just been published.[75] Related reports on joint studies at J. T. Baker Chemical Company and Fort Detrick were published elsewhere in 1966.[76]

J. K. Horner contributed to this Edgewood Arsenal Conference with his separately published report[77] on "Low-Cost Storage and Retrieval of Organic Structures by Permuted Line Notations: Small Collections." The presentations by Ernest Hyde on "A Computer-Generated Open-Ended Fragment Code" and Lucille Thomson on "Structure Display" also were elaborated elsewhere.[78, 79]

This concludes the brief and incomplete historic review of chemical notation developments that were published during the past 107 years. Other more impressive developments reported at this San Francisco meeting, such as those just announced by the Institute for Scientific Information, certainly should assure all observers that chemical notations are here to stay as long as chemists continue to read and write about the substances they are creating and studying.

## LITERATURE CITED


(1) Wiswesser, W. J., Abstracts of Papers, page 9K, 118th Meeting, ACS, Chicago, Ill., September 1950.
(2) Johnson, S. W., "Chemical Notation and Nomenclature," New York, 1871.
(3) Wiswesser, W. J., "A Line-Formula Chemical Notation," pp. 123–29, Crowell Co., New York, 1954.
(4) *Ibid.*, Abstracts of Papers, page 3K, 141st Meeting, ACS, Washington, D. C., March 1962.
(5) DeMilt, Clara, *J. Chem. Educ.* **28**, 421–4 (1951).
(6) Cannizzaro, S., "Sketch of a Course of Chemical Philosophy," Alembic Club Reprint No. 18, from *Il Nuovo Cimento* **7** 321–36 (1858).
(7) Butlerov, A. M., *Z. Chem.* **1861**, pp. 549–60.
(8) Mitchell, A. D., "British Chemical Nomenclature," p. 2, E. Arnold & Co., London.
(9) Loschmidt, Josef, "Chemische Studien," Vienna, 1861; Ostwalds Klassiker No. 190, Leipzig, 1913.
(10) Carius, L., *Ann. Chemie* **126**, 195 (1863).
(11) Erlenmeyer, E., *ibid.*, **126**, 306 (1863).
(12) Hübner, H., *ibid.*, **124**, 315–23 (1862).
(13) Kekulé, A., *ibid.*, **137**, 130–96 (1866).
(14) Debus, Henry, *J. Chem. Soc.* **1866**, pp. 17–30, 256–89.
(15) Buff, H. L., *Ann. Chemie* **140**, 156 (1866); Suppl. No. 4, 159 (1866).
(16) Erlenmeyer, E., *ibid.*, **137**, 339–54; **139**, 224–33 (1866).
(17) Frankland, E., and B. F. Duppa, *ibid.*, **138**, 328–60 (1866).
(18) Graebe, C., and C. Lieberman, *Ber.* **1**, 106 (1868).
(19) Hofmann, A. W., *Ber.* 1, 169–75 (1868).
(20) Wichelhaus, H., *Ann. Chemie* **143**, 1–22 (1867).
(21) Barnard, A. J., Jr., and W. J. Wiswesser, *Laboratory Management* **5** (10), 34–6, 38, 40–2, 44 (1967).
(22) Wiswesser, W. J., and A. J. Barnard, Jr., *Brit. Soc. Rheology Bull.* "in press," July 1968.
(23) Dyson, G. M., "A New Notation and Enumeration System for Organic Compounds," Longmans, Green, London, 1947.
(24) *Ibid.*, *Nature* **154**, 114 (1944).
(25) *Ibid.*, "A New Notation for Organic Chemistry," Roy. Inst. of Chem., 1946.
(26) *Ibid.*, *Research* **2**, 104–14 (1949).
(27) *Ibid.*, *J. Chem. Educ.* **26**, 294 (1949).
(28) *Ibid.*, *Bull. Soc. Chim. France* **19**, 20 (1952).
(29) *Ibid.*, **1957**, 45–52.
(30) *Ibid.*, *ICSU Review* **4**, 110–16 (1962).
(31) *Ibid.*, *Inf. Stor. & Retr.* **2**, 59 (1964).
(32) *Ibid.*, p. 159.
(33) Dyson, G. M., M. F. Lynch, and H. Morgan, *ibid.*, **4** (1), 27 (1968).
(34) Taylor, F. L., "Enumerative Nomenclature for Organic Ring Systems." multilith, Feb. 28, 1947; *Ind. Eng. Chem.* **40**, 734–8 (1948).
(35) Dyson, G. M., "A New Notation and Enumeration System for Organic Compounds," pp. 10, 11, 2nd ed., Longmans, Green, London & New York, 1949.
(36) Gordon, M., C. E. Kendall, and W. H. T. Davison, *Proc. 11th Int. Congr. Pure & Appl. Chem.*, London, Vol. **2**, Sec. 3, pp. 115–32, July 1947.
(37) *Ibid.*, *Royal Inst. of Chem. Monograph*, Reprint 1948.
(38) Davison, W. H. T., Abstracts of Papers, 120th Meeting, ACS, New York, September 1951.
(39) Davison, W. H. T., and M. Gordon, *Am. Doc.* 8, 202–10 (1957).
(40) Davison, W. H. T., *vs.* G. M. Dyson, *Chem. Ind.* (London) **1954**, pp. 788, 934, 1044, 1180, 1300, 1428–9.
(41) Gruber, W., *Angew. Chemie* **61**, 429–31 (1949); Beihefte Nr. 58, 1950.
(42) Wiswesser, W. J., "A Line-Formula Chemical Notation," pp. 138–40, Crowell Co., New York, 1954.
(43) NAS-NRC "Survey of Chemical Notation Systems," *Publ.* **1150**, p. 8, Washington, D. C., 1964.
(44) Silk, J. A., J. CHEM. DOC. **1**, 58–62 (1961).
(45) Silk, J. A. *ibid.*, **3**, 189–95 (1963).
(46) Zipf, G. K., "Human Nature and The Principle of Least Effort," Addison-Wesley Press, Inc., Cambridge, 1949.
(47) Wiswesser, W. J., *Chem. Eng. News* **30**, 3523–6 (1952).
(48) *Ibid.*, *ASLIB Proceedings* **5**, 137 (1953).
(49) Benson, F. R., "Recording and Recovering Chemical Information with Standard Tabulating Equipment," Remington Rand Inc., South Norwalk, Conn., 1953.
(50) Wiswesser, W. J., *Adv. Chem. Ser.* **16**, 64–82 (1956).
(51) Smith, E. G., *Science* **131**, 142–6 (1960).
(52) Smith, E. G., "The Wiswesser Line-Formula Chemical Notation," McGraw-Hill, New York, 1968.
(53) Institute for Scientific Information, "Index Chemicus Registry System (ICRS)," February 1968 announcement.
(54) Bonnett, H. T., and D. W. Calhoun, J. CHEM. DOC. **2**, 2–6 (1962).
(55) Gelberg, A., W. Nelson, G. S. Yee, and E. A. Metcalf, *ibid.*, **2**, 7–11 (1962).
(56) Wiswesser, W. J., Abstracts of Papers, p. 19F, 139th Meeting, ACS, St. Louis, Mo., March 1961.
(57) Hayward, H. W., "A New Sequential Enumeration and Line Formula Notation System for Organic Compounds," U.S. Pat. Office R & D Rept. No. **21**, Washington, D. C., 1961.
(58) I.U.P.A.C., "A Proposed International Chemical Notation, Tentative Version," Commission on Codification, Ciphering and Punched Card Techniques of the IUPAC, Longmans, Green, London, 1958.
(59) I.U.P.A.C., "Rules for I.U.P.A.C. Notation for Organic Compounds," *Ibid.*, Wiley, N. Y., and Longmans, Green, London, 1951.
(60) Hayward, H. W., H. M. S. Sneed, J. H. Turnipseed, and S. J. Tauber, J. CHEM. DOC. **5**, 183–9 (1965).
(61) Hayward, H. W., Abstracts of Papers, 155th Meeting, ACS, San Francisco, Calif., April 1968.
(62) Dyson, G. M., *ibid.*
(63) Dammers, H. F., and D. J. Polton, *ibid.* (64) Lederberg, J., "DENDRAL-64." Part I. NASA **CR-57029**, December 15, 1964.
(64) Lederberg, J., "DENDRAL-64." Part I. NASA **CR-57029**, December 15, 1964.
(65) Lederberg, J., *Proc. Nat. Acad. Sci.* **53**, 134–9 (1964).

(66) Lederberg, J., "DENDRAL-64." Part II. NASA CR-68898, December 15, 1965.
(67) Sorter, P. F., C. E. Granito, J. C. Gilmer, A. Gelberg, and E. A. Metcalf, J. CHEM. DOC. 4, 56 (1954).
(68) Granito, C. E., A. Gelberg, J. E. Schultz, G. W. Gibson, and E. A. Metcalf, *ibid.*, 5, 52-5 (1965).
(69) Granito, C. E., J. E. Schultz, G. W. Gibson, A. Gelberg, R. J. Williams, and E. A. Metcalf, *ibid.*, 5, 229-33 (1965).
(70) Landee, F. A., Abstracts of Papers, p. 3F, 147th Meeting, ACS, Philadelphia, Pa., April 1964.
(71) Landee, F. A., "Computer Methods of Handling Files of Chemically Oriented Information," Dow Chemical Co., Midland, 1965.
(72) Bowman, C. M., F. A. Landee, and M. H. Reslock, J. CHEM. DOC. 7, 43-7 (1967).
(73) Bowman, C. M., F. A. Landee, N. W. Lee, and M. H. Reslock, Abstracts of Papers, 155th Meeting, ACS, San Francisco, Calif., April 1968.
(74) Gibson, G. W., C. E. Granito, D. E. Renard, and E. A. Metcalf, "The Wiswesser Line-Notation: An Introduction," CRDL Tech. Memo 7-3, Edgewood Arsenal, Md., 1965.
(75) Mitchell, J. P., Editor, "Proceedings of the Wiswesser Line Notation Meeting of the Army CIDS Program, 6-7 Oct. 1966," EASP 400-8, Edgewood Arsenal, Md., January 1968.
(76) Barnard, A. J., Jr., C. T. Kleppinger, and W. J. Wiswesser, J. CHEM. DOC. 6, 41-8 (1966).
(77) Horner, J. K., *ibid.*, 7, 85-8 (1967).
(78) Hyde, E., F. W. Matthews, L. H. Thomson, and W. J. Wiswesser, *ibid.*, 7, 200-4 (1967)
(79) Thomson, L. H., E. Hyde, and F. W. Matthews, *ibid.*, 7, 204-9 (1967).

# Use of the IUPAC Notation in Computer Processing of Information on Chemical Structures*

H. F. DAMMERS and D. J. POLTON
"Shell" Research Limited, Woodstock Agricultural Research Centre, Sittingbourne, Kent, England.

Received May 21, 1968

**A computer-operated storage and retrieval system for chemical structures based on the use of the IUPAC notation has been in operation at Shell Research Limited, Sittingbourne, Kent, England, since 1965, involving a file of nearly 50,000 compounds. Use of the IUPAC cipher has proved advantageous as regards speed and cost of both input and searching. For most searches, scanning of the information explicit in the cipher has proved adequate. Our computer programs also enable conversion of ciphers into atom-connection tables and generation of fragmentation codes. The integrated use of these facilities and their merits relative to other approaches are discussed.**

Research activities in the agricultural chemicals and public health fields carried out by Shell companies have involved the synthesis and testing of large numbers of organic chemicals. The main part of this work has been carried out partly by Shell Research Limited at its Woodstock Agricultural Research Centre at Sittingbourne, Kent, U.K. (Table I) and partly by Shell Development Company in its Agricultural Research Division at Modesto, California.

In 1962, it was decided that computer processing would be necessary if we were to achieve the fullest possible exploitation of our large file of compound data; in addition, it was considered essential that a complete structural description of the compounds be provided to the computer system rather than one based on a fragmentation code. It was agreed to pursue two different approaches as regards the method of inputting the structural information with the aim of integrating the two methods at a later date. One of these methods was based on the use of the chemical structure typewriter.[1] This approach was followed by Shell Development Company and led to the machine described by J. M. Mullen,[2] which has been in use since late 1965 at Modesto and Sittingbourne.

The other method, followed at Sittingbourne, makes use of the IUPAC notation.[3] The main reason for adopting a notation that converts graphic structures into a linear graph (cipher) suitable for input into the computer system was that it presented the most convenient and

Table I. Shell Research Laboratories at Sittingbourne, Kent

| Laboratory | Subject field |
|---|---|
| Woodstock Agricultural Research Centre | Chemicals, in particular pesticides, for use in agriculture and public health |
| Tunstall Laboratory | Toxicology; environmental health aspects of Shell products/processes |
| Milstead Laboratory | Chemical enzymology; natural products of biological significance |

Total staff ca. 500
Graduate staff ca. 150

Technical Information Services (staff: 11 + 3 part-time)

responsible for provision of:
library services, literature searches,
research data storage and retrieval,
computer services,
translations, notification, etc.

* Presented before the Division of Chemical Literature, Symposium on Notation Systems, 155th Meeting, ACS, San Francisco, Calif., April 4, 1968.