Historic Development of Chemical Notations

WILLIAM J. WISWESSER

Weed Science Research, Agricultural Research Service, U.S. Department of Agriculture, Fort Detrick, Frederick, Maryland 21701

Received February 22, 1985

The gradual development of SYSTEMS of typed symbols for describing chemical structures and substructures (radicals) is traced from its beginning in 1787. Such symbols are emphasized because "computers are symbol processors *par excellence*." Computers are the central tools of the Information Age, but future trends in the management of chemical structure information are likely to reflect deeply rooted human habits of the past, like the use of symbols in music and algebra.

For this anniversary tribute I have compiled an on-line checklist of reports on chemical notations and their component *systems of symbols*, back to the beginning 2 centuries ago. Crosland's scholarly book[1a-y] covered most of the history from 1758 to 1944; Gelberg's report,[2a-f] citation details from 1944 to 1970; Rush's reports, many details from 1787 to 1984.[3,4] Joining Rush, "We reserve the term linear notation system for those variants of the symbolic dialect that consist of unique and unambiguous representations of chemical structure." This report is further restricted to those linear notations that utilize familiar letter symbols, excluding all-numeric notations. There are great achievements in two- and three-dimensional structure display, but the enduring value of linear descriptions is that they are, like computer commands, some 110 times less costly to process; that is the central message of this paper!

During this 200-year historic review, the rocketing surge of work in notational descriptions of chemical structures is evident from these citation counts: 50 in the first 156 years (1787–1943), and then, another 54 from 1945 to 1954; 52 from 1955 to 1964; a hefty 277 from 1965 to 1974, and 134 in the past decade.[5a-r] Copies of this on-line checklist will be added to the Repository of Chemical Notation Reprints in the newly computerized Mart Science and Engineering Library at Lehigh University, Bethlehem, PA.

OUR OVERLOOKED HISTORIC HERITAGE

Crossland showed that the inspirational taproot for chemical notations and their component systems of symbols can be traced all the way back to Linnaeus "*Systema Naturae*" of 1735–1758.[1a] Torbern Bergman, a brilliant multidisciplinary scientist and student of Linnaeus, was inspired to create similar binomial Latin names like *acidum aereum* for OCO in 1773 and scores of base + acid-adjective combinations like *argentum nitratum* ($AgNO_3$), *calx aerata* ($CaCO_3$), *cuprum salitum* ($CuCl_2$), and *zincum vitriolatum* ($ZnSO_4$) in 1775.[1b] At the same time Bergman opened the way to modern chemical SYMBOLS by designing SIGNS that could be combined like the Latin binomials, such as "fused" signs for *vitriolum zinci* (a "genus–species" reversal for $ZnSO_4$) and *luna nitrata* (an alternate for $AgNO_3$).[1c] Elsewhere, with his mathematical logic, he denoted the A + B compound simply as "AB".[1d]

The chemical revolution[1e,6] started around the same time as the American Revolution, with a sudden surge of newly discovered elements and compounds that needed systematic word and symbol identifications. In 1771, Lavoisier limited the mineral kingdom to six salts, combinations of two acids and three bases, just before Priestley discovered the third state of matter with his new "airs" or gases. In 1782, Guyton de Morveau reported some 500 systematic names for new substances in his "Tableau de nomenclature Chimique".[1f] Since then, the number of known chemicals has increased 10 000-fold—by FOUR orders of magnitude!

William J. Wiswesser graduated with a B.S. from Lehigh University in 1936 and received an honorary D.Sc. from Lehigh in 1974. He has worked for the Hercules Research Center, Trojan Power Co., Picatinny Arsenal, Copper Union, Willson Products, and the U.S. Army at Fort Detrick and is presently employed there in the Weed Science Research Laboratory, Agricultural Research Service, U.S. Department of Agriculture. His interest in simplifying chemical structure descriptions began in college when he developed a chemical shorthand based on valence-line diagrams. Today, some 50 different research organizations have more than three million Wiswesser Line Notation (WLN) records in their computers. He has written or coauthored over 50 papers, and is editor of *CWIK List News* (Chemical World Index Key) as well as the *Pesticide Index*. He is past chairman of the Lehigh Valley Section and of the History of Chemistry Division of ACS. His honors include the U.S. Army Exceptional Civilian Service award, the first "Reading Chemist of the Year" award, the last Austin M. Patterson award, the Herman Skolnik award of the ACS Division of Chemical Information, the Chemical Notation Association award, and the 1981 award of the Institute of Information Scientists.

Bergman in turn inspired Guyton de Morveau to compile his memoir of 1782 on proposed "chemical denominations".[1g] Macquer, the famed French author of the first "Dictionary of Chemistry",[1h] encouraged Guyton with a perceptive appraisal that applies to human reactions to be expected from chemical notation proposals 200 years later:[1i]

> "Your new nomenclature is excellent, and, for myself, I am ready to adopt it; but I cannot answer for everyone, because you know how much men—even the enlightened—are creatures of habit. It will only be with time that they will become familiar with names, most of which will seem at first very strange and barbaric."

Guyton de Morveau in turn was the oldest inspiring author of the historic "Methode de Nomenclature Chimique" of 1787.[6,7] The first four authors "invited Jean Henri Hassenfratz, *Sous- Inspecteur des mines*, and Pierre August Adet, *Docteur-Regent* of the Paris Faculty of Medicine—both young men with a common interest in chemistry—to join their discussions on the subject of a reform of chemical symbols. To these two chemists was delegated the task of constructing a system of symbols on similar principles to those adopted for the reform of the nomenclature....Hassenfratz and Adet deplored the diversity of the chemical symbols then in use.... They considered that the time had come to build up a new system of symbols in which simple substances were represented by simple symbols and compound substances were denoted by a combination of the symbols of the simple substances."[1j,7a]

Two years later (1789), William Higgins at Oxford discredited the phlogiston theory and started the atomic age with simple LETTER SYMBOLS like C for *an atom* of copper, I for an atom of iron, and M for an atom of mercury,[1k,13] but his atomic ideas were generally ignored.

Thomas Thomson did much to discredit Higgins and promote Dalton's graphical symbols even before Dalton published them himself. (Dalton showed Thomson his table of symbols with atomic weights in 1804.) In his 1802 "System of Chemistry",[11] Thomson used 12 letter symbols A, B, G, L, M, P, S, Y, Z, C, I, and N—nine for named components (alumina, barytes, glucinia, lime, magnesia, potash, silica, yttria, zirconia) and three for the oxides of Cr, Fe, and Ni to denote mineral notations—and "acknowledged that this system of using initial letters to denote minerals had been taken from Bergman." When Berzelius, in his turn, used his "mixed-oxide" formulas for minerals in 1814, he admitted that he was following Thomson's example, with the exception that Berzelius insisted that the initials used should be those of the Latin names.[11]

Thus, the symbol-systemizing inspirations went from Linnaeus to Bergman in Sweden, then to Bergman's friend Guyton de Morveau in France, over to Thomson in England, and back to Berzelius in Sweden.

IMPORTANCE OF SYSTEMATIC SYMBOLS

In 1787, Lavoisier gave a public lecture that might well apply to users of microcomputers or terminals today, "On the necessity of reforming and perfecting the nomenclature of chemistry," with special attention to SIGNS and SYMBOLS. The words and symbols of language, he said, do far more than express ideas; they are analytical tools like those in algebra that help mankind go from the known to the unknown. "The perfection of language was most important to those beginning to study science,..."[1m] Lavoisier also sensed, as we *must* today,[8] that a nation's security depends on its long-term educational investments. In 1791 he wrote: "I see only elderly people...who still cling to the phlogiston theory. All young people adopt the new theory and I conclude that the revolution in chemistry is complete."[1n] Just three years later, the violent French Revolutionists guillotined him. Almost a century later, historian–chemist Berthelot recalled "La Revolution Chimique, Lavoisier".[1o]

BERZELIUS BEGINS THE CHEMICAL SYMBOL SET

Berzelius introduced his proposed international atomic symbols in a preliminary report to British chemists in 1813.[1p] The poorly systemized symbols that Dalton had just published in 1810[1g] must have been the major challenge for Berzelius—such as encircled letters like (An) for Sb, (Ar) for As, (B) for Bi, (C) for Cu, (Ce) for Ce, (Cob) for Co, (G) for Au, (I) for Fe, (L) for Pb, (Ma) for Mn, (N) for Ni, (P) for Pt, (S) for Ag, (T) for Sn, (Tit) for Ti, and (Z) for Zn. Berzelius sensed the need to reserve single-letter symbols for the far more frequently cited nine nonmetals: B, C, F, H, I, N, O, P, and S.

These and other international atomic symbols became established through a reluctant acceptance of Berzelius' 1814 report "On the Chemical Signs...".[9] Here he indicated his intent to standardize the use of single-letter symbols for the common nonmetals—plus "M" for the radical in Linnaeus' *muria* (salt). His "I", "L", and "R" soon were lengthened to "Ir", "Li", and "Rh", and "Yt" for yttrium was common in 1921. Around 1871–1881, Berzelius' irregularity on the uncommon metallic symbols "U", "V", and "W" must have been recognized by Harvard College's Josiah Parsons Cooke, for in his appended tables on vapor densities he lengthened these rarely cited symbols to "Ur", "Va", and "Wo".[10]

The history of chemical notations appears to be a repeating history of oversights: Hassenfratz and Adet, the notation designers in Lavoisier's task group, overlooked the need to have conveniently printable symbols.[7] In 1814, Berzelius recognized this need:

> "The chemical signs ought to be letters, for the greater facility of writing, and not to disfigure a printed book..." "In the class which I call *metalloids*, I shall employ the initial letter only...." (as in B, C, F, H, I, N, O, P, and S).

Then, he obscured this distinction with the one-letter metallic symbols noted above. This recalls a basic principle for the writer's system of notational symbols: retain his SINGLE marks only for the frequently cited nine nonmetals defined in 1814, and for other frequently cited structural groups, including Cl, Br, and ALKYL chains; use TWO-LETTER symbols for all infrequently cited elements, with no metal exceptions.

NEW CHEMICAL SYMBOLS BY THE DOZEN: 1815–1855

A characteristic of 19th century chemistry was the appearance of comprehensive chemical tables. Berzelius started these with his massive accumulation of nearly 2000 analyses of minerals and synthetic inorganic salts in 1818.[11] When this was expanded to a third edition in 1835, his accountant required 3 years to finish the percentage composition values, based on Berzelius' newly revised set of atomic weights. In these pioneering chemical-handbook tables, he introduced dozens of shorthand symbols that other chemists declined to adopt: Berzelius showed less insight than in 1814 when he complicated those simple-letter symbols with overmarked dots, commas, minus signs, and plus signs to show the number of combined oxygen, sulfur, selenium, or tellurium atoms, respectively, and later symbol-crossing bar marks to indicate doubling of the widely confused atomic weights. Along with these likely innovations from his discovery of selenium, he identified organic radicals like acetate, benzoate, and citrate with a long bar over the capitalized initial letter of the Latin name; where later additions had to be distinguished by a second mark, this was a lower case letter. These symbols did have the advantage of being standardized substance identifications, much better than the postulated formulas: in those years most chemists overlooked Avogadro's logical conclusions from simple vapor densities, and suffered C = 6 or C = 12, along with O = 8 or O = 16, relative to H = 1. Thus, the minumum number of formulas for acetic acid during the early 19th century was *four*.[1r]

Compound radicals in these Berzelian decades were believed to characterize organic compounds, just as atomic radicals characterized inorganic compounds. The first known compound radical, the cyano CN or NC group, was denoted as *Cy* soon after its discovery by Gay-Lussac in 1815.[1s] Berzelius

used *Am* for NH_4 in his 1823 "Jahresbericht".[1t] When Liebig and Woehler characterized the benzoyl radical in 1832, the imaginative Berzelius added a footnote recommending *Bz* for this widely occurring group.[1u] He recommended *Ae* for the etheric "aethyl" radical in his 1834 "Jahresbericht", but Liebig countered with a noteworthy *E* for "Ethyle". Dumas named its simpler companion group "methylene" in 1835, but Berzelius the Organizer froze it as "methyl" and *Mt* in his 1836 "Jahresbericht".[1v] Laurent referred to benzene as *phene* (from its gas illumination), symbol *Ph*, in that same year.[1w]

Liebig symbolized the acetyl radical as *Ac* in 1839, but in his formula equivalent he overlooked its oxo-containing parallel with benzoyl's *Bz*. Gerhardt added the oxo part in 1853.[1x] In 1846 a comprehensive chemical treatise was published by Thomas Cowperthwait & Company, consisting of Edward Turner's "Inorganic Chemistry" and William Gregory's "Organic Chemistry", all revised by J. B. and R. E. Rogers.[12] The organic half abounded in new symbols besides those mentioned above: *Ad* for amido, *At* for aconitic, *Ayl* for amyle, *Ci* for cinnamyl, *Ct* for cetyle, *Fo* for formyl, *Fu* for fumaric, *Gl* for glyceryl, *Ma* for maleic, *Sa* for salicyle, and *Ul* for uryl—all with virtually no regard for a spelling uniformity of the organic radical names.

Dozens of such new symbols for compound radicals were short lived, because they were statistically unimportant and structurally unrevealing: *more effort was required to remember their meanings than was saved by writing symbols.*

"RATIONAL" LINE-FORMULA NOTATIONS: 1861–1870

The leading chemists of the western world met to resolve rampant confusions about atomic, combining, equivalent, or molecular weights at the first International Chemical Congress, held at Karlsruhe in 1860. There Cannizzaro clarified the chemistry with the simple vapor-density approach of his old teacher, Amadeo Avogadro (1776–1856).[5a] In the very next year, Alexander Butlerow published the first report on the STRUCTURE of molecules and predicted that the future task of the chemist was to determine these atom-to-atom connecting patterns.[5b] Ever since this first recognition of structural importance, chemists have identified open-chain and benzene-ring structures simply by delineating the symbols for the corresponding molecular segments, one after another as connected. These are the "constitutional" or rational formulas, commonly recognized today as line-formula notations.[5]

Josef Loschmidt led this new era of structural chemistry with his "Chemische Studien":[1y,5c] it contained 368 remarkably astute graphic diagrams, including scores of benzene rings 4 years before Kekule published his historic revelation. Loschmidt's textbook included the very first line-formula descriptions, such as C_2H_5,O,C_2H_5 (with the traditional subscript multipliers) for diethyl ether, C_2H_5,O,CH_3 for ethyl methyl ether, and C_2H_5,O,C_2H_4,O,C_2H_5 for the diethyl ether of ethylene glycol. Loschmidt also used connectivity priming marks in his C_2H_5,O'',CH_3 and C_2H_5,O'',C_2H_5 along with a new hydroxyl symbol: $CH_3'Hd'$, $C_2H_5'Hd'$, $CH'O''Hd'$, C_2H_5,O,C_2H_4Hd, and $C_2H_5,O,C_3H_4(OHd)$. The connectivities were shown in his diagrams as OVERLAPPING double-line circles for O atoms and triple-line circles for N atoms!

Loschmidt was so far ahead of his time that he was completely ignored until his privately published book was discovered in 1912 by historian–chemist Richard Anschuetz; then, it immediately was reprinted as Ostwalds Klassiker Nr. 190 and became required reading for all German chemistry students.

In 1862, the first line-formula descriptions to appear in a chemical journal were those of H. Huebner, with his discussion of the isomers of $CH_2.CN.CO.Br$ and $CH_2.Br.CO.CN$.[5d] The curious twist at the starting CH_2 link appeared in thousands of examples up to the present time and originated from the habit of picturing the carbon-skeleton segments VERTICALLY, like the human skeleton, with all of the substituents cited to the right of this vertical line. Then, the segments were delineated just like text copy—left to right and top to bottom.

L. Carius copied the comma separators by denoting methyl alcohol as CHHH,O,H in 1863, but Emil Erlenmeyer omitted the punctuation with familiar notations like C_3H_7OH, C_3H_7I, and $C_3H_7OC_3H_7$. By 1866, H. L. Buff, Henry Debus, E. Erlenmeyer, and Edward Franklin, with B. F. Duppa, and August Kekule,[5e,13] had popularized this line-formula technique with scores of lucid examples like $CH_3.CO.OH$ or $H_3C\text{-}CO\text{-}OH$, $CH_3.CO.CH_3$, $CH_3.CH_2.CO.CH_3$, $NC\text{-}CH_2\text{-}CO_2H$, and $H_2N\text{-}CH_2\text{-}CH_2\text{-}CO_2H$—with casually varied hyphen or period separators. Debus described ethylene simply as $CH_2.CH_2$ and acetylene as CH.CH.

The phenyl group or one-connected benzene ring was represented by Buff and Kekule in a space-saving manner widely used today, as in their $C_6H_5\text{-}CH_2\text{-}CH_3$, $C_6H_5\text{-}CH_2\text{-}Br$, $C_6H_5.CO.Cl$, and $C_6H_5.SO_2.OH$. Kekule also introduced a long-overlooked cost-cutting notational simplification in polysubstituted benzene ring derivatives with *lower case letter* symbols to LOCATE the ring positions.[1z,13a] Kekule abandoned letters for numbers a year later (1866), failing to realize that positions are relative rather than absolute and that letters need no added comma separators for multiple citations, at least up to the 24th–26th position.

The unsaturation mark was introduced by Erlenmeyer in this same year, 1866, along with a confusingly deformed proposed structure for naphthalene.[14] This equal sign was picked up two years later by Carl Graebe in his $C_6H_5\text{-}N{=}N\text{-}C_6H_5$ description of azobenzene, and by A. W. Hofmann in his explanation of -N=C=S and -S-CN isomers.[5f]

The word "Valenz" was introduced by H. Wichelhaus in 1867,[5g] along with another long-overlooked refinement in line notations: he suggested using PERIODS only to set off the ENDS of side groups and HYPHENS to show the directly connected groups, as in his twisted front $CO.OH\text{-}CH_2\text{-}CO.OH$ and $CH_3\text{-}CHOH.CO.OH$ (corrected from "$CH_3\text{-}CH.OH\text{-}CO.OH$"). The hyphens can be omitted, as they are in most handbook or dictionary line formulas, but peroxide-like ambiguities cannot occur with unique and unambiguous delineating segments like -CO.-, -CS.-, -PO.-, and -PS.- that were implemented in the 1954 WLN manual.[13]

THE LINE-FORMULA CONVENTION

A. D. Mitchell, author of "British Chemical Nomenclature",[15] explained the traditional convention relating to "linear expression of formulae" in his opening remarks as follows:

> "Each full point [or period mark] is regarded as separating two atoms which are directly linked in the main chain of the compound, and atoms or groups attached to each of these atoms are written immediately after it and before the next full point."

His inorganic examples are punctuated here with Wichelhaus hyphens: $NH_2\text{-}SO_2.OH$, $HO\text{-}PO.PH(OH)_2$, and $(HO)_2PO.O\text{-}PO.(OH)_2$. Mitchell also reminded formula writers that "hydrogen takes precedence of other atoms, coming immediately after the relevant atom in the main chain", not *before* the first atom in the main chain, "as is often done."[38]

The 19th century notation innovations can be concluded with the opening remark in Samuel W. Johnson's 1871 presentation of "Chemical Notation and Nomenclature":[5h]

> "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."

Chart I. Comparison of 1954 WLN, 1980 AWLN, and SEFLIN for Introductory Inorganic Molecules

1954 WLN:	EH	FH	GH	IH	QH	ZH	EIEE	OSGG	QGO	QSQO
1980 AWLN:	EH	FH	GH	IH	QH	ZH	EI*EE	OS*GG	QG''O	QS*QO
SEFLIN:	HE	HF	HG	HI	HQ	HZ	EIEE	OSGG	QGO	QSOQ
1954 WLN:	WGO	BHHH	WNQ	WGQO	WSQQ	AsHHH	GSiGGG	Ca,.O	Na,.Q	
1980 AWLN:	WG''O	BHHH	WNQ	WG*QO	WSQQ	-As*HHH	G-Si#GGG	.Ca..O	.Na..Q	
SEFLIN:	QGW	H$.BHH	QNW	QGWO	QSWQ	H$ASHH	G$SIGGG	O$CA	$NAQ	

DYSON CIPHERS START A NEW ERA

Structural challenges in 1949 were far more complicated than those in 1871: Mitchell acknowledged "that systematic nomenclature is reaching breaking point and that human ingenuity can scarcely cope with it." George Malcolm Dyson, in creating a new notation era, highlighted the problem in his second edition[5i] by citing five polysyllabic attempts to describe the six-atom molecule delineated as Cl_3C-S-Cl: perchloromethyl mercaptan, tetrachloromethyl thiol, thiocarbonyl tetrachloride, trichloromethylsulphenyl chloride, and trichloromethyl sulphur chloride, to which could be added perchloromethanethiol, trichloromethanethiol, and "sulf" equivalents of the two "sulph" departures from old Latin.

The Dyson ciphering era began with a coding proposal by a British petroleum chemist, A. R. Richards.[16] His "System of Notation for Petroleum Hydrocarbons" was essentially the international system of nomenclature reduced to a few letters and C-chain digits. G. M. Dyson immediately countered with "A Notation for Organic Compounds"[5j] and provided thousands of examples in subsequent publications[5k] that continued as the "I.U.P.A.C. Notation" in 1958 and 1961.[5l] The ciphering novelty is best explained by the first paragraph of the final version:

> "In October 1946 Dyson [reference 5m] suggested a method ('ciphering' or 'cipher notation') by which organic structures were written in linear form independently of nomenclature. The system was based on the principle of dividing the structure into individual rings, chains, functional groups, substituents, etc., and citing these portions in a fixed order by conventional symbols. The system embodied a new mode of numbering the atoms of organic structures, part of which (that dealing with polycyclic systems) was related to a system of numbering independently published by Taylor in 1948; at the suggestion of the late Dr. A. M. Patterson [personal communication] these two systems were reconciled in the I.U.P.A.C. notation."

This officially promoted Dyson–IUPAC ciphering never won many users, because the ciphers were too radically different from familiar linear notations.

NEW CHEMICAL NOTATION SYSTEMS

Gordon, Kendall, and Davison introduced their G–K–D notation in 1947.[5n] Gruber,[5o] Silk,[5p] and Wiswesser[5q] submitted theirs in 1949, first as answers in the 1949 Test List, along with several others that were considered "fragment codes" rather than notations.

The Hayward notation appeared in 1961,[5r] and its continuing reports[17-20] roughly paralleled those of the Dyson–IUPAC notation.[21-28] The Skolnik notation was the next to appear with a somewhat promising continuity of subsequent reports.[29-33] Gelberg's very comprehensive review[2] gives full documentation on all of the notation reports that appeared from 1944 to 1970, including fragment or numeric codes and isolated offerings by Bouman,[2a] Eisman,[2b] Hiz,[2c] Lefkovitz,[2d] McDonnell and Pasternack,[2e] and Welch.[2f]

The most recently noted chemical notations that employ traditional letter symbols along with other marks are those of Quadrelli et al.,[34] Lin et al.,[35-38] Dromey,[39,40] and Read.[41]

The latest formal manual on a new notation ("SEFLIN", from Taiwan)[38] provides an interesting comparison with the 1954 WLN[13] and 1980 AWLN (Advanced WLN)[42,43] descriptions of introductory inorganic molecules (all three are alike for HH, ON, OO, ONN, ONO, OOO, QG, QQ, ZZ, and QNO) (Chart I). Some of the 1954 WLN symbols also have been adopted in other notations—convincing acknowledgments that these had fairly enduring attractions. A recent team contribution on "A Convenient Notation System for Organic Structure on the Basis of Connectivity Stack"[44] has mathematically interesting foundations, but still awaits extensive user evaluation.

NEW USES FOR THE OLDEST COMPUTERIZED NOTATION

Many WLN descriptions for commonly met simple substances like those cited above have not changed since 1952; they need no change as conservative on-line standardizations of familiar line-formula notations. Our introduction to the "Symposium on Uses and Applications of the WLN Today[42] cited 36 user reports from organizations like Chemical Abstracts Service, Dow, Hoffmann-La Roche, Imperial Chemical Industries, Institute for Scientific Information, National Institute for Mental Health, Searle, and Weed Science Society of America. Additional details on how WLN can advance for the 21st century were given in a companion report.[43] Its essential usefulness is verified in other recently published reference works.[45-68] As a tool for managing chemical structure information, the final attraction is low-cost usefulness—over ONE HUNDRED TIMES less costly than graphical display processing!

LESSONS LEARNED IN THIS 200-YEAR PERSPECTIVE

Predicting things to come may be entertaining, but in this Computer Age that is revolutionizing life styles throughout the world, it is likely to be an exercise in futility. A more profitable aim is to summarize for the benefit of posterity what changes occurred thus far, beyond D. H. Rouvray's 1977 review on "The Changing Role of the Symbol in the Evolution of Chemical Notation".[69]

Chemical notation is a central tool in "The Central Science".[70] This history of chemical notation traces the growth of our knowledge about MOLECULES—the ultimate free-moving particles of matter and the central objects of all applied chemistry. Therefore, an internationally useful notation must reflect the historic facts: good traditions are attractive practices employed by many people for many years. In 1947, Dyson alerted the chemical world to the need for a computer-oriented notation, but in his "ciphers" he had a gravely mistaken idea that, as he said in 1951, "We must give the axe to tradition".

Earlier lessons learned are similar instances of overlooking the obvious: in 1787 Hassenfratz and Adet failed to see the need to use printable letter symbols for all of their tabulated substances. In 1789 Higgins failed to provide similar tables of letter symbols for all of his perceived atoms or "ultimate particles of matter". From 1811 to 1858, Dalton and Berzelius

and the great majority of chemists overlooked the simple vapor-density proofs of molecules provided by Avogadro and Gay-Lussac. Berzelius, the great "Organizer of the Science", spoiled his neat and simple letter symbols with tradition-violating overmarks. In the 1860s, the line-formula pioneers obscured the front-end groups with involuted formulas like "$CH_2Br.CH_2Br$" that persist to this day. Kekule and those notation designers overlooked the advantages of LOwer CAse letTErs to LO-CA-TE ring positions and periods to umambiguously punctuate -CO.-, -CS.-, -SO.-, -PS.- and similar side groups.

In 1939, Austin M. Patterson gave me profound personal encouragement to pursue my simplifying "continuous chains of ring positions" as a lone researcher on that idea, and I traced hundreds of such paths when his "Ring Index" appeared in 1940; but I did not sharpen the basic symbol set for the commonly met groups until 1950, with a statistical analysis of frequency data from D. E. H. Frear[71] and other sources.

Foundations for these "workhorse" atomic group symbols were laid in 1945 with a three-part report on "The Periodic System and Atomic Structure",[72] featuring color-coded explanations of the three-dimensional atom-bound electron domains. These showed that multiple bonds, as Pauling noted in 1960,[73] were extremely useful organic concepts but had no sharply defined reality like the location of electron-scattering atomic cores: unsaturation marks are undesirable and unnecessary uncertaintities when hydrogenation is detailed from physical evidence.

The heart of the chemical notation problem was noted 35 years ago, in my 1950 file memo, with supporting evidence from A. D. Mitchell:[15]

> "The greatest difficulties in notation and nomenclature are not with the acyclic and monocyclic structures, but with the multicyclic ring structures, which *seem* to contain no logical beginning-to-end sequences for simple delineation. In 1938 the writer noted his beliefs that these fundamental difficulties might be removed by firmly establishing in *ALL* structures the clear and simple atom-to-atom positional concepts that are so obvious in line-formula notations...."

THE "NEW" YET OLD APPROACH

> "Since the basic enumeration is faulty, it now seems emphatically necessary, as F. Lowell Taylor advised, [reference 74] to appeal to the logic of numbers. For the simplest visualization, we also must appeal to the logic of PICTORIALLY DIRECT NOTATIONS. [1985 emphasis] That is, the notation should be like a delineated diagram...."

The rest is WLN history: these "standardized line-formula" notations never enjoyed any IUPAC recognition, and they had no other official approval; they earned user support simply because they solved various information-managing needs with less cost and confusion than other internationally recognized alternatives.

ACKNOWLEDGMENT

I am deeply grateful to WLN investors like those represented in the earlier citations[45-68] and to the members of the Chemical Notation Association (including those in the European Chemical Structure Association) for their very impressive refinements of WLN rules. Special thanks are due to E. G. Smith, Ernie Hyde, and Peter Baker for their WLN leadership and to Alan Gelberg, James Rush, and Anders Stamvik in Sweden for their comprehensive coverage of the chemical notation literature. I am especially grateful for the continuing support of this research by the U.S. Department of Agriculture and the Data Processing Center, U.S. Army Garrison, at Fort Detrick, Frederick MD.

REFERENCES AND NOTES

(1) Crosland, M. P. "Historical Studies in the Language of Chemistry"; Dover: New York, 1978; (a) p 142; (b) p 145; (c) p 238; (d) p 267; (e) pp 126–192; (f) p 128; (g) p 154; (h) pp 120, 198; (i) p 162; (j) p 245; (k) p 267; (l) p 268; (m) p 178; (n) p 189; (o) pp 169, 357; (p) p 270; (q) p 261; (r) p 321; (s) p 222; (t) pp 302–303; (u) p 303; (v) pp 303–304, 339; (w) p 309; (x) pp 296, 304; (y) p 332; (z) p 329. Detailed page references are given for this inexpensive Dover edition—$6.00 in 1978—to help students get acquainted with this chemical heritage.

(2) Gelberg, A. "Chemical Notations". In "Encyclopedia of Library and Information Science"; Marcel Dekkert: New York, 1970; Vol. 4, pp 510–528, references (a) 11, (b) 47, (c) 73, (d) 88–93, (e) 100, and (f) 141.

(3) Rush, J. E. "Handling Chemical Structure Information". *Annu. Rev. Inf. Sci. Technol.* **1978**, 13, 209–262.

(4) Rush, J. E. "History of Chemical Notations". In "Abstracts of Papers", 188th National Meeting of the American Chemical Society, Philadelphia, PA, 1984; American Chemical Society: Washington, DC, 1984.

(5) Wiswesser, W. J. "107 Years of Line-Formula Notations (1861–1968). *J. Chem. Doc.* **1968**, *8*, 146–150, references (a) 5 and 6, (b) 7, (c) 9, (d) 12; (e) 10–17, (f) 18–19, (g) 20, (h) 2, (i) 35, (j) 24, (k) 23–33, (l) 58 and 59, (m) 25, (n) 36–40, (o) 41, (p) 44 and 45, (q) 3, 47–56, and 74–79, and (r) 60 and 61.

(6) Ihde, A. J. "Lavoisier and the Chemical Revolution". In "The Development of Modern Chemistry"; Harper and Row, New York, 1964; Chapter 3, pp 57–88.

(7) Morveau, L. G. B. de; Lavoisier, A. L.; Berthollet, C. I.; Fourcroy, A. F. de. "Methode de Nomenclature" and Hassenfratz & Adet "On y a joint Un noveau Systeme de Caracteres Chimiques, adaptes a cette Nomenclature", Paris, 1787; (a) Leicester, H. M.; Klickstein, H. S. "A Source Book of Chemistry"; Harvard University Press: Cambridge, MA, 1952; pp 181–192. See also pages 83 and 168–214 in reference 1.

(8) McCorduck, P. "The Conquering Machine". *Science 84* **1984**, *November*, 138.

(9) Berzelius, J. J. "On the Chemical Signs and the Method of Employing Them To Express Chemical Proportions". *Ann. Philos.* **1814**, *3*, 51–52 (quoted in reference 7a, pages 262–264).

(10) Cooke, J. P. "Principles of Chemical Philosophy"; Harvard: Cambridge, MA, 1881; Ur, Va, and Wo in appendix.

(11) Berzelius, J. J. "Tabell,...(formulas and percent compositions), Bihang till Tredje Delen", appended to 3rd ed. "Laerobok i Kemien"; Nordstroem: Stockholm, 1818; 100 pp.

(12) Turner, E.; Gregory Wm.; Rogers, J. B.; Rogers, R. E. "Elements of Chemistry"; Cowperthwait: Philadelphia, PA, 1846; pp 147 and 272 (Cy), 525 (Ae, Mt, Kd, Bz, Ci, Sa), 526 (Ac, Ae), 543 (Ad, Am), 590 (Ul, Bz), 601 (Sa), 605 (Ci), 607–622 (Ae), 618 (Ac), 640–643 (Mt), 643 (Fo), 646 (Ct), 647 (Ayl), and 648 (Gl).

(13) Wiswesser, W. J. "Historical Development of Chemical Symbols". In "A Line-Formula Chemical Notation"; Crowell: New York, 1954; pp 123–129, (a) pp 5, 65.

(14) Erlenmeyer, E. "Naphthalin". *Justus Liebigs Ann. Chem.* **1866**, *137*, 339–354 (ring-closing L...J in footnote diagram, p 346).

(15) Mitchell, A. D. "British Chemical Nomenclature"; Arnold: London, 1948.

(16) Richards, A. R. "A System of Notation for Petroleum Hydrocarbons". *Nature (London)* **1944**, *153*, 7.

(17) Hayward, H. W.; Sneed, H. M. S.; Turnipseed, J. H.; Tauber, S. J. "Some Experience with the Hayward Line Notation System". *J. Chem. Doc.* **1965**, *5*, 183–189.

(18) Tauber, S. J.; Bolotsky, G. R.; Fraction, G. F.; Kirby, C. L.; Reed, G. R. "Algorithms for Utilizing Hayward Chemical Structure Notations"; ICEREPAT: 1965; pp 351–378.

(19) Tauber, S. J.; Fraction, G. F.; Hayward, H. W. "Chemical Structures as Information Representations, Transformations, and Calculations". In "Colloquium on Technical Preconditions for Retrieval Center Operations"; Cheydleur, B. F., Ed.; Spartan Books: Washington, DC, 1965.

(20) Hayward, H. W. "Molecular Structure, Connected Sets, Topological Linguistics, and Chemical Notation". "Abstracts of Papers", 155th National Meeting of the American Chemical Society: San Francisco, 1968; American Chemical Society, Washington, DC, 1968.

(21) Dyson, G. M. "Some Applications of Mechanical Methods to Library Problems in Organic Chemistry". *Aslib Proc.* **1947**, *22*, 23–36.

(22) Dyson, G. M. "Codification of Organic Structures". *Research (London)* **1949**, *2*, 576, 429–431.

(23) Dyson, G. M. "Mechanical Manipulation of Chemical Structure". *Inf. Storage Retr.* **1963**, *1*, 66–69.

(24) Dyson, G. M. "Modifications and Abbreviations Recommended for Computer and Visual Handling of the IUPAC Notation". "Abstracts of Papers", 155th National Meeting of the American Chemical Society, San Francisco, 1968; American Chemical Society: Washington, DC, 1968.

(25) Dammers, H. F.; Polton, D. J. "Use of the IUPAC Notation in Computer Processing of Information on Chemical Structures". *J. Chem. Doc.* **1968**, *8*, 150–160.
(26) Polton, D. J. "The IUPAC Cipher: Optimizing Its Internal Representation for Computer Processing". *Inf. Storage Retr.* **1969**, *5*, 7–25.
(27) Dyson, G. M. "A Cluster of Algorithms Relating the Nomenclature of Organic Compounds to their Structure Matrices and Ciphers". *Inf. Storage Retr.* **1964**, *2*, 159–199.
(28) Verkade, P. E. "The IUPAC Ciphering System for Organic Compounds". *Chem. Weekbl.* **1962**, *58*, 137–143.
(29) Skolnik, H.; Clow, A. "A Notation System for Indexing Pesticides". *J. Chem. Doc.* **1964**, *4*, 221–227.
(30) Skolnik, H. "A New Linear Notation System Based on Combinations of Carbon and Hydrogen". *J. Heterocycl. Chem.* **1969**, *6*, 689–695.
(31) Skolnik, H. "A Correlative Notation System for NMR Data". *J. Chem. Doc.* **1970**, *10*, 216–220.
(32) Skolnik, H. "A Chemical Fragment Notation Index". *J. Chem. Doc.* **1971**, *11*, 142–147.
(33) Skolnik, H. "A Computerized System for Storing, Retrieving, and Correlating NMR Data". *Appl. Spectrosc.* **1972**, *26*, 173.
(34) Quadrelli, L.; Bareggi, V.; Spiga, S. "A New Linear Representation of Chemical Structures". *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 37–40.
(35) Lin, C.-H. "SEFLIN: Separate Feature Linear Notation System for Chemical Compounds". *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 41–47.
(36) Lin, C.-H. "Chemical Inference Based on SEFlIN. 1. Basic Cognizance of Molecular Shape, Fragments, and Atomic Environment of Organic Compounds". *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 47–51.
(37) Lin, C.-H.; Lee, P. L.; Lin, J.-L. "A SEFLIN Compiler—Automatic Syntactic Analysis of Separate Feature Linear Notations of Chemical Compounds". *Proc. Natl. Sci. Counc., Repub. China* **1981**, *5* (3), 165–172.
(38) Lin, C.-H.; Lin, L.-C.; Lin, J.-L. "SEFLIN—Computerized Chemical Formula"; Fen Shen: Tainan, Taiwan, 1981.
(39) Dromey, R. G. "A Structural Molecular Formula for Flexible and Efficient Substructure Searching of Large Databases". *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 163–168.
(40) Dromey, R. G. "A Simple Tree-Structured Line Formula Notation for Representing Molecular Topology". *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 225–234.
(41) 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.
(42) Wiswesser, W. J. "How the WLN Began in 1949 and How It Might be in 1999". *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 88–93.
(43) Wiswesser, W. J. "The A, B, C's of the AWLN". *J. Inf. Sci.* **1982**, *4*, 69–77.
(44) Abe, H.; Kudo, Y.; Yamasaki, T.; Tanaka, K.; Sasaki, M.; Sasaki, S. "A Convenient Notation System for Organic Structure on the Basis of Connectivity Stack". *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 212–216.
(45) Chemical Abstracts Service "Wiswesser Line Notations Corresponding to Ring Index Structures"; American Chemical Society: Washington, DC, 1968; report on NSF Contract C521, PB 180901.
(46) Institute for Scientific Information "ICRS Substructure Index" *J. Chem. Doc.* **1970**, *10*, 142.
(47) National Library of Medicine "Magnetic Tape of WLN Records in Common Data Base" and "Permuted Index of WLN in Common Data Base"; NTIS: Washington, DC, 1971; PB195547 and PB199075 microfilm.
(48) CRC Press, "CRC Atlas of Spectral Data and Physical Constants of Organic Compounds", 2nd ed.; Cleveland, OH, 1972; 6 vol., 1975.
(49) American Society for Testing and Materials "Compilation of Odor and Taste Threshold Values Data"; Stahl, W. H., Ed.; American Society for Testing and Materials: Philadelphia, PA, 1973; 260 pp.
(50) National Cancer Institute "Survey of Compounds Tested for Carcinogenic Activity"; NCI, NIH: Bethesda, MD, 1951–1973; PHS Publication 149, Supplement 5, with WLN Index.
(51) Zwolinski, B. J. "Comprehensive Index of API 44-TRC Selected Data on Thermodynamics & Spectroscopy", 2nd ed.; Thermodynamics Research Center, Texas A&M University: College Station, TX, 1974; 732 pp.
(52) Institute for Scientific Information "ISI Chemical Substructure Dictionary (CSD)"; ISI: Philadelphia, PA, 1974; 70 pp.
(53) Banks, J. E. "Wiswesser Line-Formula Notation". In "Naming Organic Compounds"; Saunders: Philadelphia, PA, 1976; WLN appendix.
(54) Granito, C. E., Gibson, G. W. "The Wiswesser Line-Notation: A Brief Introduction". In "The Merck Index", 9th ed.; Merck: Rahway, NJ, 1976; MISC. 110–115.
(55) Wiswesser, W. J. "The Pesticide Index", 5th ed.; Entomological Society of America; College Park, MD, 1976.
(56) CIMI "WLN for Compounds of Merck Index", 9th ed.; CIMI: Cherry Hill, NJ, 1978; magnetic tape file of 10 000 compounds.
(57) Blake, J. E.; Brown, S. M.; Ebe, T.; Goodson, A. L.; Skevington, J. H.; Watson, C. E. "Parent Compound Handbook—Successor to the Ring Index". *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 162–167.
(58) Rosenberg, M. D. "Introduction to the Symposium on the Uses and Applications of the Wiswesser Line Notation Today". *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 87–88.
(59) Rosenberg, M. D.; DeBardeleben, M. Z.; DeBardeleben, J. F. "Chemical Supply Catalog Indexing: Now and the Future. An Ideal Place for Use of the Wiswesser Line Notation". *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 93–98.
(60) Warr, W. A. "Diverse Uses and Future Prospects for Wiswesser Line-Formula Notation". *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 98–101.
(61) Eakin, D. R. "Graphics Challenge WLN. Can WLN Hold Fast? *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 101–103.
(62) Bond, V. M.; Bowman, C. M.; Davison, L. C.; Roush, P. F.; Young, L. F. "Applications of the Wiswesser Notation at the Dow Chemical Company". *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 103–105.
(63) Fritts, L. E.; Schwind, M. M. "Using the Wiswesser Line Notation (WLN) for Online, Interactive Searching of Chemical Structures". *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 106–109.
(64) Johns, T. M.; Clare, M. "Wiswesser Line Notation as a Structural Summary Medium". *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 109–113.
(65) Walker, S. B. "Development of CAOCI (Commercially Available Organic Chemicals Index) and Its Use in ICI Plant Protection Division". *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 3–5.
(66) Beste, C. E., Chairman "Herbicide Handbook of the Weed Science Society of America", 5th ed.; Weed Science Society of America: Champaign, IL, 1983; 515 pp. (both WLN and AWLN indexes: pp 498–508).
(67) Worthing, C. R.; Walker, S. B. Eds. "Pesticide Manual", 7th ed.; British Crop Protection Council: Lavenham, England, 1983 (WLN index, pp 613–623).
(68) Chemical Abstracts Service "Ring Systems Handbook"; American Chemical Society: Washington, DC, 1984; 3 Vol.
(69) Rouvray, D. H. "The Changing Role of the Symbol in the Evolution of Chemical Notation". *Endeavour* **1977**, *1* (1), 23–31.
(70) Kauffman, G. B.; Szmant, H. H., Eds. "The Central Science"; Texas Christian University Press: Fort Worth, TX, 1984.
(71) Frear, D. E. H.; Seiferle, E. J. "Chemical Structure and Insecticidal Efficiency". *J. Econ. Entomol.* **1947**, *40*, 736–741.
(72) Wiswesser, W. J. "The Periodic System and Atomic Structure". *J. Chem. Educ.* **1945**, *22*, 314–322, 370–379, 418–426.
(73) Pauling, L. "The Nature of the Chemical Bond", 3rd ed.; Cornell University Press: Ithaca, NY, 1960; p 218.
(74) Taylor, F. L. "Enumerative Nomenclature for Organic Ring Systems". *Ind. Eng. Chem.* **1948**, *40*, 734–738. Also, many private conversations in 1949–1951 ACS Punched Card Committee.