Methods of Writing Constitutional Formulas

KUI – 20/2012
Received December 13, 2011
Accepted February 3, 2012

N. Raos[*] and *A. Miličević*

Institute for Medical Research and Occupational Health,
Ksaverska c. 2, 10 000 Zagreb, Croatia

Chemical formulas, as well as any linguistic entity, have to fulfill two basic requirements – expressiveness and economy, *i.e.* they have to express the maximal meaning with minimal means. Besides, chemical formula, being a scientific notation, has not to convey vague and scientifically unapproved meanings. This article presents the development of various kinds of chemical formulas and discusses their meaning in the historical context. Special attention is paid to line notation, developed for computers (WLN, SMILES, InChI etc.). We also discuss Seymour B. Elk's "biparametric nomenclature", based on the concept of 3-simplex, which was claimed to be universally applicable to all classes of compounds.

Key words: *Line formula, Seymour B. Elk, chemical notation, 3-simplex*

Introduction

It is well known that modern chemical (molecular) formulas[1–3] were introduced by Jøns Jakob Berzelius (1779–1848) in 1813.[4,5] The principal aim to notate the atoms as letters seems to be quite practical:

The chemical signs ought to be letters, for the greater facility of writing, and not to disfigure a printed book. Though this last circumstance may not appear of any great importance, it ought to be avoided whenever it can be done. I shall take, therefore, for the chemical sign, the ***initial letter of the Latin name of each elementary substance****: but as several have the same initial letter, I shall distinguish them in the following manner: – 1. In the class when I call metalloids, I shall employ the initial letter only, even which this letter is common to the metalloid and some metal. 2. In the class of metals, I shall distinguish those that have the same initials with another metal, or metalloid, by writing the first two letters of the word. 3. If the first two letters be common to two metals, I shall, in that case, add to the initial letter the first consonant which they have not in common.*[5]

The symbols of the elements rendered unchanged to the present day, with a very few exceptions (e.g. So for sodium, Ur for uranium, Gl for berilium (glucinium). However, the formulas were different, at least graphically. Originally, SO_2 was written as $S\overset{2}{O}$, and later, for a very long time, as SO^2. There were also symbols for "double atoms" and oxides, a kind of chemical stenography, e.g. H̶ for H_2, or S̈ for SO_3, but they were introduced later, in 1826.[5] Berzelius symbols for "radicals", *i.e.* *Am* for NH_4,[6] survived very long and were extensively employed in organic chemistry (*Ae* for ethyl, *Bz* for benzyl, *Ph* for phenyl, *etc.*).

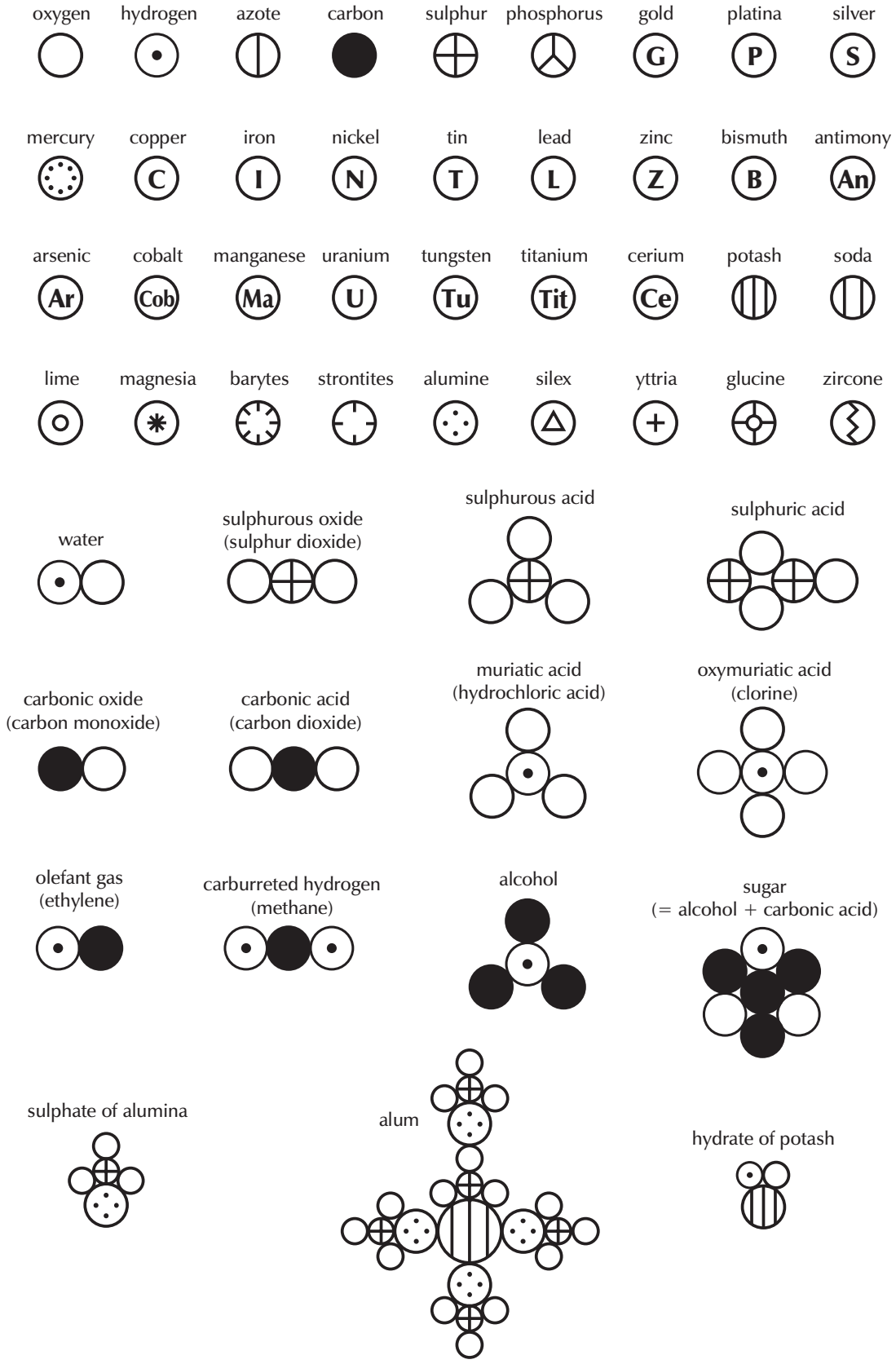
However, conceptual and graphical simplicity of the Berzelius nomenclature was not generally recognized at first. "*Berzelius symbols are horrifying*", wrote John Dalton (1766–1844). "*A young student might as soon learn Hebrew as make himself acquainted with them.*" This saying could be possibly attributed to the fact that Dalton himself introduced molecular formulas in 1810, as a consequence of his chemical atomic theory.[8,9] Dalton's formulas (Fig. 1.) were less abstract and more perceptive, but rather unpractical for larger molecules and compounds of a bigger number of elements. However, the degree of practicality is not the only difference between the two nomenclatures.

The major difference stems from the fundamental imperative of every scientific notation: it has not to imply anything that was not scientifically proved. Dalton symbols implied the real existence of atoms (that was not proved until the beginning of the twentieth century) and the molecular structure, *i.e.* bonding, which in that time was unknown even as a concept. In contrast, the Berzelius nomenclature rests on the solid concepts of constant and multiple proportions.

"*Wenn wir z. B. die Verbindungsgewichtsformel N O_6 H betrachten, so sehen wir sogleich, dass sie eine Verbindung von 14 Gewichtstheilen Stickstoff, 6 × 8 Gewichtstheilen Sauerstoff und 1 Gewichtstheilen Wasserstoff bedeutet*", was how the meaning of formulas was explained in the nineteenth century textbook,[10] published in 1878. ("*When we write e.g. compound formula N O_6 H, from this we see that the compound is composed of 14 weight parts of nitrogen, 6 × 8 weight parts of oxygen, and 1 weight part of hydrogen.*") A Berzelius formula is not, therefore, a sort of condensed structural formula, but a way of briefly writing the results of elemental analysis and molecular weight determination. As Berzelius made the first accurate determinations of atomic weights, in 1818, and especially in 1826, this interpretation is a direct consequence of his experimental work.[9]

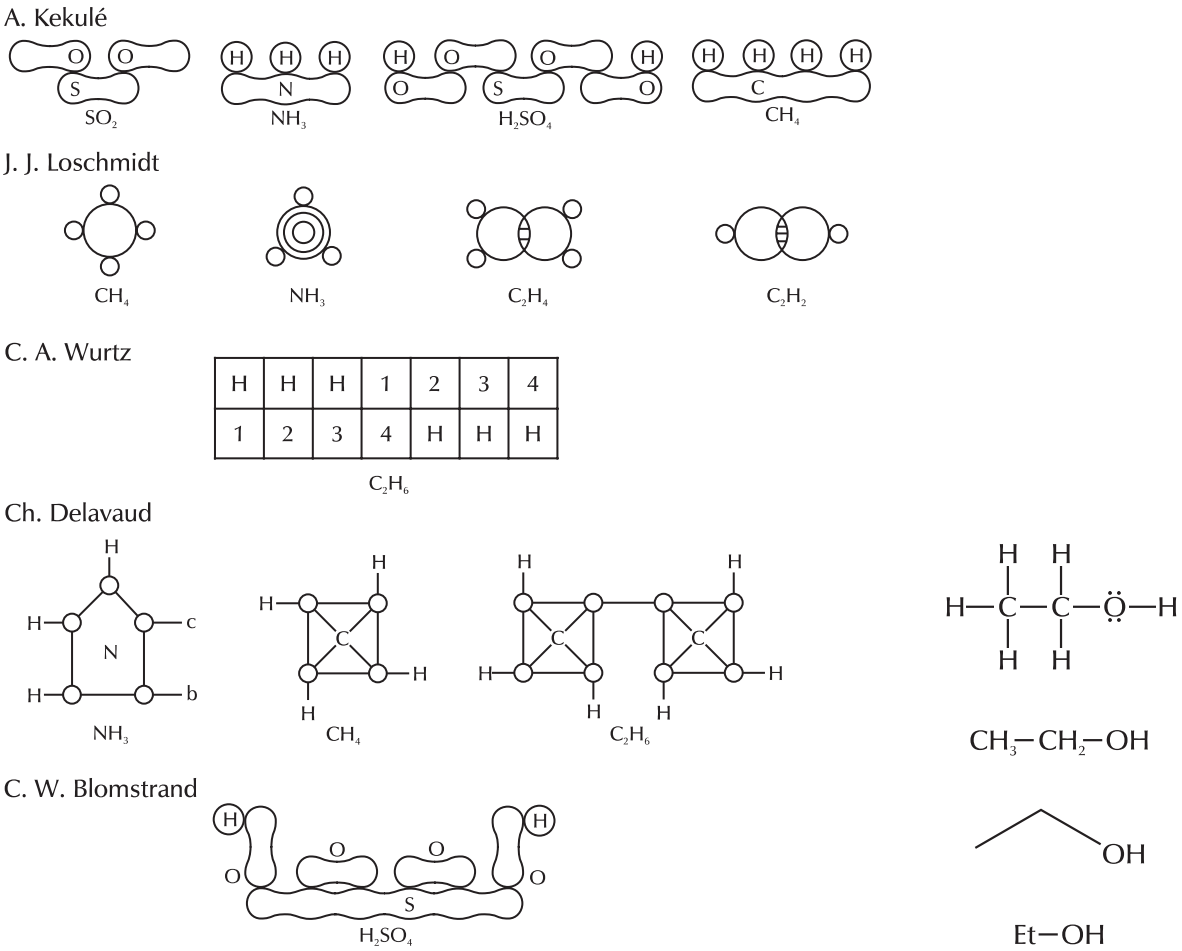
The same holds for the structural (constitutional) formula that appeared in the mid-nineteenth century, as a product of collective effort of many chemists (Edward Frankland (1825–1899), Alexander Crum Brown (1838–1922), Charles-Adolphe Wurtz (1817–1884), Archibald Scott Couper (1831–1892), and especially Friedrich August Kekulé (1829–1896).[11,12] There were many ways of writing the

[*] Corresponding author: Nenad Raos, PhD, e-mail: raos@imi.hr



F i g. 1 – *Dalton chemical formulas introduced in 1810 stem from his atomic theory*

S l i k a 1 – *Daltonove formule, koje je engleski kemičar uveo 1810. godine, proizlaze iz njegove atomske teorije*



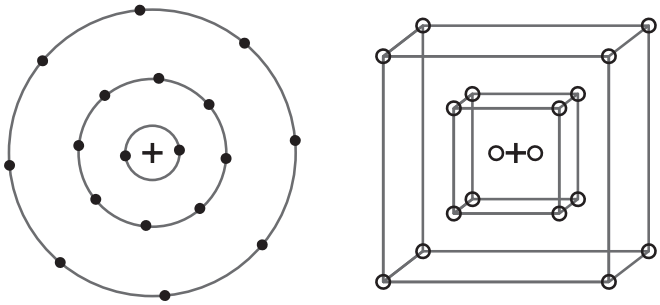
F i g. 2 – *Various methods of writing constitutional formulas in the nineteenth century (left) and today (right)*

S l i k a 2 – *Načini pisanja konstitucijskih formula u 19. stoljeću (lijevo) i danas (desno)*

same, *i.e.* constitutional formula (Fig. 2), and our present notations (using the line to mark valence) was proposed in 1858 by Couper. A constitutional formula does not imply a physical structure (geometrical-mechanical entity); it is rather a peculiar scheme of possible chemical reactions. This was first noticed by Russian chemist Alexander Mikhailovich Butlerov (1828–1886):[13–15]

To be sure, we don't know what relationship exists between the chemical influence and the mutual mechanical arrangement of the atoms in the interior of the molecule. We do not even know whether in a complex molecule two atoms which directly influence each other chemically are really situated immediately side by side. But even disregarding completely the conception of the ***physical*** *atoms we cannot deny that the chemical properties of a substance are especially conditioned by the chemical cohesion of the elements composing them. Let's start from the assumption that each chemical atom contains only a definite and limited amount of that chemical force (affinity) which it takes part in the formation of a compound. Then I would like to call* ***chemical structure*** *the chemical cohesion or the manner of the mutual binding of the atoms inside a compound substance.*[15]

In this respect constitutional formulas have the same nature as Berzelius *formules rationales* (*e.g.* $CaO.SO_3$ for calcium sulphate), introduced in 1853, and even more that of electronic formulas which Gilbert Newton Lewis (1875–1946) introduced in 1916,[16] and fully developed in 1923.[17] Interesting to note, in the same year, even before Lewis, the similar theory of chemical bonding was proposed by Walther Kossel (1888–1956),[18] but, according to L. Pauling, *"…Kossel's paper represented no significant contribution. Much of it is nonsense. He gave a long discussion of electronic valence, but nothing about covalence, although he suggested electronic structures for some molecules in which the electrons were related to two nuclei."*[19] At any rate, original Lewis formulas have less resemblance to the modern way of writing "Lewis formulas" than those proposed by Kossel (Fig. 3).



F i g. 3 – *Kossel (left) and Lewis (right) models of the argon atom*[16,18]

S l i k a 3 – *Kosselov (lijevo) i Lewisov (desno) model atoma argona*[16,18]

Formulas for computers

The advent of computers and their entrance into all spheres of life in the mid-twentieth century also produced the need to adjust chemical formulas to the age of information technology. At that time, computers used punched cards for data input. They were punched column by column, and combinations of holes in a column corresponded to the character in the ASCII code, which lacked subscripts and superscripts, Greek letters, italic *etc.* The only permitted characters were numerals, capital letters, and a limited number of other symbols.

New "computer" formulas were evolved from the short chemical formulas developed in the nineteenth century to meet type setting requirements (linotype). The formula had to be written in one line, using strictly ASCII code characters. Besides, a formula readable by computer had to be written in a very general way, and had to be as short as possible, to save computer memory and reduce input data. It had also to be friendly to humans as well as to computers.[20]

The first fully developed system of chemical formulas for computers was the Wiswesser line notation (WLN), invented in 1949 by American chemist William J. Wiswesser.[21–25] (The first computer-oriented notation was, however, introduced by G. H. Dyson[26] who had been inspired by the Richards notation for petroleum hydrocarbons.)[27] In the Wiswesser notation, acetone is written as 1V1, diethyl ether as 2O2, and acetophenone as 1VR.[28]

The Wiswesser notation, as the examples show, is based on groups of letters and numerals included in the ASCII code. In the formula for acetone, the letter "V" indicates the double bond between carbon and oxygen, while "1" means that the central C atom is also bonded to other C atoms, *i.e.* to the methyl group. In the formula of acetophenone, R denotes the aromatic ring. WLN is unique for every compound, and since it stems from regular constitutional formula, as a kind of stenography, it is easy for chemists to learn it. Although WLN was the most popular in the 1960s and 1970s, it is nowadays virtually unknown to chemists.

Another way of writing formulas for computers (line notation) is the explicit writing of short chemical formulas. Line formulas were used mostly in programs for molecular mechanics calculations, whose developments commenced in the 1960s at the Weizmann Institute of Science (Rehovot, Israel).[29–32] They are always written in brackets, *e.g.* butane is written as:

(CH3CH2CH2CH3)

The side chains are also written in brackets, but within outer brackets; *e.g.* formula for 2,3-diaminobutane is:

(CH3CH(NH2)CH(NH2)CH3)

In this notation, the atoms that close the ring are denoted with a coma, full stop, colon or any symbol other than an atomic symbol, absolute configuration designators and parentheses; formula

(,CH2CH2CH2CH2CH2,CH2)

is coding cyclohexane.

The problem of multiple bonds was solved by using different symbols for the atom of the same element. Carbonyl oxygen (=O) is denoted by Q, and hydroxyl oxygen (–O–) by O. Carbon sp^3 hybrid was coded by C, and sp^2 carbon by K. Programs for molecular mechanics were also able to construct stereoisomers by using the symbols (designators) for absolute configuration, R and S; *(S)*-proline is coded as

(.NHCH2CH2CH2.SCHKQOH).

There are many ways of writing the line formula of the same molecule, and the choice was determined by purely practical reasons.

In later years, other systems of line notation appeared, such as ROSDAL,[33] SLN[34] and SMARTS.[35] However, the line notations most widely used today are SMILES and InChI.

SMILES was developed by American chemist David Weininger and his associates in 1987.[21,36] A SMILES formula is similar to the presented line notation, but is much simpler. Brackets in a SMILES formula also denote side chains in the molecule, but the atoms effecting ring closure are indicated by 1. Atoms in aromatic rings are denoted by lowercase letters and hydrogens are omitted. Therefore, butane is CCCC, 2,3-diaminobutane CC(N)C(N)C, cyclohexane C1CCCCC1, and benzene c1ccccc1. SMILES formula is also very similar to "abbreviated chemical formula", used in molecular mechanics calculations of polypeptides; glycine is written as NCAO, valine as NC(C(C)C)AO, and histidine as NC(*BNBN*B)AO (Fig. 4).[37]

InChI notation was introduced in 2006 by IUPAC and NIST.[38,39] The special feature of InChI is writing the chemical formulas on the basis of layers of structure complexity. The first and simplest layer is connectivity. Then follow charge layer, isotopes layer, stereochemical layer, *etc.* The layers are separated by a delimiter (/). This method of writing allows omission of irrelevant layers of complexity, and writing of the formula in the simplest form. Every formula begins with "InChI", followed by number 1 (current version of InChI) and the letter S, denoting the standard version of the InChI program. The rest of the formula consists of a series of layers and sublayers. The InChI formula for ethanol is:

InChI=1S/C2H6O/c1-2-3/h3H,2H2,1H3

From this formula, one can learn that the molecular formula of ethanol is C_2H_6O, that the first, second, and third atom in the molecule are non-hydrogen atoms, and that three hydrogen atoms are bound to the first, two to the second, and one hydrogen atom is bound to the third non-hydrogen atom.

Elk's universal formulas

All the methods of presenting molecular constitution rest on two concepts. The first is the concept of coordination, the second is the concept of chain. Both concepts rest on the systematic chemical nomenclature; trioxosulphato(2–) is equivalent to formula SO_3^{2-} (sulphur is coordinated by three oxygen atoms), as well as 2,3-dimethylbutane is equivalent to $CH_3CH(CH_3)CH(CH_3)CH_3$ (two methyl groups are bonded to the butane chain).*

* Different writing constitutional formulas led to the different, colloquial names of compounds, *e.g.* dimethyldiketone, $(CH_3)_2(CO)_2$, for butane-2,3-dione, $CH_3COCOCH_3$.

(NC(CCCC)AONC(C(C)C)AONC(C*BBBBB*B)AONCAONC(CCCNB(N)N)AONC(C4S)AONC(CCAOO)AONC(C
C(C)C)AONC(C)AONC(C)AONC(C)AONC(CC.S)AONC(CCCCN)AONC(CCCNB(N)N)AONC(C*BNBN*B)AON
CAONC(CC(C)C)AONC(CC)AONC(CAON)AONC(C*BBBBOB*B)AONC(CCCNB(N)N)AONCAONC(C*BBBBOB*
B)AONC(CO)AONC(CC(C)C)AONCAONC(CAON)AONC(C*BBN≸BBBBB≸*B)AONC(C(C)C)AONC(C5S)AONC
(C)AONC(C)AONC(CCCCN)AONC(C*BBBBB*B)AONC(CCAOO)AONC(CO)AONC(CAON)AONC(C*BBBBB*B)
AONC(CAON)AONC(AOC)AONC(CCAON)AONC(C)AONC(AOC)AONC(CAON(AONC(CCCNB(N)N)AONC(CAON
)AONC(AOC)AONC(CAOO)AONCAONC(CO)AONC(AOC)AONC(CAOO)AONC(C*BBBBOB*B)AONCAONC(C(C)
CC)AONC(CC(C)C)AONC(CCAON)AONC(C(C)CC)AONC(CAON)AONC(CO)AONC(CCCNB(N)N)AONC(C*BB
NB*B)AONC(C*BBN≸BBBBB≸*B)AONC(C6S)AONC(CAON)AONC(CAOO)AONCAONC(CCCNB(N)N)AONC(AO
C)AC*NC(CC*C)AONCAONC(CO)AONC(CC)AONC(CAON)AONC(CC(C)C)AONC(C7S)AONC(CAON)AONC(C
(C)CC)AO*NC(CC*C)AONC(C6S)AONC(CO)AONC(C)AONC(CC(C)C)AONC(CC(C)C)AONC(CO)AONC(C)
AONC(CAOO)AONC(C(C)CC)AONC(AOC)AONC(C)AONC(CO)AONC(C(C)C)AONC(CAON)AONC(C7S)AONC
(C)AONC(CCCCN)AONC(CCCC)AONC(C(C)CC)AONC(C(C)C)AONC(CO)AONC(C)AONCAONC(CAOO)AONC
AONC(CC.SC)AONC(CAON)AONC(C)AONC(S*BBN≸BBBBB≸*B)AONC(CC)AONC(C)AONC(C*BBN≸BBBBB≸
*B)AONC(CCCN)AONC(CAON)AONC(CCCNB(N)N)AONC(C5S)AONC(CCCCN)AONCAONC(AOC)AONC(CAOO
)AONC(C(C)C)AONC(CC)AONC(C)AONC(C*BBN≸BBBBB≸*B)AONC(C(C)CC)AONC(C)AONCAONC(C4S)A
ONB(C)AONC(CC(C)C)AOO)

Fig. 4 – *Complete "abbreviated chemical formula" of lysozyme*[37]
Slika 4 – *Potpuna "skraćena kemijska formula" lizozima*[37]

A quite different concept was the basis for the new nomenclature proposed by American scientist Seymour B. Elk and summed up in two monographs.[40,41] He claims that his nomenclature is "a common nomenclature that would be applicable ***to ALL chemistry***" (Ref. 41, p. 2). Why? Because it is based on the universal mathematical concept of 3-simplex, *i.e.* the most simple body in 3D space (tetrahedron).

As every 3D body can be constructed from one or more tetrahedrons, Elk nomenclature is principally applicable to every molecular structure. The general formula for any molecular structure is V_4, V_4E_6, $V_4E_6F_4$, or any combination of these, where V denotes either a single atom or a large module located at the vertex of such a theoretical tetrahedron, while E denotes a module located at the centre of each of the edges of that tetrahedron. F denotes yet another atom or module situated at the centre of each of the faces of that tetrahedron. The simplest example is white phosphor, P_4, which in the Elk notation has the formula $(P1)_4$:$^{(1–5,3–7)}$(1), meaning that four phosphorous atoms form a chain with two bridges, between positions 1 and 5, and positions 3 and 7. Similarly, the Elk formula of phosphor(III) oxide, P_4O_6, is $(P1O1)_4$:$^{(1–9,5–13)}$(1O1); phosphor and oxygen atoms are situated at the vertices and the edges, respectively, of the tetrahedron (Fig. 5).

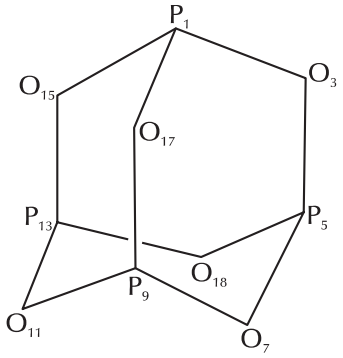


Fig. 5 – *Structural formula of phosphor(III) oxide, P_4O_6, with Elk numbering*
Slika 5 – *Strukturna formula fosforova(III) oksida, P_4O_6, s Elkovim oznakama*

Elk formulas for very complex molecules look very simple, *e.g.* C2(C1C2)$_3$C1C2C1:$^{(1–15)}$(1) for bicyclo[6.2.0]deca-1,3,5,7,9-penataene, (C1)$_8$:$^{(1–7,3–13,5–11,9–15)}$(1) for cubane, and [1/2C1/2):$^{(2,2)}$(1/2)]$_\infty$ for diamond (marks C and C mean position on the vertex and the edge of the tetrahedron, respectively). Elk also developed a new stereochemical nomenclature, which was illustrated on isomers of phosphorous sulphide, P_4S_3.[42] Besides the ordinary single (n = 1), double (n = 2), and triple bonds (n = 3), Elk also proposed the signs for the bonds with intermediate, non-integer bond

numbers: alpha bond ($0 < n < 1$, usually $n \approx 0.5$), beta bond ($2 < n < 3$) and gamma bond ($n > 3$, usually $n \approx 2.5$). There are also aleph and bet bonds, which are very close to single and double bonds, respectively, but they are not these bonds exactly. For instance, the "canonical name" for acetylene (ethyne) is H1C3C1H, for diborane $(\underline{B}\alpha H\alpha)_2$ and for naphthalene $(C\beta)_{10}{}^{(1-11)}(\beta);{}^{(3,5,7,9,12,15,17,19)}(1H)$. The name for "lithium acetylide", $(C\alpha Li\alpha)^{(1-5)}(3)$ can be easily read as "two CLi moieties bonded by a triple bond".

In discussing the new method of writing chemical formulas, one has to keep in mind two aspects. The first is practicality:[43] to be widely accepted, someone has to learn and someone has to teach the new nomenclature. Chemists also have to be willing to use it. It is not, however, very probable in spite of the opinion that the new nomenclature would possibly be proved practicable in the long run: "*With the set of known chemicals numbering over 42,000,000 (in Chemical Abstract's base) and continually growing (about 2,000 new additions every day), the desirability for a unified system for naming all chemicals simultaneously grows*".[44] The problem is further aggravated because the Elk formulas correspond to no systematic name; for the proper use of his formulas it should be necessary to find very different names for chemical compounds and also to introduce a new classification (e.g. "vertecides", "edgides", "facides").

The second problem is that the molecule is regarded primarily as a 3D object, not as a chemical entity. Therefore, Elk nomenclature provides no new insight into chemical reactivity, as did the valence theory and consequent constitutional formulas. Thus, the wider application of the Elk "biparametric nomenclature" in the future has to be regarded as dubious.

List of abbreviations
Popis kratica

ASCII	– American Standard Code for Information Interchange
InChI	– International Chemical Identifier
NIST	– National Institute of Standards and Technology
ROSDAL	– Representation of Organic Structure Descriptions Arranged Linearly
SLN	– Sybil Line Notation
SMARTS	– SMiles ARbitrary Target Specification
SMILES	– Simplified Molecular-Input Line-Entry Specification

References
Literatura

1. *M. P. Crosland,* Historical Studies in the Language of Chemistry, Dover, New York, 1978.
2. *W. J. Wiswesser,* Historical development of chemical notations, J. Chem. Inf. Comput. Sci. **25** (1985) 258–263.
3. *J. V. Knop, I. Gutman, N. Trinajstić,* Primjene teorije grafova u kemiji. VII. Prikazivanje kemijskih struktura u dokumentaciji, Kem. Ind. **24** (1975) 505–510.
4. *J. J. Berzelius,* Ann. Phil. **3** (1813) 359.
5. *J. J. Berzelius,* Ann. Phil. **3** (1814) 51–52.
6. *J. J. Berzelius,* Über die Bestimung der relativen Anzahl von einfachen Atomen in chemischen Verbindungen, Pogg. Ann. **8** (1826) 1–24.
7. *J. J. Berzelius,* Jahresbericht (1823).
8. *J. Dalton,* A New System of Chemical Philosophy, 1808.
9. *A. Lundgren,* Berzelius, Dalton, and the Chemical Atom, in E. M. Malhado, T. Frängsmyr (eds.): Enlightenment Science in the Romantic Era. The Chemistry of Berzelius and Its Cultural Setting, Cambridge University Press, Cambridge, 1992.
10. *E. F. v. Gorup-Besanez,* Lehrbuch der Chemie für den Unterricht auf Universitäten, Technischen Lehranstalten und für das Selbstudium, I. Band, Anorganische Chemie, Braunschweig, 1873.
11. *N. Raos,* Što je dvodimenzijska struktura, in N. Raos (ed.), Nove Slike iz kemije, Školska knjiga and Hrvatsko kemijsko društvo, Zagreb, 2004, pp. 63–74.
12. *D. Grdenić,* Povijest kemije, Novi Liber and Školska knjiga, Zagreb, 2001, pp. 678–690.
13. *A. M. Butlerov,* Einiges über chemische Structure der Körper, Z. Chem. Pharm. **4** (1861) 549.
14. *G. V. Bykov,* The origin of the theory of chemical structure, J. Chem. Educ. **39** (1962) 220–224.
15. *D. F. Larder, F. F. Kluge,* Alexander Mikhailovich Butlerov's theory of chemical structure, J. Chem. Educ. **48** (1971) 287–291.
16. *G. N. Lewis,* The atom and the molecule, J. Amer. Chem. Soc. **38** (1916) 762–785.
17. *G. N. Lewis,* Valence and the Structure of Atoms and Molecules, The Chemical Catalog Co., New York, 1923.
18. *W. Kossel,* Über Molekülbildung als Frage des Atombaus, Annal. Phys. **354** (1916) 229–362.
19. *L. Pauling,* G. N. Lewis and the chemical bond, J. Chem. Educ. **61** (1984) 201–203.
20. *J. Gastaiger, T. Engel,* Chemoinformatics, Wiley, Weinheim, 2003, pp. 15–157.
21. *W. J. Wiswesser,* 107 years of line-formula notations (1861–1968), J. Chem. Soc. **8** (1968) 146–150.
22. *W. J. Wiswesser,* A line-formula chemical notation, Crowell, New York, 1954.
23. *E. G. Smith,* The Wiswesser Line-Formula Chemical Notation, McGraw-Hill, New York, 1968.
24. *T. M. Johns, M. Clare,* Wiswesser line notation as structural summary medium, Chem. Inf. Comput. Sci. **22** (1982) 109–113.
25. *R. P. Swanson*, The Entrance of Informatics into Combinatorial Chemistry, in W. B. Rayward, M. E. Bowden (eds.), The History and Heritage of Scientific and Technological Information Systems, American Society for Information Science and Technology, 2004.
26. *G. H. Dyson,* Codification of organic structures, Research (London) **2** (1949) 429–431.
27. *A. R. Richards,* A system of notation for petroleum hydrocarbons, Nature **153** (1944) 715.
28. *R. Apodaca,* Everything Old is New Again – Wiswesser Line Notation (WLN), Depth-First, 20. 7. 2007. URL: http://depth-first.com/articles/2007/07/20/everything-old-is-new-again-wiswesser-line-notation-wln/
29. *S. Lifson,* J. Chim. Phys. Physiochim. Biol. **65** (1968) 40.
30. *S. Lifson, A. Warshel,* Consistent force field calculations on conformations, vibrational spectra, and enthalpies of cycloalkane and *n*-alkane molecules, J. Chem. Phys. **49** (1968) 5116–5128.
31. *S. R. Niketić, K. Rasmussen*, The Constant Force Field: A Documentation, in G. Berthier *et al.*, Lecture. Notes on Chemistry, Vol. 3, Springer, Berlin, 1977.

32. *N. Raos, Vl. Simeon*, Konformacijska analiza metodom usklađenog polja sila (CFF), Kem. Ind. **28**(11) (1979) 511–517.
33. *M. F. Lynch, J. M. Barnard, S. M. Welford,* Generic Structure. Storage and Retrieval, J. Chem. Inf. Comput. Sci. **25** (1985) 264–270.
34. *S. Ash, M. A. Cline, R. W. Homer, T. Hurst, G. B. Smith,* SYBYL Line Notation (SLN): A Versatile Language for Chemical Structure Representation, J. Chem. Inf. Comput. Sci. **37** (1997) 71–79.
35. SMARTS Theory Manual, Daylight Chemical Information Systems, Santa Fe, New Mexico. URL: http://www.daylight.com/dayhtml/doc/theory/theory. smarts.html
36. *D. Weininger,* SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comp. Sci. **28** (1988) 31–36.
37. *M. Levitt, S. Lifson,* Refinement of protein conformations using a macromolecular minimization procedure, J. Mol. Biol. **46** (1969) 269–279.
38. The IUPAC International Chemical Identifier (InChI). URL: http://www.iupac.org/inchi/.
39. InChI converter. URL: http://www.inchi.info/
40. *S. B. Elk,* A New Unifying Biparametric Nomenclature that Spans all of Chemistry, Elsevier, Amsterdam, 2004.
41. *S. B. Elk,* The Structure-Nomenclature Cycle in Chemistry, Mathematical Chemistry Monographs, 11, Kragujevac, Serbia, 2011.
42. *S. B. Elk*, Replacing traditional (chirality-based) stereochemical nomenclature with a system based solely on stereogenecity, MATCH Commun. Math. Co. **59** (2008) 453–492.
43. *N. Raos*, Seymour B. Elk, The Structure-Nomenclature Cycle of Chemistry, Croat. Chem. Acta, **84** (4) (2011) CCCCXI–CCCCXII.
44. A New Unifying Biparametric Nomenclature that Spans all of Chemistry. URL: http://www.elsevierdirect.com/product.jsp?isbn=9780444516855.

SAŽETAK

Načini pisanja konstitucijskih formula

N. Raos i A. Miličević*

Kemijska formula, kao i svaki lingvistički entitet, treba zadovoljiti dva temeljna zahtjeva, ekspresivnost i ekonomičnost, a to znači da treba izraziti što više značenja što manjim sredstvima. Kemijska formula usto ne smije, budući da je znanstveno sredstvo izražavanja, prenositi neodređena ili znanstveno neutemeljena značenja. U ovom se članku prikazuje razvoj mnogih vrsta kemijskih formula, a njihovo se značenje razmatra i u povijesnom kontekstu. Posebna se pozornost daje linijskoj notaciji (retčanim formulama) razvijenoj za potrebe rada s računalima (sustavi WLN, SMILES, InChI itd.). Razmotrili smo i "biparametarsko nazivlje" Seymoura B. Elka, utemeljeno na pojmu 3-simpleksa, koje bi trebalo biti prikladno za sve vrste spojeva.

Institut za medicinska istraživanja i medicinu rada,
Ksaverska c. 2, 10 000 Zagreb

Prispjelo, 13. prosinca 2011.
Prihvaćeno, 3. veljače 2012.