

Načini pisanja konstitucijskih formula

KUI – 19/2012
Prispjelo 13. prosinca, 2011.
Prihvaćeno 3. veljače, 2012.

N. Raos* i A. Miličević

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

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.

Ključne riječi: *Linijska formula, Seymour B. Elk, kemijske oznake, 3-simpleks*

Uvod

Svim je kemičarima poznato da je moderne kemijske (molekularne) formule^{1–3} uveo Jøns Jakob Berzelius (1779. – 1848.), 1813. godine.^{4,5} Osnovni razlog za prikazivanje atoma slovima čini se da je bio posve praktičan:

Kemijske oznake trebaju biti slova, radi olakšavanja pisanja, a ne radi smanjenja broja slika u tiskanoj knjizi. Premda se čini kako ovo posljednje možda i nije toliko važno, ipak to treba provoditi gdje god se može. Stoga ću za kemijsku oznaku uzeti početno slovo latinskog imena svake elementarne tvari: no kako ih nekoliko počinje s istim slovom, razlikovat ću ih na sljedeći način: – 1. U razredu koji zovem metaloidima [nemetallima] upotrijebit ću samo početno slovo; čak i kada je to slovo zajedničko metaloidu i nekom metalu. 2. U razredu metala, razlikovat ću one koji imaju isto početno slovo s drugim metalom, ili metaloidom, pišući prva dva slova riječi. 3. Ako su prva dva slova zajednička dvama metalima, prvom ću slovu dodati prvi suglasnik koji im nije zajednički.

Simboli elemenata ostali su nepromijenjeni do naših dana uz vrlo malo iznimaka (primjerice So za natrij, Ur za uranij ili Gl za berilij (glucinijski)). No formule su bile drugačije, barem grafički. Izvorno se SO₂ pisao kao S^oO², a dugo potom kao SO². Bilo je i oznaka za "dvostruke atome", neka vrsta kemijske stenografije, primjerice H za H₂ ili Š za SO₃, no takav je način pisanja uveden tek kasnije, 1826. godine.⁶ Berzeliusovi simboli za "radikale", npr. Am za NH₄,⁶ vrlo su se dugo održali i često su se upotrebljavali u organskoj kemiji (Ae za etil, Bz za benzil, Ph za fenil itd.).

No suštinska i grafička jednostavnost Berzeliusove nomenklature nije dugo bila prepoznata. "Berzeliusovi su simboli užasni", pisao je John Dalton (1766. – 1844.). "Mladi bi student lakše naučio hebrejski nego se na njih navikao." Rečeno se vjerojatno može pripisati činjenici da je i Dalton, 1810. godine, predložio molekularne formule koje su pro-

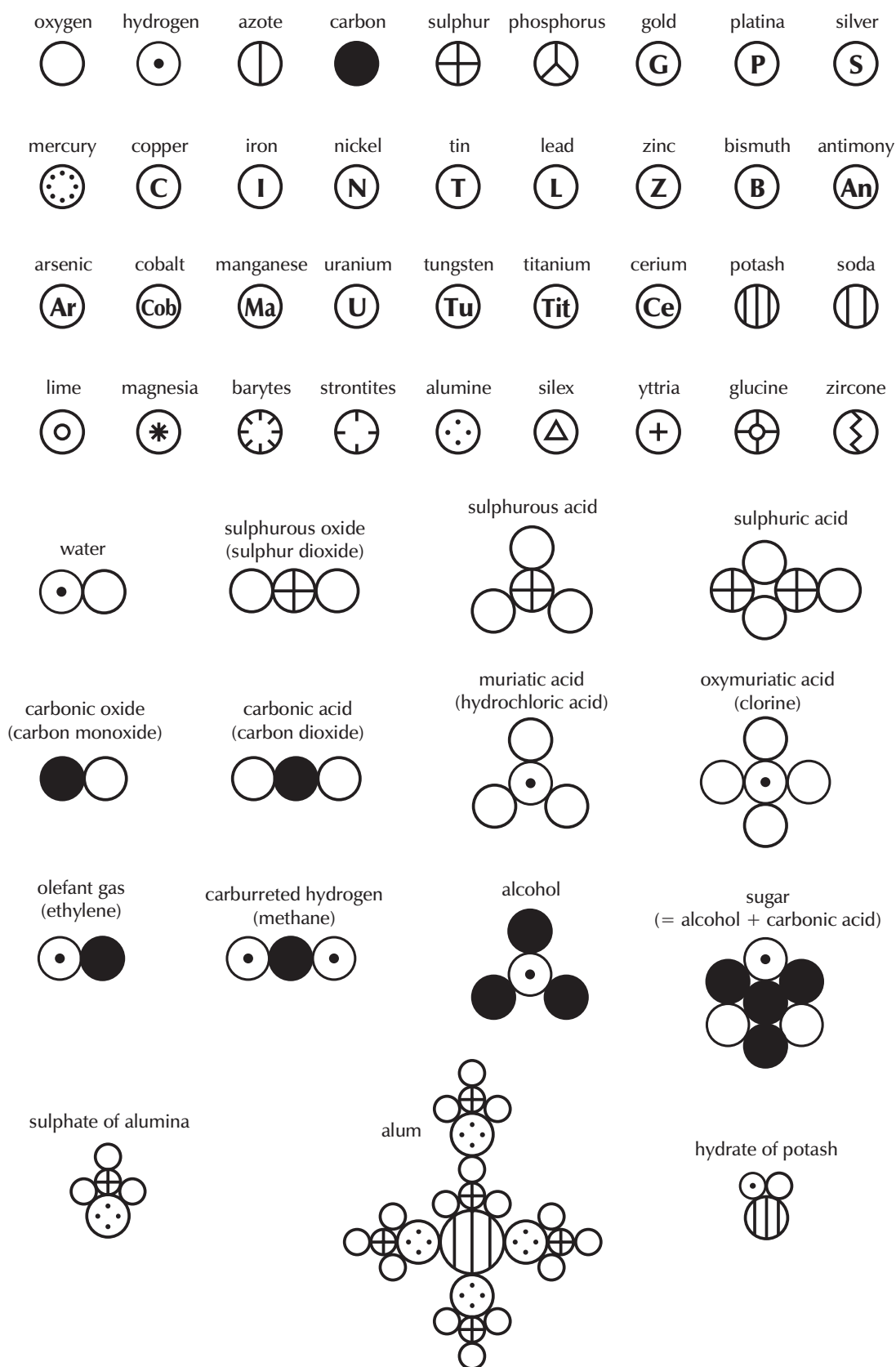
izlazile iz njegove teorije kemijskog atoma.^{8,9} Daltonove formule (slika 1) bile su manje apstraktne i lakše predočive, no bile su nepraktične za velike molekule te spojeve više elemenata. No praktičnost nije jedino po čemu se razlikuju te dvije nomenklature.

Najveća razlika proizlazi iz temeljnog zahtjeva koji se postavlja pred svaki sustav znanstvenih oznaka: one ne smiju implicirati ništa što nije znanstveno dokazano. Daltonove su oznake podrazumijevale zbiljsko postojanje atoma (što nije dokazano sve do početka prošlog stoljeća) a usto i molekularnu strukturu, tj. vezivanje, o čemu se u to vrijeme nije razmišljalo ni kao o pretpostavci. Nasuprot tome, Berzeliusova se nomenklatura osnivala na čvrstim temeljima zakona stalnih i višekratnih omjera.

"Wenn wir z. B. die Verbindungsgewichtsformel N O₆ H betrachten, so sehen wir sogleich, dass sie eine Verbindung von 14 Gewichtstheilen Stickstoff, 6 × 8 Gewichtstheilen Sauerstoff und 1 Gewichtstheilen Wasserstoff bedeutet", objašnjava se značenje formula u udžbeniku iz 19. stoljeća,¹⁰ objavljenom 1878. godine. ("Kad primjerice pogledamo težinsku formulu spoja N O₆ H, odmah vidimo da ona znači spoj 14 težinskih dijelova dušika, 6 × 8 dijelova kisika te 1 težinskog dijela vodika.") Berzeliusove formule stoga nisu neka vrsta sažetih konstitucijskih formula, nego sažet način pisanja rezultata elementne analize i određivanja molekularne mase. Kako je Berzelius napravio prvo precizno određivanje atomskih masa elemenata – 1818., a posebice 1826. godine – takva je interpretacija izravna posljedica njegova eksperimentalnog rada.⁹

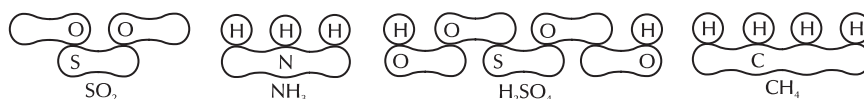
Isto se može reći i za strukturne (konstitucijske) formule koje su sredinom 19. stoljeća proizišle iz kolektivnog napora mnogih kemičara (Edwarda Franklanda (1825. – 1899.), Alexandera Cruma Browna (1838. – 1922.), Charles-Adolphe Wurtza (1817. – 1884.), Archibalda Scotta Coupera (1831. – 1892.), a posebice Friedricha Augusta Kekuléa (1829. – 1896.).^{11,12} Bilo je mnogo načina da se napiše isto, tj. konstitucijska formula (slika 2), no našu je današnju notaciju (upotrebu crtice za označavanje valencije) predložio

* Autor za dopisivanje: dr. sc. Nenad Raos, e-pošta: raos@imi.hr

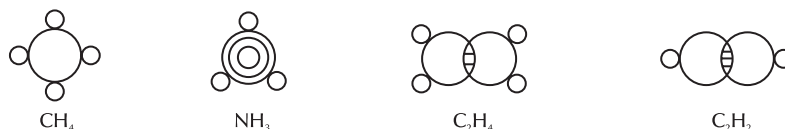


Slika 1 – Daltonove formule, koje je engleski kemičar uveo 1810. godine, proizlaze iz njegove atomske teorije
 Fig. 1 – Dalton chemical formulas introduced in 1810 stem from his atomic theory

A. Kekulé



J. J. Loschmidt

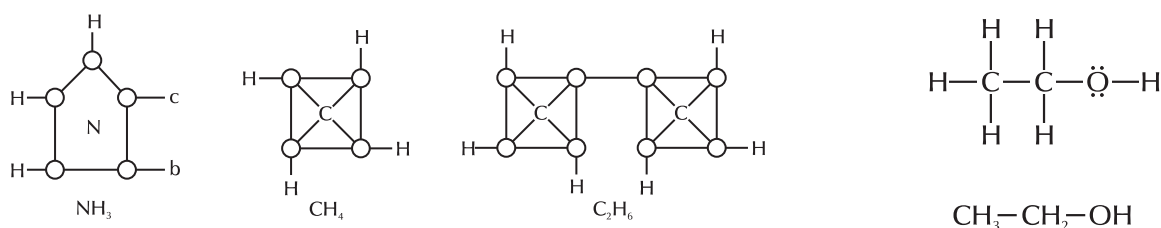


C. A. Wurtz

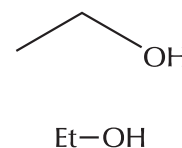
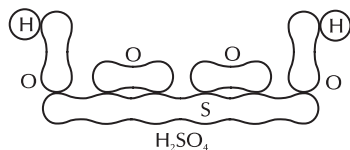
H	H	H	1	2	3	4
1	2	3	4	H	H	H

C₂H₆

Ch. Delavaud



C. W. Blomstrand



Slika 2 – Načini pisanja konstitucijskih formula u 19. stoljeću (lijevo) i danas (desno)

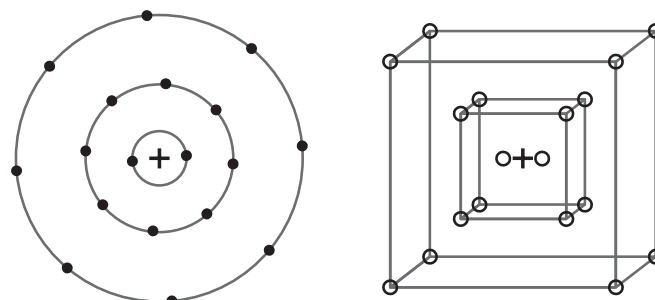
Fig. 2 – Various methods of writing constitutional formulas in the nineteenth century (left) and today (right)

Couper 1858. godine. Konstitucijska formula ne podrazumijeva nikakvu fizičku strukturu (geometrijsko-mehanički entitet); prije bi to bila osebujna shema mogućih kemijskih reakcija. To je prvi uočio ruski kemičar Aleksandr Mihailovič Butlerov (1828. – 1886.):^{13–15}

Posve je izvjesno da ne znamo koji odnos vlada između kemijskog utjecaja i uzajamnog mehaničkog rasporeda atoma unutar molekule. Ne znamo čak ni to da li se u složenoj molekuli dva atoma koji izravno kemijski utječu jedan na drugi doista nalaze jedan do drugog. No čak i ako potpuno odbacimo pojam **fizičkih** atoma, ne možemo zanijekati da su kemijska svojstva tvari ponajviše uvjetovana kemijskom kohezijom elemenata koji ih sačinjavaju. Počnimo od pretpostavke da svaki kemijski atom sadrži samo određenu i ograničenu količinu kemijske sile (afiniteta) koja sudjeluje u stvaranju nekog spoja. Tada bih volio zvati **kemijskom strukturom** kemijsku koheziju ili način uzajamnog vezivanja atoma unutar složene tvari.¹⁵

U tom bi pogledu konstitucijske formule bile blizu Berzeliovim *formules rationales* (npr. CaO.SO₃ za kalcijev sulfat) uvedenima 1853. godine, a još više elektronskim formulama koje je uveo Gilbert Newton Lewis (1875. – 1946.) 1916.¹⁶ te ih do kraja razvio 1923. godine.¹⁷ Zanimljivo je napomenuti da je iste godine, čak prije Lewisa, sličnu teori-

ju kemijskog vezivanja predložio Walther Kossel (1888. – 1956.),¹⁸ ali, kako kaže L. Pauling, "...Kosselov članak nije donio ništa bitnoga. Većinom su u njemu besmislice. Dao je dugu raspravu o elektronskoj valenciji, no ništa o kovalenciji, iako je predložio i elektronske strukture molekula u kojima elektroni pripadaju obim jezgrama."¹⁹ Ipak, izvorne Lewisove formule manje slične suvremenom načinu pisanja "Lewisovih formula" od formula koje je predložio Kossel (slika 3).

Slika 3 – Kosselov (lijevo) i Lewisov (desno) model atoma argona^{16,18}Fig. 3 – Kossel (left) and Lewis (right) models of the argon atom^{16,18}

Formule za računala

Pojava elektroničkih računala i njihov ulazak u sve sfere života, sredinom 20. stoljeća, dovela je do potrebe prilagodbe kemijskih formula informatičkom dobu. U to su se doba za unos podataka u računala upotrebljavale bušene kartice. Bile su bušene stupac po stupac, a kombinacija rupa u stupcu odgovarala je znaku koda ASCII, kojemu su nedostajali supskripti i superskripti, grčka i kurzivna slova itd. Jedini dopušteni znakovi bile su brojke, velika slova te ograničen broj drugih simbola.

Nove, "kompjutorske" formule potekle su iz sažetih formula koje su razvijene u 19. stoljeću radi prilagodbe tiskarskom slogu (linotip). Formulu treba napisati u jednom retku, primjenjujući strogo simbole definirane kodom ASCII. Formula koju može čitati računalo treba usto biti napisana vrlo općenito, a mora biti napisana i što je moguće kraće, kako bi se uštedjelo na računalnoj memoriji te smanjila količina ulaznih podataka. Formula također mora biti prilagođena i čovjeku i računalu.²⁰

Prvi potpuno razvijen sustav kemijskih formula za računala bila je Wiswesserova linijska notacija (*Wiswesser line notation*, WLN), koju je 1949. iznašao američki kemičar William J. Wiswesser.^{21–25} (Prvu notaciju za računala uveo je međutim G. H. Dyson,²⁶ koji je pak bio potaknut Richardsovom notacijom za ugljikovodike iz nafte.)²⁷ U Wiswesserovoj notaciji aceton se piše 1V1, dietil-eter 2O2, a acetofenon 1VR.²⁸

Kako nam ovi primjeri pokazuju, Wiswesserova se notacija temelji na skupinama slova i brojaka iz ASCII-koda. U formuli acetona slovo "V" ukazuje na dvostruku vezu između ugljika i kisika, dok "1" znači da je i središnji C-atom vezan za drugi C-atom, tj. metilnu skupinu. U formuli acetofenona, R označava aromatski prsten. WLN je jedinstvena za svaki spoj, a budući da proizlazi iz regularne konstitucijske formule, kao neka vrsta stenografije, kemičaru nije teško da je nauči. Iako je WLN bila vrlo popularna u 1960. i 1970. godinama, za nju danas malo tko zna.

Drugi je način pisanja formula za računala (linijska notacija) eksplicitni način pisanja sažetih formula. Linijske (retčane) formule najviše su se primjenjivale u programima za molekularno-mehaničke proračune, čiji je razvoj započeo u 1960-im godinama na Weizmannovu znanstvenom institutu (Rehovot, Izrael).^{29–32} One se uvijek pišu unutar zagrade, butan se primjerice piše kao:



Pobočni lanci također se pišu u zagradi, no unutar glavne zagrade; formula je 2,3-diaminobutana primjerice:



U tom sustavu oznaka atomi koji zatvaraju prsten označavaju se zarezom, točkom, dvotočkom ili bilo kojom drugim znakom koji nije simbol atoma, oznaka apsolutne konfiguracije ili zagrade; formula



kodira cikloheksan.

Problem višestrukih veza riješen je upotrebom različitih simbola za atom istog elementa. Karbonilni kisik (=O) označava se slovom Q, a kisik hidoksilne skupine (–O–)

simbolom O. Ugljikov sp³-hibrid kodira se slovom C, a sp²-ugljik slovom K. Programi za molekularnu mehaniku mogli su konstruirati i stereoisomere upotrebom simbola (oznaka) za apsolutnu konfiguraciju, R i S; (S)-prolin piše se kao

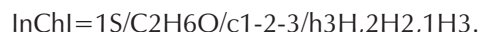


Retčana formula iste molekule može se napisati na mnogo načina, a izbor ovisi o čisto praktičkim razlozima.

Posljednjih godina pojavili su se i drugi sustavi linijske notacije poput sustava ROSDAL,³³ SLN³⁴ i SMARTS.³⁵ Ipak, sustavi linijske notacije koja se danas najviše primjenjuju su SMILES i InChI.

Sustav SMILES razvio je američki kemičar David Weininger sa suradnicima 1987. godine.^{21,36} Formula SMILES slična je prikazanoj linijskoj notaciji, no mnogo je jednostavnija. I u formuli SMILES zagrade označavaju pobočne lance u molekuli, ali su atomi koji zatvaraju prsten označeni brojkom 1. Atomi aromatskog prstena označavaju se malim slovima dok se vodikovi atomi ne pišu. Stoga je butan CCCC, 2,3-diaminobutan CC(N)C(N)C, cikloheksan C1CCCCC1, a benzen c1ccccc1. Formula SMILES vrlo sliči "skraćenoj kemijskoj formuli" koja se primjenjuje u molekularno-mehaničkim proračunima polipeptida; glicin se piše NCAO, valin NC(C(C)C)AO, a histidin NC(*BNBN*B)AO (slika 4).³⁷

Notaciju InChI uveli su 2006. godine IUPAC i NIST.^{38,39} Posebno obilježje InChI-ja pisanje je kemijskih formula na temelju slojeva strukturne kompleksnosti. Prvi i najjednostavniji sloj je sloj povezanosti. Potom slijedi nabojni sloj, izotopni sloj, stereokemijski sloj itd. Slojevi su odijeljeni kosom crtom (/). Takav način pisanja omogućuje ispuštanje irelevantnih slojeva kompleksnosti, pa se formula može napisati posve jednostavno. Svaka formula počinje s "InChI", nakon čega slijedi broj 1 (trenutna verzija InChI-ja) te slovo S, koje označava standardnu verziju programa InChI. Ostatak formule sastoji se od niza slojeva i podslojeva. Formula u kodu InChI za etanol je:



Iz te se formule vidi da je molekularna formula etanola C₂H₆O, da prvi, drugi, i treći atom u molekuli nisu vodikovi atomi te da su tri vodikova atoma vezana za prvi, dva za drugi, a jedan za treći nevodikov atom.

Elkove univerzalne formule

Svi načini prikazivanja konstitucije molekula temelje se na dva koncepta. Prvi je koncept koordinacija, drugi je koncept koncept lanca. Oba su koncepta u tijesnoj vezi sa sustavnom kemijskom nomenklaturom: triokso sulfato(2–) odgovara formuli SO₃^{2–} (sumpor je koordiniran sa tri kisikova atoma), kao što je 2,3-dimetilbutan drugi način pisanja formule CH₃CH(CH₃)CH(CH₃)CH₃ (dvije metilne skupine vezane su za lanac molekule butana).*

Iz posve drugog polazišta potekla je nova nomenklatura, koju je predložio američki znanstvenik Seymour B. Elk te je

*Različiti načini pisanja konstitucijskih formula vode do različitih, kolo-kvijalnih, imena spojeva, npr. dimetildiketon, (CH₃)₂(CO)₂, za butan-2,3-dion, CH₃COCOCH₃.

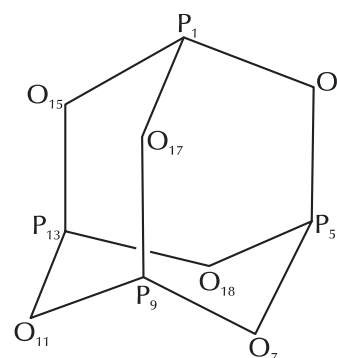
(NC(CCCC)AONC(C(C)C)AONC(C*B*BBBB*B)AONCAONC(CCCNB(N)N)AONC(C4S)AONC(CCA00)AONC(C(C)C)AONC(C)AONC(C)AONC(C)AONC(CC.S)AONC(CCCCN)AONC(CCCNB(N)N)AONC(C*B*BNB*N*B)AONCAONC(C(C)C)AONC(CC)AONC(CAON)AONC(C*B*BBBBOB*B)AONC(CCCNB(N)N)AONCAONC(C*B*BBBBOB*B)AONC(CO)AONC(C(C)C)AONCAONC(CAON)AONC(C*B*BNB*BBBBB*B)AONC(C(C)C)AONC(C5S)AONC(C)AONC(C)AONC(CCCCN)AONC(C*B*BBBB*B)AONC(CCA00)AONC(CO)AONC(CAON)AONC(C*B*BBBB*B)AONC(CAON)AONC(AOC)AONC(CCAON)AONC(C)AONC(AOC)AONC(CAON(AONC(CCCNB(N)N)AONC(CAON)AONC(AOC)AONC(CA00)AONCAONC(CO)AONC(AOC)AONC(CA00)AONC(C*B*BBBBOB*B)AONCAONC(C(C)CC)AONC(C(C)C)AONC(CCAON)AONC(C(C)CC)AONC(CAON)AONC(CO)AONC(CCCNB(N)N)AONC(C*B*BNB*N*B)AONC(C*B*BNB*BBBBB*B)AONC(C6S)AONC(CAON)AONC(CA00)AONCAONC(CCCNB(N)N)AONC(AOC)AC*NC(CC*C)AONCAONC(CO)AONC(CC)AONC(CAON)AONC(C(C)C)AONC(C7S)AONC(CAON)AONC(C(C)CC)A0*NC(CC*C)AONC(C6S)AONC(CO)AONC(C)AONC(C(C)C)AONC(C(C)C)AONC(CO)AONC(C)AONC(CA00)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(CA00)AONC(AONC(CC.SC)AONC(CAON)AONC(C)AONC(S*B*BNB*BBBBB*B)AONC(CC)AONC(C)AONC(C*B*BNB*BBBBB*B)AONC(CCCN)AONC(CAON)AONC(CCCNB(N)N)AONC(C5S)AONC(CCCCN)AONCAONC(AOC)AONC(CA00)AONC(C(C)C)AONC(CC)AONC(C)AONC(C*B*BNB*BBBBB*B)AONC(C(C)CC)AONC(C)AONCAONC(C4S)AONB(C)AONC(C(C)C)A00)

Slika 4 – Potpuna “skraćena kemijska formula” lizozima³⁷Fig. 4 – Complete “abbreviated chemical formula” of lysozyme³⁷

detaljno izložio u dvije monografije.^{40,41} Elk tvrdi da je njegova nomenklatura “zajednička nomenklatura koja se može primijeniti u **CJELOKUPNOJ kemiji**” (ref. 41, str. 2). Zašto? Zato što se zasniva na univerzalnom matematičkom pojmu 3-simpleksa, tj. najjednostavnijem tijelu u trodimenzijskom prostoru (tetraedru).

Kako se svako trodimenzijsko tijelo može konstruirati iz jednog ili više tetraedara, Elkova se nomenklatura može načelno primijeniti na svakoj molekularnoj strukturi. Opća formula molekularne strukture je V_4 , V_4E_6 , $V_4E_6F_4$, ili njihova kombinacija, gdje V označava pojedini atom ili pak veliki modul smješten na vrhu (vertex) takvog teoretskog tetraedra, dok E označava modul smješten u središtu svakog brida (edge) takvog tetraedra. F označava još jedan atom ili modul smješten na stranici (face) tetraedra. Najjednostavniji primjer je bijeli fosfor, P_4 , koji po Elku ima formulu $(P1)_4^{:(1-5,3-7)}(1)$, što znači da četiri atoma fosfora čine lanac s dva mosta, između položaja 1 i 5 te 3 i 7. Slično tome, Elkova formula fosforova(III) oksida, P_4O_6 , je $(P1O1)_4^{:(1-9,5-13)}(1O1)$; atomi fosfora smješteni su na vrhovima, a atomi kisika na bridovima tetraedra (slika 5).

Elkove formule za vrlo kompleksne molekule su jednostavne, primjerice $C2(\underline{C1C2})_3C1\underline{C2C1}^{:(1-15)}(1)$ za biciklo[6.2.0]deka-1,3,5,7,9-penataen, $(\underline{C1})_8^{:(1-7,3-13,5-11,9-15)}(1)$ za kuban, a $[1/2C1/2]^{:(2,2)}(1/2)_\infty$ za dijamant (oznaka \underline{C} označava položaj na vrhu, a $\underline{\underline{C}}$ na bridu tetraedra). Elk je

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

razvio i novu stereokemijsku nomenklaturu, koju daje na primjeru fosforovih sulfida, P_4S_3 .⁴² Uz jednostruke ($n = 1$), dvostruke ($n = 2$) i trostruke veze ($n = 3$), Elk predlaže i oznake za veze s racionalnim veznim brojem: alfa-vezu ($0 < n < 1$, obično $n \approx 0,5$), beta-vezu ($2 < n < 3$) i gama-vezu ($n > 3$, obično $n \approx 2,5$). Usto uvodi alef-vezu i bet-vezu, veze slične jednostrukim i dvostrukim vezama – prve prvima, druge drugima – no nisu njima identične. “Kanonsko

ime" za acetilen (etin) je primjerice H1C3C1H, za diboran ($B\alpha H\alpha$)₂ a za naftalen ($C\beta$)₁₀^{(1-11)(\beta)},_{(3,5,7,9,12,15,17,19)(1H)}. Ime "litijeva acetilida" ($CaLi\alpha$)⁽¹⁻⁵⁾⁽³⁾ može se jednostavno pročitati kao "dvije skupine CLi vezane trostrukom vezom".

Razmatrajući nov način pisanja kemijskih formula moramo imati na umu dva njegova vida. Prvi je praktičnost:⁴³ da bi se mogao naširoko primjenjivati, netko ga mora naučiti i netko ga treba učiti. Kemičari moraju biti motivirani da se njime služe. To pak nije izgledno unatoč mišljenju da bi se nova nomenklatura dugoročno mogla pokazati praktičnom: "Sa skupom od više od 42 000 000 poznatih kemikalija (u datoteci *Chemical Abstracts*) koji usto neprestano raste (oko 2000 novih članova svakog dana), poželjnost jedinstvenog sustava za imenovanje svih kemikalija raste zajedno s njime".⁴⁴ Problem je još teži zbog toga što Elkova formula ne odgovara nijednom sustavnom imenu; za ispravnu upotrebu njegovih formula trebalo bi iznaći posve drugačija imena za kemijske spojeve i uvesti novu klasifikaciju (primjerice "vršnjaci", "bridnjaci" i "plošnjaci").

Druga je poteškoća što se molekula promatra prije svega kao trodimenzijski objekt, a ne kako kemijski entitet. Stoga Elkova nomenklatura ne pruža nov uvid u kemijsku reaktivnost, kao što je to učinila teorija valencije i iz nje izvedena konstitucijska formula. Stoga dvojimo u širu primjenu Elkove "biparametrijske nomenklature" u budućnosti.

Popis kratica

List of abbreviations

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

Literatura

References

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. 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 Bestimmung 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, u E. M. Malhado, T. Frängsmyr (ur.), *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 Selbststudium*, I. Band, *Anorganische Chemie*, Braunschweig, 1873.
11. N. Raos, Što je dvodimenzijska struktura, u N. Raos (ur.), *Nove Slike iz kemije*, Školska knjiga i Hrvatsko kemijsko društvo, Zagreb, 2004., str. 63–74.
12. D. Grdenić, Povijest kemije, *Novi Liber i Školska knjiga*, Zagreb, 2001., str. 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, *Cheminformatics*, Wiley, Weinheim, 2003., str. 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, u W. B. Rayward, M. E. Bowden (ur.), *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, u G. Berthier et al., *Lecture Notes on Chemistry*, Vol. 3, Springer, Berlin, 1977.
32. N. Raos, VI. 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, 2011.
42. S. B. Elk, Replacing traditional (chirality-based) stereochemical nomenclature with a system based solely on stereogenicity, *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>.

SUMMARY

Methods of Writing Constitutional Formulas

N. Raos* and A. Miličević

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.

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

*Received December 13, 2011
Accepted February 3, 2012*