PIERRE LASZLO

CONVENTIONALITIES IN FORMULA WRITING

INTRODUCTION

Chemical formulas, those small icons which chemists are wont to scribble in their notebooks and in odd places, such as the back of an envelope, and which to the general public have become emblems of their profession, are an excellent topic for history. These artefacts remain today tools for communication within the community of chemists. They continue serving as didactic instruments in teaching. The establishment of an individual formula for a chemical compound or a substance chronicles the laboratory methods, both routine and specific, which came into play in order for it to be written down and to assume the status of the analog of a word, to be stored within the growing lexicon of chemistry.

When addressing this topic, the historical narrative, besides its usual needs for accuracy and for an unerring sense of the strange and original taste of the bygone, demands the twin crutches of philosophical and linguistic inquiries. I wish to provide these complements if not in full, at least in a manner suggestive of some of the main issues.

I shall concern myself with the period of consolidation, when structural formulas entered the language of organic chemistry and started becoming stereotyped, the approximate period 1865–1905.[1] Why choose such a periodization? Because it brackets, approximately, the birth of the modern chemistry journal, *JACS,* which was started in 1879, and that of the modern comprehensive repertory of new chemical compounds, *Chemical Abstracts,* which was launched in 1907. Kekulé announced in 1865 his cyclic structure of benzene. The Chemical Society published in London, in 1882, *Nomenclature and Notation,* the first guidelines for establishing systematic and uniform practice. And the American Chemical Society followed suit in 1884 by establishing its Committee on Nomenclature and Notation. The international conference convened in Geneva in 1892 established norms for chemical nomenclature.[2] And Alfred Werner, in 1895, gave a systematic nomenclature for coordination complexes. Key milestones in the history of molecular formulas – so-called "structural formulas"; I favor the adjective "molecular" since the meaning of "structural" has changed considerably over the twentieth century – include the serendipitous synthesis of mauveine (1857), the first synthesis of alizarin (1868) and the identification of ibogaine (1905). The forty years 1865–1905 were thus for molecular formulas of organic compounds those of the rise in their practical use, of their standardization and also of the first challenges to the rules governing them.

As always in history of science, the risk of Whig history lurks at every corner of the retrodictive narrative. The danger is to read into the structural formulas, as they

U. Klein (ed.), Tools and Modes of Representation in the Laboratory Sciences, 47–60.

were used at the end of the nineteenth/-beginning of the twentieth century, meanings which they had yet to acquire in the post-Gilbert N. Lewis and post-Linus C. Pauling eras. Examples of such potential anachronisms are: (i.) viewing benzene rings as *ipso facto* synonyms of "aromaticity"; (ii.) reading double bonds as implying shorter and stronger interatomic linkages; (iii.) interpreting loss of a water molecule in a dehydration process as a thermodynamic driving force for the observed conversion. The eerie superficial similarity of these late nineteenth-century formulas to our early twenty-first century formulas can easily become misleading.

1. ON THE NECESSITY OF CONVENTIONS

Structural formulas came of age in the 1860s.[3] With the advent of graphical representations for chemical compounds, came the attendant need for arbitrary conventions in making such graphemes become integral and essential components of publications. One witnesses indeed a stabilization of such graphic notation for organic compounds during the period of consideration, 1865–1905. Moreover, the formulas rapidly settled into a format which would endure spectacularly, to the extent that, more than a century later, present-day formulas continue bearing a strong resemblance to their forebears. Such a success begs for consideration of the means by which it came to be achieved.

These means were, first and foremost, the anchoring in a set of routine laboratory procedures. Justus von Liebig's Giessen laboratory in its very organization had pioneered such an analytical methodology.[4] The first step was to obtain the analysis of the unknown compound, and to derive from it the elemental composition. The molecular weight was obtained from measurement of the density, and later on from measurement of the melting point depression or of the elevation of the boiling point. These assumed explicitly or implicitly the validity of the Avogadro-Ampère hypothesis. Consideration of the molecular weight turned the elemental composition into a compositional formula such as $C_2H_4O_2$, where the subscripts denoted the numbers of atoms of each element denoted by letters. Establishment of the compositional formula was a well-trodden and an established routine. Empirical analytical data were thus converted into a sequence of alphanumerical symbols, the formula. But to what extent did chemists who followed this routine believe in atoms?[5] Were they only paying lip service to the tenets of positivism by denying actual existence to the atoms while their very behavior demonstrated a blind faith in their existence?

The establishment of the structural formula was an exercise in representation which cannot be considered outside its historical context. During the period under consideration, the late nineteenth century, the structural (or constitutional) formula was a mapping-out of the relative location of the atoms, based upon a series of reactions the substrate molecule had undergone.

Thus structural formulas postulated the at least empirical validity of atomic theory.[6] They paradoxically embodied reactivity data as a static structure. And the latter was to be considered as a map,[7] rather than as an actual spatial geometry for the molecule depicted. This last point is crucial. Structural formulas during at least the period 1865–1905 were much closer to a topological than to a geometrical representation. I can do no better than to quote to this effect Ira Remsen's textbook, first published in 1885:

> In studying the chemical conduct of these compounds, their decompositions, and the modes of preparing them, we become familiar with many facts which it is desirable to represent by means of the formulas.[8]

Remsen, at the time one of the best German-trained American organic chemists, kept reiterating throughout his textbook the two key concepts in the above definition of the structural formula: (i.) it embodies the familiarity of the chemist with the compound under study; (ii.) such familiarity is gained from application to that compound of well-understood reactions. To quote again Remsen:

> The formulas are but the condensed expressions of the conclusions which are drawn from the reactions.[9]

Thus to read an actual molecular geometry into a turn of the twentieth century structural formula would be an anachronism.

The phrase "condensed expressions" in this last quotation of Remsen's begs for attention. Indeed, understanding of structural formulas is improved when one realizes that chemists were intent upon concision, that they aimed at shorthand notation of the empirical evidence. This explains one of the features in the standardization of the writing of structural formulas, the recourse to compositional formulations in the writing of structural formulas: for instance notations such as C_6H_5 or C_6H_4 refer unambiguously to presence of a benzene ring in a formula.

The concept of a formula thus aptly summed-up by Remsen is of a chemical "word" (more on that aspect later) which, somehow, represents the sum total of a set of chemical transformations. The chemical formula is close to being an analogon of the bottom-line in a financial recapitulation from an accountant. But, if indeed it is the sum total of the transformations undergone by a molecule, does it have predictive value: is it at the same time the sum total of all the other reactions, yet to be actualized, a particular molecule is capable of? Such a conceptual leap, viz. reading in a formula both the real and the virtual, with the effect of virtualizing the real and of prognosticating the realizing of the virtual, slowly came to pass.

An epistemic switch indeed took place between 1854 and 1858: the formula transformed from a retrodictive device to a predictive tool. A key ingredient was the gradual integration of isomer count as a mapping instrument. The Kekulé benzene formula was swiftly adopted by the chemical industry which very greatly helped its acceptance by the academic community, and this was another key ingredient.

Molecular formulas have become an important part of the language of chemistry but they do not constitute all of it. Truly, they are ideograms in that they may be interpreted directly as pictures or can be named.[10] I can look at a pair of fused six-membered rings, one of which bears oxygen atoms, and the word "naphtoquinone" jumps to mind. Such a duality of pictures and names is essential to the language of chemistry. It obeys Peirce's second trichotomy, which delineates iconic, indexical and symbolic relationships between signs and objects.

The concept of semiotic iconicity,[11] which posits resemblance, formal or actual, between sign and object is relevant here. Was the iconicity of molecular formulas a necessity or a contingence? This is an interesting question for the historian and for the philosopher. It may well have been a fortuitous occurrence. Whatever the case, its existence tends to dehistoricize the devising and the development of molecular formulas, on the one hand, and those of the attendant chemical nomenclature, on the other.

2. HISTORICAL CONTINUITY

I have drawn attention already to the remarkable endurance of formulas for organic molecules. They look remarkably alike at the turns of the twentieth and of the twenty-first centuries. Yet, so many changes befell and nourished chemistry during the intervening twentieth century. The main such change, relevant to our topic, is that structural elucidation underwent during the 1930s and the 1940s a major revamping of its methods, in the process jettisoning "wet chemistry" (chemical reactions performed at the bench) in favor of a gamut of physical methods, consisting at first of molar refraction, parachor, measured dipole moments, followed later on by X-ray diffraction and by the various spectroscopies: infrared, uv-visible, nuclear magnetic resonance, mass spectrometry, circular dichroism, optical rotatory dispersion, etc.

Why did paper representations of the structural formula remain near-invariant during that whole period, when the methodology for determining the structure of an organic molecule underwent such revolutionary changes? The answer to this question involves at least these three factors: the force of convention, already alluded to in the previous section; substantialism, i.e., the implicit notion of a molecule embodying somewhat magically the qualities of the corresponding substance; and the cumulative nature of the program for chemistry, as it was set in the eighteenth century, at the time when Venel wrote his entry "Chymie" for the *Encyclopédie,* and at the time of Lavoisier.

Structural formulas are written in a highly arbitrary manner, subject to conventions: the student of organic chemistry is first exposed to the paradigmatic teaching of nomenclature, he or she is apprenticed into the correct writing of structural formulas at the same time when basic laboratory skills are inculcated to the novice, such as distillation, melting point determination, running a Grignard reaction, etc. Such an apprenticeship also changed surprisingly little for the duration of the twentieth century. Hence, there was a need for fixing the conventions in structural representation very early on, because otherwise the language of chemistry would quickly become an unwieldy tool for communication, simply because of the cornucopia of organic molecules to be thus represented. Structural formulas were set early on, already before World War I, into such a *lingua franca* format, enforced by international conferences and committees, as well as by the house style of publications such as first and foremost the *Beilstein* and the *Chemical Abstracts.*

Substantialism was (and may remain to this day) a mythical belief in a full continuity between perceived qualities of substances at the macroscopic scale (such as color, smell, aspect of the crystals, biological activity, ...) and at the microscopic scale of the molecule. Such an act of faith into the homogeneity of chemical matter, at whatever scale of spatial co-ordinates from the centimeter to the nanometer, went hand in hand with naïve realism regarding the molecules of chemistry, to be considered as molecular objects, no different from those in the ordinary, macroscopic world: objects with a given shape, sets of tiny balls connected with springs as in the sub-discipline of the aptly termed molecular mechanics. Substantialism was part and parcel of the ideology which claimed semi-complete autonomy of chemistry from physics.

The third factor responsible for the near-invariance of structural representation of organic molecules for the duration of the twentieth century has been the perceived program of natural-products chemistry. As it was conceived at the time of Gabriel Venel, this was to be a natural history of substances extracted from plants predominantly. Each such substance had to be isolated, characterized with physical data, named, and related to kindred substances. Elucidation of its structure was tacked on to the list of its physical characteristics, when it started becoming feasible: it made it easier to thus distinguish a substance S, identified by its structural formula, from the manifold of its innumerable isomers T, U, V, W, X, Y, Z, ...

But is it really the case that structural formulas remained invariant throughout the twentieth century? Of course not. Two significant changes, however minute and discreet they may appear, were the assumption of formulas from their initial textual assimilation to iconic status; and their spatial orientation on the printed page. We shall come to the other, typographical changes in representation at a later stage.

A very important point to note is that there is significant evidence for structural formulas to have continued being considered by quite a few chemists, at least during the period 1865–1905, as an integral part of the text of a chemical publication. The typographer would do his/her best to set the formula along the line of text it belonged with. To many organic chemists, the formula was an integral part of the sentence it was included in. Close scrutiny of publications of this period reveals, as the proverbial smoking gun, that very often a structural formula is followed indeed by a punctuation mark, such as a comma or a period (dot, or full stop), as befits a word in a text: *the formula was then word-like*. Only later, much later, did it wrap itself into blank space and started being released from textual into iconic status.

Such typographic conventions endured during the whole period under consideration. One finds them, whether in an article by Emil Fischer and Otto Fischer in 1879,[12] or as late as 1903 in the contribution to the first synthesis of indigo by Bamberger and Elger.[13]

Such a move was of crucial importance. Textual and iconic registers differ markedly. The former partakes of authoritative discourse, the latter provides illustrative value. The former makes the structural formula into a narrative element, the latter presents the structural formula as a piece of evidence for consideration by the reader. The typographic change is symptomatic of the transition which chemical publications undergo at the turn of the twentieth century, when figures and images in the chemical paper start becoming enframed and captioned, from the earlier natural historical narrative to the latter brief (almost in the legal sense) presented in front of the court of opinion of the chemical community.

Iconic messages differ from linguistic messages in that the former, even more than the latter, are polysemic, admit of several meanings simultaneously. A written, a textual message guides its reader toward its intended meaning: hence, to follow Roland Barthes, two of the anchoring functions of a textual message are identification and interpretation.[14] By stepping into the void in-between blocks of text, the structural formula would divest itself of such indexing, deictic functions of

http://www.springer.com/978-1-4020-0100-0

Tools and Modes of Representation in the Laboratory Sciences
Klein, U. (Ed.)
2001, XVI, 264 p., Hardcover
ISBN: 978-1-4020-0100-0